eJHI electronic Journal of Health Informatics
www.eJHI.net
2012; Vol 7(1):e9

Agreeing on Meaning: A Fundamental of Sharing Health Information

Michael J Lawley, David P Hansen, Madonna Kemp, and Deanne M Vickers

The Australian e-Health Research Centre, CSIRO ICT Centre, Brisbane, Qld, Australia

Abstract

Topic: *A preliminary study on the reproducibility of results when mapping terms from an existing terminology to SNOMED CT post-coordinated expressions is described.* **Background:** *Implementing SNOMED CT requires a strategy for migrating existing systems and data that currently use other terminologies as well as ensuring that SNOMED CT contains suitable content that covers the domain. Mapping terms from these terminologies to SNOMED CT is one element of such a strategy. Snapper is a tool designed to assist in this complex task and enable the creation of quality mappings.* **Methods:** *Ten terms from the ANZICS diagnosis codes were selected for mapping according to specified guidelines. The resulting mapping expressions were compared with each other and discussions were conducted with the mapping participants to determine issues they encountered during the process.* **Results:** *Consistency was easily achievable with mapping to single concepts, but was more difficult when mapping to post-coordinated expressions. The difficulties were traced to a lack of specificity in the supplied guidelines resulting in uncertainty in structuring the representation of compound concepts.*

Keywords: Clinical Terminology; Systematized Nomenclature of Medicine

1 Introduction

SNOMED CT is being introduced as the lingua franca of health information in Australia [1]. As a clinical terminology with a formal semantics based on description-logic, it could provide a perfect underpinning for the implementation of clinical decision support which is both precise and can cope with the increasing knowledge base of the medical community. With so much data already collected using existing terminologies (formal or ad-hoc), the Australian E-Health Research Centre has developed the *Snapper* tool for, among other things, creating mappings from an existing terminology to SNOMED CT, and therefore allowing for existing clinical data to be treated as if it was collected in SNOMED CT.

In this paper we extend the work begun with an evaluation of Snapper itself [2], and describe a preliminary comparison of how different experts map an existing terminology to SNOMED CT. In this case the authors of the paper each mapped a very small subset (10 terms) from the Australia New Zealand Intensive Care diagnosis codes [3] to SNOMED CT. This mapping was carried out in mid-2009 using the then-current version of Snapper. Various aspects of the tool are described followed by a comparison of the resulting mappings. These are then used to inform a more general discussion of the task of mapping itself.

2 Background

Mapping or linking existing terminologies to SNOMED CT is traditionally an arduous and time consuming exercise [4, 5], requiring a relatively high level of knowledge of the clinical domain and a high level knowledge of the structure, content and concept model guidelines of SNOMED CT. The CSIRO Australian E-Health Research Centre has developed Snapper, which is a software tool to support such mapping or linking and has been specifically designed with the non-expert in mind.

With its detailed underlying concept model, documented in the User Guide [4], the detailed definition of SNOMED CT concepts can be a daunting task for those attempting to map an existing terminology to post-coordinated expressions. Indeed, the sheer size of SNOMED CT itself and the quantity of written documentation can inhibit their ability to produce semantically correct content suitable for clinical use [2]. To mitigate this, the Snapper tool provides search and browse facilities and a configurable *automap* feature to help users start mapping from an existing terminology to SNOMED CT. Advanced features include a concept model constraint checker, for determining conformance of an expression with the SNOMED CT Concept Model and provide feedback on how an incorrect expression may be corrected; a graphical expression view that displays additional inferred relationships for the mapping expression; and the ability to (incrementally) classify the entire mapping for feedback on the positioning of the mapped terms within the SNOMED CT hierarchy.

3 Methods

One author had previously completed a fully modelled mapping of terms from ANZICS to SNOMED CT [2]. Ten source terms were chosen to give a mix of terms that were correctly and incorrectly mapped by the automap feature and terms that would require post-coordination or specialization relationships. Each author was then given these 10 terms to map to SNOMED CT. A brief editorial policy was agreed in advance; that the mapping would be used for secondary purposes, such as counting patients who had had procedures or injuries of particular sites of the body. The policy also laid out the purpose of the original terminology; the codes were used to describe the reason that a patient is in the ICU, rather than to describe any procedures that the patient might have undergone [6].

An important aspect of comparing the resulting mappings is determining whether two expressions have the same meaning. While this is easy in the case of single concepts, it is more difficult – even by eye – in the case of expressions that combine multiple concepts using the SNOMED CT post coordinated expression syntax. Snapper's graphical expression view was used to compare the expressions including inferred relationships.

The authors also tracked how long the mapping took and provided feedback on the tool and their thought processes while performing the mapping and reflecting on the results.

4 Results

All authors used the automap functionality to do a preliminary mapping, with all reporting that two terms automapped to a single concept while two others mapped to concepts close in the hierarchy to the concept eventually mapped. One term mapped to a child of the correct map while another term mapped to a related term in a different part of the hierarchy. The time taken by the authors to perform the mapping ranged from 32 minutes to 1 hour.

In all, the authors mapped five terms identically, all of which mapped to a single concept. The other five concepts required a post-coordinated expression to be built. In many cases they built different expressions. The visualization capabilities of Snapper were used to manually inspect and compare the expressions. Future functionality in Snapper will allow for expressions to be automatically compared for formal equivalence or subsumption using the Snorocket classifier [7].

As an example of the kinds variations found, the four expressions modelled for the case of "*Head/pelvis trauma, surgery for*" are shown in Figure 1 with the subsumption relationship between each pair of expressions shown in Table 1. While all mappings were different, they all contain the detail suitable for answering queries about the focus of injuries or the procedure performed. However, the expressions are to different hierarchies (*Clinical finding* vs *Procedure*), and thus the queries would need to be structured differently. Fortunately, where this conflict exists, the use of the *Clinical finding* hierarchy was rather trivial and thus the relevant *Procedure* content could be easily extracted without a significant change in intended meaning.

The astute reader may also notice that both expressions C and D are invalid with respect to the SNOMED CT Concept Model; the *Due to* relationship for a *Clinical finding* is only permitted to have another *Clinical finding* as its value, and not a *Procedure*. In this instance, the correct relationship should have been either *Associated with* or the more specific *After*. This error was not caught be the constraint checker at the time due to a bug; because *Due to* is a sub-role (specialisation) of *Associated with* relationship was being seen as satisfying the range restriction of *Associated with* even though the range restriction of *Due to* should have taken priority.

5 Discussion

It became clear when comparing the mappings and discussing the motivations behind the different approaches

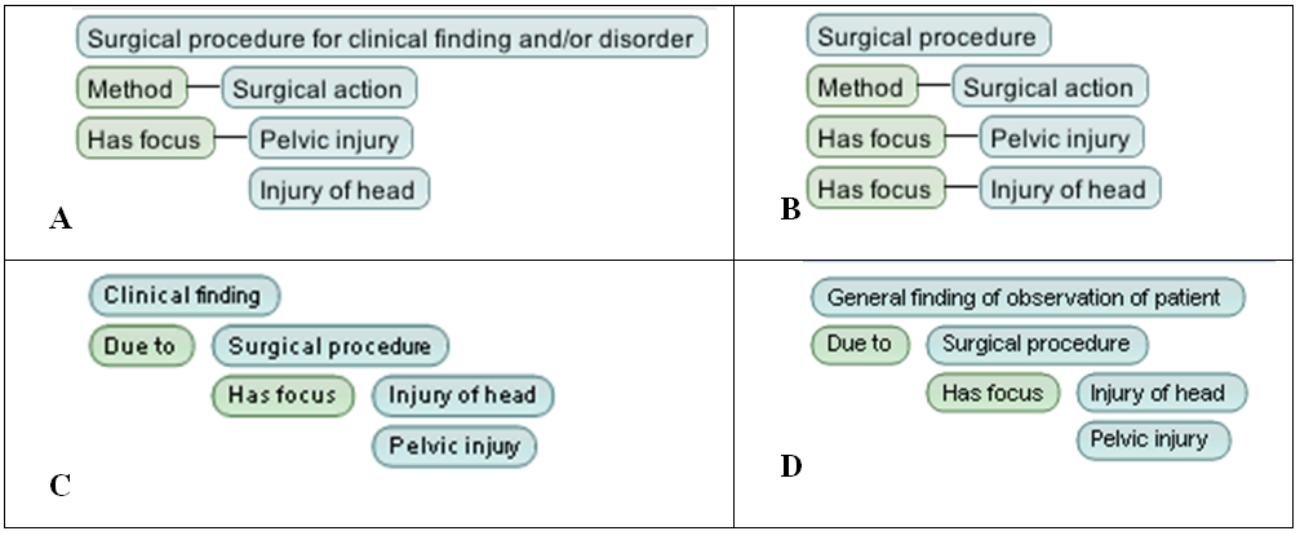

Figure 1: Different modelling of "Head/pelvis trauma, surgery for"

	A	B	C	D
A	≡	<	≠	≠
B	>	≡	≠	≠
C	≠	≠	≡	>
D	≠	≠	<	≡

Table 1: Subsumption relationships between the expressions of Figure 1

with their authors that the editorial guidelines supplied were insufficient. Specifically, because the source terms did not all clearly fall into a single SNOMED CT hierarchy (e.g., *Clinical finding*, *Event* or *Procedure*) and they represented compound things involving multiple aspects, it became difficult to maintain consistency. This was compounded by the lack of clear and specific examples of the intended use of the resulting mapping.

In situations such as this, there are two main approaches: map the terms as accurately as possible to their meaning as intended in the original context, or map the terms sufficiently accurately to answer specific kinds of queries. The implication is that the latter approach may result in some loss of precision, but that this is insignificant in the context of the intended use of the mapping. As noted above, in the case of our mapping case study this distinction was not made clear and this ambiguity was compounded by lack of clarity in kinds of queries that would be asked.

Another aspect that came out in the debriefing discussions was the role of *context* and the applicability or appropriateness of mapping terms explicitly to SNOMED CT's *Situation with explicit context* hierarchy. This hierarchy is used to represent greater contextual detail of findings and procedures. This includes scheduled or not-performed procedures, negative or suspected findings, and other aspects such as historical findings (i.e., patient history), and those associated with someone other than the patient (i.e., family history). The use of a finding or procedure concept in a record without any explicitly stated context is considered to have a "soft-default" context, for example that the finding *is present*, is about *this patient*, and is *current at this time*. Thus, when creating a mapping, context should also be taken into account in case any of the terms to be mapped do not align with the default context. At the least, the default context should be used when comparing the mappings and should be supported by tooling.

For this reason it is felt that the mapping guidelines should at least have specified that all mappings be explicitly fully-contextualised, or treated as if a specified default context has been applied. That is, all mappings would be to the *Situation with explicit context* hierarchy. However, this still leaves some leeway when mapping compound terms such as "*Head/pelvis trauma, surgery for*"; should they be mapped as above with the *Has focus* relationship (Figure 2), or should both the findings and procedure instead be mapped using the *Associated finding* relationship (Figure 3).

To deal with the compound term situation, the guidelines should also identify a model for consistently combining the multiple concepts. This then effectively specifies an information model or a fragment of one. While *Situation with explicit context* does represent a suitable

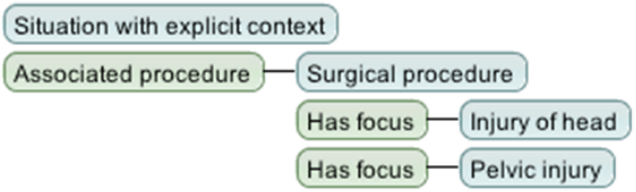

Figure 2: Compound terms mapped with Has focus *relationship.*

Figure 3: Compound terms mapped with Associated finding *relationship.*

example, there is a lack of clear guidelines on the appropriate use of the model in both the IHTSDO's User Guide [4] and Style Guide [8]; specifically, whether it is valid to use both the *Associated procedure* and the *Associated finding* relationships, and when it is appropriate to use *Associated finding* with a *Clinical finding* or other non-*Procedure* value. An alternate approach would be to rely on an external information model such as OpenEHR[1]/CEN 13606[2], CDA[3], or a context-specific one. In this case the mapping would become a one-many mapping and would need to rely on an associated information model binding mechanism to capture its full semantics.

6 Conclusions

This study is a preliminary study for a formal comparison of mappings between experts. Snapper is now licensed to terminology experts around the world. It is planned to repeat this study with a number of these experts, a larger subset of terms from the ANZICS sub diagnosis codes (approximately 60), and to provide more detailed and explicit instructions as to the purpose of the original terms and, following the discussion above, the way in which the mapping should be done when compound terms are encountered.

The results above show the importance of understanding the meaning of both (a) the original terms and (b) the way in which a mapping will be used. We speculate that even given ideal circumstances, where comprehensive and explicit specifications of both use cases are available, achieving a shared and unambiguous understanding of a medical term will be difficult. The key issue identified here is not terminology or tooling performance (both are considered well developed). Rather, our experience reveals the inherent difficulty in successfully negotiating meaning and dealing with multiple and different uses and re-uses of clinical data itself.

Due to the difficulties associated with creating accurate maps, the authors recommend extreme caution if attempting to build and use such a map for clinical purposes such as a decision support system. Rather maps are really only suitable for statistical analysis or as part of a process of migration to and adoption of SNOMED CT.

[1] http://www.openEHR.org
[2] http://www.en13606.org
[3] http://www.hl7.org/implement/standards/cda.cfm

References

1. National E-Health Transition Authority, Available from http://www.nehta.gov.au

2. Vickers DM, Lawley MJ. Mapping Existing Medical Terminologies to SNOMED CT: An Investigation of the Novice User's Experience. In: Sintchenko, Vitali (Editor); Croll, Peter (Editor). HIC 2009: Proceedings; Frontiers of Health Informatics - Redefining Healthcare, National Convention Centre Canberra, 19-21 August 2009. Brunswick East, Vic.: Health Informatics Society of Australia (HISA), 2009: 46-51. http://search.informit.com.au/documentSummary;dn=497370883409153;res=IELHEA

3. ANZICS Data Dictionary, Australia New Zealand Intensive Care Society, http://www.anzics.com.au/uploads/CORE/ANZICS_APD_Data_Dictionary_Version2.3_Sep08.pdf (last accessed 29 June, 2009). Appendix E.

4. The International Health Terminology Standards Development Organisation: SNOMED CT User Guide, International Release, July 2008.

5. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, Elkin PL, "Using SNOMED CT to Represent Two Interface Terminologies", Journal of American Medical Informatics Association. 2009; 16:81–88. DOI 10.1197/jamia.M2694.

6. Personal communication, Cook D, PA Hospital, Annerly, Qld. 2009.

7. Lawley MJ "Exploiting Fast Classification of SNOMED CT for Query and Integration of Health Data", Proceedings of the 3rd International Conference on Knowledge Representation in Medicine (KR-MED) 2008. Cornet R, Spackman KA (Eds)

8. The International Health Terminology Standards Development Organisation: SNOMED CT Style Guide: Situations with Explicit Context, January 2008

9. Lawley MJ, Vickers DM, Ireland D, Hansen DP, "Terminology recycling with the Snapper Platform: SNOMED CT adoption by evolution not revolution", submitted, Journal of Biomedical Semantics

Correspondence

Michael Lawley, PhD

Australian e-Health Research Centre
Level 5
UQ Health Sciences Building 901/16
Royal Brisbane and Women's Hospital
Herston, Queensland 4029
Australia

Phone: +61 7 3253 3609
Fax: +61 7 3253 3690
http://aehrc.com

michael.lawley@csiro.au