

# Maintaining simple SNOMED CT reference sets

M. Cordell#, R. Richards, C. Richardson

National Clinical Terminology and Information Service (NCTIS) - NEHTA

# Corresponding author: Matt Cordell Email matthew.cordell@nehta.gov.au

## Introduction

SNOMED CT® reference sets are initially developed against a specific edition. However, with content being updated every six months there is always potential for referenced components to change during any of these updates. When used as a mechanism for subsetting, it's generally desired that only active components are referenced within. Maintenance considerations are important for organisations publishing formal reference sets and end users who may create bespoke subsets.

## Background

The workbench used by the NCTIS currently provides two different methods of creating such reference sets:

### Specific selection of components

Concepts may be individually selected, possibly from the result of mapping to some reference list or by systematically adding and removing sub-hierarchies.

For example:

- First including all the [125605004 *Fracture of bone (disorder)*] concepts;
- Then excluding *the* [439987009 *Open fracture of bone (disorder)*] and [213346004 *Fractures involving multiple body regions (disorder)*] subclasses.

This approach is often chosen when a specific selection of concepts has been identified for inclusion in the reference set.

### Generation via a rules based specification

The specification is simply a list of criteria connected by logical operators “AND” and “OR” from which the reference set is calculated.

For example:

- Concepts that are: Types of - [125605004 *Fracture of bone (disorder)*]
- AND NOT Types of ( [439987009 *Open fracture of bone (disorder)*] OR [213346004 *Fractures involving multiple body regions (disorder)*]).

A specification may be created when reference set constraints or requirements may be described as distinct rules to be applied to SNOMED CT.

## Methods

Both the current SNOMED CT-AU release in development and the last published release (RF1 and RF2) along with the last two releases of the international edition of SNOMED CT (RF1) are imported into a relational database.

### Quality Checks

For most of the NCTIS reference sets some common quality checks to determine validity are automatically performed daily:

- Concepts are active (current or pending move status)
- Concepts are from specific top level hierarchies
- Reference sets do not contain concepts identified as non-human

Conformance to these constraints is monitored routinely, with particular attention committed following the six monthly updates.

### Reports

A report is generated for reference sets created through the specific selection mechanism to guide content authors as to the changes required following an international update. These reports currently detail the:

- Concepts that have been retired
- Potential replacement candidates

There is effectively a query used for each of the details above. Retired members are simply identified by checking which of the active members in each reference set have been retired.

```
SELECT * FROM referenceSet
WHERE referencedComponentId NOT IN (SELECT conceptId FROM RF1_Concepts
                                     WHERE status IN (0,11))
```

The result set from this query, “RetiredMembers”, are the concepts that require action. The RF1 historical relationships are leveraged to identify the candidate replacements for each of these concepts.

```
SELECT * FROM RF1_Relationships
WHERE characteristicType = 2
AND conceptId1 IN ("RetiredMembers")
```

Finally, to improve human useability of the report, a sequence of join queries are added that provide descriptions for all the concept ID's – refsetId, conceptId1, relationshipType, conceptId2.

### Actioning the maintenance report

The report is reviewed by the terminology team to confirm the actions suggested by the report, particularly when there is more than one option.

Option	Comment
MAY BE A	For concepts identified as ambiguous. There are usually multiple candidate replacements. Which candidates are accepted is resolved through consensus or discussion.
MOVED FROM	Not encountered.
MOVED TO	Not encountered.
REPLACED BY	Erroneous concepts are often replaced by more accurate candidates.
SAME AS	For concepts identified as duplicates. Only a single, non-contentious candidate replacement is reported.
WAS A	Candidate replacements are the less specific supertype of the retired concept. The candidate may be accepted, or a more specific replacement manually identified.
“not stated”	Not encountered.

These reports are team reviewed to determine required content changes, and the required edits are then allocated to content authors for further action. The collaborative process is fully documented with all decisions, actions and peer reviews added to the maintenance report.

nehta

National E-Health Transition Authority

www.nehta.gov.au

### Maintaining specification based reference sets

Recalculation of specification based reference sets is mandatory. Concepts are automatically added or removed from the reference set as a result of any changes to core tables. Consequently, reference sets generated by a specification are at the mercy of SNOMED CT modelling. Changes to reference set membership following an update are monitored using difference reports. The changes may be investigated to identify the root cause of the changes – for example the addition or removal of a single IS A relationship can potentially result in a change of membership for hundreds of concepts. The cause should be assessed to determine subsequent actions.

Both modelling AND reference set membership changes are acceptable	No action
Modelling changes are acceptable BUT reference set membership changes are not acceptable	Revision of the reference set specification is required
Modelling changes are questionable AND reference set membership changes are not acceptable	Direct feedback to the international release querying the change. The specification may be modified to compensate for the change either permanently or temporarily.

## Results

Reference sets developed by the specific selection of concepts tend to contain content that is less subject to change compared with the specification reference sets. The concepts chosen for inclusion are typically very close to the use case, unambiguous and devoid of aggregating concepts that are not of clinical value - characteristics that may warrant a concept's retirement. The table below is an example of the information provided in the maintenance reports that are generated.

Reference set	Member ID	Member Term	Option	Candidate ID	Candidate Term
Specimen anatomical site	71993006	Capsule of distal interphalangeal joint of great toe	REPLACED BY	363672001	Structure of capsule of interphalangeal joint of great toe
Emergency department diagnosis	162345005	Blocked ear	MAY BE A	445849005	Obstruction of external auditory canal
Emergency department diagnosis	162345005	Blocked ear	MAY BE A	103281005	Sensation of blocked ears
Specimen anatomical site	244692006	Neck of glenoid of scapula	SAME AS	63572008	Structure of neck of scapula

- Twenty reference sets are based on this method covering over 40,000 concepts.
- Less than twenty concepts were affected in each of the last two maintenance cycles (< 0.05% of members impacted)

Content requirements for specification reference sets are often broad sweeping and maximally inclusive and therefore more exposed to potential changes in the terminology. The current suite of reference sets developed by this approach are sufficiently broad that the changes to date have been considered to be tolerable. The table below illustrates the variety of changes that were observed in some of the specification reference sets following the July 2010 update:

1. Completely unaffected
2. New members added and no existing ones removed
3. Existing members removed and no new ones added
4. Both new members added and existing ones removed

Additionally, the table shows that the changes in membership may be caused by a combination of changes to the definitions of existing content, and the creation or retirement of new concepts. In context, the July 2010 release of SNOMED CT introduced 2331 new concepts and retired 288.

Reference Set	Total Active	New Members	New Concepts	Removed Members	Retired Concepts
Imaging procedure <sup>3</sup>	4201			3	1
Musculoskeletal finding <sup>4</sup>	10579	81	80	73	8
Neoplasm and/or hamartoma <sup>4</sup>	6652	24	24	1	
Respiratory finding <sup>2</sup>	3944	24	24		
Skeletal system <sup>1</sup>	3743				

## Discussion

The current maintenance processes are mostly a undertaken manually. There is potential for the partial automation of each process to improve efficiency. For example:

- Recalculation of specification reference sets could be automatically triggered upon update of the underlying data.
- For non-contentious decisions in the maintenance report, such as “SAME AS” and “REPLACED BY” changes, automation of the required authoring could be investigated.

Both release formats have been leveraged in generating the maintenance reports because of the progressive refinement of the technical specifications of the new release format. Now that the RF2 has been ratified, the NCTIS maintenance processes will be modified to remove dependency on the old format.

The current experience of tolerable content changes for specifications may change when the results of some of the IHTSDO project groups work progresses into the international release. Reference set authors should be aware of, and prepared for, these changes in order to anticipate future maintenance requirements.

## Conclusion

The reference sets developed from specifications are reliant on the authoring within SNOMED CT, and changes to the hierarchy can have significant impacts on the reference set content, even though the maintenance process may not require manual input. Additional maintenance activities may be required, depending on the publisher's tolerance for changes in reference set content.

For reference sets where each member has been included in response to specific requirements, the priority of maintenance activities is to ensure that each reference set maintains at least the same degree of content coverage between releases whilst also removing referenced components that are no longer valid. This process is more involved than for specification reference sets. However, the extent of maintenance required has been shown to be limited.

For any organisation developing reference sets the impact of maintenance demands, upon both resources and content, should be considered when initially determining the development process.

\* SNOMED CT® and IHTSDO® are registered trademarks of the International Health Terminology Standards Development Organisation.