MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-790

Semantic Alignment between ICD-11 and SNOMED CT

Jean-Marie Rodrigues[a,b], David Robinson[c], Vincenzo Della Mea[d], James Campbell[e], Alan Rector[f], Stefan Schulz[g], Hazel Brear[h], Bedirhan Üstün[i] , Kent Spackman[c] , Christopher G. Chute[j] , Jane Millar[c], Harold Solbrig[k], Kristina Brand Persson[l]

[a] *INSERM U1142, LIMICS, Paris, France*
[b] *Department of Public Health and Medical Informatics, Univ. Jean Monnet of Saint Etienne, France*
[c] *Intl Health Terminology Standards Development Organization, Copenhagen Denmark*
[d] *Department of Mathematics and Computer Science, University of Udine, Italy*
[e] *Department of Internal Medicine University of Nebraska Medical Center Omaha USA*
[f] *University of Manchester, United Kingdom*
[g] *IMI, Medical University of Graz, Austria*
[h] *Health and Social Care Information Centre. United Kingdom*
[i] *World Health Organization, Geneva, Switzerland*
[j] *Division of General Internal Medicine, Johns Hopkins University, Baltimore, MD, 21287, USA*
[k] *Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA*
[l] *National Board of Health and Welfare, Stockholm, Sweden*

Abstract

Due to fundamental differences in design and editorial policies, semantic interoperability between two de facto standard terminologies in the healthcare domain – the International Classification of Diseases (ICD) and SNOMED CT (SCT), requires combining two different approaches: (i) axiom-based, which states logically what is universally true, using an ontology language such as OWL; (ii) rule-based, expressed as queries on the axiom-based knowledge. We present the ICD–SCT harmonization process including: a) a new architecture for ICD-11, b) a protocol for the semantic alignment of ICD and SCT, and c) preliminary results of the alignment applied to more than half the domain currently covered by the draft ICD-11.

Keywords:

ICD, SNOMED CT, Standards, Ontology, Terminology, Classification.

Introduction

The project to achieve semantic alignment between these two standards in the healthcare clinical vocabulary began with an agreement signed in 2010 between the World Health Organization (WHO) and the International Health Terminology Standards Development Organization (IHTSDO). ICD[1], currently published as ICD-10, is the most important worldwide standard for mortality and morbidity statistics. However, it is is also used – in several national modifications and extensions – for health care documentation and billing. The international clinical terminology standard SCT[2,3] has been expanding under the management of the IHTSDO. SCT promises to provide an international standard for codes, terms and formalisms to represent details of the health care process.

The current ICD – SCT alignment efforts occur at a time when clinicians, documentation specialists, epidemiologists, health care administrators and health service researchers identify more and more use cases in which SCT is used in parallel with ICD and local procedure and medication terminology systems. This alignment is driven by requirements for increasing granularity of clinical content to record expanding medical knowledge arising from genomic and related research. To ensure full semantic interoperability between ICD and SCT, a semantic alignment policy was developed which relates ICD classes to rule-based queries depending upon an ICD-11–SCT Common Ontology (CO) [4]. Here we report on the current state of this harmonization effort.

This harmonization requires an innovative architecture for ICD-11 because, in the past, the two standards have been based on different semantics: SCT on axioms that express universal truths (e.g. that all instances of *Thrombosis* affect the vascular system); ICD on rule-based knowledge that introduce class definition (e.g. thrombosis in pregnancy falls into a different class for public health reporting).

Materials and Methods

ICD-11 – SCT Harmonization

In 2007, the WHO launched the revision of ICD[5]. After the agreement between WHO and IHTSDO, a Joint Advisory Group (JAG) was established in 2010. There was consensus within JAG that the harmonization could not simply be a mapping between representational entities (classes and concepts) of both systems. The consensus approach was to base the alignment around a Common Ontology following widely acknowledged principles [6-10].

ICD-11 was designed as a multi-component architecture[4]. The first component is a set of "linearizations" for different uses cases – mortality, morbidity, primary care – that are organised as a single hierarchy with disjoint, exhaustive classes taking origin in previous versions of ICD. A second component, named the Foundation Component (FC), contains all of the ICD-11 classes organised according to new, more flexible principles.

This foundation component has at its core, a model of meaning based on description logic [11], using formalisms and language equivalent to those of the semantic web community deployed in OWL[12] and SNOMED CT. This model was named the ICD – SCT Common Ontology (CO)[4, 13]. Fig. 1 illustrates how the common ontology is related to: (i) SNOMED CT, (ii) ICD-11 linearizations and (iii) contingent

knowledge in the ICD content model, such as diagnostic criteria or therapies, originating with WHO class definitions.

The Common Ontology is a subset of the international release of SNOMED CT (hereafter abbreviated to "SNOMED") expanded and revised for ICD convergence. The Common Ontology has been harmonized with ICD text definitions supplied by the WHO. The CO drew primarily from SNOMED *Clinical Findings* hierarchy, which includes findings, disorders and diseases. The CO has minor components from other SNOMED hierarchies including *Situations*, *Events* and *Social context* and will have defining attributes taken from *Body Structure, Organisms, Physical agents* and others.

JAG had concluded during convergence discussions that these concepts denote clinical situations, i.e. phases of a patient's life, in which a given condition of clinical relevance is present [14].

The ICD class definitions and metadata were assembled using the ICD URI API[15, 16]. SNOMED normal forms and definitions were provided by IHTSDO from the 2015-01-30 international release. The two terminologies were lexically mapped and managed with an Equivalence Table (ET), which was the worksheet for semantic analysis as described below. A Sequel Pro API was used by the IHTSDO to interface with a DL classified developmental version of SNOMED.

This ET contained stated normal forms of pre-coordinated concepts as well as proposed additions to SNOMED. Referential quality assurance rules ensured consistency between chapters and tracked changes to SNOMED across developmental releases.

The architecture of the system was built around a web-accessible MySQL ET data base that could be fed with Excel files or SQL. The database could generate output in any of these modes or as an OWL file [17]. The database is synchronised with IHTSDO equivalence matching tools using a customized exchange format. In this database, we used a double browser (ICD-11/SNOMED) with graphical interface connected by equivalences links.

The web application was able to maintain multiple equivalences, recorded by author, in order to also study inter-observer agreement in equivalence identification. When testing semantic alignment required reclassification of the common ontology, we exported an OWL version to Protégé [18] for description logic classification and comparison of inheritance.

Methods for semantic alignment

1 For a defined subset of ICD beta foundation hierarchy (roughly equivalent to a chapter in ICD-10), generate a candidate map from ICD-11 classes to concepts in "Clinical findings", "Situations", "Events" or "Social context" branches of the SNOMED hierarchy. To identify the map, consider the SNOMED fully specified name (FSN), ICD short text definition, the SNOMED logical definition, and the SNOMED Short Normal Form. (Class M, Table1)

2 For ICD-11 classes without corresponding SNOMED content, mark as Unmatched (U). Develop when possible a candidate pre-coordinated SNOMED concept node to be added to core. Use the new SNOMED concept's normal form as the Common Ontology (CO) concept. (class U/A see Table 1)

3 If ICD class is too complex for a single or pre-coordinated SNOMED concept, try to express the ICD 11 class as a Boolean Logical expression within the constraints of SNOMED model of meaning. Identify the expression as the CO entry. (class U/E, Table 1)

4 Bypass ICD-11 residual classes (NEC) but check if there is a broader match (Parent) (Class U/R, Table 1)

5 If none of the above is possible, propose added SNOMED attributes (U/X, Table 1) or new attribute values (U/EX, Table 1) to create the CO concept.

Table 1– Types of match of ICD Common Ontology concepts to SNOMED CT (SCT)

Match Type & Meaning	**Action in SNOMED**	**Common Ontology Axiom**
Match (M)	–	SCT Short Normal Form
Unmatched/A (U/A)	Add appropriate pre-coordinated concept to SCT	New SCT precoordinated Normal Form
Unmatched/E (U/E)	Post-coordinated expression without change in the model of meaning	SCT post-coordinated Logical expressions
Unmatched/R (U/R)		None
Unmatched/X (U/X)	Potential to add with change to SCT model of meaning	Discussion with IHTSDO
Unmatched /EX	Potential to add with change to content model-object/value	Discussion with IHTSDO

Table 1 summarizes the different types of SNOMED (SCT) candidate matches to Common Ontology

6 For each pair of ICD-11 class/subclass and SNOMED concept in the equivalence table: a) check the WHO short text definition for content, consistency and meaning; b) check the semantics of the SNOMED concept or expression including FSN and description logic (DL) definition (short normal form) to assess the alignment of the meanings of the ICD and SNOMED definitions; c) flag all discrepancies and send them to the WHO/IHTSDO interdisciplinary team for:

a) modification of ICD-11 text definition by a Joint Advisory Group definitions workgroup, or

b) changes to SNOMED description logic definition by SNOMED editors

7 For revisions to SNOMED concept definitions, recompile the DL classification of the edited SNOMED content including expressions (U/E). From the reclassified SNOMED, enumerate the set of all subsumed SNOMED concepts corresponding to each equivalent ICD-11 class and assure that the subsumed set has a one-to-one match within the set of subsumed SNOMED mapped concepts and expressions. Identify discrepancies between the subsumed sets.

8 Evaluate the discrepant class/concept pairs for FSN and logic definitions and determine the root cause of the mismatch. Is this is a misalignment of the ICD-11 subclass with the definition of the ICD-11 class or a difference in concept or attribute definition in SNOMED?

Submit the discrepancies to an interdisciplinary team to identify the root cause and propose:

a) JAG workgroup to revise ICD-11 class definition

b) IHTSDO Editorial staff to revise or expand SNOMED content

9 After resolution of all issues in this chapter of ICD-11, return to step 1 for analysis of next domain.

Results

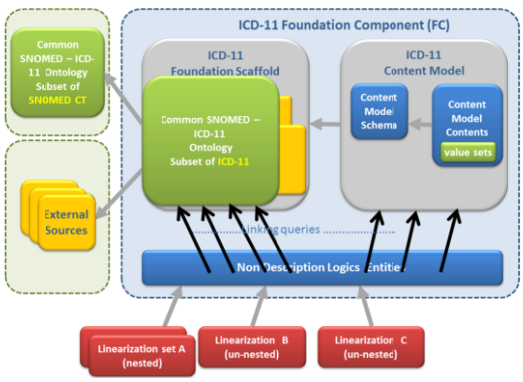

Figure 1 – Architecture of the new ICD-11 architecture and its relationship to SNOMED CT

Updated architecture

Figure 1 is a graphical representation of the new ICD-11 architecture. The outer light blue box represents the Foundation Component (FC), which contains all of the ICD-11 classes or entities, regardless of source. The two smaller boxes on the left represent the external sources for concepts, primarily SNOMED. The Common Ontology is shown in green, a subset of SNOMED participating in ICD-11.

Linearizations are shown in red below the Foundation Component. Each code in each linearization points to a "non-DL entity" in the Foundation Component (FC) whose meaning in terms of the Ontology Component is expressed in a linking query referencing the Ontology Component.

Additional facts about the entities in the Foundation Component are contained in the Content Model. These might include signs and symptoms of diseases, indications and contraindications of drugs among many. These facts are "contingent" (or in traditional philosophical language "particular") rather than universal. That is, they are usually true, or true under certain conditions, but not necessarily "true by definition".

Semantic alignment

The semantic alignment process is ongoing; to date: 16 751 classes/subclasses of the ICD-11 Foundation component have been studied from approximately 39,000 entities (42,9%) .

Table 2 summarizes the percentage of each ICD-11 chapter that has been analysed.

Tables 3, 4 and 5 provide examples of semantic alignment activities to date.

Table 6 presents the summary results of match types for the portion of ICD-11 evaluated thus far.

Table 2– Semantic alignment by ICD-11 chapter (percentage on 01-01-2015)

Chapter	(%)
Infectious diseases	90%
Diseases of the blood and blood-forming organs	30%
Endocrine, nutritional and metabolic diseases	55%
Diseases of the ear and mastoid process	60%
Diseases of the circulatory system	100%
Diseases of the respiratory system	100%
Diseases of the digestive system	100%
Diseases of the skin	80%
Diseases of the musculoskeletal system and connective tissue	90%
Diseases of the genitourinary system	100%
Pregnancy, childbirth and the puerperium	100%
Certain conditions originating in the perinatal and neonatal period	100%

Table 3– Exact match examples

ICD11 Rubric	Common Ontology (Short Normal Form)
Cerebral venous thrombosis	95464572001|Disease (disorder)|:{116676008|Associated morphology (attribute)|=396339007|Thrombus (morphologic abnormality)|363698007|Finding site (attribute)|=68351006|Structure of cerebral vein (body structure)|}
Coronary vaso spastic disease with angina	194828000|Angina (disorder)|+23687008|Coronary artery spasm (disorder)|:{363698007|Finding site (attribute)|=74281007|Myocardium structure (body structure)|}

Discussion

The overall results show that among 16,751 ICD-11 FC entities 14,348 (85.5%) (Table 6) can be represented by the SCT model of meaning either directly or through a pre-coordinated/post–coordinated alignment between the two systems. Very few are residuals which have to be cleaned from the FC or need a revision of SCT formal model of meaning. We propose three examples (Tables 3, 4 and 5) to show how the two systems share the same universal knowledge and how they differ in their contingent knowledge:

1. The defining relationships of SNOMED allow the full representation that can be used for both the ICD-11 Foundation Component class and the SNOMED concept. For example, the ICD Foundation Component entity, "Coronary vaso-spastic disease with angina", is necessarily and sufficiently defined by SNOMED "87343002 | Prinzmetal angina (disorder)".
2. The SNOMED defining relationships do not provide a complete ontological representation of both entities in a single SNOMED concept, in which case the concept model or pre-coordinated SNOMED content must be

Table 4– Pre coordination examples

ICD-11 Rubric	Common Ontology
Aldosterone-producing carcinoma	FSN Primary hyperaldosteronism due to aldosterone-secreting malignant neoplasm of adrenal gland (disorder) Short Normal Form 116680003 | Is a |88213004 | Hyperaldosteronism, |42752001 | Due to|255035007 | Adrenal carcinoma|
Acute myocardial infarction, STEMI, anterior wall	FSN Acute ST segment elevation myocardial infarction of anterior wall(disorder) Short Normal Form 401303003 | Acute ST segment elevation myocardial infarction | + 54329005 | Acute anterior myocardial infarction

Table 5– Post coordination examples

ICD-11 Rubric	Common Ontology
Asymptomatic stenosis of extracranial carotid artery	116680003 | Is a |230738008 | Asymptomatic cerebrovascular disease, |363698007 | Finding site|17999001 | Structure of cervical portion of internal carotid artery, |116676008 | Associated morphology|415582006 | Stenosis
Internal auditory artery occlusion	116680003 | Is a |2929001 | Occlusion of artery, |363698007 | Finding site|89471000 | Structure of labyrinthine artery

expanded. An Example is ICD-11 "Acute myocardial infarction, STEMI anterior wall" that can only be represented in SNOMED by pre-coordinating "401303003 |Acute ST segment elevation myocardial infarction (disorder)" with "54329005 | Acute anterior myocardial infarction".

Conclusion

The essence of the ICD-11 SCT semantic alignment is the establishment of a SNOMED subset with its logical or model of meaning representation that precisely formalizes the meaning of the content of the ICD-11 Foundation Component, following principles of formal ontology and logic *i.e.* that is restricted to axioms that express universal truths in terms of SNOMED concepts. This is clearly distinguished from the ICD content model on the one hand, which represents contingent knowledge at the level of Foundation Component entities, and the rules base (Fig. 1, "non-DL entities"), that cannot be expressed directly in the SNOMED compositional grammar (or any similar logical formalism) and which contains queries on the common ontology that assure the disjointness principle in the linearizations created out of them.

Thus, all content of ICD-11, the semantic standard for health statistics in mortality, morbidity, primary care documentation and billing, will be linked to SCT, the most fine grained medical terminology system, each of which keeps its own profile as a distinct terminology artifact.

Table 6– Match types overall results

Match Type	Number	Common Ontology
Match M	8354 (49.8%)	SCT Short Normal Form
U/A	4933 (29.4%)	To be developped with SCT grammar and pre-coordination
U/E	1061 (6.3%)	To be developped with SCT grammar and post coordination
U/R	1487 (8.8%)	Navigational/residual concepts
U/X and U/EX	916 (5.4%)	Requires clarification

This will require certain refinement and redesign efforts increasing the quality on both ICD-11 and SNOMED, but this is an advantage in itself. When finished, users will have at their disposal two semantically interoperable terminology systems, each tuned for its specific purposes. In the longer term, sharing the maintenance between WHO and the SNOMED authority, IHTSDO, will ease the introduction of new knowledge sources into the heathcare community.

Further on, this common ontology shall be used for the maintenance of all of the existing WHO ICD as well as the ICD-(10/11) national modifications, thereby easing international comparisons and backward compatibility with current systems.

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).

References

[1] The World Health Organization, The International Classification of Diseases 11th Revision is due by 2017, http://www.who.int/classifications/icd/revision/en/ (last accessed 15 Dec 2014).

[2] International Health Terminology Standards Development Organisation (IHTSDO): SNOMED CT: http://www.ihtsdo.org/snomed-ct/ (last accessed 16 Dec 2014).

[3] SNOMED CT browser: https:/uts.nlm.nih.gov/ (last accessed 15 Dec 2014)

[4] J.M.Rodrigues, S.Schulz, A.Rector, K.Spackman, T.B.Üstün, C.G.Chute, V.Della Mea, J.Millar, K.Brand Persson, Sharing Ontology between ICD-11 and SNOMED CT will enable seamless re-use and semantic interoperability. Studies in Health Technology and Informatics 2013;192:343-346.

[5] J.M.Rodrigues, A.Kumar, C.Bousquet, B.Trombert, Using the CEN/ISO standard for categorial structure to harmonise the development of WHO international terminologies, Studies in Health Technology and Informatics 150 (2009), 255-259.

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

[7] M.Ashburner et al. Gene Ontology: tool for the unification of biology. Nature Genetics 25(2000), 25:29.

[8] OBO foundry principles: http://obofoundry.org/ontologies.shtml (last accessed 5 Oct 2014).

[9] A.Rector, A.Rossi, M.F. Consorti, P.Zanstra, Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation, International Journal of Medical Informatics 48 (1998), 71-84.

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

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

[12]W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview 2009, http://www.w3.org/TR/owl2-overview/ (last accessed 28 Oct 2014).

[13]Rodrigues, J. M., Schulz, S., Rector, A., Spackman, K., Millar, J., Campbell, J., & Persson, K. B. (2013). ICD-11 and SNOMED CT Common Ontology: Circulatory System. Studies in health technology and informatics, 205, 1043-1047.

[14]S.Schulz, A.Rector, J.M.Rodrigues, K.Spackman, Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012, 819-827.

[15]The World Health Organization, ICD-11 beta draft: http://www.who.int/classifications/icd11 (last accessed 28 Apr 2014).

[16]G.Jiang, H. Solbrig, C.G. Chute, "Building Standardized Semantic Web RESTful Services to Support ICD-11 Revision", Proc of SWAT4LS, 2012, CEUR-WS 952.

[17]V.Della Mea, O.Vuattolo, B.Ustun ,A.Rector, S.Schulz, J.M.Rodrigues,K.Spackman, D.Robinson, J.Millar, J.R.Campbell,C.G.Chute, H.Solbrig, K.B.Persson, A web-based tool for supporting the development of a Common Ontology between ICD11 and SNOMED-CT. Proceedings of 2nd IEEE International Conference on Healthcare Infomatics (ICHI2014), 144-148, IEEE, 2014

Address for correspondence

RODRIGUES Jean-Marie <rodrigues@univ-st-etienne.fr>