

SNOMED CT Implementation: Implications of Choosing Clinical Findings or Observable Entities

Anne Randorff RASMUSSEN^{a,1}, Kirstine ROSENBECK^a

^a*Department of Health Science and Technology, Medical Informatics,
Aalborg University, Denmark*

Abstract. Internationally, it is a priority to develop and implement semantically interoperable health information systems.[1] One required technology is the use of standardised clinical terminologies. The terminology, SNOMED CT, has shown superior coverage compared to other terminologies in multiple clinical fields. The aim of this paper is to analyse SNOMED CT implementation in an Electronic Health Record (EHR). More specifically, differences and consequences of applying clinical findings (CFs) as an alternative to observable entities (OEs) is analysed. Results show that CFs represents the content of the templates with better coverage, with more parent concepts and with a higher degree of fully defined terms than the OEs. We discuss the possibility to further evaluate the observable entity hierarchy to overcome a potential overlapping use of the two hierarchies.

Keywords. Clinical terminology, Implementation, SNOMED CT, Observable entity, Clinical finding, Electronic Health Record

1. Introduction

Multiple definitions of identical concepts are a challenge in data communication in health care. Use of standardised clinical terminologies has the potential to ensure unambiguous data definition. This is a prerequisite in achieving semantic interoperability between health information systems. There exists numerous clinical terminologies, but SNOMED CT has shown to be superior regarding coverage in multiple clinical fields.[2,3] Therefore, SNOMED CT is chosen as the point of departure in this study.

SNOMED CT is maintained and refined by the International Health Terminology Standardisation Organisation (IHTSDO). The organisation has published strategies and rules for the implementation of SNOMED CT to unify future implementation [4]. However, these are mostly theoretical as only few SNOMED CT implementation projects are documented.[5] Inevitably, there will be deviations between the way SNOMED CT is implemented in real-life projects and the theoretical recommendations. These deviations are important to report, since they increase knowledge on possible implementation strategies for SNOMED CT. To support this, Alan Rector has argued that the goal of clinical terminologies is implementation in clinical information systems. In addition, he doubted that all terms currently part of SNOMED CT was actually

¹ Anne Randorff Rasmussen, Fr. Bajers Vej 7 C2-, DK-9220 Aalborg Ø, arra@hst.aau.dk.

operational: “It is a significant clinical task to find out what situations the term is intended to cover which might actually be recorded in an operational record”.[6]

This study is based on implementation of SNOMED CT in an EHR-system in the Northern Jutland Region in Denmark. The terminology is implemented alongside the configuration the EHR-system. Our point of departure is two locally designed clinical templates “nursing status” and “physical examination”. As they are clinical notes, a structured narrative approach was chosen. Structured narratives combine the familiarity, ease of use and freedom of expression of the narrative with the ability to browse data based on the gross structure represented by sections, fields and paragraphs.[7]

In the specifications provided by IHTSDO, it is stated that the OE hierarchy in SNOMED CT should be used for coding sections, fields and paragraphs. “Concepts in this hierarchy can be thought of as representing a question or procedure which can produce an answer or a result.”² However, when mapping expressions from the respective templates to SNOMED CT, a lack of quality and comprehensiveness was found in the OE hierarchy. The aim of this paper is to systematically analyse the implications of applying CFs as an alternative to OEs when configuring the “nursing status” and “physical examination” templates.

2. Method

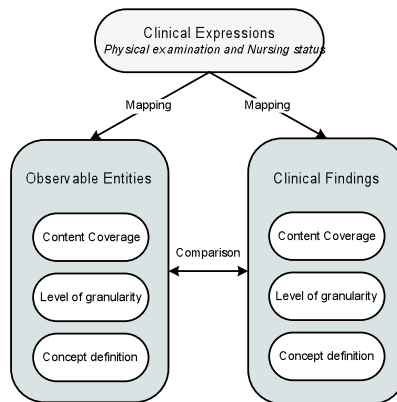


Figure 1. Overview of the method applied to compare OEs and CF in this study

The method applied is illustrated in Figure 1. Templates that represent two clinical domains are included in this study to achieve expressions with varied characteristics. The data set consists of a total of 34 clinical expressions: Physical examination (22 expressions) and Nursing status (12 expressions). 8 cases of compounded terms exist in the data set, e.g. 'skin and mucosa finding', and 'respiration and circulation'. 7 of these are found in 'Nursing status'. The clinical expressions were mapped to SNOMED CT OEs and CFs respectively. When mapping the compounded terms we initially strived to find a pre-coordinated concept covering both expressions, otherwise post-coordination by combination is used to represent the expressions.

The analysis framework was developed to systematically evaluate the usefulness of a set of SNOMED CT concepts. In the research literature there are rather few

² <http://www.ihtsdo.org/snomed-ct/snomed-ct0/snomed-ct-hierarchies/observable-entity/#c1513>

methods for analysing SNOMED CT. An exception is [8] where an information-content measure is developed. This measure is based on the analysis of the parents, pathways and branches of SNOMED CT. However, we want to analyze retrieval and reuse potential, therefore the exact measure of [8] is not applicable, but our approach is similarly based on these core-characteristics of SNOMED CT.

In the analysis, we compared and assessed the potential of each hierarchy to represent the clinical expressions. The analysis is conducted within the following areas; content coverage, level of granularity and concept definition. These are described in details below. The *content coverage* is analyzed to assess whether concepts in SNOMED CT are able to represent clinical expressions. Also the use of pre- and post-coordinated concepts is stated.

The *level of granularity* is examined, defined as the level of detail associated with each concept. Hence, the number of parent concepts is measured, as shown in Figure 2. This measure is chosen, as it expresses the potential of the concept to be used for data retrieval purposes, as search strategies can be based on either one of the parents or the concept itself. The parents make it possible to retrieve data based on a more granular level based on inherited meaning only. A Wilcoxon signed-rank test is performed to assess whether there is a significant difference between the number of parent concepts in the CF and OE hierarchy. This test is done contrary to a paired t-test, as we cannot assume normal distribution.

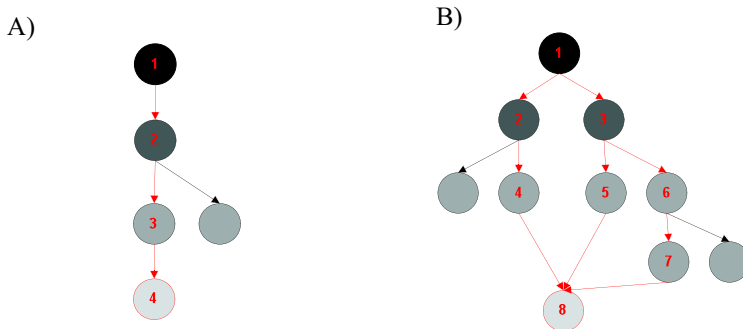


Figure 2 A) 4 parent nodes and B) 8 parent nodes. Identical parent concepts are only included once

The *concept definition* is examined, defined as whether the concept is primitive or fully defined. A concept is primitive when its logic definition does not sufficiently express its meaning. Further, primitive concepts do not have the defining relationships needed to computably distinguish them from their parent or sibling concepts.[9] For fully defined concepts, aggregated data can be based on characteristics that are stated by other expressions than the inherited meaning.

3. Results

The results of analysing the content coverage, level of granularity and concept definition is presented in following tables;

The *content coverage* is shown in Table 1. A coverage of 100% for the CF and 94 % for OE is achieved. Post-coordination is used more frequently to represent the clinical expression in OE than in CF and more pre-coordinated concepts of the compounded expressions was found for the CF hierarchy than in the OE. In CF the

compounded expressions 'Skin and mucous membranes' and 'sleep and rest' exist. In the OE hierarchy these expressions exist as separate concepts only.

The tables (Table 2a and Table 2b) show results of analysing *level of granularity* and *concept definition* for each expression.

Table 1. SNOMED CT coverage for the clinical expressions in nursing status and physical examination.

Nursing Status and Physical Examination	Pre-Coordination	Post-Coordination	Total
CF	31 (91%)	3 (9%)	34/34 (100%)
OE	25 (74%)	7 (20%)	32/34 (94%)

Table 2a Results assessing the number of parents (NoP) and definition for nursing status.

Clinical expressions	functional perf. and activity	Nutrition	Defecation	Micturition	Respiration	Circulation	Skin	Mucosa	Pain / sensation	Sleep (and rest)	Rest	Psychosocial	Cognitive funct.	Communication	Value belief	Sexuality	Reproduction	NoP		Def	
																		OE	CF	OE	CF
	5	5	16	16	4	4	4	4	4	7	7	5	5	4	5	7	4	5	5	1	1
	5	5	7	12	4	4	6	4	5	6	0	7	7	4	6	5	5	6	5	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	0	1	1	0	0	0	0	1	1	-	1	0	1	1	1	1	1	1	1	1

Table 2b Results assessing the number of parents (NoP) and definition for nursing status.

Clinical expressions	General	Physics	Mental state	Head and neck	Skull	Eye/Vision	Oral cavity	Ear	Nose	Neck	Endocrine	Lymphoid system	Truncus	Cardiac ausc.	Respiratory ausc.	Breast	Abdominal	Urogenital	Rectum	Beck structure	Limb	Neurological	Skin	NoP		Def	
																								OE	CF	OE	CF
	4	5	4	5	7	4	4	6	8	6	5	5	4	0	0	4	6	4	6	0	5	4	4	4	4	4	
	4	5	6	5	9	4	10	9	8	6	4	7	4	13	10	7	6	7	14	5	5	3	6	6	6		
	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	1	1	1	1	-	1	1	1	1	1		
	1	1	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0		

The Wilcoxon signed-rank test shows a significant difference in the number of parent concepts for the two hierarchies with $p=0.031$. The average number of parents for the concepts in OE is 5.15 and for CF 6.33, looking at the concepts for the physical examination only, the difference in number of parents increase. This means that for the whole dataset and especially for the physical examination, the CF hierarchy has more granulated concepts than the OE hierarchy. A similar difference is obtained when comparing the level of definition for each concept. All OEs are primitive, whereas 56% of CF is fully defined. Also, it is observed that the proportion of fully defined concepts is higher if we look at the physical examination alone.

4. Discussion

In this study a comparison was performed by mapping concepts from two clinical templates to concepts from the OE and the CF hierarchies. The aim was to investigate whether the CF contribute with a higher quality and comprehensiveness than the OEs.

Existing literature lacks focus on the usage of the specific SNOMED CT hierarchies. The main objective in the scientific literature is to investigate the potential of SNOMED CT to cover the content of different clinical domains.[5]

The results show that the needed concepts can be found in both the OE hierarchy (94%) and CF hierarchy (100%), which potentially can induce ambiguous encoding. The problem is not merely redundant concepts, but that two hierarchies can be used interchangeably. IHTSDO suggest that each hierarchy has a certain purpose, but our study and a study by Lee et al. suggest that the stated purposes are not clear enough to allow consistent mapping. [10] Lee et al. find overlaps between e.g. the “clinical finding” and “morphologic abnormality” hierarchies when mapping a palliative care dataset. To keep the mapping consistent, they introduce guidelines. However, local guidelines cannot handle terminology inconsistencies between organisations. Therefore, in time, improving the consistency of SNOMED CT itself would be preferable.

Improving the consistency to avoid redundant use of hierarchies is not a simple task. In our study, it is suggested that the parameters: content coverage, granularity and definition might be useful in determining which hierarchy reaps most benefits in terms of retrieval and reuse purposes. Using these parameters and the “nursing status” and “physical examination” datasets, the CF hierarchy is superior to the OE hierarchy. However, these are only two clinical examples and even among these there are differences in the results. To the authors knowledge similar studies examining the same dataset using two different SNOMED CT hierarchies is not available. Therefore, more studies are needed on the implementation of SNOMED CT with the focus of analysing the usage of the hierarchies.

Acknowledgement. This research is part of our PhD-studies that are co-financed by Region Northern Jutland, CSC Scandihealth and Trifork A/S.

References

- [1] Garde S, Knaup P, Hovenga EJS, Heard S. Towards Semantic Interoperability for Electronic Health Records. *Methods Inf.Med.* 2007;3:332.
- [2] Brown SH, Rosenbloom ST, Bauer BA, et al. Direct Comparison of MEDCIN® and SNOMED CT® for Representation of a General Medical Evaluation Template. *AMIA.Annu.Symp.Proc.* 2007:75.
- [3] Chute CG, Cohn SP, Campell KE, Oliver DE, Campell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. *JAMIA* 1996;3(3):224.
- [4] International Health Terminology Standards Development Organisation. *IHTSDO*. Available at: <http://www.ihtsdo.org/>. Accessed 11/18, 2009.
- [5] Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008 Oct 27;8 Suppl 1:S2.
- [6] Rector AL. Clinical terminology: why is it so hard? *Methods Inf Med* 1999 Dec;38(4-5):239-252.
- [7] Johnson SB, Bakken S, Dine D. An Electronic Health Record Based on Structured Narrative. *Journal of the American Medical Informatics Association :JAMIA* 2010 21:54.
- [8] Cornet R. Information-content-based measures for the structure of terminological systems and for data recorded using these systems. *Stud.Health Technol.Inform* 2010:1075.
- [9] IHTSDO. SNOMED Clinical Terms. *User Guide*. 2010 January.
- [10] Lee DH, Lau FY, Quan H. A method for encoding clinical datasets with SNOMED CT. *BMC Med Inform Decis Mak* 2010, 10:53