Linking Substances to Organisms

- Key Challenges & Discussion Questions

Prepared from: "Linking substances to organisms.pptx" — April 2026 Business Meetings, Vienna, Austria

---

Background

This work stems from a member forum request to improve the consistency of modeling biological substances in SNOMED CT, specifically by providing explicit links from substances to organisms. Internal testing revealed several interconnected challenges that should be resolved in order to develop a sound logical model that can be implemented.

Key Challenges

1. Ambiguous Semantics of "Substance"

The current Editorial Guide (2026) defines the Substance hierarchy as a container for chemical constituents used in medicinal/non-medicinal products, allergies, adverse reactions, and clinical reporting. However, the formal definition from the 2014 Substance Redesign project group introduces precise ontological language (collectives, grains, compound of fractions) that is not uniformly applied.

Key tension: a substance's identity is maintained after division (divisibility), whereas dividing an organism or product typically changes the class of the resulting parts. This distinction is central to whether organism-derived materials (e.g., cannabis leaf, pollen, fruit) qualify as substances in the strict sense.

Core problem: Without a shared, operationally clear definition of "substance," it is impossible to determine what should be included in the hierarchy, what relationship attributes apply, or how to classify borderline entities like plant parts, extracts, and metabolites.

---

2. Choosing the Right Attribute for Organism Links

Three candidate attributes exist from different ontology sources:

Attribute	Source	Intended Meaning
`Source organism`	Relation Ontology / BFO	Biological association between a material entity and a taxon
`Derives from`	Relation Ontology / BFO	One material entity ceases and another begins, inheriting matter
`Source material`	IDMP	Material from which a substance is derived, defined by taxonomic and anatomical origin

It is unclear which attribute(s) best capture the range of substance–organism relationships (e.g., chemical compound extracted from a plant vs. a plant part vs. an animal metabolite vs. a product derived from a plant), and whether these relationships are **defining** or **additional** (non-defining).

---

3. Quality and Accuracy of AI-Generated Links

An AI mapping of 1,314 substances under `289958009 |Organic natural material (substance|` to organisms produced three categories of results (validated using cannabis as a test case):

- **Correctly exclusive mappings** (e.g., Cannabis resin, Cannabidivarin, Nabiximols → Cannabis sativa)
- **Non-exclusive mappings** — substances now confirmed in other organisms (e.g., CBD found in *Trema micrantha*, CBG in *Helichrysum umbraculigerum*, CBC in *Rhododendron anthopogon*)
- **Biologically incorrect mappings** — animal metabolites misclassified as plant substances (e.g., 11-hydroxy-delta-9-THC and THC-COOH are liver metabolites, not plant constituents)

This illustrates that organism links require expert biological review and cannot rely on AI alone.

---

4. Naming and Descriptions for Organism-Derived Substances

When the preferred term of a substance could duplicate an organism concept without a semantic tag, four description patterns are under consideration:

Pattern	Example contrasting |Cannabis (organism)|
X (substance) — no change	Cannabis (substance)
X material	Cannabis material (substance)
X substance (word substance always present)	Cannabis substance (substance)
Substance derived from X	Substance derived from Cannabis
Substance with <organism> source	Substance with Cannabis source

Existing FSN guidance (organism scientific name) and PT guidance (common names) would still apply, but need to be reconciled with the chosen pattern.

---

5. Dispositions, Roles, and Problematic Groupers

Classifying substances by role or disposition (e.g., hallucinogenic agent, psychoactive substance, cannabinoid receptor agonist) creates overlapping hierarchies and logical inconsistencies:

- `Cannabis (substance)` currently appears under both material-based groupers (`Plant material`, `Organic natural material`) and disposition-based groupers (`Hallucinogenic agent`, `Psychoactive substance`).
- If a new grouper `Substance from Cannabis` is created, its placement is ambiguous: under `Material`? `Biological substance`? `Organic natural material`?
- Cannabis oil and extracts may not be purely "organic natural material" if they include synthetic analogues or processing agents.
- It is unclear whether a substance group should inherit the disposition of one of its subtypes (e.g., should all cannabis substances be modeled as hallucinogenic, even non-intoxicating CBD?)
- `Cannabis (substance)` may need inactivation and replacement with a more specific concept such as `Cannabis with cannabinoid receptor agonist mechanism of action (substance)`

---

6. Classifying Diverse Substance Types from a Single Organism

The cannabis example alone yields at least five distinct substance types, each may require a different modeling approach:

Substance Type	Example	Proposed Modeling Pattern
Chemical compound	Tetrahydrocannabinol (THC), CBD	Direct chemical definition
Substance from plant	Substance from Cannabis	source material some Cannabis (organism)

Constitutional/regional part	Cannabis leaf	source material some (constitutional part of some Cannabis (organism))
Secretion/produced by organism	Cannabis resin	source material some (secretion produced by some Cannabis (organism))
Substance of product from plant	Cannabis extract, cannabis oil	source material some (Cannabis extract (physical object) and has substance some substance from Cannabis)

There is currently no consensus on whether `substance from product` should be a subtype of `substance from plant`, given that products may contain synthetic substances not found in the source organism.

---

7. SEP Model Issues and the "Clay Statue" Problem

For substances like pollen, which can be modeled as both a structural entity (pollen as a botanical structure) and a material entity (pollen as a substance), there is potential conflict with the SEP (Structure–Entire–Part) model. The "clay statue" issue arises: is pollen structure the same entity as pollen substance, or are they distinct things that co-occupy the same space? This affects whether new concepts are needed and how they relate.

---

8. Organism Hierarchy Considerations

Organisms themselves present modeling challenges that propagate into substance links:

- Taxonomic names and classifications change over time (nomenclature instability)
- Grouper concepts exist outside standard taxonomic ranks
- Common names need to be assigned consistently

- Substances may be linked at different taxonomic levels (genus vs. species), creating inconsistency (e.g., cannabis parts linked to `Cannabis (organism)` at genus level vs. specific compounds linked to `Cannabis sativa` at species level)

---

9. Scope and Business Case Clarity

Substances are tightly coupled to other high-level groupers and used across multiple hierarchies (disorders, findings, products, procedures, observables). Any change to substance modeling has cascading implications across SNOMED CT. A clear scope definition and detailed business cases are needed to ensure the smooth progress of the project.

Discussion Questions

Appraisal of the business case project document

On Substance Semantics

Q1. Should the 2014 Substance Redesign definition (collective / compound of collectives) be formally adopted as the operational definition in the Editorial Guide? If so, which borderline cases (plant parts, metabolites, extracts) does this include or exclude?
This is noted in the project charter partially. The proposal focuses on biological substances, metabolites, and extracts. Borderline cases like Fish oil, Apple, and Cocoa are used as examples of associations the project aims to create.

Q2. How should the boundary between a physical object and a substance (e.g., a cannabis oil vs. cannabis substance) be drawn in practice?
The proposal partially addresses this issue by using examples like Theobroma cacao (substance) vs. Cocoa (substance) (the edible substance from the bean), highlighting the distinction that needs to be modeled. That said, it doesn't answer the specific question.

Q3. Does divisibility — the property that each divided portion retains the same class — serve as a reliable test to distinguish substances from organisms and products in all cases?
This specific philosophical aspect is not addressed in the proposal.

On Attribute Selection

Q4. Which attribute — Source organism, Derives from, or Source material — best represents the relationship between a biological substance and its organism of origin? Should more than one be used, and if so, under what conditions?
The project charter is not clear on this one. It proposes one attribute: "something similar to, 'has source organism' or 'has biological origin'". It does not compare it against other options like "Derives from".

Q5. Should organism links be defining (part of the necessary and sufficient conditions) or additional (non-defining)? What are the downstream reasoning implications of each choice?
The project charter answers the first question. It uses "Exact" (corresponding to Defined status) and "Broader" (corresponding to Primitive status) mappings, which relates to defining vs. primitive links. The additional relationship (non-defining) was not considered. If a concept can be sufficiently defined, would it have complete classification of all substances. For example, cannabis substance, would it be possible to classify all substances that can be originated from the Cannabis organism? The impact (including potential inferences and/or maintenance of intermediate parents) is not accessed.

Q6. Where a single substance occurs in multiple organisms (e.g., CBD in Cannabis sativa and Trema micrantha), how should this be modeled: multiple relationship values, a class-level organism link, or a higher-level taxon?
Partial answer - The approach allows for "Broader" mapping.

On Descriptions and Naming

Q7. Which description pattern should be adopted for organism-derived substances: X (substance), X material, X substance, or Substance derived from X? What are the display, search, and disambiguation implications of each?
Partially answered - Objective 2 is to "Create appropriate editorial guidance around the use of the attribute". A side benefit is resolving ambiguity and updating FSN based on the naming of the attribute for these groupers.

Q8. Should FSNs for organism-derived substances follow the organism's scientific name (aligning with organism FSN conventions), or should a different rule apply?
Not specifically addressed, though it mentions updating FSNs.

On Groupers and Dispositions

Q9. Where should a new grouper concept like |Substance from Cannabis| sit in the hierarchy — under Material (substance), Biological substance (substance), or Organic natural material (substance)? What criteria should determine placement?
Further clarification needed. The project scope is defined nominally by 115668003 |Biological substance (substance)|.

Q10. Should disposition-based groupers (e.g., Hallucinogenic agent) be modeled as necessary conditions for substances that have that disposition, or only as additional classification? How should a substance that has a disposition only at high doses (e.g., cannabis) be handled?
Further clarification needed. The proposal states that "Addition of other attributes (disposition or modification)" is Out of scope and notes the current "has disposition" property is "unpredictable and inconsistent". Also refers to inference of the disposition for the children, which poses the issue raised in this presentation for some of Cannabis existing children.

Q11. Should Cannabis (substance) be inactivated and replaced with a disposition-specific concept? What replacement concept(s) would be needed, and how would existing downstream content (e.g., Psychotic disorder caused by cannabis) be managed?
This is not addressed in the proposal.

Q12. Should substance from product be a subtype of substance from plant when the product may also contain synthetic substances? How should the Open World Assumption be applied here?
Partially answered - The use case includes linking "extracts and other edible substances back to their originating organism". It does not address synthetic substances or the Open World Assumption.

On Biological Accuracy and AI-Generated Links

Q13. What governance process should be established to validate AI-generated substance–organism links before they are published? Who has the domain expertise to review across all organism types?
This is not addressed in the proposal.

Q14. How should substances that are animal or human metabolites (e.g., 11-hydroxy-delta-9-THC, THC-COOH) be distinguished from primary plant-derived substances in the model? Should a new attribute or parent concept be introduced?
This is not addressed in the proposal. Although, the scope includes 91720002 |Body substance (substance)| (and therefore, 767279008 |Metabolite (substance)|).

Q15. When a substance is found in a new organism after initial modeling (e.g., CBD now confirmed in Trema species), what is the editorial change management process — update the attribute, add a new concept, or both?
Not specifically answered - The project aims for a system that allows the substance hierarchy to "respond to changes to the organism hierarchy". An "Authoring Template" for ongoing work is also considered.

On Specific Examples and Broader Application

Q16. For pollen: should Pollen (substance) and Pollen structure (substance) be distinct concepts, and how does the SEP model apply? Does the "clay statue" problem require resolution before pollen substances and pollen can be properly represented?
Not addressed in the proposal.

Q17. For blueberries (and similar edible substances): should the approach be to (a) adjust the parent hierarchy of the existing concept, (b) rename the existing concept to Blueberry fruit and add a new grouper, or (c) retire the existing concept and create two distinct concepts? What factors should determine this choice for other fruit/food substances?
Partially - The proposal highlights the issue for edible substances (cocoa, paw-paw) but defers the specific approach.

Q18. How should the modeling of organism-derived substances be prioritized — by clinical use frequency, by member request, by organism kingdom (e.g., plants first), or another criterion?

Partially - The project approach suggests prioritization will be done by tranches based on existing concept aggregations (e.g., Nucleic acid, Microbial antigen, Edible substance).

Q19. What is the intended scope of organism links — all substances in the Substance hierarchy, only those under Organic natural material, or a defined subset? What business cases justify the scope boundary?
This is answered, but further clarification might be required - The scope is explicitly defined and the business cases are detailed under "PROJECT INITIATION" and "USE CASE(S)".

On Cross-Hierarchy Impact

Q20. How should changes to the Substance hierarchy be coordinated with dependent hierarchies (Clinical findings, Pharmaceutical products, Procedures) to avoid breaking existing classification, maps and reference sets?
Partially - The project acknowledges the need for review of classification results, specifically "changes to concepts outside the substance hierarchy".

Q21. Should there be a formal impact assessment process before any substance grouper is inactivated or restructured?
Not part of proposal - The evaluation of groupers is out of scope; however, the need for a review of cross-hierarchy impacts (Q20) implies a form of impact assessment.