

SNOMED CT's RF2: Is the Future Bright?

Werner CEUSTERS^{a,1}

^a*New York State Center of Excellence in Bioinformatics & Life Sciences, Buffalo, USA*

Abstract. SNOMED CT's new RF2 format is said to come with features for better configuration management of the SNOMED vocabulary, thereby accommodating evolving requirements without the need for further fundamental change in the foreseeable future. Although the available documentation is not yet convincing enough to support this claim, the newly introduced Model Component hierarchy and associated reference set mechanism seem to hold real promise of being able to deal successfully with a number of ontological issues that have been discussed in the recent literature. Backed up by a study of the old and new format and of the relevant literature and documentation, three recommendations are presented that would free SNOMED CT from use-mention confusions, unclear referencing of real-world entities and uninformative reasons for change in a way that does not force SNOMED CT to take a specific philosophical or ontological position.

Keywords. SNOMED CT, RF2, change management, meaning

1. Introduction

SNOMED CT is a clinical reference terminology designed to enable electronic clinical decision support, disease screening and enhanced patient safety. It was first released in 2002 following the merger of SNOMED-RT and Clinical Terms Version 3. In 2010, the International Health Terminology Standards Development Organization (IHTSDO) announced the future distribution of SNOMED CT under a new format called 'RF2' [1] of which more detail became officially available with the January 2011 version [2-4]. The RF2 format is claimed to offer greater flexibility and more explicit and comprehensive version control than RF1 with new features for configuration management thereby accommodating evolving requirements without a need for further fundamental change in the foreseeable future [4]. One such feature is that RF2, through the introduction of a new hierarchy called the '*SNOMED CT Model Component*' [2] which includes the existing Concept Model, allows SNOMED CT to be described in terms of its own structure thereby reducing, so it is hoped, the burden and costs incurred by content developers, implementers and release centers while at the same time improving product functionality and quality. The current documentation of RF2 is marked by a focus on making language- and realm extensions as well as mappings towards other terminologies more manageable. It introduces in addition a number of merely cosmetic changes to the existing history mechanism. But at first sight, it seems also to hold much promises to deal with a number of issues concerning the ontological underpinnings of SNOMED CT that have been reported upon in the literature such as,

¹ Corresponding Author: Werner Ceusters. Ontology Research Group, New York State Center of Excellence in Bioinformatics & Life Sciences, University at Buffalo, 701 Ellicott street, Buffalo NY 14204, USA; E-mail: ceusters@buffalo.edu.

for example, the underspecification of reasons for change [5], the (in)adequacy of SNOMED's intensional and extensional definitions [6], its still incoherent ontological commitment [7], and the ambiguities and conflation in its conceptual structures and in its treatment of terms proposed as 'synonyms' [8]. The goal of the work reported on here was to assess whether RF2 represents an opportunity to resolve these issues whether immediately or in the foreseeable future.

2. Methods

SNOMED CT's documentation and its Concept Model as reflected in the Linkage Attributes were studied for all releases from January 2002 to July 2010. To assess the evolution of the Concept Model, we generated from the relationship tables included in each version a graph representing the relationships actually used in linking conceptIDs from one hierarchy to conceptIDs from the same or another hierarchy, thereby keeping track in each version of the number of times a specific relation, e.g. 'USING DEVICE' was used in relation to the status, e.g. 'current', 'ambiguous', etc., between specific hierarchies. As an example, the relationship 'Computerized tomography guided biopsy of brain (procedure) → METHOD → Biopsy – action (qualifier value)' in version V would increment the occurrence count of the 5-tuple 'procedure – (0) → METHOD → qualifier value – (0)' for version V where '0' indicates the status 'current'. For each tuple, 10 examples of relationships for further inspection – specifically those that revealed astonishing results such as 'substance (2) → SAME AS → procedure (0)' – were selected to find commonalities in the underlying causes for error and of assessing to what extent they relate to the issues described in the introduction. Finally, the new Model Component hierarchy was investigated to see whether it could be expanded with additional entries capable of either solving the issues, or if not, making them explicit.

3. Results: Three Recommendations

The data upon which our analysis and recommendations are based can be downloaded from [9]. They indicate that many problems can be traced back to underlying causes: (1) a mixing of object and meta-language and use-mention confusions, (2) unclarity about what some conceptIDs exactly denote, and (3) use of ambiguous and uninformative codes for the reasons why concepts are inactivated.

Unfortunately, the documentation of RF2 is not yet explanatory enough and lacks clearly worked out examples to assess for each issue identified whether it can be resolved by merely introducing new Model Component entries and associated data types or whether other measures are required as well. Our first – and by far not exhaustive – proposal is therefore formulated in terms of the following *three recommendations* which experts in RF2 can then implement more adequately in the new format they have designed:

1. do not make double use of the ConceptID as an identifier for the concept and an identifier for the Concept Component;
2. add to each Concept Component a field that indicates to what broad category the intended referent of that concept belongs;
3. expand the Concept Inactivation Value sub-hierarchy with concepts that reference whether a change in SNOMED CT is motivated by (1) a change in

reality, (2) the SNOMED CT authors' or users' understanding of reality as reflected in the advance of the state of the art in the biomedical domain, or (3) a mistake that is strictly internal in SNOMED CT as an information artifact [10].

4. Discussion

SNOMED CT is in its Technical Reference Guide described as '*a concept-based terminology which means that each medical concept is uniquely identified and can have multiple descriptions*'. Readers are further told that '*concepts are related to each other by hierarchical relationships*' and that '*relationships are also defined to describe additional attributes of concepts*' [11].

Until the January 2010 version, SNOMED CT's authors defined a concept as '*a clinical idea to which a unique ConceptId has been assigned*' thereby further specifying that '*each Concept is represented by a row in the Concepts Table*' [12]. In 2010, in line with earlier critiques about the ambiguities concept-based systems in general suffer from [13], the glossary of the Technical Reference Guide marks the word '*Concept*' as '*an ambiguous term. Depending on the context, it may refer to: a clinical idea to which a unique ConceptId has been assigned; the ConceptId itself, which is the key of the Concepts Table (in this case it is less ambiguous to use the term "concept code"); the real-world referent(s) of the ConceptId, that is, the class of entities in reality which the ConceptId represents (in this case it is less ambiguous to use the term "meaning" or "code meaning")*' [14]. However, merely pointing this out, however true it might be, does not yet solve the problem. For one could still read in the same document, for example, that a SNOMED CT term is '*a text string that represents the Concept*'. So what is it then that is represented by a term: (1) the clinical idea, (2) less likely, but nevertheless in line with the expressed ambiguity – the ConceptId, or (3) the real-world referent(s)? The same question must then be asked for the several hundred occurrences of the word '*concept*' throughout the SNOMED CT documentation. In some cases, readers can infer from the context which meaning is intended, but in most cases, only the SNOMED CT authors can provide the answer by rewriting the entire documentation.

Unfortunately, as inspection reveals, it is very hard for readers and even for SNOMED CT authors, to disambiguate on the basis of the minimal context provided in sentences in which the word '*concept*' appears between *concept as clinical idea* and *concept as meaning*, i.e. *as real-world referent*. This is not only because clinical ideas are real-world entities themselves – although of a different nature than, for example, persons, viruses and surgical procedures, and some being such that they are about other real-world entities while others are about nothing at all [8] – but also because SNOMED CT authors have not yet made it clear what sorts of real-world entities their concepts represent: denoting real-world entities unambiguously requires ontological commitment and it has been shown that SNOMED CT is incoherent in this respect [7].

Relying on '*meaning*' unfortunately doesn't help much. According to SNOMED CT's glossary definition for '*concept*' discussed above, the meaning of a concept(Id) would correspond to what Frege referred to as the '*Bedeutung*' ('*reference*', '*extension*') of a term [15]. However, in the User Guide, it is specified that '*a "concept" is a clinical meaning identified by a unique numeric identifier (ConceptId) that never changes. The concepts are formally defined in terms of their relationships with other concepts. These logical definitions give explicit meaning which a computer*

can process and query on' [16]. Here, the word 'meaning' corresponds rather to Frege's 'Sinn' ('sense', 'intension') [15]. And finally, in the SNOMED-CT Editorial Guide, a document that became part of the official documentation only since the latest release (although parts of it existed earlier in the form of drafts for comments), SNOMED CT is described as a 'terminological resource' which 'consists of codes representing meanings expressed as terms, with interrelationships between the codes to provide enhanced representation of the meanings' [17]. As a result, the reader is not only left with the question what sort of meaning is discussed each time the word 'meaning' is used – the Editorial Guide is indeed more about 'meanings' than 'concepts' – but also what actually is represented in SNOMED-CT: (1) clinical ideas – in people's minds or concretized in writings, software programs and presentations, respectively called L2 and L3-entities in [8], (2) a broader group of real-world referents that includes not only tangible entities such as patients and knives but also the processes in which the latter participate and the forces they undergo, or (3) 'meanings'.

Without a clear answer to these questions, an answer that might be different for each individual occurrence of the word, SNOMED CT users will make interpretations in different ways, thereby rendering their data mutually incompatible. It will be difficult also to grasp, yes, the meaning of statements such as 'The meaning of a Concept does **not** change [emphasis added]', when immediately followed by the sentence '***If the Concept's meaning changes*** because it is found to be ambiguous, redundant or otherwise incorrect, the Concept is made inactive [emphasis added]' [11]. For the same reason, probably, it has escaped the attention of the SNOMED CT authors that relationships of the sort 'event → MAY BE → navigational concept', 'person → MOVED TO → namespace concept' and, indeed 'physical object → IS A → inactive concept' do not have the same sort of meaning as 'procedure → METHOD → physical object' [9]. The former are statements about the concepts as representational units in SNOMED CT itself (i.e. meta-language statements), while the latter is a statement about the referents of these concepts (an object-language statement). The problem arises because SNOMED CT does not assign, in contrast to entries in the Description and Relationships Table, a separate component ID to an entry in the Concept Table.

5. Conclusion

The three recommendations, despite being very modest, address the issues sufficiently. The first solves the object-/metalanguage confusion. The second solves the problem of what sort of entity in each individual case is referenced by a conceptId. Potential values for the proposed field can be based not only on the L1/L2/L3 distinction [8] – roughly: first-order entities that are not about anything (e.g. person, scalpel) / beliefs, desires, intentions whether about something (e.g. a diagnosis) or about nothing (e.g. some psychotic beliefs) / and information artifacts such as staging scales, guidelines, and, indeed, SNOMED CT itself – but also on whether a universal or defined class is referenced [18], and potentially even on the putative 'possibilia' and 'non-existing entities' [19] endorsed by terminology and ontology developers who do not wish to be hampered by the complexity of Ontological Realism [20]. By doing so, SNOMED CT can even maintain a philosophically rather neutral position even though a clear shift towards OBO Foundry compatibility is observable. And finally, the rather ad hoc motivation for inactivating concepts is catered for by our third recommendation.

Acknowledgements

The work described was funded in part by grant R21LM009824 from the National Library of Medicine. The content of this paper is solely the responsibility of the author and does not necessarily represent the official views of the NLM or the NIH.

References

- [1] International Health Terminology Standards Development Organisation. SNOMED Clinical Terms® Technology Preview Guide - January 2010 International Release (US English) 2010.
- [2] International Health Terminology Standards Development Organisation. SNOMED CT® Release Format 2.0 Reference Set Specifications - Version 1.0a (January 2011 International Release) 2011.
- [3] International Health Terminology Standards Development Organisation. SNOMED Clinical Terms® Release Format 2.0 Data Structures Specification - Version 1.0a (January 2011 International Release) 2011.
- [4] International Health Terminology Standards Development Organisation. SNOMED CT® Release Format 2.0 Guide for Updating from RF1 to RF2 - Version 1.0a (January 2011 International Release) 2011.
- [5] Ceusters W, Spackman KA, Smith B, editors. Would SNOMED CT benefit from Realism-Based Ontology Evolution? American Medical Informatics Association 2007 Annual Symposium Proceedings, Biomedical and Health Informatics: From Foundations to Applications to Policy; 2007 November 10-14; Chicago IL: American Medical Informatics Association.
- [6] Mougín F, Bodenreider O, Burgun A. Looking for Anemia (and Other Disorders) in SNOMED CT: Comparison of Three Approaches and Practical Implications. AMIA Annual Symposium Proceedings. 2010;527-31.
- [7] Schulz S, Cornet R. SNOMED CT's Ontological Commitment. In: Smith B, editor. ICBO: International Conference on Biomedical Ontology. Buffalo NY: National Center for Ontological Research; 2009. p. 55-8.
- [8] Ceusters W, Smith B. A Unified Framework for Biomedical Terminologies and Ontologies. In: Safran C, Marin H, Reti S, editors. Proceedings of the 13th World Congress on Medical and Health Informatics (Medinfo 2010), Cape Town, South Africa, 12-15 September 2010. Amsterdam: IOS Press; 2010. p. 1050-4.
- [9] Ceusters W. Additional Data for MIE2011; www.referent-tracking.com/CeustersMIE2011AddData.zip. 2011.
- [10] Ceusters W. Applying Evolutionary Terminology Auditing to SNOMED CT. American Medical Informatics Association 2010 Annual Symposium (AMIA 2010) Proceedings. Washington DC 2010. p. 96-100.
- [11] International Health Terminology Standards Development Organisation. SNOMED CT® Technical Reference Guide - January 2011 International Release - (US English) 2011.
- [12] The International Health Terminology Standards Development Organisation. SNOMED CT® Technical Reference Guide – July 2009 International Release 2009.
- [13] Smith B. Beyond concepts: ontology as reality representation. Proceedings of the third international conference on formal ontology in information systems. Amsterdam: IOS Press; 2004. p. 73-84.
- [14] International Health Terminology Standards Development Organisation. SNOMED CT® Technical Reference Guide - July 2010 International Release (US English) 2010.
- [15] Frege G. Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik. 1892;100:25-50.
- [16] International Health Terminology Standards Development Organisation. SNOMED Clinical Terms® User Guide - January 2011 International Release - (US English) 2011.
- [17] International Health Terminology Standards Development Organisation. SNOMED CT® Editorial Guide - January 2011 International Release - (US English) 2011.
- [18] Smith B, Ceusters W. Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies. Applied Ontology. 2010;5(3-4):139-88.
- [19] Ceusters W, Elkin P, Smith B. Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach. International Journal of Medical Informatics. 2007 March;76:326-33.
- [20] Lord P, Stevens R. Adding a Little Reality to Building Ontologies for Biology. Plos ONE. 2010;5(9):e12258.