# Normal forms

Every application has its own data entry screens, workflow, internal database design, and other nuances, and yet despite this, we talk of semantic interoperability. In order to achieve interoperability, and enable a receiver to aggregate data coming from any number of applications, it must be possible to compare data generated on any of these applications. In order to compare data, it helps to imagine a canonical or normal form. [A canonical form is defined as a serialized representation of a SNOMED CT expression which follows the normal form and in which the refinements, attributes and attribute groups are arranged in a standard order.] If all data, regardless of how it was captured, can be converted into a common representation, it becomes possible to compare.

To that end, we differentiate the "model of use" from the "model of meaning", where the former represents the way in which the data was captured, and the latter represents a common representation. All representations recommended in this guide can be converted into a common model of meaning. This common model of meaning can be expressed in a SNOMED CT normal form and/or a RIM Normal Form, thereby enabling data comparisons.

SNOMED CT Normal Forms

The text below is taken from the introduction to the document 'SNOMED CT Transformations to Normal Forms', and outlines the purpose of transformations and the general method of transformation. Related information can be found in the Technical Implementation Guide ([TIG](http://www.ihtsdo.org/fileadmin/user_upload/doc/en_us/tig.html)). From the perspective of integration of SNOMED CT expressions in HL7 communications the assumption is that in most cases a "close to user" form will be stored and communicated. The normal form transformation provides a method that enables consistent comparison of these expressions with one another and with retrieval queries.

The purpose of generating normal forms is to facilitate complete and accurate retrieval of pre- and post-coordinated SNOMED CT expressions from clinical records or other resources.

The approach described is based on the description logic definitions of SNOMED CT concepts which are used when recording clinical statements in an electronic records system. Using this approach, expressions that are authored, stored and/or communicated in a relatively informal close-to-user form are logically transformed into a common normalized form. In this normalized form it is possible to apply simple rules to test subsumption between expressions.

The simplest case of a valid close-to-user expression is a single conceptId, and the approach described can be applied to these simple pre-coordinated expressions, as well as to more complex expressions that include multiple conceptIds and refinements (qualifiers).

Likewise, transformations and normalisations can be both simple and complex, however the general principle is that the normalisation process will restate a SNOMED CT expression in terms of the 'primitive' concepts with which it is associated in the reference data. By example, the SNOMED CT concept [ 80146002 | appendectomy |] would, in essence, transform under normalisation to [ 71388002 | procedure |: { 260686004 | method | = 129304002 | excision - action |, 405813007 | procedure site - Direct | = 66754008 | appendix structure |} ] ("a procedure that consists of excising an appendix").

The approach to normalization may be applied to the specific SNOMED CT expressions but may also be extended to take account of contextual information derived from the information model in which the expression is situated. Therefore, the normal form may include SNOMED CT context information, even if this is not present in the initial SNOMED CT expression. As such the result of transformation of [ 80146002 | appendectomy |] is a simplification (the additional contextual/situation information is missing), but it is hoped that the example sufficiently illustrates the principle of normalisation.

The algorithm extends earlier work on normal forms as follows:

* Normalizes fully-defined values within definitions or expressions producing nested expressions that are fully normalized.
* Merges refinements stated in an expression with definitional relationships present in the definitions of the concepts referenced by the expression.
* The merge process takes account of refinements that may not be grouped or nested in a manner that precisely reflects the structure of a current (or future) concept definition.
* This avoids the need to add, store and communicate potentially spurious detail from current definitions to the expression recorded by a user or software application.
* Takes account of context rules including soft default context and a preliminary approach to moodCode mapping and handling of procedures with values (present in algorithm but not yet easily visible in test environment).

Transformations to Normal Forms

The requirements for full normalization of alternative representation using different combinations of SNOMED CT and HL7 RIM artifacts requires an agreed upon comprehensive reference normal form. This is beyond the scope of this document. However, the rules and guidance in Guidance on Overlaps between RIM and SNOMED CT Semantics (§ **Error! Reference source not found.** ) provide the foundations for specifying some of the more common transformation requirements.

In particular the following types of transformation may be required

* Transforming deprecated patterns using the Observation.code and Observation.value to the preferred pattern. See Act.code (§ **Error! Reference source not found.**) and Observation.value (§ **Error! Reference source not found.**) and Observation code and value (in event mood) (§ **Error! Reference source not found.**)
* Transforming based on moodCode and statusCode to apply specified contexts to SNOMED CT expressions, where these expressions do not state an explicit context. See Act.moodCode (§ **Error! Reference source not found.**) and Act.statusCode (§ **Error! Reference source not found.**).
* Transforming any deprecated uses of the negationInd attribute to an appropriate SNOMED CT expression that explicitly state appropriate "finding context" or "procedure context". See Act.negationInd (§ **Error! Reference source not found.**).
* Transforming any information in specific HL7 methodCode, targetSiteCode and approachSiteCode attributes into the appropriate refinements of the associated SNOMED CT expression. See Procedure.methodCode and Observation.methodCode (§ **Error! Reference source not found.**), Procedure.targetSiteCode and Observation.targetSiteCode (§ **Error! Reference source not found.**) and Procedure.approachSiteCode and SubstanceAdministration.approachSiteCode (§ **Error! Reference source not found.**).
* Transforming the representation of "subject" participation and SNOMED CT "subject relationship context" into a single coherent form. See Subject Participation and Subject Relationship Context (§ **Error! Reference source not found.**).

Additional documentation on this topic will be added based on experience with use of this specification.