

- *PROPOSAL FOR PRESENTATION TO THE CONTENT MANAGERS ADVISORY GROUP*
- *REQUIREMENT TO BE FULFILLED PRIOR TO SUBMITTING A PROJECT PROPOSAL TO SENIOR MANAGEMENT TEAM*
- *SEEKING SUPPORT AND ENDORSEMENT OF THE FOLLOWING PROJECT:*

1. PROJECT IDENTIFICATION	
Project Name:	Addition of new attribute to link substances to the organisms from which they originate.
Description:	Create a new attribute for the substance hierarch labeled, something similar to, “has source organism” or “has biological origin” that links a substance concept to the organism it is sourced from.
Strategic Alignment	This aligns with the strategic direction of SNOMED CT to provide a more computable definition for each concept in the hierarchy. While this initial attribute may not be sufficient to provide a full definition for a substances, it is a valuable step and it will provide a computable association between biological substances (i.e. antigens, nucleic acids, venoms etc) back to the originating organism.

PROJECT INITIATION, PROBLEM STATEMENT AND/OR BACKGROUND
<p>Currently, there is no way to programmatically identify associations between Organisms (<410607006 Organism (organism)) and Substances (<105590001 Substance (substance)) concepts within SNOMED CT.</p> <ul style="list-style-type: none"> ● This limits the ability of terminology implementers and data analysts to leverage the terminology to appropriately identify the originating organism – for analytics or user experience. ● The absence of such an association is particularly limiting in the laboratory space. ● The “biological substances” are manually organized and completely independent of the organism hierarchy – raising the potential for inconsistencies between the two. ● The terms used for some substances may be internationally ambiguous, associating organism names can mitigate some of this risk. <p>Examples of the types of associations (and definitions) that are can be made through this proposal:</p> <ul style="list-style-type: none"> ● Taipan venom (substance): IS A Venom (substance) , has originating organism = Oxyuranus scutellatus (organism) ● Fish oil (substance): IS A Oil (substance), has originating organism = Fish (organism) ● Apple (substance): IS A Fruit (substance), has originating organism = Malus domestica (organism) ● Pork insulin (substance): IS A Insulin (substance), has originating organism = Sus scrofa scrofa (organism) ● RNA of Chikungunya virus (substance): IS A Ribonucleic acid (substance), has originating organism = Chikungunya virus (organism) ● Antigen of Ebolavirus (substance): IS A Antigen (substance), has originating organism = Genus Ebolavirus (organism) <p>Introduction of the association allows some substances to be sufficiently defined and classified, reduces ambiguity and allows the substance hierarchy to respond to changes to the organism hierarchy. (Which in turn may flow through to Procedures and Clinical Findings)</p> <p>Aside from some minor improvements to support Medicinal Products, the substance hierarchy has had little attention for decades and is well overdue. It is expected that as this proposal is actioned, further ambiguities and issues will be identified and documented for subsequent consideration.</p>

USE CASE(S) and SCENARIO supported
<ul style="list-style-type: none"> ● Geisinger Health System (U.S.) has provided a letter of support for the Australian use case as well as an additional use case in the same domain. ● State reportable laboratory test results that identify an organism based on DNA or RNA will return a result from the substance hierarchy. The CDC and the state want to receive the organism concept from which the DNA or RNA came to be the reported concept. When PCR and other MALDI-TOF type diagnostic tests are being employed, it requires manually mapping from the DNA/RNA concept

back to the organism concept in order to send the correct concept. Another option is to map the test result directly to the organism, bypassing the fact the test was for the DNA/RNA, but this creates a loss of granularity in the result value.

- Linking of the extracts and other edible substances back to their originating organism for Kingdom plantae and other organisms. For example:
 - Prior to the February 2025 edition the international concept 1268856000 |Theobroma cacao (substance)| was active and contained the synonym cocoa. This is an incorrect representation because 901000161107 |Cocoa (substance)| is the edible substance produced from the bean of 71950006 |Theobroma cacao (organism)| plant.
 - A similar issue occurred in Jan 2025 with 227439007|Paw-paw (substance)| being *replaced* with 735212003|Carica papaya (substance)| OR 1354825004|Asimina triloba (substance)|. Arguably this is still not quite right, as the substance is (presumably) really referring to the “fruit” of both these plants as per the previous example.
- Australia (CSIRO) has provided a documented use case for that need for this attribute, summarized simply as the absence of any relationship between these three concepts:
 - 28531000087107|Vaccine product against severe acute respiratory syndrome coronavirus 2 (medicinal product) |
 - 1240591000000102|Severe acute respiratory syndrome coronavirus 2 not detected (finding)|
 - 840539006|Disease caused by severe acute respiratory syndrome coronavirus 2 (disorder)|

Having such a link within the terminology use useful for analytics, and designing responsive clinical interfaces/decision support.

Project Objectives – Indicate the key objectives to be accomplished through a successful completion of the proposed project.

PROJECT OBJECTIVES
<ol style="list-style-type: none"> 1. Create an appropriately named attribute concept in SNOMED to represent a “source organism”. Associated MRCM rules to appropriately validate the attribute domain and range The domain constraint of at least <105590001 Substance (substance) (or potentially The range constraint will <410607006 Organism (organism) . 2. Create appropriate editorial guidance around the use of the attribute 3. Phased implementation of the addition of the attribute (described further in the procedure approach). 4. Improved quality throughout the terminology - Substances and dependent hierarchies 5. Improved implementation usefulness supporting programmatic associations between organisms, substances and findings/procedures.

Project Benefits

PROJECT BENEFITS:
<ol style="list-style-type: none"> 1. The immediate benefit is that a more reliable/predictable Substance hierarchy. Leveraging organisms – while not perfect, is mostly a reliable Linnean taxonomy – could sufficiently define a significant portion of content. Changes can be made in one place (the organism) and the improvements cascade through to substances and other top-level hierarchies that reference these substances. 2. Furthermore, it provides a link between the investigations (with the property “has component= some substance”) and diseases (“has causative agent = some organism”). Presently substances and organisms have no relationship so there’s no ontological association between <i>Syphilis (disorder)</i> and <i>Treponema pallidum nucleic acid assay (procedure)</i>. Such an association is analogous to that between Medicinal Products and Allergies. 3. An additional minor side benefit is the cross-cultural ambiguity of some substances, particularly foods (e.g. “Pepper”) can be resolved through sufficient modelling and updated FSNs.

Project Approach – Internal and/or external resources; Is collaborative authoring proposed?

PROJECT APPROACH

The number of affected concepts is large - potentially directly impacting around 14 thousand concepts (<115668003|Biological substance| OR <91720002|Body substance|). However, the change should be introduced incrementally, identifying tranches of work based of the existing aggregations, for example:

- <27380003 |Nucleic acid (substance)| (466 concept)
- < 116633006 |Microbial antigen (substance) | (790 concepts)
- < 256363008|Animal material (substance)| (338 concepts)
- < 67866001 |Insulin (substance)| (28 concepts)
- < 762766007 |Edible substance (substance)| (2323 concepts)
- And so on..

These examples can be broken down further as required, with specific tranches identified as part of the elaboration phase. This project can also help to clean up the high level concepts that are either ambiguous or questionable. For example the concepts 91720002|Body substance (substance)|, 413674002|Body material (substance)| and 115668003|Biological substance (substance)| - what do these mean exactly and how are they different?

Executing the change should be automated where possible with the following high level approach proposed:

1. For each tranche, the set of concepts identified should be mapped to a corresponding organism within SNOMED (using Snap2Snomed). Each mapping should be two one of two relationship types:
 2. "Exact" = The organism is the exact one from which the substance originates
E.g. 783431009| Ribosomal ribonucleic acid of Campylobacter (substance) | & 35408001| Genus Campylobacter (organism)|
 3. "Broader" = The organism is less specific (higher in linean taxomony) from which the substance originates
E.g. 735030001|Garlic (substance)| & 39293009| Allium (organism)|
Since SNOMED CT does not contain "*Allium sativum (organism)*".
4. Creation of an "Input Table" generate with fields
 - ConceptId = The source substance concept to be remodelled.
 - Parent - Ideally the "Parent" will be the concept from which the tranche was identified. For example, the parent for all the "Microbial antigens" will be 116633006| Microbial antigen|.
 - SourceOrganism – The target of the mapping exercise.
 - Definition Status – Defined/Primitive. Corresponding to the map types "Exact"/ "Broader" respectively.
 - Note: It is recognised that some concepts may currently have a "has disposition" property that will not be included in the defining axiom with this approach. However the property would be inferred from the stated parent.
 - For example "Mycotoxin" IS A Toxin, hasSourceOrganism some "Fungus".
 - The end result upon classification is identical.
 - The currently application of "has disposition" is unpredictable and inconsistent.
5. Scripted authoring – Creation and execution of a script to process the "Input Table". This script will leverage the Snowstorm API.
6. Classification – After authoring, the classification results should be reviewed. Of particular interest are changes to concepts outside the substance hierarchy. Substance concepts should classify as per the modelling of the organism hierarchy.

Resources: Australia is prepared to provide the resources to develop the associated transform script. At least 1 Australian Terminologist can contribute to the development of the "Input Table" (specifically the mapping exercise). Interest from other members (e.g. US) may provide collaboration partners. There is no expectation from SI staff to perform this work, beyond the necessary review processes.

Development of an Authoring Template can also be considered if required for ongoing terminology authoring.

It has already been identified that there are ambiguities and editorial issues within the substance content. In particular the use of "agent", "product" & "Material" are used throughout, but the exactly meaning of these concepts is debatable (though you can see how they are used). For example: What is the difference between 105899005| Animal agent (substance) | and 256363008| Animal material (substance)|. It may be worthwhile to

Project Scope- Indicate the scope of the project which includes specifically what is in scope for the project and also any relevant scope exclusions that may apply.

PROJECT SCOPE
<p>In Scope:</p> <ol style="list-style-type: none"> 1. Creation of domain attribute added to the substance hierarchy that has an attribute range value from the organism hierarchy. 2. Application of this new attribute to concepts within the nominal scope of “substances of biological origin” - (<115668003 Biological substance OR <91720002 Body substance (depending on 1) <p>Out of scope</p> <ol style="list-style-type: none"> 1. Substances that are not associated with an organism 2. Updating descriptions. Concepts will be taken at face value, and lineage will be based on existing parentage. 3. Addition of other attributes (disposition or modification). 4. Evaluation of existing high level groupers and tidy up. This may be a preliminary project before work commences in anger. 5. Antibody substances are out of scope. Though a future project should create an association between antibodies and antigens/organisms for similar reasons as described here.