Towards Semantic Interoperability in Healthcare: Ontology Mapping from SNOMED-CT to HL7 version 3

Amanda Ryan

School of Economics and Information Systems
The University of Wollongong,
Northfields Avenue, Wollongong, NSW, 2522,
Australia
Email: ajr883@uow.edu.au

Abstract

One of the most successful Healthcare Information Models is version 2 of the Health Level 7 (HL7) standard. However, this standard has various problems, mainly its lack of semantic interoperability. This shortfall was addressed in HL7 Version 3, a newer standard which has been designed to solve this problem. Total semantic interoperability cannot be achieved without defined terminology, and to this end the use of the Systemised Nomenclature of Medicine - Clinical Terms (SNOMED-CT) is proposed. The difficulty arrives when deciding how to integrate the information model and the terminology. The line between where one ends and the other begins is often indistinct. This paper describes a proposal for normalising the two using ontology mapping and basing HL7 message models on SNOMED-CT concepts and their relationships, in an effort to further total semantic interoperability and seamless communication between healthcare entities.

Keywords: Ontology Mapping, Interoperability, HL7, SNOMED-CT, Health Informatics

1 Introduction

Information in the Healthcare domain is enormously complex, covering many different types of data. Patient administration, organisational information, clinical data and laboratory/pathology data are different but must be compensated. Add to this the integration of all of these areas, and storage in Electronic Health Records. As a result of this diversity and richness of the data, and also due to the fragmented nature of Health Informatics' implementation and research efforts, many different models have been designed to represent information in this field.

One of the most successful of these information models is version 2 of the *Health Level 7* (HL7) standard. However, this standard does not achieve plug-and-play interoperability (Huff, Bidgood, Cimino & Hammond 1998, Klein 2005). In fact, any kind of interoperability between communicating HL7 systems can be time-consuming to establish and require careful attention to detail, especially between existing systems that have not been set up to communicate by design. Also, because HL7 was first developed in 1987 to suit technology at that time, it is in an out-dated format of different fields separated by "pipes" (|) and "carets" (∧), where precision in the number of fields is a must - a single missing pipe could convert an otherwise meaningful message into a meaningless stream of text. The HL7 Organisation says this about the version 2 messaging standard:

> These messages evolved over several years using a "bottom-up" approach that has addressed individual needs through an evolving ad-hoc methodology. There is neither a consistent view of that data that HL7 moves nor that data's relationship to other data (HL7 2006).

As such, no semantic inferences of the information can be determined automatically.

To overcome these problems, Version 3 of the HL7 standard was developed (HL7 2006). This standard incorporates a new paradigm for information representation and messaging in comparison to HL7 version 2 in terms of a new information model and intrinsic extensibility. This newer version is geared towards semantic interoperability between systems and is consistent with the technology developments, purpose and recommendations emerging from the Semantic Web. As such, implementing HL7 version 3 can be seen as a platform for the implementation of modern distributed systems standards, not just in Health Informatics, but as a platform to experiment with emerging distributed systems standards more generally.

The difficulty arrises when trying to represent clinical concepts and constructs within the HL7 framework. One of the biggest obstructions to communication occurs when there are multiple ways of describing a single concept. For example, one person may write potassium as "pot" and another may write potassium as "K", which can make inferring semantics a complex problem. To overcome this, a standard vocabulary is needed to populate the model with meaningful data. If there existed a standard list of codes to represent many different clinical concepts, both people in the previous example could look up the code for potassium and use it, thus not allowing misunderstanding to occur. Huff *et al.* (1998) say "it is only by integrating the structure of a message with the vocabulary sent in the message that unambiguous information exchange between systems can be acheived."

To this end, there has been suggestions of HL7 consulting with domain experts to create standard vocabulary tables. In the same way that there exists many different information models for healthcare, this has already been done many times by domain experts who have created different terminologies for clinical and other healthcare data. Rather than repeating this time-consuming and exhaustive process, an existing terminology created by domain experts could be integrated with the HL7 information model.

In light of emerging Health Informatics standards use in Australia (NEHTA 2006), The terminology chosen to use in this project is the *Systemised Nomenclature of Medicine - Clinical Terms* (SNOMED-CT). SNOMED-CT is a result of a combination by the College of American Patholigists and the U.K. Department of Health of two existing terminologies (*SNOMED RT* and *Clinical Terms Version 3*) to create a unified terminolgy with a greater depth and coverage of healthcare data (SNOMED 2006). SNOMED-CT is widely regarded as the most comprehensive clinical healthcare terminology in the world and is even multilingual, covering both English and Spanish language concepts. The idea of using this terminology within the HL7 information model is not as simple as it sounds, as SNOMED-CT has its own rich information model. This paper outlines a strategy for normalisation between the two models involving mapping of concepts and relationships in SNOMED-CT to classes and attributes in HL7.

2 HL7 version 3

HL7 Version 3 uses an object-oriented development methodology and a Reference Information Model (RIM) to create messages (HL7 2006). The RIM is a UML[1]-style diagram based around six core types of classes and rules governing how they relate to each other. These rules, as well as further restrictions on allowable attributes for each class, make up the information model of HL7 version 3. Cardinality and optionality constraints also exist on relationships between classes and on attributes.

The six core classes of the RIM are:

- Act - an action of interest
- Entity - a class or instance of a specific thing capable of participating in Acts
- Role - An Entity, in a particular Role, can participate in an Act
- Participation - an association between a Role and an Act
- ActRelationship - an association between a pair of Acts
- RoleLink - a connection between two roles expressing a dependency

Three of these classes – Act, Entity and Role – are further represented by a set of specialized classes, or sub-types. E.g. specialisations of the Act class include Observation, Procedure and Substance Administration. As can be seen by the nature of the RIM classes, the HL7 information model takes an act-centred view, with processes and information in healthcare represented primarily in terms of the acts performed within an organisational context(Vizenor 2004).

Acts, Entities and Roles all have an attribute called `code`. This attribute could be populated with a SNOMED-CT code corresponding to the kind of Act, Entity or Role it is. This is a current subject for discussion by the TermInfo Project, sponsored by the HL7 Vocabulary Technical Committee(HL7 2006, SNOMED 2006). This will be taken into account as part of the ontology mapping process.

3 SNOMED-CT

SNOMED-CT is made up of concepts and their attributes, which consist of relationships to other concepts. As such, relationships in SNOMED-CT are modelled as a triple of (concept, attribute, concept). For example, Figure 1 shows the concept "O/E - Blood Pressure Reading 163020007" in the centre with arrows to other concepts showing its attributes. "O/E" stands for *On Examination*, and belongs to the concept group *Clinical Finding*. The number after the name of the concept is the *SNOMED-CT Concept ID* (SCTID). The SNOMED-CT relationships from Figure 1 have been tabulated in Table 1.

Some examples of concept groups in SNOMED-CT are as follows:

- Clinical Findings - the results of a clinical observation, assessment or finding
- Procedures - purposeful activities performed in the provision of health care
- Body Structures - normal and abnormal body structures
- Substances - active chemical constituents of drug products, food, chemical allergens, toxicity information, etc
- Physical Objects - natural and man-made objects
- Events - occurences that result in injury
- Observable Entities - procedures or questions which, when combined with a result, constitute a finding
- Qualifier Values - concepts not contained elsewhere in SNOMED-CT which are required for attributes e.g. open, left, right, etc

These "concept groups" are concepts themselves and are the top-level concepts in SNOMED-CT, as SNOMED-CT is structured as a multiple-inheritance hierarchy of concepts. Top-level concepts are concepts that are the direct children of the root concept. The root concept is the single, topmost concept in SNOMED-CT and is "SNOMED CT Concept" and has the SCTID 138875005.

SNOMED-CT models concepts and their relationships to each other in clinical constructs. As can be seen from the example concept groups, in particular Observable Entities, this can involve information on procedures which do not necessarily have to occur. This is in contrast to the HL7 act-centred view of healthcare information, where a procedure will only be recorded if it is intended to occur, or has occured. The impact of this will be assessed in the mapping process.

[1]Unified Modeling Language - http://www.uml.org/

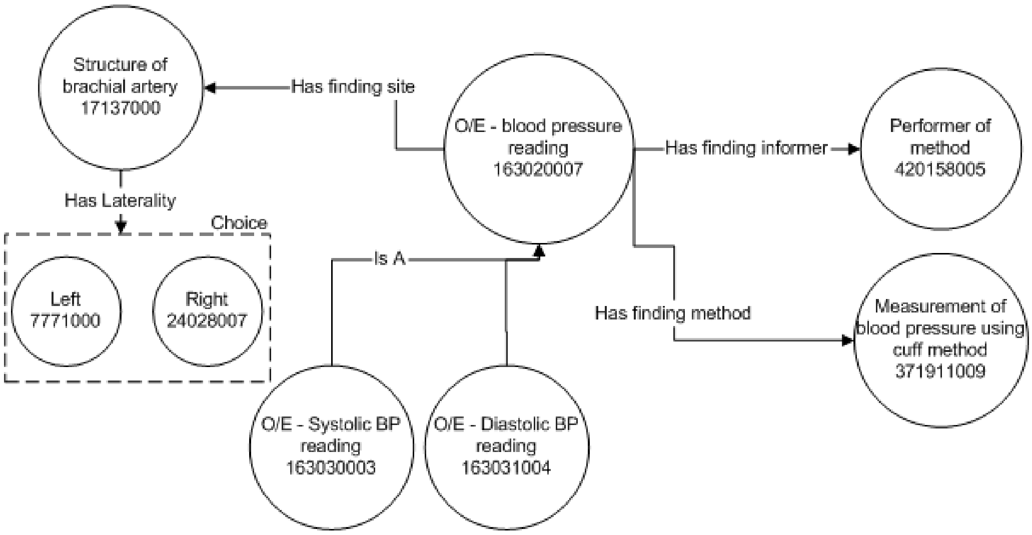

Figure 1: Blood Pressure Reading Concepts and Attributes in SNOMED-CT

Concept	Attribute	Concept
O/E - blood pressure reading	Finding site	Structure of brachial artery
Structure of brachial artery	Laterality	Left/Right
O/E - blood pressure reading	Finding informer	Performer of method
O/E - blood pressure reading	Finding method	Measurement of blood pressure using cuff method
O/E - systolic BP reading	Is a	O/E - blood pressure reading
O/E - diastolic BP reading	Is a	O/E - blood pressure reading

Table 1: SNOMED-CT relationships shown in Figure 1. Relationships in SNOMED-CT are modelled as a triple of (concept, attribute, concept).

4 Ontology Mapping

An ontology O is defined by its set of Concepts C with a corresponding subsumption hierarchy H_C. Relations R exist between single concepts, which also have a corresponding hierarchy H_R (Ehrig & Staab 2004). Both the HL7 and SNOMED-CT information models can then be said to be ontologies.

Ontology mapping takes two ontologies as input and creates a semantic correspondence between the entities in the two input ontologies (Rahm & Bernstein 2001). Ehrig and Staab(2004) define ontology mapping:

> Given two ontologies O_1 and O_2, mapping one ontology onto another means that for each entity (concept C, relation R, or instance I) in ontology O_1, we try to find a corresponding entity, which has the same intended meaning, in ontology O_2.

There are two kinds of conflicts between heterogeneous ontologies (Tang, Liang & Li 2005). The first conflict occurs when ontologies for the same domain knowledge have different semantic structures, which has been shown to occur between HL7 and SNOMED-CT in at least one respect - HL7's act-centred view and SNOMED-CT's main aim of modeling clinical constructs. This is to be expected as the two information models have different aims and purposes.

The second conflict occurs when either the same concept has different names in both ontologies, or the same name refers to different concepts in both ontologies. Both of these situations occur between the HL7 and SNOMED-CT information models. For example, one of the concept groups in SNOMED-CT is 'Event', referring to an occurrence which results in injury. Examples of this are "Accidental Fall", "Flood" and "Motor Vehicle Accident". HL7 also contains an 'Event', but this time it refers to an Act which has occured. This could be that a patient's blood pressure was taken, or that a patient was admitted to hospital. In SNOMED-CT, a patient's blood pressure that has been taken is a Clinical Finding. These situations and conflicts will have to be taken into account during the mapping process.

Because SNOMED-CT models actual clinical constructs, the HL7 models will be based on existing SNOMED-CT models and the mapping will be in the direction of SNOMED-CT $\rightarrow$ HL7.

As an example, the blood pressure reading concept from SNOMED-CT in Figure 1 has been mapped manually to HL7. Figure 1 could be represented in HL7 as shown in Figure 2.

The mapping of SNOMED-CT concepts to their container fields in HL7 for blood pressure reading is shown in Table 2. The first five rows show a relatively straightforward mapping between the two. E.g., the SNOMED-CT attribute `Finding site` is captured within an attribute of a similar name in HL7 (`targetSiteCode`), and the SNOMED-CT concept will become the data to populate this field. The last three rows in the table refer to attributes in SNOMED-CT which are mapped to classes in HL7- `ComponentOf` and `Informant`. This mapping is almost a reversal of the first mappings, with the SNOMED-CT attributes mapped to HL7 classes and the SNOMED-CT concepts mapped to an attribute in a separate HL7 class.

Another example, this time of the cannula insertion procedure, is shown in Figures 3 and 4. This example is more complex than the blood pressure

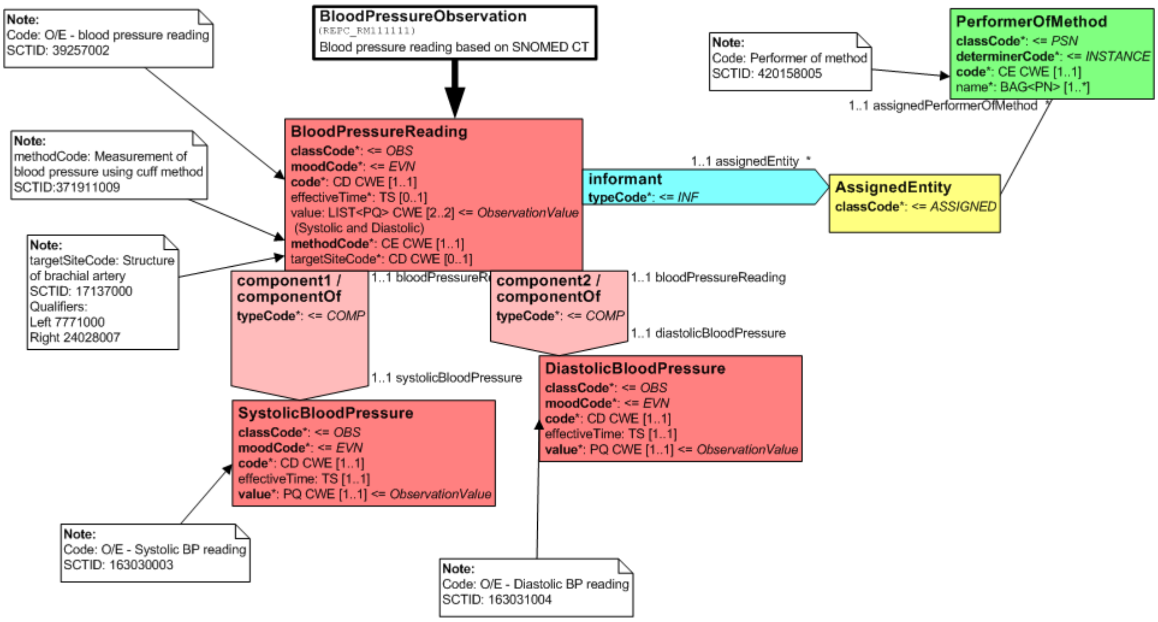

Figure 2: Blood Pressure Concept and its attributes from SNOMED-CT represented in HL7

SNOMED-CT		HL7	
Concept	Attribute	Class	Attribute
O/E - blood pressure reading	n/a	Observation	code
Structure of brachial artery	Finding Site	Observation	targetSiteCode
Left	Finding Site - Laterality	Observation	targetSiteCode
Right	Finding Site - Laterality	Observation	targetSiteCode
Measurement of blood pressure using cuff method	Finding method	Observation	methodCode
Performer of method	Finding Informer	Informant	Person::code
O/E - Systolic BP reading	Is a	ComponentOf	Observation::code
O/E - Diastolic BP reading	Is a	ComponentOf	Observation::code

Table 2: Mapping from SNOMED-CT to HL7 in Figures 1 and 2.

reading example and raises new issues. Blood pressure reading is an observation, and was relatively simple to model. Cannula Insertion is a procedure, and the differences in modelling between an observation and a procedure suggest that different considerations will have to be taken into account for every type of HL7 Act. The mapping of SNOMED-CT concepts to their container fields in HL7 for cannula insertion is shown in Table 3. Note that the mapping in this case is mostly in the form of SNOMED-CT codes fitting into different `code` attributes of the one HL7 Act subclass (procedure). The SNOMED-CT concepts marked with an asterisk (*) denote higher level concepts in the SNOMED-CT hierarchy. In the blood pressure example, all concepts were at the lowest level, so no further choices or specialisations of the concepts could be made. For cannula insertion this was not possible, so the concepts shown in the diagram as "choice" are specialisations of the SNOMED-CT concepts shown in Table 3. This suggests that automatic mapping of this concept to HL7 may not be acheivable, as some decisions will need to be human-made.

As can be seen from these simple examples, careful thought and study of the SNOMED-CT dataset will have to be put into the mapping process. To further complicate matters, HL7 has some vocabulary tables of its own, used in fields such as `classCode` and `moodCode`. One suggestion that has been made in trying to draw the terminology and the message structure together is the inclusion of the HL7 vocabulary terms and concepts as an addition to SNOMED-CT.

5 Conclusion

The HL7 RIM and SNOMED-CT are ontologies which need to be able to work together, so what is needed is a mapping from one ontology to the other. The results of this mapping could facilitate the automatic generation of HL7 messages from the structure of SNOMED-CT's concepts and relationships. If fully automated generation of HL7 messages cannot be achieved, it may still be attainable with only a few human-made decisions, which would be as acceptable in a real world setting.

In trying to achieve this goal, the question of whether HL7 Version 3 can effectively model clinical concepts and their relationships can also be answered.

The ultimate goal of this research is to create a model for combining messaging and terminologies in a seamless way that promotes total semantic interopability between systems and ease-of-use for healthcare systems developers and users.

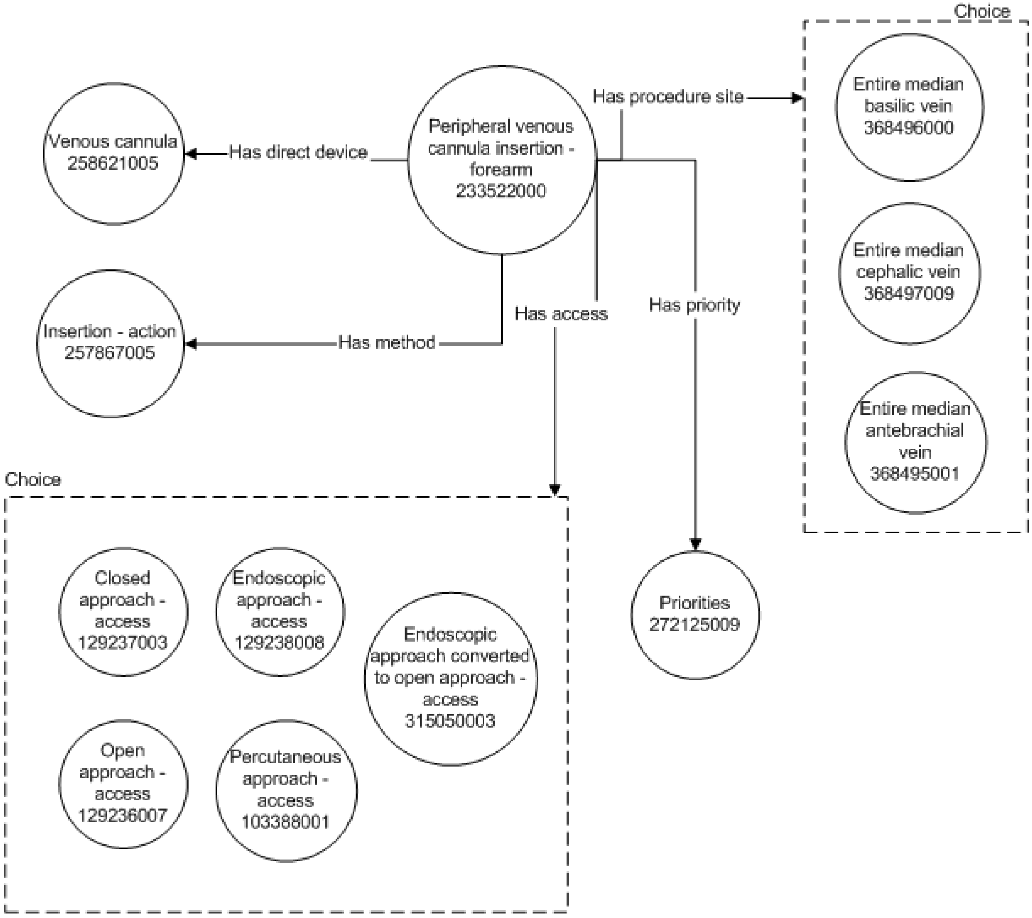

Figure 3: Cannula Insertion Concepts and Attributes in SNOMED-CT

References

Ehrig, M. & Staab, S. (2004), QOM - Quick Ontology Mapping, *in* S.A. McIlraith, ed, 'ISWC 2004', LCNS 3298, ppp. 683-697.

Health Level 7, 'About HL7' & 'HL7 Terminfo', *Health Level 7 webpage*, `http://www.hl7.org/`, last accessed 07/09/06.

Huff, S. M., Bidgood, W. D, Cimino, J. J. & Hammond, W. E. (1998), A Proposal for Incorporating Health Level Seven (HL7) Vocabulary in the UMLS Metathesarus, *in* 'Journal of American Medical Informatics Association, AMIA Annual Fall Symposium Supplement', pp. 800-804.

Klein, J. (2005), Integrating Electonic Health Records using HL7 Clinical Document Architecture, *in* 'Business Integration Journal'.

National E-Health Transition Authority (2006), *National E-Health Standards Development: A Management Framework*, version 1.0.

Rahm, E. and Bernstein, P.A (2001), A survey of approaches to automatic schema matching, *in* 'The VLDB Journal', Vol. 10, pp. 334350.

SNOMED International (2006), 'What is SNOMED-CT' & 'Terminfolink', *SNOMED International webpage*, `http://www.snomed.org/snomedct/what_is.html` & `http://www.snomed.org/Terminfo/Terminfolink.html`, last accessed 17/09/06.

Tang, J., Liang, B.Y. & Li, J.Z., Toward Detecting Mapping Strategies for Ontology Interoperability, *in* 'WWW 2005', May 10 - 14, 2005, Chiba, Japan.

Vizenor, L., Actions in Health Care Organizations: An Ontological Analysis, *in* M. Fieschi et al, eds, 'Medinfo 2004', pp. 1403-1407, 2004, IOS Press, Amsterdam,

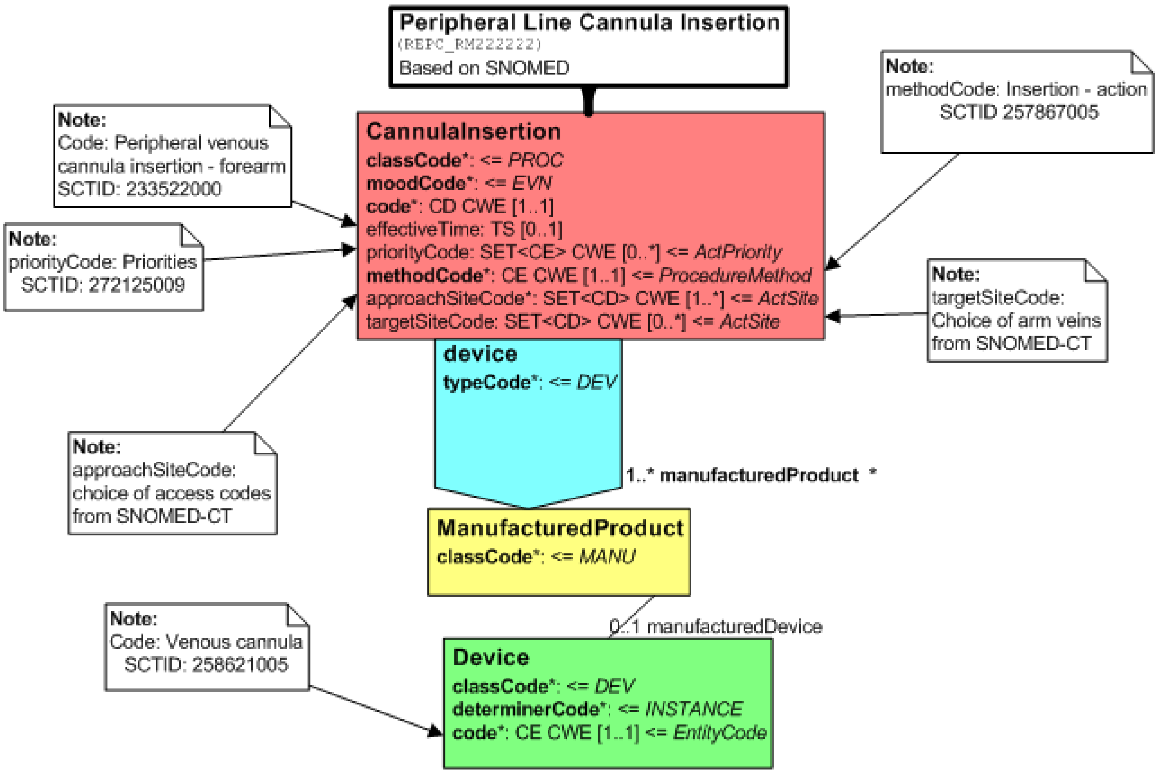

Figure 4: Cannula Insertion Concept and its attributes from SNOMED-CT represented in HL7

SNOMED-CT		HL7	
Concept	Attribute	Class	Attribute
Peripheral venous cannula insertion - forearm	n/a	Procedure	code
Insertion - Action	Method	Procedure	methodCode
Structure of superficial forearm vein*	Procedure Site	Procedure	targetSiteCode
Priorities*	Priority	Procedure	priorityCode
Surgical access values*	Access	Procedure	approachSiteCode

Table 3: Mapping from SNOMED-CT to HL7 in Figures 3 and 4.