1. Detailed aspects of issues with a vocabulary specification formalism
	1. Introduction

SNOMED CT Concept Domain Constraints (§ **Error! Reference source not found.**) specifies SNOMED CT value sets using a ‘simple notation'. As noted in Approach to Value Set Constraint Specifications (§ **Error! Reference source not found.**) this simple notation may result in certain error patterns in membership testing.

Two related requirements for value set specification need to be considered. These, and the dominant error patterns that will be encountered if they are not addressed, are as follows:

1. The ability to specify post-coordination-accommodating ‘implicit Expression’ value sets. It is already a common convention in HL7 V3 messaging specifications to exploit the hierarchical nature of certain terminologies by referring to a value set member and indicating that this member and the members that it subsumes should constitute an allowable value set. An extended requirement refers to the ability to define value sets that will test the suitability of candidate SNOMED CT expressions which may have become valid as a result of post-coordinated attribute refinement, but would fail ‘simple implicit’ validity testing against ‘node’ subsumption ('simple subsumption'). Failure to satisfy this requirement will principally result in false negative (erroneous rejections) treatment of suitable post-coordinated expressions.
2. The ability of value set specifications to allow the communication of 'Finding/Procedure/Observable entity'-based expressions as well as their post-coordinated ‘situation/context wrapped’ and pre-coordinated 'context-dependent' or 'situation' counterparts [SNOMED CT has recently renamed the top-level concept of 'context-dependent categories' (where can be found 'finding' and 'procedure' concepts where nuances of status and state are explicit in the reference data) - the top-level concept in this chapter is now named 243796009 | situation with explicit context (situation |)|]. Failure to satisfy this requirement will commonly result in false negative (erroneous rejection) treatment of suitable post-coordinated expressions and either false positive (erroneous inclusion) or false negative treatment of pre-coordinated expressions (depending on the value set specification strategy adopted).

Neither of the above requirements can be satisfactorily supported by value set specifications that either simply enumerate ‘valid codes’ or ‘valid subsumption nodes’ ('codes and logical descendants').

These requirements are now explored in more detail, and are followed by (1) an enumeration of the desirable features of a specification formalism and (2) an explanation for the inclusion of a normalization step in value set testing.

* 1. ‘Implicit Expression’ value sets

Whilst ‘simple implicit’ (subtype testing) value set specifications are suitable for ‘Primitive’ SNOMED CT concepts (even if post-coordination is allowed), in those situations where value sets are specified by reference to ‘Fully Defined’ concepts, a ‘simple’ solution is inadequate.

 E.2.1 Requirements for ‘abstract or Primitive SNOMED CT concepts’

As with (presumably) all vocabularies organized by subsumption hierarchies, SNOMED CT includes a number of abstract[[1]](#endnote-2) ‘high-level’ concepts that can be thought of as organizing the content into coherent ‘chapters.’ By example, SNOMED CT has high-level concepts such as [ 404684003 | clinical finding |], [ 71388002 | procedure |] and [ 105590001 | substance |], each correspondingly subsuming thousands of pre-coordinated concepts that are deemed to be ‘Findings’, ‘Procedures’ or ‘Substances.’

It is a property/requirement of the SNOMED CT classification process that a distinction is made between ‘Primitive’ and ‘Defined’ concepts (put simply, only Defined [in terms of other concepts] Concepts can acquire new, inferred sub-types as a result of the classification process), and whilst including a high number of Defined Concepts is desirable for more complete classification, it is an inevitable feature of SNOMED CT that a number of concepts need to be regarded as Primitive (to introduce nuances of the world against which ‘Defined’ content can be formally differentiated).

For this guide, the importance of the Primitive/Defined[[2]](#endnote-3) distinction is that as long as value sets are defined by reference only to Primitive Concepts, we can be confident that, even where post-coordination is allowed[[3]](#endnote-4), expressions cannot logically be ‘made’ members of the value set (§ Figure 1). This suggests that for many coarse-grained ‘universal’ value set specifications there is little need for a specification form of greater sophistication than:

“this field may communicate concepts subsumed by [SNOMED CT Primitive]a OR subsumed by [SNOMED CT Primitive]b OR subsumed by [SNOMED CT Primitive]…n”

which would appear to be satisfactorily supported by a notation similar to the current HL7 documentation convention of:

|  |
| --- |
| Example 32.  |
| Act.code <= [SNOMED CT Primitive]a OR [SNOMED CT Primitive]b OR [SNOMED CT Primitive]…n |

 E.2.2 Requirements for ‘detailed or fully-defined SNOMED CT categories’

Whilst many ‘universal’ value sets can be specified by the mechanism above, as concept domains are progressively constrained we may reach a point where a detailed SNOMED CT-derived value set is specified by reference to one or more Fully-Defined Concepts[[4]](#endnote-5). In this setting, where post-coordination is allowed, it will be possible to 'create' expressions that are now members of the value set but whose ‘focus concepts’ would not be members according to ‘simple’ subsumption testing (Figure 2). The '(clinical) focus concept' (often singular but strictly-speaking a set of 'focus concepts') is the concept that to a large degree characterizes the type of expression being documented or communicated. Reference to the SNOMED CT concept model when the nature of the 'focus concept' is known will indicate which types of refinement and which axes of 'context modification' can be applied. For example, if the 'focus concept' is a member of the set specified by [ <<404684003 | clinical finding |], inspection of the concept model will tell us that the concept can be modified by selecting/refining values for defining characteristics with attribute names such as [ 363698007 | finding site |], [ 246112005 | severity |], [ 116676008 | associated morphology |] etc., and that the focus concept can serve as the value of an [ 246090004 | associated finding |] attribute of a 'context/situation' wrapped post-coordinated expression. Additional information which may influence appropriate aspects of model application are (1) whether a concept chosen from the sets specified by [ ((<<363787002 | observable entity |) OR (<<386053000 | evaluation procedure |)) ] is accompanied by a value (determining whether it should be treated by the concept model as a 'finding' or a 'procedure') and (2) the moodCode value of the relevant HL7 V3 class (as this will determine the detailed value applied to the respective attribute names [ 408729009 | finding context |] or [ 408730004 | procedure context |]).

In order to avoid false rejection of valid ‘post-coordination by refinement’ expressions, value set specifications need to be modified to allow their inclusion. Consistent with the guidance that is currently offered for normal form generation for data retrieval, the following general modifications to each specification (and, for comparison purposes, each 'candidate' expression) should be considered:

* A ‘relaxing’ of each ‘Focus Concept’ to its proximal primitive supertype(s)
* Explicit reference to the required attributes of valid expressions

By example, such a transformation would result in the 'simple' value set ‘predicate’

33149006 | Pancreatectomy |

being rephrased as

71388002 | procedure |:
{ 260686004 | method |= 129304002 | excision - action |,
363704007 | procedure site |= 15776009 | pancreatic structure |}

and the value set ‘candidate’

9524002 | Total pancreatectomy |

being rephrased as

71388002 | procedure |:
{ 260686004 | method |= 129304002 | excision - action |,
363704007 | procedure site |= 181277001 | entire pancreas |}

As indicated in Figure 2, failure to include this transformation step would result in inappropriate rejection of a valid post-coordinated representation of 'total pancreatectomy'.

The tests performed to determine value set membership against 'complex' value set predicates are still 'subsumption' tests, but here might be regarded as 'complex subsumption', insofar as testing will include steps such as candidate expression normalisation, structural predicate/candidate comparison (or the equivalent) and multiple 'simple subsumption' tests for focus concept, attribute name and attribute value comparisons.



Figure 1: The consequences of refinement post-coordination on valid value set membership for sets defined by reference to Primitive Concepts



Figure 2: The consequences of refinement post-coordination on valid value set membership by reference to Fully Defined or 'Definable’ Concepts

* 1. Pre- and Post-Coordinated Concepts and Expressions

For many 'universal' specifications it will be possible to anticipate and provide a set of appropriate value set clauses that would adequately accommodate, in a general sense, paired 'plain' pre-coordinated (concepts specified in the sets [ <<404684003 | clinical finding |] and [ <<71388002 | procedure |]) and 'context/situation' ([ <<413350009 | finding with explicit context |] and [ <<129125009 | procedure with explicit context |]) pre-coordinated concepts. It may well be desirable to do this, as there are many concepts in the set specified by [ <<243796009 | situation with explicit context |] that represent useful clinical notions, and also happen to represent some nuance of status or state (e.g. 'dizziness' is a 'finding' and 'dizziness present' is an 'explicit situation' - the latter represented in SNOMED CT by [ 162260006 | dizziness present |]).

For example we may wish to specify that both 'plain' clinical findings or their 'context/situation' counterparts are valid value set members:

|  |
| --- |
| Example 33.  |
| Observation.code ((<<404684003 | clinical finding |) OR (<<413350009 | finding with explicit context |)) |

Whilst not worrying about the details of moodCode/context bindings (not all pre-coordinated 'context/situation' concepts would be valid for all moodCode values), this notation would appear to suffice (specifying the paired ‘clinical finding’ and the ‘explicit situation' finding). However, consider the following more precise/refined 'simple' value set, consisting of a set of 'plain' concept clauses:

* ((<<50043002 | disorder of respiratory system |) OR
* (<<49601007 | disorder of cardiovascular system |) OR
* (<<119292006 | disorder of gastrointestinal tract |))

To reproduce the paired ‘simple & situation’ pattern here we also need pre-coordinated concepts of the form:

* Respiratory system disorder with explicit context OR
* Cardiovascular system disorder with explicit context OR
* Gastrointestinal tract disorder with explicit context

Even if these did exist as pre-coordinated concepts (currently they do not) we may well run into the same pattern of problem as in ‘Implicit Expression’ value sets, since it will be possible to ‘make’ a ‘Cardiovascular system disorder with explicit context’ expression by post-coordinated refinement of a suitable 'context/situation' supertype.

The second pattern of 'context/situation' representation (for example stating in a SNOMED CT expression that a finding is 'present', or that a procedure was 'performed in the past on a family member') is the use of SNOMED CT's 'context/situation' wrapper in the creation of expressions. This has the effect of introducing a 'focus concept' from the set specified by [ <<243796009 | situation with explicit context |] and the nesting (as the value of the respective attributes [ 246090004 | associated finding |] or [ 363589002 | associated procedure |]) of the 'clinical kernel' expression.

Re-using the 'dizziness' example, it is possible to document 'dizziness present' by both the use of the existing pre-coordinated 'context/situation' concept ([ 162260006 | dizziness present |]) and by the following expression [ 373573001 | clinical finding present |: 246090004 | associated finding | = 404640003 | dizziness |]). A value set specification that is looking for subtypes of 'dizziness' (the 'finding') and 'dizziness present' (the 'situation') (by the specification [ ((<<404640003 | dizziness |) OR (<<162260006 | dizziness present |))] ) would inappropriately reject this expression if only the 'focus concept' [ 373573001 | clinical finding present |] was tested.

 E.3.1 Context or situation wrapping, refinement and normal forms

We therefore have two patterns of problem:

1. For pre-coordinated content, where explicit context/situation variants should also be allowed, we will need pre-coordinated 'situation with explicit context' concepts that may well not exist
2. Where post-coordination is allowed we also need specifications to accommodate content that has 'become valid' as a result of
* refinement of 'plain' concepts
* refinement of 'context/situation' concepts
* 'context/situation wrapping' of 'plain' concepts

It therefore seems reasonable to regard value set specifications where post-coordination (by sub-type refinement or context/situation wrapping) is taking place as similar to predicate specification for post-coordinated data retrieval.

With some modifications and additional tuning (see below) such ‘value set predicates’ can be generated by processing pre-coordinated ‘simple’ concepts according to the published rules for SNOMED CT expression transformation to normal forms. Without loss of precision this will result in specifications that will appropriately allow the communication of expressions that would have been missed by simple subsumption testing.

* 1. End Result

Taking the above suggestions to their conclusion, it is recommended that even for the most abstract value set specification, an inclusive value set representation will need to be modified from:

|  |
| --- |
| Example 34.  |
| Observation.code ((<<404684003 | clinical finding |) OR (<<413350009 | finding with explicit context |)) |

To a form that states 'following value set normalization, valid expressions will be those with a focus concept in the descent of Finding with explicit context ( [ <<413350009 | finding with explicit context |] ) and a value for the attribute [ 246090004 | associated finding |] from the descent of Clinical finding ([ <<404684003 | clinical finding |]). For example:

|  |
| --- |
| Example 35.  |
| Observation.code << [Following ‘value set’ normal form transformation]:413350009 | Finding with explicit context |: 246090004 | associated finding | = <<404684003 | clinical finding | |

For more refined/precise value sets the change would be from:

|  |
| --- |
| Example 36.  |
| Observation.code (( <<50043002 | disorder of respiratory system |) OR  ( <<NO CODE | respiratory system disorder with explicit context |) OR  ( <<49601007 | disorder of cardiovascular system |) OR  ( <<NO CODE | cardiovascular system disorder with explicit context |) OR  ( <<119292006 | disorder of gastrointestinal tract |) OR  ( <<NO CODE | gastrointestinal system disorder with explicit context |)) |

To a form more like

|  |
| --- |
| Example 37.  |
| Observation.code [Following ‘value set’ normal form transformation]: <<413350009 | Finding with explicit context |:  246090004 | associated finding | = ( <<64572001 | disease |: 363698007 | finding site | =  <<(( 20139000 | respiratory system structure |) OR  (113257007 | cardiovascular structure |) OR  (122865005 | gastrointestinal tract structure |)) |

* 1. Transformation rules.

General purpose SNOMED CT expression transformation rules published by IHTSDO apply default context values to un-specified axes e.g. the full transformation of ‘Disorder of cardiovascular system’ is:

243796009 | situation with explicit context|:
{ 246090004 | associated finding | = ( 64572001 | disease |:
363698007 | finding site |= 113257007 | cardiovascular structure |)
,408729009 | finding context |= 410515003 | known present |
,408731000 | temporal context |= 410512000 | current or specified |
,408732007 | subject relationship context |= 410604004 | subject of record |}

The use of these defaults is appropriate for retrieving and analyzing data that does not include any explicit indication of context. However, within HL7 artefacts, context may be denoted by other attributes or by associated classes in the model. The SNOMED CT expression transformation rules described a process for recombining information model contextual information to provide accurate values for unspecified attributes. Explicit context attribute values, such as those shown in the expression below, are considered prior to merging information model context. The defaults only apply to missing context information after explicit context information provided by the information model has been taken into account.

243796009 | situation with explicit context |:
{ 246090004 | associated finding |= ( 64572001 | disease |:
363698007 | finding site |= 113257007 | cardiovascular structure | )}

* 1. Representation concept model constraints

The IHTSDO publishes human and machine-readable representations of the main elements of the SNOMED CT Concept Model. Further work is continuing to formalize an extended expression constraint syntax which is closely related to the constraint representation in this document. Enhancements to constraint representation arising from that work will be applied to future releases of this guide.

* 1. Schematic Illustrations of SNOMED CT Expressions

These following informal representations of expression help to explain some of the language used elsewhere in this guide. In November 2013, IHTSDO adopted a set more fomal diagramming guidelines for SNOMED CT definitions and expression and this will be used in future versions of this document.



Figure 3: Illustration of names used to refer to general elements of an expression



Figure 4: Illustration of the names used to refer to parts of a nested expression



Figure 5: Illustration of the names used to refer to parts of an expression that represent context

1. The distinction between ‘abstract’ and ‘detailed’ (e.g. between ‘procedure’ and ‘total pancreatectomy’) might be better articulated in alternative ways (e.g. ‘narrow’ and ‘broad intension’), but it is hoped that the point is clear. [↑](#endnote-ref-2)
2. Whilst it is fair to say that many ‘abstract’ SNOMED CT Concepts are ‘Primitive’, it should also be noted that many ‘detailed’ Concepts – such as the vast majority of concepts in the descent of [ <<105590001 | substance |] are also Primitive. [↑](#endnote-ref-3)
3. With the exception of ‘context/situation wrapping.’ [↑](#endnote-ref-4)
4. Or ‘potentially’ Fully Defined – that is, Concepts that could be modeled as Fully Defined within the published SCT concept model, but where either the modelling is insufficient, or where the modelling is sufficient but the 'Fully Defined' assertion has not been made by the editors of the terminology. [↑](#endnote-ref-5)