Towards an Ontology of Biomedical Relations:

A Preliminary Analysis

Lowell Vizenor

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Overview

Part 1: Ontology 101

Part 2: Enhancing Biomedical Ontologies through Alignment of Semantic Relationships

Ontology

Word Origin

ontologia = ont + logia

literally: the study of being(s)

coined in 1613:

Rudolf Göckel (Goclenius)
Lexicon philosophicum

Jacob Lorhard (Lorhardus)
Theatrum philosophicum.

Aristotle (384-322 BC)

CATEGORIAE (*Categories*)

1. Substance
2. Quantity
3. Quality
4. Relation
5. Place
6. Time
7. Position
8. State
9. Action
10. Passion

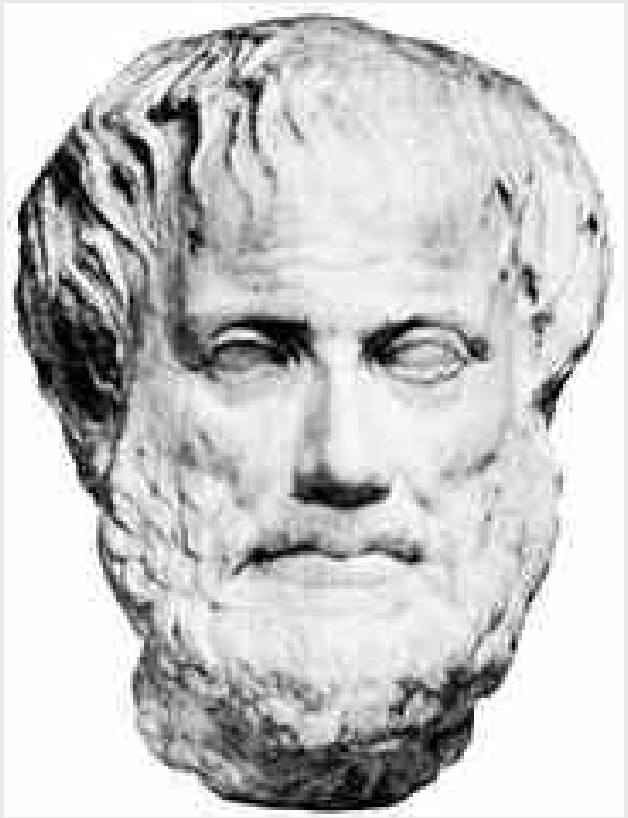
"A five-foot tall (*quantity*) man (*substance*) who was a thinker (*quality*) sat (*position*) on a bus (*place*) one morning (*time*), feeling hungry (*state*), but continuing to do a crossword puzzle (*action*) enthusiastically (*passion*)."

First Philosophy (the science of being *qua* being) seeks to provide a *definitive* and *exhaustive* classification of entities in all areas of being.

What is (an) ontology?

Philosophical Ontology

Aristotle (384-322 BC)



W. V. O. Quine (1900-2000)

Philosophical Ontology

is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality

But how do we go about discovering *what there is*?

Aristotle: Ontology is a *sui generis* science, distinct from the special sciences (e.g. physics, chemistry, and biology)

Quine: the way to do ontology is exclusively through the investigation of scientific theories.

Quine

the ontologist's task is to establish what kinds of entities scientists are committed to in their theorizing (i.e. to find the ontology *in* scientific theories).

How do we do this?

We define the vocabulary of the corresponding scientific theory and give it its canonical formalization (i.e. put it in the language of first-order logic).

"It is then, Quine argues, only the bound variables of a theory that carry its definitive commitment to existence. It is sentences like 'There are horses,' 'There are numbers,' 'There are electrons,' that do this job"

Smith (2003) "Ontology"

To be is to be the value of a bound variable.

What is (an) ontology?

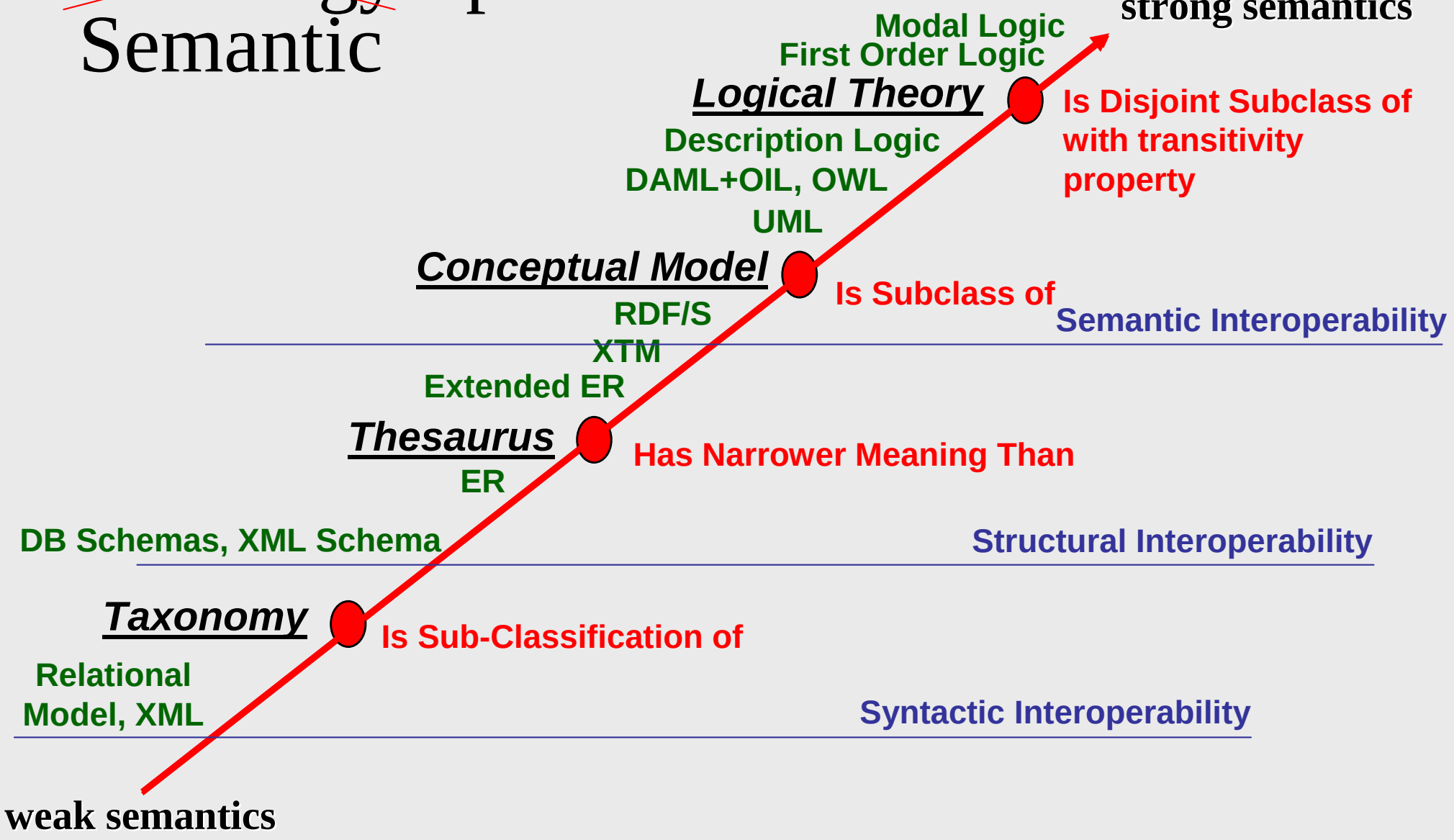
Ontology in Information Systems

Gruber (1993): “an **ontology** is an explicit specification [i.e. formalization] of a conceptualization”

A *conceptualization* is an abstract, simplified view of the world that we wish to represent for some purpose

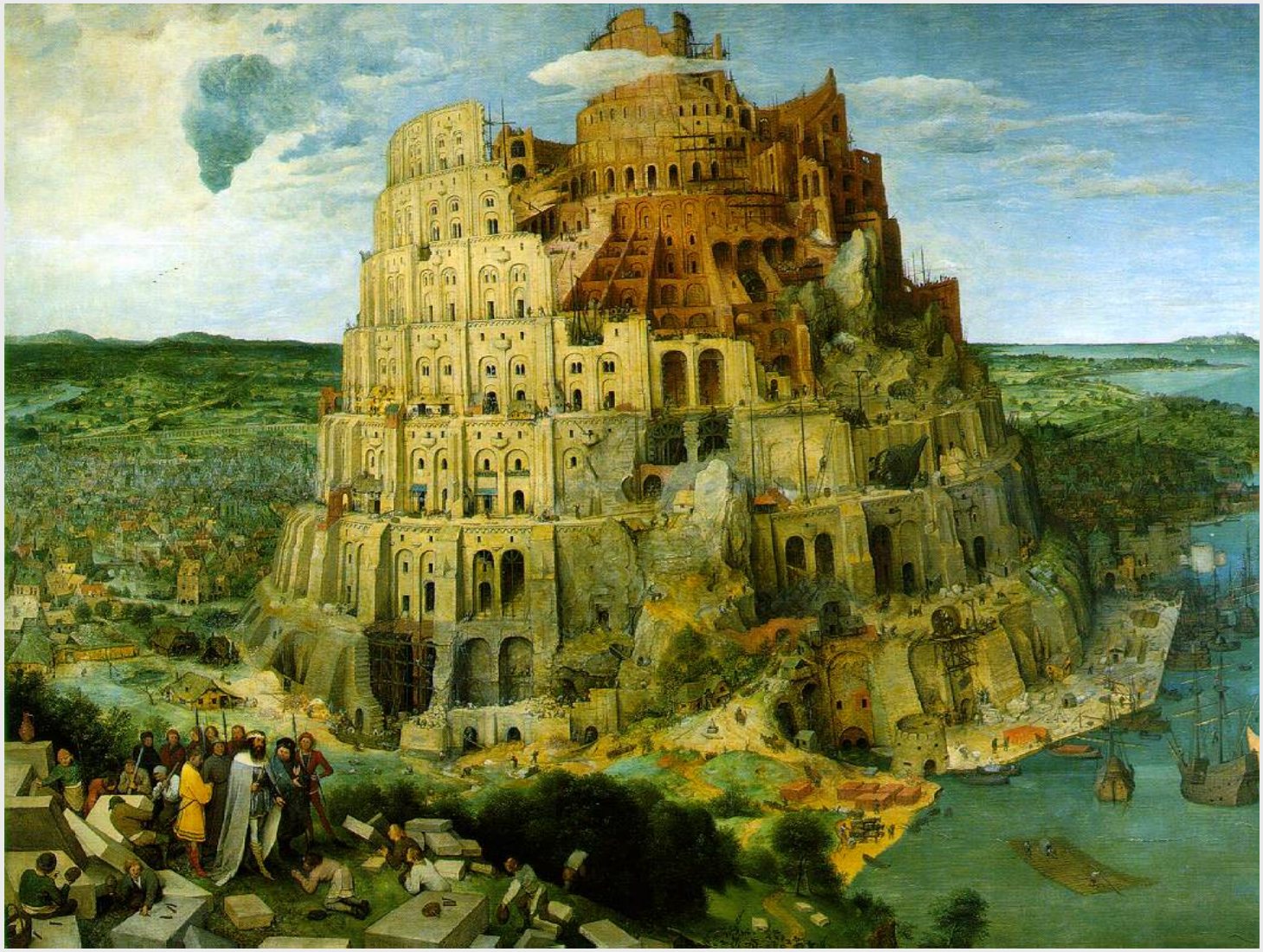
[i.e.] … the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them.

~~Ontology~~ Semantic Spectrum



From Leo Orbst "Ontologies and the Semantic Web for Semantic Interoperability"

The Tower of Babel Problem: Semantic Mismatch



Sources of the problem

Data- and knowledge-base systems contain idiosyncratic terms and concepts by means of which they build frameworks for information representation.

Use identical terms but with different meanings …

Use different terms to express the same meaning.

can give rise to terminological and conceptual incompatibilities.

Gruber’s approach to ontology represents a partial solution to the problem.

But …

not all conceptualizations are created equal

Borges' Animals

The *Celestial Emporium of Benevolent Knowledge* divides animals into...

1.those that belong to the Emperor,
2.embalmed ones,
3.those that are trained,
4.suckling pigs,
5.mermaids,
6.fabulous ones,
7.stray dogs,
8.those included in the present classification,
9.those that tremble as if they were mad,
10.innumerable ones,
11.those drawn with a very fine camelhair brush,
12.others,
13.those that have just broken a flower vase,
14.those that from a long way off look like flies.

From "The Analytical Language of John Wilkins“
Jorge Luis Borges

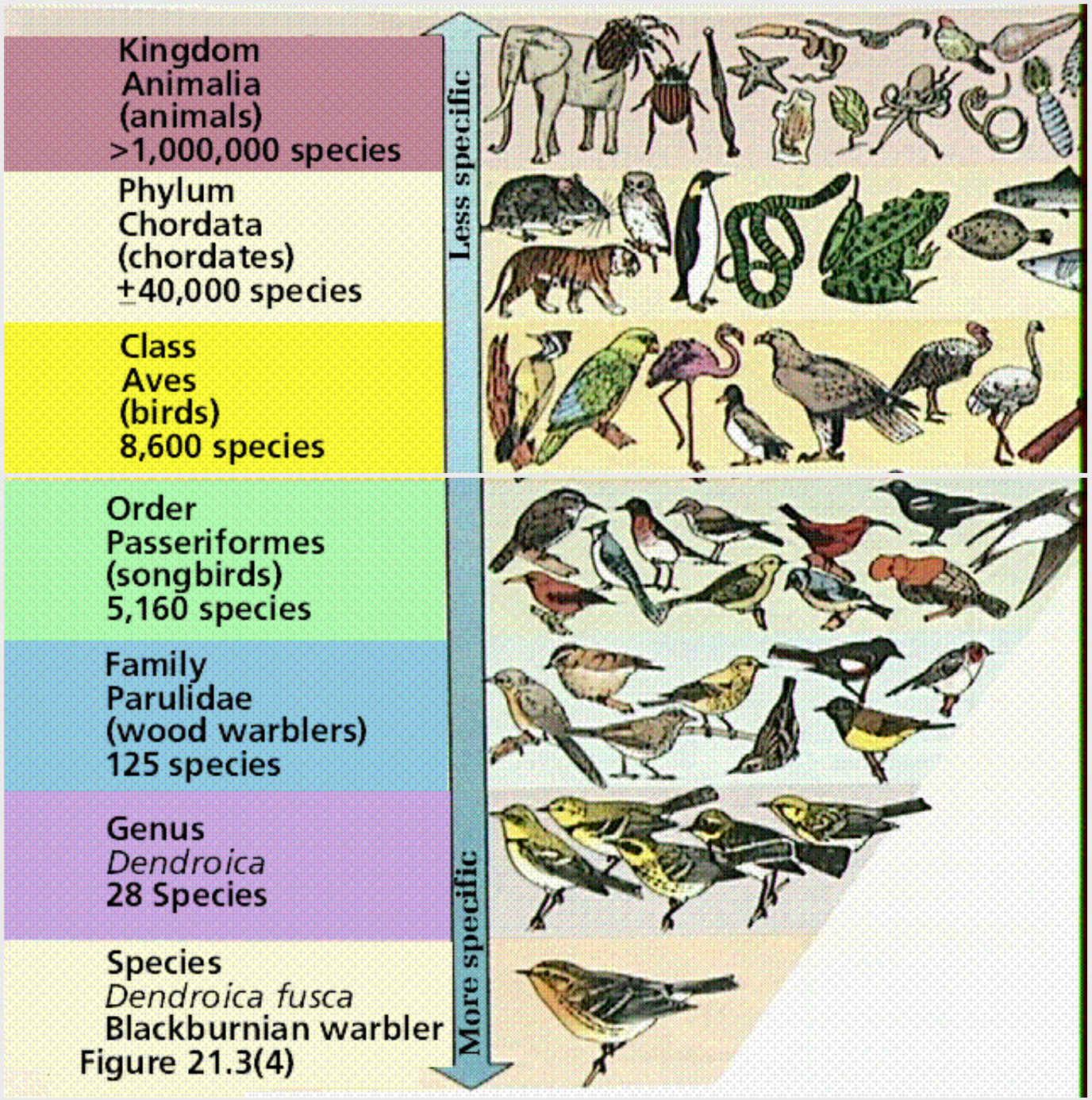


Image from Purves et al., *Life: The Science of Biology*, 4th Edition, by Sinauer Associates

The Knowledge Engineering Paradox

Waterman (1986): the more competent domain experts become, the less able they are to describe the knowledge they use to solve problems.

Declarative versus procedural (or tacit) knowledge

… expert at what they *do* and not (necessarily) at what they *know*

The moral of the story is:

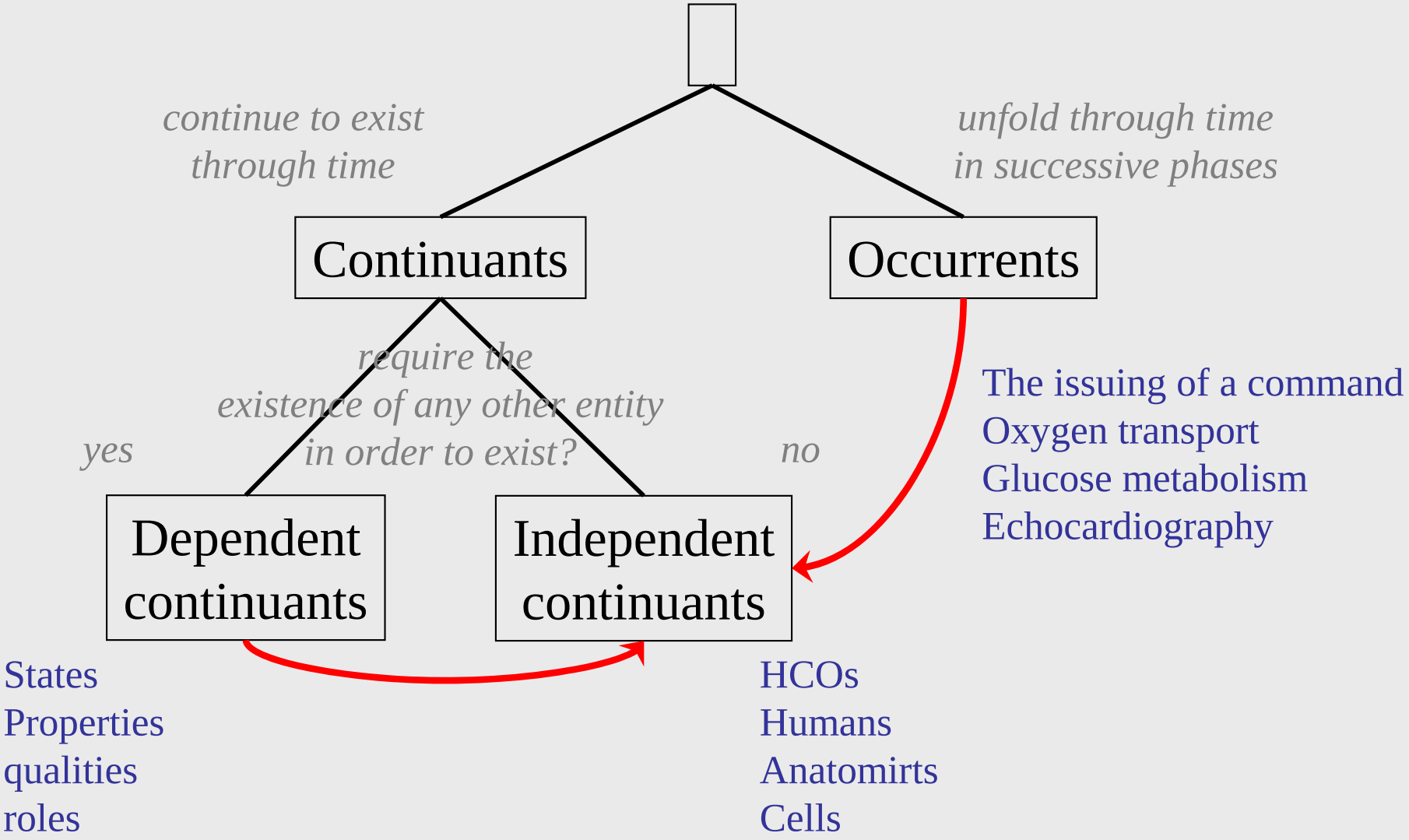
We need to think about the *content* of our ontology and not only the format in which it is expressed.

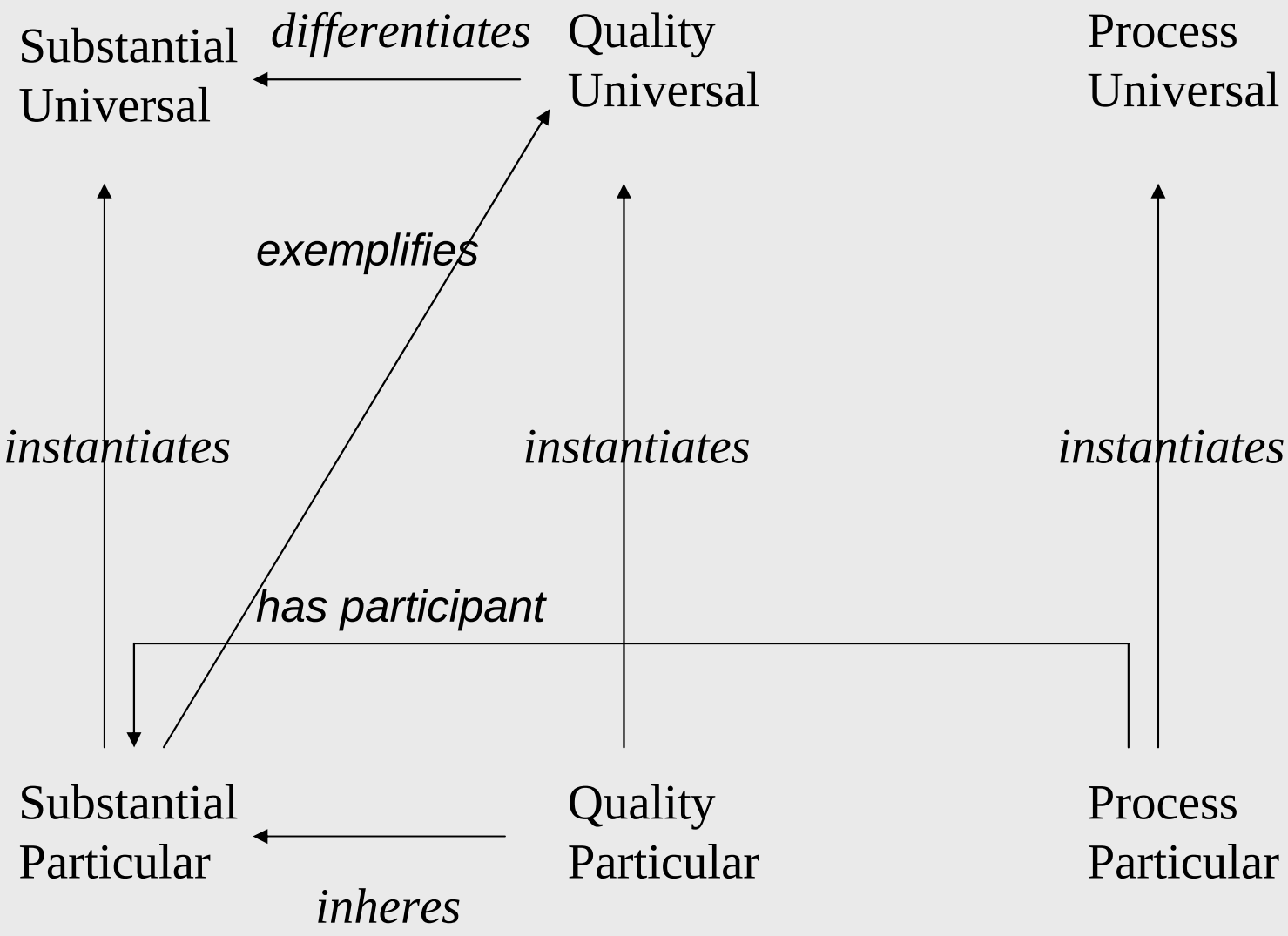
Ontology Revisited

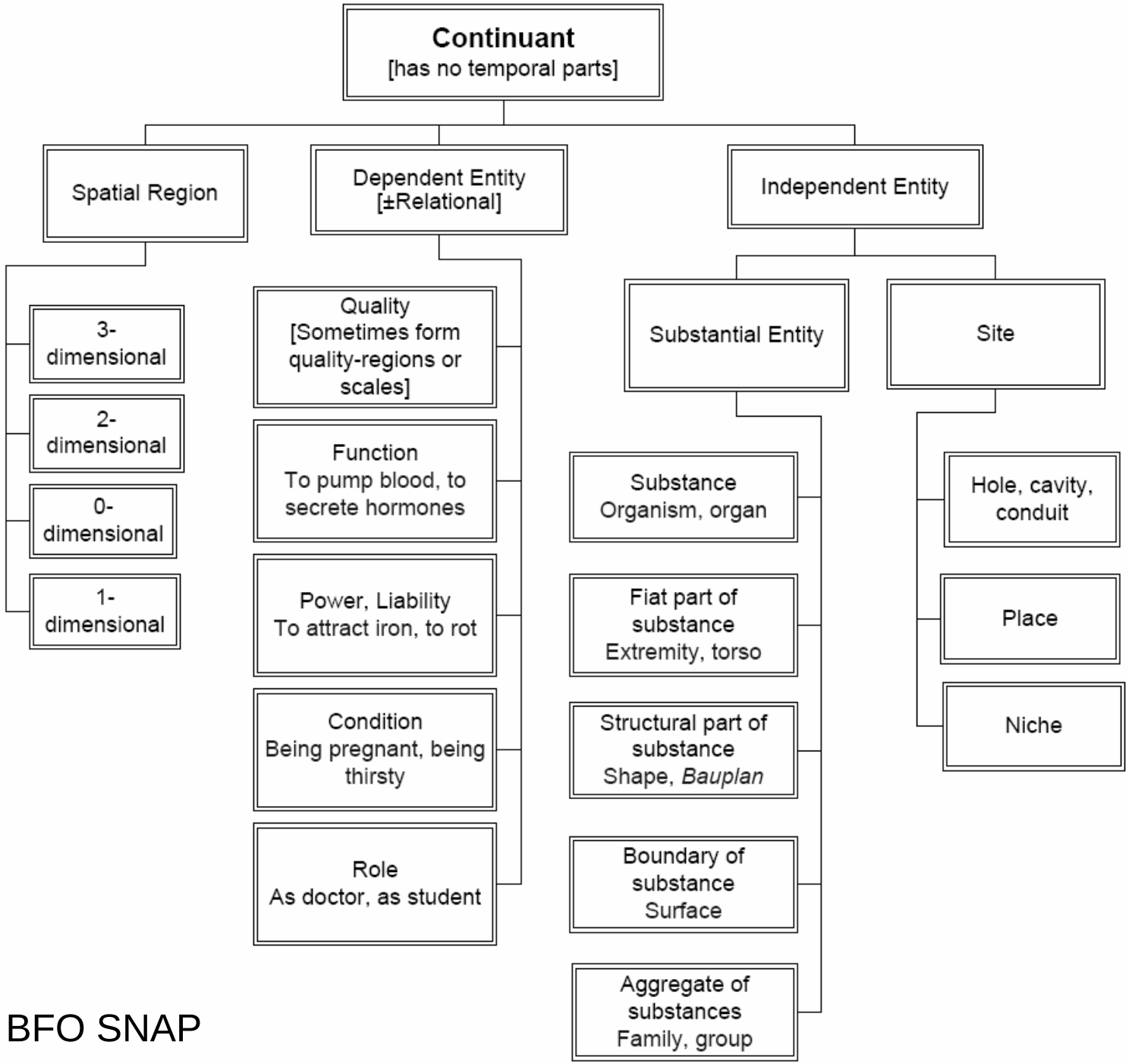
Domain expertise needs to be used in combination with ontological principles.

We can’t simply model what domain experts tell us, we need to situate that knowledge within a comprehensive system.

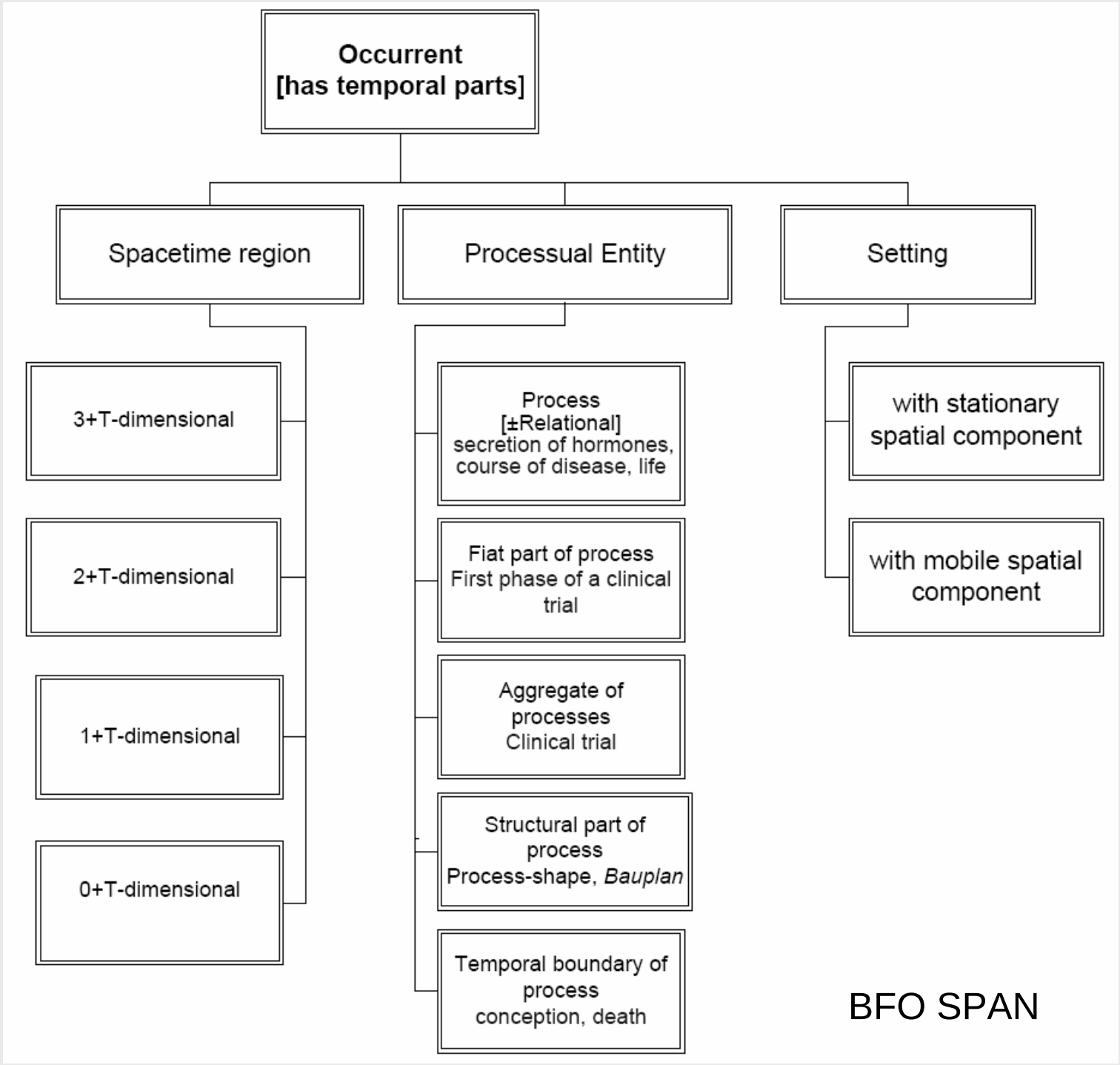
Some formal-ontological distinctions



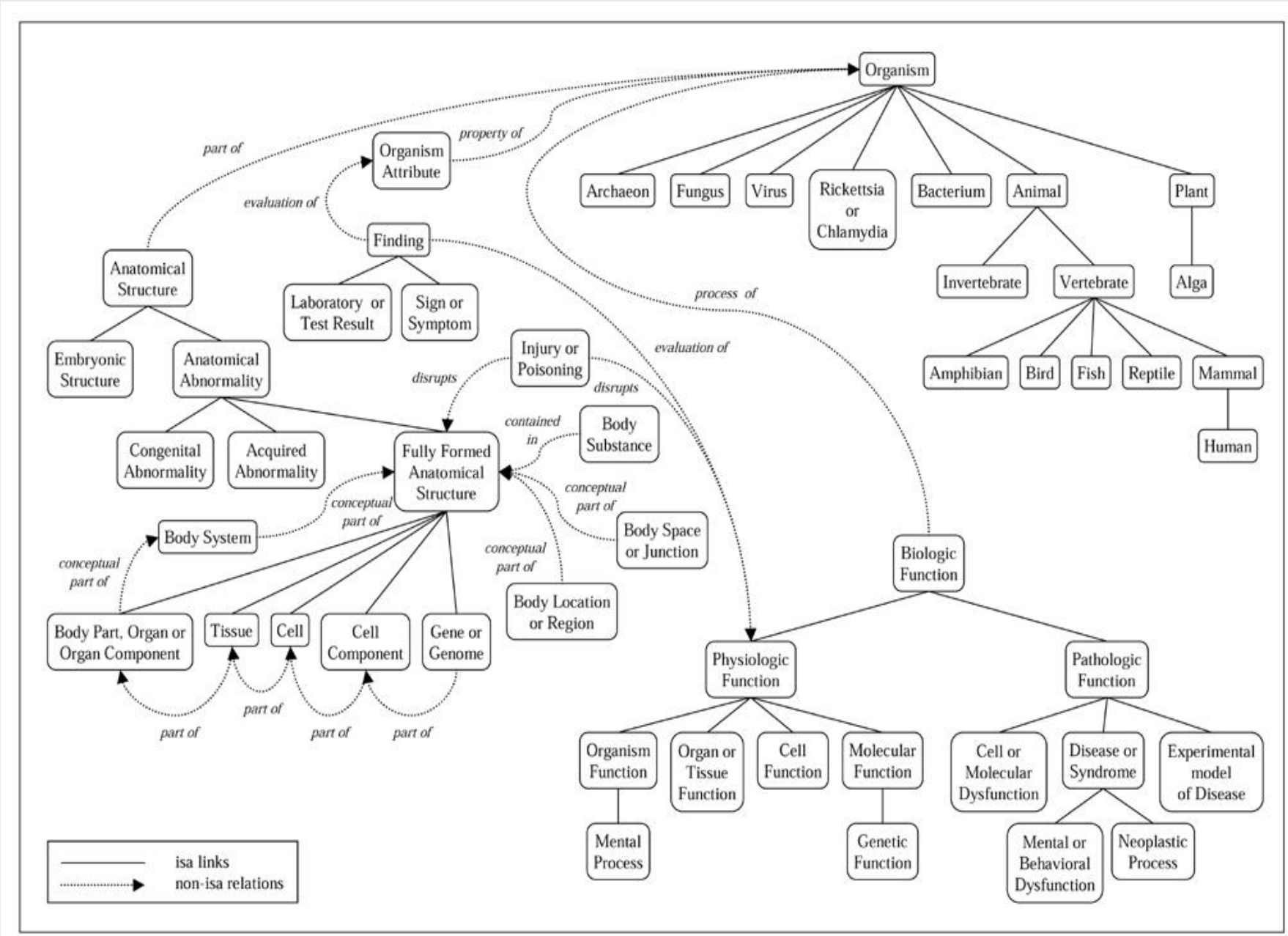

Substantial Universal
differentiates
Quality Universal
Process Universal
exemplifies
instantiates
instantiates
instantiates
has participant
Substantial Particular
Quality Particular
Process Particular
inheres


Continuant
[has no temporal parts]
Spatial Region
Dependent Entity
[±Relational]
Independent Entity
3-dimensional
2-dimensional
0-dimensional
1-dimensional
Quality
[Sometimes form quality-regions or scales]
Function
To pump blood, to secrete hormones
Power, Liability
To attract iron, to rot
Condition
Being pregnant, being thirsty
Role
As doctor, as student
Substantial Entity
Site
Substance
Organism, organ
Fiat part of substance
Extremity, torso
Structural part of substance
Shape, Bauplan
Boundary of substance
Surface
Aggregate of substances
Family, group
Hole, cavity, conduit
Place
Niche

BFO SNAP


Occurrent
[has temporal parts]
Spacetime region
Processual Entity
Setting
3+T-dimensional
2+T-dimensional
1+T-dimensional
0+T-dimensional
Process
[±Relational]
secretion of hormones, course of disease, life
Fiat part of process
First phase of a clinical trial
Aggregate of processes
Clinical trial
Structural part of process
Process-shape, *Bauplan*
Temporal boundary of process
conception, death
with stationary spatial component
with mobile spatial component

BFO SPAN



A portion of the current Semantic Network

Enhancing Biomedical Ontologies through Alignment of Semantic Relationships:

Exploratory Approaches

Lowell Vizenor

Olivier Bodenreider

Lee Peters

Alexa T McCray

Submitted to AMIA 2006

❒ **Objective:** investigates several methods for aligning Metathesaurus relationships with their counterparts in the UMLS Semantic Network.

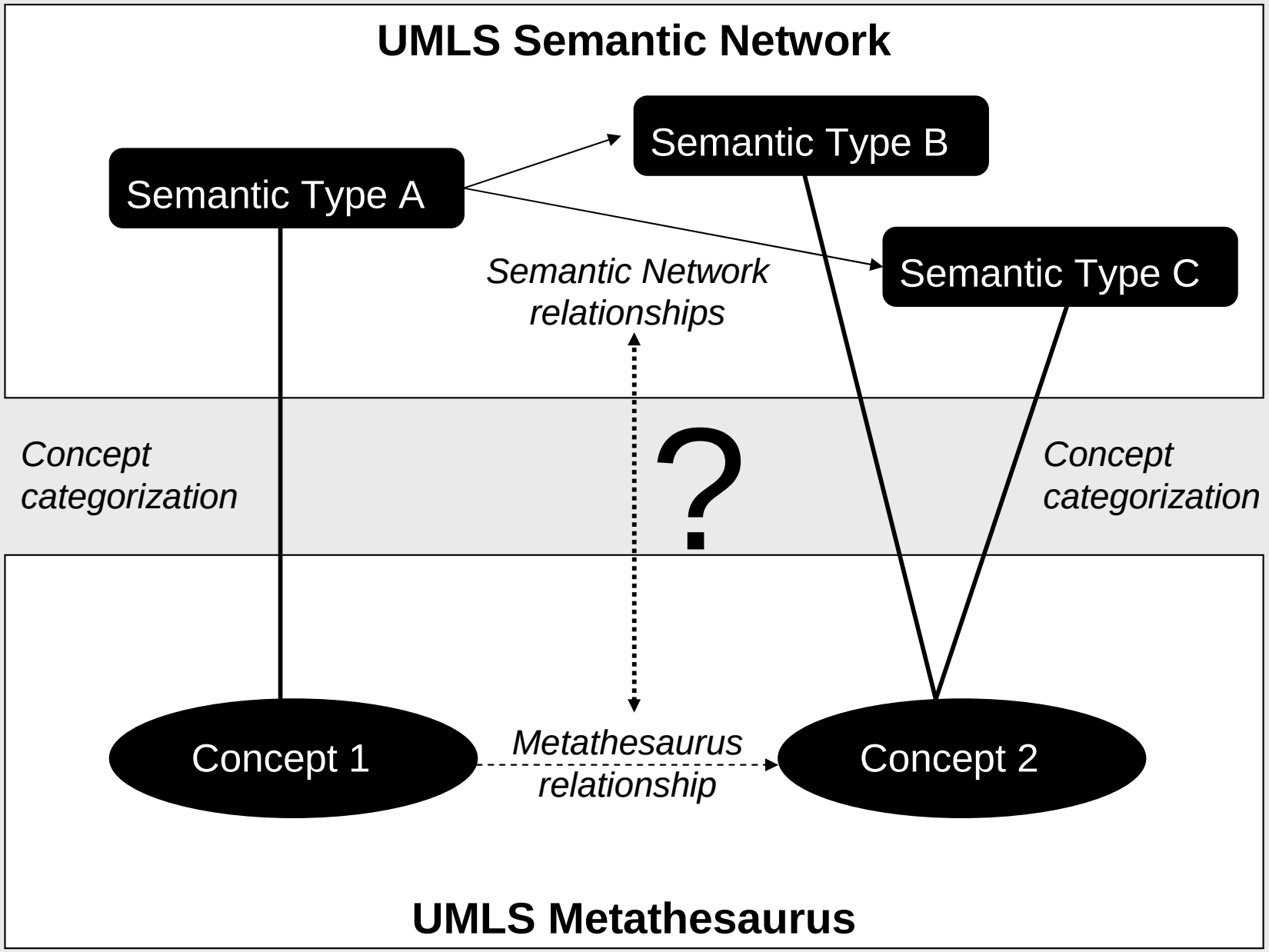
❒ **Ontology Alignment**

The UMLS is a two level structure

1. Metathesaurus
2. Semantic Network

Alignment of relationships (not concepts) across ontologies

Ontologies represent knowledge at widely different levels of granularity


UMLS Semantic Network
Semantic Type A
Semantic Type B
Semantic Type C
Semantic Network relationships
Concept categorization
?
Concept categorization
Concept 1
Metathesaurus relationship
Concept 2
UMLS Metathesaurus

UMLS Metathesaurus

- Large repository of interrelated concepts coming from one hundred biomedical vocabularies
- Over 1 million concepts
- 139 (unique) relationships

Thesaural Relationships

E.g., parent/child, broader than/narrower than

Specified Relationships

E.g., *isa*, *location_of*, *ingredient_of*, *manisfestation_of*, *mapped_to*

The semantics of metathesaurus relationships is implicit; no definitions are given

UMLS Semantic Network

- A small, manually curated high-level network
- 135 Semantic Network Types
- 54 Semantic Network Relationships

Each SN Rel has an inverse, a textual definition, and a list of Semantic Types that are linked by the relationship.

isa

associate_with

- functionally_related_to
- physically_related_to
- spatially_related_to
- temporally_related_to
- conceptually_related_to

Methods

Four Methods for eliciting the Semantics of Metathesaurus Relationships

Metathesaurus-Centric:

1. Manual elicitation
2. Abstraction at the level of high level concepts
3. Abstraction at the level of Semantic Types.

Semantic Network-Centric

4. Top-down elicitation

Metathesaurus-Centric

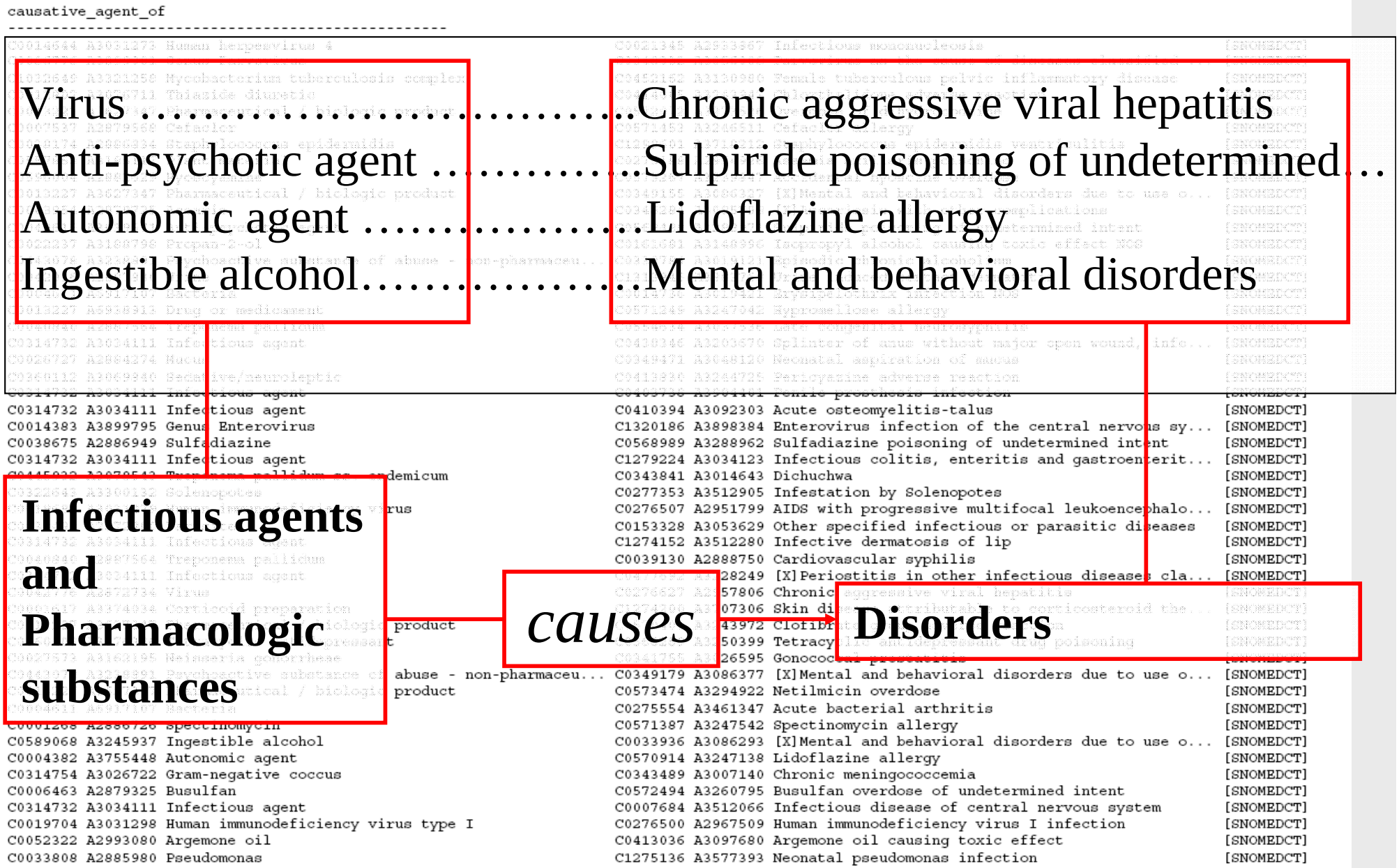
1) Manual elicitation

2 random samples of a maximum of 50 relations per Metathesaurus relationship

Link the MT Rel to SN Rel (when possible)

and

Identify the type of relationship (e.g. semantically equivalent, narrower than or broader than).


causative_agent_of
Virus ……Chronic aggressive viral hepatitis
Anti-psychotic agent ……Sulpiride poisoning of undetermined…
Autonomic agent ……Lidoflazine allergy
Ingestible alcohol……Mental and behavioral disorders
C0314732 A3034111 Infectious agent
C0014383 A3899795 Genus Enterovirus
C0038675 A2886949 Sulfadiazine
C0314732 A3034111 Infectious agent
C0589068 A3245937 Ingestible alcohol
C0004382 A3755448 Autonomic agent
C0314754 A3026722 Gram-negative coccus
C0006463 A2879325 Busulfan
C0314732 A3034111 Infectious agent
C0019704 A3031298 Human immunodeficiency virus type I
C0052322 A2993080 Argemone oil
C0033808 A2885980 Pseudomonas
C0410394 A3092303 Acute osteomyelitis-talus [SNOMEDCT]
C1320186 A3898384 Enterovirus infection of the central nervous sy... [SNOMEDCT]
C0568989 A3288962 Sulfadiazine poisoning of undetermined intent [SNOMEDCT]
C1279224 A3034123 Infectious colitis, enteritis and gastroenterit... [SNOMEDCT]
C0343841 A3014643 Dichuchwa [SNOMEDCT]
C0277353 A3512905 Infestation by Solenopotes [SNOMEDCT]
C0276507 A2951799 AIDS with progressive multifocal leukoencephalo... [SNOMEDCT]
C0153328 A3053629 Other specified infectious or parasitic diseases [SNOMEDCT]
C1274152 A3512280 Infective dermatosis of lip [SNOMEDCT]
C0039130 A2888750 Cardiovascular syphilis [SNOMEDCT]
C0349179 A3086377 [X]Mental and behavioral disorders due to use o... [SNOMEDCT]
C0573474 A3294922 Netilmicin overdose [SNOMEDCT]
C0275554 A3461347 Acute bacterial arthritis [SNOMEDCT]
C0571387 A3247542 Spectinomycin allergy [SNOMEDCT]
C0033936 A3086293 [X]Mental and behavioral disorders due to use o... [SNOMEDCT]
C0570914 A3247138 Lidoflazine allergy [SNOMEDCT]
C0343489 A3007140 Chronic meningococcemia [SNOMEDCT]
C0572494 A3260795 Busulfan overdose of undetermined intent [SNOMEDCT]
C0007684 A3512066 Infectious disease of central nervous system [SNOMEDCT]
C0276500 A2967509 Human immunodeficiency virus I infection [SNOMEDCT]
C0413036 A3097680 Argemone oil causing toxic effect [SNOMEDCT]
C1275136 A3577393 Neonatal pseudomonas infection [SNOMEDCT]
Infectious agents and Pharmacologic substances
causes
Disorders

Methods

Four Methods for eliciting the Semantics of Metathesaurus Relationships

Metathesaurus-Centric:

1. Manual elicitation
2. Abstraction at the level of high level concepts
3. Abstraction at the level of Semantic Types.

Semantic Network-Centric

4. Top-down elicitation

Metathesaurus-Centric

2) Abstraction at the level of high-level concepts

- Compute the lowest common ancestor for the domain and range of each MT relationship in a given source
- Graph/Frequency/etc.

REL access_instrument_of [SNOMEDCT] (50)
DOM 50 119 49881 C0014243: *Endoscope*
RNG 50 230 49770 C0282493: *Procedure by method*

frequency *total distance* *score*

In some cases, the prototypical relation is uninformative, because the lowest common ancestor is the root of the terminology:

REL associated_finding_of [SNOMEDCT] (50)
DOM 49 215 48785 C1136258 *SNOMED CT Concept*
RNG 31 192 30808 C1136258 *SNOMED CT Concept*

In other cases, there is so much dispersion that the relationship does not have much semantics at all:

REL associated_with [BI] (5)
DOM 1 1 999 C0034951 refractive disorder
DOM 1 1 999 C0020699 hysterectomy
DOM 1 1 999 C0521346 respiratory
DOM 1 1 999 C0235480 paroxysmal atrial fibrillation
DOM 1 1 999 C0014118 endocarditis
RNG 1 1 999 C0155685 acute bacterial endocarditis
RNG 1 1 999 C1270947 partial hysterectomy
RNG 1 1 999 C0004238 atrial fibrillation
RNG 1 1 999 C0205481 ophthalmologic
RNG 1 1 999 C0032739 ppd positive

Methods

Four Methods for eliciting the Semantics of Metathesaurus Relationships

Metathesaurus-Centric:

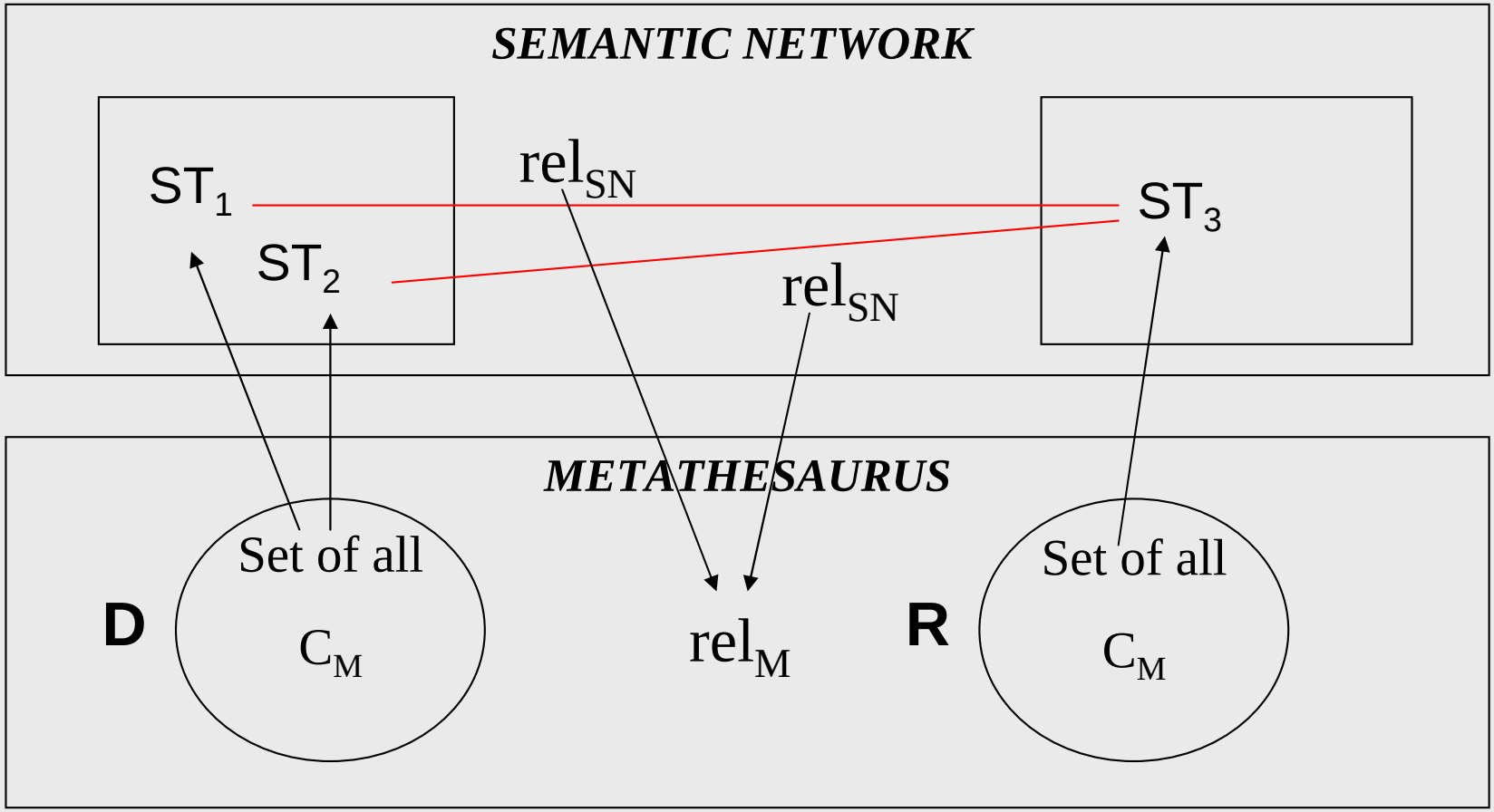
1. Manual elicitation
2. Abstraction at the level of high level concepts
3. Abstraction at the level of Semantic Types.

Semantic Network-Centric

4. Top-down elicitation

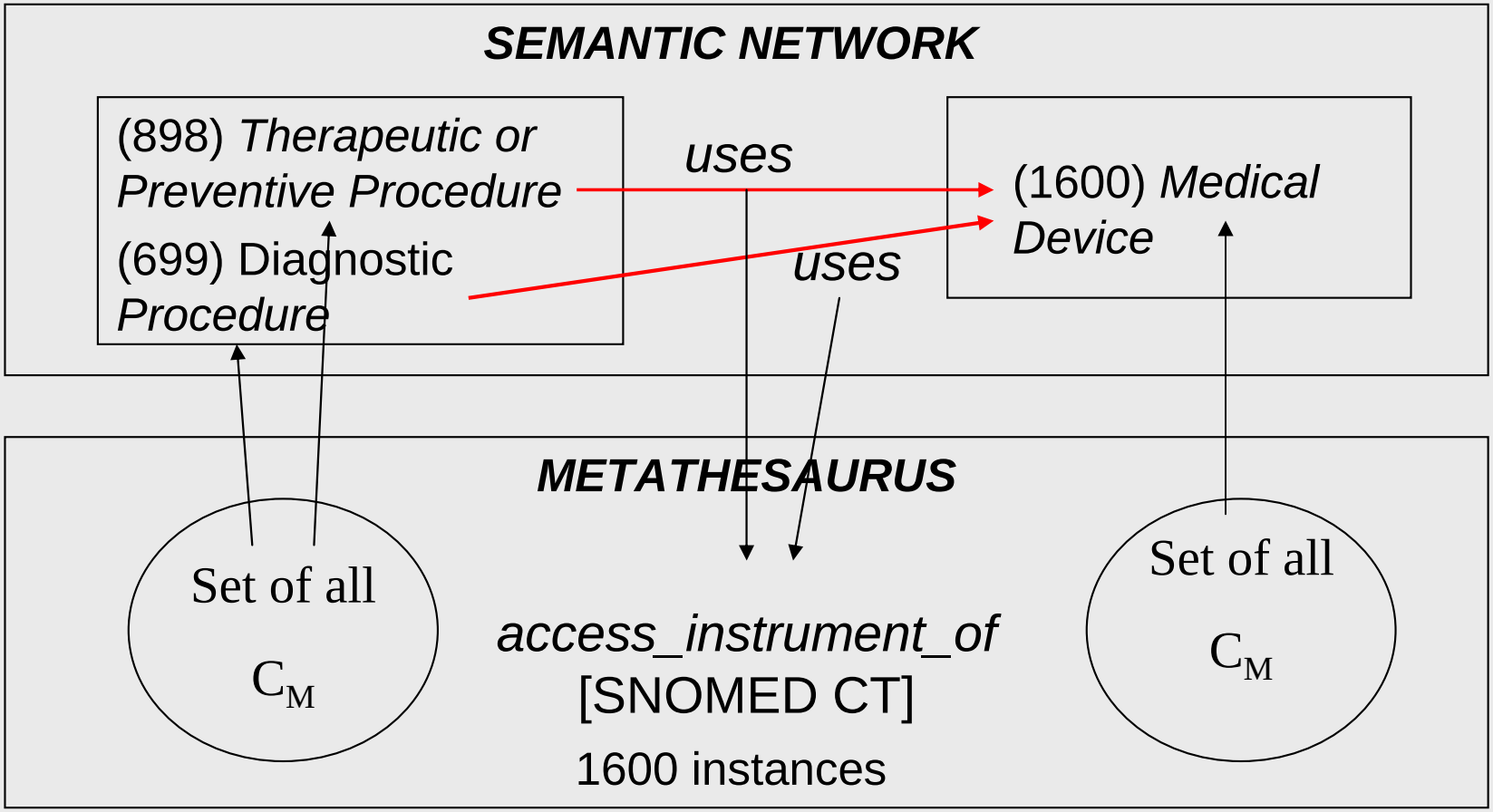
Metathesaurus-centric

3) Abstraction at the level of Semantic Types



Metathesaurus-centric

3) Abstraction at the level of Semantic Types



Methods

Four Methods for eliciting the Semantics of Metathesaurus Relationships

Metathesaurus-Centric:

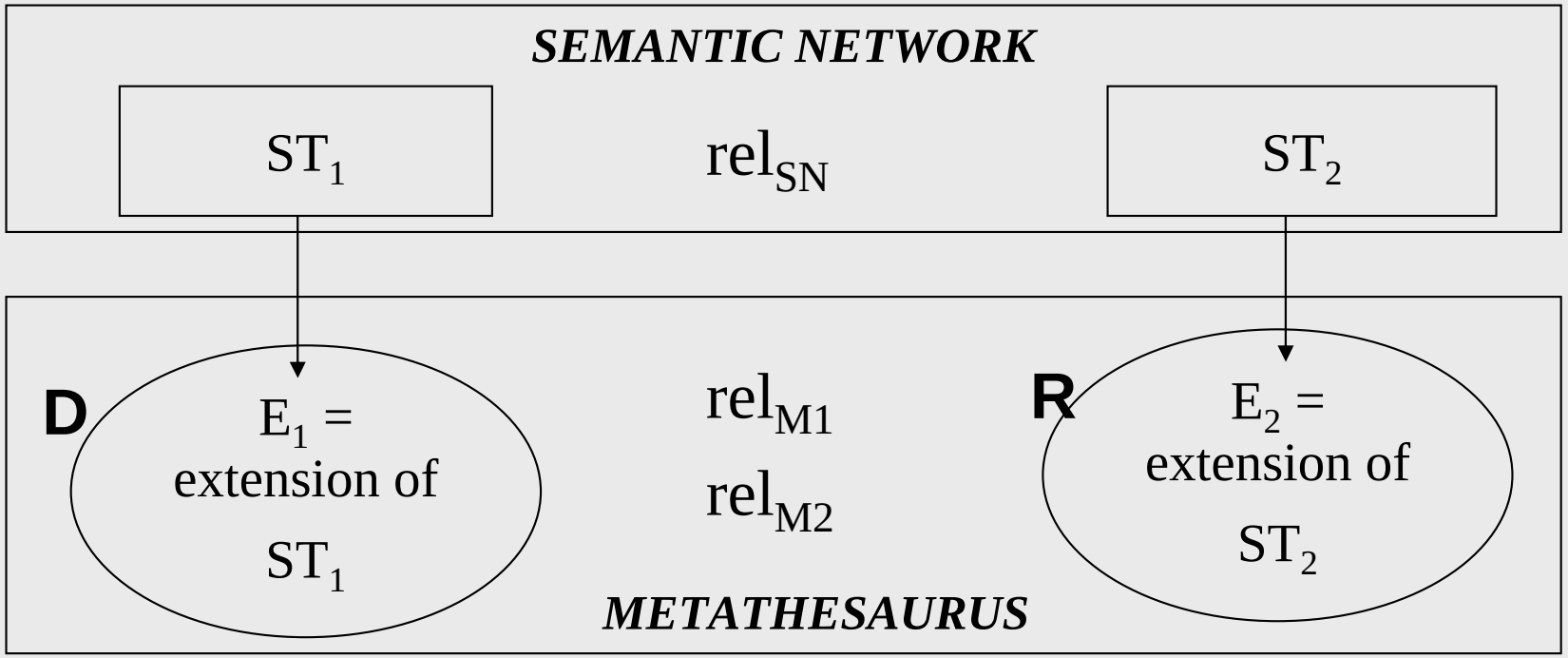
1. Manual elicitation
2. Abstraction at the level of high level concepts
3. Abstraction at the level of Semantic Types.

Semantic Network-Centric

4. Top-down elicitation

Semantic Network-Centric

4)Top-down elicitation



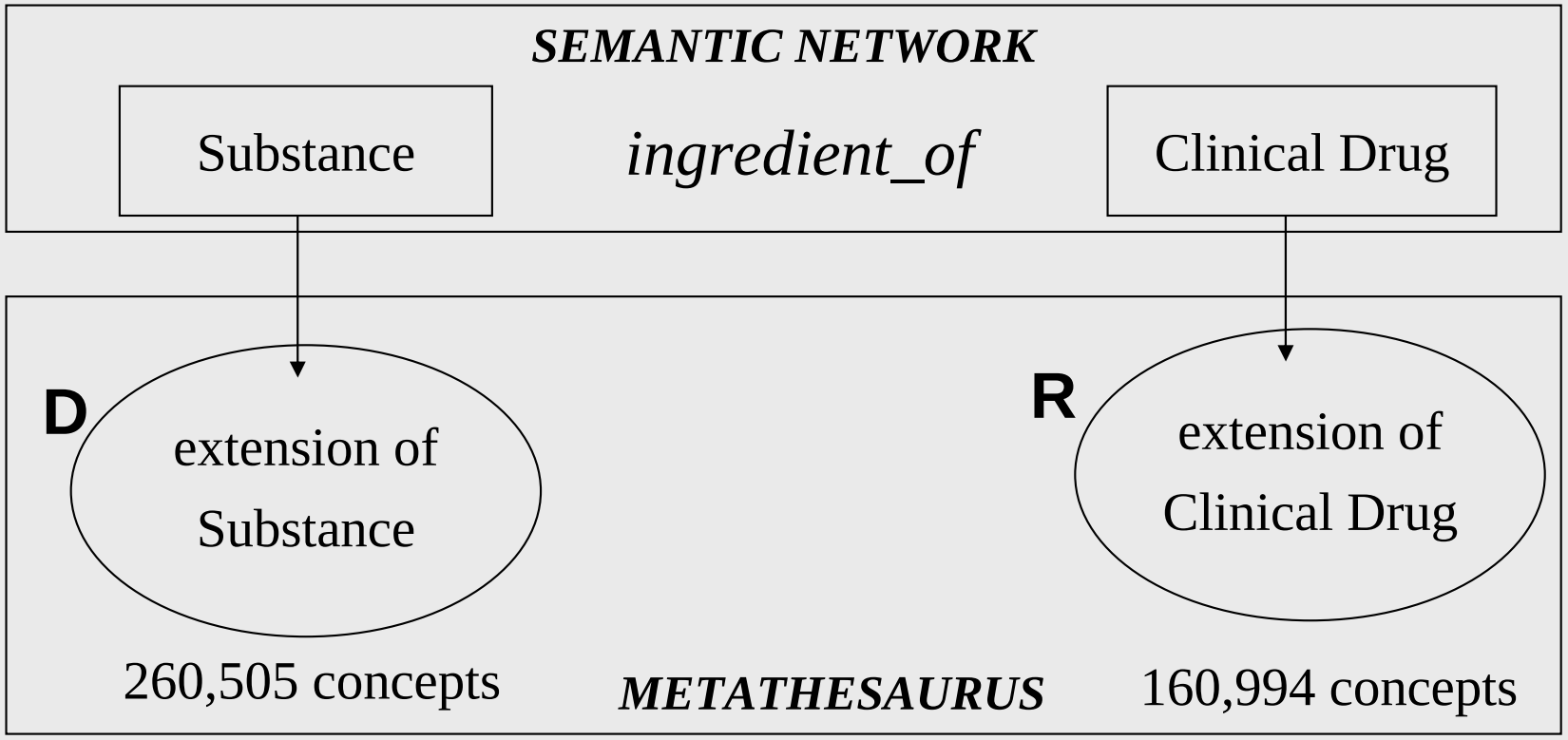
We examine what relations are represented in the Metathesaurus, pairwise, between a concept from E_1 and a concept from E_2 and obtain a set of Metathesaurus relationships

$$\{rel_{M1}, rel_{M2}, \ldots\}$$

along with frequency info for each rel_M

Semantic Network-Centric

4)Top-down elicitation



High frequency:

active_ingredient_of
dose_form_of
ingredient_of

Of note:

metabolizes
has_contraindication

AN EXTENDED EXAMPLE

finding_site_of (SNOMED CT)

I. Manual elicitation

Brain tissue structure | *Trace alternate EEG pattern*

Endocrine structure | *External endometriosis*

Gallbladder structure | *Malignant tumor of gallbladder*

Skin structure | *Epithelioma based cell*

Stomach wall structure | *Gastromalacia*

A specification of the SN relationship *location_of*

domain: *anatomical structure*

range: *disorders*

AN EXTENDED EXAMPLE

finding_site_of (SNOMED CT)

II. High level concepts

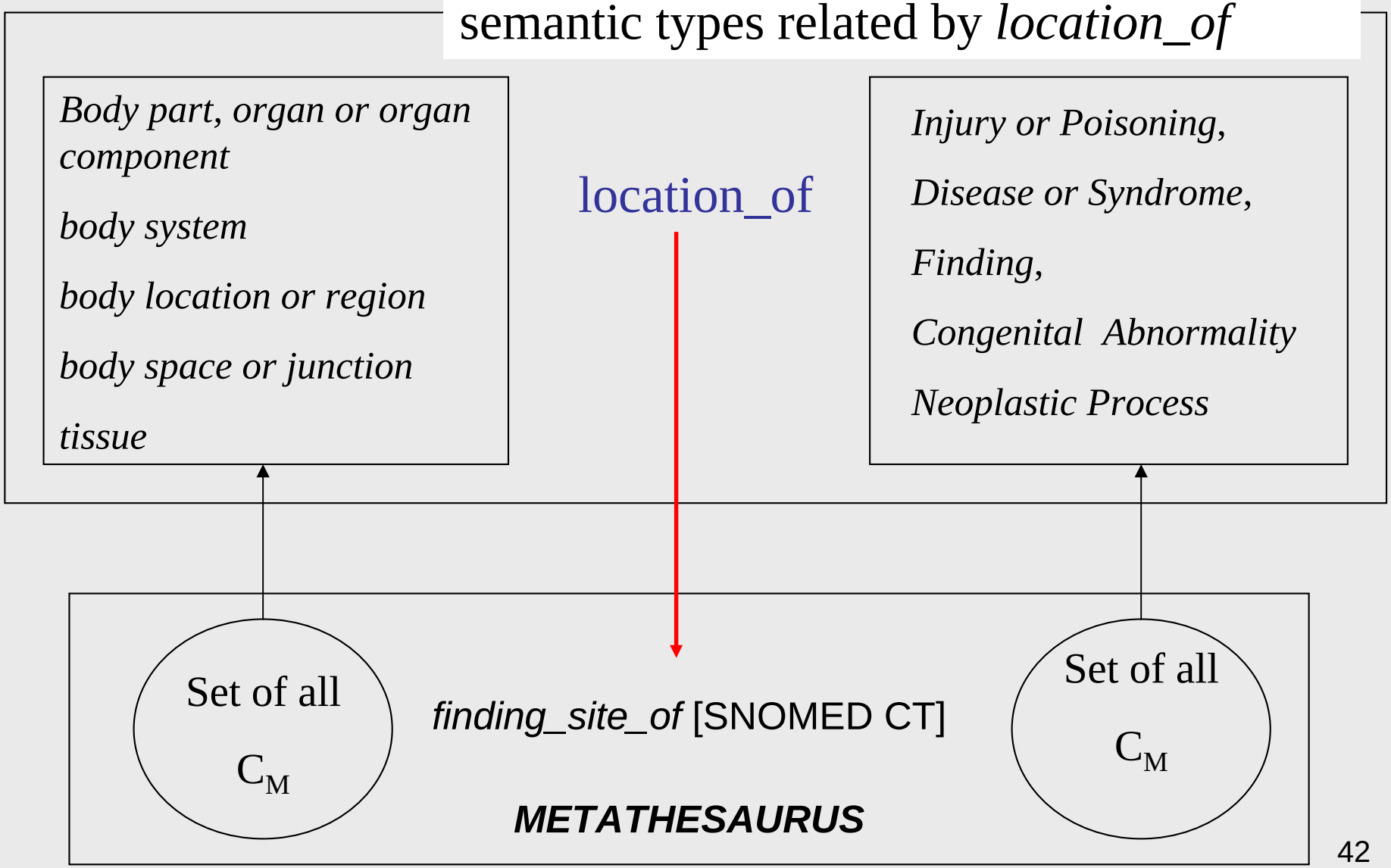
REL finding_site_of [SNOMEDCT] (50)
DOM 50 xxx xxxxx C… Anatomical structure
RNG 50 xxx xxxxx C… *SNOMED CT Concept*

The range concepts belong to several distinct hierarchies in SNOMED CT

EXTENDED EX

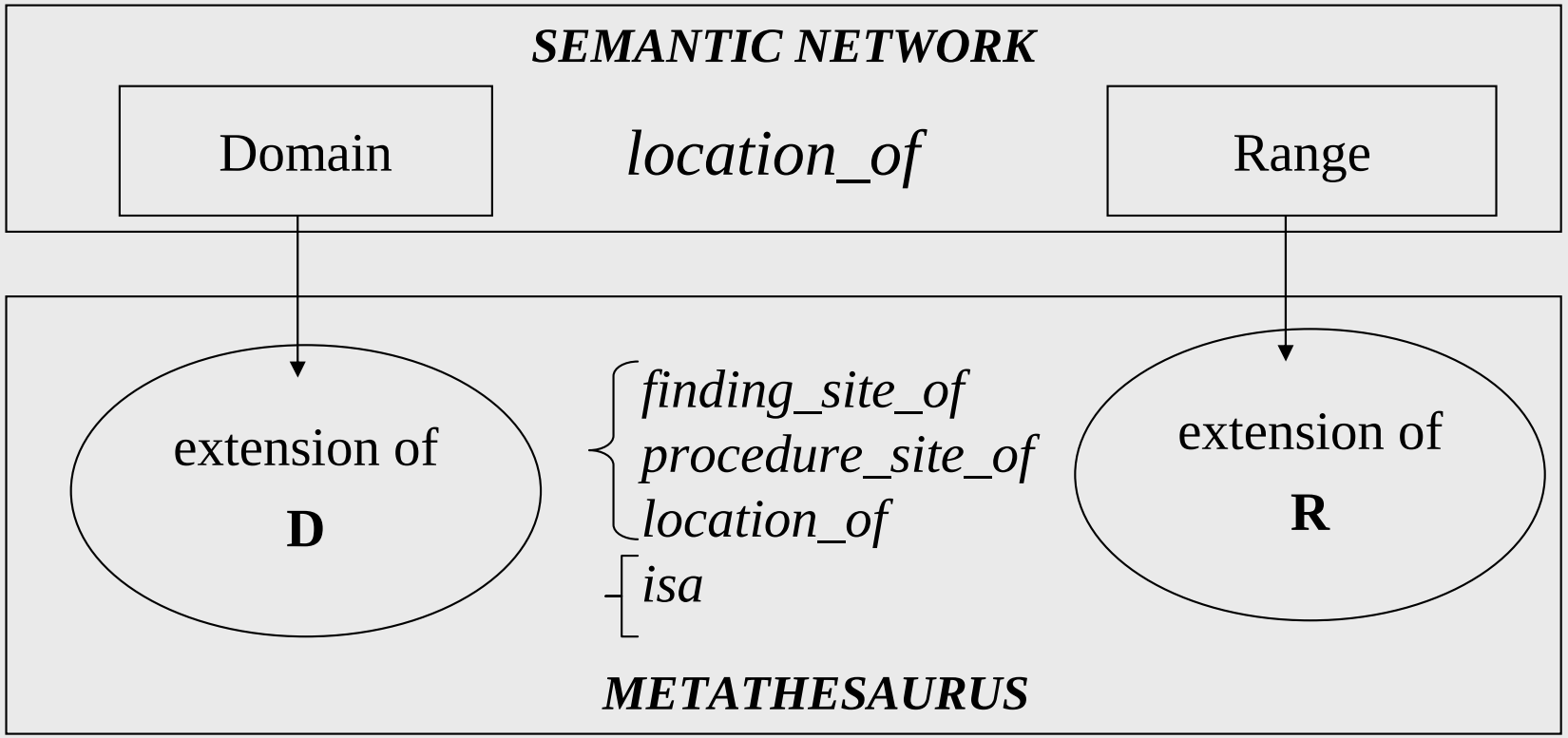
3) Abstraction at

99.5% of the 63,655 pairs of MT concepts related by *finding_site_of* have their semantic types related by *location_of*



EXTENDED EXAMPLE

4)Top-down elicitation



NOTE: some abnormal anatomical structures are considered diseases

(e.g. *Bladder fistula* isa *Bladder disease)*

RESULTS

139 relationships present in the UMLS Metathesaurus

- SNOMED CT (62)
- LOINC (15)
- NDFRT (15)
- FMA (8)
- RxNorm (7)

116 unique to a specific vocabulary

23 are found in two or more vocabularies

RESULTS

We aligned 80 (58%) of the Metathesaurus relationships with Semantic Network relationships

Alignment at a course level of granularity

e.g. *metabolic_site_of < functionally_related_to*

At a more fine grained level of granularity

e.g. *focus_of < issue_in*

27 cases identical relationship

e.g. *affects, process_of, ingredient_of*

RESULTS

59 Metathesaurus relationships fall into a number of additional categories

Lexical relations

e.g. *british_form_of, xml_form_of, suffix_of*

Mapping relations

e.g. *see_from, uniquely_mapped_from*

Vocabulary Management relations

e.g. *classifies, moved_from, replaces*

CONCLUSIONS

The methods used here are good indicators of the meaning of a relationship, but they are not a substitute for an explicit definition.

Defined relationships would make it easier to integrate into other vocabularies.

2 objectives

First, improve the usefulness of vocabulary-specific relationships in the context of the UMLS

Second use these methods as a starting point for the development of a comprehensive ontology of biomedical relationships.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.

A special thanks to May Cheh for her support during my stay at the NLM.