

Description Logic Enhancements - Briefing Paper

Description Logic Enhancements - Community of Practice Consultation

- [Description Logic Enhancements - Community of Practice Consultation](#)
 - [Introduction](#)
 - [Rationale](#)
 - [Example of new modeling capabilities](#)
 - [Proposed Changes](#)
 - [New Reference Sets](#)
 - [New Logic Features](#)
 - [New Refset File - Example](#)
 - [New Classification Service](#)
 - [Roadmap of Changes](#)
 - [Technology Preview](#)
 - [Proposed Schedule for Change](#)
 - [CheckList to Determine Impact of Proposed Changes](#)
 - [Impact on SNOMED Consumers](#)
 - [Impact on SNOMED Producers](#)
 - [Feedback](#)
 - [Footnotes](#)

Introduction

This briefing paper details changes proposed by SNOMED International's Modeling Advisory Group to enhance SNOMED CT's Description Logic capabilities. The proposal has been accepted by SNOMED International's Management Team to be put forward for consultation with the wider Community of Practice - including Member Countries, Vendors and adopters of SNOMED CT. This document is intended to be read by technical staff in organisations who produce SNOMED CT content as an extension to the SNOMED CT International Edition. It should be read in conjunction with the attached [Executive Summary](#) which gives an overview of the background and drivers for proposing enhancements to SNOMED CT at this time.

On completion, readers are invited to submit questions and comments via [this form](#) before the cut off on 28 February 2018. Responses to this feedback will be posted on [this confluence page](#).

Rationale

The motivation for introducing these enhancements to SNOMED CT's Description Logic capabilities is to enable improvements to the quality and analytics capabilities of SNOMED CT.

The use of Description Logic together with a computer program that performs Logic Reasoning (referred to as a "classifier") allows concept hierarchies to be automatically created and maintained based on the logical definition of each concept. This automation improves the quality of the clinical information represented by SNOMED CT as well as reducing the maintenance burden inherent in modifying an increasing volume of content. With accurate logical definitions for each concept, authors can take advantage of the classifier's ability to infer new parent/child relationships automatically, rather than being required to state every valid relationship explicitly - a task which is both immensely time consuming and error-prone.

The proposed changes will benefit:

- Content authors, who will be able to achieve greater productivity and reach a higher quality at a lower cost
- Implementers (e.g. vendors), who will receive a more consistent product
- End users (e.g. clinicians, researchers), who will use a more complete and consistent product which ultimately provides them with better tools
- SNOMED International and members, who will be better able to achieve their objectives, such as improving the quality of the product, aiding adoption and attracting new members

Example of new modeling capabilities

An example of concept modeling that will become available with the new description logic capabilities is shown in the diagram below. It will become possible to declare that a concept can be either one thing or another and have each type of concept classify correct. In the example of

[Secondary diabetes mellitus \(disorder\)](#)

, any concept which IS A

[Diabetes mellitus \(disorder\)](#)

and is *either*

[Due to \(attribute\)](#)

a type of

[Disease \(disorder\)](#)

or has a

|Causative agent (attribute)|

which is a type of

|Pharmaceutical / biologic product (product)|

will be classified as a type of

|Secondary diabetes mellitus (disorder)|

automatically, rather than requiring this hierarchy to be manually maintained.

[blocked URL](#)

Proposed Changes

The proposal is to make more types of information available to the classifier than can currently be expressed within the limitations of the RF2 Stated Relationship file format. In the current classification process, the Stated Relationship file is used as input to the classifier to enable it to calculate an inferred hierarchy. The output of this process is converted into Distribution Normal Form (DNF)¹ and represented using the RF2 Relationship File. In the proposed classification process, logical definitions utilizing a wider range of Description Logic features will be available to the classifier, which will allow more accurate automatic inferences to be performed.

One of the principle considerations in designing this proposal is to minimize the negative impact for organisations who do not wish to (or are not in a position to) take full advantage of the new capabilities. The use of reference sets to replace the stated relationship file, while preserving the format of the more commonly used (inferred) RF2 Relationship File, will allow most users to benefit from the improvements in classification, without requiring changes to existing systems.

New Reference Sets

A new RF2 reference set file containing OWL Statements in Functional-Style syntax will initially augment, and subsequently replace the existing Stated Relationship file. The name of the new reference set file is expected to follow the format "der2_sRefset_StatedOWLFULL_INT_YYYYMMDD.txt", and be published as part of the SNOMED CT International Edition. The OWL statements will be organised into two refsets as follows:

Refset	Usage
762103008 OWL ontology reference set	Contains the 'setup' information for the ontology, static 'headers' that aren't expected to change from one release to another.
733073007 OWL axiom reference set	Contains information relating to the definition of individual concepts

The header information in the OWL ontology reference set as well as some of the attribute logical properties (such as the fact that the

123005000 |Part of|

attribute is transitive in nature) will now make explicit behaviour that has previously been somewhat hidden in existing software tools.

New Logic Features

Allowing SNOMED CT to use the full range of logical expression that is possible with Description Logic would cause a dramatic increase in classification times¹. It would also increase the complexity for content authors in their mission to correctly define the meaning of clinical concepts. As such, the Modeling Advisory Group has recommended the addition of only those Description Logic features that are essential to support the proposed modeling of new content in specific priority hierarchies. This aims to successfully strike the balance of being able to express essential meaning, while not introducing unnecessary complexity.

The proposed new logic features are explained in the table below:

New Feature	Explanation	Use Case example
-------------	-------------	------------------

Property Characteristics	Allows SNOMED CT to specify which attributes should have transitive and reflexive properties.	<p>In the Body Structure hierarchy, there is a requirement to make the</p> <p> Part of </p> <p>attribute transitive. So if</p> <p> Entire finger </p> <p>is</p> <p> Part of </p> <p> Entire hand </p> <p>and</p> <p> Entire hand </p> <p>is</p> <p> Part of </p> <p> Entire upper limb </p> <p>, then we want the classifier to be able to infer that</p> <p> Entire finger </p> <p>is</p> <p> Part of </p> <p> Entire upper limb </p>
Property Chains	Allows attributes to be linked together such that additional logical inferences can be made where both attributes are used in combination.	<p>This logic feature would allow</p> <p> Has active ingredient </p> <p>to be linked with</p> <p> Is modification of </p> <p>such that a medicinal product that has an active ingredient which is the modification of another substance, could classify as a child of a product containing the less modified substance.</p> <p>This behaviour is critical to controlling the effect of changes to the</p> <p> Substance (substance) </p> <p>hierarchy on other hierarchies.</p>
Additional Axioms including General Concept Inclusion	<p>Currently all attributes stated for a concept must be present for another concept to be considered a descendent in the hierarchy. General concept inclusions with additional axioms provide authors with greater flexibility in how concepts are defined, and provide classifiers with a greater ability to make appropriate inferences.</p> <p>For example, these features can be used to state that a subset of attributes is sufficient to define a concept, and therefore that this sufficient set can be used to classify other concepts as descendants in the hierarchy.</p>	<p>In the</p> <p> Clinical finding (finding) </p> <p>hierarchy, there is a requirement to define</p> <p> Secondary diabetes mellitus </p> <p>such that it can either be due to</p> <p> Disease </p> <p>OR caused by a</p> <p> Drug or medicament </p> <p>. Current restrictions in logic capabilities prevent subtypes of Diabetes mellitus that are due to a</p> <p> Disease </p> <p>or caused by a</p> <p> Drug or medicament </p> <p>from being classified as a subtype of</p> <p> Secondary diabetes mellitus </p> <p>.</p>

New Refset File - Example

An example of the proposed refset file format with some sample rows is shown here (preferred terms have been included to aid readability):

moduleId	refsetId	referencedComponentId	owlExpression
900000000000012004 SNOMED CT model component module	762103008 OWL ontology reference set	734146004 OWL ontology namespace	Prefix(=< http://snomed.info/id/ >)
900000000000012004 SNOMED CT model component module	762103008 OWL ontology reference set	734146004 OWL ontology namespace	Prefix(owl:=< http://www.w3.org/2002/07/owl# >)
900000000000207008 SNOMED CT core module	733073007 OWL axiom reference set	404684003 Clinical finding	SubClassOf(404684003 :138875005)
900000000000012004 SNOMED CT model component module	733073007 OWL axiom reference set	123005000 Part of	SubObjectPropertyOf(123005000 :733928003)

New Classification Service

To make it easier for organisations with a requirement to classify SNOMED CT, SNOMED International will make an open source classification service available to download. This service will classify SNOMED CT based on the new OWL reference set file. As well as performing classification, the service will calculate changes to the (inferred) relationship file in the Distribution Normal Form², and will (eventually) produce a complete OWL Ontology file which can be imported into tools such as Protégé. This will replace the existing PERL script that is provided for converting RF2 file to OWL.

The new SNOMED CT classification service will be used internally by SNOMED International for authoring terminology and producing releases. External organisations will be able to run the service locally, or use it as a reference implementation to implement these features in their own software.

Roadmap of Changes

Technology Preview

A technical preview of the new RF2 OWL Refset file will be published by the end of February 2018, including examples of the new logic features used in modeling. The technical preview will be revised and reissued as required should issues arise.

Proposed Schedule for Change

The table below shows the features that will be introduced over the next few releases of the SNOMED CT International Edition. This is an optimistic timeline that has been driven by the desire to fulfil obligations for successful delivery of work being done in the

373873005 |Pharmaceutical / biologic product (product)|

and

105590001 |Substance (substance)|

hierarchies:




International Release	Stated Relationship File	Additional Logic Features in International Release	Production OWL Refset	Technical preview OWL refsets
20180131	Unchanged.	Nil	N/A	Snapshot of stated relationships plus: <ul style="list-style-type: none"> General Concept Inclusions Property Axioms Expected publication early February 2018.
20180731	Unchanged. Will not include additional logic features. Insufficient on its own for correct classification.	<ul style="list-style-type: none"> General Concept Inclusions Property Axioms 	First official release of OWL refset. To be used as the primary input into the classification process. Full, Snapshot and Delta of all content.	N/A

20190131	Deprecated.	Support all features in SNOMED CT logic profile.	Content changes authored as normal.	N/A
----------	-------------	--	-------------------------------------	-----



CheckList to Determine Impact of Proposed Changes

SNOMED Implementers can be considered in two broad categories - those who consume SNOMED as it is published without adding further concepts (Consumers), and those who add concepts to an extension (Producers). The following two checklists address both these categories of implementers in turn. Since Producers make content for use by Consumers, both tables will be of interest to them.

Impact on SNOMED Consumers

Your Current or Planned Usage	Impact of Change	Detail of Impact
Adding translated content (and /or additional synonyms) as descriptions.		This type of content will be unaffected by the proposed changes. The new OWL Refset will only contain logical definitions. Descriptions and Language Refset Acceptability will remain unchanged.
Adding members to an existing reference set, such as a Simple or Map Reference Set		This type of content will be unaffected by the proposed changes.
Subsumption Testing of pre-coordinated content using the Is a (attribute) hierarchy - for example, using a transitive closure table to determine if one concept is a descendent or ancestor of another.		The Is a (attribute) hierarchy will continue to be represented in the (inferred) Relationship File and so no changes will be required to satisfy this use case.
Testing subsumption between postcoordinated expressions by transformation to normal forms. This approach to testing subsumption is also referred to as "Structural Subsumption Testing".	blocked URL	It will no longer be possible to test subsumption between postcoordinated expressions using transformation to normal forms. Instead, subsumption testing of expressions will require use of a classifier. SNOMED International know of a small number of organisations who currently use Structural Subsumption Testing and are keen to hear from others who do so, to ensure that appropriate guidance can be provided. Any one who is aware of implementations that perform structural subsumption testing should notify SNOMED International via an email to techsupport@ihtsdotools.org with the subject line of "Structural Subsumption Testing".

Impact on SNOMED Producers

Your Current or Planned Usage	Impact of Change	Detail of Impact
Representing the "Stated" view of concept definitions in RF2 based software tools.	blocked URL	While the (inferred) Relationship file (which is most often consumed by SNOMED CT implementations) will not change in format, software which currently uses data from the Stated Relationship file will need to be re-engineered to instead pull this data from the OWL Axiom Refset.
Adding extension concepts and/or defining relationships to SNOMED CT		Each new concept added to SNOMED CT requires at least one defining relationship. As the Stated Relationship File in the International Edition will be deprecated and replaced by the OWL Axiom Reference Set, any new defining relationships will need to be added to this new reference set. The Relationships File will need to be automatically generated from the OWL Axiom Reference Sets. This can be done using the free service provided by SNOMED International.
Validating SNOMED CT release artifacts.	blocked URL	Validation rules will need to be modified to allow for the absence of the Stated Relationship file, and to instead check that the OWL Reference set has been correctly formed.
Classifying SNOMED CT.		SNOMED International will publish an open source classification service, which will perform this task. This will support extensions that add axioms to the OWL Axiom Refset in an extension module. The input to this service will be RF2 Concept File, Description File and the two new OWL reference sets. The output will be the inferred relationships, as defined in RF2 format (using Distribution Normal Form ¹). It is also intended that a full snapshot OWL Ontology file could be produced for use in other tools such as Protégé.

Authoring content using OWL based software tools.	blocked URL	Once the OWL Refsets are available it will be much easier to author content directly in OWL. However, some transformation will still be required to combine axioms from the OWL reference sets to form a full snapshot OWL ontology file (as required by existing OWL tools).
---	-----------------------------	---

Feedback

Feedback on this proposal is invited via [this form](#). SNOMED International will respond to feedback (within two weeks of receipt) by posting responses on [this page of our confluence site](#). The consultation period finishes on 28th February 2018.

Further Reading

[SNOMED CT Logic Profile Enhancements](#)

[OWL Reference Set Specification](#)

[Logic Profile Specification](#)

[Classification Service with documentation available in GitHub](#)

- Makes use of the [Snomed Owl Toolkit](#) which is capable of converting SNOMED CT Snapshot files into an OWL ontology file (functional syntax) for use in 3rd party tools such as Protégé.
- Documentation for generation of [Necessary Normal Form](#).

Footnotes

¹ Testing shows classification time increasing from 50 seconds to 1.5 hours, which is considered unacceptable in a realtime authoring environment.

² The term "Distribution Normal Form" (DNF) will be changed to Necessary Normal Form (NNF) to better reflect the content of the (inferred) Relationship file.