

Aligning the Top Level of SNOMED-CT with Basic Formal Ontology

William R. Hogan, MD, MS

Background

- Translational research involves computer processing of large datasets
- Datasets will be collected by different researchers at different times at different locations
- Need to annotate data with controlled, structured vocabularies known as ontologies to enable algorithmic processing
- Ontologies used to annotate genotypes must be compatible with ontologies used to annotate phenotypes
- Leading candidate phenotype ontology—SNOMED-CT—is not compatible with the Gene Ontology (GO) and other bioinformatics ontologies
- GOAL: align top-level of SNOMED-CT with Basic Formal Ontology (BFO), the same top-level ontology with which GO is compatible

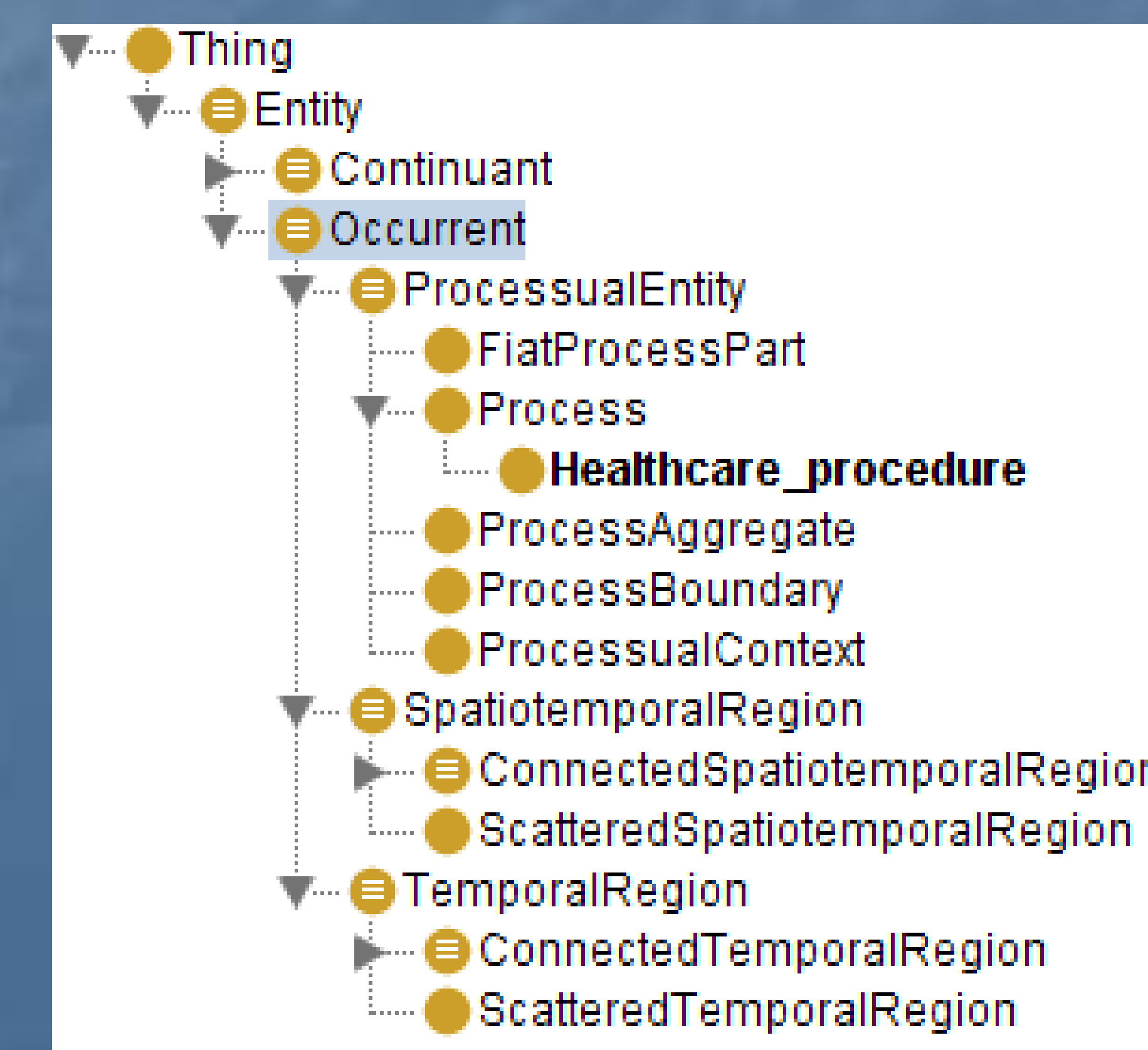
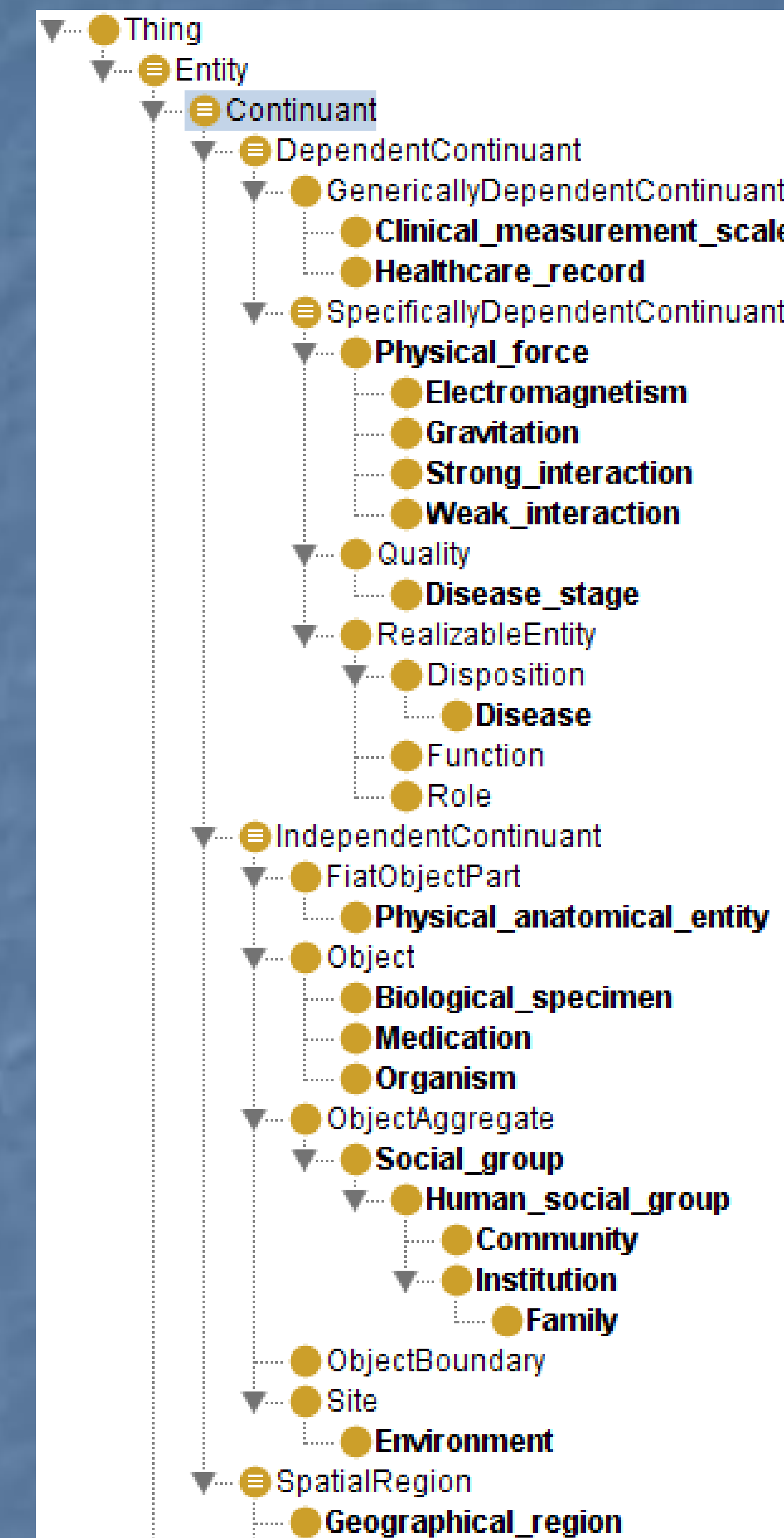
Methods

- Materials
 - SNOMED-CT, January, 2008
 - Basic Formal Ontology version 1.1 OWL file
- Procedure
 - Review SNOMED-CT User Guide
 - Review 19 top-level concepts and their children
 - Place each of the 19 top-level concepts and ten key children (so far) into BFO
 - Text definitions for new terms
 - Assign SNOMED-CT Concept Ids as appropriate
 - Fill in missing terms in hierarchy

Table: Final placement of SNOMED-CT concepts

Concept	Placement
Body structure	<i>is_a</i> FiatObjectPart
Clinical finding	Deprecated
Disease	<i>is_a</i> Disposition
Environment or geographical region	Environment <i>is_a</i> Site Geographical region <i>is_a</i> Spatial Region
Event	<i>equals</i> Process Boundary
Linkage concept	Deprecated
Observable entity	<i>equals</i> Specifically Dependent Continuant
Function	<i>equals</i> Function
Process	<i>equals</i> Process
Organism	<i>is_a</i> Object
Pharmaceutical / biologic product	Medication <i>is_a</i> Object
Physical force	<i>is_a</i> Specifically Dependent Continuant
Electromagnetism	<i>is_a</i> Physical force
Gravitation	<i>is_a</i> Physical force
Physical object	<i>equals</i> Object
Procedure	<i>is_a</i> Process
Qualifier value	<i>equals</i> Dependent Continuant
Record artifact	<i>is_a</i> Generically Dependent Continuant
Situation with explicit context	Deprecated
Social context	Deprecated
Community	<i>is_a</i> Human social group
Family	<i>is_a</i> Institution
Institution	<i>is_a</i> Human social group
Special concept	Deprecated
Specimen	<i>is_a</i> Object
Staging and scales	Disease stage <i>is_a</i> Quality Clinical measurement scale <i>is_a</i> Generically Dependent Continuant
Substance	<i>equals</i> Object

Extended Basic Formal Ontology



Results

- 31 terms aligned with BFO
 - 19 top-level terms
 - 8 of their descendants
 - 4 terms from split of two top-level terms
- Destinations
 - 7 equated with existing BFO terms
 - 17 placed as children of BFO or one another
 - 7 terms deprecated
- Added two fundamental forces of nature and two ancestor terms for institution for completeness
- Overall: 27 terms -> 31 terms -> 35 terms -> 30 aligned with BFO + 5 that cannot be aligned (and thus were deprecated)

Conclusions

- Even at the top level, SNOMED-CT has:
 - Ambiguous terms
 - Terms that do not refer to any entity
 - Epistemology
 - Arbitrary logical combinations
- Alignment with BFO:
 - Identifies and helps remove mistakes
 - Builds out a logically coherent hierarchy
 - Harmonizes with top-level of Gene Ontology
- Once a concept has been placed into BFO, its children do NOT necessarily follow
- Final result is available as OWL file