

# 1 Introduction

## Background

SNOMED CT is a clinical terminology with global scope covering a wide range of clinical specialties and requirements. The use of SNOMED CT expressions in Electronic Health Records (EHRs) provides a standardized way to represent clinical meanings captured by clinicians and enables the automatic interpretation of this meaning. SNOMED CT expressions are a structured combination of one or more concept identifiers used to represent a clinical idea in a logical manner.

The SNOMED CT Composition Grammar is a lightweight syntax for the representation of [SNOMED CT expressions](#), which has proven to be both human readable and machine parsable.

## History

The SNOMED CT Composition Grammar was initially specified as part of the document "*SNOMED Clinical Terms Abstract Logical Models and Representational Forms, External Draft for Comment Version*". This version (v1) was used extensively as both a human readable and machine parsable syntax.

A revised version (v2.0 to v2.2) of this specification was adopted as an IHTSDO standard in 2010. This version followed the prior version in most details, with the following enhancements:

- The specification of the grammar was defined in Augmented Backus-Naur Form (ABNF)<sup>1</sup>. This provided a formal standards-based reference for the grammar's structure.
- Unnecessary whitespace designators (i.e. <ws>) were removed from several places in the grammar.
- A maximum length constraint for SNOMED CT identifiers ([SCTIDs](#)) was added. [SCTIDs](#) consist of a sequence of digits, which must be between 6 and 18 digits in length.
- The hex code for carriage return (CR) was incorrectly given as '0C' in the previous version. It was corrected to '0D'.
- Detailed character encoding information for [UTF-8](#) was added.
- The ABNF definition of [term](#) was amended to allow correct parsing by the ABNF parser generator.

In 2015, the original IHTSDO standard was revised (v2.3 and v2.3.1). The new specification is completely backwardly compatible with the prior standard, in that any expression written using the prior standard *will* necessarily conform to the new standard. However, two significant enhancements were made:

- Concrete values (e.g. integers, decimals and strings) are now permitted as attribute values.
- A definition status may now (optionally) be included at the start of an expression to indicate whether the clinical idea being expressed is 'equivalent to' or a 'subtype of' the expression.

In 2020, concrete attribute values were extended to permit booleans (v2.4).

These enhancements will be described and explained further within this guide.

## Purpose

The purpose of this document is to define and describe a formal compositional grammar which is used to represent SNOMED CT expressions. SNOMED CT expressions are a structured combination of one or more concept identifiers used to express a clinical idea. SNOMED CT expressions may exist either independently or within the scope of an information model, message instance, EHR system or clinical repository. This document also provides examples and guidance to assist in the implementation of this syntax.

## Scope

This document presents the specification of a compositional grammar, which can be used to represent SNOMED CT expressions. This document also provides a logical model for the grammar, discusses a set of example expressions and describes some implementation considerations.

The revised compositional grammar specified in this document is the first of a consistent set of computer processable languages designed to meet a broader set of requirements related to the use of SNOMED CT. Other SNOMED CT computable languages, which are either available or under development include:

- **Expression Constraint Language:** used to define a *bounded set* of clinical meanings represented by either precoordinated or postcoordinated expressions; and
- **Templates:** which allow slots to be added to expressions, expression constraints or queries, which can be filled with specific values at a later time.

[SNOMED CT Compositional Grammar](#) is designed to provide a common foundation for the additional functionality added by these other languages. However, the compositional grammar itself does not provide this added functionality.

This document does not include a full description of how to implement a compositional grammar parser, classifier or interpreter. It also does not describe how to implement an EHR which uses compositional grammar to represent clinical ideas. Instead, it provides general guidance to assist in the implementation of compositional grammar in any of these applications.

# Audience

The target audiences of this document include:

- SNOMED National Release Centers;
- SNOMED CT designers and developers, including designers and developers of EHR systems, information models, data entry interfaces, storage systems, decision support systems, retrieval and analysis systems, communication standards and terminology services;
- SNOMED CT terminology developers, including concept model designers, content authors, map developers and release process managers.

# Document Overview

This document defines, describes and provides implementation guidance for the use of [SNOMED CT Compositional Grammar](#). [Chapter 2](#) begins by describing the use cases in which it is anticipated that SNOMED CT compositional grammar will be used. [Chapter 3](#) then describes the requirements used to guide the definition of this language. In [Chapter 4](#), the logical model of compositional grammar is presented, while in [Chapter 5](#) the syntax is defined, in terms of an ABNF serialisation of the logical model. [Chapter 6](#) then presents some examples of expressions that conform to SNOMED CT compositional grammar, and [Chapter 7](#) discusses some implementation considerations. Finally, [Appendix A](#) explains how to represent concept definition and expression relationship statements.

# Glossary

The following table contains the definition of terms used within this document. Please refer to the [SNOMED Glossary](#) for additional definitions.

Term	Definition
Augmented Backus-Naur Form	A language used to define the formal syntax of another language (as defined by Internet Standard 68, RFC 5234).
<a href="#">Compositional Grammar</a>	The set of rules that govern the way in which SNOMED CT expressions are represented as a plain text string.
<a href="#">Concept Model</a>	A set of rules that determines the permitted sets of relationships between particular types of concepts.
Equivalent To (definition status)	Indicates that the given expression is both necessary and sufficient to fully define the clinical meaning being expressed.
<a href="#">Expression</a>	A structured combination of one or more concept identifiers used to express a clinical idea.
Expression Constraint	A computable rule that can be used to define a set of clinical meanings.
<a href="#">Machine Readable Concept Model</a>	A representation of the rules that comprise the SNOMED CT Concept Model in a form that can be processed by computer software and applied to validate content.
<a href="#">Postcoordinated Expression</a>	Representation of a clinical meaning using a combination of two or more concept identifiers is referred to as a postcoordinated expression.
<a href="#">Precoordinated Expression</a>	Representation of a clinical meaning using a single concept identifier is referred to as a precoordinated expression.
Subtype Of (definition status)	Indicates that the given expression is necessary but not necessarily sufficient to define the clinical meaning being expressed.

---

## Footnotes

### RefNotes

- 1 ABNF as defined by Internet Standard 68, RFC 5234