THE BOLZANO SUMMER OF KNOWLEDGE

2020 and 2021

ICBO / FOIS Keynote:

SNOMED CT x BFO: can the gap between legacy terminology and foundational ontology be bridged?

Stefan Schulz

Medical University of Graz (Austria)
Averbis GmbH (Freiburg im Breisgau, Germany)
Member of Modelling Advisory Group of SNOMED International

September 16, 2021

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

[2] Health/disease-related phenomena with [1] as participants or as locations

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

[2] Health/disease-related phenomena with [1] as participants or as locations

[3] Carers participating in [2] and their roles

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

[2] Health/disease-related phenomena with [1] as participants or as locations

[3] Carers participating in [2] and their roles

[4] External objects that participate in [2]

SNOMED CT – a large clinical terminology

What does it represent?

[1] Humans, their material and immaterial constituents, qualities, functions etc.

[2] Health/disease-related phenomena with [1] as participants or as locations

[3] Carers participating in [2] and their roles

[4] External objects that participate in [2]

[5] Observations, interpretations, actions and plans made by [3]

SNOMED CT – a large clinical terminology

What does it represent?

ウイルス

فايروس

вірус

U07.1

イルス

ριό

病毒

[1] Humans, their material and immaterial constituents, qualities, functions etc.

[2] Health/disease-related phenomena with [1] as participants or as locations

[3] Carers participating in [2] and their roles

[4] External objects that participate in [2]

[5] Observations, interpretations, actions and plans made by [3]

[6] The languages in which [1] and [3] communicate (including codes)

[7] Health records and their components describing [1]...[5] using [6]

SNOMED CT's ambition is to provide codes, terms and formal descriptions

SNOMED CT in a nutshell

- Merger of a large, informal clinical term collection (UK) with pre-existing huge "clinical nomenclature" (2002)

- International standard since 2007, maintained by SNOMED International, introduced in 40 countries

- Created and maintained by clinicians and computer scientists (not by ontologists)

- Addition of description-logics-like semantics, later clear commitment to OWL-EL

- Crisp separation of "conceptual" layer and "terminological layer"

- Currently around 350,000 concepts

- More than one million axioms

- Divided into 17 class hierarchies

SNOMED CT in a nutshell

Body structure ....
　　　　Morphological Abnormality .....
　　　　Cell Structure ....
Clinical finding ....
　　　　　Disorder ....
Environment or geographical location ....
Event ....
Observable entity...
Organism ....
Pharmaceutical / biologic product ....
　　　　medicinal product form ....
　　　　medicinal product ....
　　　　　　clinical drug ....
Physical object ....
Physical force ....
Procedure ....
Qualifier value ....
　　　　Administration method ....
　　　　Disposition ....
　　　　Dose form ....
　　　　Property ....
　　　　Product name ....
　　　　Process ....
　　　　Unit of presentation ....
Record artifact ....
Situation with explicit context ....
Social context ....
　　　　Ethnic group ....
　　　　Religion / philosophy ....
　　　　Occupation ....
　　　　Person ....
　　　　Life style ....
Specimen ....
Staging and scales ....
　　　　Tumor staging ....
Substance ....

ウイルス

病毒

ριό

イルス

- Merger of a large, informal clinical term collection (UK) with pre-existing huge "clinical nomenclature" (2002)

- International standard since 2007, maintained by SNOMED International, introduced in 40 countries

- Created and maintained by clinicians and computer scientists (not by ontologists)

- Addition of description-logics-like semantics, later clear commitment to OWL-EL

- Crisp separation of "conceptual" layer and "terminological layer"

- Currently around 350,000 concepts

- More than one million axioms

- Divided into 17 class hierarchies

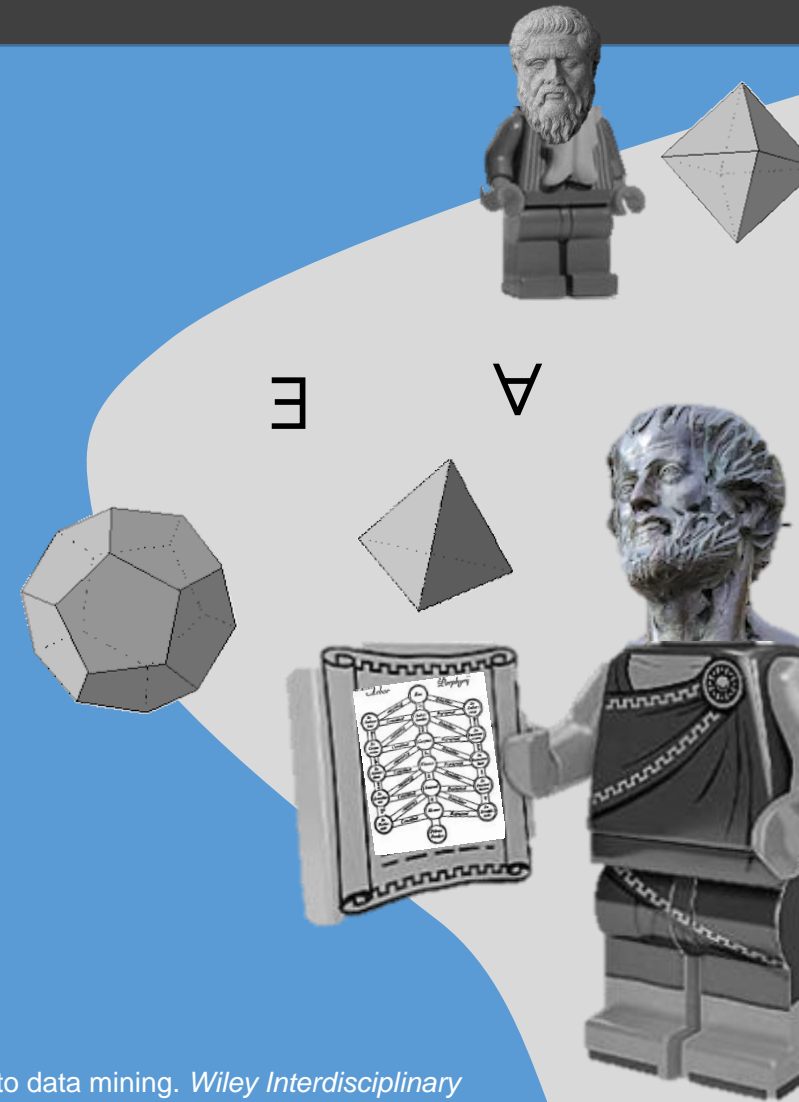# Domain ontology vs. Foundational Ontology

What a foundational ontology represents

- Foundational ontology = Upper level ontology
- "...philosophically well-founded axiomatic domain-independent categories and their ties (e.g., objects, events, causality, parthood, spatial–temporal connections, dependencies, etc.) that can be used to articulate the representation of phenomena in different material domains" *
- has "no overlap with domain ontologies"**
- combines "mathematical logic with philosophy to produce the most general abstraction" ***

* Amaral, G., Baião, F., & Guizzardi, G. (2021). Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1408.
** Smith, B. Information Artifact Ontology: General Background.   https://stids.c4i.gmu.edu/papers/STIDSPresentations/STIDS2013_Tutorial1_p1_Smith.pdf
*** Cummings, J., & Stacey, D. Lean Ontology Development: An Ontology Development Paradigm based on Continuous Innovation. KEOD. 2018.

BFO in an nutshell

BFO, BFO 1
- Very small upper-level ontology for science
- Single hierarchy of types (universals), with free-text definitions / elucidations
- Created by philosophers and logicians
- Continuant – occurrent dichotomy

BFO 2
- Addition of binary and ternary relations
- Addition of axioms in FOL
- Version in Common Logic
- Version in OWL-DL (approximated)
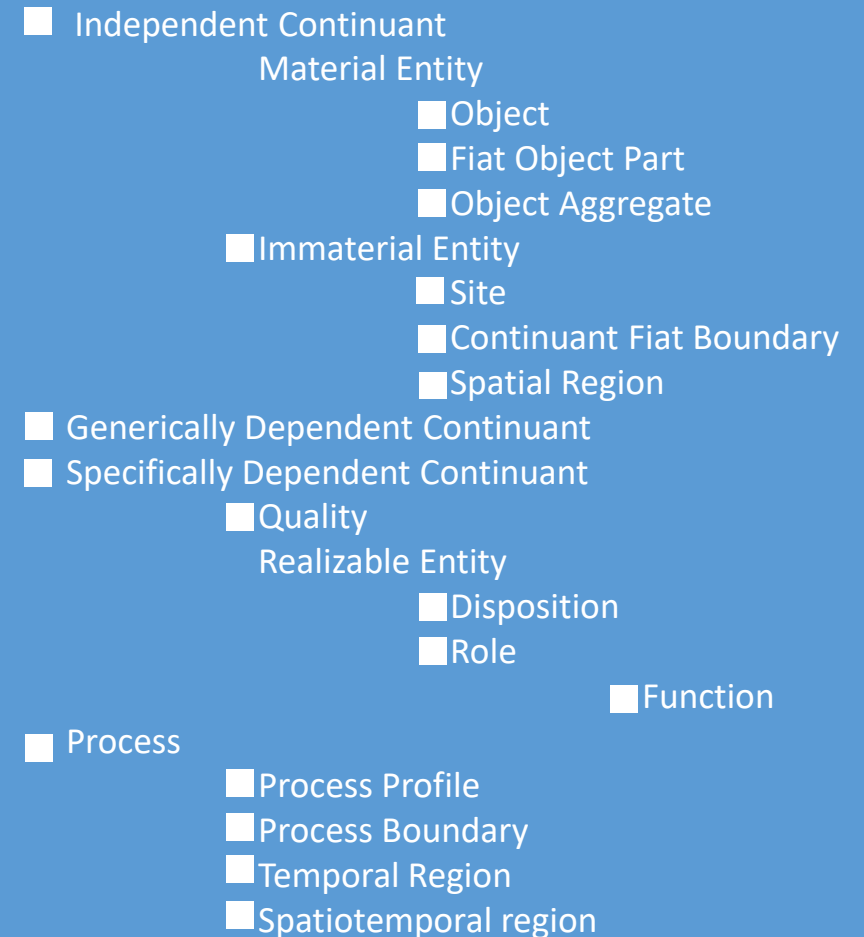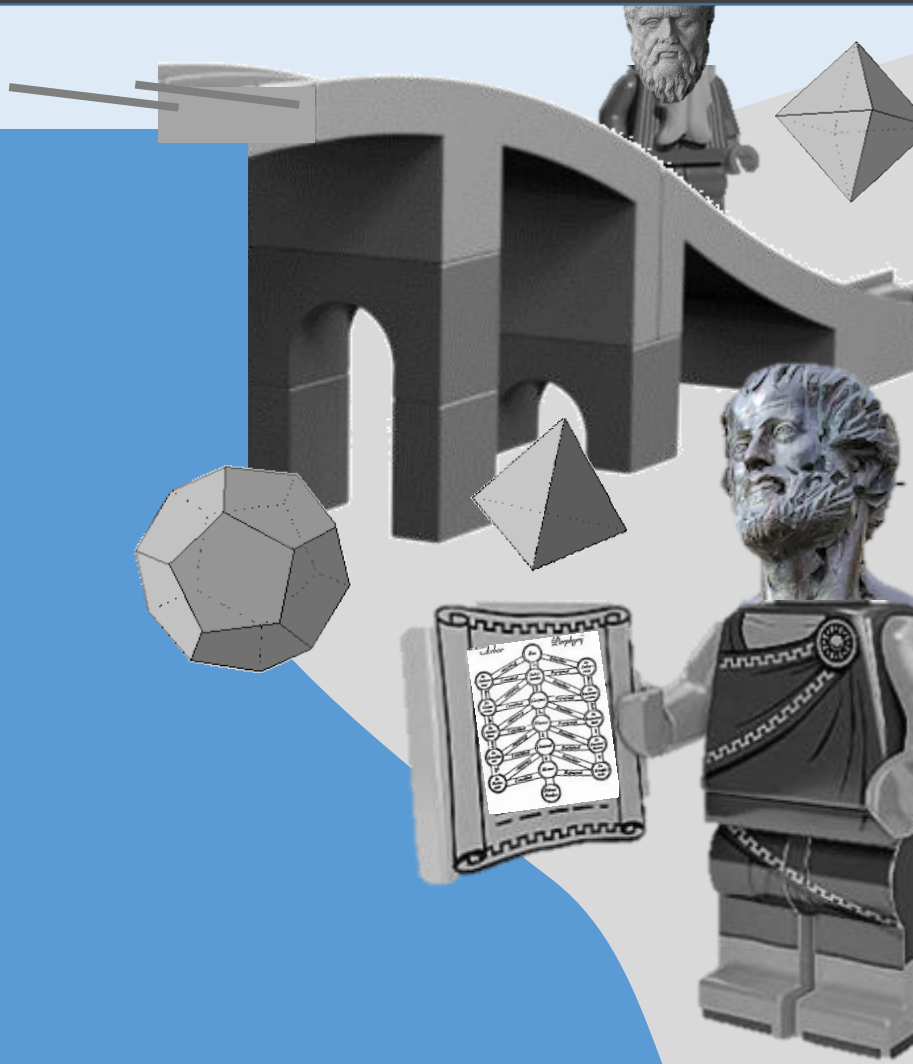- Towards ISO standardisation (currently 60%)

BFO in an nutshell

BFO, BFO 1
- Very small upper-level ontology for science
- Single hierarchy of types (universals), with free-text definitions / elucidations
- Created by philosophers and logicians
- Continuant – occurrent dichotomy

BFO 2

- Addition of binary and ternary relations
- Addition of axioms in FOL
- Version in Common Logic
- Version in OWL-DL (approximated)
- Towards ISO standardisation (currently 60%)

- ■ Independent Continuant
  - Material Entity
    - ■ Object
    - ■ Fiat Object Part
    - ■ Object Aggregate
  - ■ Immaterial Entity
    - ■ Site
    - ■ Continuant Fiat Boundary
    - ■ Spatial Region
- ■ Generically Dependent Continuant
- ■ Specifically Dependent Continuant
  - ■ Quality
  - Realizable Entity
    - ■ Disposition
    - ■ Role
      - ■ Function
- ■ Process
  - ■ Process Profile
  - ■ Process Boundary
  - ■ Temporal Region
  - ■ Spatiotemporal region

(simplified)

Driver of this work. interest in bridging between two standards

What's the problem

- SNOMED CT's use case:
  - Interoperable descriptions of all content of health records
  - No claim to "represent the world"
  - Major use case: content retrieval in electronic health records
- Architecture
  - Upper level consolidated, but not informed by ontology
  - Textual descriptions of upper-level concepts lack precision
  - Modelling guided by set of domain/range constraints ("SNOMED concept model")
- Use of description logics OWL EL aiming at:
  - Guided content creation and maintenance
  - Supporting plausible entailments between propositions
    - "coronavirus" entails "virus"
    - "injury of toe" entails "injury at lower extremity"
    - but "amputation of toe" does not entail "amputation of lower extremity"
    - "combined fracture of forearm" entails "fracture of ulna" and "fracture of radius"

Perception of foundational ontology by terminologists

- The current SNOMED CT upper level
  - It works, so it's sufficient - why invest effort for upper-level alignment?
  - People are used to how it's structured, complete redesign unfeasible

- Continuants / occurrent dichotomy seen as superfluous: "diseases are both"

- Medical "concepts" ill-defined and short-lived
  - (e.g., Rheumatoid Arthritis now is different from what it meant 10 years ago)
  - Some diseases are seen more as social constructs than crisp biological entities

- Medicine works well with fuzzy language:
  - e.g. in medical practice, "diagnosis", "disorder" and "disease" seen as synonyms

- Additional fuzziness when translating in other languages
  - e.g. "symptom" means also "sign"
  - e.g. "finding" means (i) sth. in the patient, (ii) result of an observation, (iii) a document
  - e.g. "biopsy" means (i) medical procedure, (ii) specimen obtained from that procedure

Harmonizing SNOMED CT with BFO

- Pilot study
  - 2 – 3 hierarchical levels (SNOMED CT, covering all semantic tags)
    3 hierarchical levels (BFO)
- Analysing
  - Text definitions / elucidations of upper-level concepts
  - Extension (descendant concepts)
  - Patterns of formal definitions
- First high-level alignment
  - Manually, approximative

BFO
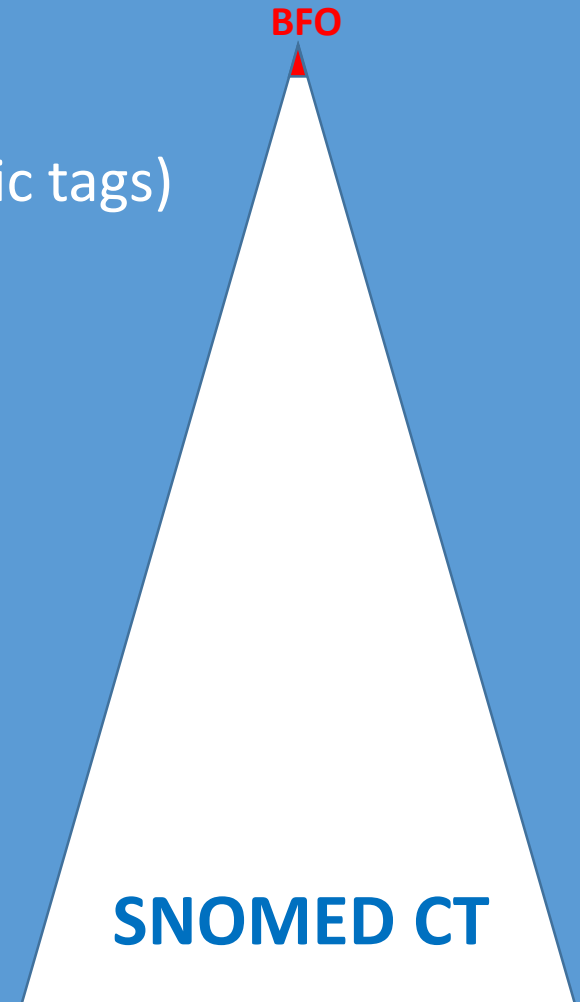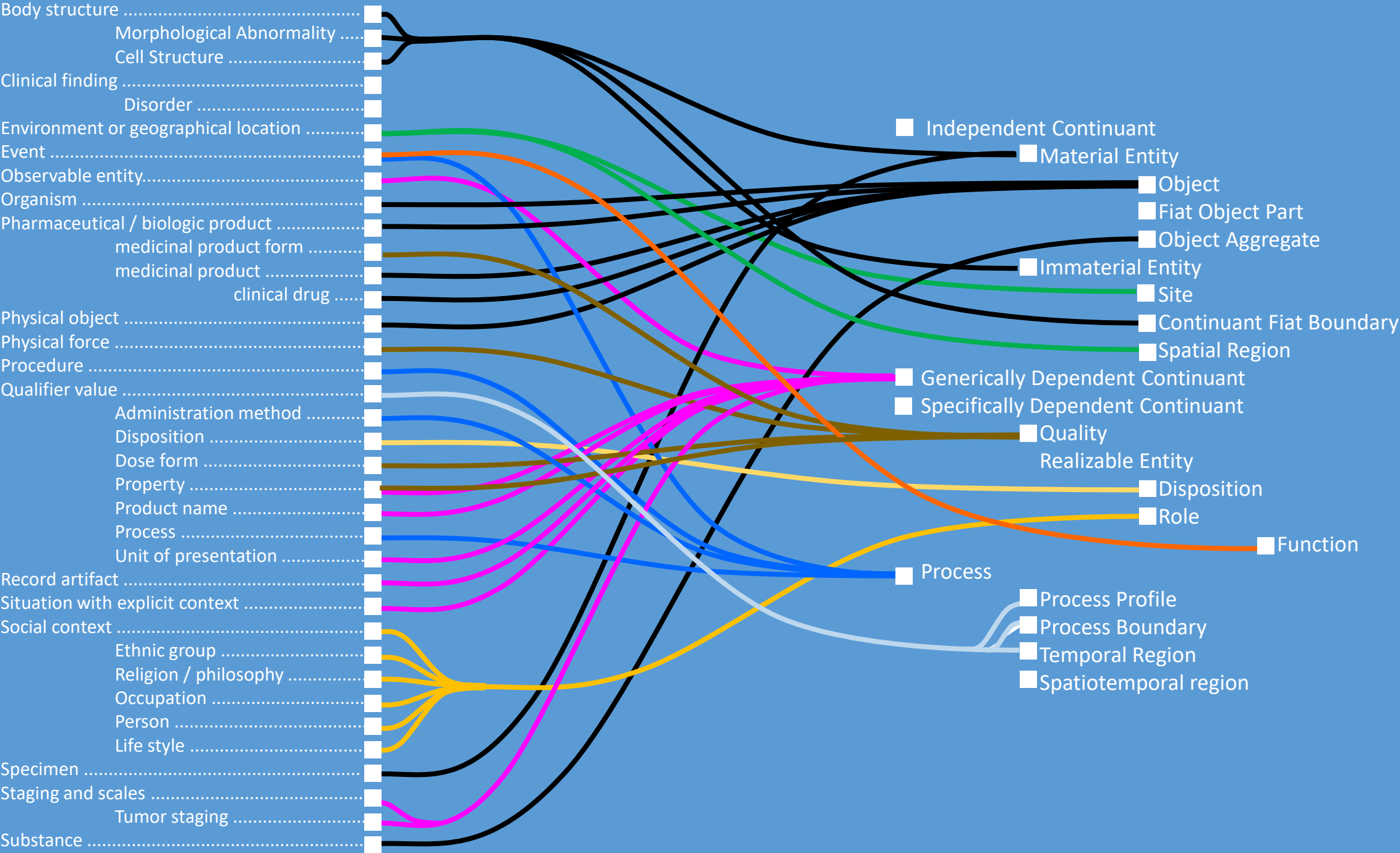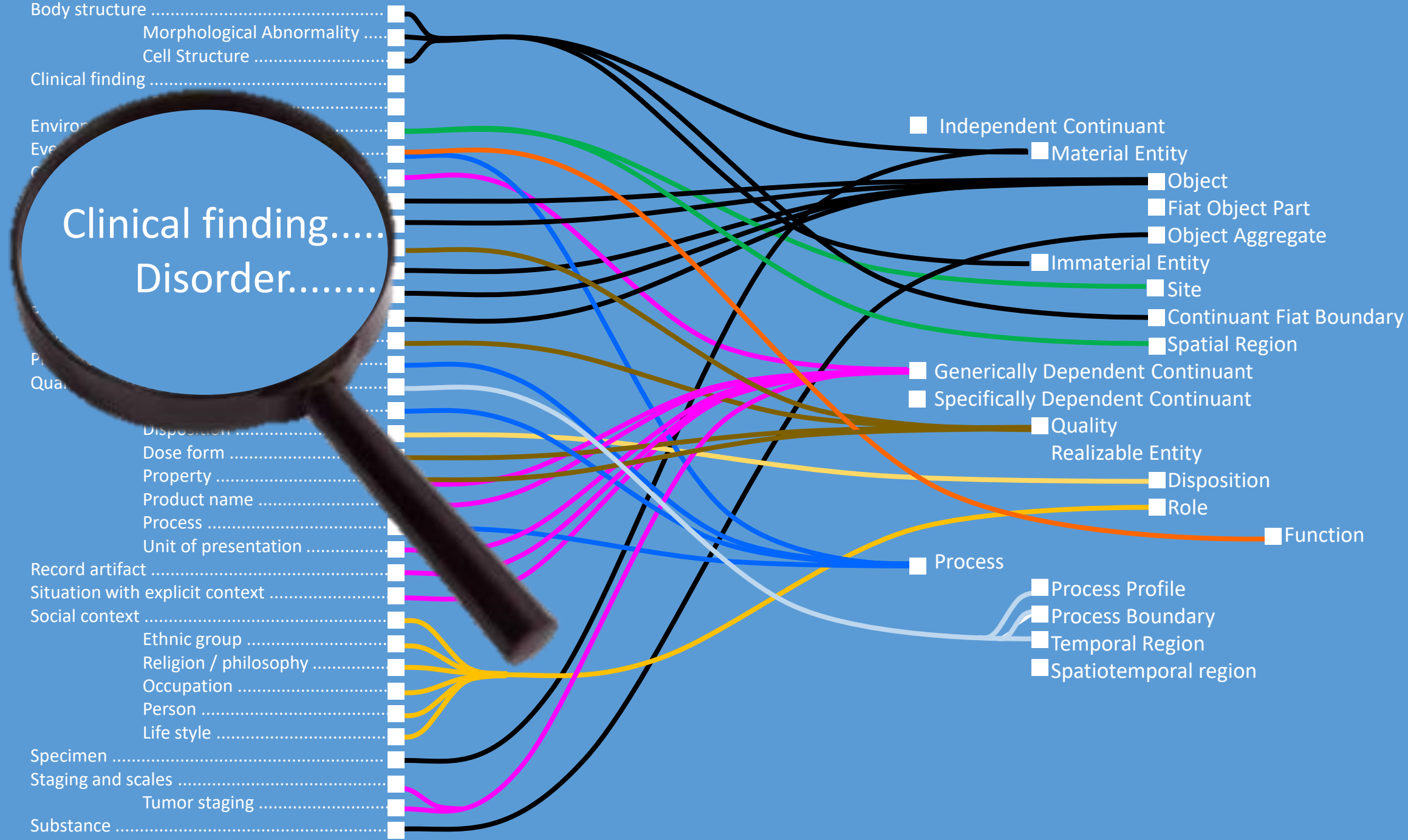
SNOMED CT

Analysing Clinical Findings / Disorders

Clinical finding.....
Disorder........

⅓ of SNOMED concepts

Text definition:

"|Clinical finding| represents the result of a clinical observation, assessment or judgment and includes normal and abnormal clinical states e.g. |asthma|, |headache|, |normal breath sounds|). The |clinical finding| hierarchy includes concepts used to represent diagnoses."

Controversial: that all diseases are disorders, all of which are clinical findings

Analysing Clinical Findings / Disorders

Clinical finding.....
Disorder........

Extension (example of taxonomic descendants):

Pacemaker battery depletion, Yawning absent, Hypercholesterolemia, Financially poor, Appendicitis, Anxiety about going crazy, Fracture of radius AND ulna, brown hair, Azygos lobe of lung, Post-acute COVID-19, Lack of drug action, Diastolic rumble, Twin pregnancy, Does not move, Unpleasant dream, Viral hepatitis type B, Dog bite of toe, Branchioskeletogenital syndrome, Late effect of child abuse, Poisoning caused by barbital, Addiction to sun exposure, epidermal nevus, Bipolar 2 disorder, Cutaneous Munchausen syndrome by proxy, Headache, ...

Analysing Clinical Findings / Disorders

ulna
radius

Formal definitions

Example:

'Fracture of Radius AND Ulna' EquivalentTo:
    'Clinical Finding'
    and 'role group' some
        (('finding site' some 'Bone structure of radius') and
        ('associated morphology' some Fracture))
    and 'role group' some
        (('finding site' some 'Bone structure of ulna') and
        ('associated morphology' some Fracture))

- Does it make sense that a combined fracture class is a subclass of either simple fracture?

- What is the meaning of the object property 'role group' ?

- By which BFO object properties 'finding site' and 'associated morphology' can be subsumed?

Analysing Clinical Findings / Disorders

Dissecting the pattern

- Anatomical sites (range of 'finding site') are independent continuants as well as morphologically abnormal structure (range of 'associated morphology'), according to mapping done before

Formal definitions

Pattern:

'Finding 1' EquivalentTo:
   'Clinical Finding'
   and 'role group' some
      (('finding site' some 'Anatomical site 1') and
      ('associated morphology' some 'Morphology 1'))
   and 'role group' some
      (('finding site' some 'Anatomical site 2') and
      ('associated morphology' some 'Morphology 2'))

Analysing Clinical Findings / Disorders

Dissecting the pattern

- Anatomical sites (range of 'finding site') are independent continuants as well as morphologically abnormal structure (range of 'associated morphology'), according to mapping done before

- Under the assumptions 'finding site' subPropertyOf ' bfo:'located in' 'associated morphology' subPropertyOf ' bfo:'has participant' the range of 'role group' must be an occurrent

Formal definitions

Pattern:

'Finding 1' EquivalentTo:
  'Clinical Finding'
   and 'role group' some
     (('finding site' some 'Anatomical site 1') and
     ('associated morphology' some 'Morphology 1'))
   and 'role group' some
     (('finding site' some 'Anatomical site 2') and
     ('associated morphology' some 'Morphology 2'))

Analysing Clinical Findings / Disorders

Dissecting the pattern

- Anatomical sites (range of 'finding site') are independent continuants as well as morphologically abnormal structure (range of 'associated morphology'), according to mapping done before

- Under the assumptions
'finding site' subPropertyOf ' bfo:'located in'
'associated morphology' subPropertyOf ' bfo:'has participant'
the filler of 'role group' must be an occurrent

- From a domain perspective, most plausible:
'role group' subPropertyOf bfo:'has occurrent part'
'Clinical finding' SubClassOf bfo:occurrent

Formal definitions

Pattern:

'Finding 1' EquivalentTo:
    'Clinical Finding'
      and 'role group' some
        (('finding site' some 'Anatomical site 1') and
        ('associated morphology' some 'Morphology 1'))
      and 'role group' some
        (('finding site' some 'Anatomical site 2') and
        ('associated morphology' some 'Morphology 2'))

Further consequences

Difference between

- Pattern 1: "strict" clinical occurrent:
  class of disorders characterised by
  fracture morphology located at the radius.
  Not in SNOMED CT pre-coordinated content but
  valid expression for concept post-coordination

'Fracture of Radius 1' EquivalentTo:
    'Clinical Finding' and
      (('finding site' some 'Bone structure of ulna') and
      ('associated morphology' some Fracture))

- Pattern 2: "broad" clinical occurrent:
  class of disorders characterised
  by including
  fracture morphology located at the radius

'Fracture of Radius 2' EquivalentTo:
    'Clinical Finding'
    and 'role group' some
      (('finding site' some 'Bone structure of radius') and
      ('associated morphology' some Fracture))

Proposal for reinterpretation

'Fracture of Radius' EquivalentTo:
  'Clinical Finding'
   and 'role group' some
    (('finding site' some 'Bone structure of radius') and
    ('associated morphology' some Fracture))

'Having fracture of Radius' EquivalentTo:
  'Clinical Occurrent'
   and 'bfo:has occurrent part' some
    ((bfo:'located in' some 'Bone structure of radius') and
    (bfo:'has participant' some Fracture))

- Relabelling the top concept "Clinical finding → Clinical occurrent"

- Relabelling all descendants, e.g. "Having X" or "Clinical occurrent with" (probably not accepted)

- Substituting SNOMED CT relations by bfo relations (at least the relation 'role group')

Next steps

- Assessment for plausibility from a domain point of view
    - Sample of maximally diverse Clinical finding subclasses
    - Analysis of other design patterns in the Clinical finding subhierarchy
    - Focus on terms that refer to realizables
    (e.g. allergy to pollen, as a disposition)
    ...clinical occurrent that has participant that
    is bearer of disposition (?)
    - Comparison with other finding / disease ontologies
    ... e.g. human phenotype ontology

Human phenotype ontology:

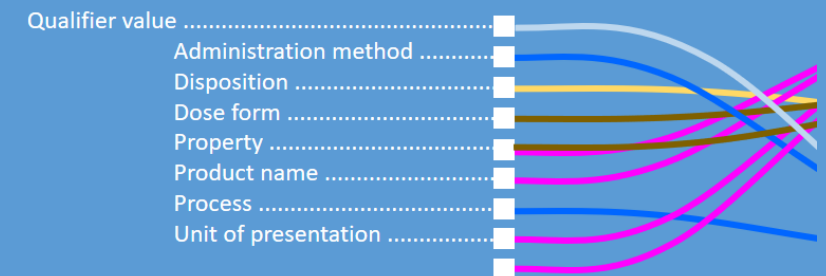'Fractured radius' EquivalentTo:
    and 'has part' some
        (('inheres in' some 'radius bone') and fractured)

Next steps

- Assessment for plausibility from a domain point of view
  - Sample of maximally diverse Clinical finding subclasses
  - Analysis of other design patterns in the Clinical finding subhierarchy
  - Focus on terms that refer to realizables
    (e.g. allergy to pollen, as a disposition)
    ...clinical occurrent that has participant that
    is bearer of disposition (?)
  - Comparison with other finding / disease ontologies
    ... e.g. human phenotype ontology

- Extend analyses to other subhierarchies
  - Qualifier values

- Focus on alignment of BFO object properties
  to SNOMED "linkage concepts"

Human phenotype ontology:

'Fractured radius' EquivalentTo:
    and 'has part' some
        (('inheres in' some 'radius bone') and fractured)

Qualifier value ....
Administration method ....
Disposition ....
Dose form ....
Property ....
Product name ....
Process ....
Unit of presentation ....

Wrapping up (I)

- SNOMED CT legacy:
  - all kinds of entities have accumulated in the clinical finding hierarchy
  - mereological relations between pathological body parts are expressed as taxonomical hierarchies between clinical findings
  - ambiguous label "clinical finding", questionable placement of disorders under clinical finding
- Justification
  - Whether a clinical finding or disorder concept is a continuant or occurrent does not matter in clinical discourse. Often two mutually dependent entities, i.e. tumour as mass and tumour as growth process (logical polysemy*)
  - Current architecture privileges high-recall retrieval use cases

* Arapinis, Alexandra and Vieu, Laure. A Plea for Complex Categories in Ontologies. 1 Jan. 2015 : 285 – 296.
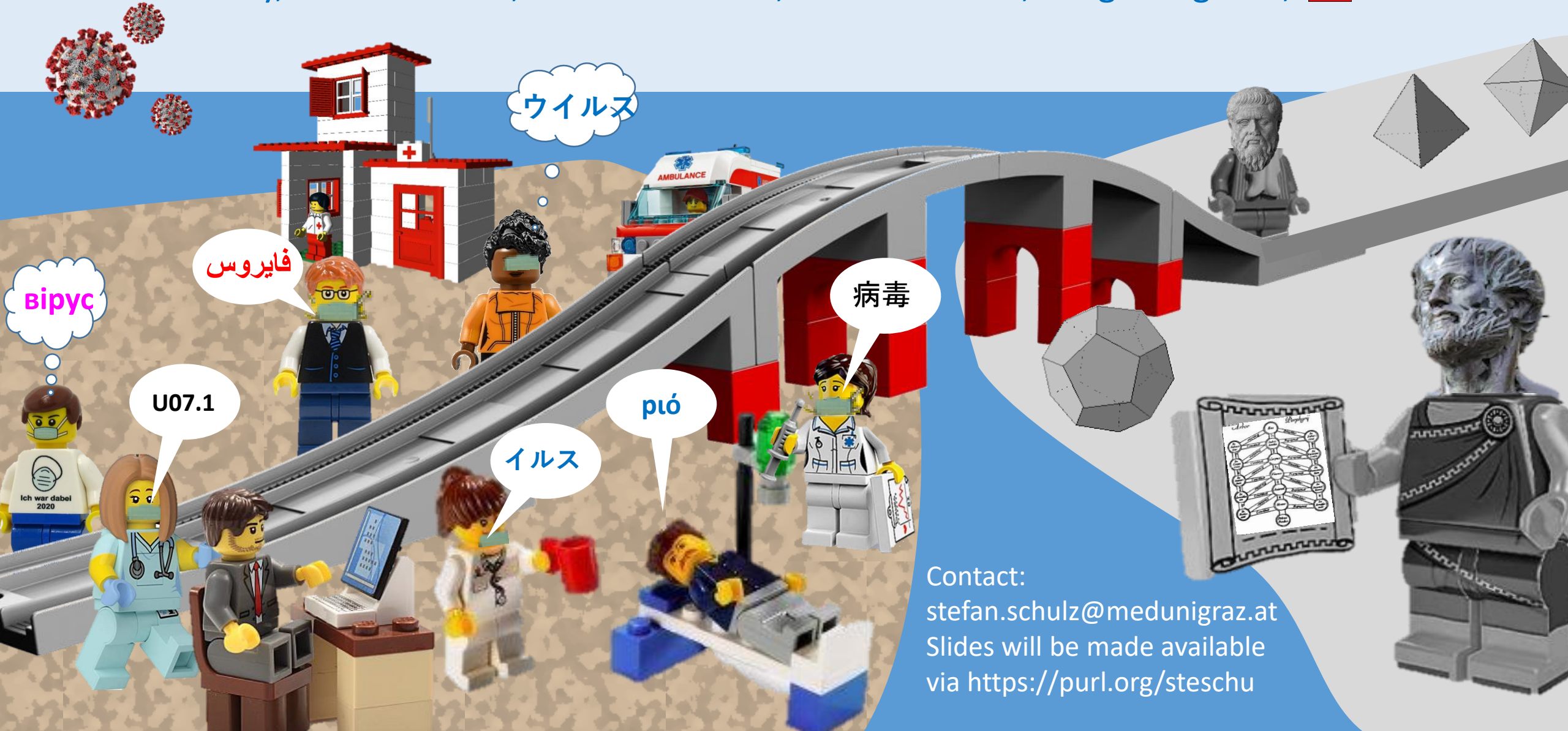
Wrapping up (II)

- Case study shows feasibility of non-disruptive upper-level alignment of a large domain ontology
- Instead of redesign from scratch: re-interpretation of existing content, with minimal interventions:
  - Better labelling
  - More precise free-text definitions / elucidations
  - Giving a precise meaning to "black-box" relation "role group"
- Further investigation
  - Extension to other segments of SNOMED CT
  - Mapping of SNOMED object properties to BFO (OWL) object properties
  - Other domain ontologies / other foundational ontologies
- Far-reaching goal: seamless integration