New Jersey's Science
and Technology University
Computer Science Department

October 23, 2007

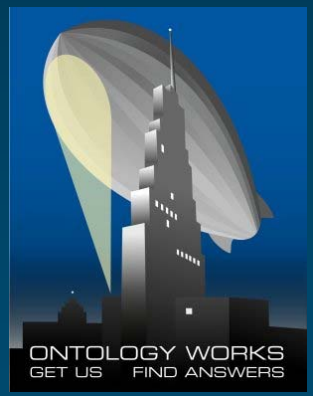
Aligning relationships in the UMLS
Methods and preliminary results

Olivier Bodenreider

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

Acknowledgments

- Alexa T. McCray
  Harvard Medical School
  Boston, Massachusetts



- Lowell Vizenor
  OntologyWorks, Inc
  Baltimore, Maryland

Overview

- The UMLS: A two-level structure
- Aligning relationships
  - Multiple approaches
  - Preliminary results
- Towards an ontology of relationships

The UMLS

A two-level structure

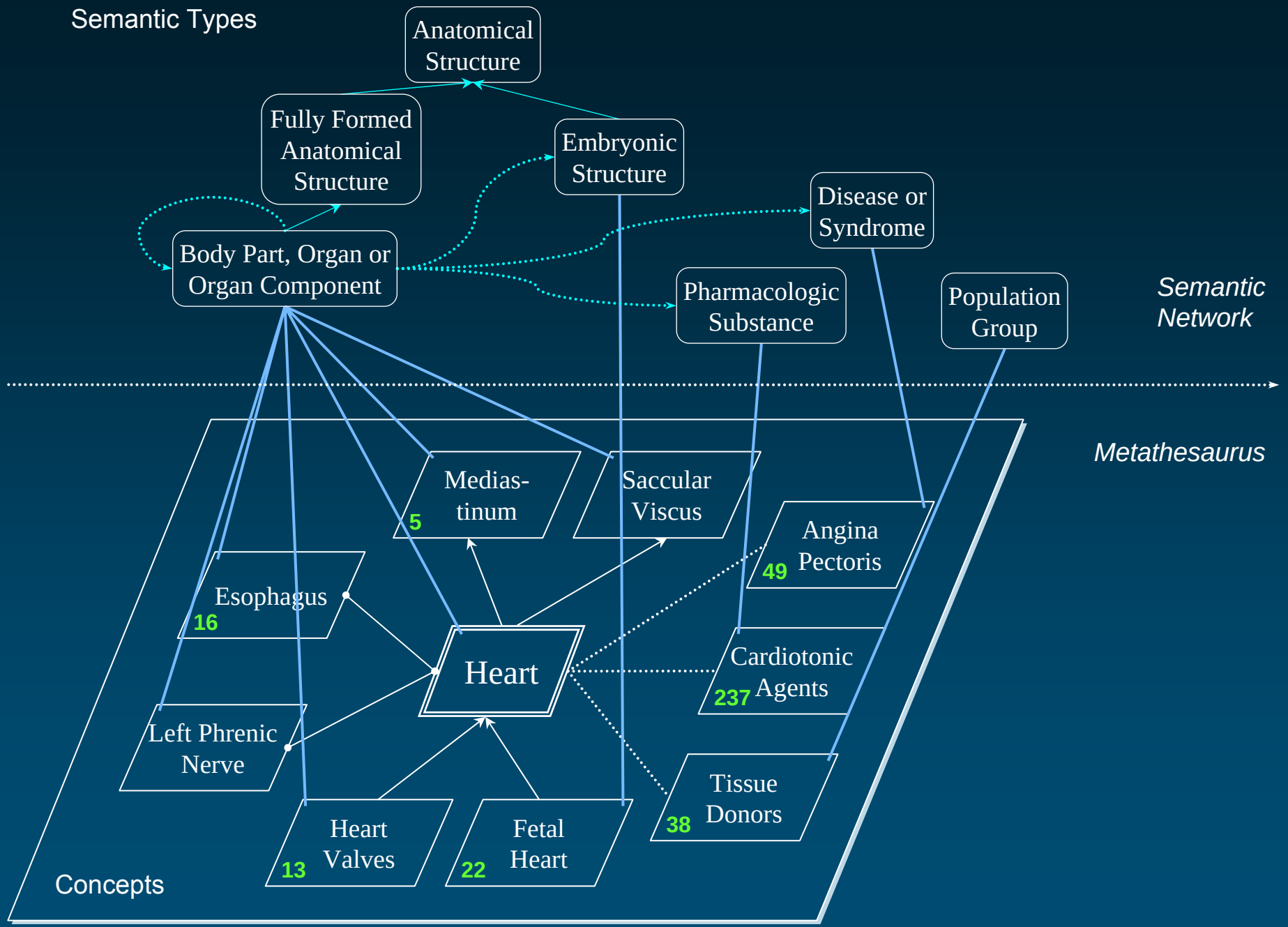
Unified Medical Language System

- SPECIALIST Lexicon
  - 360,000 lexical items
  - Part of speech and variant information
- Metathesaurus
  - 6M names from over 100 terminologies
  - 1.5M concepts
  - 8M relations
- Semantic Network
  - 135 high-level categories
  - 7000 relations among them

Lexical resources

Terminological resources

Ontological resources


Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
5
Saccular Viscus
Angina Pectoris
49
Esophagus
16
Heart
Cardiotonic Agents
237
Left Phrenic Nerve
Tissue Donors
38
Heart Valves
13
Fetal Heart
22
Concepts

UMLS Metathesaurus

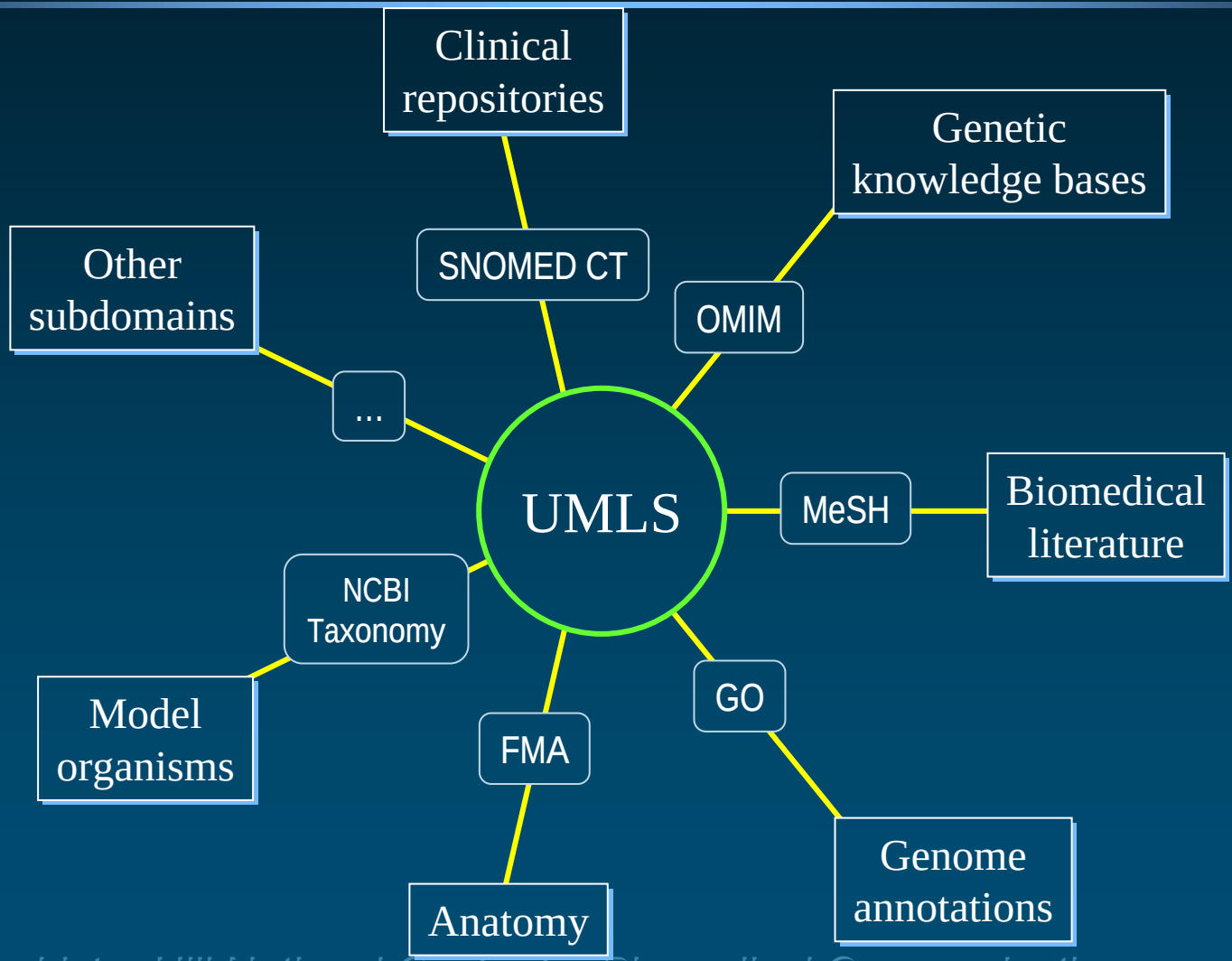
Source Vocabularies (2007AB)

- 143 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.9M names
  - 1.4M concepts
  - 8M relations
- Common presentation

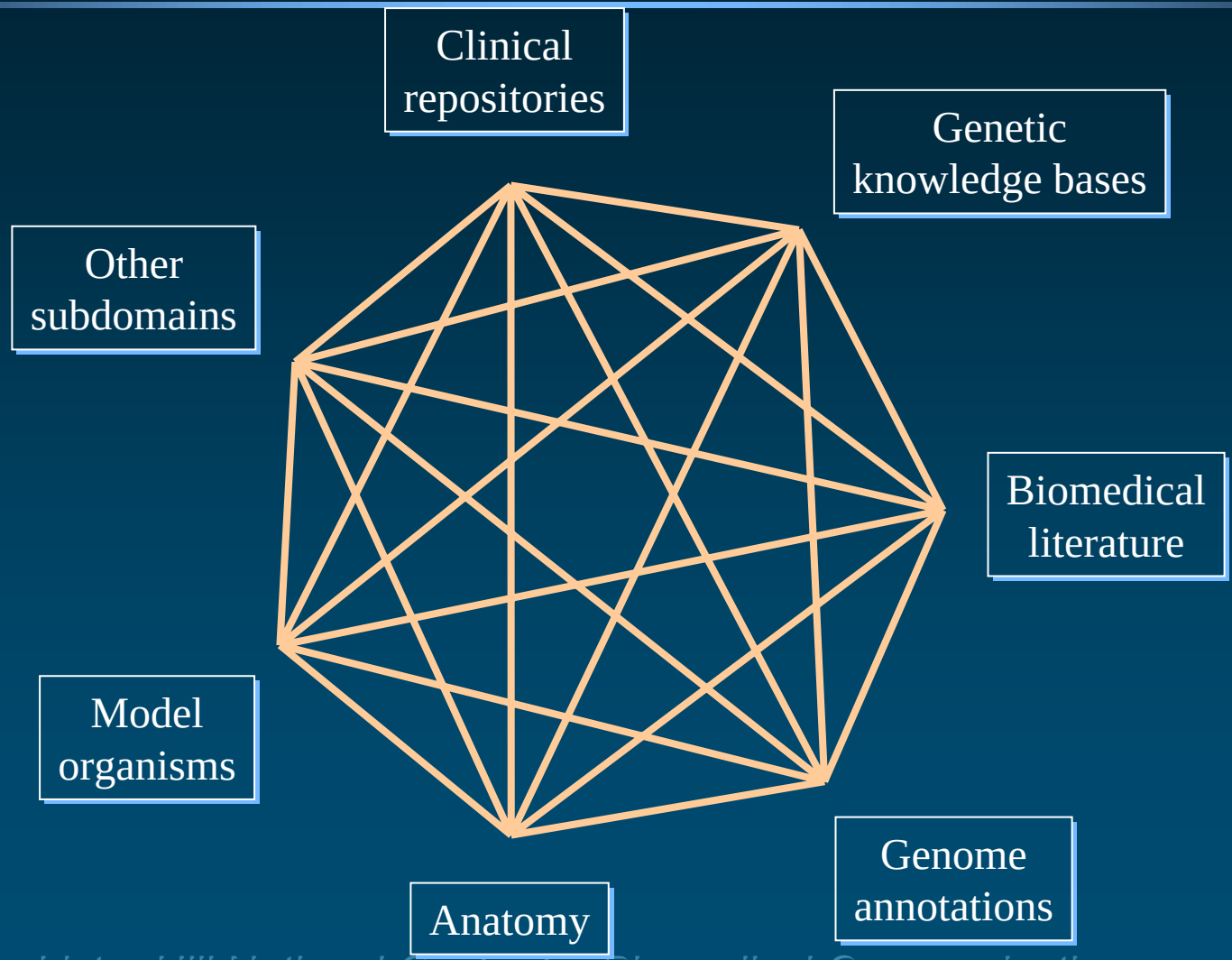
Metathesaurus Basic organization

- Concepts
  - Synonymous terms are clustered into a concept
  - Properties are attached to concepts, e.g.,
    - Unique identifier
    - Definition
- Relations
  - Concepts are related to other concepts
  - Properties are attached to relations, e.g.,
    - Type of relationship
    - Source

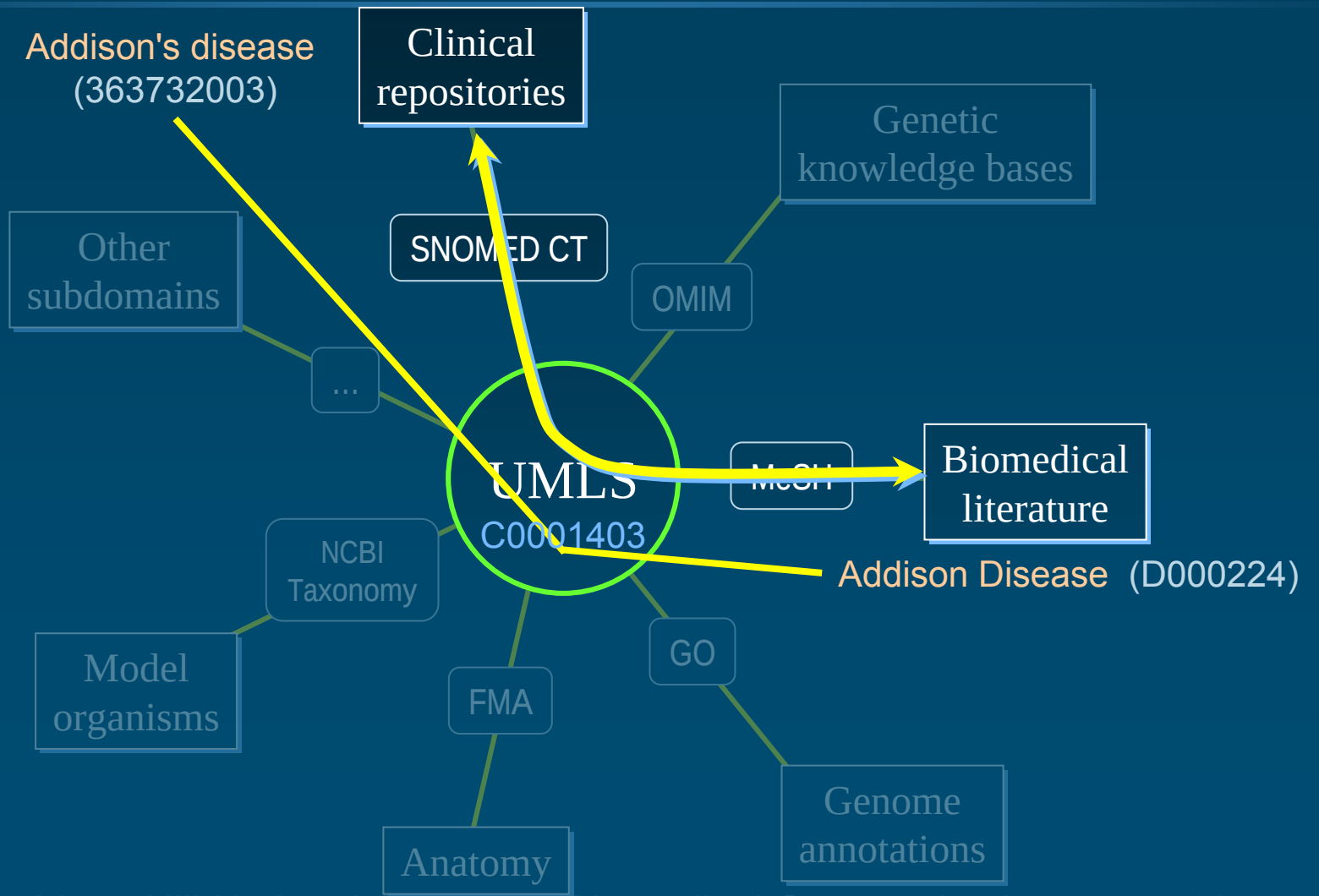
Integrating subdomains



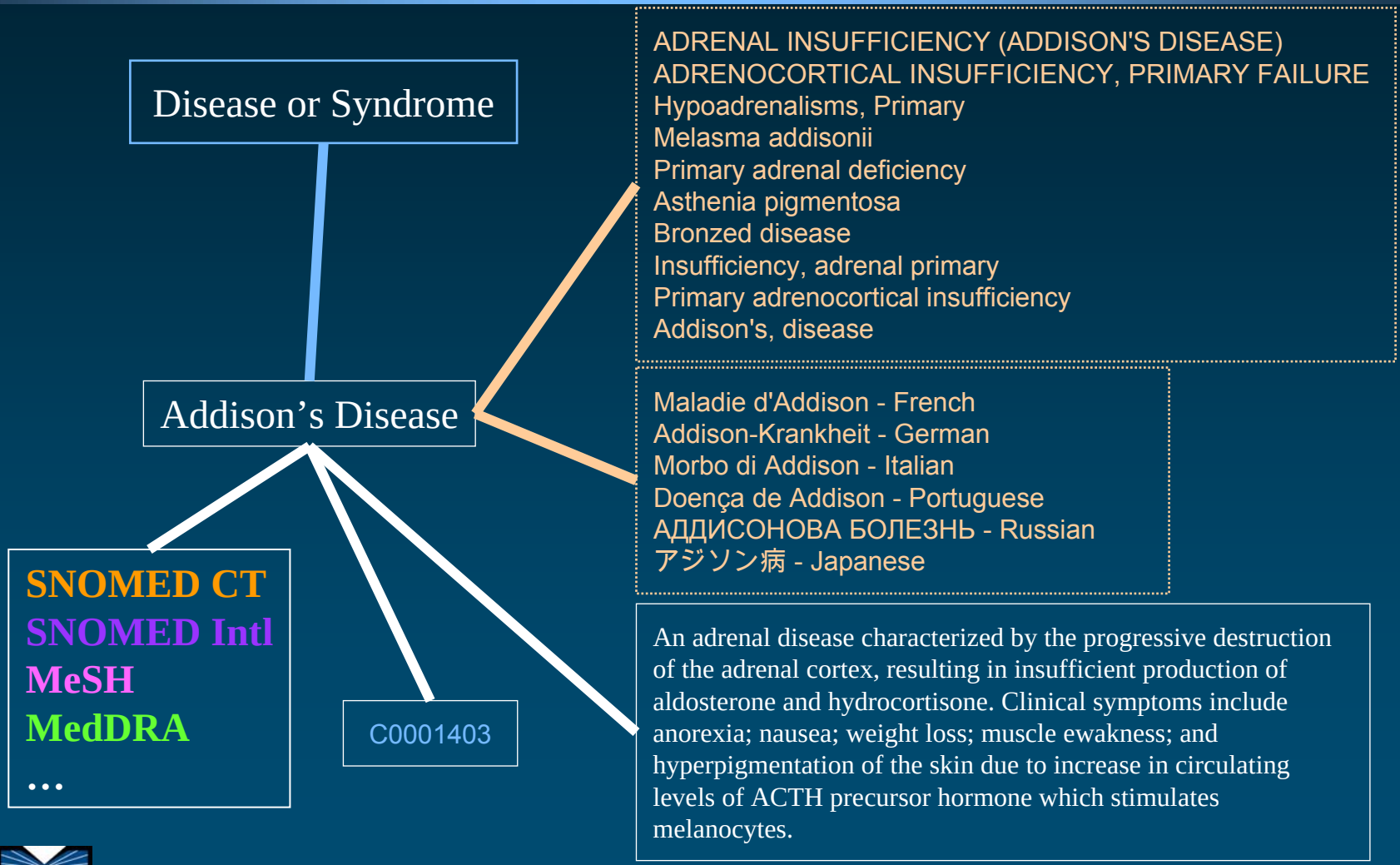
Integrating subdomains



Trans-namespace integration

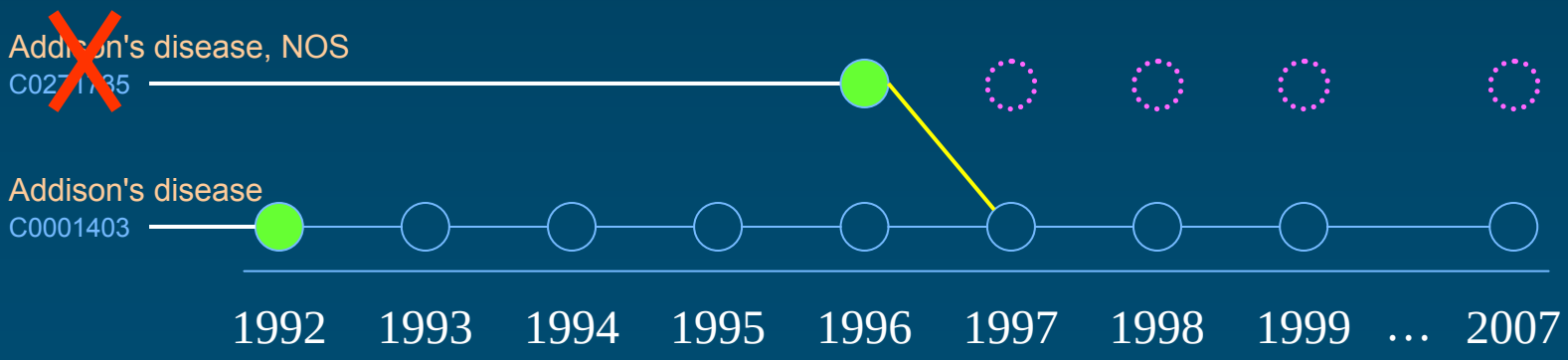


Addison's Disease: Concept



Metathesaurus Evolution over time

- Concepts never die (in principle)
  - CUIs are permanent identifiers
- What happens when they do die (in reality)?
  - Concepts can merge or split
  - Resulting in new concepts and deletions



Metathesaurus Relationships

- Symbolic relations: ~8 M pairs of concepts
- Statistical relations : ~6 M pairs of concepts (co-occurring concepts)
- Mapping relations: ~150,000

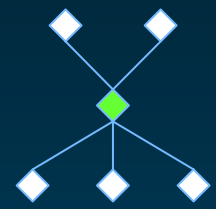
- Categorization: Relationships between concepts and semantic types from the Semantic Network

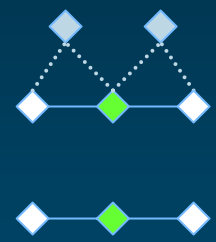
Symbolic relations

- Relation
  - Pair of “atom” identifiers
  - Type
  - Attribute (if any)
  - List of sources (for type and attribute)
- Semantics of the relationship: defined by its type [and attribute]
- Recorded bidirectionally

Symbolic relationships Type

- Hierarchical
  - Parent / Child — **PAR/CHD**
  - Broader / Narrower than — **RB/RN**
- Derived from hierarchies
  - Siblings (children of parents) — **SIB**
- Associative
  - Other — **RO**
- Various flavors of near-synonymy
  - Similar — **RL**
  - Source asserted synonymy — **SY**
  - Possible synonymy — **RQ**



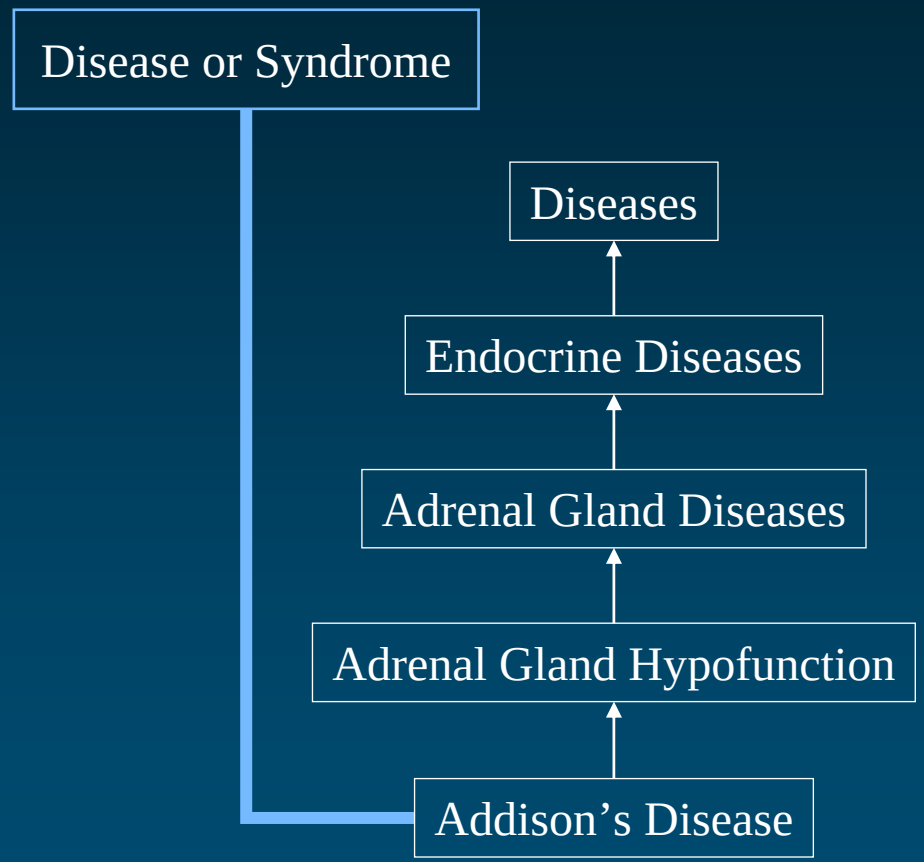


Symbolic relationships Attribute

- Hierarchical
  - isa (is-a-kind-of)
  - part-of
- Associative
  - location-of
  - caused-by
  - treats
  - …
- Cross-references (mapping)

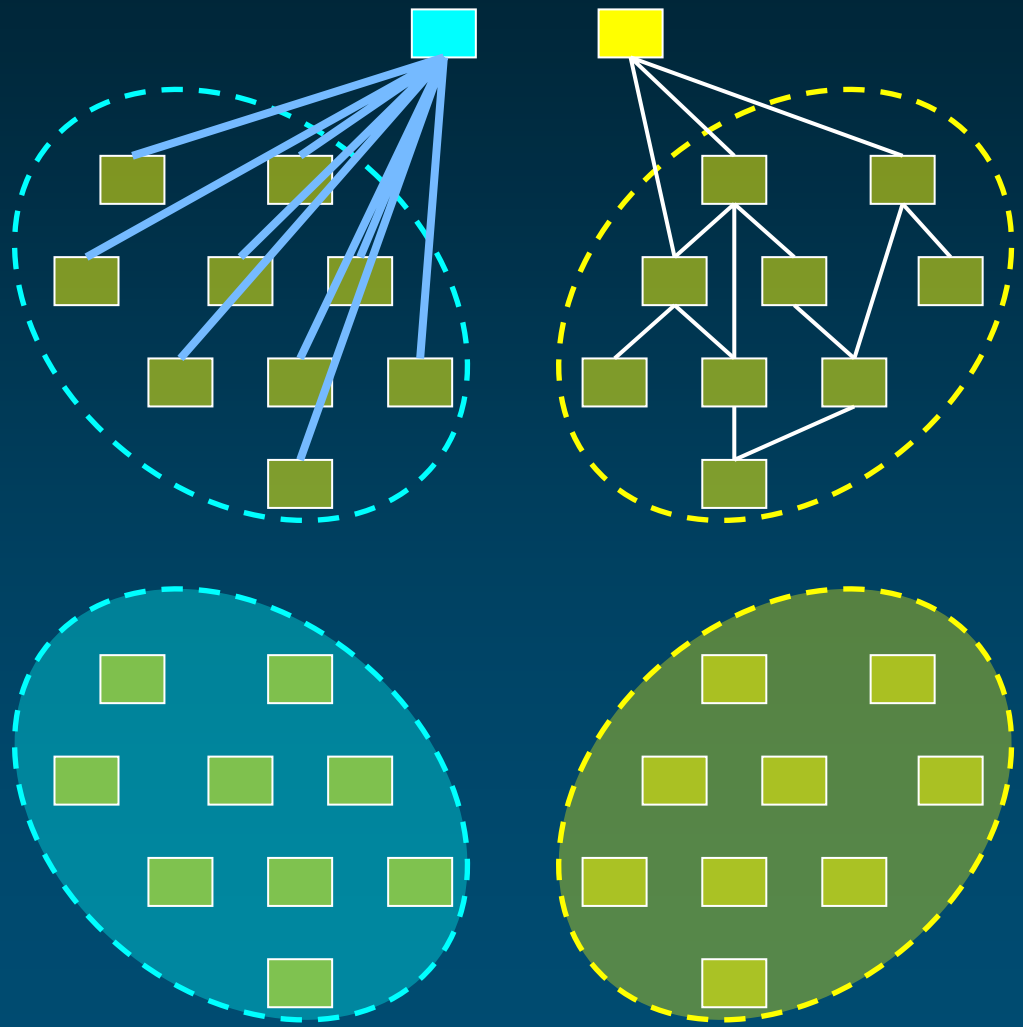
Categorize concepts

- High-level categories (semantic types)
- Assigned by the Metathesaurus editors
- Independently of the hierarchies in which these concepts are located



Categorization vs. hierarchies

- Semantic type
  - List of all concepts having this semantic type
- Concept
  - List of all descendants



UMLS Semantic Network

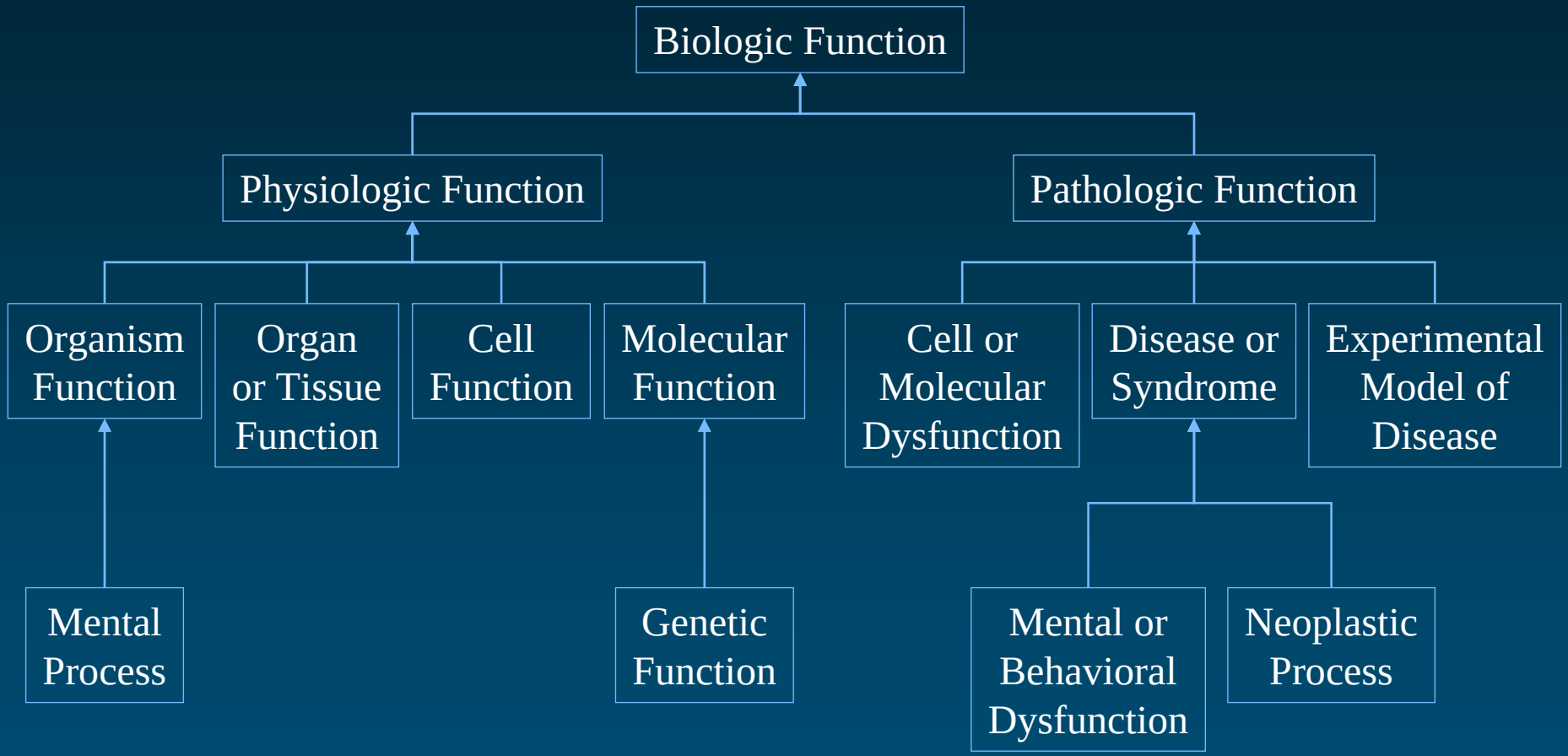
Semantic Network

- Semantic types (135)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process

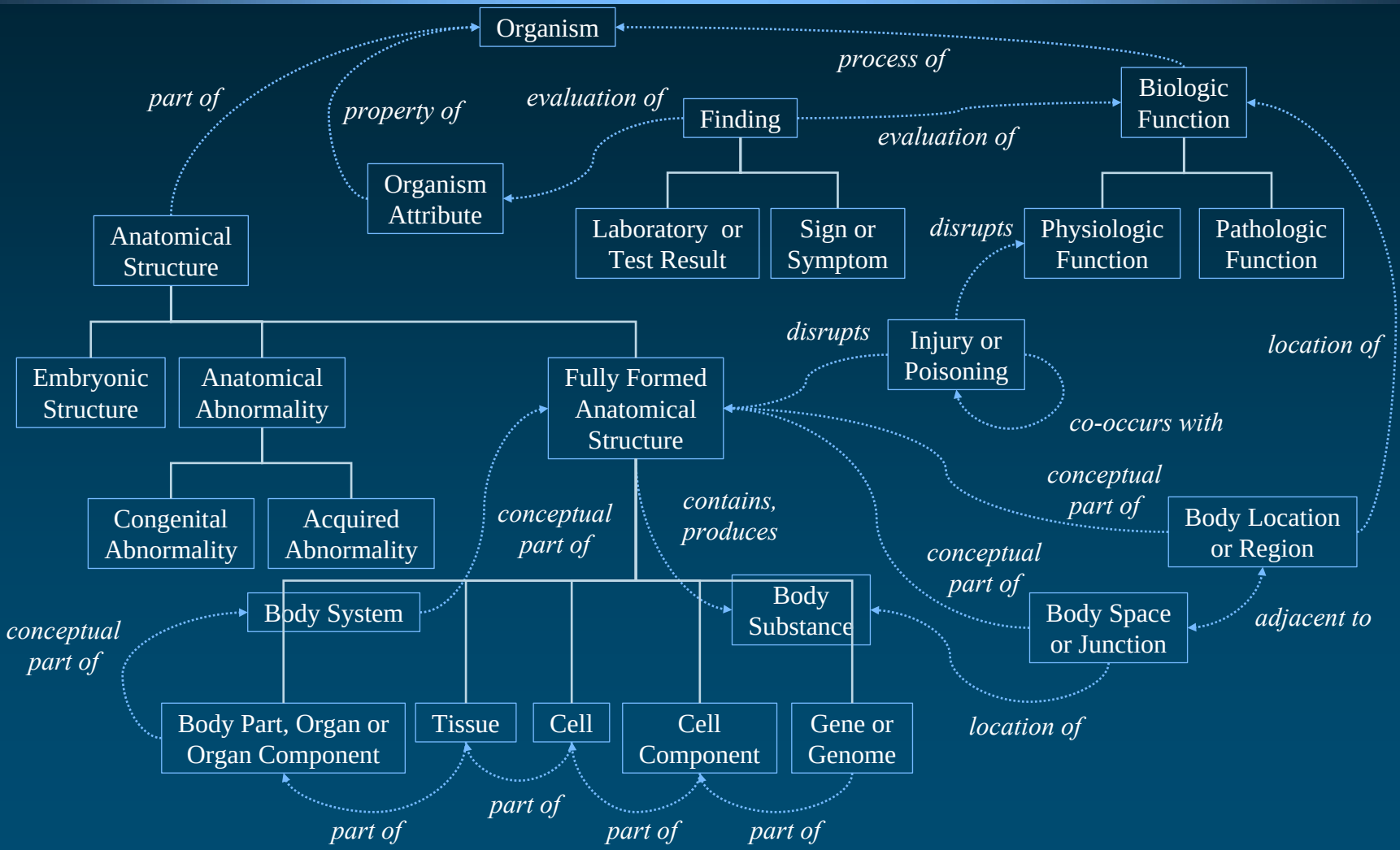
Semantic Network

- Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - Animal *isa* Organism
      - Enzyme *isa* Biologically Active Substance
    - among relations
      - treats *isa* affects
  - non-hierarchical
    - Sign or Symptom *diagnoses* Pathologic Function
    - Pharmacologic Substance *treats* Pathologic Function

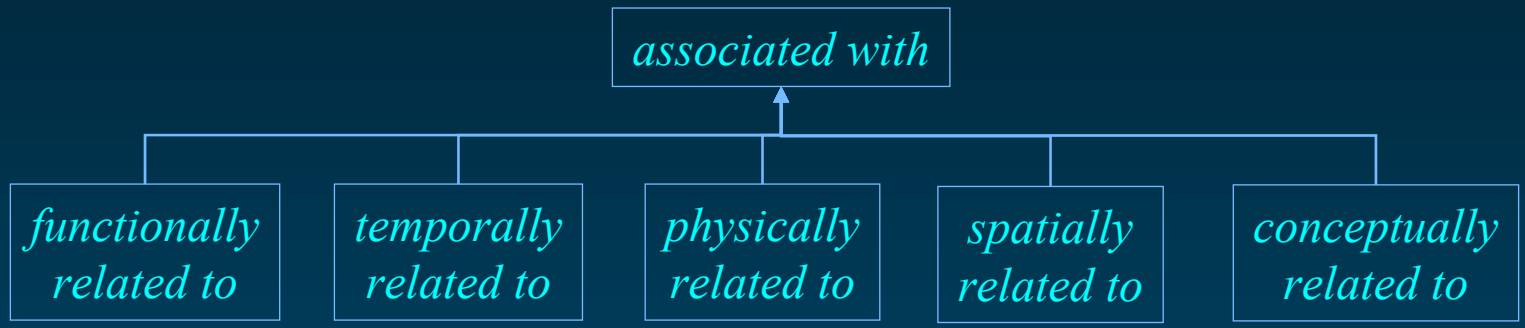
“Biologic Function” hierarchy (isa)

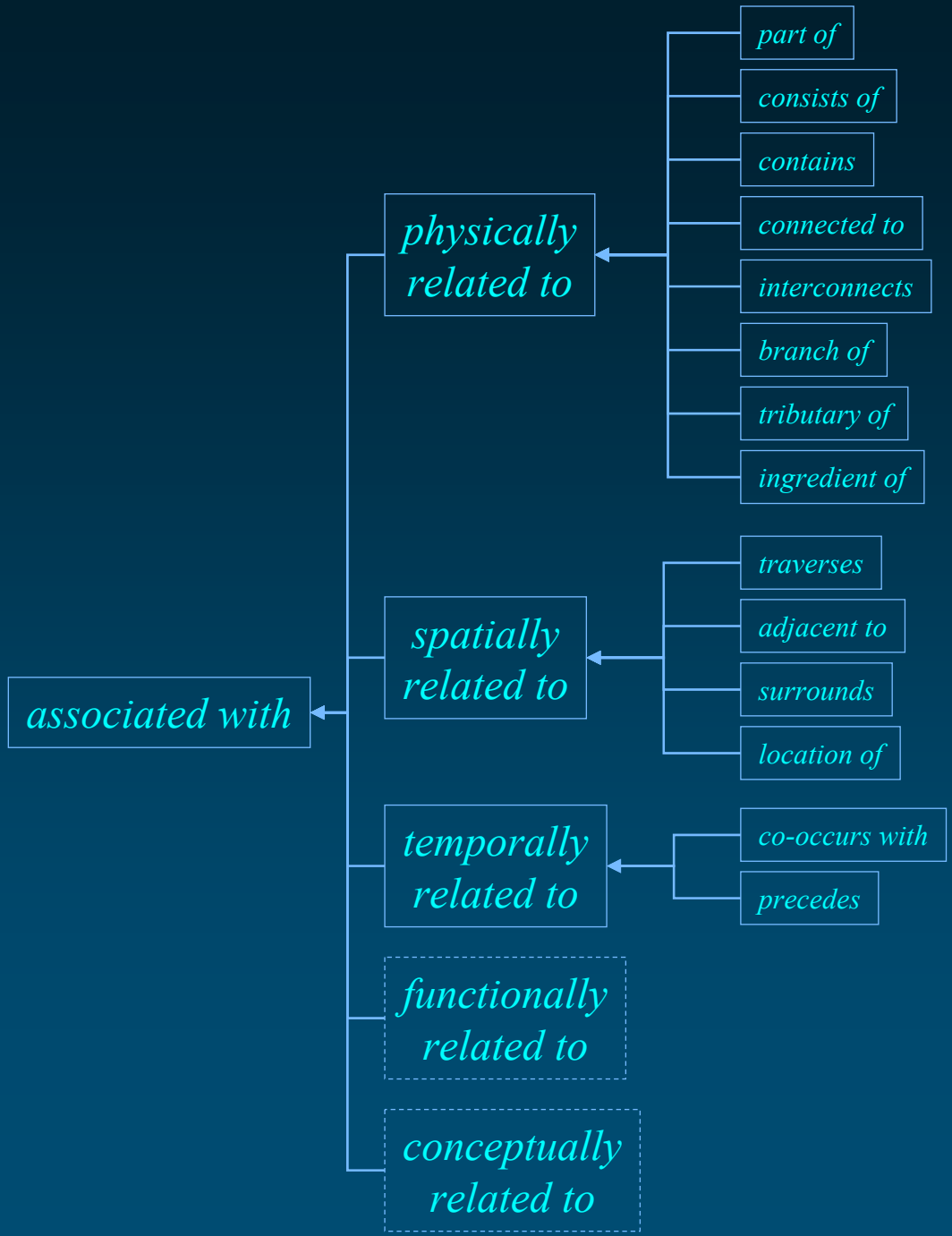


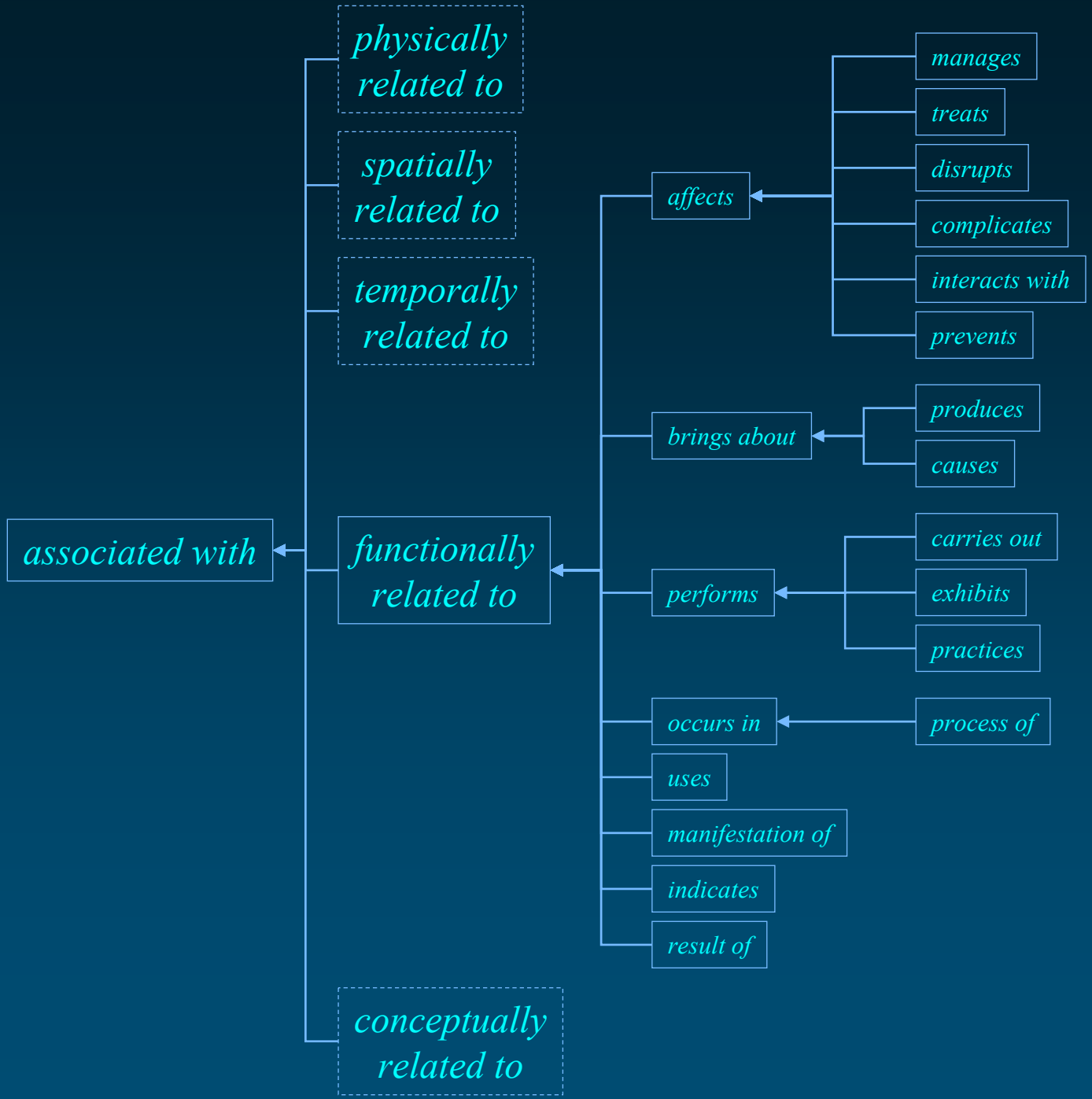
Associative (non-isa) relations

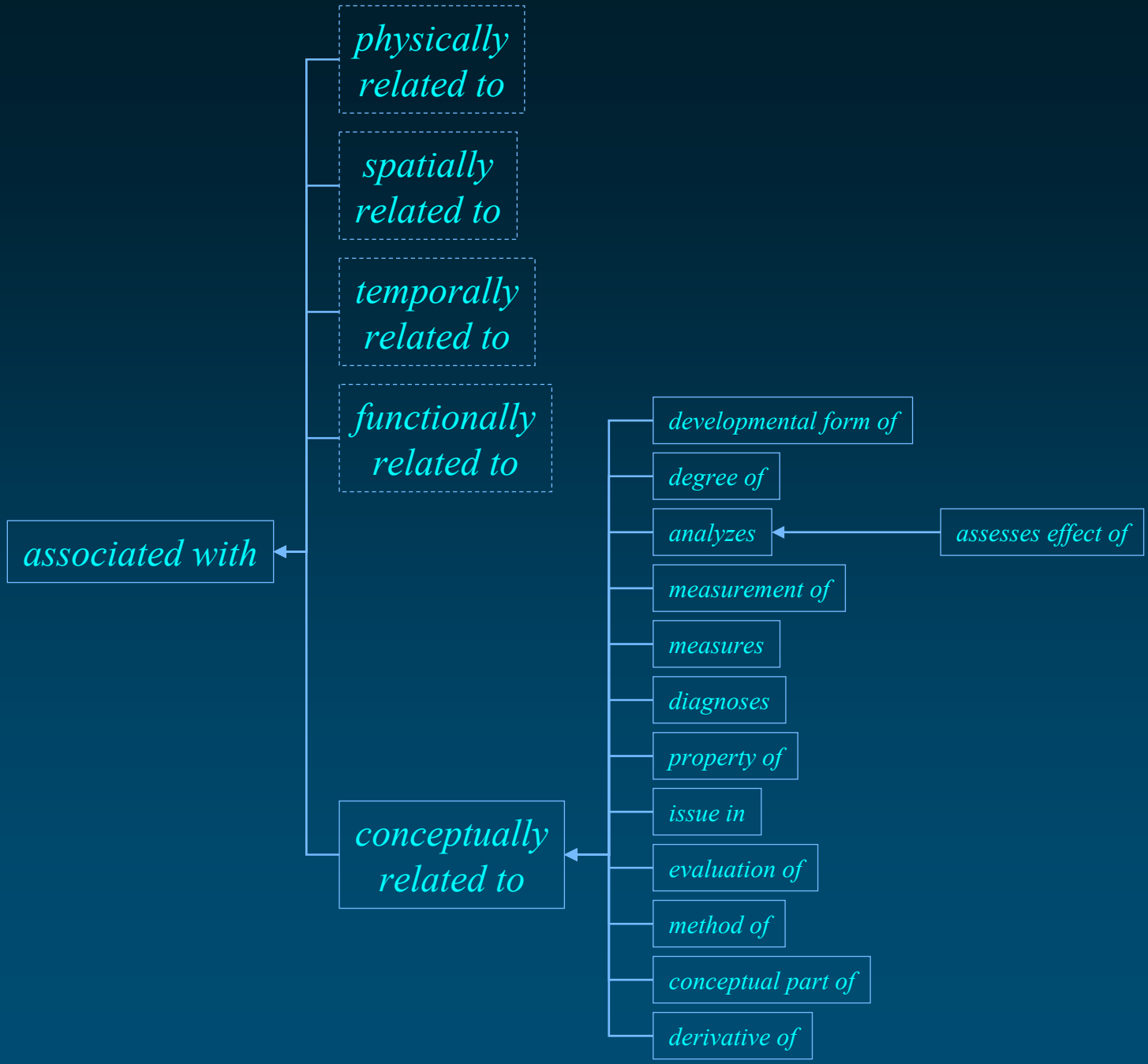


Relationship hierarchy




associated with
physically related to
part of
consists of
contains
connected to
interconnects
branch of
tributary of
ingredient of
spatially related to
traverses
adjacent to
surrounds
location of
temporally related to
co-occurs with
precedes
functionally related to
conceptually related to


associated with
physically related to
spatially related to
temporally related to
functionally related to
conceptually related to
affects
manages
treats
disrupts
complicates
interacts with
prevents
brings about
produces
causes
performs
carries out
exhibits
practices
occurs in
process of
uses
manifestation of
indicates
result of

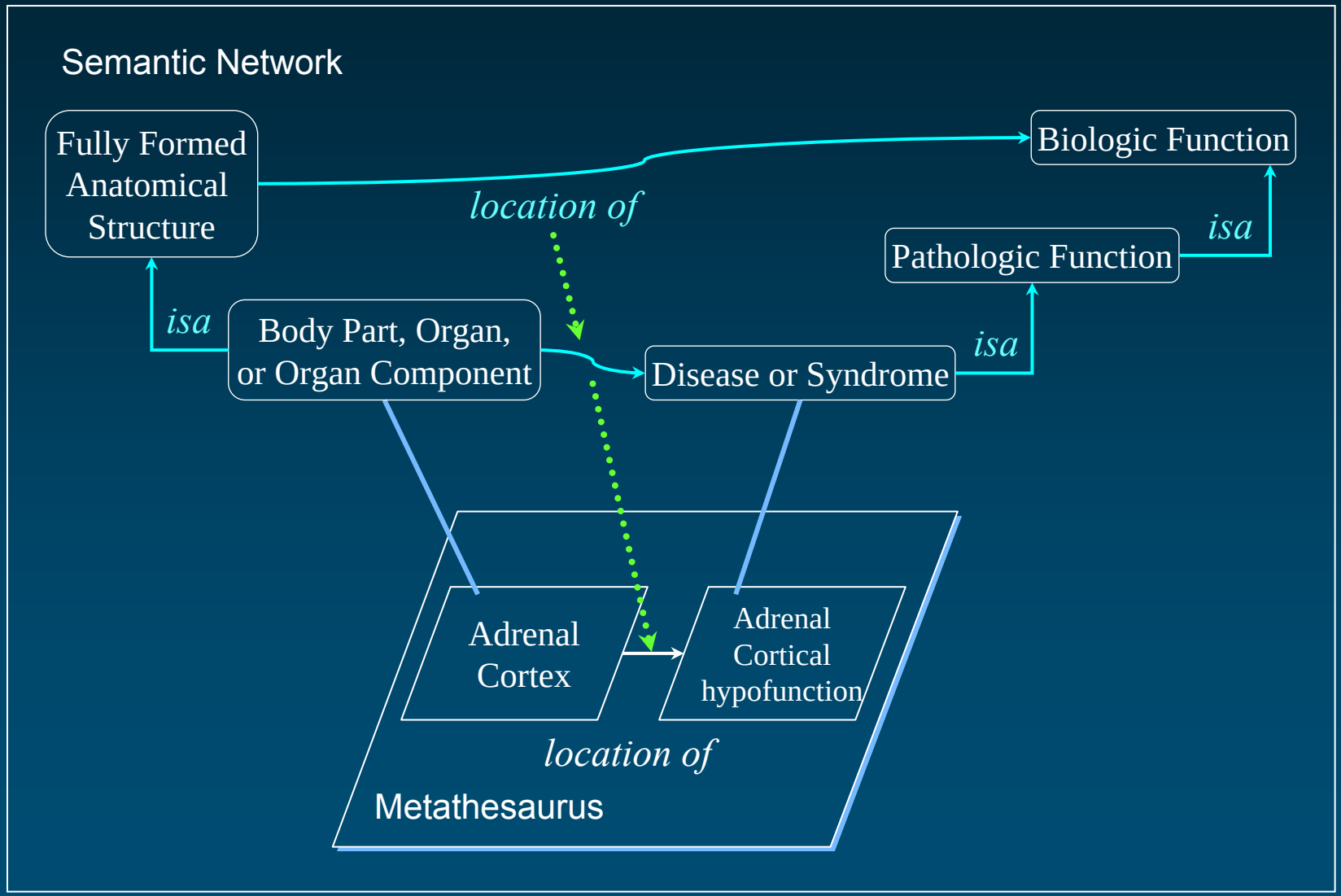

physically related to
spatially related to
temporally related to
functionally related to
associated with
conceptually related to
developmental form of
degree of
analyzes
assesses effect of
measurement of
measures
diagnoses
property of
issue in
evaluation of
method of
conceptual part of
derivative of

Why a semantic network?

- Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently of their position in a hierarchy*

- A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level

Relationships can inherit semantics



Aligning relationships

Relationships in the UMLS

(2005AC)

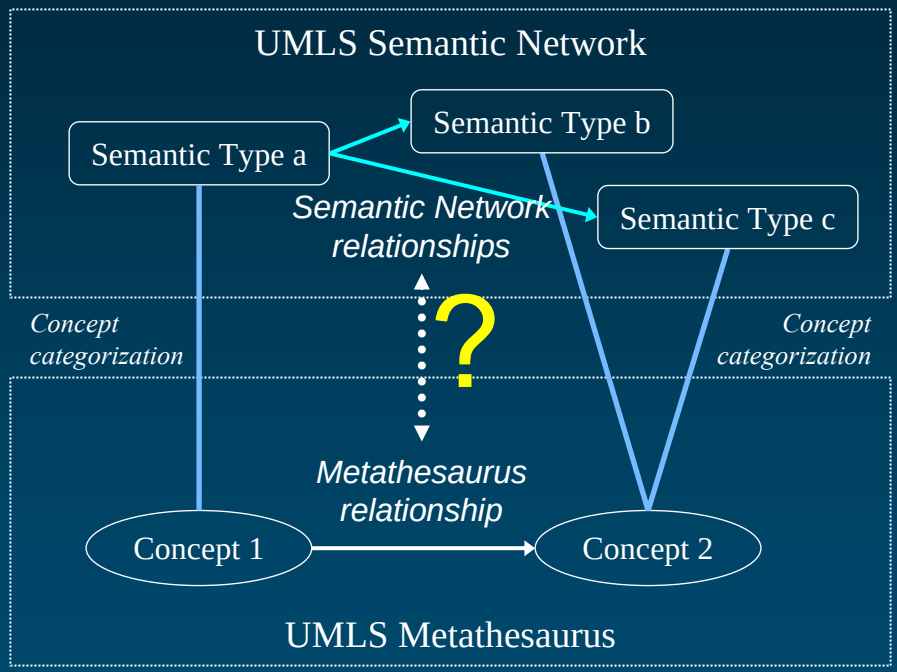
- Metathesaurus
  - 139 relationships
    - Thesaural
    - Specified
  - No definition
  - No organization

    - *isa*
    - *causes*
    - *due to*
    - *manifestation of*
    - *icd asterisk*
    - *associated finding of*
    - *may treat*

- Semantic Network
  - 54 relationships
  - Textual definition
  - Organized in 5 hierarchies

    - *isa*
    - *associated with*
      - *functionally related to*
      - *physically related to*
      - *spatially related to*
      - *temporally related to*
      - *conceptually related to*

Concepts/Types vs. Relationships



Objectives

- Elicit the meaning of Metathesaurus relationships
- Establish
  - A correspondence between
    - Each of the 139 relationships in the Metathesaurus
    - And relationships in the Semantic Network
  - The nature of the correspondence
    - Equivalence, specialization, other
- Enrich SN relationships with relationships from the Metathesaurus
- Develop an ontology of relationships

Multiple complementary approaches

- Metathesaurus-centric
  - Manual elicitation
  - Abstraction at the level of high-level concepts
  - Abstraction at the level of Semantic Types
- Semantic Network-centric

1a

Manual elicitation

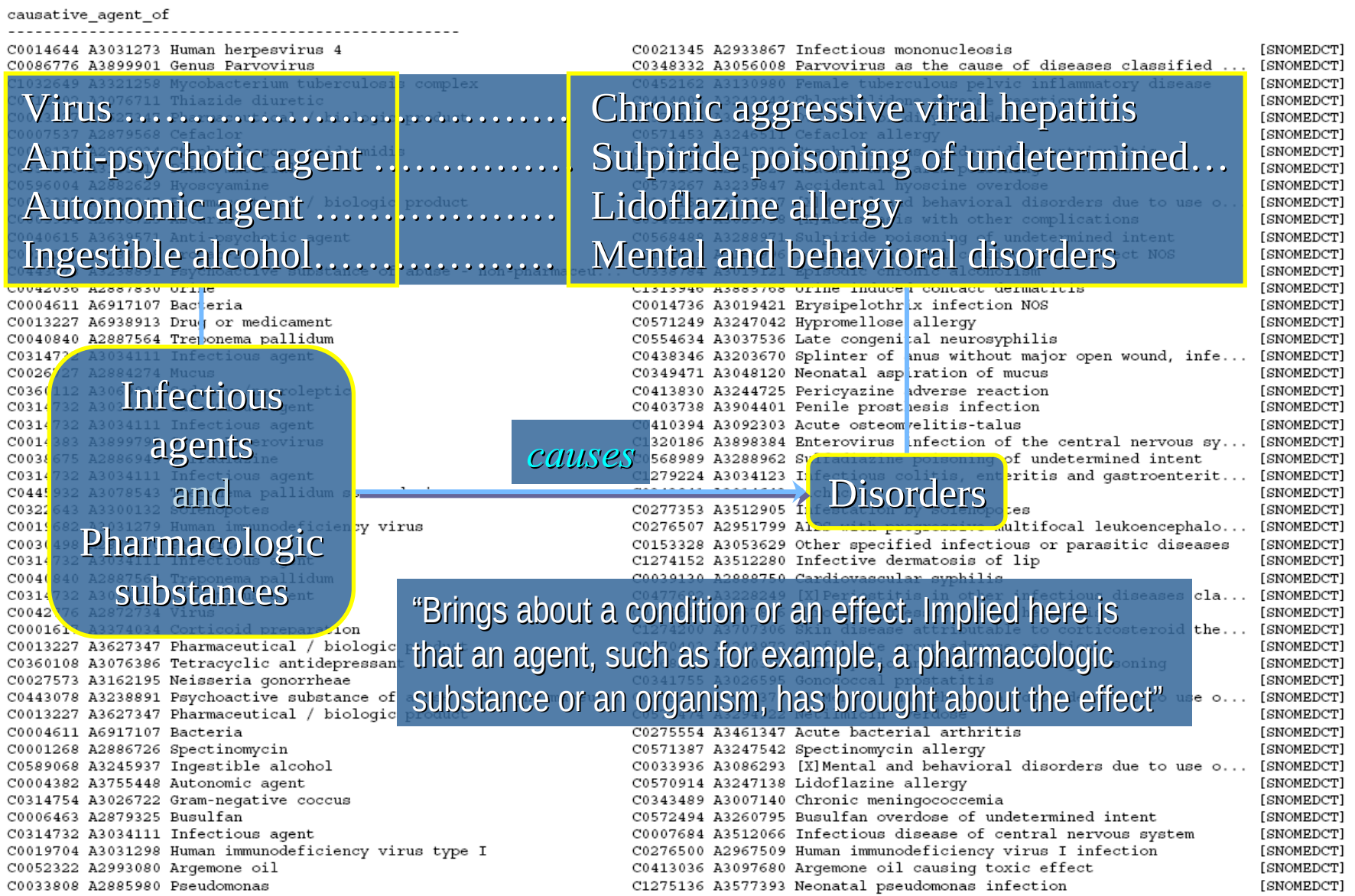
- Create random samples of a maximum of 50 relations per Metathesaurus relationship
- Review manually

causative agent of

causative_agent_of

Domain		
C0014644	A3031273	Human herpesvirus 4
C0086776	A3899901	Genus Parvovirus
C1032649	A3321258	Mycobacterium tuberculosis complex
C0012802	A3076711	Thiazide diuretic
C0013227	A3627347	Pharmaceutical / biologic product
C0007537	A2879568	Cefaclor
C0038174	A2886834	Staphylococcus epidermidis
C0557858	A3185117	Plant material
C0596004	A2882629	Hyoscyamine
C0013227	A3627347	Pharmaceutical / biologic product
C0003954	A2878725	Ascaris
C0040615	A3639571	Anti-psychotic agent
C0022237	A3188798	Propan-2-ol
C0443078	A3238891	Psychoactive substance of abuse - non-pharmaceu..
C0042036	A2887830	Urine
C0004611	A6917107	Bacteria
C0013227	A6938913	Drug or medicament
C0040840	A2887564	Treponema pallidum
C0314732	A3034111	Infectious agent
C0026727	A2884274	Mucus
C0360112	A3069840	Sedative/neuroleptic
C0314732	A3034111	Infectious agent
C0314732	A3034111	Infectious agent
C0014383	A3899795	Genus Enterovirus
C0038675	A2886949	Sulfadiazine
C0314732	A3034111	Infectious agent
C0445932	A3078543	Treponema pallidum ss. endemicum
C0322643	A3300132	Solenopotes
C0019682	A3031279	Human immunodeficiency virus
C0030498	A2872800	Parasite
C0314732	A3034111	Infectious agent
C0040840	A2887564	Treponema pallidum
C0314732	A3034111	Infectious agent
C0042776	A2872734	Virus
C0001617	A3374034	Corticoid preparation
C0013227	A3627347	Pharmaceutical / biologic product
C0360108	A3076386	Tetracyclic antidepressant
C0027573	A3162195	Neisseria gonorrheae
C0443078	A3238891	Psychoactive substance of abuse - non-pharmaceu..
C0013227	A3627347	Pharmaceutical / biologic product
C0004611	A6917107	Bacteria
C0001268	A2886726	Spectinomycin
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

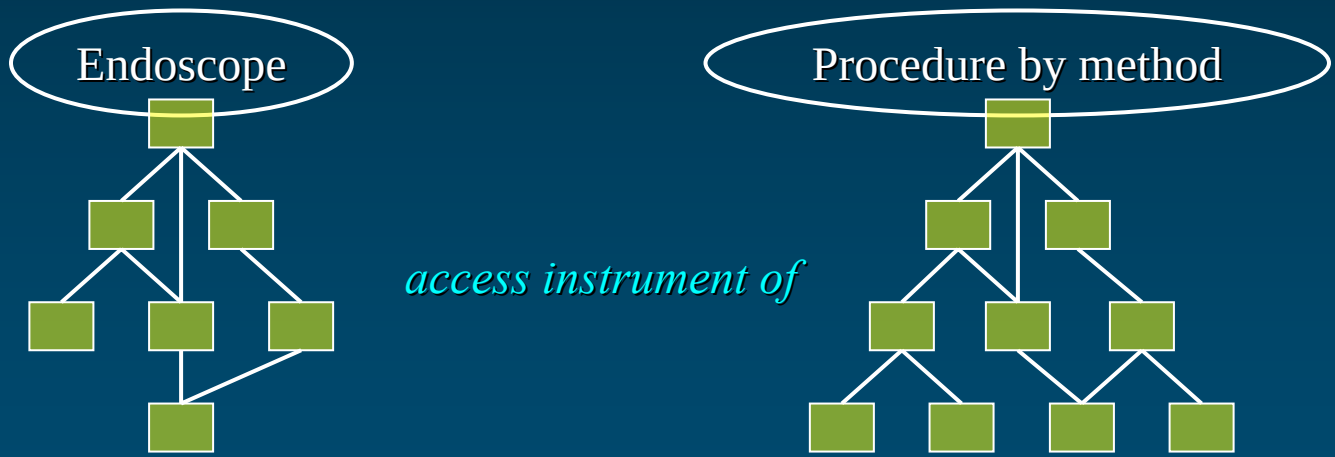
Range			
C0021345	A2933867	Infectious mononucleosis	[SNOMEDCT]
C0348332	A3056008	Parvovirus as the cause of diseases classified ...	[SNOMEDCT]
C0452162	A3130980	Female tuberculous pelvic inflammatory disease	[SNOMEDCT]
C0414005	A3243943	Chlorthalidone adverse reaction	[SNOMEDCT]
C0572072	A3273526	Overdose of dihydrocodeine	[SNOMEDCT]
C0571453	A3246511	Cefaclor allergy	[SNOMEDCT]
C1299601	A3718212	Staphylococcus epidermidis ventriculitis	[SNOMEDCT]
C0275188	A2954926	Asaemia axillaris poisoning	[SNOMEDCT]
C0573267	A3239847	Accidental hyoscine overdose	[SNOMEDCT]
C0349155	A3086327	[X]Mental and behavioral disorders due to use o...	[SNOMEDCT]
C0348284	A3085798	[X]Ascariasis with other complications	[SNOMEDCT]
C0568488	A3288971	Sulpiride poisoning of undetermined intent	[SNOMEDCT]
C0161681	A3148996	Isopropyl alcohol causing toxic effect NOS	[SNOMEDCT]
C0338784	A3019121	Episodic chronic alcoholism	[SNOMEDCT]
C1313946	A3883768	Urine induced contact dermatitis	[SNOMEDCT]
C0014736	A3019421	Erysipelothrix infection NOS	[SNOMEDCT]
C0571249	A3247042	Hypromellose allergy	[SNOMEDCT]
C0554634	A3037536	Late congenital neurosyphilis	[SNOMEDCT]
C0438346	A3203670	Splinter of anus without major open wound, infe...	[SNOMEDCT]
C0349471	A3048120	Neonatal aspiration of mucus	[SNOMEDCT]
C0413830	A3244725	Pericyazine adverse reaction	[SNOMEDCT]
C0403738	A3904401	Penile prosthesis infection	[SNOMEDCT]
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]
C0477692	A3228249	[X]Periostitis in other infectious diseases cla...	[SNOMEDCT]
C0276627	A2957806	Chronic aggressive viral hepatitis	[SNOMEDCT]
C1274200	A3707306	Skin disease attributable to corticosteroid the...	[SNOMEDCT]
C0570403	A3243972	Clofibrate group adverse reaction	[SNOMEDCT]
C0568289	A3250399	Tetracyclic antidepressant drug poisoning	[SNOMEDCT]
C0341755	A3026595	Gonococcal prostatitis	[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]



$$causative\ agent\ of_{SNOMED\ CT} \equiv causes_{SN}$$

1b Abstraction at the level of high-level concepts

- Compute the lowest common ancestor(s) for concepts in the domain and range of the relationship, respectively

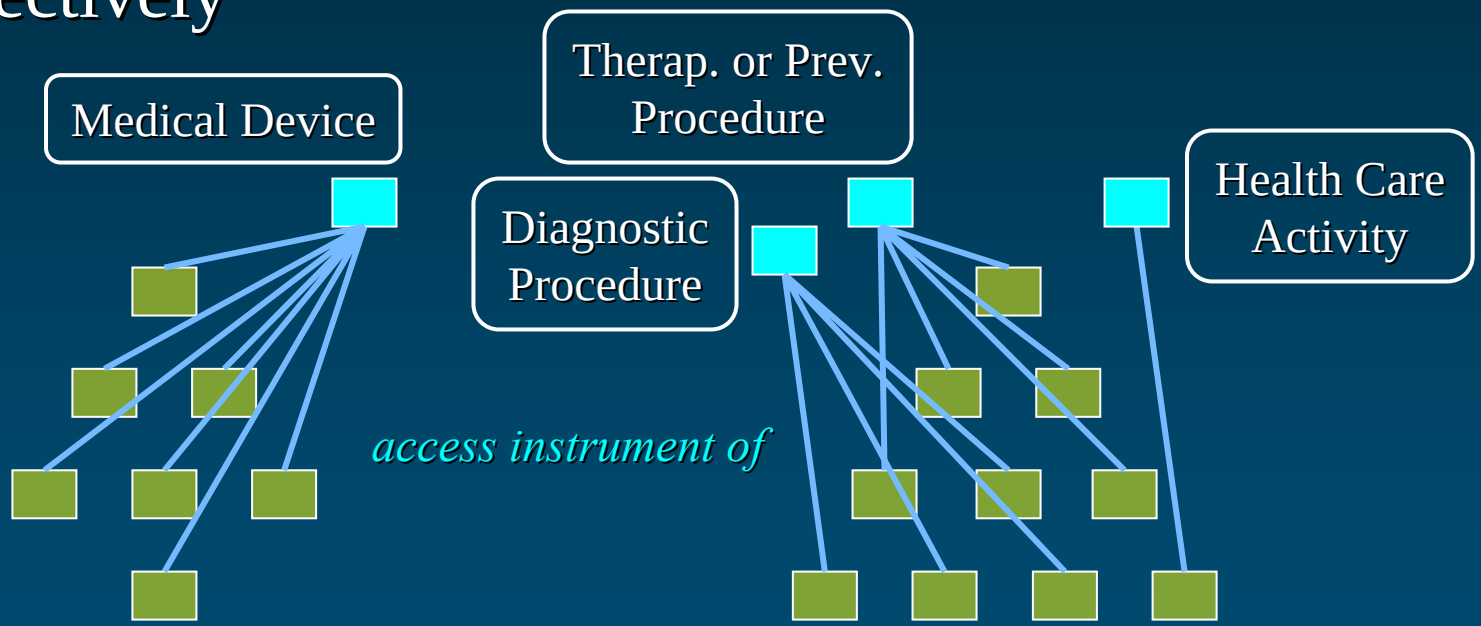


access instrument of (Endoscope, Procedure)

1c

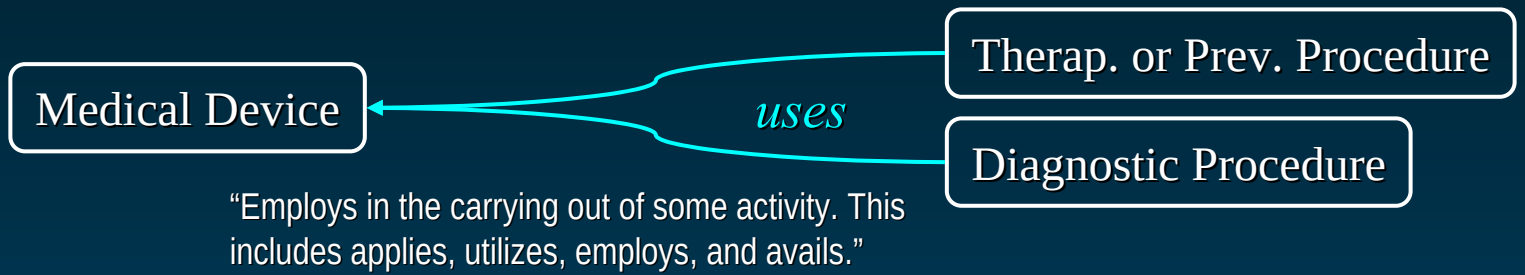
Abstraction at the level of STs

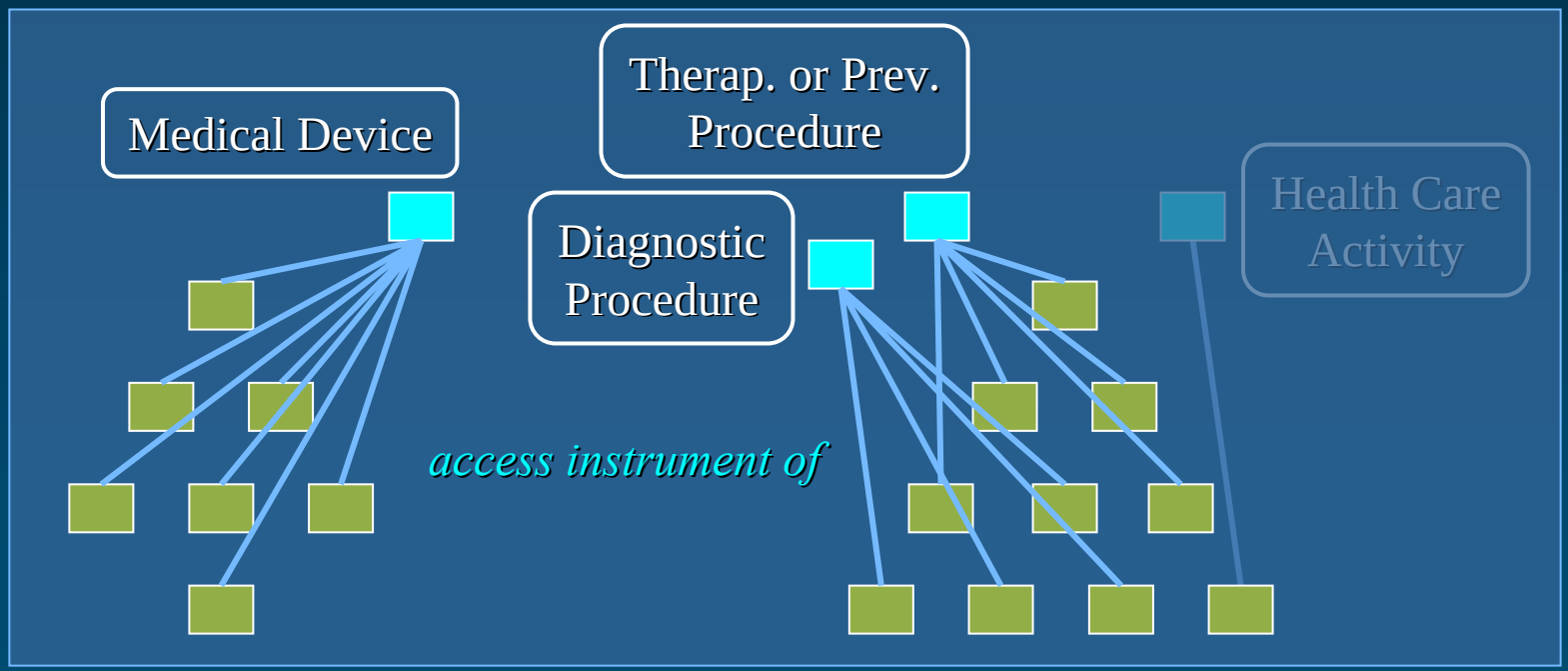
- Compute the distribution of the STs for concepts in the domain and range of the relationship, respectively



access instrument of (Medical Device, Diagnostic/Therap. or Prev. Procedure)

Identifying candidates in the SN

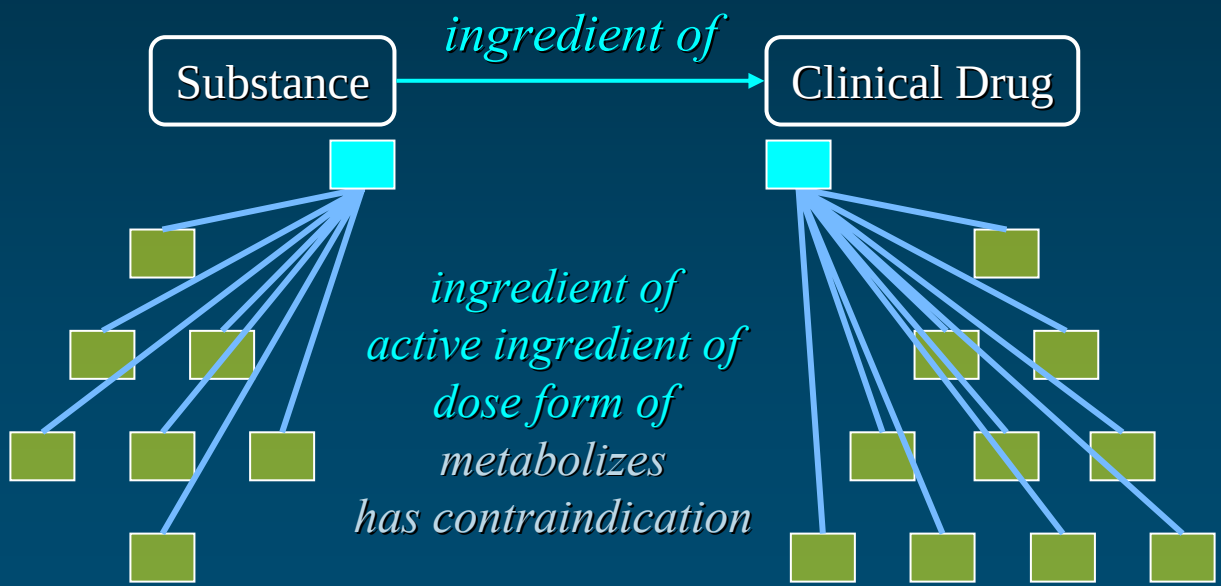




$$access\ instrument\ of_{SNOMED\ CT} \equiv inv(uses_{SN})$$

2 Semantic Network-centric approach

- For a given relationship between two STs in the SN, list all the relationships between concepts categorized by these STs



$\{ingredient\ of_{*},\ active\ ingredient\ of_{SNOMEDCT},\ dose\ form\ of_{*},\ \} \equiv ingredient\ of_{SN}$

Results Materials

(2005AC)

- 139 specified relationships (RELA) in the Metathesaurus
- Major contributors
  - SNOMED CT (62)
  - LOINC (15)
  - NDF-RT (15)
- Most relationships are specific to one vocabulary
  - 116 specific to one vocabulary
  - 23 found in at least 2 vocabularies

Results Alignment

- 80 relationships (58%) could be aligned
  - Identical
    - *affects, process of, ingredient of*
  - Roughly equivalent to
    - *focus of*$_{SNOMEDCT}$ ≡ *issue in*$_{SN}$
  - More specific than
    - *metabolic site of*$_{NDFRT}$ < *functionally related to*$_{SN}$
- Non aligned (non-semantic relationships)
  - Lexical relatioships (e.g., *british form of*$_{*}$, *suffix of*$_{LOINC}$)
  - Mapping relationships (e.g., *see from*$_{CRISP}$, *mapped to*$_{*}$)
  - Vocabulary management (e.g., *replaces*$_{SNOMEDCT}$)

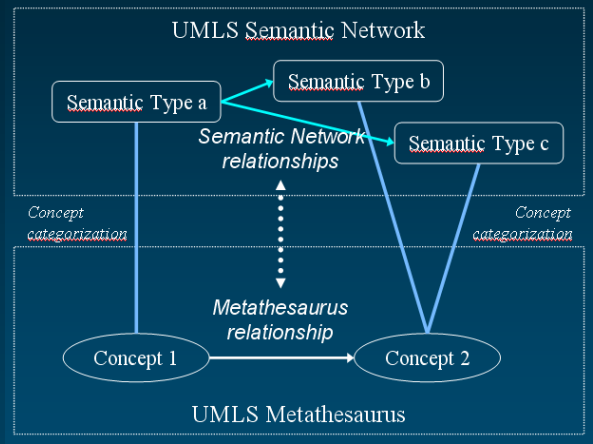
Some issues

- Manual elicitation
  - Manual review by experts required
- Abstraction at the level of high-level concepts
  - Lowest common ancestor = root of the terminology (heterogeneous set of concepts)
- Abstraction at the level of Semantic Types
  - Multiple STs for a set of concepts
  - Multiple relationships for a given pair of STs
- Semantic Network-centric approach
  - Multiple relationships in the Metathesaurus between the concepts corresponding to a given pair of STs

Towards an ontology of relationships

Motivation

- Consistency checking
  - Relations asserted between STs in the SN should provide constraints for relations between Metathesaurus concepts categorized by these STs

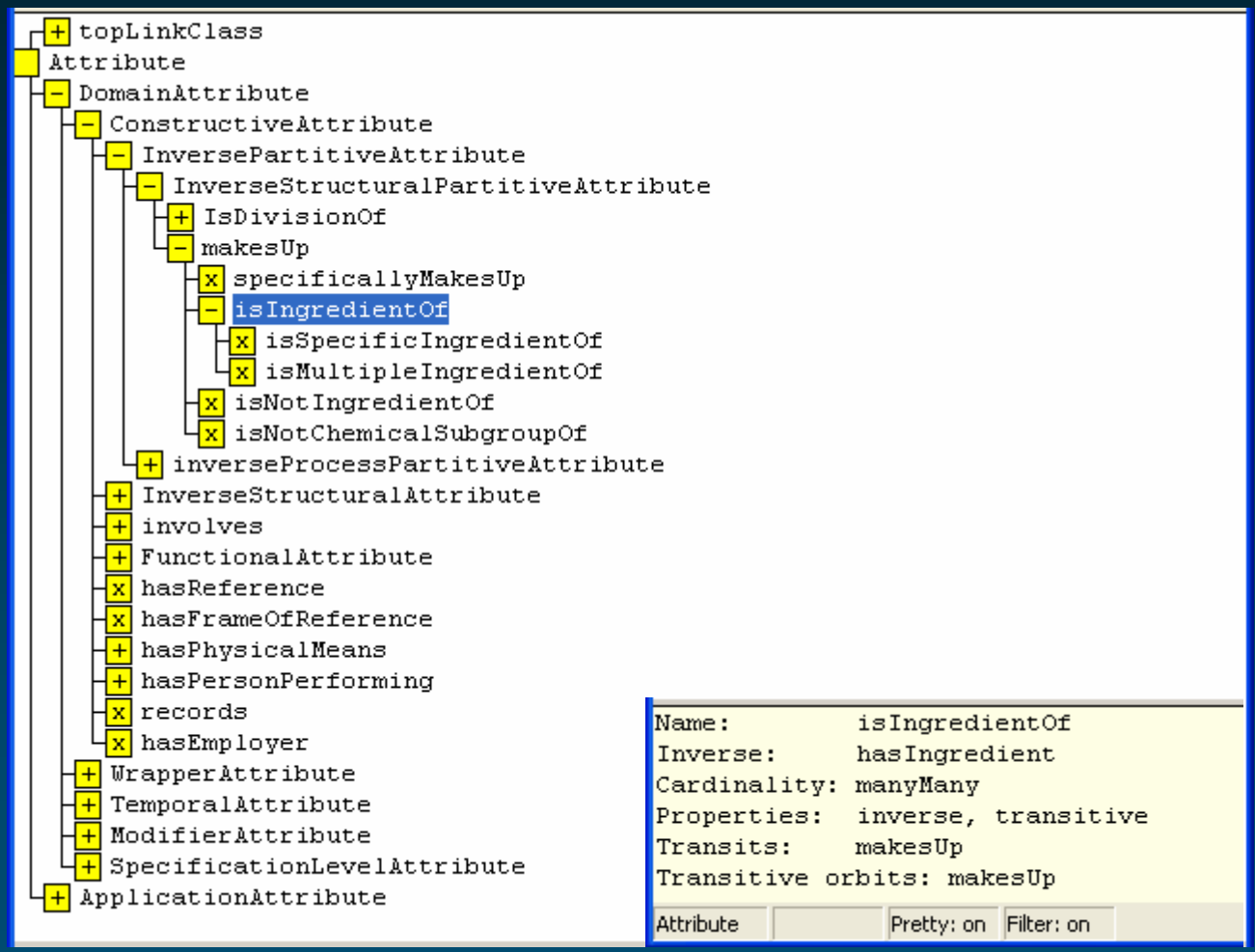


- Effective subsumption reasoning
  - Requires explicit equivalence between identical relationships (e.g., inv(*due to*) = *cause of* ≡ *causes*)
  - Requires explicit subProperty relations

Existing resources

- Semantic Network relationships
  - 54 relationships, not formally defined, organized into a shallow hierarchy
  - http://semanticnetwork.nlm.nih.gov/
- GALEN relationships
  - Over 500 relationships, semi-formally defined, organized into a hierarchy
  - http://www.opengalen.org/
- OBO relation ontology
  - 10 relationships, formally defined
  - http://obofoundry.org/ro/

GALEN relations *isIngredientOf*

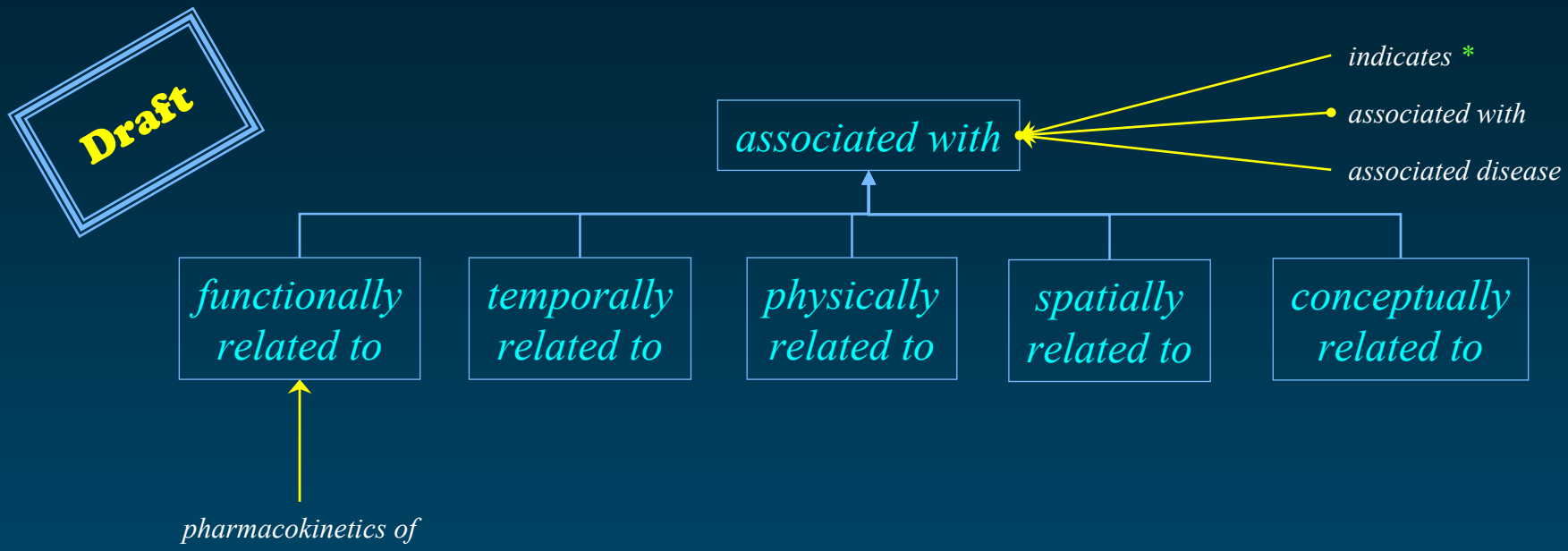


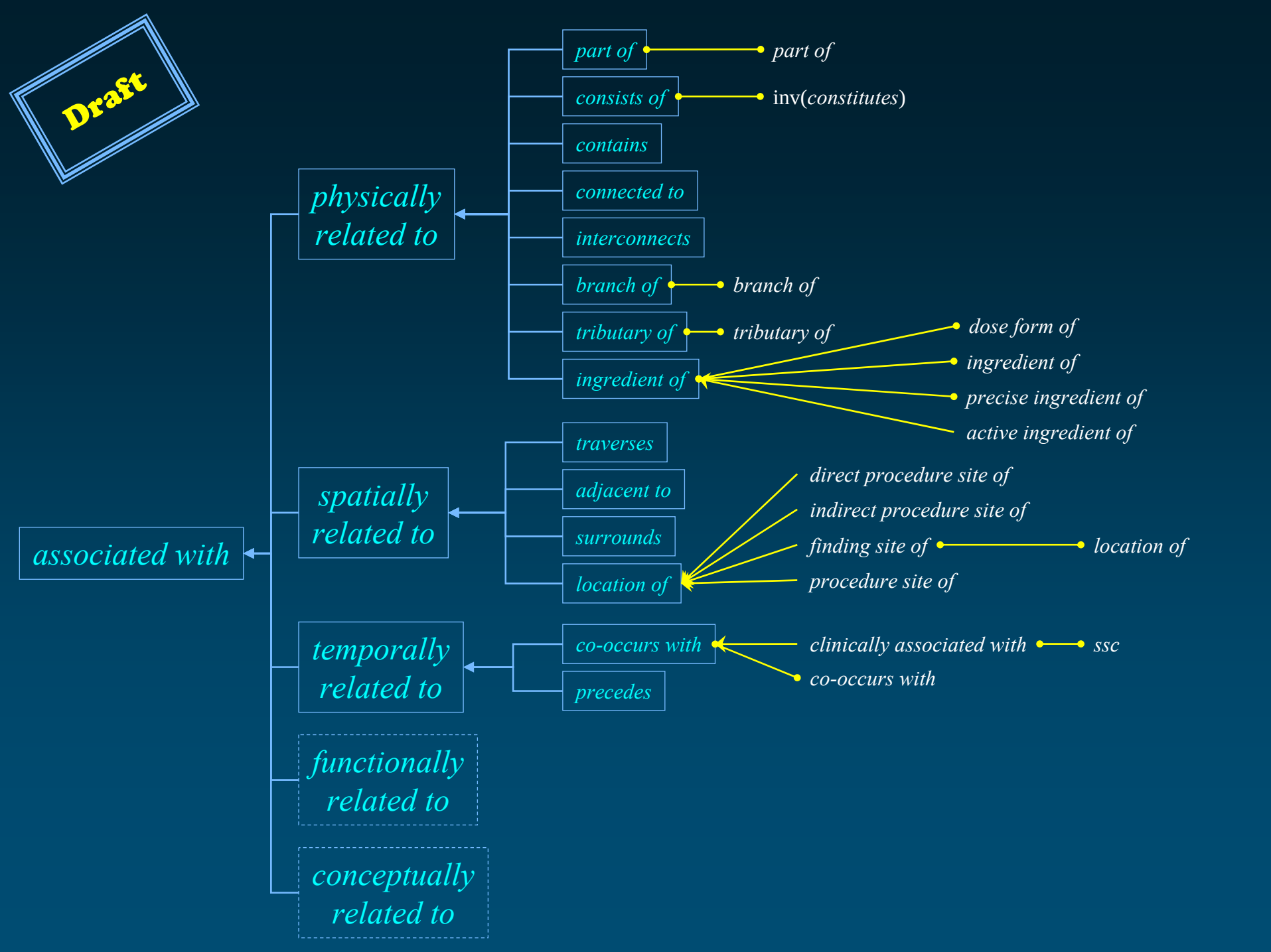
OBO relations

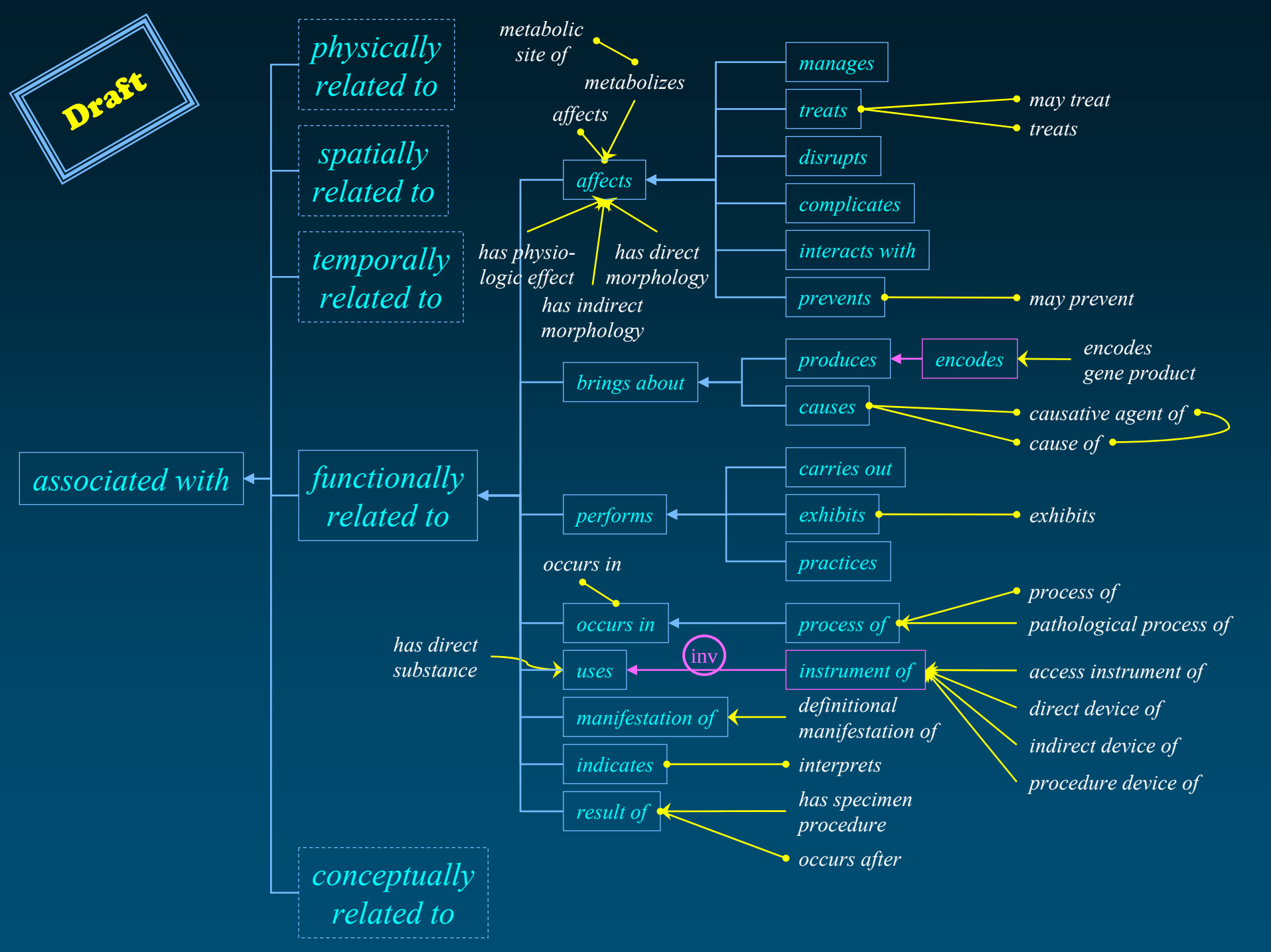
Summary Table

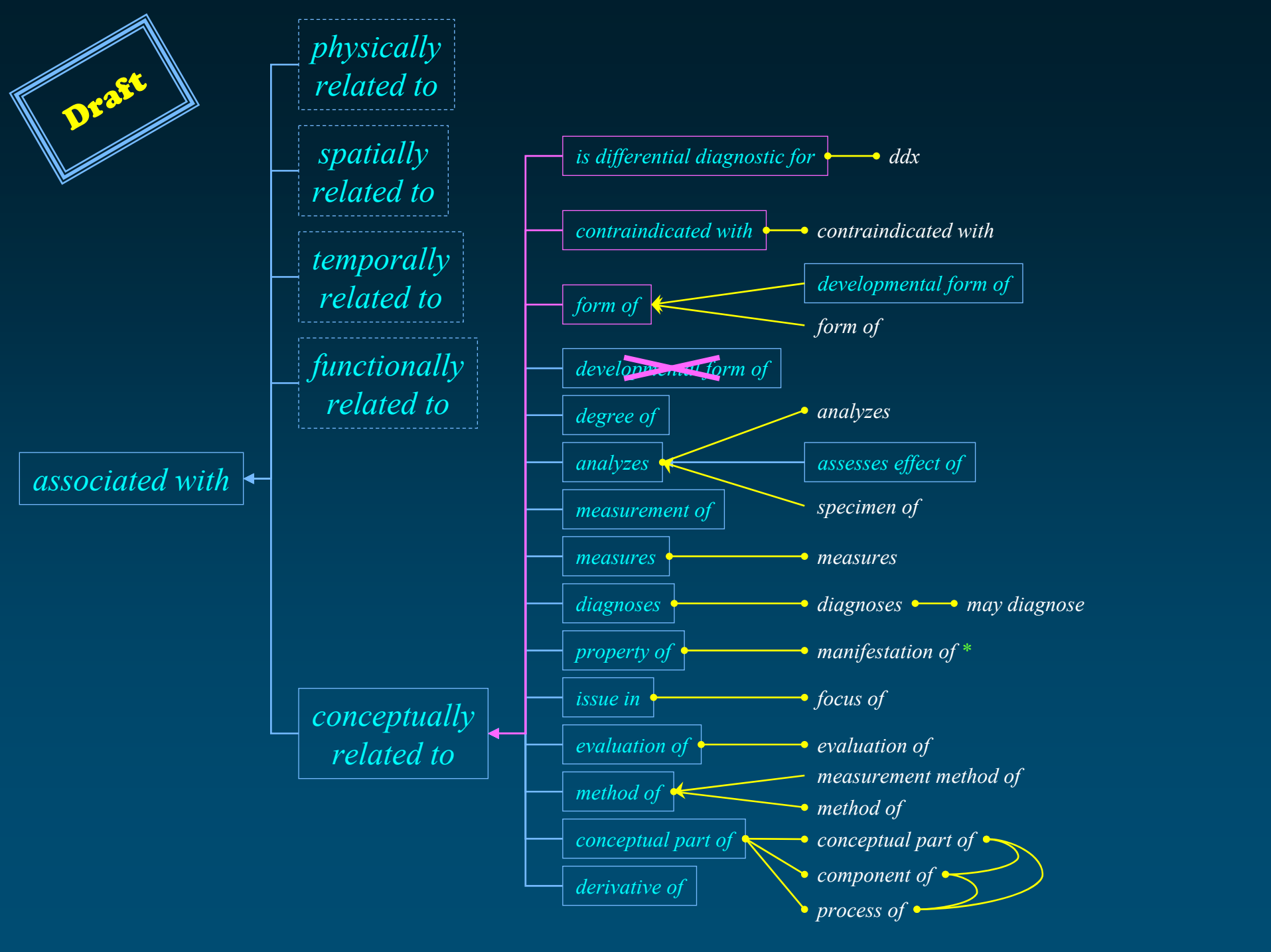
name	transitive	symmetric	reflexive	anti-symmetric	documentation
is_a	+		+	+	View detailed summary
part_of	+		+	+	View detailed summary
integral_part_of	+		+	+	View detailed summary
proper_part_of	+				View detailed summary
located_in	+		+		View detailed summary
contained_in					View detailed summary
adjacent_to					View detailed summary
transformation_of	+				View detailed summary
derives_from	+				View detailed summary
preceded_by	+				View detailed summary
has_participant					View detailed summary
has_agent					View detailed summary
instance_of					View detailed summary

Extending the SN relationship hierarchy




Draft
associated with
physically related to
part of
part of
consists of
inv(constitutes)
contains
connected to
interconnects
branch of
branch of
tributary of
tributary of
ingredient of
dose form of
ingredient of
precise ingredient of
active ingredient of
spatially related to
traverses
adjacent to
surrounds
location of
direct procedure site of
indirect procedure site of
finding site of
location of
procedure site of
temporally related to
co-occurs with
precedes
clinically associated with
ssc
co-occurs with
functionally related to
conceptually related to


Draft
associated with
physically related to
spatially related to
temporally related to
functionally related to
conceptually related to
metabolic site of
metabolizes
affects
affects
has physio-logic effect
has direct morphology
has indirect morphology
manages
treats
may treat
treats
disrupts
complicates
interacts with
prevents
may prevent
brings about
produces
encodes
encodes gene product
causes
causative agent of
cause of
performs
carries out
exhibits
exhibits
practices
occurs in
occurs in
process of
pathological process of
has direct substance
uses
inv
instrument of
access instrument of
direct device of
indirect device of
procedure device of
manifestation of
definitional manifestation of
indicates
interprets
result of
has specimen procedure
occurs after


Draft
associated with
physically related to
spatially related to
temporally related to
functionally related to
conceptually related to
is differential diagnostic for
ddx
contraindicated with
contraindicated with
form of
developmental form of
form of
degree of
analyzes
assesses effect of
specimen of
measurement of
measures
measures
diagnoses
diagnoses
may diagnose
property of
manifestation of *
issue in
focus of
evaluation of
evaluation of
measurement method of
method of
conceptual part of
conceptual part of
component of
process of
derivative of

References

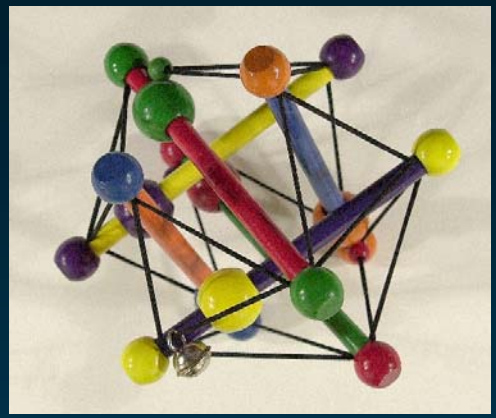
- Vizenor L, Bodenreider O, Peters L, McCray AT. *Enhancing biomedical ontologies through alignment of semantic relationships: Exploratory approaches.* AMIA Annu Symp Proc 2006:804-808.

- McCray AT, Bodenreider O. *A conceptual framework for the biomedical domain.* In: Green R, Bean CA, Myaeng SH, editors. The semantics of relationships: an interdisciplinary perspective. Dordrecht; Boston: Kluwer Academic Publishers; 2002. p. 181-198.

References

- McCray AT, Nelson SJ.
  The representation of meaning in the UMLS.
  Methods Inf Med 1995;34(1-2):193-201

- McCray AT.
  An upper level ontology for the biomedical domain.
  Comp Funct Genom 2003; 4:80-4.



Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

NATIONAL LIBRARY OF MEDICINE

Olivier Bodenreider

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