Can SNOMED CT be harmonized with an upper-level ontology?

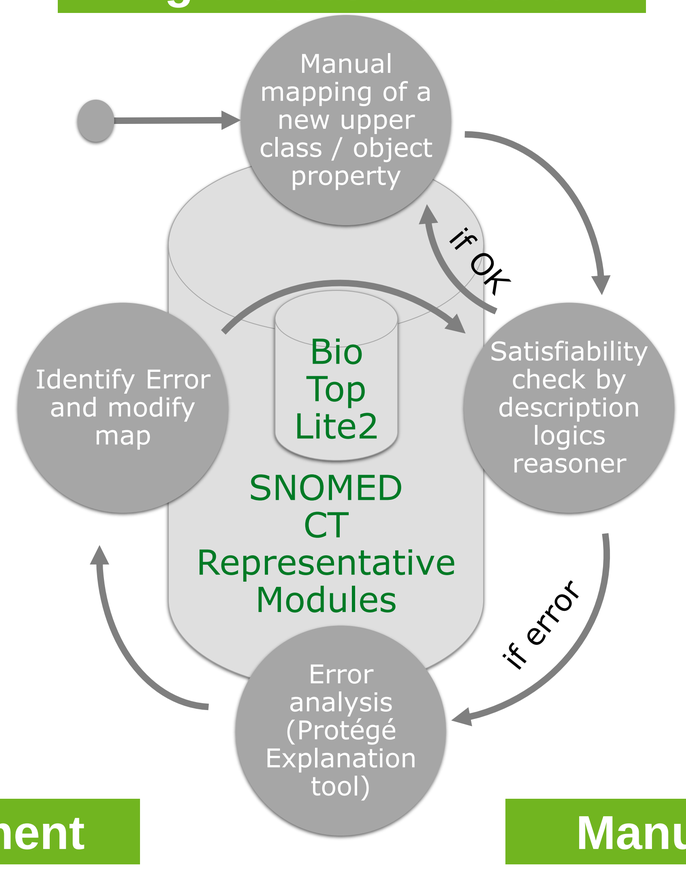
Stefan Schulz, Catalina Martínez-Costa

Medical University of Graz

Context

Many **ontologies** have been developed bottom-up in different contexts. They do not share an joint upper-level. Key terms (e.g. disorder, animal, drug, situation, condition) have different meanings and often lack explicit definitions. Alignment of lexical labels does not guarantee alignment of meaning. It is sensible to assume that **interoperability** between semantic artefacts is facilitated by (i) a well-understood and well performing representational language; and (ii) by a top-level layer of shared categories and relations (**upper level ontology**).

Alignment / Validation



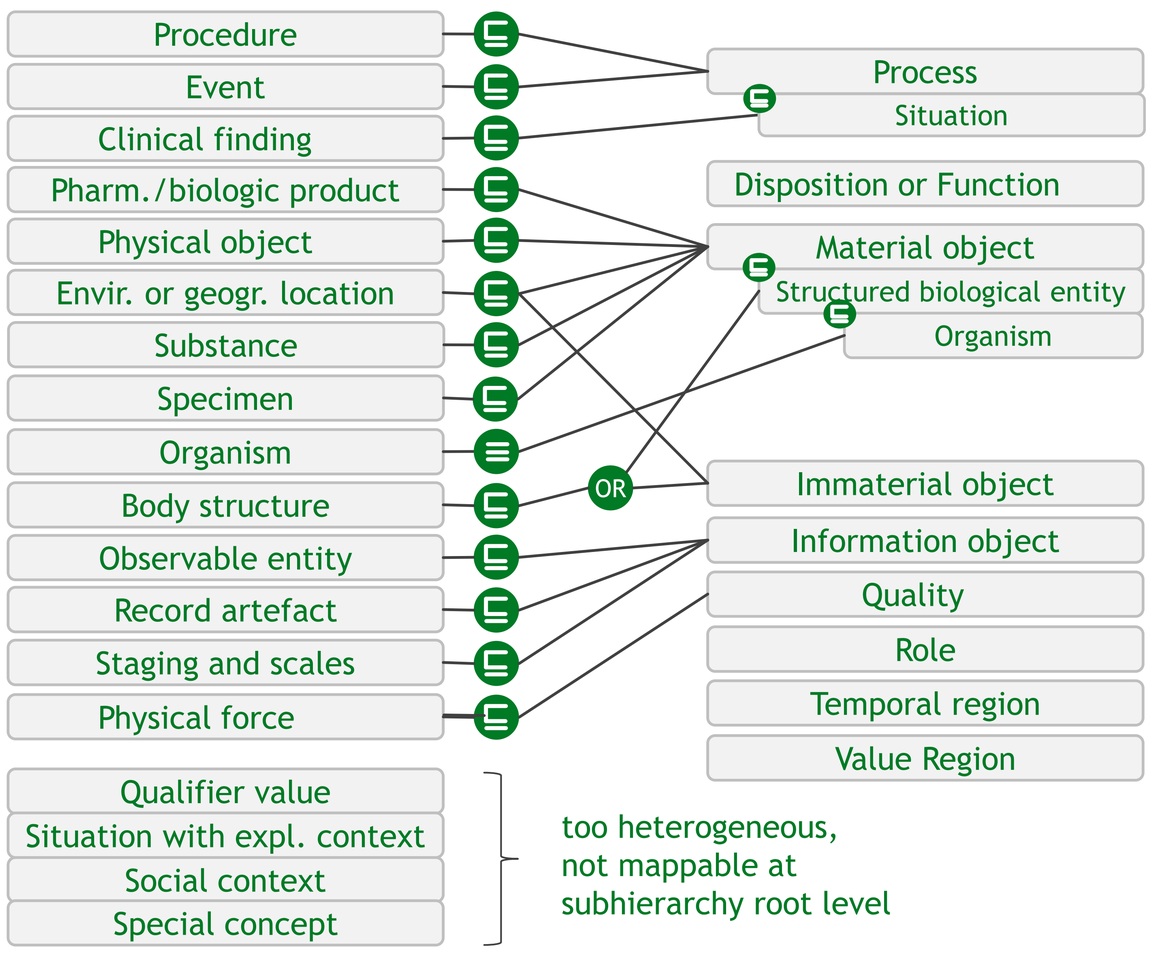
Goal

The goal is to analyse the ontological structure of the **OWL version** of **SNOMED CT** (generated by PERL script) in terms of

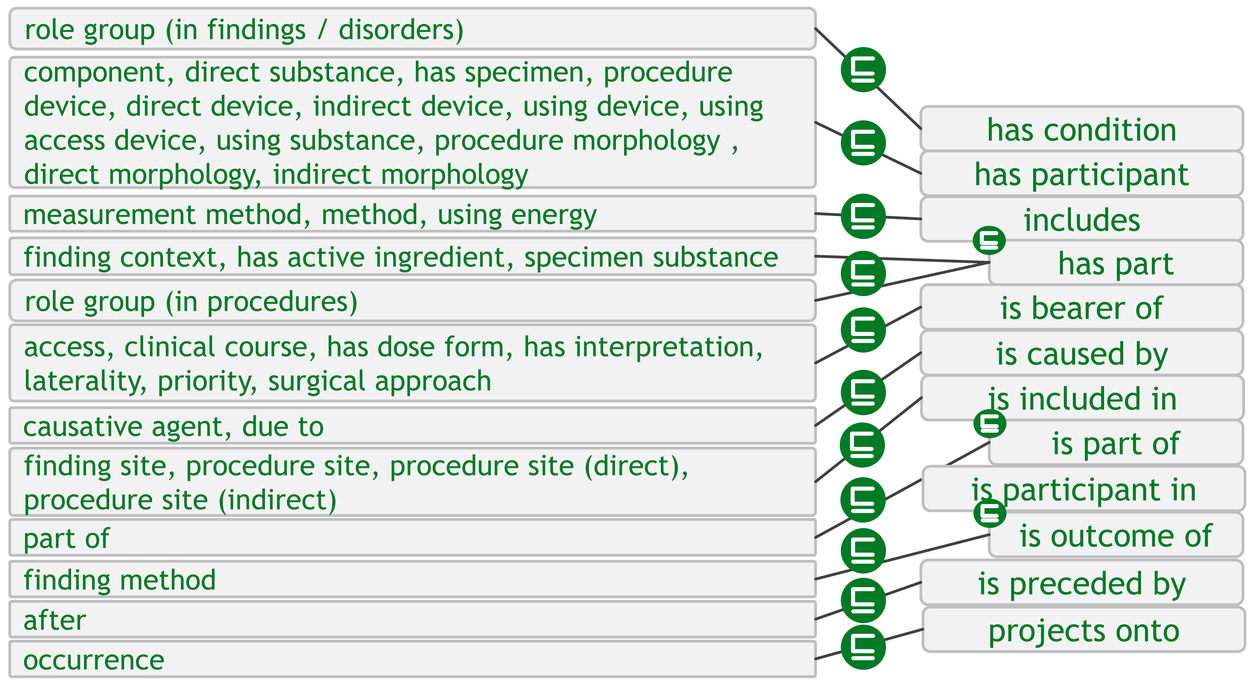
- Upper level concepts (classes)
- Relations (object properties)
- Constraints (axioms)

A preliminary manual alignment is done with the Upper-Level Ontology **BioTopLite2**. Consistency and performance are checked and the feasibility of moving to a richer language are assessed.

Manual upper level class alignment



Manual upper level relation alignment



Exercise limited to most frequent relations, accounting for 95% of all relational statements in SNOMED CT

Results

- **Class alignments** possible for all subhierarchy roots but four that were heterogeneous and semantically shallow. All but one used subclass axioms.
- **Relation alignments**: complex, still ongoing, with all relations but one using subproperty statements.
  - Many abstractions (e.g., 'finding site' → 'is included in') would be sufficient for fully defining concepts
  - Some relation abstraction are lossy (e.g. 'has active ingredient' as a subproperty of 'has part')
  - The RoleGroup relation had to be split in several subrelations (e.g. 'has condition', 'has part'…)
  - A few relations are rather complex, e.g. 'has focus', which expresses intentionality
- **Classification time** of maximally diverse modules with approx. 11,000 concepts: max 15 min (Fact++ on high performance laptop)
- Debugging of insatisfiable classes time consuming

Conclusions

- Feasibility study on aligning SNOMED CT with a principled, highly axiomatised upper level ontology shows value but also complexity
- Reveals ontological shallowness of some hierarchies
- Shows route towards simplification of SNOMED CT relations
- Relevant in light of current redesign discussions in the IHTSDO Modeling Advisory Group
- OWL-DL expressiveness allows reasonable classification performance only with small modules.

Poster presented at SNOMED CT Expo 2015, Montevideo, Uruguay, October 29 – 10, 2015.