Healthcare Ontology

Personal Health Ontology

Additions / comments by Stefan Schulz

SPHN classes

SNOMED CT – a large clinical terminology
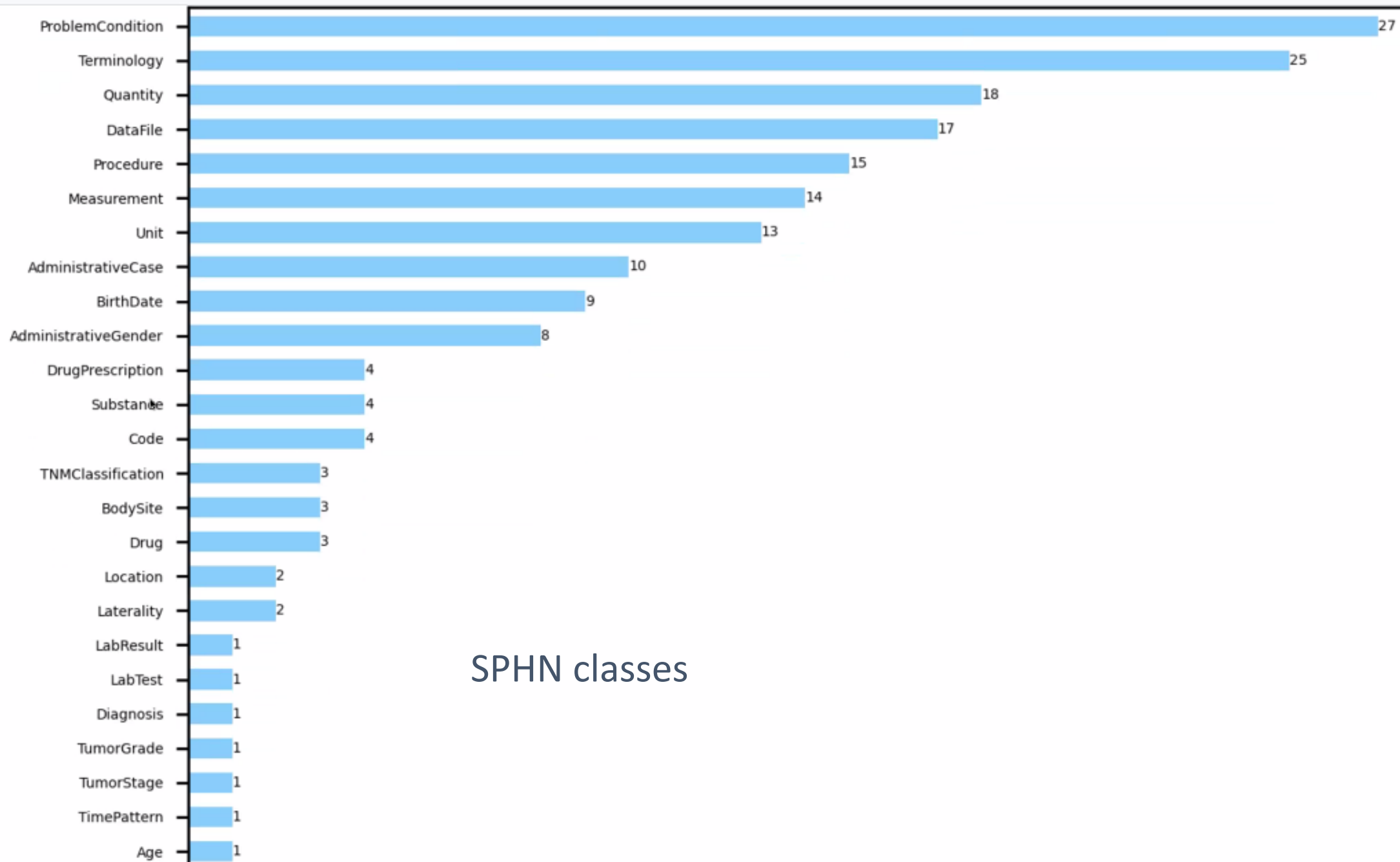
SNOMED CT's ambition is to provide codes, terms and formal descriptions

Specific Considerations

- The inclusion of knowledge elements pertaining to the **delivery of healthcare, clinical studies**, data from **consumer health devices** with a <u>central focus and relation to the individual patient</u> (as opposed to the population).
- The ability to express statements that pertain to:
    - **facts** (e.g. the observation of systolic blood pressure for a particular patient at a particular time, place, and using a specified procedure)
    - **provenance** (e.g. specifics of the data collection, the set of data transformations)
    - **claims** (e.g. the diagnosis of alzheimer's is a claim about the underlying disease) and their evidence.
    - **vocabulary management** (e.g. specification of identifiers and their issuers, and mappings between vocabulary concepts)
- *Alignment* to a **top level (foundation) ontology** to guide knowledge construction in **a principled and extensible manner.**
- *Alignment* to types and type hierarchies defined in other **vocabularies/ontologies**, where possible.
- The use of **design patterns** for all types of closed world knowledge in the PHKG, implemented as **SHACL rules/SPARQL queries**, to constrain the composition of class expressions as well as the validation of their instantiation.

Simplification strategies in healthcare ontology

- Fundamental distinction:
  - Clinical entities (which characterise the patient and related processes)
  - Information entities
- The class "Condition" as an overarching node including diseases, disease processes, physiologic processes (e.g. birth), injuries, signs, symptoms, life events, allergies. Modelled as processes (can be static, e.g. absence of thumb, or highly dynamic, e.g. epileptic seizure).
  Condition = "having sth. of clinical interest", in which a human body or parts thereof participate. (cf. https://dl.acm.org/doi/10.3233/AO-230018)
  In SNOMED CT parent of the hierarchies Finding and Event. No ontological distinctions à la OGMS. Only distinction between disposition and manifestation in terms of "having a disposition", "having a manifestation", conditions are agnostic regarding abnormal / normal : boundary fuzzy, subject to individual judgement
- Procedure: all processes with a health professional as agent and some clearly defined goal (treatment, prevention, investigation)
- Distinction of Condition from Observable: Observables are information entities that together with a value (quantitative or qualitative) (+ unit of measurement) (+ comparator) represent some condition.
- Minimum of binary relations (ideally only ontological relations, others to be expressed as processes). Additional predicates ("reason for procedure", "evidence for diagnosis") may be demanded by the use case, but should at least theoretically be defined
- Qualifiers, "positive", "high", "elevated": avoid proliferation (concentrate on HL7 value sets and corresponding SNOMED CT mappings)
- Each piece of information: to be traced back to patient ID, stay, institution / department ID, IDs of carer(s). IDs link to sensitive data (names, date of birth)
- Consensus on type of information not to be represented:
  - descriptions of details of still images or videos
  - descriptions of procedure details, such as surgical intervention or autopsies

  Instead: representation of summarization of too detailed diagnostic investigation / therapeutic procedure statements only (what is of real interest for decisions): type and outcome of intervention, diagnosis of pathological or radiological examination

What do elements in a clinical knowledge graph represent?

- Nodes represent
  - Instances of information entity (IE) classes
  - Instances of clinical entity (CE) classes
  - CE classes (in case instantiation cannot be taken for granted)
  - Literals (numbers, boolean)
- Edges link
  - IE instances to IE classes (i rdf:Type I)
  - IE / CE instances to IE / CE instances (i rel j)
  - IE / CE instances to literals
  - IE to CE classes (next slide)

How to interpret Instance - rel- Class triples?

Representational pattern: instance *i* **represents** a class *C*

1. Punning: "**i**; **represents**; *C*" - Classes are treated like instances.
   Problem: no clear semantics, no correct reasoning
2. Value restrictions: the triple "**i**; **represents**; *C*" is interpreted as:
   i subClassOf **represents** only *C*
   Not expressible in OWL-EL. Correct reasoning in OWL-DL*
3. Indirect reference to Universals: "**i**; **represents**; **c**"

   "**c** rdf:Type (*Universal* and **extendsTo** some C) "
   Makes an implicit distinction between universals and classes). Universals extend to entities to a certain class, of which individuals of a certain kind are assumed to exist but not necessary related to i.
   Example the IE i = "suspected stroke" represents the universal "Stroke". It has "stroke" individuals in its extension, but this does not mean that i represents any of these stroke entities. i is only about the type / universal **Stroke**, which is modelled as an OWL individual

Schulz, S; Martínez-Costa, C; Karlsson, D; Cornet, R; Brochhausen, M; Rector, A .
An Ontological Analysis of Reference in Health Record Statements. Proceedings of the 8th International Conference on Formal Ontology in Information Systems . 2014; 267:
-8th International Conference on Formal Ontology in Information Systems (FOIS); SEPT 22-25, 2014; Rio de Janeiro

What should elements in a clinical knowledge graph represent?

- Variant 1
  - Nodes
    - Instances of information entity classes (such as FHIR resources, SPHN), additionally Patient, Institution, Stay
    - Domain ontology classes (SNOMED CT, LOINC, …)
    - Literals (numbers, Boolean)
  - Edges
    - Relations between information entity classes and instances using rdf:type, e.g.
      aidava:Condition#3456789 rdf:Type sphn:Condition
    - Relation between information entity instances and domain ontology classes ("binding*")), e.g.
      - relations instance – class (beyond OWL DL), e.g.
        Triple: (aidava:Condition#345678, sphn:hasCode snomed:50920009 |Myocarditis (disorder)|)
      - or approximation in OWL DL using the value restriction operator:
        OWL expression: isAbout (instance_x, Class_y) , e.g.
        instance_x rdf:Type sphn:hasCode only (Class_y or (not Upperclass)),e.g.
        aidava:Condition#345678 rdf:Type sphn:hasCode only (snomed:50920009 |Myocarditis (disorder)|) or (not 404684003 |Clinical finding (finding)|)
- Variant 2 – based on Variant 1
  - Nodes

    > Allowing universals as entities in the range of represents:
    > diagnosticStatementOnStroke123 represents value (Type and extendsTo some Stroke)
    > (not necessarily the stroke in the patient 123 but some other on which the Universal is grounds)
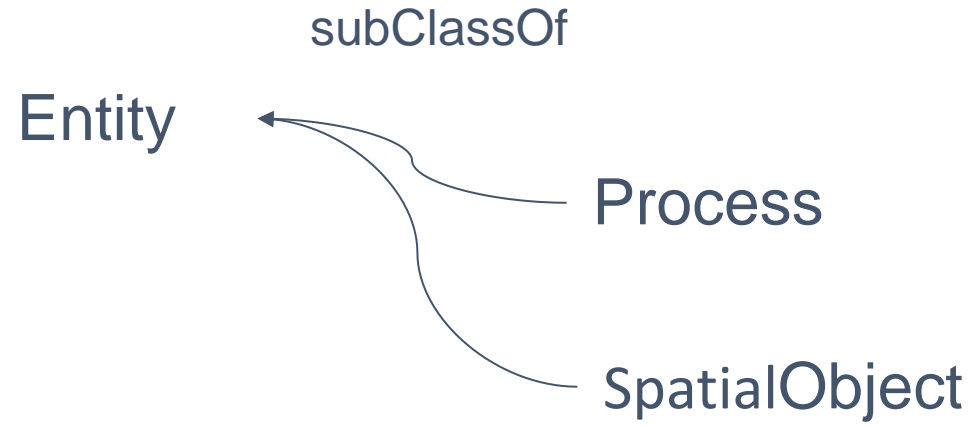
    - additionally: instances of domain classes for all those of which an instantiation in clinical reality can be assumed:
      Confirmed diseases, signs, symptoms, performed procedures, ingested medications
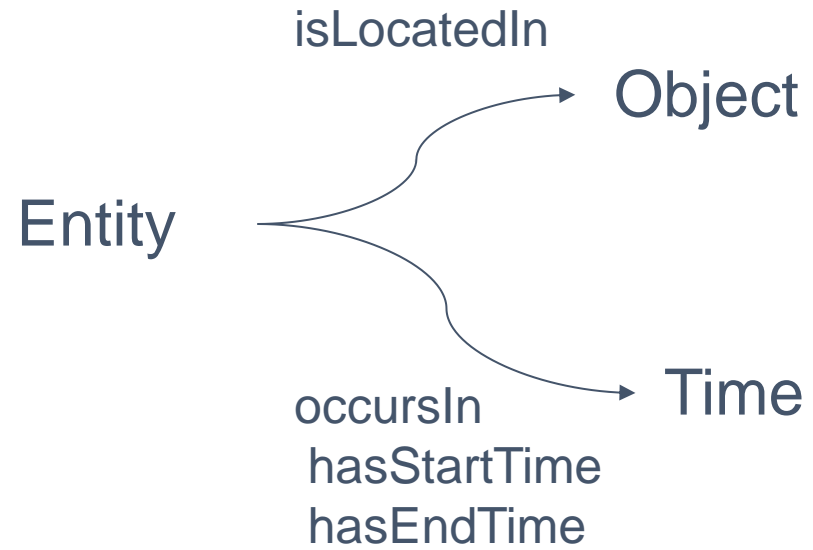- Pros / cons:
  - V1 refers to domain terms (things coded by SNOMED etc.) only indirectly as classes: clear-cut split between instance-to-instance and instance-to-concept links.
    V2 adds more complexity, but requires drawing a line between what does exist (or does not) in clinical reality and what may exist (and is therefore not instantiated)
    V1 makes it complicated to assert co-references. E.g. three information entities are about a tumor: are there three tumors or is the same tumor referenced, tree times

subClassOf

Entity

Process

SpatialObject

Object = Continuant
Also information entities
are continuants

Entity

isLocatedIn → Object

occursIn
hasStartTime
hasEndTime → Time

other relations:

isLocatedIn
isPartOf
hasParticipant

Foundational Ontology
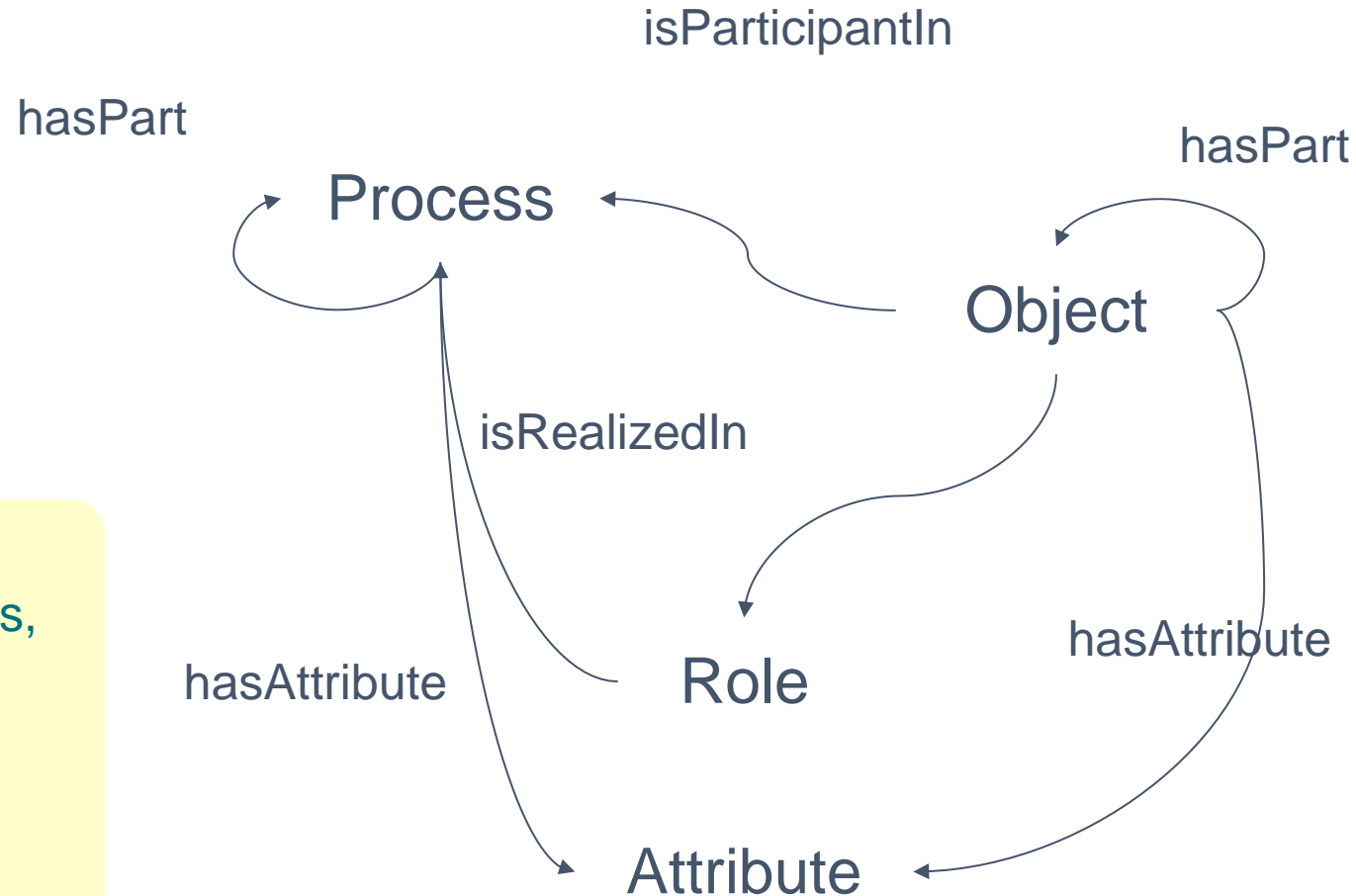
Object
    Organism
        Person
    Anatomy
    Organization
    Medication
    Instrument/Device
    Information Content Entity
        representing:
            Observation
            Quantity
             Concept
            Concept Mapping
            Identifier
            Identifier Mapping
            Data Source Description

Quality
Role
        Subject
        Patient
        Care Provider
            Physician
        Data Provider

Process
        Procedure
        Event
        Clinical Study
        Data Transformation

- information object
  - ClinicalStatement
  - Datatype
    - Complex datatype
      - Numeric complex datatype
        - Numeric interval datatype
          - Decimal interval datatype
          - Integer interval datatype
    - Primitive datatype
      - Non numeric primitive datatype
        - Boolean datatype
        - Date datatype
        - Datetime datatype
        - String datatype
        - Time datatype
      - Numeric primitive datatype
        - Decimal number primitive datatype
          - Decimal datatype
          - Double datatype
          - Float datatype
        - Integer number primitive datatype
          - Int datatype
          - Integer datatype
          - Long datatype
          - Short datatype
  - QuantitativeResultValue

Metadata
Entity

Entity

hasIdentifier

uri

name@lang

description@lang

uri

name@lang

description@lang

identifier

terminologySystem

identifier

version

url

mapping-type
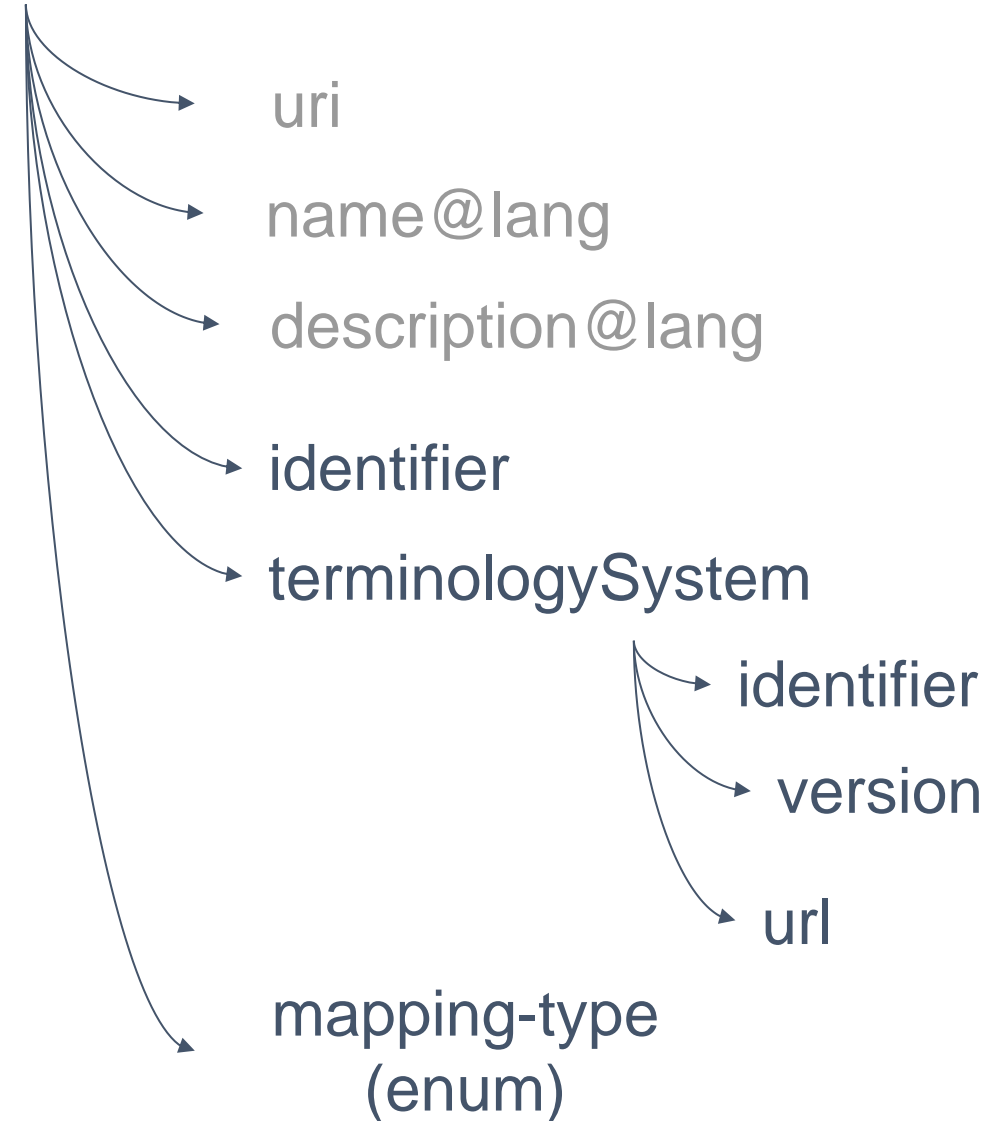(enum)
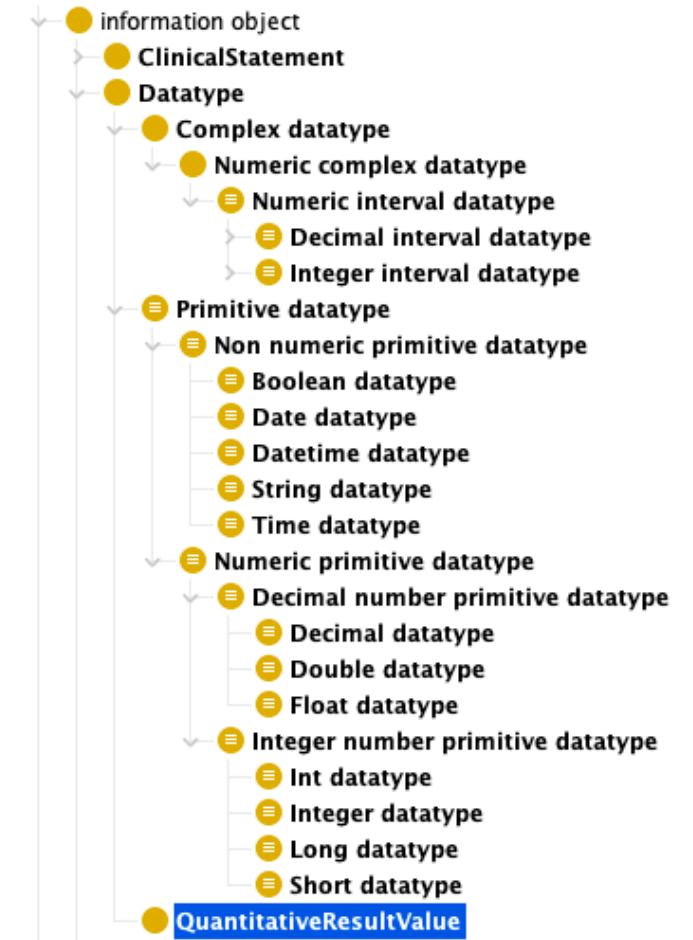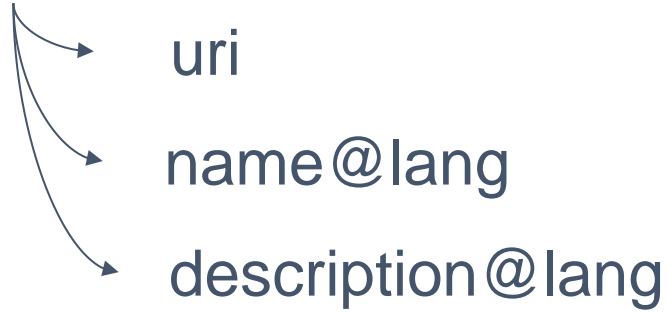
I see the following distinctions of entities, which should not be confounded:
1. Entities of language vs. other entities
2. Entities of language (spans) and the symbols used in annotation
3. Symbols used in annotation and their referents in the domain
4. Referents that are types (classes) and referents that are individuals

Procedure

Action

difference between procedure and action?

Subject

Site

Agent

Instrument

Time

Place

difference between subject and agent?

Quantity

Procedure

Value

Unit

Time

also: value comparator, e.g. ">"
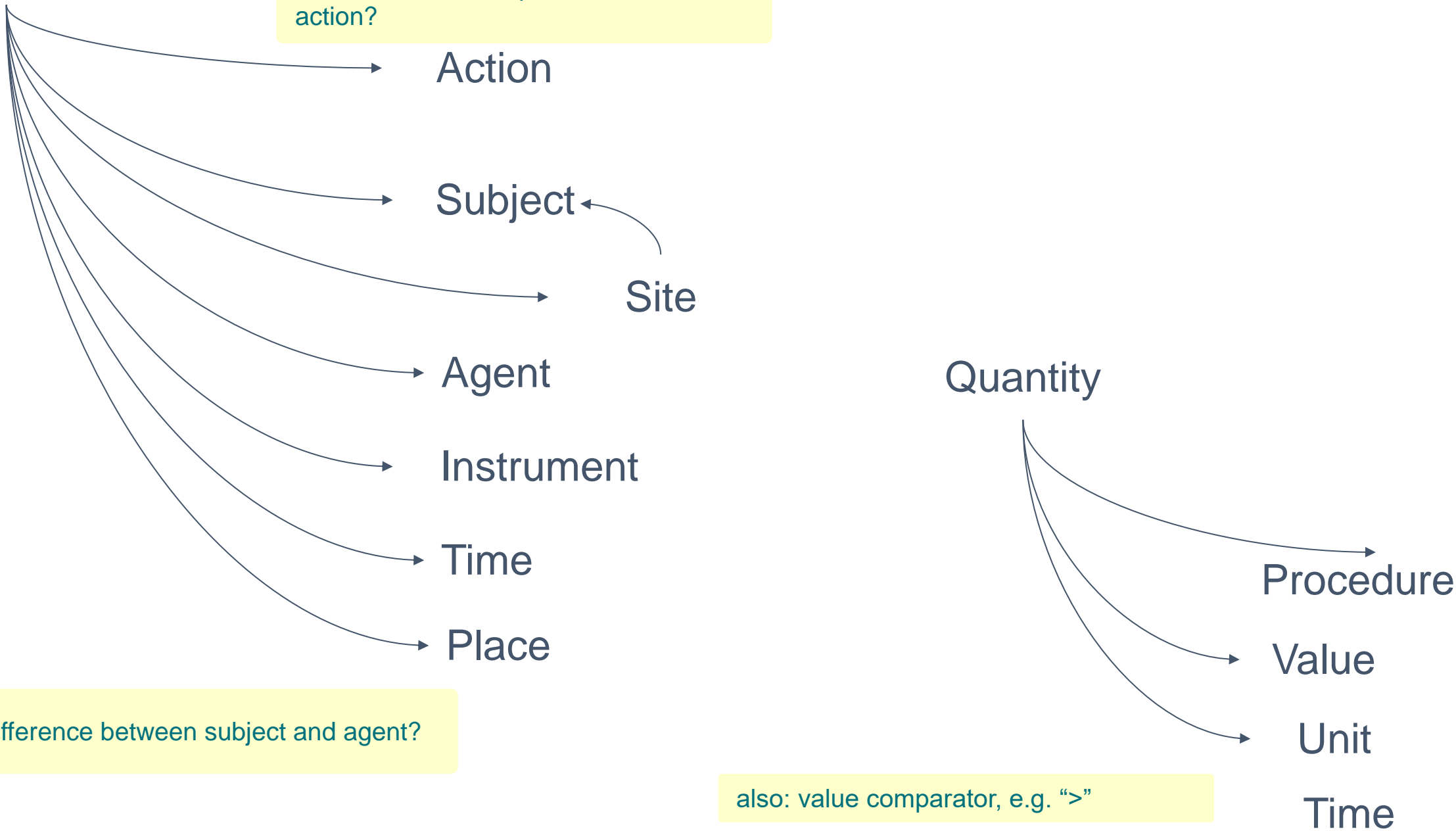
Quantity

Value

Unit

Time

Process

Time

instant of time
vs interval of time

instant of time
- datetime

relative time
- age

relative time: duration

SWAT4HCLS Biohackathon

Goal will be to develop the ontology, and be able to transform other representations.

Key idea is that the transformation is expressible via SPARQL construct queries (provided the data are in graph normal form).

- OMOP to HO
- SPHN to HO
- FHIR to HO

Methodology

-> convert datatype properties to classes; with single value datatype property ("hasValue")

-> transform object properties to classes,

Prototypical modelling examples, demonstrating that competency questions can be answered (see following). Suggestion: modeling in OWL-DL (using Protégé) ABox and TBox. Demonstrating reasoning examples using SPARQL and/or OWL query

Example 1

- Alma Alpha, born 2001-08-07 was admitted to emergency room on 2023-02-29-13:00, 30 min after road traffic accident. Conscious, oriented. Bleeding wound at skull. Abdominal pain. No evidence of fractures. Hb 7.0 g/dl, HR 120, BP 90/60. Ultrasound: massive liquid in peritoneal space. Suspected spleen rupture. Ringer lactate iv 30 ml/min. Prepared for laparotomy.
- Should be retrieved when querying for
  - no loss of conciousness
  - Less then one hour after accident
  - Tachykardia
  - Anemia
  - Suspected organ lesion
  - No laparotomy done before 2023-02-29-13:00
  - Head injury
  - No femur fractur

Example 2

- Bertha Beta, born 2014-01-14, type 1 diabetes known for 4 years, no diabetes in family history, was seen in paediatric endocrinology clinic on 2023-02-23 2pm by Dr John Jenkins, accompanied by her mother. According to mother, repeatedly refused insulin injections due to pain and erythema at injection site. On examination no skin abnormalities.
- Should be retrieved when querying for
  - diabetic child
  - diabetes diagnosed more than one year ago
  - seen in a paediatric institution
  - parent reported pain
  - seen in an endocrinology clinic
  - skin alteration in the past, information provenance: family member
  - No abnormalities of skin of extremities
  - No abnormalities of skin of abdomen.
  - No Type 1 diabetes in family history

More examples… more AIDAVA-related

SNOMED CT – a large clinical terminology

?

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

SNOMED CT – a large clinical terminology

?

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

?

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

[2] Health/disease-related phenomena with [1] as participants or as locations
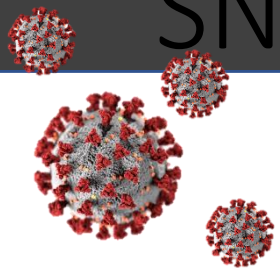
SNOMED CT – a large clinical terminology

?

[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

Knowledge graphs rooted in interoperability standards: preferred representation of clinical information

revealed purpura
swelling of the
showing the
Hematological
...
...2.0%,
lymphocytes
...b 7.1 g/dl
...unt of 9.1 ×
...evealed
...creased number
...es and an

„canonical" representation of clinical information

SNOMED CT – a large clinical terminology

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

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 ....

(simplified)

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" ***

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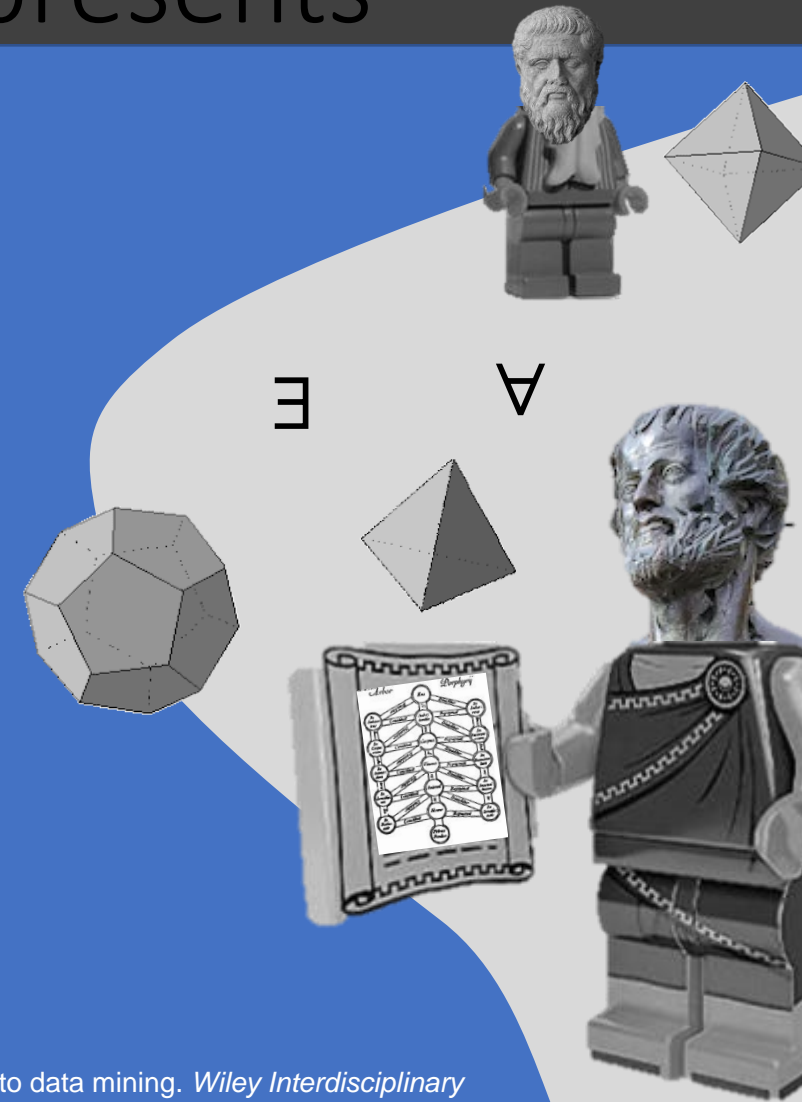
* Amaral, G., Baião, F., & Guizzardi, G. (2021). Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1408.
** Smith, B. Information Artifact Ontology: General Background.   https://stids.c4i.gmu.edu/papers/STIDSPresentations/STIDS2013_Tu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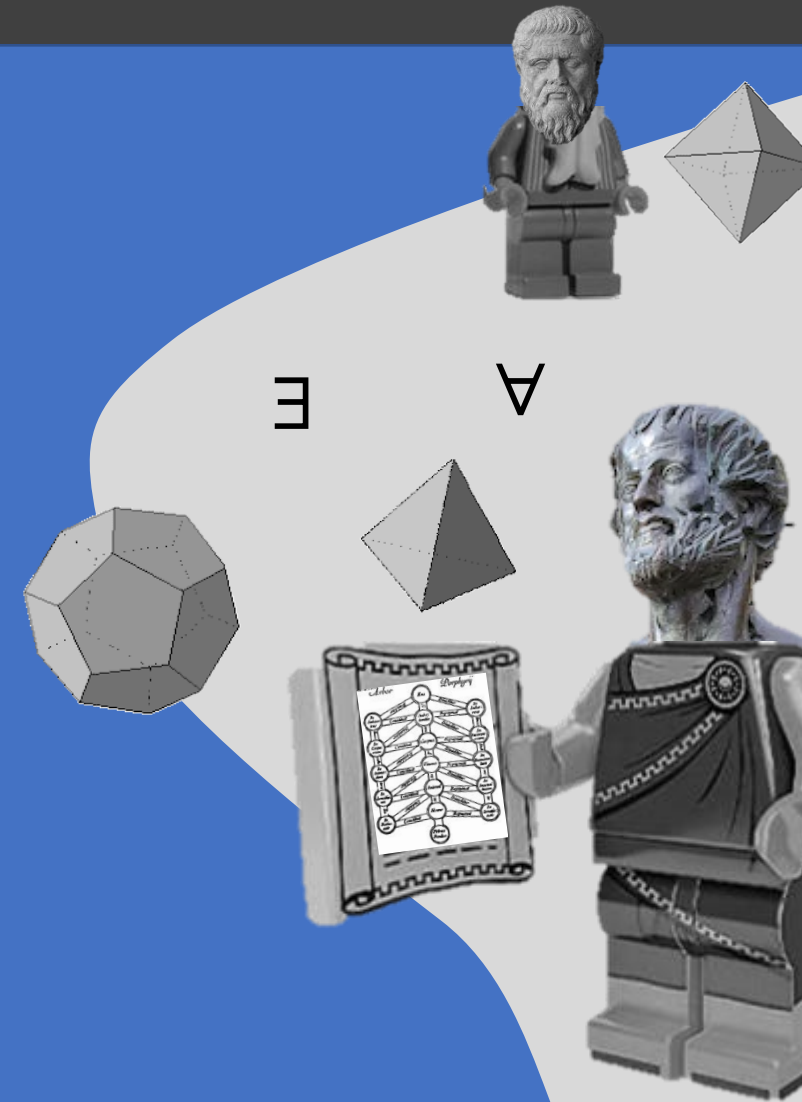
BFO in an nutshell

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

Generi
Specifi

Proces

Driver of this work. interest in bridging between two standards

What's the problem
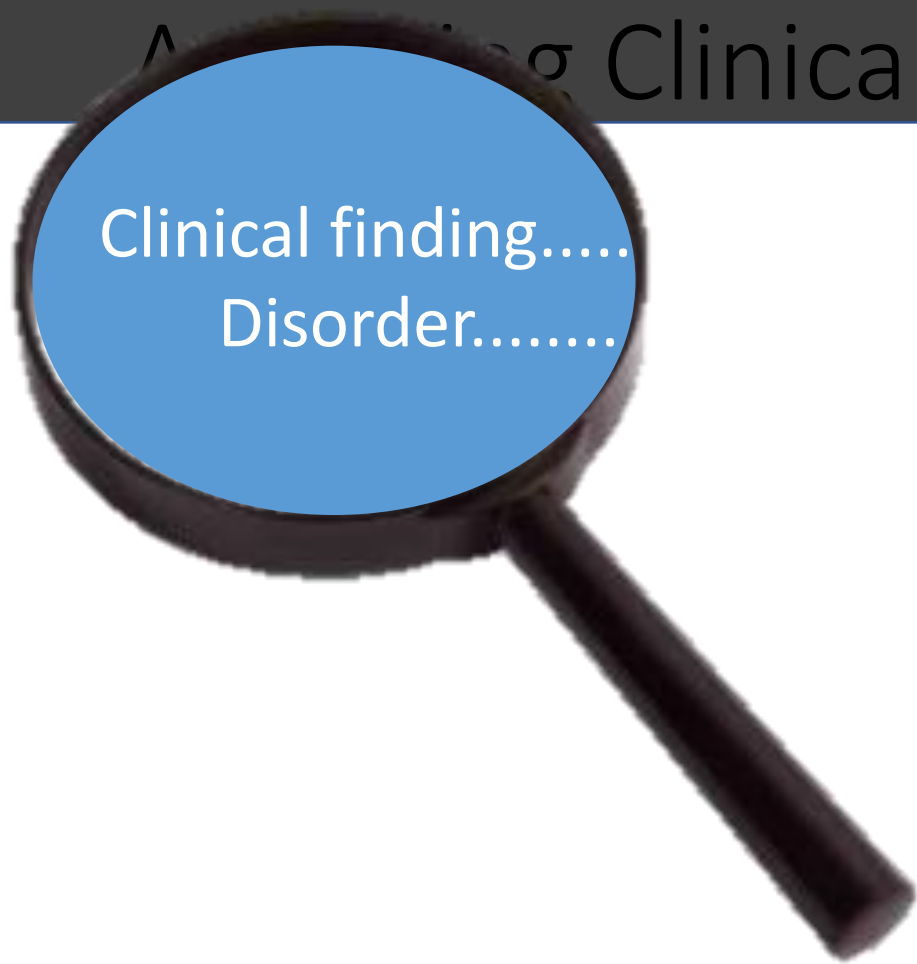
Perception of foundational ontology by terminologists

Harmonizing SNOMED CT with BFO

BFO

SNOMED CT

Analysing Clinical Findings / Disorders
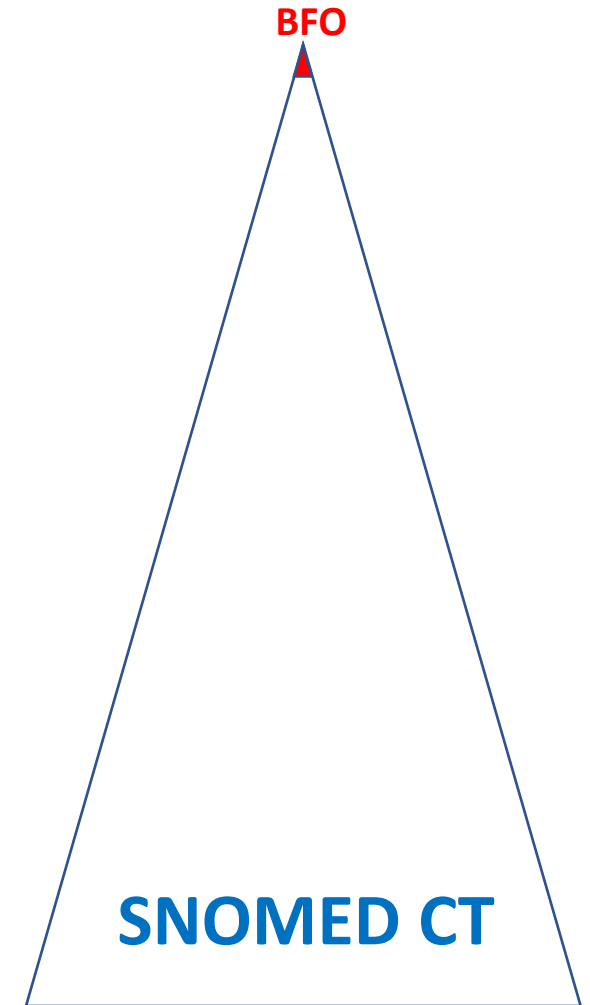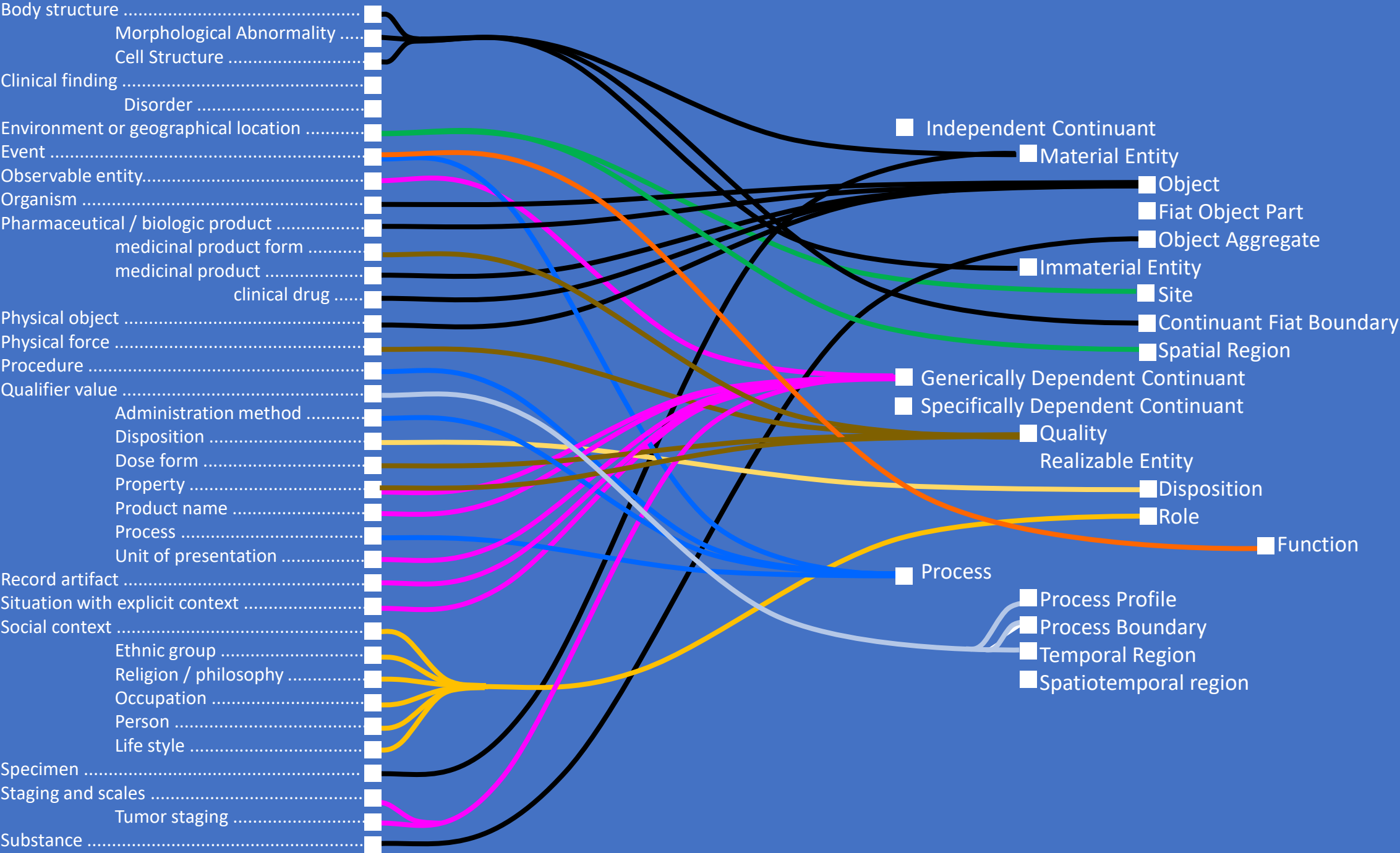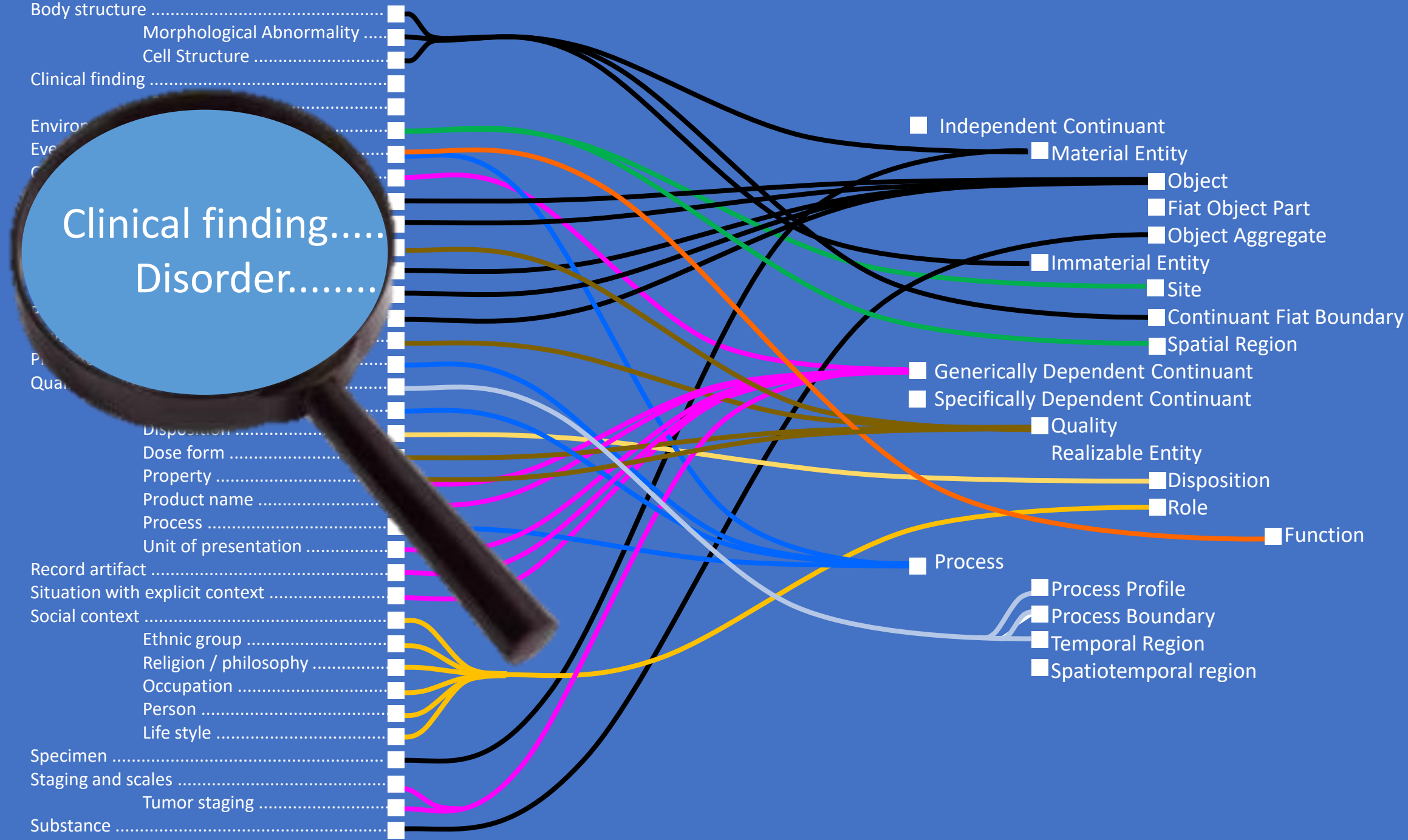
Clinical finding.....
Disorder........

clinical

states

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

Accessing Clinical Findings / Disorders

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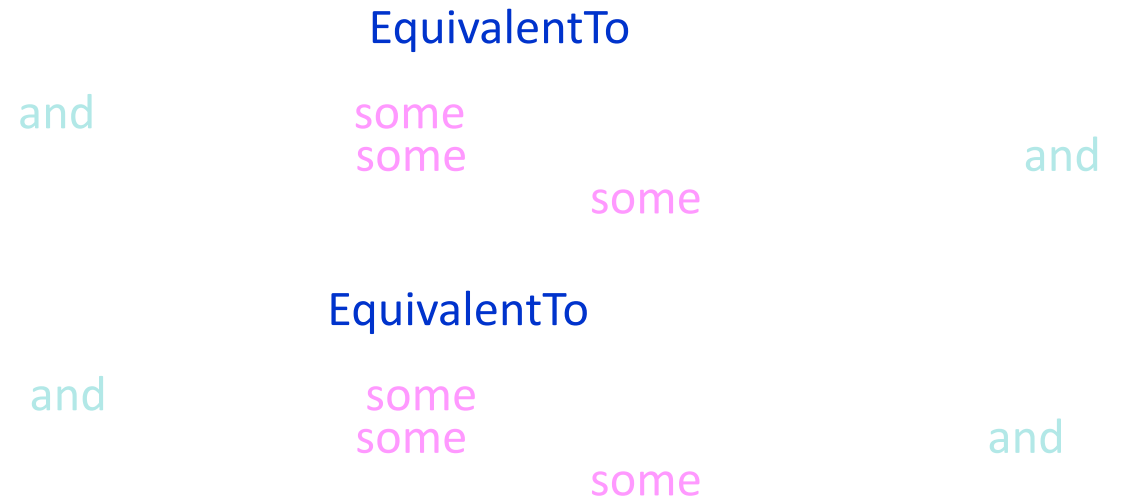Disorder........

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

EquivalentTo

and   some
       some   and
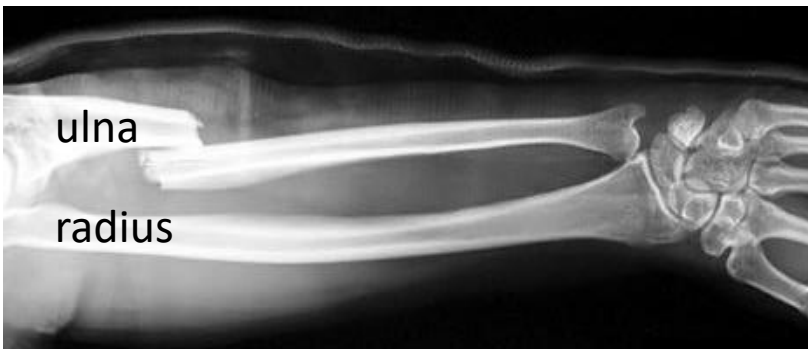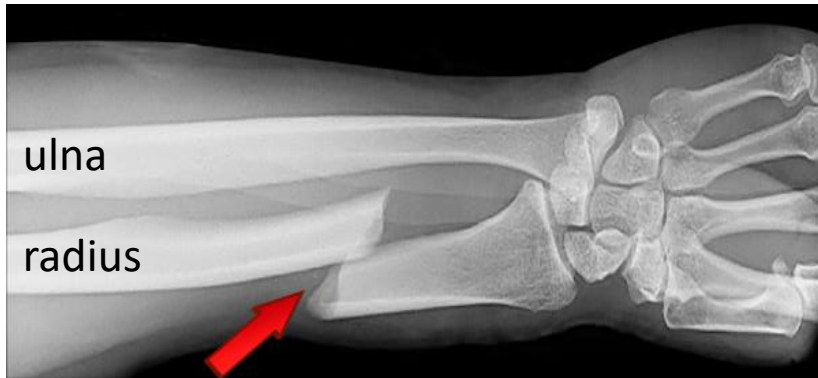   some

and   some
       some   and
            some

and   some   and
            some

'role group' – black box object property intended to structuring axioms like this

EquivalentTo

and   some

some   and

some

EquivalentTo

and   some

some   and

some

EquivalentTo

ulna
radius

and   some
             some   and

                    some

and   some
             some   and

                    some

- 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?

Anatomical sites

abnormal structure

morphologically

EquivalentTo

and some

some 'Anatomical site 1' and

some 'Morphology 1'

and some

'Anatomical site 2' and

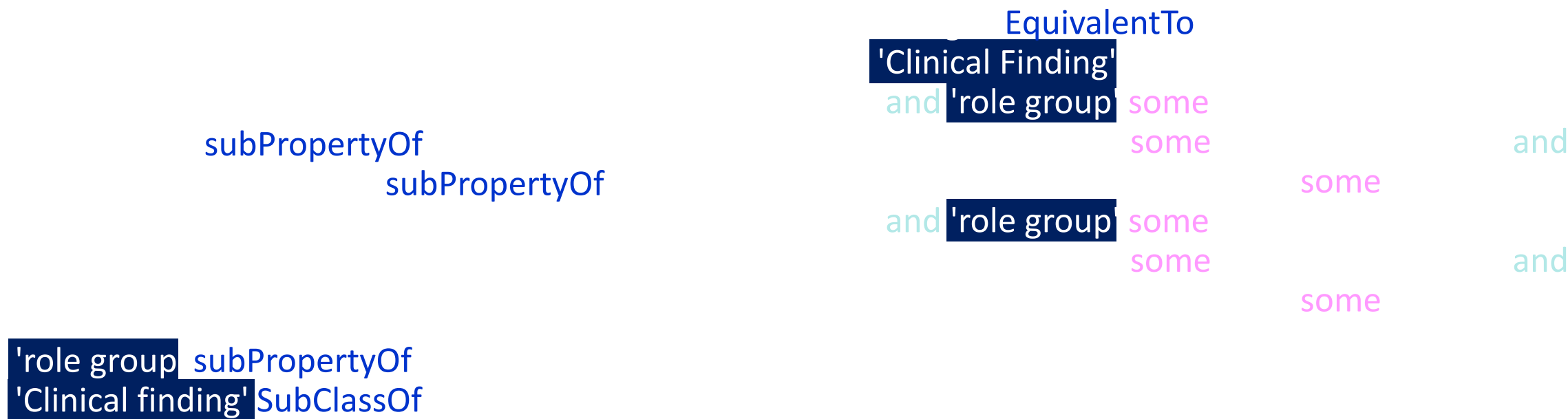some 'Morphology 2'

Analysing Clinical Findings / Disorders

EquivalentTo

and   some

(('finding site' some 'Anatomical site 1') and
('associated morphology' some 'Morphology 1'))

subPropertyOf

subPropertyOf

and   some

(('finding site' some 'Anatomical site 2') and
('associated morphology' some 'Morphology 2'))

the range of 'role group'

Analysing Clinical Findings / Disorders

EquivalentTo

'Clinical Finding'
and 'role group' some
some   and
                          some
and 'role group' some
          some   and
                  some

subPropertyOf
        subPropertyOf

'role group' subPropertyOf
'Clinical finding' SubClassOf

Further consequences

EquivalentTo

and

some

and

some

EquivalentTo

and

some

some

and

some

by including
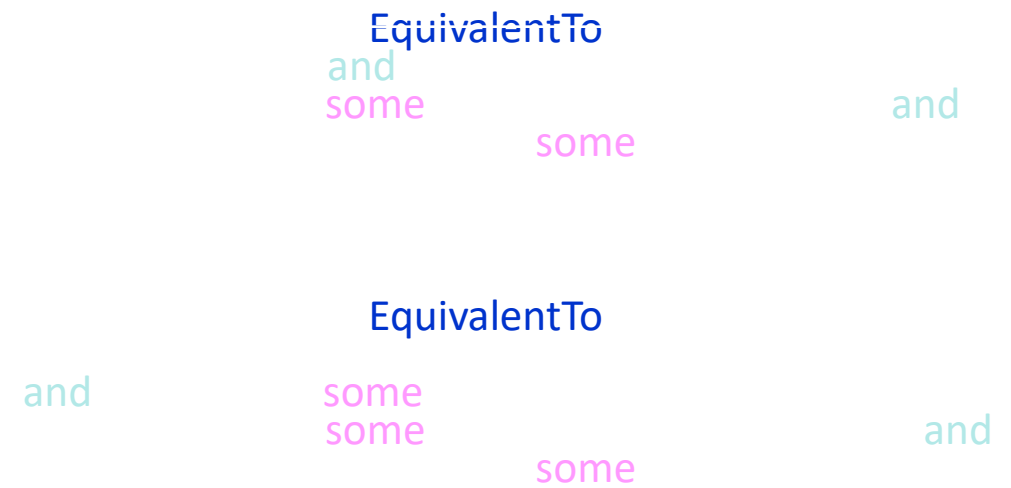
Proposal for reinterpretation

EquivalentTo

Having fracture of Radius EquivalentTo
Clinical Occurrent

and   some

and bfo:has occurrent part some

some

and   bfo:'located in' some   and

some

bfo:'has participant' some

Next steps

EquivalentTo

and   some

some   and

Next steps

EquivalentTo

and   some

some   and

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

Wrapping up (I)

Wrapping up (II)
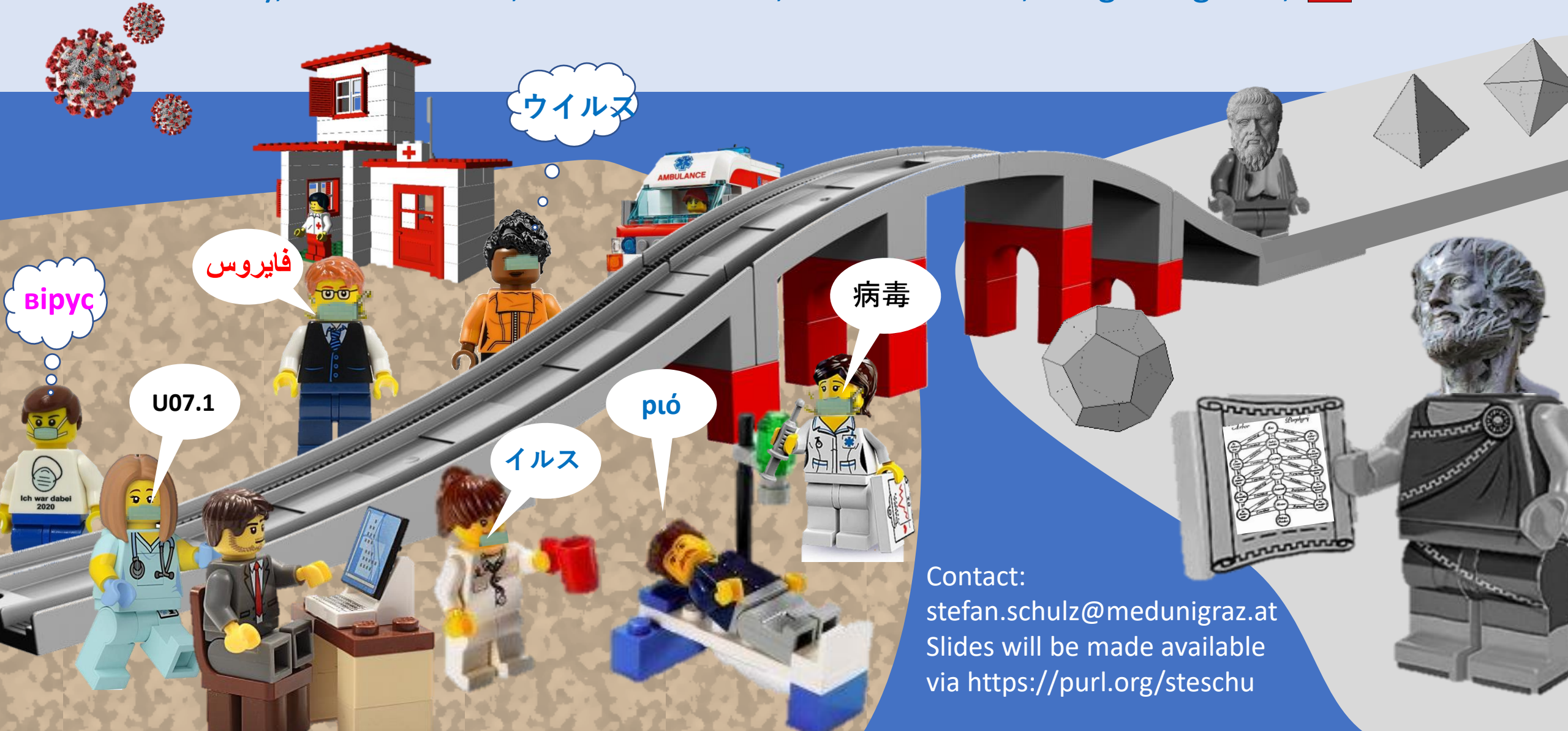
Contact:
stefan.schulz@medunigraz.at
Slides will be made available
via https://purl.org/steschu