Kent Spackman

International Healthcare Terminology Standards Development Organisation (**IHTSDO**)

Stefan Schulz

Medical Informatics Research Group
University Medical Center
Freiburg, Germany

KR-MED 2008

Representing and Sharing Knowledge Using SNOMED

SNOMED CT:

Ontological, Terminological, and Knowledge Representation Aspects

May 31, 2008, Phoenix, Arizona, USA

Purpose of the Tutorial (I)

- 1: Theoretical underpinnings
  - Get aware of the enormous variety of biomedical vocabularies and their diverging architectural principles
  - Comprehend the current structure of SNOMED CT as a result of its evolution
  - Understand the nature of terminologies in contrast to classifications, nomenclatures, and ontologies
  - Understand the basic principles of formal ontology as a foundation of modern vocabulary development
  - Envisage the limitations of terminological / ontological knowledge representation related to the representation of domain knowledge in a broader sense

Purpose of the Tutorial (II)

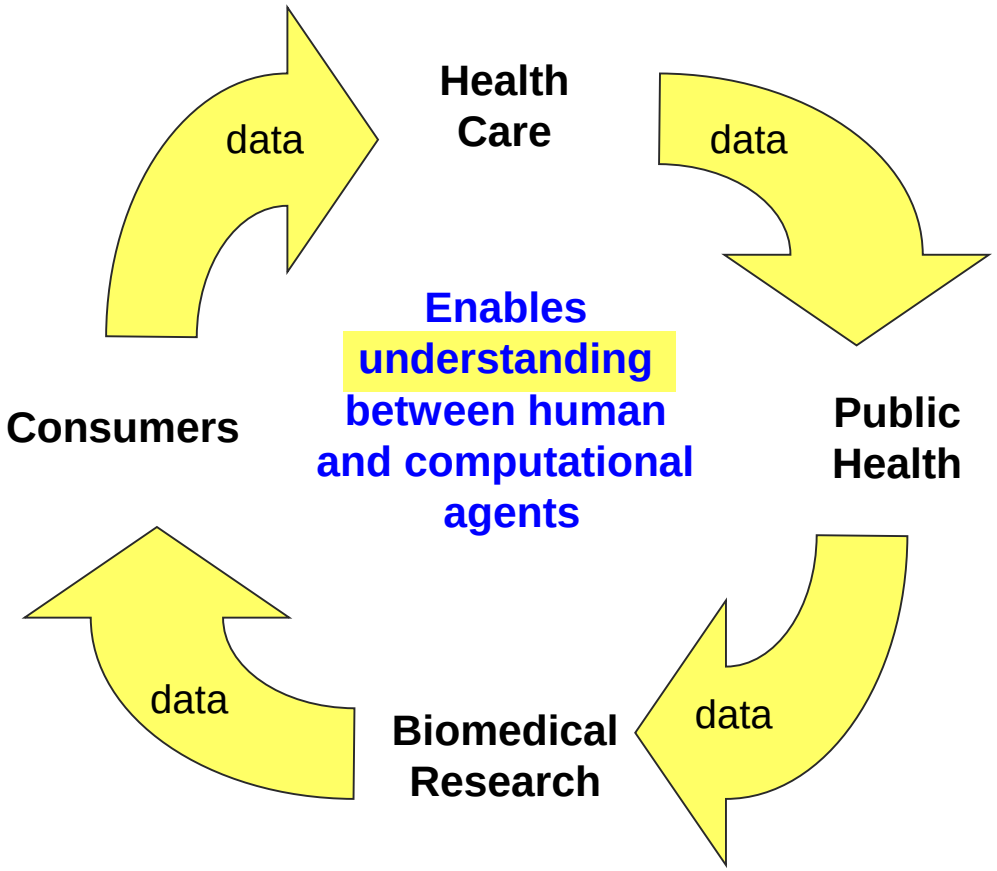
- 2: The practice of SNOMED CT
  - Understand the description logics used for representing SNOMED CT
  - Apply the description logics to special requirements: partonomies, complex procedures
  - Understand Pre-coordination and the SNOMED CT compositional syntax
  - Get insight into current redesign efforts (e.g. substance redesign)
  - Discuss the SNOMED CT context model and the terminology / information model interface

Preliminary remarks

- Attendees:
  - Heterogeneous
  - Experts: please challenge our viewpoints
  - Novices: please ask if you don’t understand a term
  - All: participate actively, feel free to interrupt us
- We have enough time for (moderated) discussions
  - 1st half: Presenter: Stefan, Moderator: Kent
  - 2nd half: Presenter: Kent, Moderator: Stefan
- Download tutorial slides from: http://www.kr-med.org/2008/tutorial/tutorial1.zip

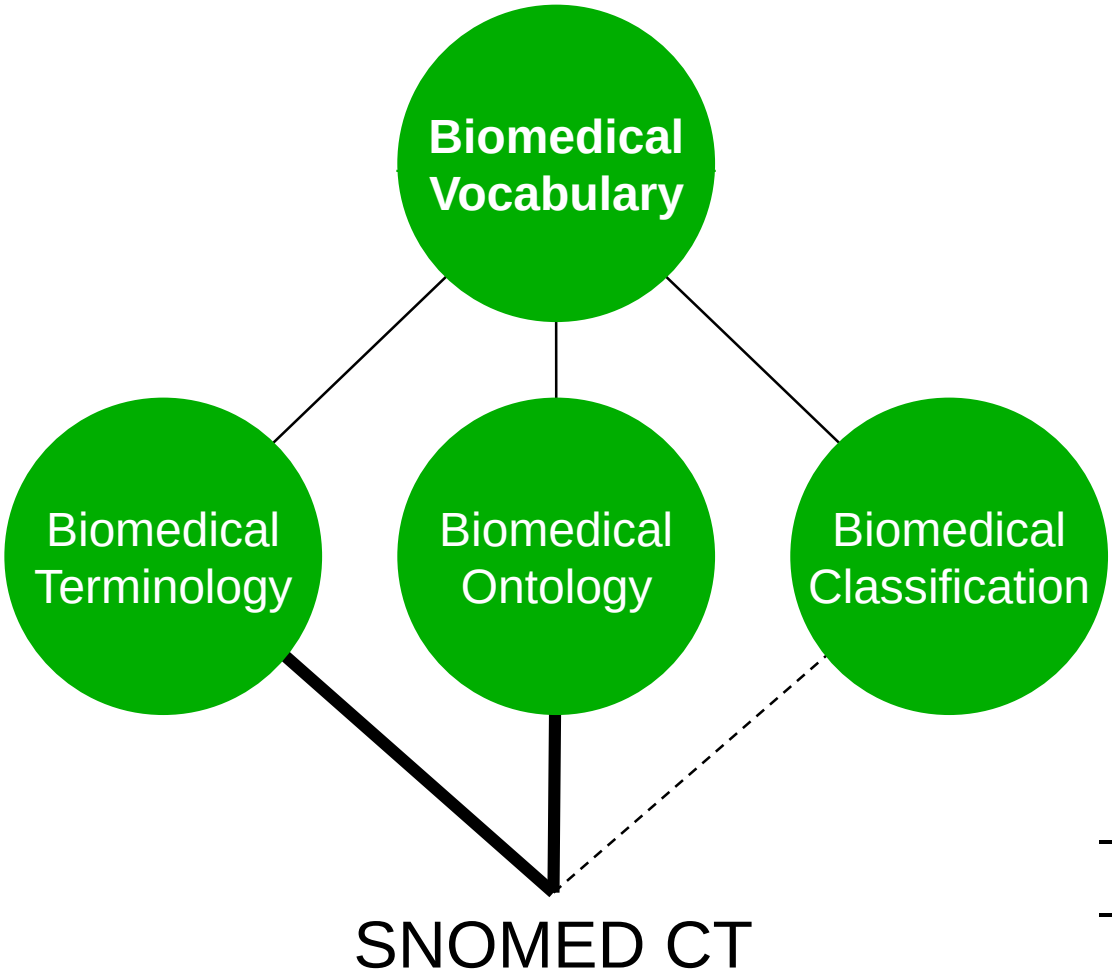
Biomedical Vocabularies

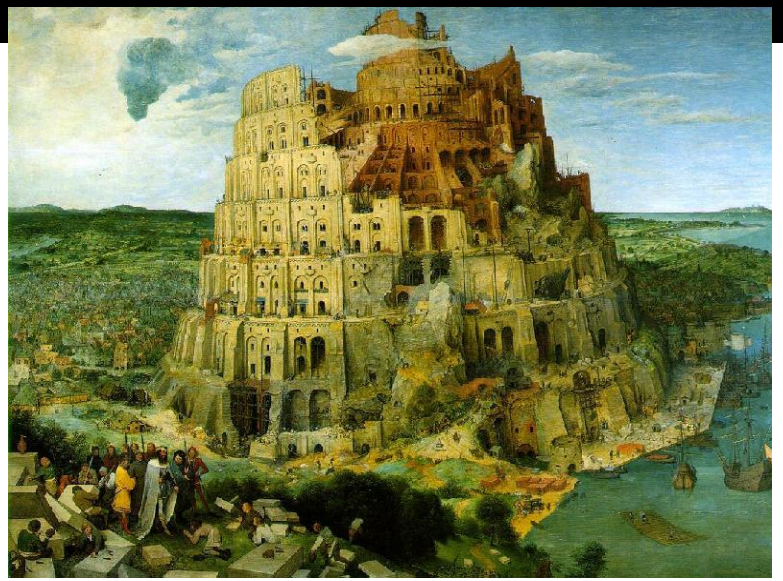
Understanding / Semantic Interoperability



Common languages: *Biomedical Vocabularies*

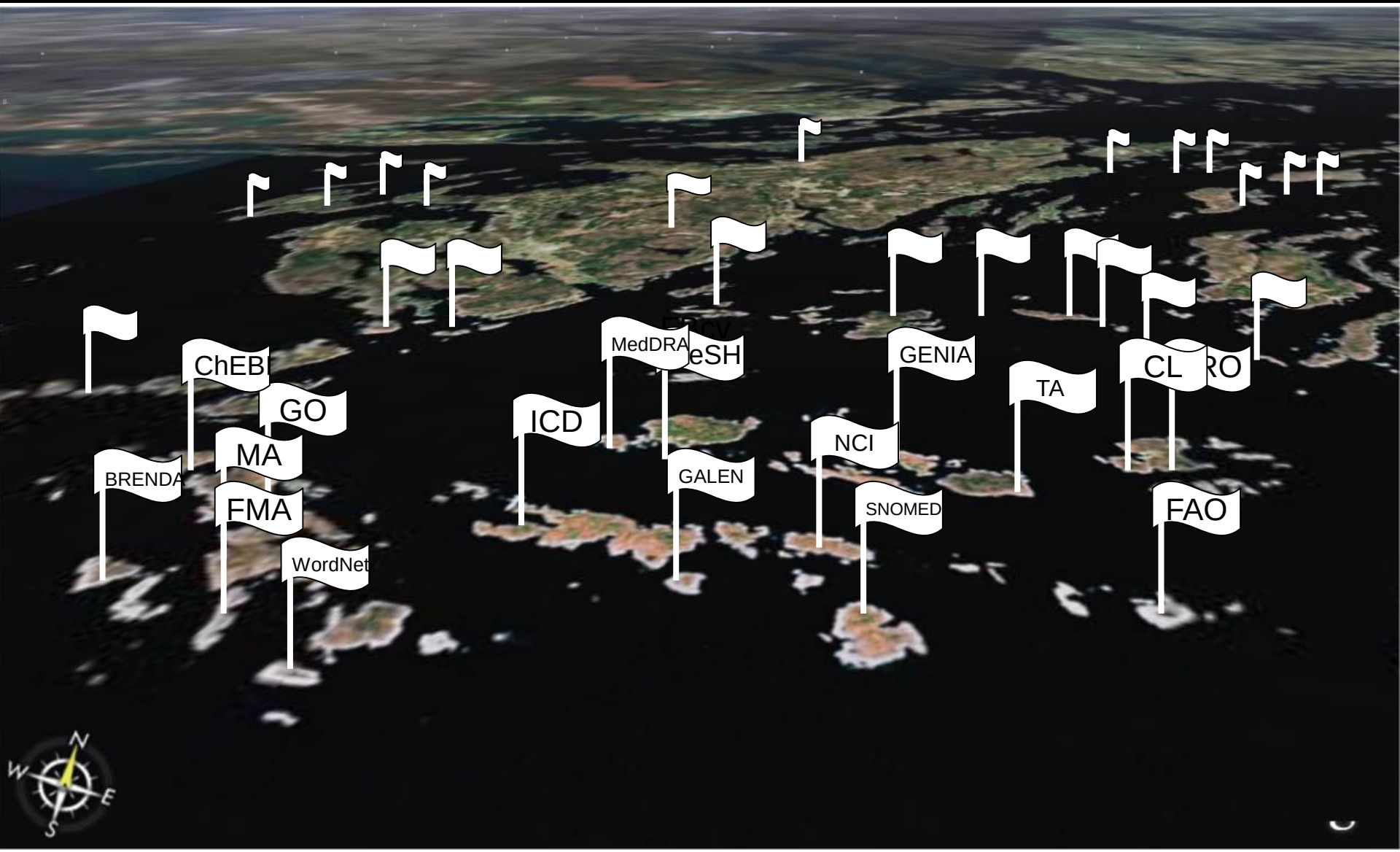
Meta – Terminological Issues



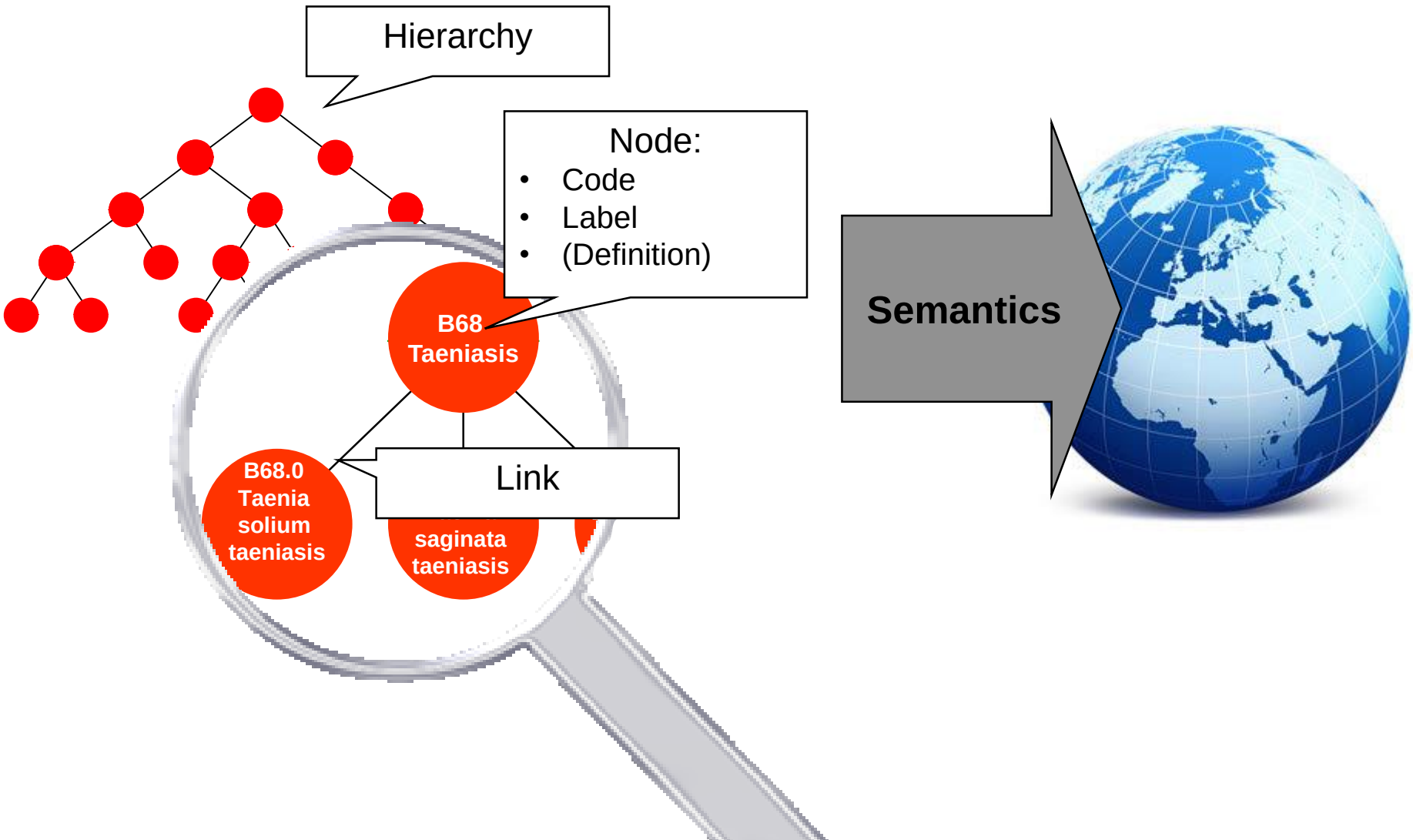


- Real systems are not ideal
- Real systems are often hybrids

A cruise through the archipelago of biomedical vocabularies



What biomedical vocabularies have in common




ICD
International Classification
of Diseases

CLASSIFICATION: ICD-10

International Statistical Classification of Diseases and Related Health Problems
10th Revision
Version for 2007

Tabular List of inclusions and four-character subcategories

Chapter List

Chapter	Blocks	Title
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mec
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy, childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality

CLASSIFICATION: ICD-10

B67 **Echinococcosis**

Includes: hydatidosis

B67.0 **Echinococcus granulosus infection of liver**

B67.1 **Echinococcus granulosus infection of lung**

B67.2 **Echinococcus granulosus infection of bone**

B67.3 **Echinococcus granulosus infection, other and multiple sites**

B67.4 **Echinococcus granulosus infection, unspecified**

B67.5 **Echinococcus multilocularis infection of liver**

B67.6 **Echinococcus multilocularis infection, other and multiple sites**

B67.7 **Echinococcus multilocularis infection, unspecified**

B67.8 **Echinococcosis, unspecified, of liver**

B67.9 **Echinococcosis, other and unspecified**

Echinococcosis NOS

B68 **Taeniasis**

Excludes: cysticercosis (B69.-)

B68.0 **Taenia solium taeniasis**

Pork tapeworm (infection)

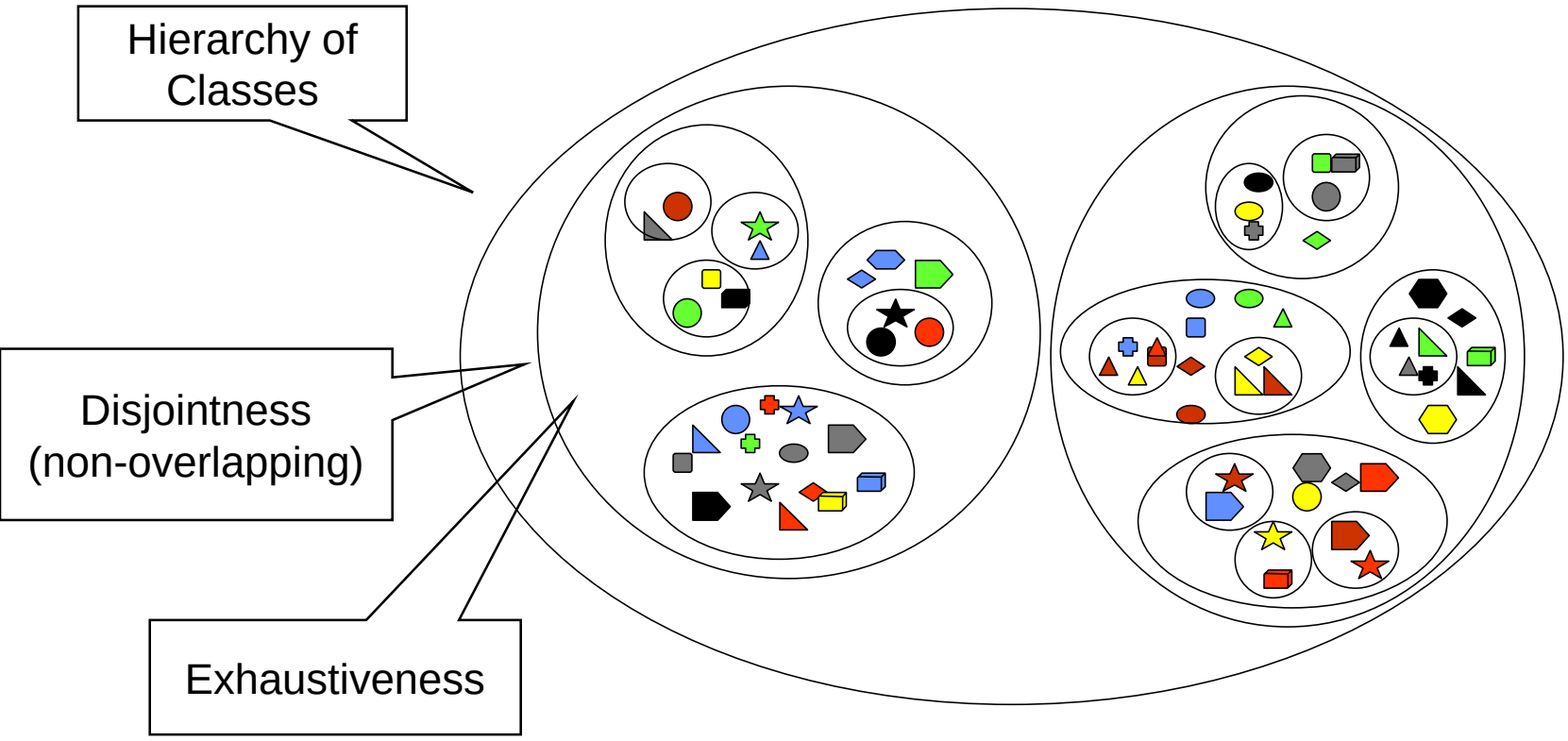
B68.1 **Taenia saginata taeniasis**

Beef tapeworm (infection)

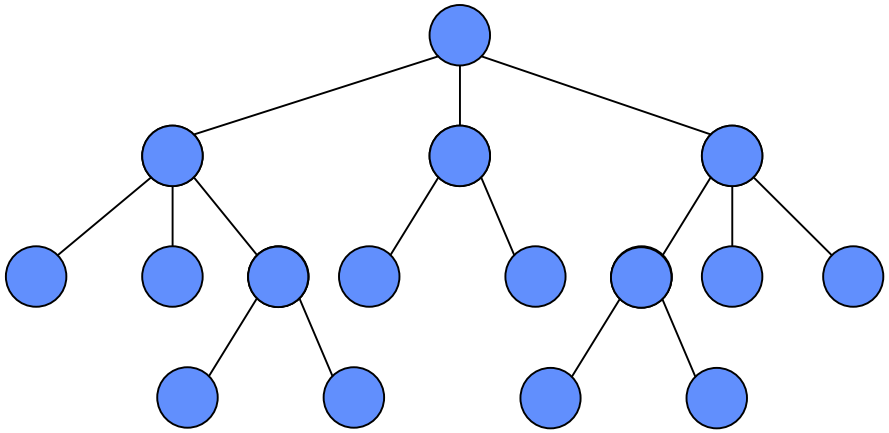
Infection due to adult tapeworm Taenia saginata

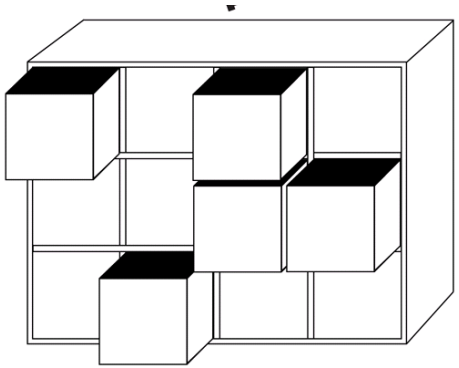
B68.9 **Taeniasis, unspecified**

CLASSIFICATION: ICD-10



CLASSIFICATION: ICD-10





CLASSIFICATION: ICD-10

- Nodes represent:
  - Mutually disjoint classes of particular disease entities
  - Often Idiosyncratic classification criteria
    "I83 Varicose veins of lower extremities –
    Excludes: complicating: pregnancy (O22.0), puerperium (O87.8)"
  - Classification criteria mix inherent properties of entities with epistemic information
    A15.1 Tuberculosis of lung, confirmed by culture only
  - Labels: explanatory
  - Terms: quasi-synonymous entry terms in different languages (alphabetical index)
- Links:
  - Connect classes with superclasses (taxonomy)
- Semantics:
  - Taxonomy: All particular entities that instantiate one class, also instantiate all superclasses


MeSH
Medical Subject Headings

THESAURUS: Medical Subject Headings

MeSH Tree Structures - 2006

Return to Entry Page

1. Anatomy [A]
2. Organisms [B]
   - Animals [B01] +
   - Algae [B02] +
   - Bacteria [B03] +
   - Viruses [B04] +
   - Fungi [B05] +
   - Plants [B06] +
   - Archaea [B07] +
   - Mesomycetozoea [B08] +
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Physical Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology and Food and Beverages [J]
11. Humanities [K]
12. Information Science [L]
13. Persons [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographic Locations [Z]

THESAURUS: Medical Subject Headings

Bacteria [B03]

- Atypical Bacterial Forms [B03.110] +
- Bacteria, Aerobic [B03.120]
- Bacteria, Anaerobic [B03.130]
- Bacteroidetes [B03.140] +
- Biofilms [B03.150]
- Blood-Borne Pathogens [B03.165]
- Chlorobi [B03.250] +
- Chloroflexi [B03.275] +
- Cyanobacteria [B03.280] +
- Endospore-Forming Bacteria [B03.300] +
- Fusobacteria [B03.370] +
- Gram-Negative Bacteria [B03.440] +
- Gram-Positive Bacteria [B03.510]
  - Actinobacteria [B03.510.024] +
  - Gram-Positive Cocci [B03.510.400] +
  - Gram-Positive Endospore-Forming Bacteria [B03.510.415] +
  - Gram-Positive Rods [B03.510.460] +
- Proteobacteria [B03.660] +
- Spirochaetales [B03.851] +
- Spores [B03.867] +
- Sulfur-Reducing Bacteria [B03.900] +

Hierarchical principle: broader term / narrower term

Return to Entry Page

THESAURUS: Medical Subject Headings

Bacteria [B03]

Gram-Positive Bacteria [B03.510]

Gram-Positive Cocci [B03.510.400]

Staphylococcaceae [B03.510.400.790]

Staphylococcus [B03.510.400.790.750]

Staphylococcus aureus [B03.510.400.790.750.100]

Staphylococcus epidermidis [B03.510.400.790.750.343]

Staphylococcus haemolyticus [B03.510.400.790.750.400]

Staphylococcus hominis [B03.510.400.790.750.425]

Return to Entry Page

THESAURUS: Medical Subject Headings

2006 MeSH

MeSH Descriptor Data

Return to Entry Page

MeSH Heading	Staphylococcus aureus
Tree Number	B03.510.400.790.750.100
Annotation	infection = STAPHYLOCOCCAL INFECTIONS & do not bother to coord with S. aureus unless particularly discussed (index IM); DF: STAPH AUREUS
Scope Note	Potentially pathogenic bacteria found in nasal membranes, skin, hair follicles, and perineum of warm-blooded animals. They may cause a wide range of infections and intoxications.
Allowable Qualifiers	CH CL CY DE EN GD GE IM IP ME PH PY RE UL VI
Entry Version	STAPH AUREUS
Previous Indexing	Staphylococcus (1966-1974)
Online Note	use STAPHYLOCOCCUS AUREUS to search MICROCOCCUS PYOGENES 1975-91; use STAPHYLOCOCCUS 1966-74
History Note	76; was MICROCOCCUS PYOGENES see under STAPHYLOCOCCUS 1963-75; MICROCOCCUS PYOGENES was see STAPHYLOCOCCUS AUREUS 1976-91
Unique ID	D013211

MeSH Tree Structures

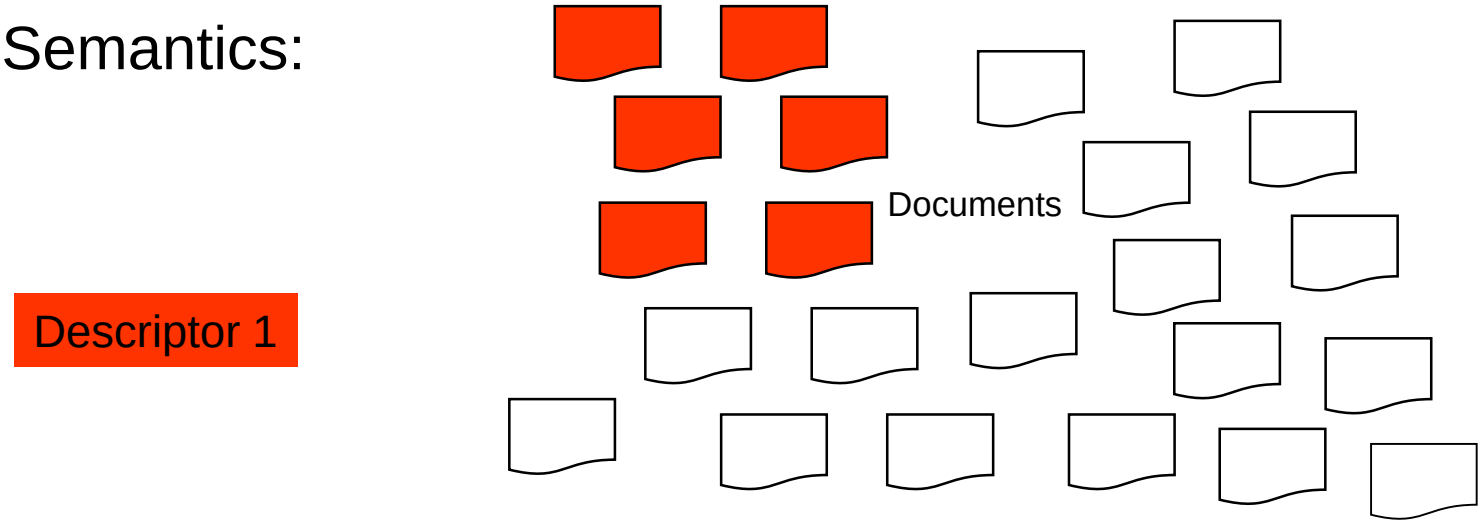
Bacteria [B03]
- Gram-Positive Bacteria [B03.510]
  - Gram-Positive Cocci [B03.510.400]
    - Staphylococcaceae [B03.510.400.790]
      - Staphylococcus [B03.510.400.790.750]
        - ▶ Staphylococcus aureus [B03.510.400.790.750.100]
        - Staphylococcus epidermidis [B03.510.400.790.750.343]
        - Staphylococcus haemolyticus [B03.510.400.790.750.400]
        - Staphylococcus hominis [B03.510.400.790.750.425]

Return to Entry Page

Link to NLM Cataloging Classification

THESAURUS: Medical Subject Headings

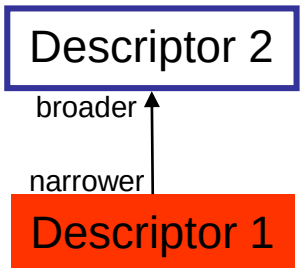
- Nodes:
  - Descriptors for content of biomedical publications
  - Labels: Common, Unambiguous Terms; Definitions (scope notes)
  - Terms: entry terms (synonyms, more specific terms) , translations
- Links:
  - Polyhierarchical connections of “broader” with “narrower terms”
- Semantics:

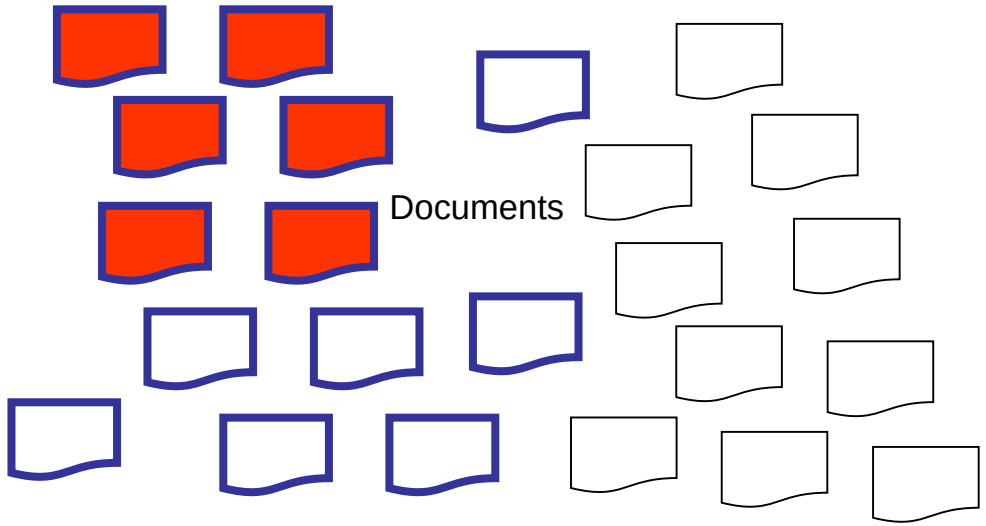


Descriptor 2 is broader than descriptor 1

THESAURUS: Medical Subject Headings

- Nodes:
  - Descriptors for content of biomedical publications
  - Labels: Common, Unabbiguous Terms, Definitions (scope notes)
  - Terms: entry terms (synonyms, more specific terms) , translations
- Links:
  - Polyhierarchical connections of “broader” with “narrower terms”
- Semantics:

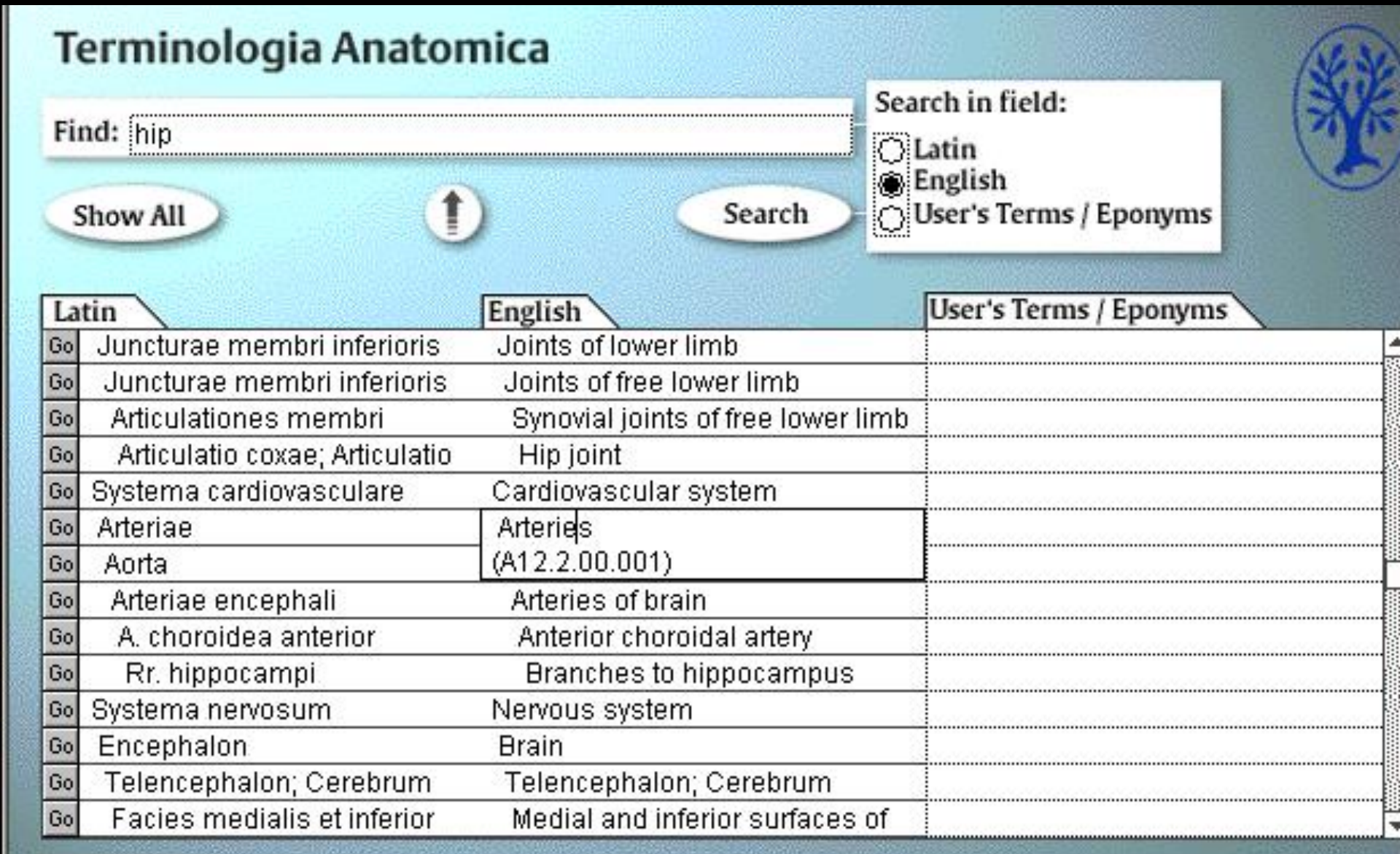




Descriptor 2 is broader than descriptor 1

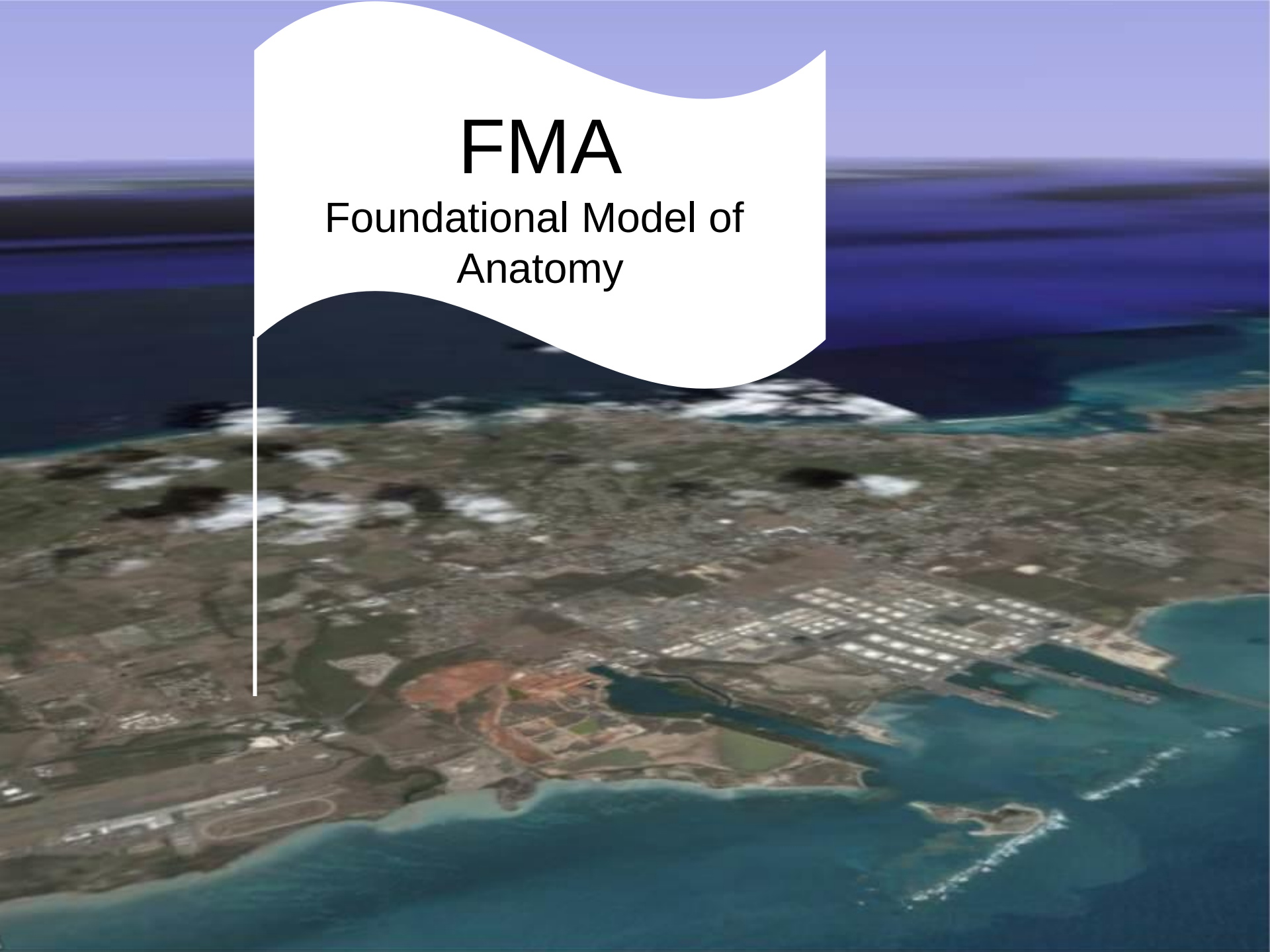

TA
Terminologia Anatomica

NOMENCLATURE: Terminologia Anatomica

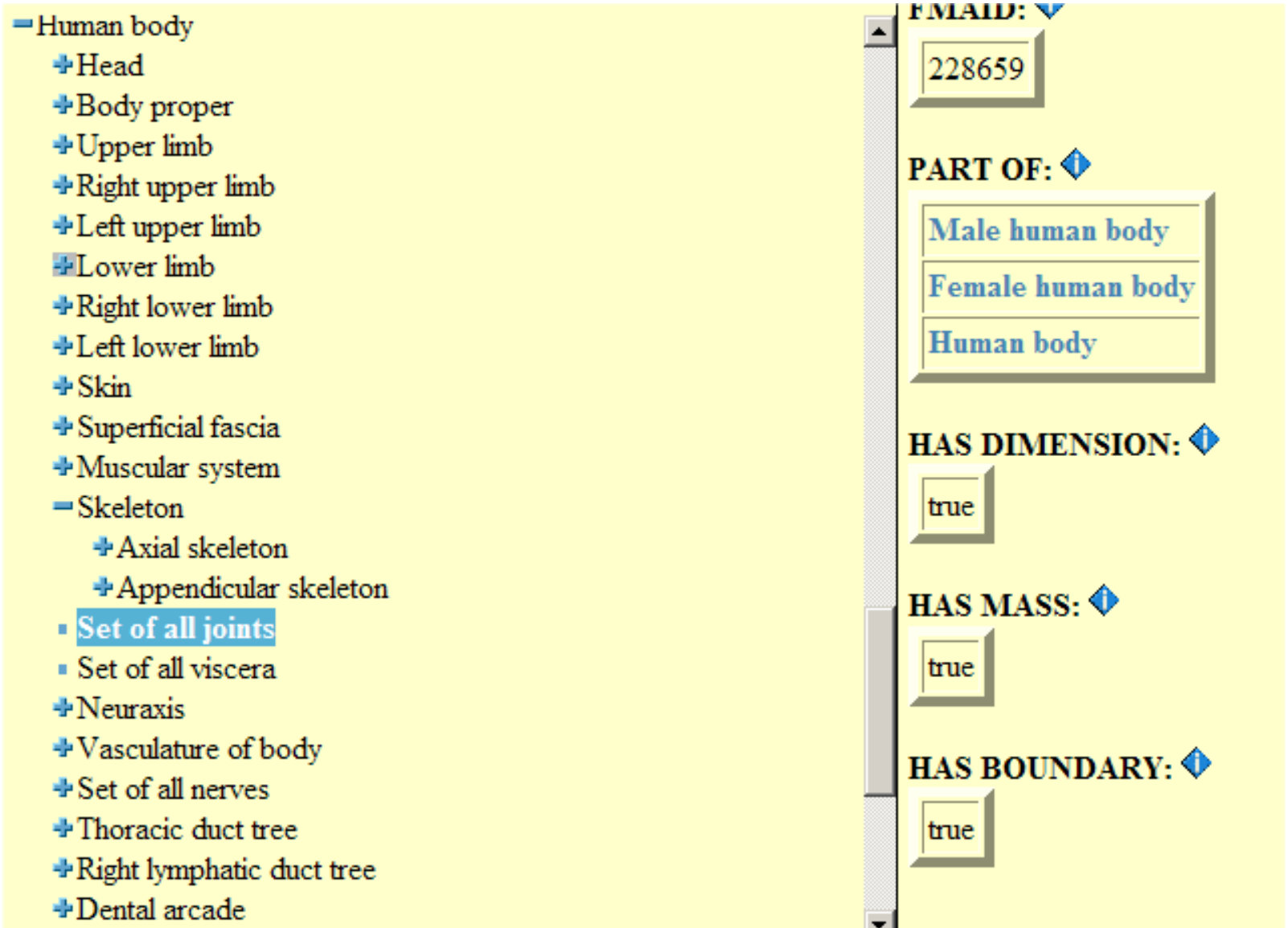


NOMENCLATURE: Terminologia Anatomica

- Nodes:
  - Standardized Anatomical Terms (English / Latin)
    “Junctura Membris Inferioris”- “Joints of Lower Limb”
- Links:
  - Partonomic
- Semantics:
  - A part of B: In an canonic instance of a human body the anatomical structure denoted by A is included into the anatomical structure denoted by B – and vice versa


FMA
Foundational Model of Anatomy

ONTOLOGY: Foundational Model of Anatomy

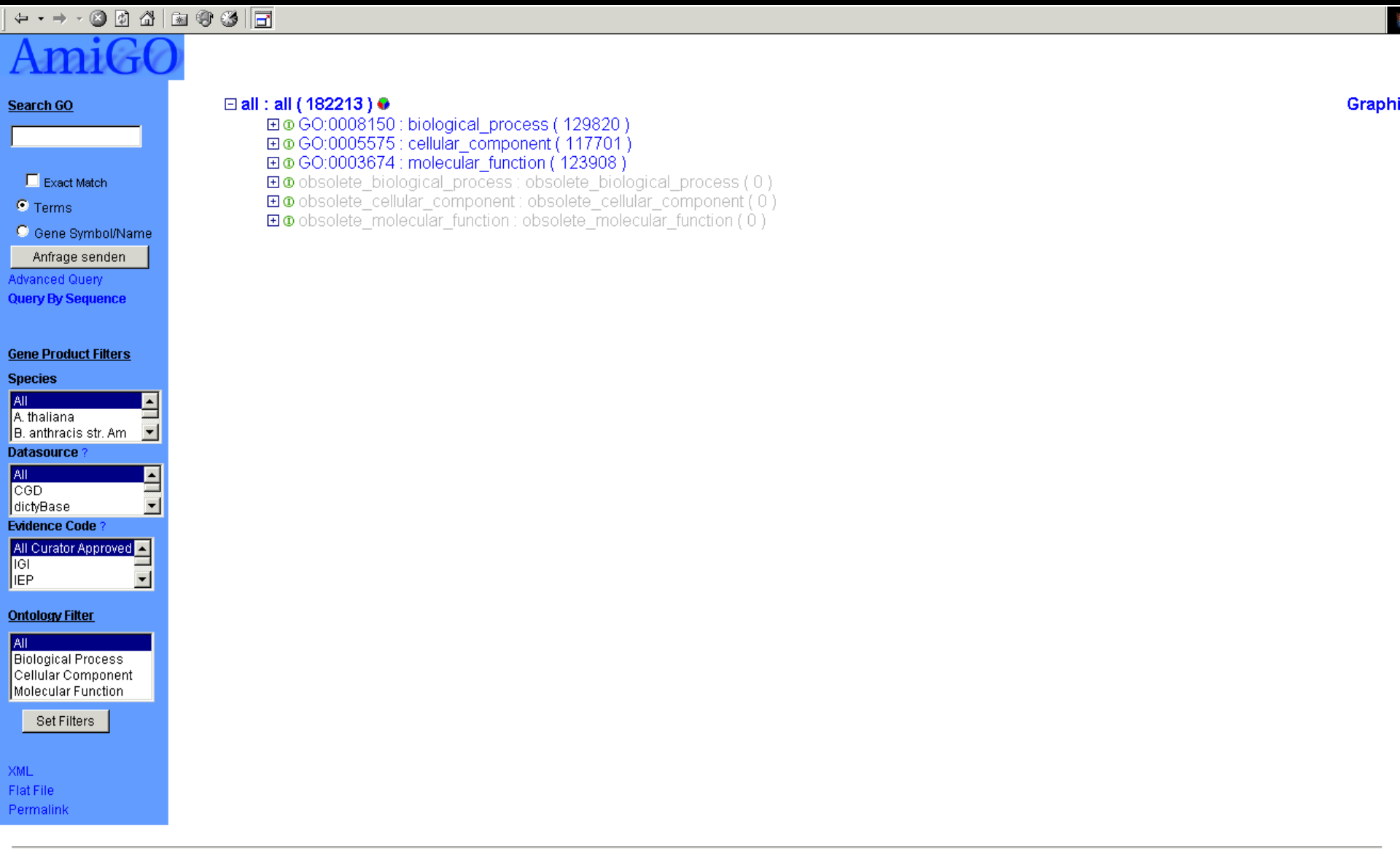


ONTOLOGY: Foundational Model of Anatomy

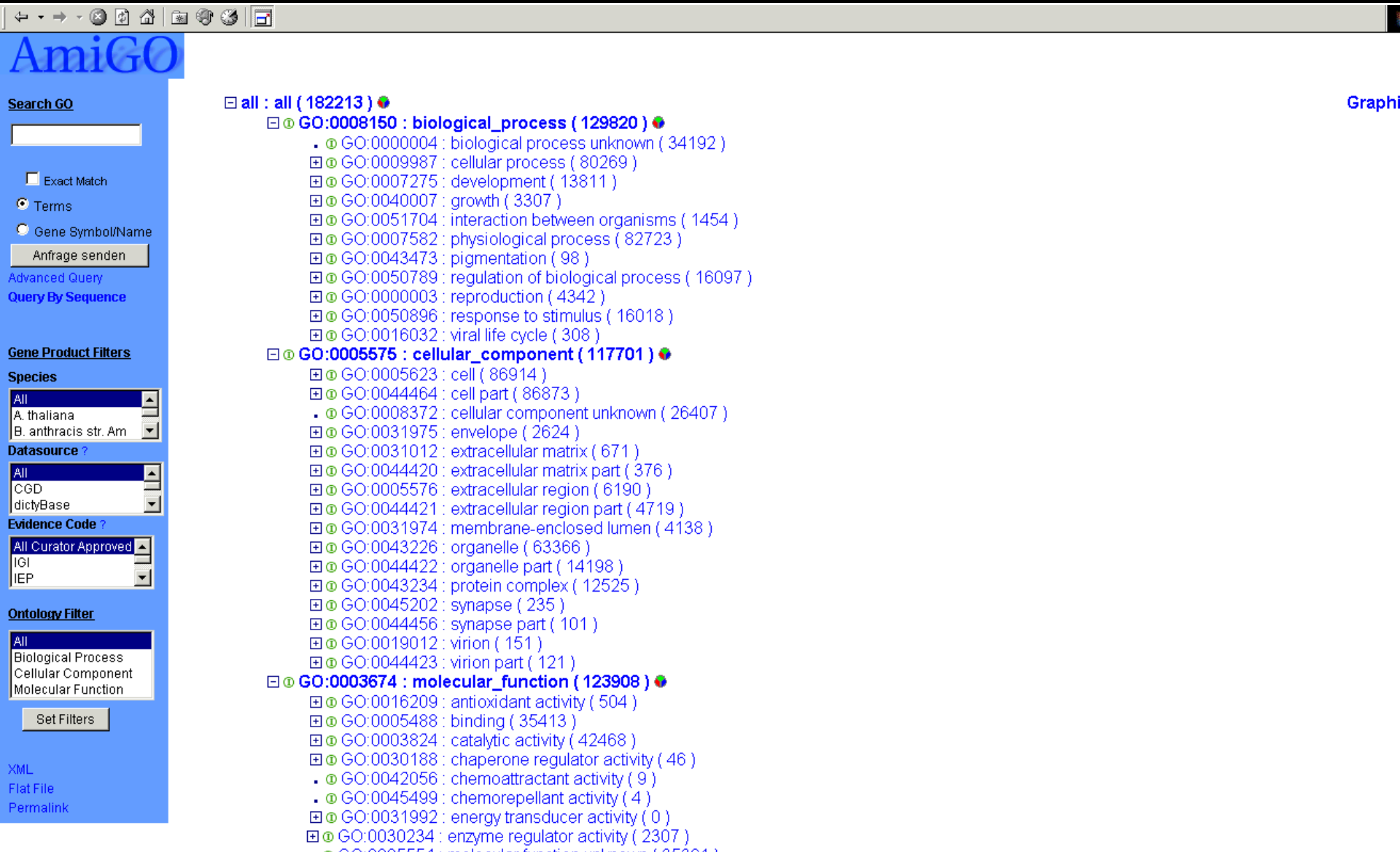
- Nodes:
  - Classes of anatomical entities that constitute a canonic human body
  - Labels: Exact anatomical terms, compatible with TA
    "Posterior ramus of third thoracic nerve"
  - Terms: Synonyms and Translations
- Links:
  - Taxonomic, Partonomic, Topological
- Semantics:
  - Frame-based
  - Taxonomy: All particular entities that instantiate one class, also instantiate all superclasses
  - A part-of B: In all canonic instances of a human body the anatomical structure that instantiates A is included into the anatomical structure that instantiates B
    and vice versa


GO
Gene Ontology

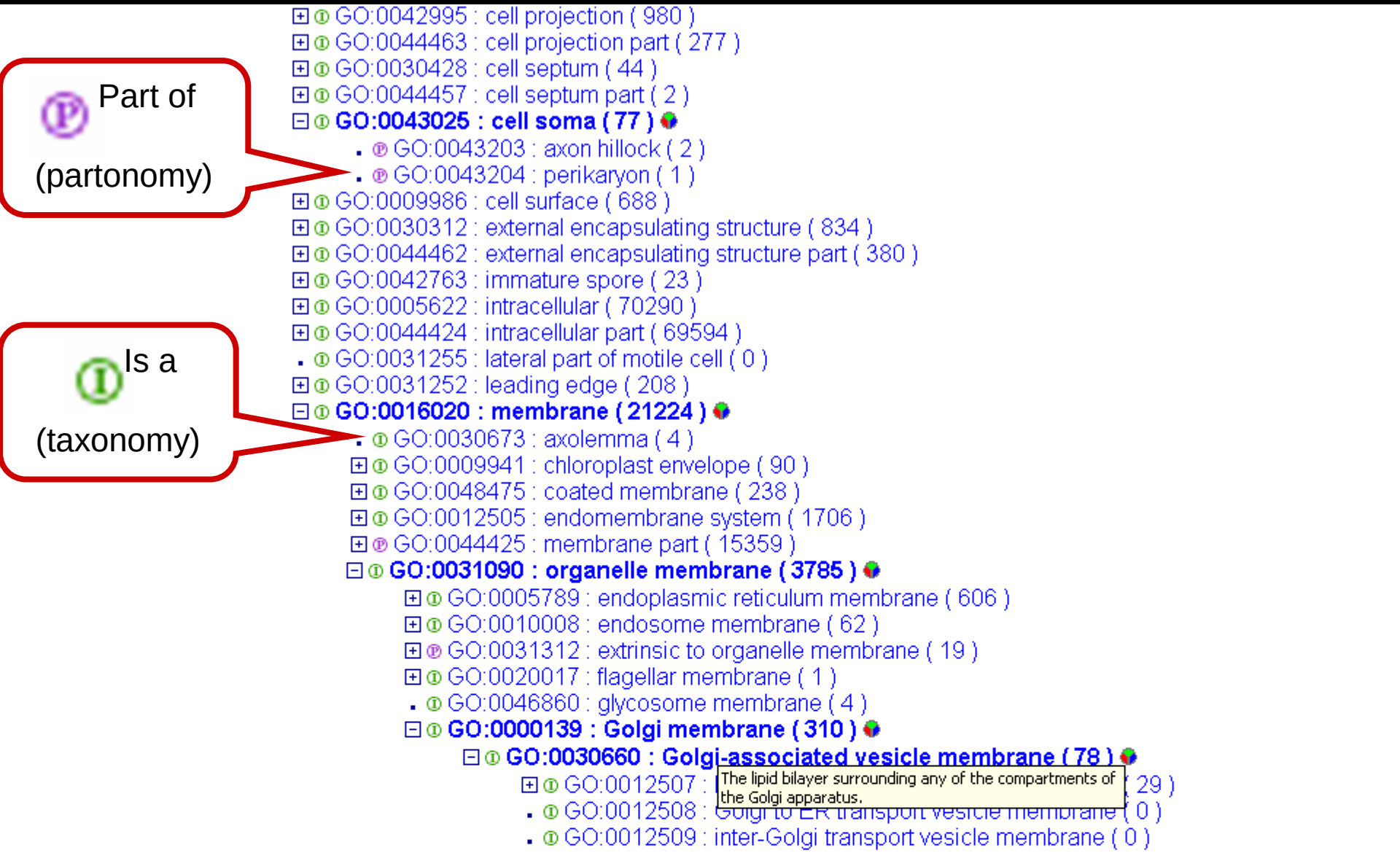
ONTOLOGY: Gene Ontology



ONTOLOGY: Gene Ontology



ONTOLOGY: Gene Ontology



ONTOLOGY: Gene Ontology

- Nodes stand for:
  - Originally: document/resource descriptors like MeSH, now: classes of particular entities as delineated by the meaning of the ontology labels
  - Labels: unambiguous, self-explaining noun phrases
    "low voltage-gated potassium channel auxiliary protein activity"
- Links:
  - Connect classes with superclasses (taxonomy)
  - Connect parts with wholes (partonomy)
- Semantics:
  - Taxonomy: All particular entities that instantiate one class, also instantiate all superclasses
  - A part of B: All particular entities that instantiate A are part of at least one particular entity that instantiates B

openGALEN

ONTOLOGY: OpenGALEN

('SurgicalProcess' which
IsMainlyCharacterisedBy
{Performance
IsEnactmentOf ('SurgicalFixing' which
hasSpecificSubprocess ('SurgicalAccessing'
hasSurgicalOpenClosedness
(SurgicalOpenClosedness which hasAbsoluteState
surgicallyOpen))
actsSpeclflcallyOn (PathologlcalBodyStructure which <
Involves Bone
hasUniqueAssociatedProcess FracturingProcess
hasSpecificLocation (Collum whlch
IsSpecificSolidDivisionOf (Femur whlch
hasLeftRlghtSelectorleftSelect!on))>))))

ONTOLOGY: OpenGalen

- Nodes:
  - Medical Concepts
  - Labels: Artificial, Self-Explaining:
    "*SurgicalOpenClosedness*"
- Links:
  - Taxonomic, partonomic, other relations
- Semantics:
  - Description Logics T-Box (unary and binary predicates)
  - Non partonomic relations as existential restrictions
  - Sanctioning
  - Closed-world semantics

Better understanding SNOMED CT

SNOMED CT

SNOMED since 1965

Nomenclature / Pathology

multiaxial nomenclature of medicine

Logic-based descriptions

Fusion with CTV 3

Principles of Formal Ontology

Context Model

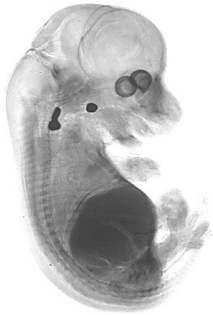
SNOMED im UMLS

IHTSDO

1965 1970 1975 1980 1985 1990 1995 2000 2005

SNOP SNOMED SNOMED II SNOMED 3.0 SNOMED 3.5 SNOMED RT SNOMED CT

Embryo *Fetus* *Infant* *Child* *Adolescence*



SNOMED CT

- The current structure of SNOMED CT is a result of its evolution
- Represents several tendencies from decades of nomenclature, terminology, ontology, and classification system development

SNOMED CT

Nomenclature: Multiaxial Structure

- SNOMED CT Concept
  - Body structure
  - Clinical finding
  - Environment or geographical location
  - Event
  - Linkage concept
  - Observable entity
  - Organism
  - Pharmaceutical / biologic product
  - Physical force
  - Physical object
  - Procedure
  - Qualifier value
  - Record artifact
  - Situation with explicit context
  - Social context
  - Special concept
  - Specimen
  - Staging and scales
  - Substance

Formal Language

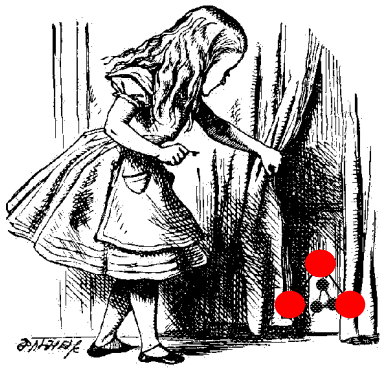
Benign neoplasm of heart =
64572001|disease|:
{116676008|hasasociated morphology|
=3898006|neoplasm, benign|
,363698007|finding site|=80891009|heart structure|}

Thesaurus

Hierarchy: CTV3 Navigation (v8)

416118004	administration
255203001	additional values
213769004	causes of injury and poisoning
71388002	procedure
254291000	staging and scales
14679004	occupation
373873005	pharmaceutical / biologic product
334251007	appliance
258666001	unit
410607006	organism

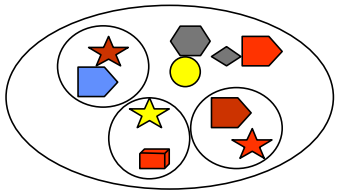
Ontological Principles



Sanctioning

- *Qualifiers*
  - severity
    - severities
  - episodicity
    - episodicities
  - clinical course
    - courses

Clinically relevant classes



SNOMED CT

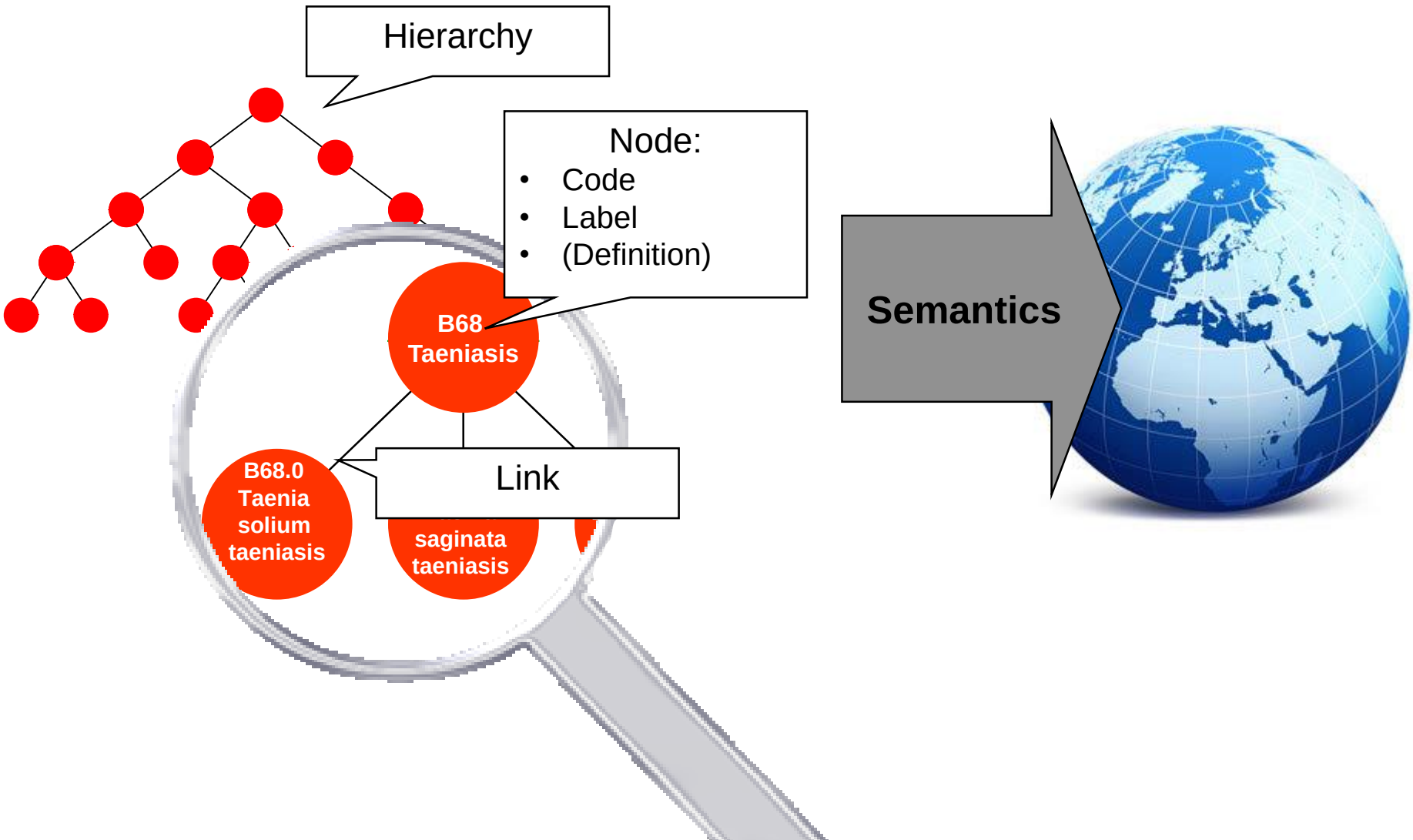
- The current structure of SNOMED CT is a result of its evolution
- Represents several tendencies from decades of nomenclature, terminology, ontology, and classification system development
- Identification of elements of
  - Terminology
  - Ontology

SNOMED CT

- The current structure of SNOMED CT is a result of its evolution
- Represents several tendencies from decades of nomenclature, terminology, ontology, and classification system development

Terminology vs. Ontology

What biomedical vocabularies have in common



Terminology vs. Ontology

Terminology

Set of terms representing the system of concepts of a particular subject field. (ISO 1087)

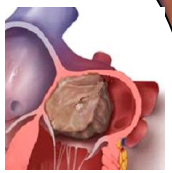
Formal Ontology

Ontology is the study of what there is. Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities.
(Stanford Encyclopedia of Philosophy)

Terminology

Entities of Language (Terms)

Shared / Meanings / Entities of Thought (Concepts)



„benign neoplasm of heart“
„gutartige Neubildung des Herzmuskels”
“neoplasia cardíaca benigna”

Example: UMLS (mrconso table)

Entities of Language (Terms)

```
C0153957|ENG|P|L0180790|PF|S1084242|Y|A1141630||||MTH|PN|U001287|benign neoplasm of heart|0|N||
C0153957|ENG|P|L0180790|VC|S0245316|N|A0270815||||ICD9CM|PT| 212.7|Benign neoplasm of heart|0|N||
C0153957|ENG|P|L0180790|VC|S0245316|N|A0270817||||RCD|SY|B727.| Benign neoplasm of heart|3|N||
C0153957|ENG|P|L0180790|VO|S1446737|Y|A1406658||||SNMI|PT|    D3-F0100|Benign neoplasm of heart, NOS|3|N||
C0153957|ENG|S|L0524277|PF|S0599118|N|A0654589||||RCDAE|PT|B727.|Benign tumor of heart|3|N||
C0153957|ENG|S|L0524277|VO|S0599510|N|A0654975||||RCD|PT|B727.| Benign tumour of heart|3|N||
C0153957|ENG|S|L0018787|PF|S0047194|Y|A0066366||||ICD10|PS|D15.1|Heart|3|Y||
C0153957|ENG|S|L0018787|VO|S0900815|Y|A0957792||||MTH|MM|U003158|Heart <3>|0|Y||
C0153957|ENG|S|L1371329|PF|S1624801|N|A1583056|||10004245|MDR|LT|10004245|Benign cardiac neoplasm|3|N||
C0153957|GER|P|L1258174|PF|S1500120|Y|A1450314||||DMDICD10|PT| D15.1|Gutartige Neubildung: Herz|1|N||
C0153957|SPA|P|L2354284|PF|S2790139|N|A2809706||||MDRSPA|LT| 10004245|Neoplasia cardiaca benigna|3|N||
```

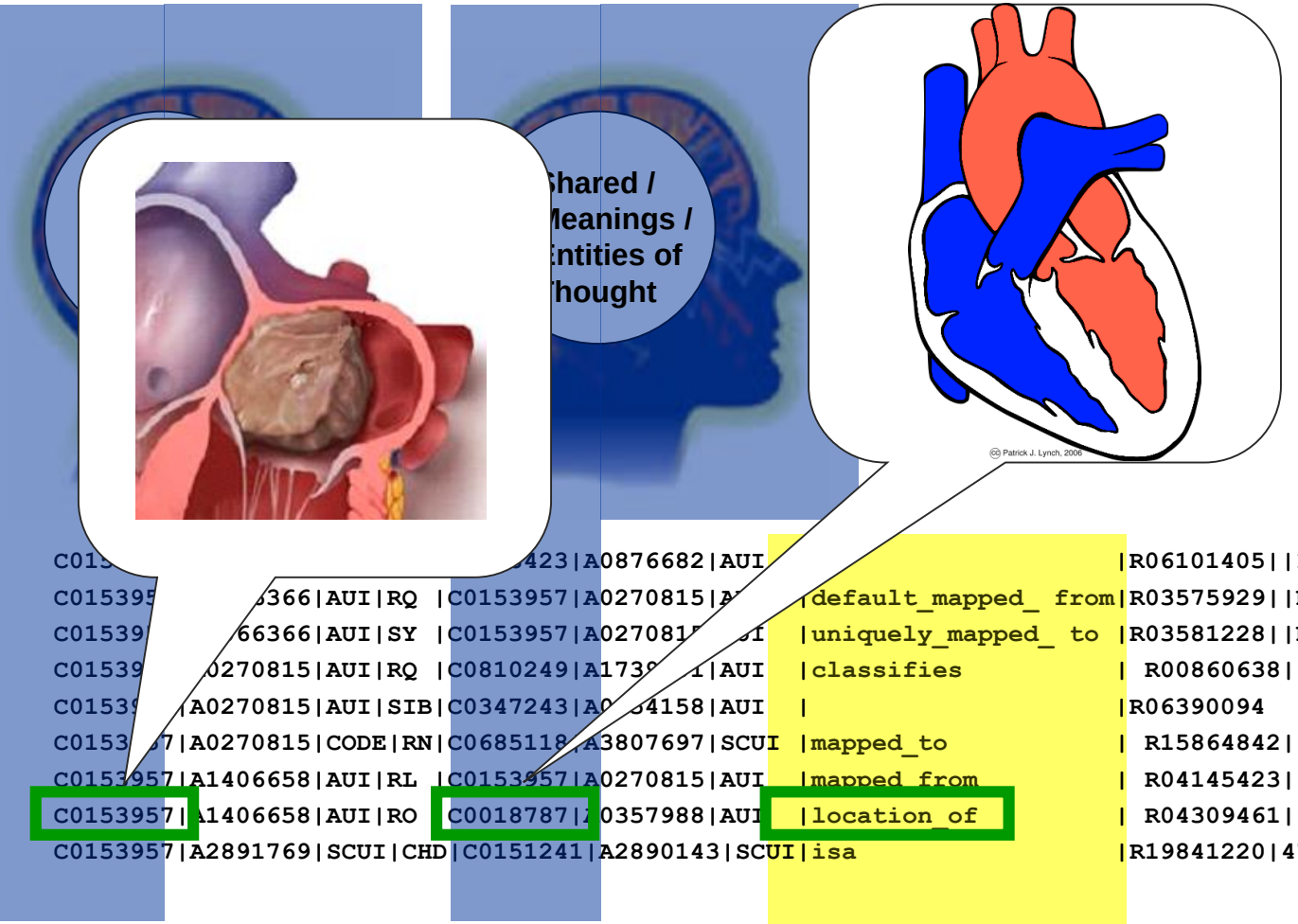
Unified Medical Language System, Bethesda, MD: National Library of Medicine, 2007: http://umlsinfo.nlm.nih.gov/

Example: UMLS (mrrel table)


```
C0153957|A0066366|AUI|PAR|C0348423|A0876682|AUI  |                    |R06101405||ICD10|ICD10|||N||
C0153957|A0066366|AUI|RQ |C0153957|A0270815|AUI  |default_mapped_ from|R03575929||NCISEER|NCISEER|||N||
C0153957|A0066366|AUI|SY |C0153957|A0270815|AUI  |uniquely_mapped_ to |R03581228||NCISEER|NCISEER|||N||
C0153957|A0270815|AUI|RQ |C0810249|A1739601|AUI  |classifies          | R00860638||CCS|CCS|||N||
C0153957|A0270815|AUI|SIB|C0347243|A0654158|AUI  |                    |R06390094          || ICD9CM|ICD9CM||N|N||
C0153957|A0270815|CODE|RN|C0685118|A3807697|SCUI |mapped_to           | R15864842||SNOMEDCT|SNOMEDCT||Y|N||
C0153957|A1406658|AUI|RL |C0153957|A0270815|AUI  |mapped_from         | R04145423||SNMI|SNMI|||N||
C0153957|A1406658|AUI|RO |C0018787|A0357988|AUI  |location_of         | R04309461||SNMI|SNMI|||N||
C0153957|A2891769|SCUI|CHD|C0151241|A2890143|SCUI|isa                 |R19841220|47189027|SNOMEDCT|SNOMEDCT|0|Y|N||
```

Semantic relations

Example: UMLS




```
C015                 423|A0876682|AUI                    |R06101405||ICD10|ICD10|||N||
C015395      366|AUI|RQ  |C0153957|A0270815|A           |default_mapped_ from|R03575929||NCISEER|NCISEER|||N||
C01539       66366|AUI|SY |C0153957|A027081?  ?I        |uniquely_mapped_ to |R03581228||NCISEER|NCISEER|||N||
C01539       0270815|AUI|RQ |C0810249|A173?  ?|AUI      |classifies          | R00860638||CCS|CCS|||N||
C0153?  |A0270815|AUI|SIB|C0347243|A0  34158|AUI      |                    |R06390094          || ICD9CM|ICD9CM||N|N||
C0153 7|A0270815|CODE|RN|C0685118|A3807697|SCUI       |mapped_to           | R15864842||SNOMEDCT|SNOMEDCT||Y|N||
C0153957|A1406658|AUI|RL |C0153957|A0270815|AUI       |mapped_from         | R04145423||SNMI|SNMI|||N||
C0153957|A1406658|AUI|RO |C0018787|A0357988|AUI       |location_of         | R04309461||SNMI|SNMI|||N||
C0153957|A2891769|SCUI|CHD|C0151241|A2890143|SCUI     |isa                 |R19841220|47189027|SNOMEDCT|SNOMEDCT|0|Y|N||
```

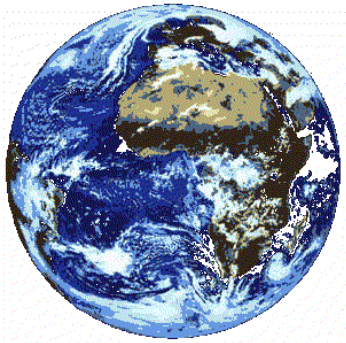
INFORMAL **Semantic relations**

Formal Ontology represents the world

ology

Set of terms representing the system of concepts of a particular subject field. (ISO 1087)

Formal Ontology



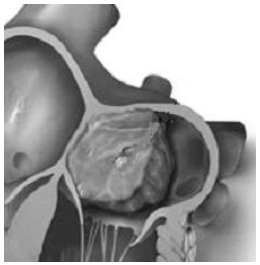
Ontology is the study of what there is (Quine).
Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities.
(Stanford Encyclopedia of Philosophy)

Formal Ontology

Organizing Entities

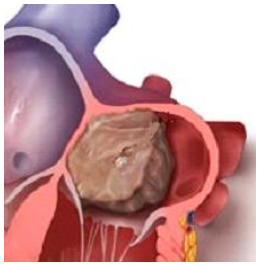
Entity Types

The type “benign neoplasm of heart”

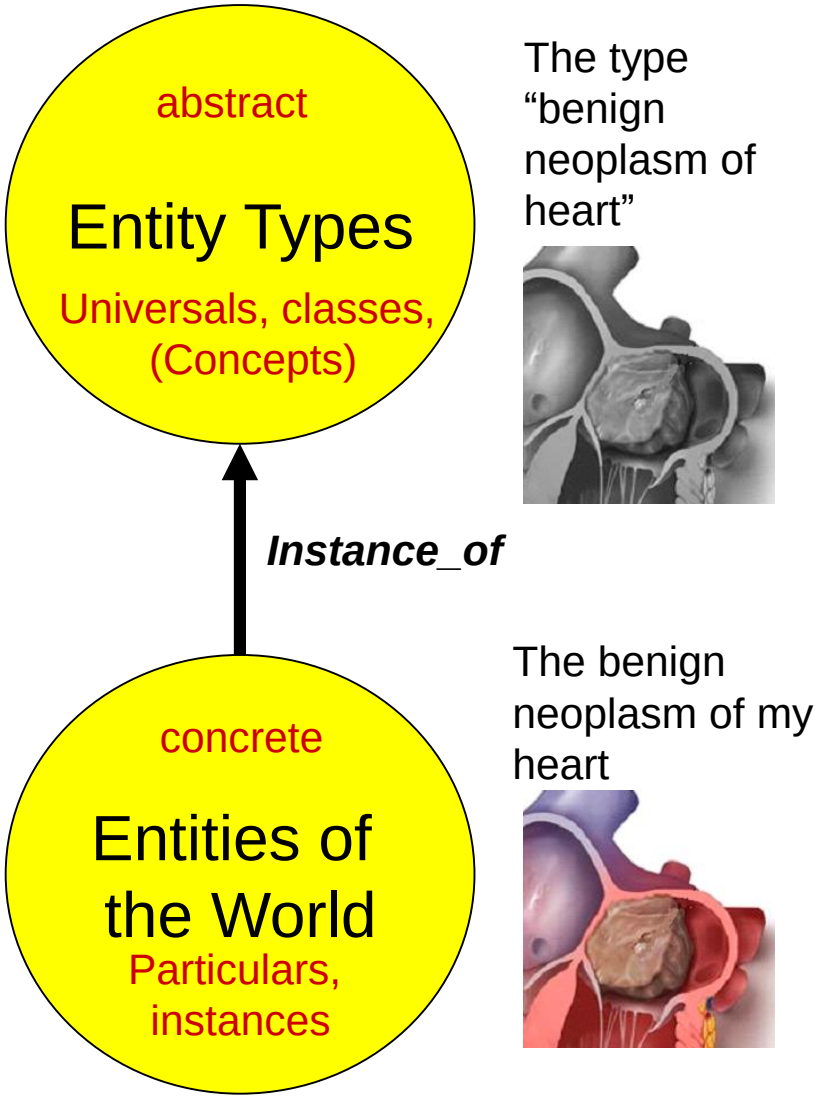


Entities of the World

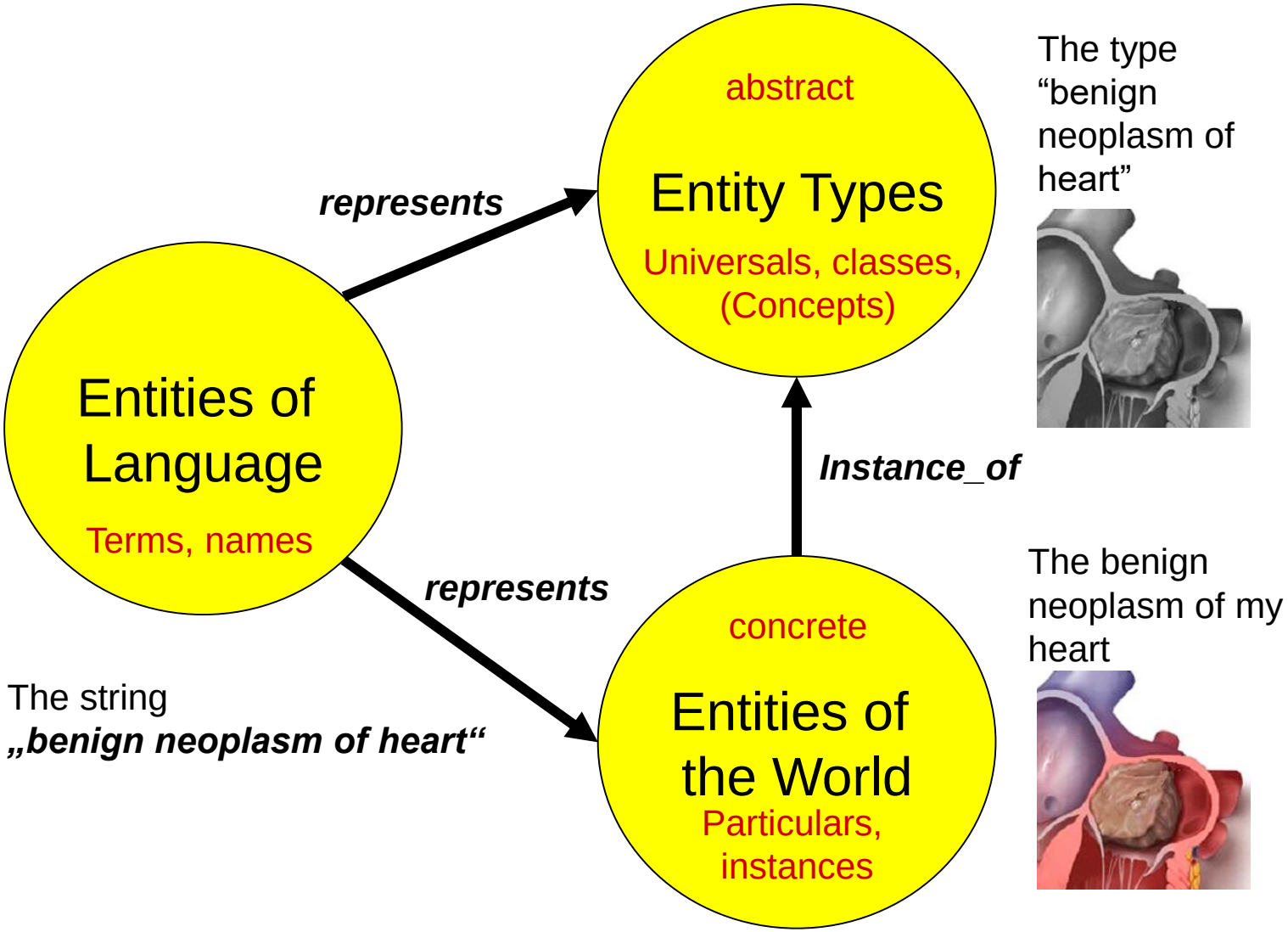
My benign neoplasm of heart



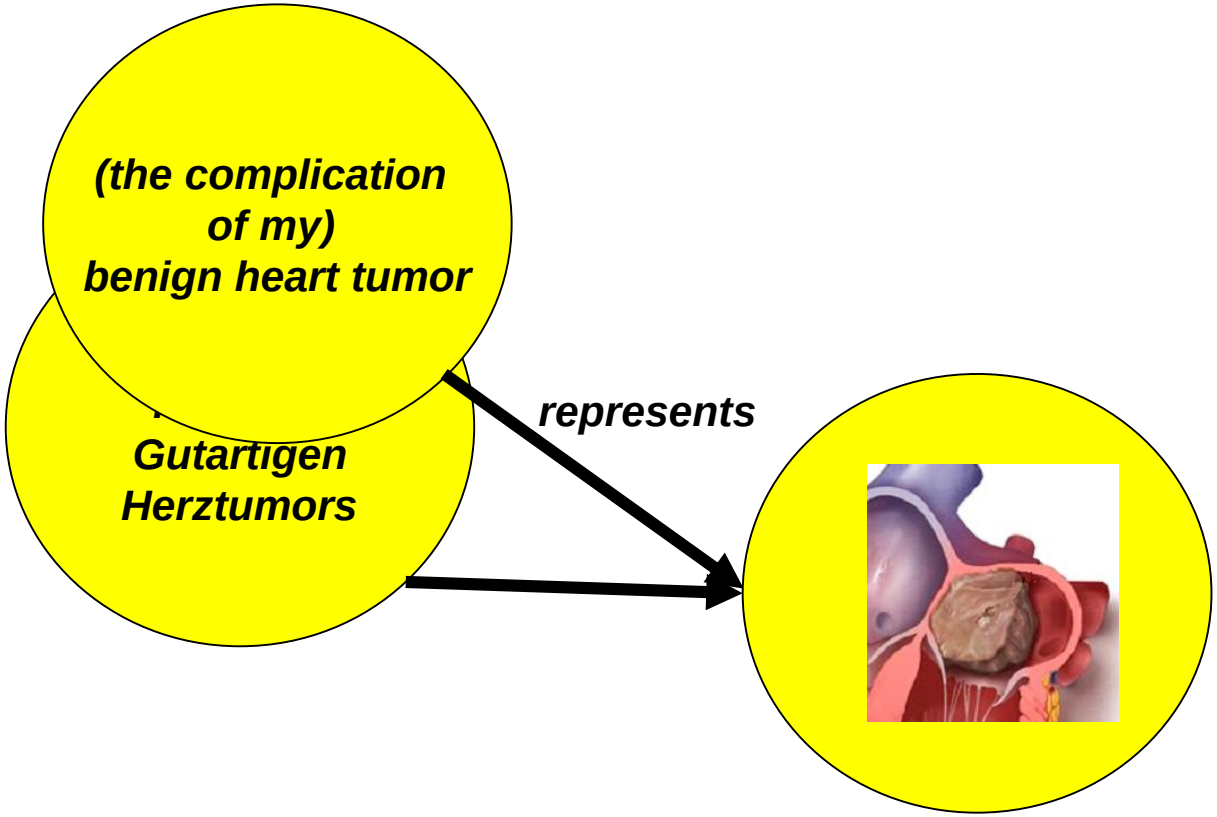
Organizing Entities



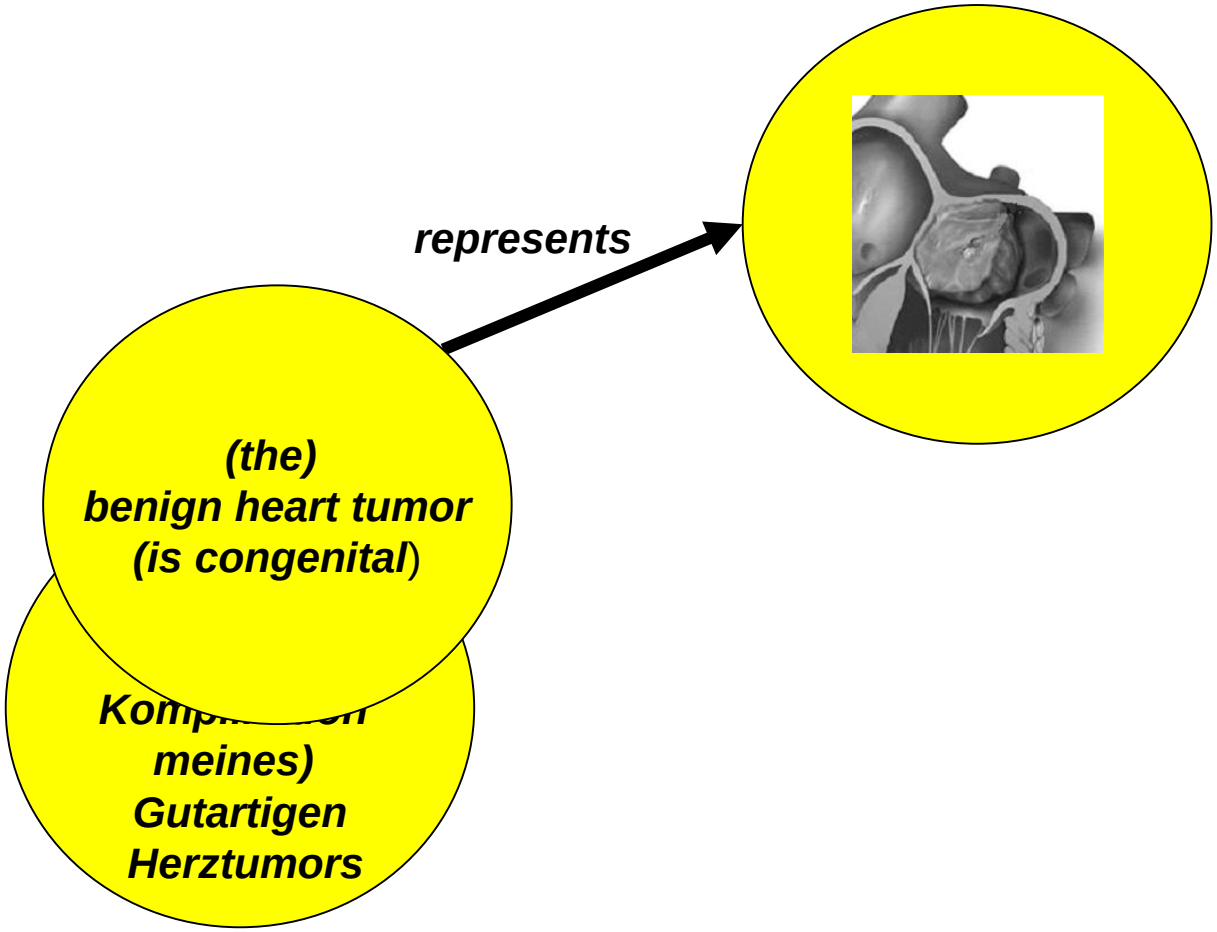
Organizing Entities

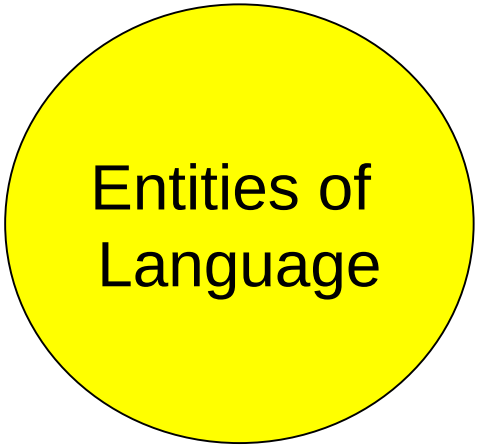


Organizing Entities



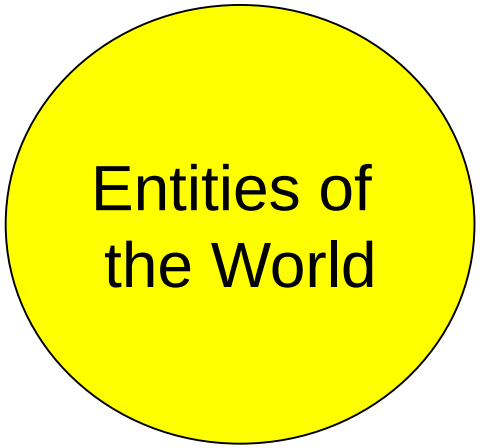
Organizing Entities

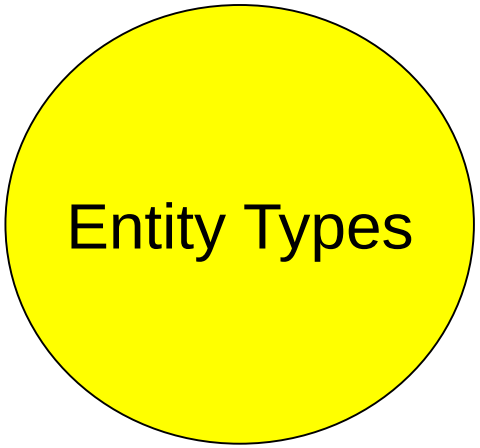




…are represented by terminologies

Databases systems represent …

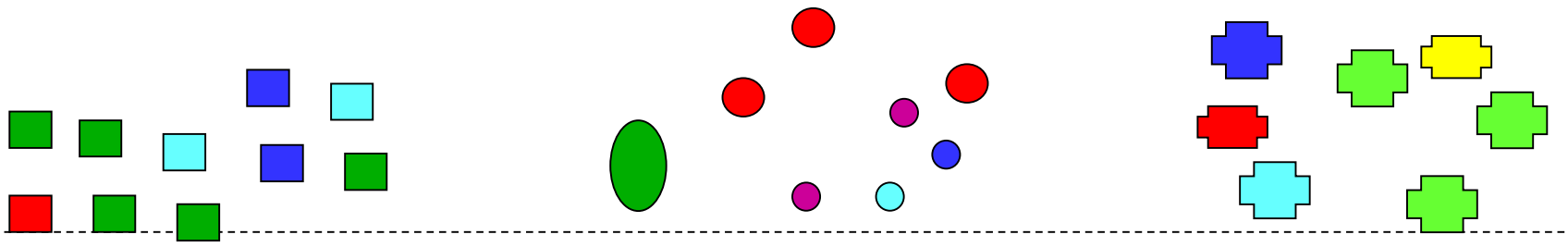




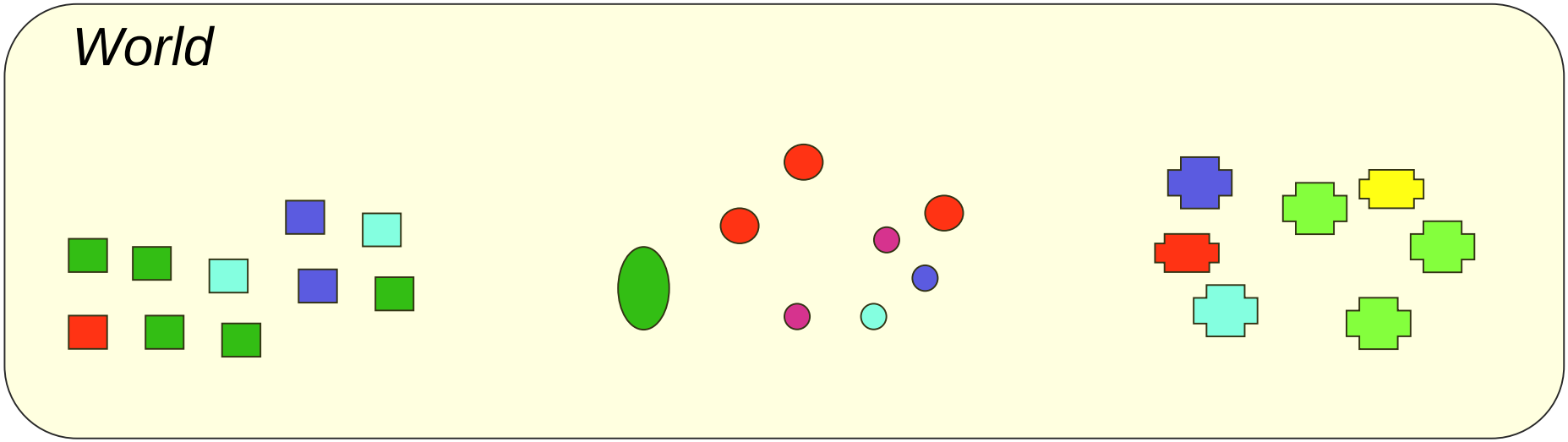
… are organized in formal ontologies

Hierarchies, Types, Classes, Individuals

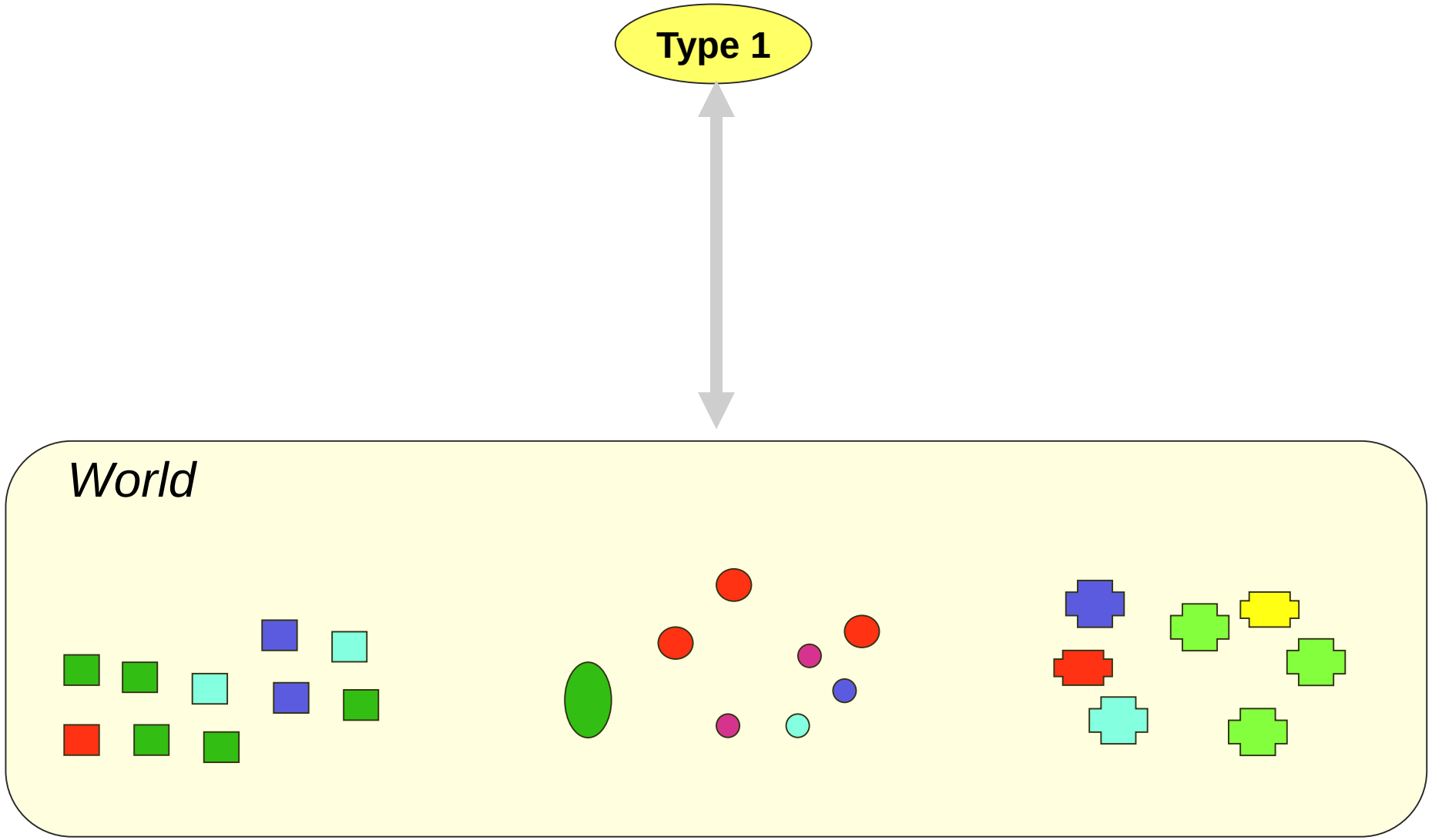
World



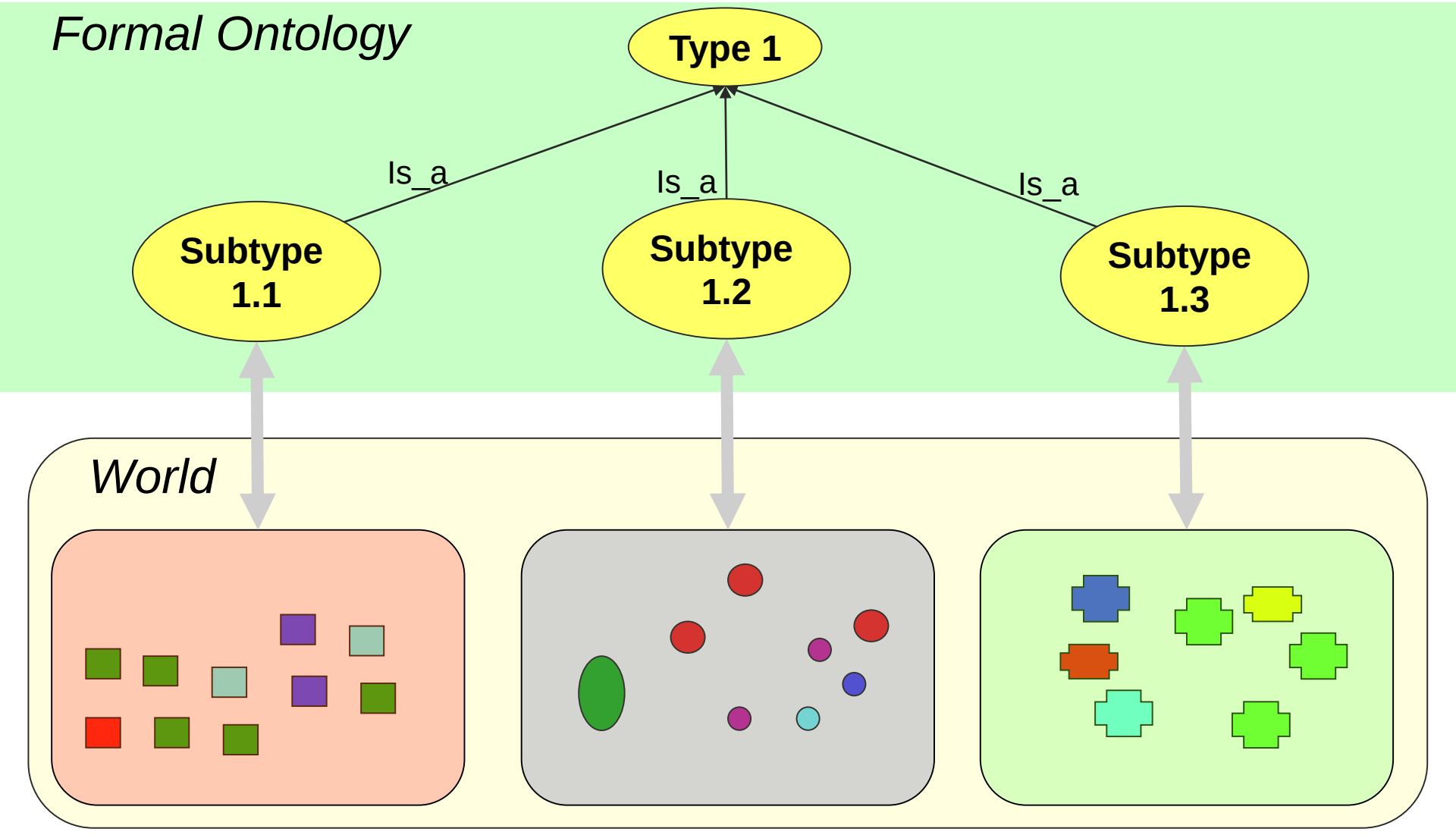
Hierarchies, Types, Classes, Individuals



Hierarchies, Types, Classes, Individuals

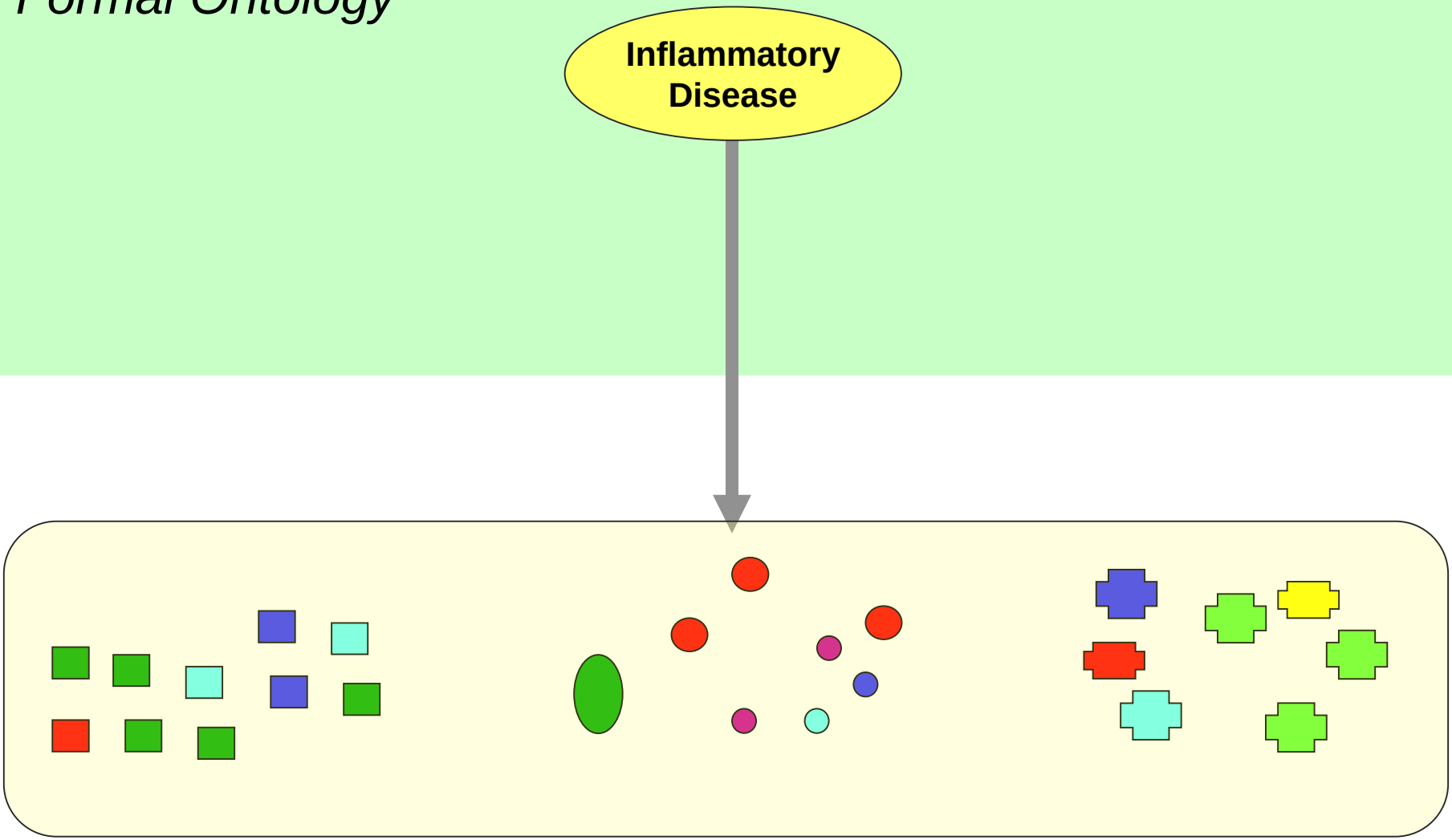


Hierarchies, Types, Classes, Individuals

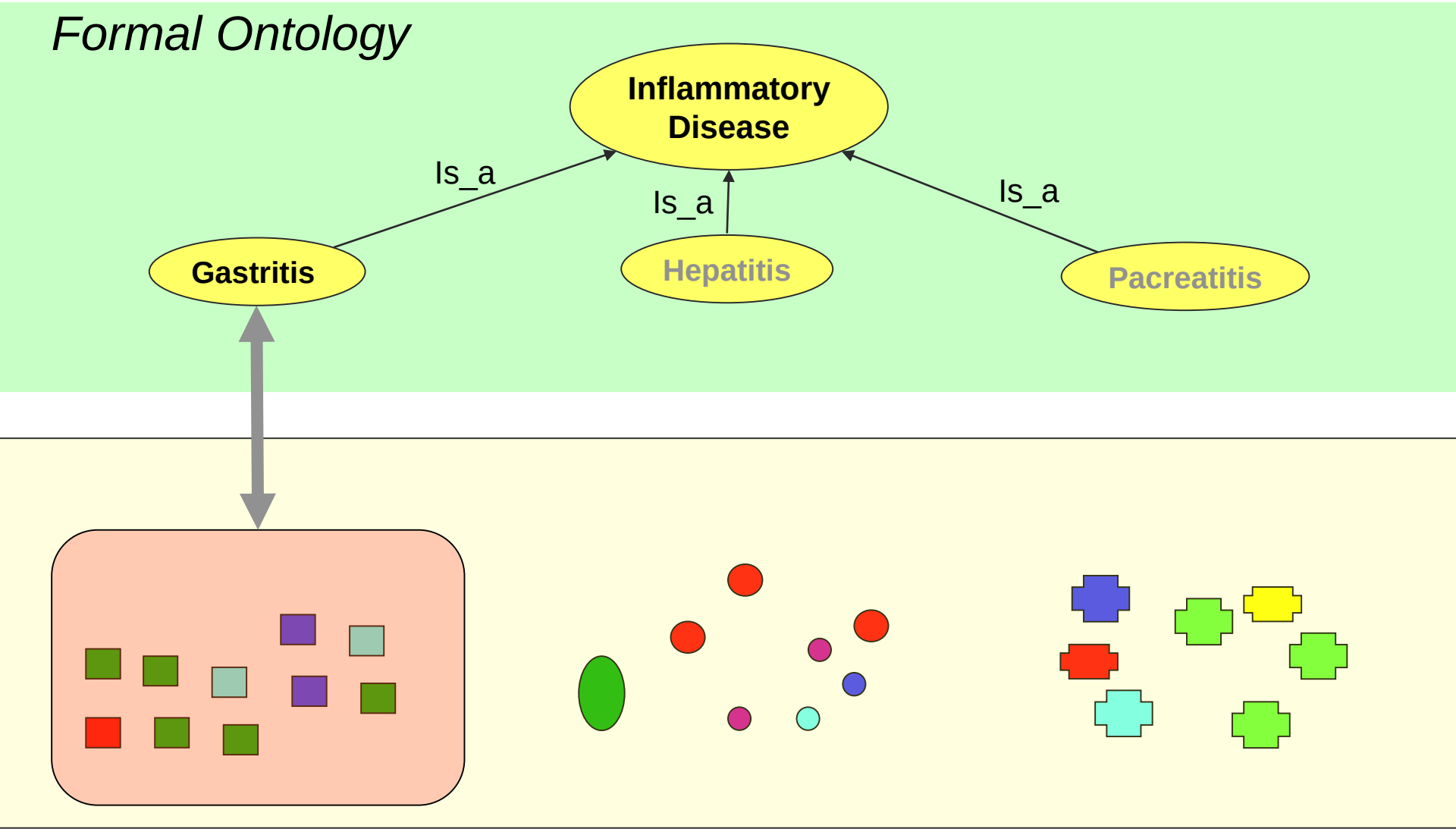


Hierarchies, Types, Classes, Individuals

Formal Ontology

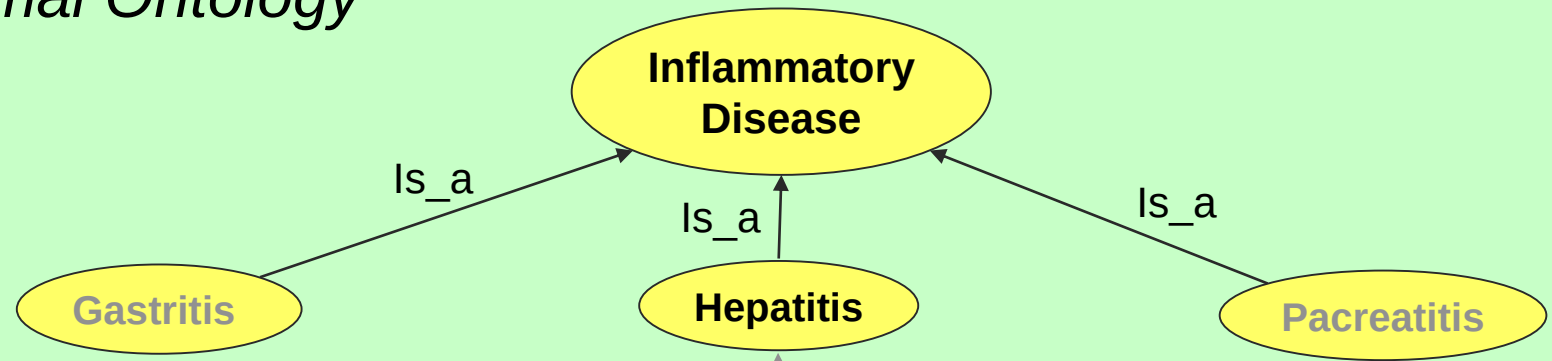


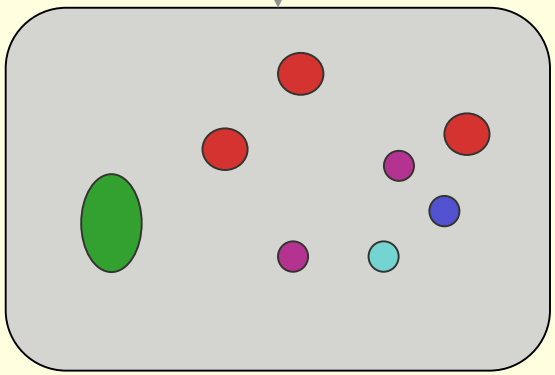
Hierarchies, Types, Classes, Individuals

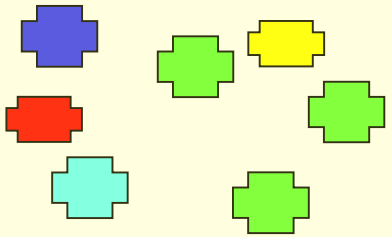


Hierarchies, Types, Classes, Individuals

Formal Ontology

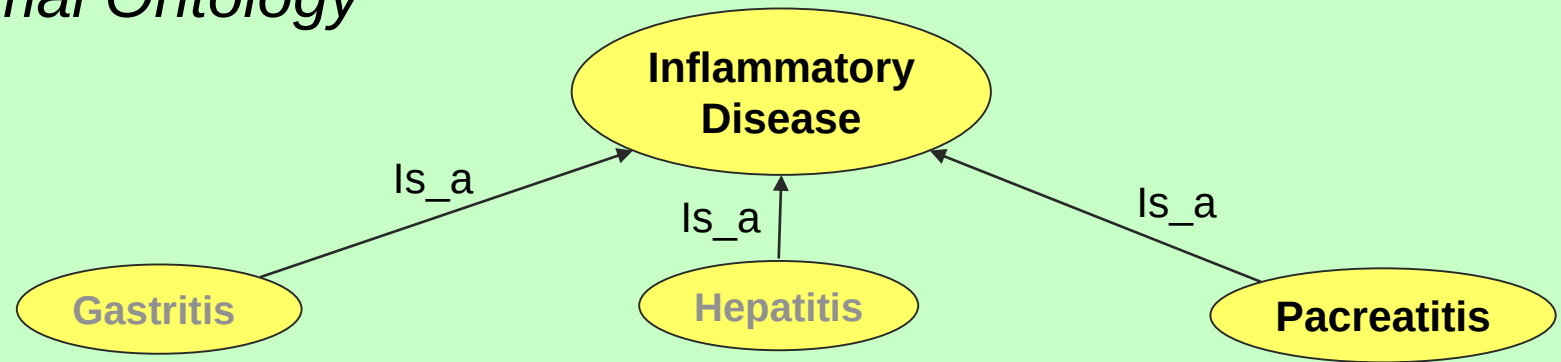


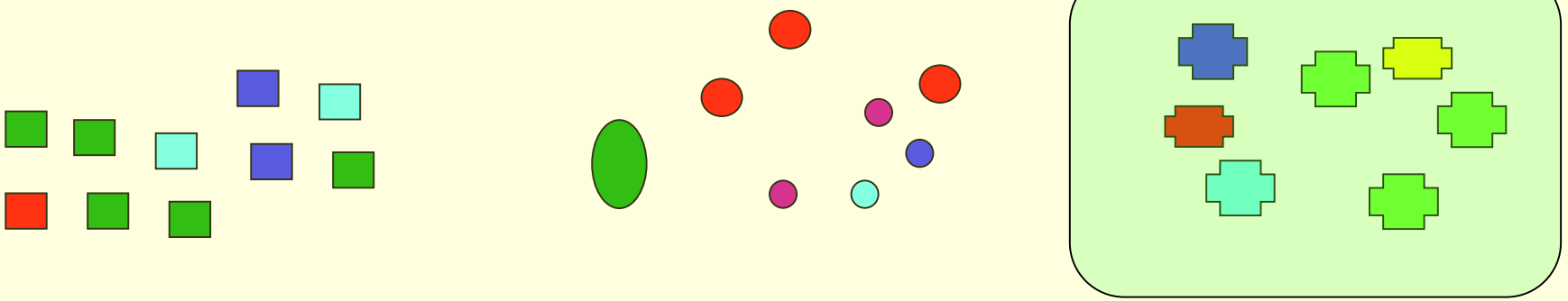




Hierarchies, Types, Classes, Individuals

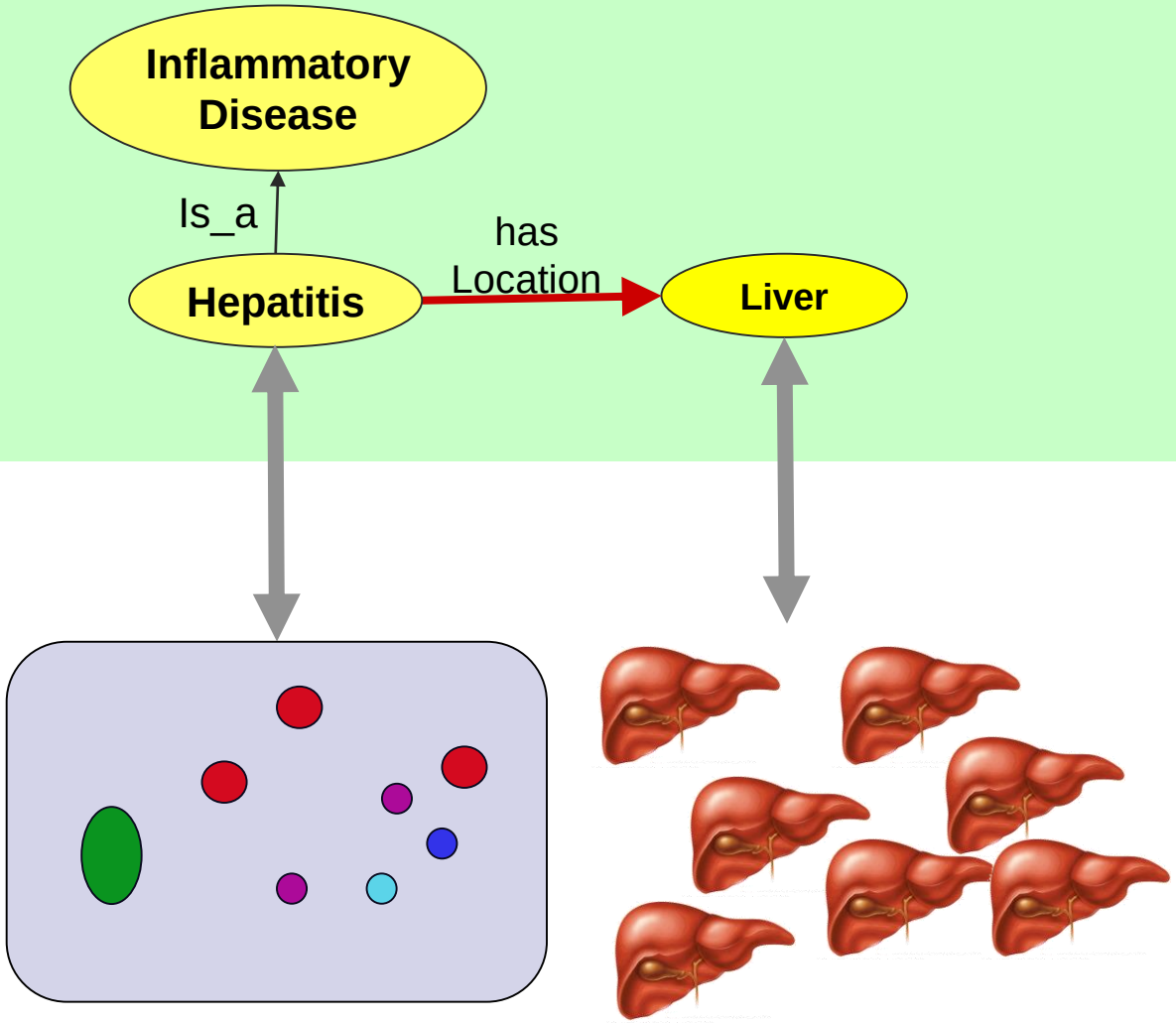
Formal Ontology





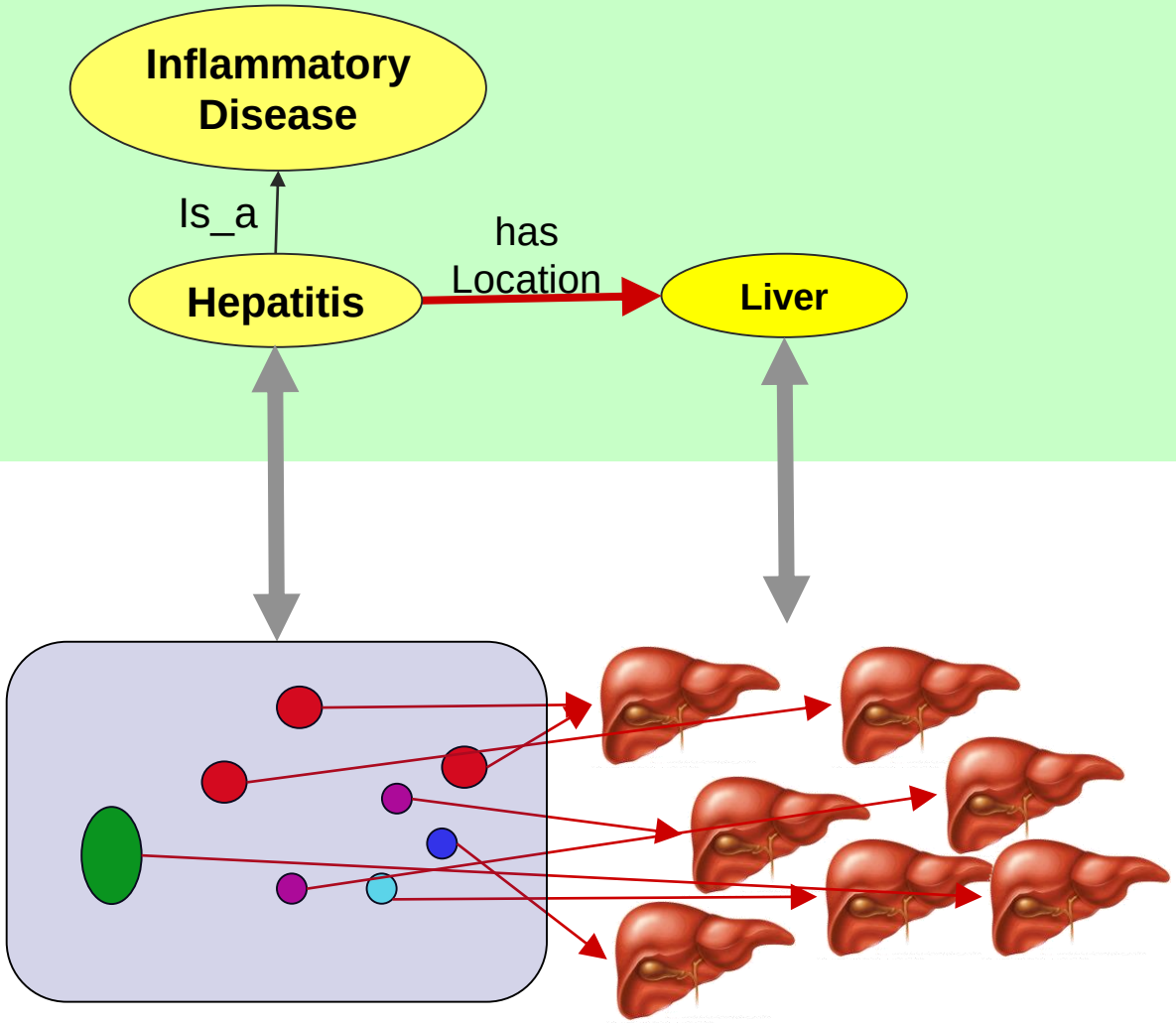
Relations and Definitions

Formal Ontology

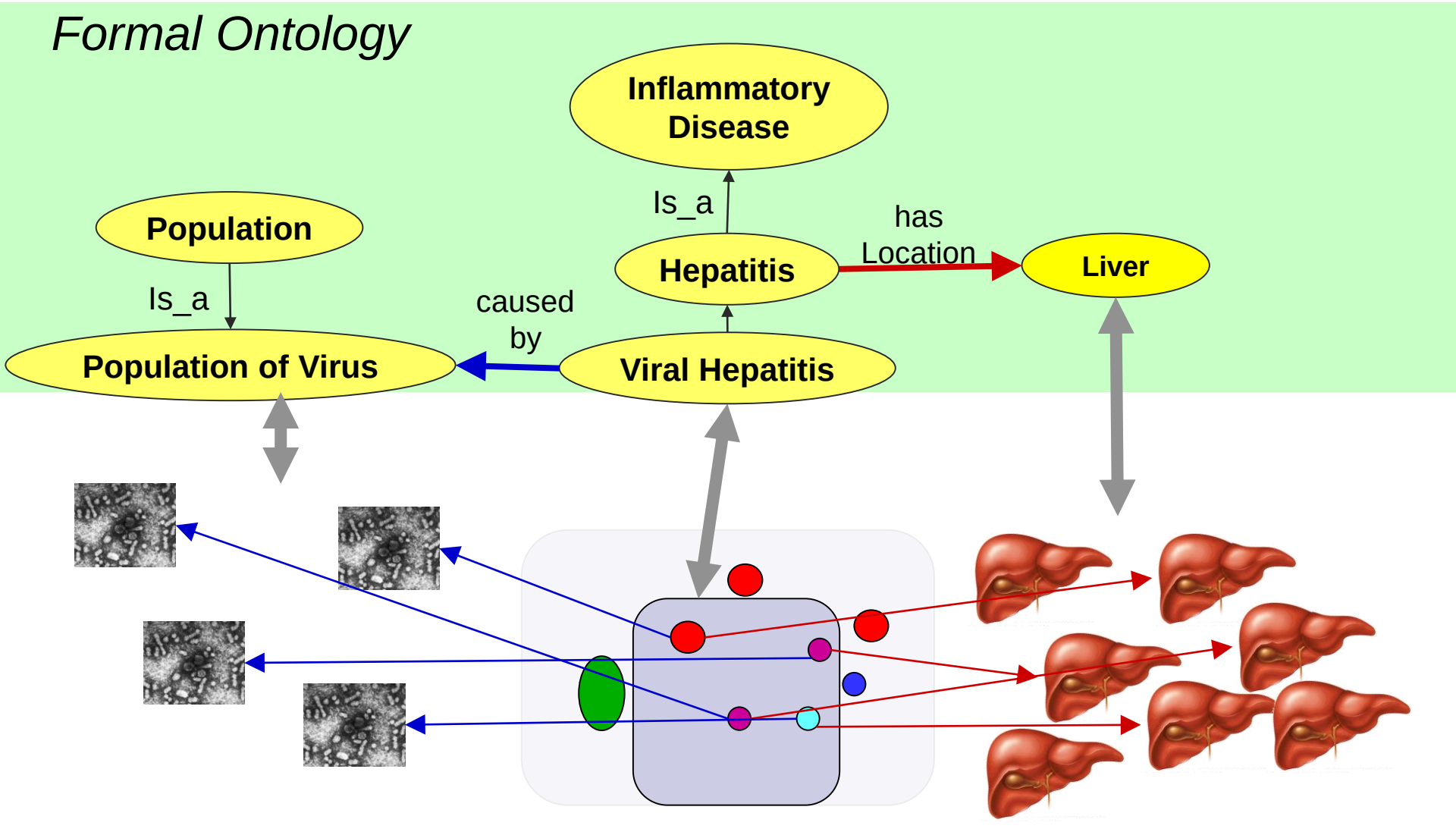


Relations and Definitions

Formal Ontology



Relations and Definitions



Languages for formal ontologies

- Natural Language

  "Every hepatitis is an inflammatory disease that is located in some liver"
  "Every inflammatory disease that is located in some liver is an hepatitis"

- Logic

$$\forall x: instanceOf(x, Hepatitis) \Leftrightarrow instanceOf(x, Inflammation) \wedge \exists y: instanceOf(y, Liver) \wedge hasLocation(x,y)$$

Logic is computable: it supports machine inferences but…

it only scales up if it has a very limited expressivity

SNOMED CT: Terminology and Ontology aspects

Fully Specified Name

Preferred Term

- Descriptions
  - F fracture of upper limb (disorder)
  - P fracture of upper limb
  - S arm fracture
  - S fracture of arm
  - S fracture of bone of upper limb

Synonyms

Same structure for other languages

- Definition: Fully defined by ...
  - is a
    - D fracture of bone
    - D injury of upper extremity
    - D finding of bone of upper limb
  - Group
    - associated morphology
      - D fracture
    - finding site
      - D bone structure of upper limb

Taxonomic Parents (isA)

Logical Restrictions

Full-text definitions mostly missing

Terminology

Formal Ontology

Terminologies vs. Formal Ontologies

Terminologies

- Describe: Meaning of human language units
- "Concepts": aggregate (quasi)-synonymous terms
- Relations: informal, elastic Associations between Concepts
- Description pattern: $Concept_1$ Relation $Concept_2$

Formal Ontologies

- Describe: entities of reality as they generically are – independent of human language
- "Types": represent the generic properties of world entities
- Relations: rigid, exactly defined, quantified relationships between particulars
- Description pattern: for all instance of $Type_1$: there is some…

Example Hepatitis - Liver

Terminologies

- Concept *Hepatitis*:
  {Hepatitis (D), Leberentzündung (D), hepatitis (E), hépatite (F)}
- Concept *Liver*:
  {Leber (D), liver (E), foie (F)}
- Relations:
  - *Hepatitis – hasLocation – Liver*
  - *Hepatitis – isA - Inflammation*

Formal Ontologies

- Type: *Hepatitis*:
- Description:

  ”*Every hepatitis is an inflammatory disease that is located in some liver*”
  “Every inflammatory disease that is located in some liver is an hepatitis”

Example Hand - Thumb

Terminologies

- Concept *Hand*:
  {Hand (D), hand (E), main (F)}
- Concept *Thumb*:
  {Daumen (D), thumb (E), pouce (F)}
- Relations:
  - *Hand – hasPart – Thumb*
  - *Thumb – partOf – Hand*

Formal Ontologies

- Type: *Thumb*:
- Description:

”Every thumb is part of some hand”
“Every hand has some thumb as part”

?

Example Aspirin - Headache

Terminologies

- Concept *Aspirin*:
  {Aspirin (D,E), Acetylsalicylsäure (D), ASS (D), acetylsalicylic acid (E), Acide acétylsalicylique(F)}
- Concept *Headache*:
  {Kopfschmerz (D), headache (E), céphalée(F)}
- Relation:
  - *Aspirin – treats – Headache*

fuzzy

Formal Ontologies

- Type: *Aspirin*:
- Description:
- *”For every portion of aspirin there is some disposition for treating headache”*

complicated !

Strengths of Formal Ontologies

- Exact, logic-based descriptions of entity types that are instantiated by real-world objects, processes, states
- Representation of stable, context-independent accounts of reality
- Use of formal reasoning methods using tools and approaches from the AI / Semantic Web community

Formal Ontologies: Limitations (I)

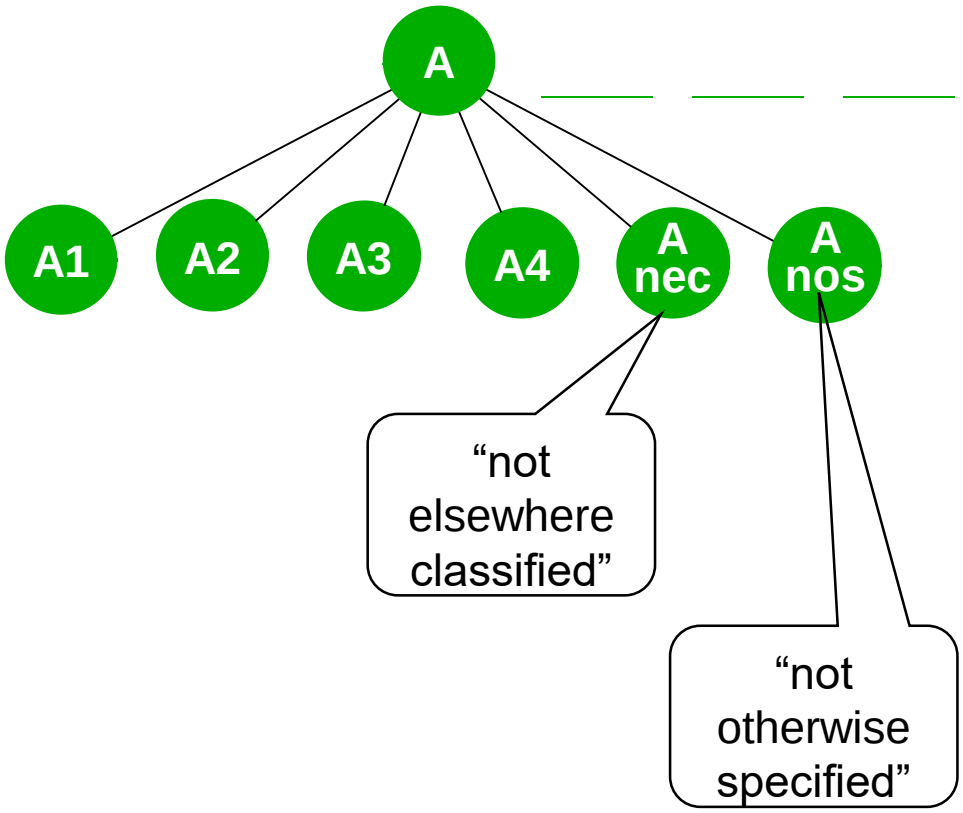
- Only suitable to represent shared, uncontroversial meaning of a domain vocabulary
- Supports universal statements about instances of a type:
  - All Xs are Ys
  - For all Xs there is some Y
- Properties of types are properties of all entities that instantiate these types (strict inheritance)

Classification vs. Ontology

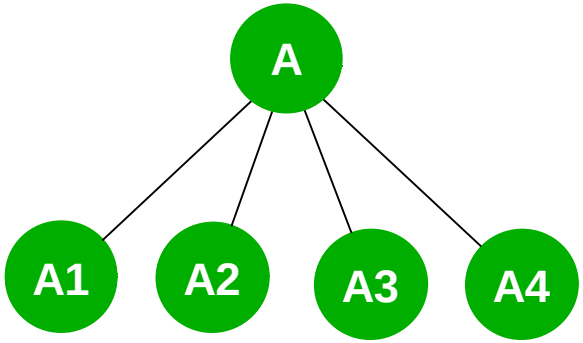
Classification systems vs. Ontologies

Classifications vs. Formal Ontologies

Classifications



Formal Ontologies

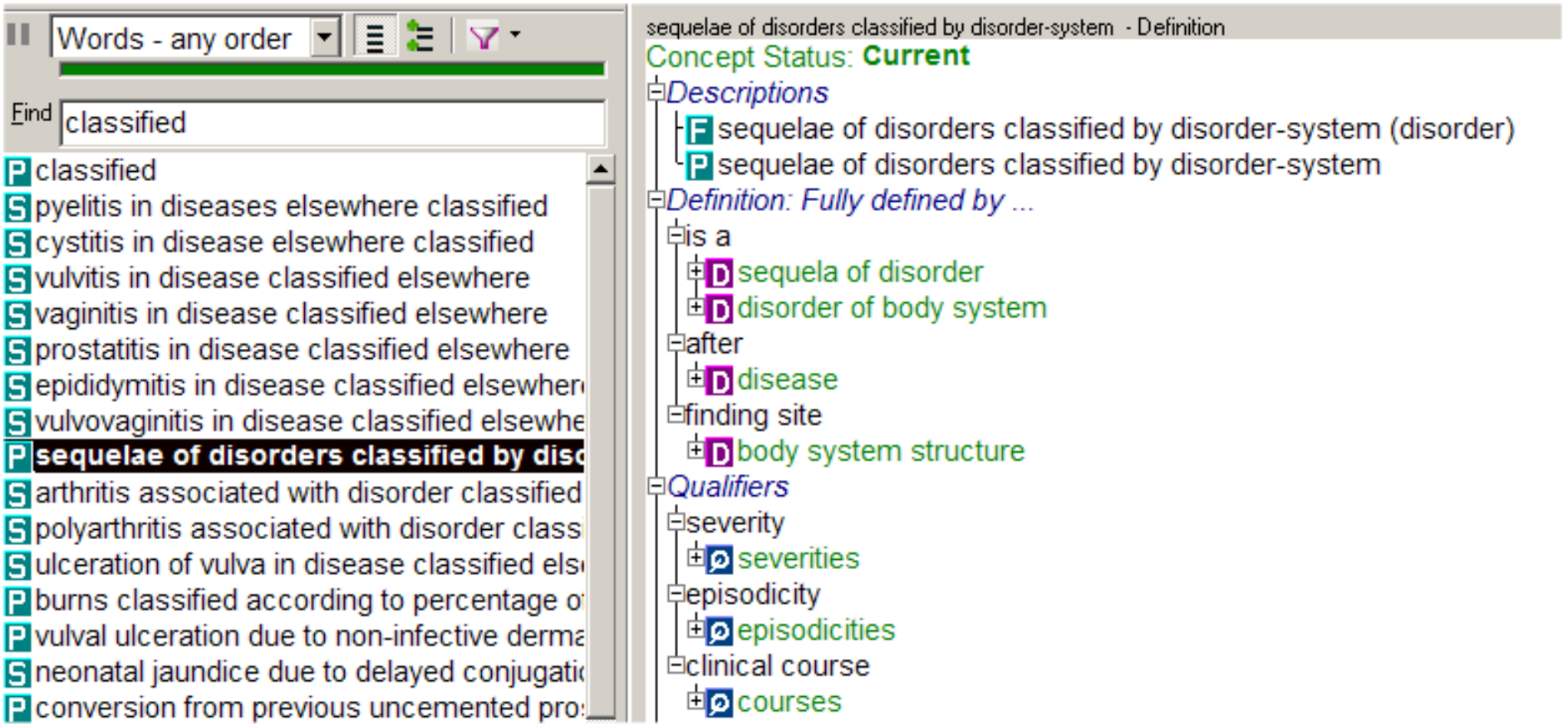


Classifications vs. Formal Ontologies

Classifications

Formal Ontologies

SNOMED CT: Classification aspects



SNOMED CT and Classifications

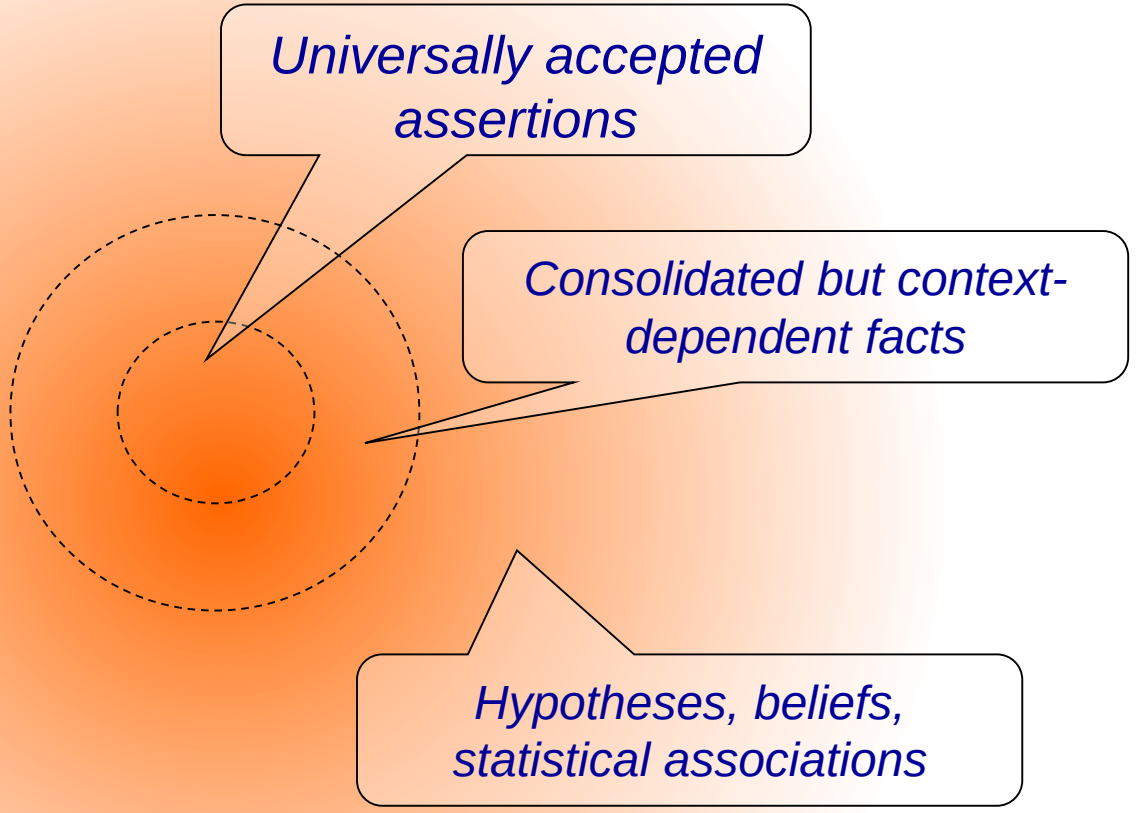
- Many classes in classification systems cannot be adequately expressed in SNOMED
- Problem:
  - SNOMED supports existential quantification and conjunction, but not negation
  - Classifications contain classes defined by negation:

Viral hepatitis (B15-B19)
Excludes: cytomegaloviral hepatitis (B25.1)
herpesviral [herpes simplex] hepatitis (B00.8)
sequelae of viral hepatitis (B94.2)

B17 Other acute viral hepatitis
B17.0 Acute delta-(super)infection of hepatitis B carrier
B17.1 Acute hepatitis C
B17.2 Acute hepatitis E
B17.8 Other specified acute viral hepatitis Hepatitis non-A non-B (acute)(viral) NEC

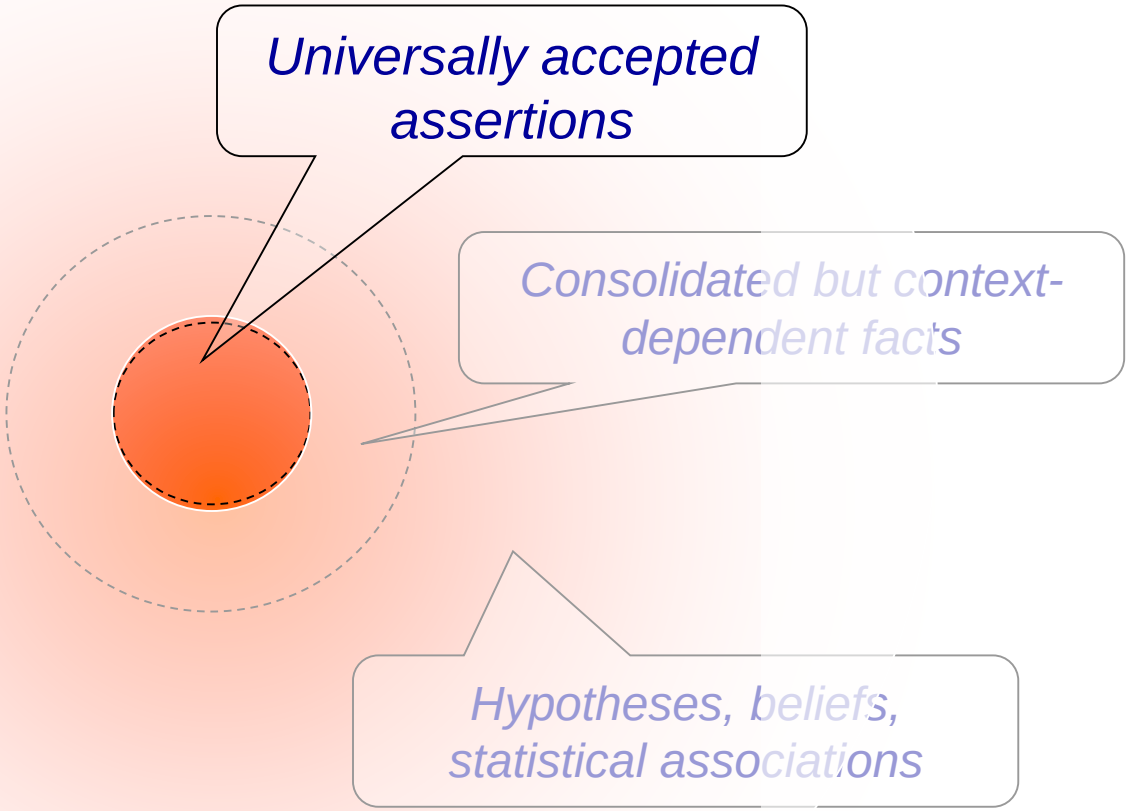
Knowledge Representation

Continuum of knowledge



Domain Knowledge

Formal Ontology !



Domain Knowledge

Instance-level Knowledge / Belief

- Working Hypothesis
  The patient was admitted with suspected appendicitis
- Unknown facts
  Allergies unknown
- Ruled-out facts
  No Pregnancy
  Absent corneal reflex
- Imprecise
  Patient reports “liver disease”
- Epistemic
  The diabetes was recently diagnosed
- Classification-related:
  Cause of death: A09 - Diarrhoea and gastroenteritis of presumed infectious origin
  Diagnosis: B37.8 - Candidiasis of other sites

Domain Knowledge

- Facts that are known to be true under certain circumstances:
  Excessive alcohol consumption can cause gout
- Context dependent facts:
  Hg_2Cl_2 is a diuretic drug
  Aspririn is an analgetic drug
- Facts about populations:
  Malaria is endemic in Mozambique
- Recommendations / Guidelines:
  Old patients with newly diagnosed Hypertension should be treated with diuretics or Ca channel blockers
- Basic scientific facts
  Many urokinase-type plasminogen activators are expressed in the kidney
- Results from clinical trials:
  One-lung overventilation does not induce inflammation in the normally ventilated contralateral lung.
- Default / canonic knowledge
  „Adult humans have 32 teeth“

Take home messages

- Ontologies describe classes of domain entities (ideally) by their inherent properties
- Classifications classify entities according to well-defined criteria
- Terminologies relate words and terms
- SNOMED CT is a hybrid terminology / ontology with elements of classifications
- Knowledge representation extends terminology / ontology by large
- (Computable) Ontologies are restricted to make universal statements of the type for all… some

Practice of Good Ontology

Practice of Good Ontology

Learning good ontology practice from bad ontologies…

Don’t mix up universals (Concepts, Classes) with individuals (Instances)

- *subclass-of (Motor Neuron, Neuron)* (FMA, OpenGALEN)
- *Is_a (Motor Neuron, Neuron)*
- *instance-of (Motor Neuron, Neuron)* (FlyBase)

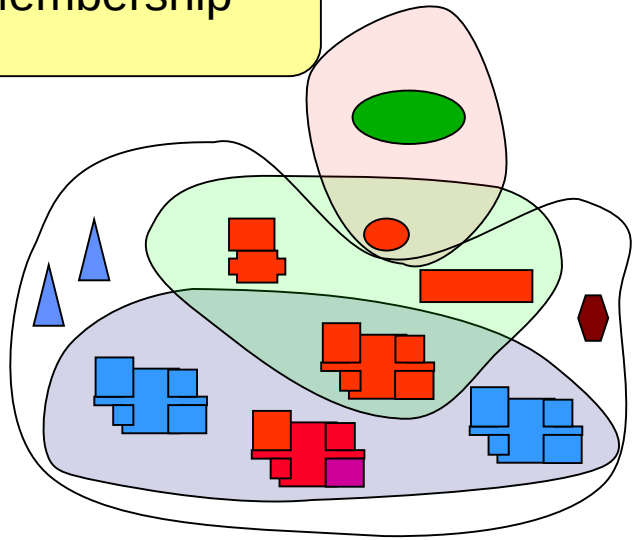
Is_a = subclass_of:

Taxonomic Subsumption

But:

- *instance-of (my Hand, Hand)*
- *instance-of (this amount of insulin, Insulin)*
- *instance-of (Germany, Country)*
- *not: instance of (Heart, Organ)*
- *not: instance of (Insulin, Protein)*

Instance_of Class Membership



Don’t use superclasses to express roles

- *Is_a (Fish, Animal)*
- *Is_a (Fish, Food)* ??

- *Is_a (Acetylsalicylic Acid, Salicylate)*
- *Is_a (Acetylsalicylic Acid, Analgetic Drug)* ??

Be aware of the “rigidity” of entity types

Partition the ontology by principled upper level categories

Example: DOLCE's Upper Ontology

Endurant (Continuant)
- ***Physical***
  - **Amount of matter**
  - **Physical object**
  - **Feature**
- ***Non-Physical***
  - **Mental object**
  - **Social object**
- **…**

Perdurant (Occurrent)
- ***Static***
  - **State**
  - **Process**
- ***Dynamic***
  - **Achievement**
  - **Accomplishment**

Quality
- ***Physical Qualities***
  - **Spatial location**
  - **…**
- ***Temporal Qualities***
  - **Temporal location**
  - **…**
- ***Abstract Qualities***
  - **…**

Abstract
- ***Quality region***
  - **Time region**
  - **Space region**
  - **Color region**

Source: S. Borgo ISTC-CNR

Limit to a parsimonious set of semantically precise Basic Relations

First version of the OBO Relation Ontology

Foundational relations

is_a

part_of

Spatial relations (connecting one entity to another in terms of relations between the spatial regions they occupy)

located_in

contained_in

adjacent_to

Temporal relations (connecting entities existing at different times)

transformation_of

derives_from

preceded_by

Participation relations (connecting processes to their bearers)

has_participant

has_agent

Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L Rector and Cornelius Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(5), 2005.

Avoid idiosyncratic categorization

- Body structure (10)
  - Acquired body structure
  - Anatomical organizational pattern
  - (…)
- Clinical finding (22)
  - Administrative statuses
  - Adverse incident outcome categories
  - (…)
- Environment or geographical location
  - Environment
  - Geogr. and/or political region of the world
- Event (19)
  - Abuse
  - Accidental event
  - Bioterrorism related event
  - (…)
- Linkage concept
  - Attribute
  - Link assertion
- Observable entity
  - Age AND/OR growth period
  - Body product observable
  - (…)
- Clin. history / examination observable (21)
  - Device observable
  - Drug therapy observable
  - Feature of Entity
  - (…)
- Organism (11)
  - Animal
  - Chromista
  - Infectious agent
  - (…)
- Pharmaceutical / biologic product (58)
  - Alcohol products
  - Alopecia preparation
  - Alternative medicines
  - (…)
- Physical force (21)
  - Altitude
  - Electricity
  - (…)
- Physical object (8)
  - Device
  - Domestic, office and garden artefact
  - Fastening
  - (…)
- Procedure (23)
  - Administrative procedure
  - Community health procedure
  - (…)
- Qualifier value (52)
  - Action
  - Additional dosage instructions
  - (…)
- Record artifact
  - Record organizer
  - Record type
- Situation with explicit context (17)
  - A/N risk factors
  - Critical incident factors
  - (…)
- Social context (10)
  - Community
  - Family
  - Group
  - (…)
- Special concept
  - Namespace concept
  - Navigational concept
  - Non-current concept
- Specimen (45)
  - Biopsy sample
  - Body substance sample
  - Cardiovascular sample
  - (…)
- Staging and scales (6)
  - Assessment scales
  - Endometriosis classification of American Fertility Society
  - (…)
- Substance (11)
  - Allergen class
  - Biological substance

The Celestial Emporium of Benevolent Knowledge

Jorge Luis Borges

"On those remote pages it is written that animals are divided into:

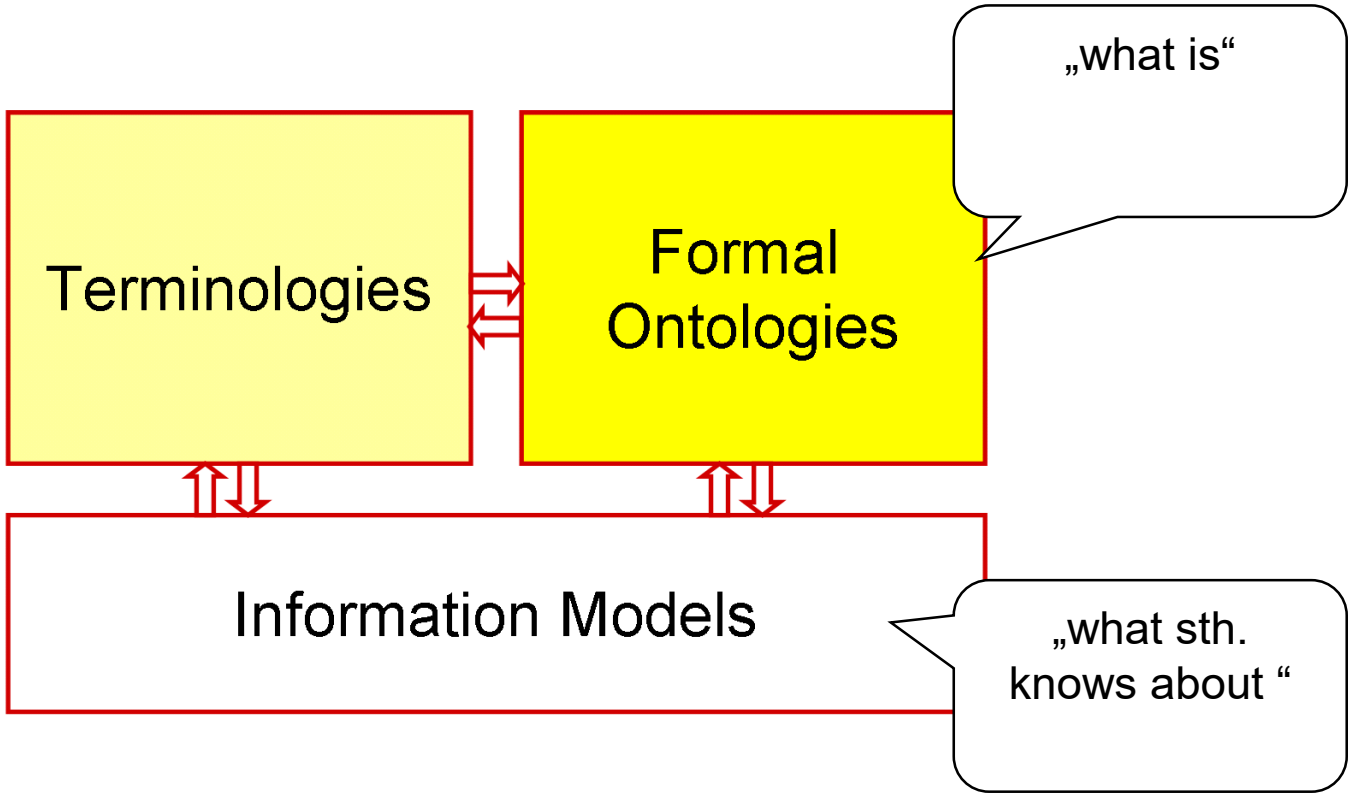
a. those that belong to the Emperor
b. embalmed ones
c. those that are trained
d. suckling pigs
e. mermaids
f. fabulous ones
g. stray dogs
h. those that are included in this classification
i. those that tremble as if they were mad
j. innumerable ones
k. those drawn with a very fine camel's hair brush
l. others
m. those that have just broken a flower vase
n. those that resemble flies from a distance"

Be aware of ambiguities

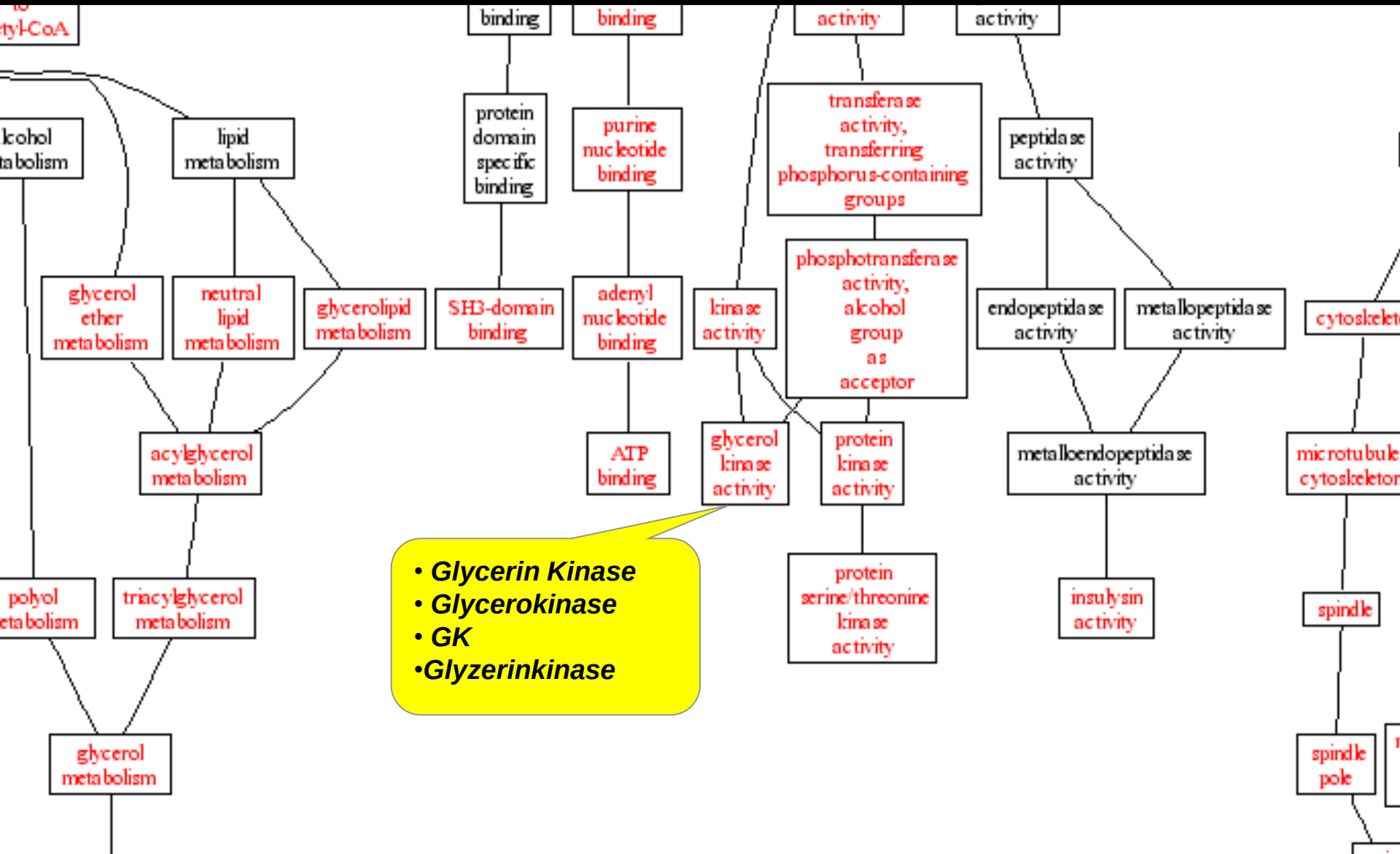
- "Institution" may refer to
  1. (abstract) institutional rules
  2. (concrete) things instituted
  3. act of instituting sth.
- "Tumor"
  1. evolution of a tumor as a disease process
  2. having a tumor as a pathological state
  3. tumor as a physical object
- "Gene"
  1. a (physical) sequence of nucleotides on a DNA chain
  2. a collection of (1)
  3. A piece of information conveyed by (1)

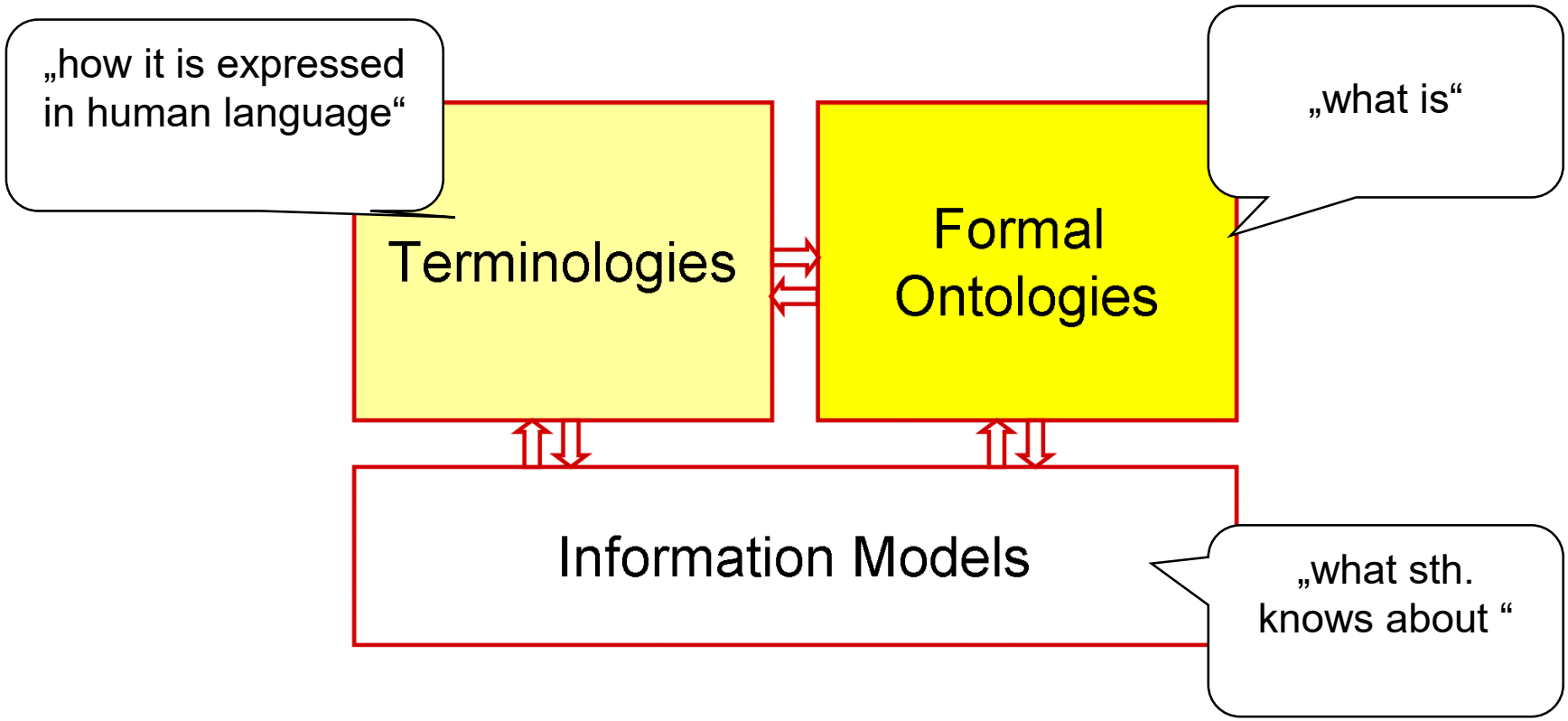
Don‘t mix up ontology with epistemiology

- *Is_a (Infection of unknown origin, Infection)*
- *Is_a (Newly diagnosed diabetes, Diabetes)*
- *Is_a (Family history of diabetes, Diabetes)*



Don‘t mix up Ontology IDs with Terms




„how it is expressed in human language“
Terminologies
Formal Ontologies
„what is“
Information Models
„what sth. knows about “

Don’t underestimate Ontology Maintenance

- Formal Ontologies must always be maintained
  - consistent (free of logic contradiction): prerequisite for machine reasoning
  - adequate (correctly describe the domain) prerequisite to prevent erroneous deductions
- Maintenance load is much higher than with terminologies.
- Ontology maintenance is mainly task of domain experts. IT staff has supportive function
- Typical design and maintenance errors

Aspects of Knowledge Representation

- Terminological Knowledge: What is known about the meaning of terms in a domain
  "neoplasm" has a broader meaning as "sarcoma"
- Ontological "Knowledge": What is univocally accepted as generic properties of types of entities of a domain (often definitional or trivial):
  every hepatitis is located in some liver
  every cell has some cell membrane
- Terminologies and Ontologies represent this kind of Knowledge, but…
- Knowledge representation is more: