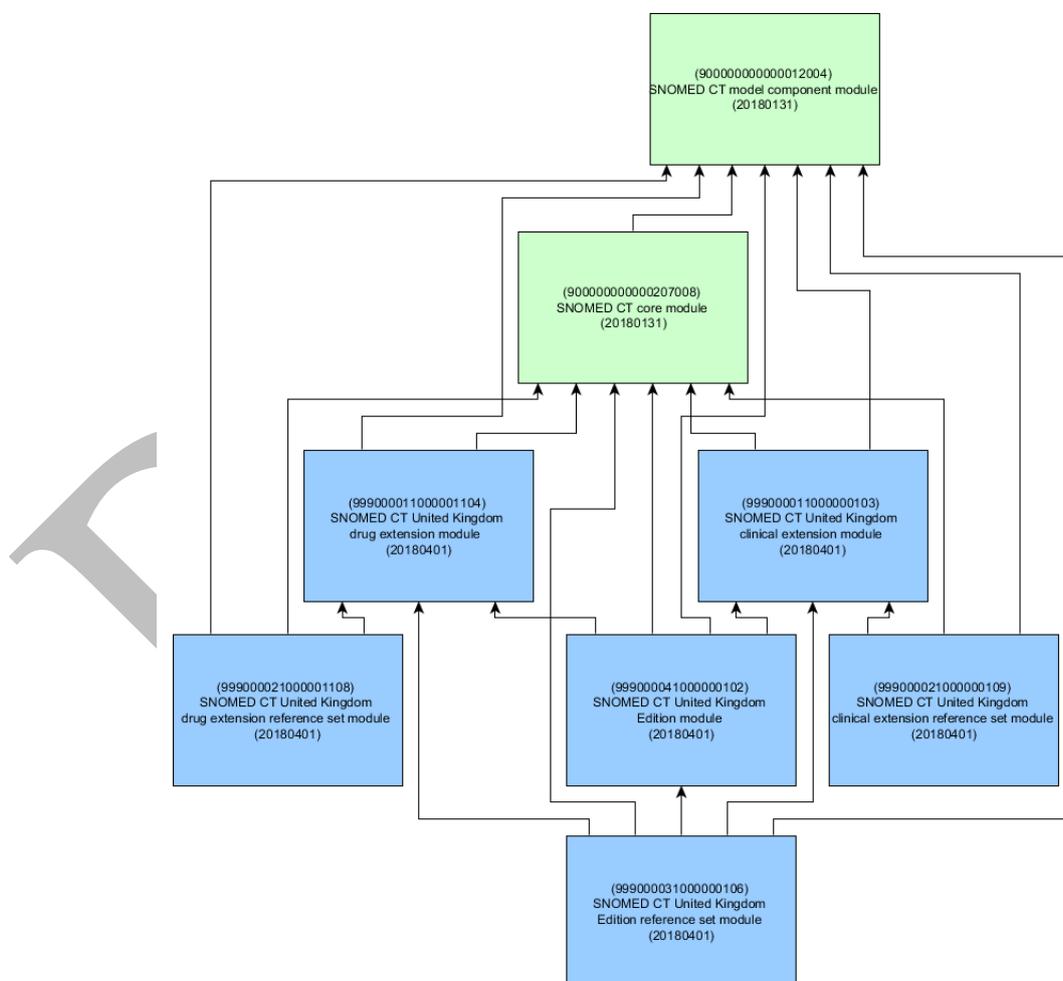


# Edition Composition Reference Set (ECRS) proposal

## Purpose

The Module Dependency Reference Set allows dependencies between *module versions*<sup>1</sup> to be stated. It is also the basis for interpreting SNOMED Version URIs.

This leads to a conflict because a moduleId (e.g., *United Kingdom Edition reference set module*) that represents the “UK Edition” which depends on the modules that contain concept definitions (*United Kingdom Edition, United Kingdom drug extension, United Kingdom clinical extension, and International Core and SNOMED CT model component*), but it does not depend on the other reference set modules even though their content is considered to be part of the “UK Edition”.



This conflict has arisen partly because the definition of the MDRS required stating all transitive dependencies (which somewhat undermines its usefulness), and partly because it was co-opted

<sup>1</sup> In this document, “module” is to be interpreted in terms of the RF2 file format. It does **not** refer to the Description Logic notion of a module, which is variously a syntactically or semantically-complete set of axioms.

for use by the URI specification as the only existing source of machine-readable information that could describe the scope of content in any specific Edition.

With the maturation of tooling that exchanges RF2 content, the introduction of the Classification Service, the modularisation of International content, and the increasing use of the SNOMED URI specification, there is a need to resolve the dual role that the MDRS is playing.

## Overview

This document proposes three things:

1. an Edition Composition Reference Set (ECRS) to identify the set of modules that contain content for an Edition, and thus supports the needs of the URI specification,
2. a backwards compatible update to the URI Specification accounting for ECRS content, and
3. a relaxation of some of the constraints on the Module Dependency Reference Set so that it can better fill its original design objective of documenting content-based dependencies between modules.

Specifically, the intent of the ECRS is to state which modules make up the content of given Versioned Edition of a SNOMED CT release. The ECRS leverages the existing content of the MDRS where dependencies between *versions* of modules are documented.

To avoid redundancy and potentially conflicting statements, the ECRS does not re-state or in any way alter the dependencies recorded in the MDRS. It is designed to be low maintenance and should only need updating if and when (leaf) modules are added / removed from an Edition. In contrast, the MDRS is updated for every Version since it records version-level dependencies.

## Reference Set Data Structure

Field	Data type	Purpose	Mutable	Part of Primary Key
id	<a href="#">UUID</a>	A 128 bit unsigned <a href="#">Integer</a> , uniquely identifying this <a href="#">reference set member</a> .  Different versions of a <i>reference set member</i> share the same <a href="#">id</a> but have different <a href="#">effectiveTime</a> . This allows a <i>reference set member</i> to be modified or made <a href="#">inactive</a> (i.e. removed from the active set) at a specified time.	NO	YES (Full /Snapshot)
<a href="#">effectiveTime</a>	<a href="#">Time</a>	The inclusive date or time at which this version of the identified <a href="#">reference set member</a> became the current version.  The current version of this <a href="#">reference set member</a> at time <i>T</i> is the version with the most recent <a href="#">effectiveTime</a> prior to or equal to time <i>T</i> .	YES	YES (Full)  Optional (Snapshot)
<a href="#">active</a>	<a href="#">Boolean</a>	The state of the identified <a href="#">reference set member</a> as at the specified <a href="#">effectiveTime</a> .	YES	NO

		If <a href="#">active</a> = 1 (true) the <a href="#">reference set member</a> is part of the current version of the set, if <a href="#">active</a> = 0 (false) the <a href="#">reference set member</a> is not part of the current version of the set.		
<a href="#">moduleId</a>	<a href="#">SCTID</a>	Identifies the <a href="#">SNOMED CT module</a> that contains this <a href="#">reference set member</a> as at the specified <a href="#">effectiveTime</a> .  The value must be a <a href="#">subtype</a> of 900000000000443000   Module (core metadata concept)   within the metadata <a href="#">hierarchy</a> .  Identifies the SNOMED CT Edition that includes the content of the <a href="#">referencedComponentId</a> module and that of all of its dependencies (as detailed in the Module Dependency Reference Set).  <b>Note:</b> In almost all other situations <a href="#">moduleId</a> is mutable. However, as in the Module Dependency Reference Set a change to the <a href="#">moduleId</a> would also change the Edition being described. Therefore, it must <b>not</b> be treated as mutable.	NO	NO
<a href="#">refsetId</a>	<a href="#">SCTID</a>	Identifies the <a href="#">reference set</a> to which this <a href="#">reference set member</a> belongs.  In this case, always ???   Edition composition reference set   as there is only one Module Dependency Reference set.	NO	NO
<a href="#">referencedComponentId</a>	<a href="#">SCTID</a>	A reference to the <a href="#">SNOMED CT component</a> to be included in the <a href="#">reference set</a> . This refers to the target of the composition (i.e. the root module that the Edition identified by <a href="#">moduleId</a> includes).	NO	NO

## ECRS Format

The MDRS has no additional columns beyond the minimal set for a Reference Set. It is a standard RF2 file and thus conforms to the normal rules for computing the Snapshot View. However, because Snapshot computation requires knowledge of the set of module-effectiveTime pairs, and that list of pairs depends on the ECRS (and MDRS) content there is a cyclic dependency. To avoid ambiguity we need to ensure that, like the MDRS, the content of the ECRS that is required for identifying these pairs is independently discoverable.

To determine the set of modules that comprise the content of an Edition,  $E$ , for a specific release date,  $T$ , the most recent rows where the  $moduleId == E$ , and the  $effectiveTime \leq T$ , are identified.

Then  $E$  and all the modules that are referenced components in *active* rows become the source module set. These module-effectiveTime pairs are then used in conjunction with the MDRS to determine all the transitively referenced module-effectiveTime pairs.

Looking again at the UK Edition example, the following two entries in the ECRS would be sufficient to indicate that, since 20040131, all three modules (and each of the modules they in turn depended on) made up the content of the UK Edition.

...	effectiveTime	active	moduleId	referencedComponentId
	20040131	1	999000031000000106  UK Edition	999000021000001108  UK drug extension ref set
	20040131	1	999000031000000106  UK Edition	999000021000000109  UK clinical extension ref set

## Design and consistency criteria

Only the owner of the moduleId should be able to make assertions about the set of modules that comprise the content of the Edition identified by the moduleId.

The moduleId column of the ECRS identifies the Edition module so that, by the rules of RF2, others cannot make assertions about Editions that are not theirs.

For consistency, it is invalid for an MDRS component identifier (i.e., the UUID) to be associated with more than one moduleId.

The complete set of module-version dependencies must be consistent.

Consider the case where the ECRS states that Edition *E* comprises modules *E* and *F*, and the MDRS says that E-20180131 depends on M-20180131 and F-20170731 depends on M-20170731.

When resolving the set of module-effectiveTime pairs for the Version E-20180131, one would end up with:

E-20180131, M-20180131, F-20170731, and M-20170731

Since it is not well-formed to include content from two differently-timestamped modules, the Version E-20180131 is itself not well-formed. Either the ECRS is wrong and would need to be updated to exclude module *F*, or (the more likely case) the MDRS is wrong and the content of module *F* would need to be updated to account for changes in the later module, M-20180131, and the MDRS dependencies also updated accordingly.

## MDRS changes

Two changes are proposed for the MDRS:

1. The first change is to remove the requirement that transitive dependencies are explicitly stated. For example, where module A explicitly depends on module B and module B explicitly depends on module C, the current MDRS requirement would require a row for the transitive A to C dependency. This change would remove that requirement. Note that if there was an explicit dependency from A to C, then that must still be explicitly recorded, even though it is transitively implied. Dependencies that are *only* transitively implied should not be explicitly recorded.
2. The second change is to remove the restriction on cyclic dependencies. In theory it is perfectly okay for modules to be mutually dependent (and thus they cannot be used

independently). So it should be possible to record this mutual dependency. Indeed, the *SNOMED CT model component module* and the *SNOMED CT core module* currently exhibit this mutual dependency because of 138875005 |SNOMED CT Concept|.

## URI Specification Impact

Section 2.1 of the URI Standard describes how the interpretation of a Version URI links to the dependencies specified in the MDRS to determine the computable content of a Version of an Edition.

This text would need to be updated to reference the combined use of the ECRS and the MDRS as well as indicating that, if ECRS content does not exist for a given module, then the MDRS alone is used. This provides for a backwards compatible interpretation of URIs and supports incremental adoption of the ECRS by organisations maintaining SNOMED CT extensions.

## Document History

Date	Version	Reason
2018/05/15	0.1	Initial draft
2018/05/28	0.2	Revised draft based on internal feedback