electronic Journal of Health Informatics
http://www.ejhi.net
2011; Vol 6(1): e1

Building SNOMED CT Reference Sets for Use as Interface Terminologies

David P Hansen, Marcel Giermanski, Monika Dujmovic, Josh Passenger and Michael J Lawley

Australian E-Health Research Centre, CSIRO ICT Centre, Brisbane, QLD, Australia

Abstract

The SNOMED CT Clinical Terminology has been made available to the Australian health community as the clinical terminology to be used to standardise the collection of clinical data. Reference Sets are a mechanism to identify a specific subset of SNOMED CT concepts relevant to a particular medical domain. They also provide alternate organisation or navigation hierarchies. As such, they attempt to bridge the gap between a general-purpose reference terminology and an interface terminology.

In this paper we describe the process of building a Reference Set that can be used to describe colonic pathologies and how that Reference Set can be used as an interface terminology in a specific software system. Ideally the Reference Set will be generally useful for other applications, such as to support synoptic reporting. This would also help to meet the aim of recording clinical information in a standard vocabulary.

The process described uses the Reference Set Editor in the Snapper Platform from the Australian e-Health Research Centre (AEHRC), which already supports the building of SNOMED CT extensions through a reverse mapping module. Pathology terms from a standard text on performing endoscopies formed the basis of a terminology for describing colonic pathologies. The CSIRO Colonoscopy Simulator at the AEHRC provides an ideal case study for the issues of using Reference Set in specialised health software.

For Australian Health IT providers to successfully implement the vision of recording all clinical information using a single terminology, tool support for the creation of Reference Sets will need to be flexible enough to handle many issues.

Keywords: Clinical Terminology, Health Data Mining, Electronic Health Records

1. Introduction

In pursuing the ideal of a single terminology in which to collect clinical information there are a number of practical implementation issues which need to be overcome. One important issue is providing user interfaces which guide the expert, either clinician or information manager, to capture the correct terms when entering data. The ideal situation would involve health information systems being user friendly enough for data to be collected by the clinician. Again ideally, the clinician should be presented with a context aware list of terms to select from and should be unaware that they are terms from a particular terminology.

The clinical terminology now being introduced in Australia is SNOMED CT, distributed by the National e-Health Transition Authority [1]. SNOMED CT is a large and sophisticated terminology with a formal semantics grounded in description logic. Subsets of the full clinical terminology are known as "Reference Sets", which are documented fully in the SNOMED CT documentation [2]. The documentation describes four different types of Reference Sets, as shown in Figure 1. These include the Simple Reference Set, which is essentially just a collection of concepts. There are then three specialised Reference Set types which allow for particular uses of subsets of clinical terminology terms. Of interest to us are Navigation Reference Sets. These support the intuitive navigation of the concepts.

The *electronic Journal of Health Informatics* is an international journal committed to scholarly excellence and dedicated to the advancement of Health Informatics and information technology in healthcare. ISSN: 1446-4381
© Copyright of articles is retained by authors; originally published in the *electronic Journal of Health Informatics* (http://www.ejhi.net). This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License (http://creativecommons.org/licenses/by-nc-sa/2.5/au).

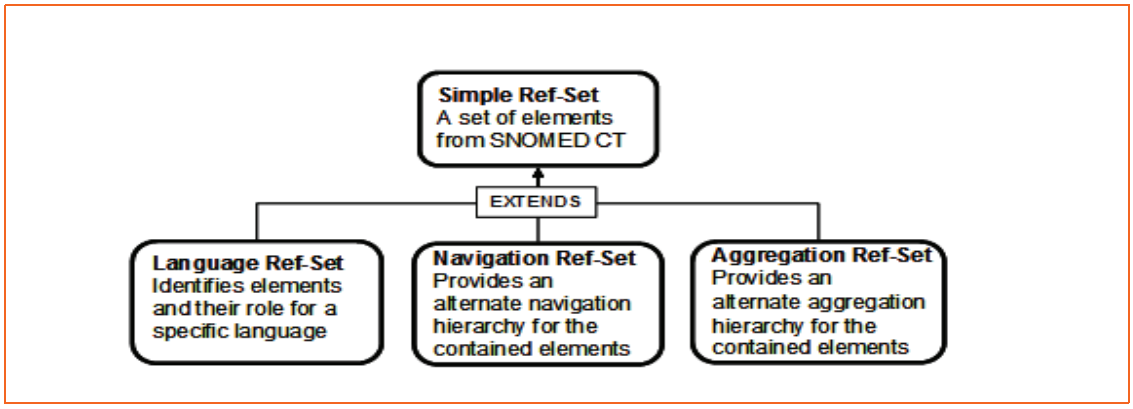

Figure 1: *Reference Set Implementations as Suggested by "SNOMED Clinical Terms Reference Set Specification", by IHTSDO.*

To support the implementation of SNOMED CT in Australia, the Australian E-Health Research Centre (AEHRC) is developing a software platform, the Snapper Platform, which supports various use cases for terminology. The primary use case is Reverse Mapping: the process of mapping concepts in existing terminologies to a base ontology, in this case SNOMED CT. Other components in the platform include search, browse, annotate, drag and drop, which can be plugged together to support a wide range of use cases, including building an extension based on an existing terminology, annotating text and images and building a Reference Set. Current tools for building reference sets tend to be ad-hoc use of browsers and spreadsheets or in-house developments.

The AEHRC also has a terminology server for publishing the extensions and Reference Sets which we describe in this paper. The Snapper Platform and the terminology server both use the snorocket classification engine [3] to maintain the stated and inferred relationships.

This paper describes the Reference Set Editor which has been added as an additional module to the Snapper Platform. A description is then given of the process of building a Reference Set which can be used in other software systems to capture clinical data. In this case the software system is a colonoscopy simulator. Having a simulator use terms from a clinical terminology starts the process of having clinicians get used to those clinical terms as well as performing a colonoscopy.

2. Methods

Figure 2 shows a screen shot of the Snapper Platform with the Reference Set Editor. The right hand pane contains the search functionality while the bottom pane contains a visual representation of the terminology which allows a user to browse through the hierarchy of clinical concepts.

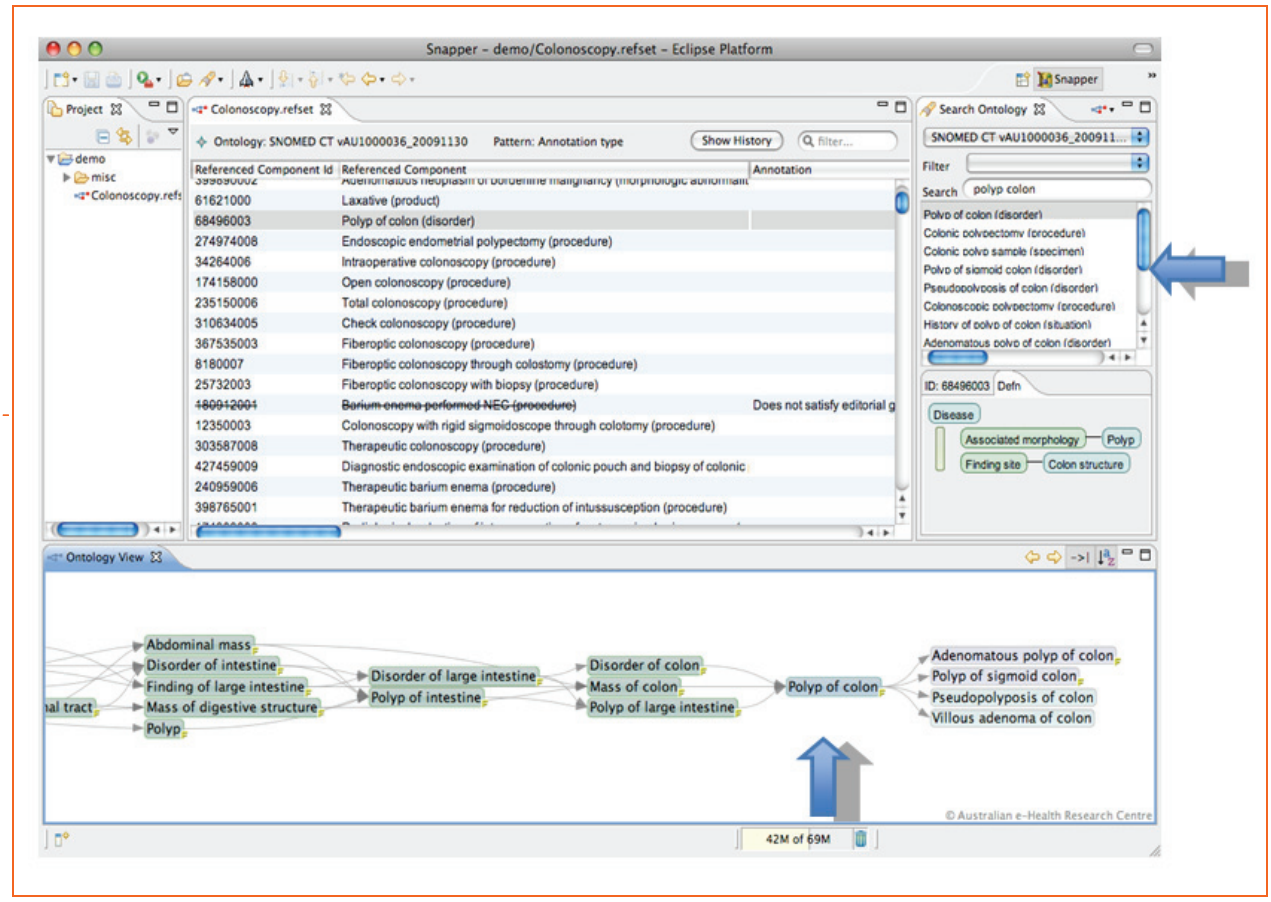

Figure 2: *The Snapper Platform, with Reference Set Editor.*

The left hand pane allows for the creation of a project and associated Reference Sets or Extensions. Once a Reference Set has been created, concepts can be added to the Reference Set by searching or browsing to the appropriate concepts and dragging the concepts on to the Reference Set pane in the middle of the tool. The user can also import the direct children (or specialisations), or all of the descendants of a concept already contained within the Reference Set. This is supported using a right mouse click on the concept and selecting the appropriate option in the pop-up menu.

Reference Sets in Snapper can have one of two types: simple, which is a list of concepts, or navigation, which allows for hierarchies to be created. If child concepts are added to a Navigation Reference Set then they maintain their existing hierarchy in SNOMED CT.

Reference sets are created in groups. That is, a Reference Set Group is initially created within the editor and individual Reference Sets are created therein. Furthermore, a Reference Set Group can be created from an existing mapping, where the mapped terms are used to populate a simple Reference Set, along with their SNOMED CT counterparts.

Navigation Reference Sets allow the user to define a custom hierarchy by dragging and dropping members contained therein; also, the removal of a member from the Reference Set "promotes" all of its children, rather than deleting them, within the hierarchy, i.e. each parent of a removed member acquires the removed member's children as additional children of its own. In addition to these manual actions, the user has at their disposal the ability to automatically create hierarchies based on either the SNOMED CT hierarchy (Top-Down) or the members' Fully Specified Names (Alphabetical) via the "Create Hierarchy" Menu and Toolbar as shown in Figure 3.

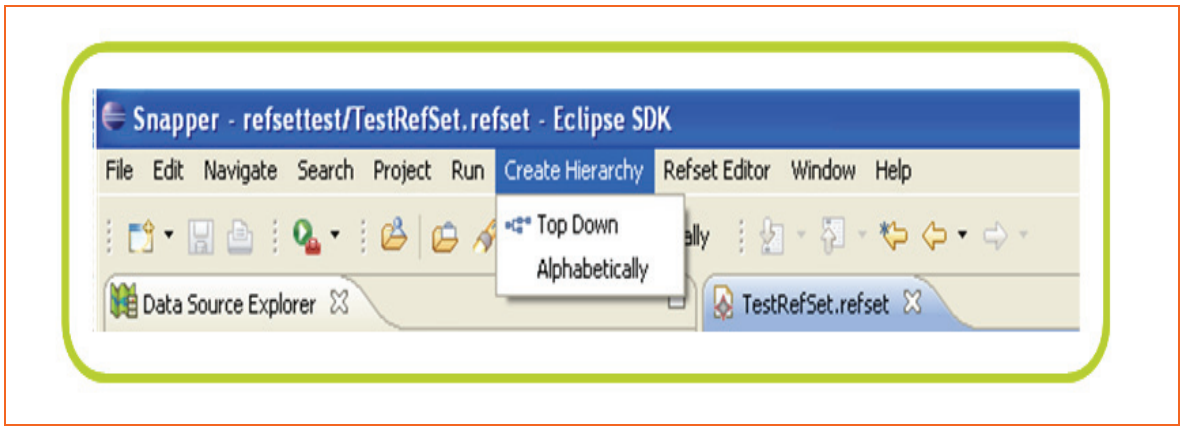

Figure 3: *The Reference Set can follow the SNOMED CT hierarchy, a hierarchy introduced by the user or simply list the terms alphabetically on the Fully Defined Name of the term.*

3. Results

The CSIRO Colonoscopy Simulator project [4] at the AEHRC provides an ideal case study for the issues of using Reference Set in specialised health software. The simulator is aware of the position of the endoscope in the colon. The developers wanted to be able to use that context to display a different set of pathologies to the clinician at the different positions the endoscope is in the colon. The first step in creating the Reference Set was to find the terms which need to be in the Reference Set. Forty-five terms were selected from a standard text on endoscopy to form the initial terminology [5]. The Reverse Mapping functionality of the Snapper Platform was then used to create an extension to SNOMED CT [6]. Not being particularly complex, all 45 terms were able to be mapped to single concepts from the clinical terminology. This is not always the case and the Snapper Platform contains much functionality to support the building of post-coordination expressions to describe complex clinical concepts in the case that a one-to-one mapping is not possible. Additionally, a term can be mapped as a *specialisation* of a concept or post-coordinated expression if it is not possible (or desired) to map to an exact equivalent.

Once this initial mapping was complete, the Reference Set could be built, which in this case was simply a matter of converting the mappings to a Reference Set And the Snapper tool can automatically create a navigation hierarchy based on the SNOMED CT "IS A" hierarchy. Additionally, for a term like *colorectal* which was mapped to the concept *110797007 | Colon and rectum (CS)* |, which sits outside the normal body structure hierarchy of SNOMED CT, it can be manually placed in an appropriate location in the navigation reference set hierarchy.

Figure 4 shows the current pathology capture interface in colonoscopy simulator. The interface the Colonoscopy Simulator required the pathologies to be divided into 'Pathology Type', 'Pathology Location' and 'severity' – concepts which are not terms that currently exist in the terminology. These terms were added to the local terminology and published them as an extension to SNOMED CT in the terminology server. Now that these high level concepts were available, the required concepts could be dragged onto the high level concepts to create the navigational hierarchies which support the required user interface.

Once a final Reference Set was created, it was possible to make it available for potential use in the Colonoscopy Simulator application by either exporting a file or via the terminology server discussed earlier. Accessing it programmatically via the terminology server would allow multiple uses of the same Reference Set and supports the move to fully interoperable clinical information systems.

Now that core clinical terminology has been extended, the new concepts could be used to provide a hierarchy within the Reference Set that can be used to support the interface in future versions of the simulator.

Figure 4: *The Colonoscopy Simulator, with selection of pathologies which can be supported through a Navigation Reference Set.*

4. Discussion

The creation of a Reference Set is not just a matter of making a list of SNOMED CT codes for use in other software. For example, the SNOMED Navigation Reference Set format does not directly allow for specific concepts to be included merely for navigational grouping purposes and not selectable in the user interface (one could use a second, simple, reference set to indicate which concepts are acceptable choices). Other studies [7] have suggested that additional semantics would improve the ability to use SNOMED CT Reference Sets as Interface Terminologies. Building a Reference Set which supports multiple use cases will require the addition of extra concepts and perhaps the addition of new relationships. The case study reported here was relatively simple, but as more clinical information systems move to use the SNOMED CT more complex issues will arise. However, the advantages of moving to a single clinical terminology with a strong underlying description logic are well documented [8]. Tool support for building Reference Sets must also support the addition of new concepts to the core terminology and the creation of new hierarchies to support the use case being implemented.

The IHTSDO has recently changed the proposed specification for how to build nd describe reference sets. Support for Interface Terminologies will need to consider issues such as how to specify leaf nodes which are for navigation only, but shouldn't be able to be selected.

With Reference Sets being an important aspect in building usable information systems, Reference Set Editor will be further developed to enable ease of building the sets and exporting them to other tools. Current developments are the automated building and organisation of Reference Sets based on the occurrence and use of concepts in existing electronic health records.

5. Conclusion

For Australian Health IT providers to successfully implement the vision of recording all clinical information using a single terminology, tool support for the creation of Reference Sets will need to be flexible enough to handle many issues. The IHTSDO Global Health Terminology Workbench is a tool for the building of SNOMED CT content, and includes a complex rule-based Reference Set builder. in contrast, the Snapper Platform provides an

easy-to-use stand-alone Reference Set builder that is designed to be used without requiring a detailed technical knowledge of the inner workings of SNOMED CT.

Acknowledgements

We thank the Scientific Committee of HIC 2009: Frontiers in Health Informatics for selecting this paper for this eJHI special issue.

References

1. National E-Health Transition Authority, Fact sheet: A national approach to clinical terminologies, August 2006, Available from http://www.nehta.gov.au

2. College of American Pathologists, SNOMED Clinical Terms Reference Sets Technical Guide, July 2006.

3. Lawley M, Vickers D, Hansen D. Converting Ad Hoc Terminologies to SNOMED CT Extentions Health Informatics Conference 2008: The Person in the Centre 31 Aug - 02 Sep, pp 133.

4. Hellier D, Passenger J, Appleyard MN, Ourselin S. Development of a new Colonoscopy Skills Trainer using Virtual Reality and Haptic Feedback. In SimTecT 2007 Healthcare Simulation Conference, p.8.

5. Cotton PB, Williams CB. Practical Gastrointestinal Endoscopy: The Fundamentals. Blackwell Publishing, 5th Edition, 2003.

6. Lawley MJ. Exploiting Fast Classification of SNOMED CT for Query and Integration of Health Data In: Cornet R, Spackman K, editors. Representing and sharing knowledge using SNOMED. Proceedings of the 3rd International Conference on Knowledge Representation in Medicine KR-MED 2008, Phoenix (Arizona), USA, May 31-June 2, 2008, p. 8.

7. Rosenbloom, ST, Brown, SH, Froehling, D, Bauer, BA, Wahner-Roedler, DL, Gregg, WM and Elkin, PL : Using SNOMED CT to Represent Two Interface Terminologies, Journal of the American Medical Informatics Association. 2009; 16 (1): 81-92.

8. Rector AL and Brandt S: Why do it the hard way? The case for an expressive description logic for SNOMED. Journal of the American Medical Informatics Association. 2008; 15 (6): 744-751.

Correspondence

Dr. David Hansen
E-Health Research Centre
CSIRO ICT Centre
Royal Brisbane and Womens Hospital
Herston, Qld 4027, Australia

Phone: +61 (0)7 3253 3600
Fax: +61 (0) 7 3253 3690
http://aehrc.com/

David.Hansen@csiro.au