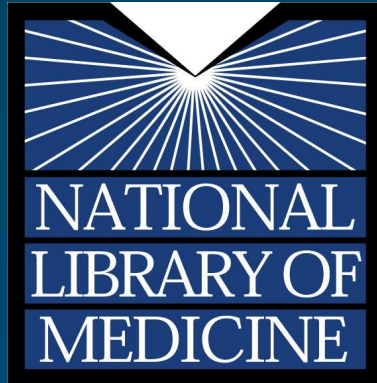# Ontological Spring
Naumburg, Germany - April 17-20, 2002

Examples of ontologies
in the biomedical domain

Anita Burgun
Medical School / Univ. Hospital
Rennes, France

Olivier Bodenreider
National Library of Medicine
Bethesda, Maryland - USA

NATIONAL LIBRARY OF MEDICINE

HTTP://WWW.MED.UNIV-RENNES1.FR

UNIVERSITE DE RENNES 1   FACULTÉ DE MÉDECINE

Introduction

- Broadness of the domain
  - 800,000 concepts in the UMLS
- No complete theory of the domain
  - Empirical knowledge
    - Signs or symptoms
  - New knowledge
    - Molecular Biology

Overview of our presentation

- Illustrate from examples rather than report on all aspects
- Some existing "ontologies"
  - How the biomedical domain is represented in existing "ontologies" (medical or not)
- Compatibility among ontologies
- Discussion inspired by the representation of the biomedical domain in several systems ("Blood")

Existing ontologies

Examples in the biomedical domain

Knowledge oriented: General ontologies

- Concepts are prototypes of exemplars
- Concepts are models for types of individuals
- Cyc
- #$HumanChild
  - (#$isa #$HumanChild #$ExistingObjectType)
  - (#$genls #$HumanChild (#$JuvenileFn #$Person))
- #$Infection
  - (#$isa #$Infection #$PhysiologicalConditionType)
  - (#$genls #$Infection #$AilmentCondition)
- #$InfectionFn
  - (#$isa #$InfectionFn #$CollectionDenotingFunction)
  - (#$resultIsa #$InfectionFn #$InfectionType)
  - (#$resultGenl #$InfectionFn #$Infection)
  - (#$arg1Isa #$InfectionFn #$ExistingObjectType)
  - (#$arg1Genl #$InfectionFn #$AnimalBodyPart)

5

Domain ontologies

- GALEN

- E.U. project

- U. Manchester

- compositional models of medical concepts with formal properties

- language independent concept systems which are interpreted through separate grammars and lexicons

Domain ontologies

◆ GALEN

◆ Model for medical procedures

◆ SurgicalDeed which

isCharacterisedBy (performance whichG

  isEnactmentOf ((Excising which playsClinicalRole SurgicalRole)
 whichG <

      actsSpecificallyOn (NeoplasticLesion whichG

        hasSpecificLocation AdrenalGland)

    hasSpecificSubprocess (SurgicalApproaching whichG

      hasSurgicalOpenClosedness (SurgicalOpenClosedness
        whichG hasAbsoluteState surgicallyOpen))>))

◆ Surgically open extracting of an adrenal gland neoplastic lesion

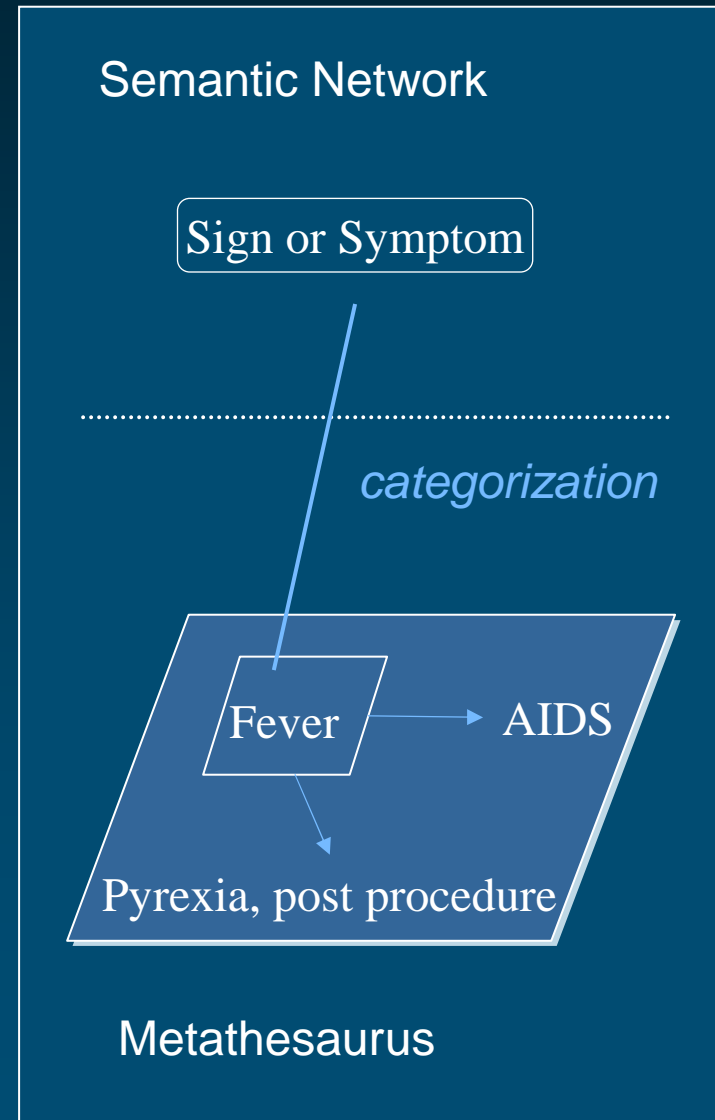# Language oriented : General ontologies

◆ Concepts are clusters of terms
◆ Meanings of the words are concepts
◆ WordNet
  ● U. Princeton
  ● 100,000 synsets
◆ Hyponymy: A concept represented by the synset {x,x', …} is said to be a hyponym of the concept represented by the synset {y,y',…} if native speakers of English accept sentences constructed from such frames as « An x is a kind of y »
◆ Fever is a Symptom > Evidence > Information > Cognition/Knowledge > Psychological feature

Domain ontologies

- Unified Medical Language System (US. National Library of Medicine)
- 50 families of vocabularies
- 800,000 concepts of the biomedical domain, 134 semantic types
- Concepts are clusters of terms
- 10 M interconcept relationships inherited from the source vocabularies.
- Hierarchical relation: A concept represented by the cluster{x,x', …}is said to be a child of the concept represented by the cluster {y,y',…} if any of the source terminologies shows a hierarchical relationship between x and y.

Knowledge organization in the UMLS

- ◆ UMLS knowledge sources
  - ● Metathesaurus
  - ● Semantic Network
  - ● Lexical resources
    - ■ SPECIALIST Lexicon
    - ■ Lexical tools
- ◆ Two-level structure

Semantic Network

Sign or Symptom

categorization

Fever → AIDS

Pyrexia, post procedure

Metathesaurus

New generation coding systems

- SNOP Systematized Nomenclature of Pathology
  - 1965 College of American Pathologists
  - Multiaxial
  - TOPO   MORPHO   ETIO   FUNCTION
  - Lung   inflammation staph   fever
- SNOMED Systematized Nom. Of Medicine (79)
  - DISEASE   PROCEDURE   OCCUPATION
  - Cross references
  - Add relationships between terms

New generation coding systems

- ◆ SNOMED-RT
  - ● Reference terminology
  - ● Electronic Patient Record - Interoperability
  - ● « a common reference point for comparison and aggregation of data throughout the entire healthcare process »
  - ● Multiaxial 10 axes
  - ● 121,000 concepts,   340,000 relationships
  - ● SNOMED-CT (CAP - UK NHS) combines SNOMED-RT and Clinical Terms (Read codes)

New generation coding systems

- Fully Specified Name: Nephrectomy (procedure)
- Concept ID: 85250002   SNOMED ID: P1-71340
- Definition:

  Is a Kidney excision (procedure)

  Associated topography   Kidney (body structure)

  Has action Excision - action

- Parent(s): Kidney excision (procedure)
- Child(ren) (N=8)
  - Bilateral nephrectomy
  - Donor nephrectomy (procedure)
  - Heminephrectomy (procedure)
  - Nephroureterocystectomy

« Specialized » ontologies

- Digital Anatomist
- GeneOntology
  - currently acts as a controlled vocabulary for annotating the genes as well as an ontology
- Application ontologies, e.g., MENELAS

Compatibility among ontologies

The example of UMLS and WordNet

Compatibility among ontologies

Compatibility in depth:
lower levels of ULO and
upper domain categories
(e.g., Disease)

Upper Level
Ontology
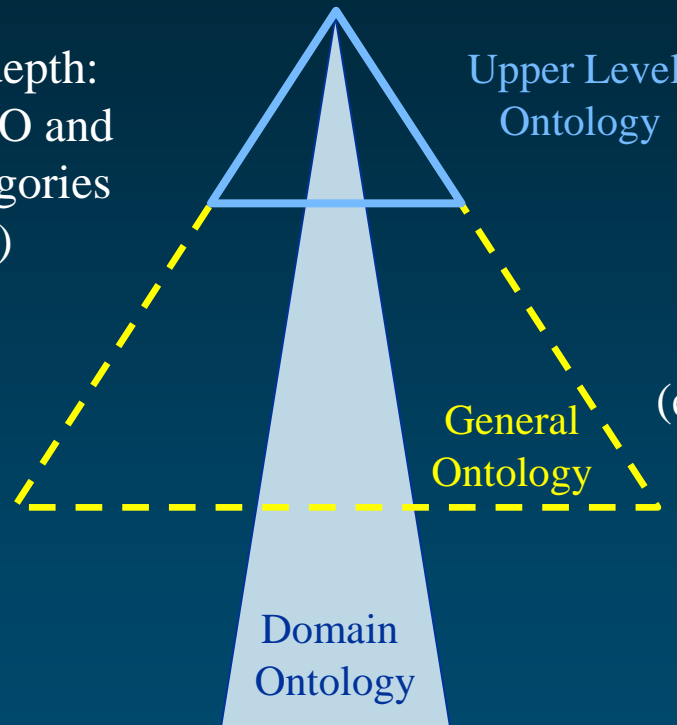
Compatibility in breadth:
Categories that do not
specifically belong to D
(e.g., Manufactured Object)

General
Ontology

Domain
Ontology

Universal Compatibility :
Generic theories, e.g., time, space
Meta-level categories, e.g., properties, roles

Compatibility UMLS/ WordNet

Upper Level Ontology

General Ontology

Domain Ontology

Animals
Diseases

17
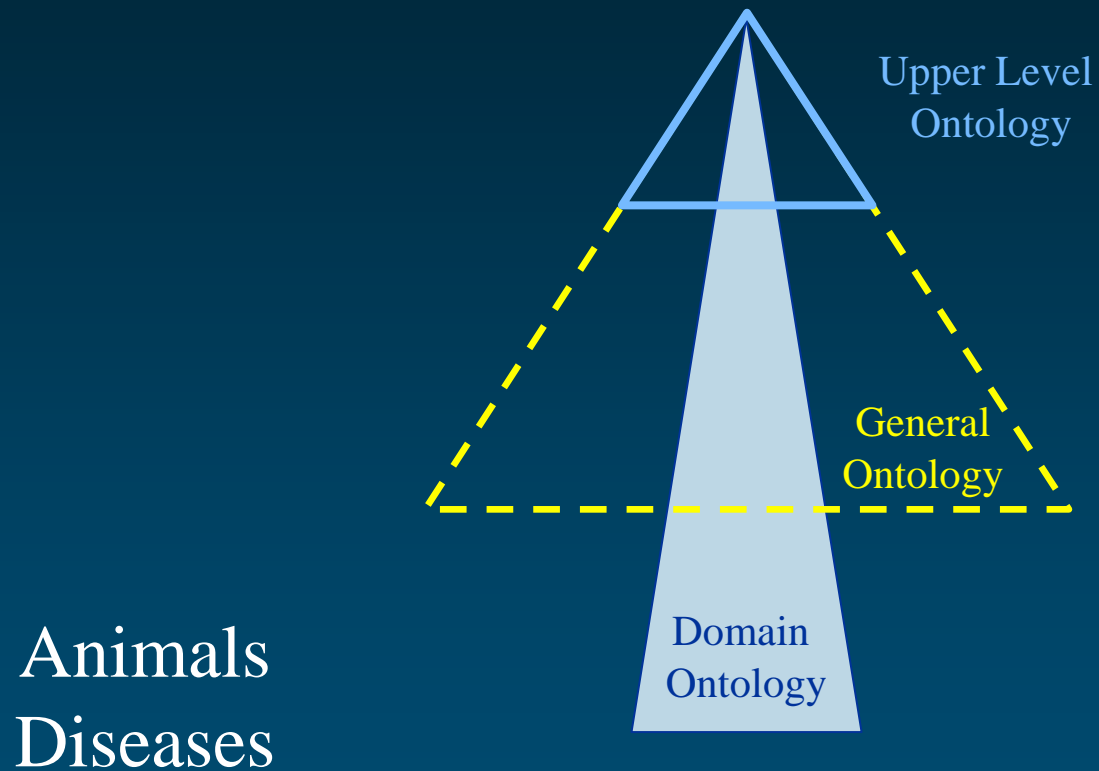
UMLS (2000)

- Domain
  - Biomedicine
- Terms
  - Lexical variants
  - Multiple languages
- Concepts
  - Clusters of synonymous terms
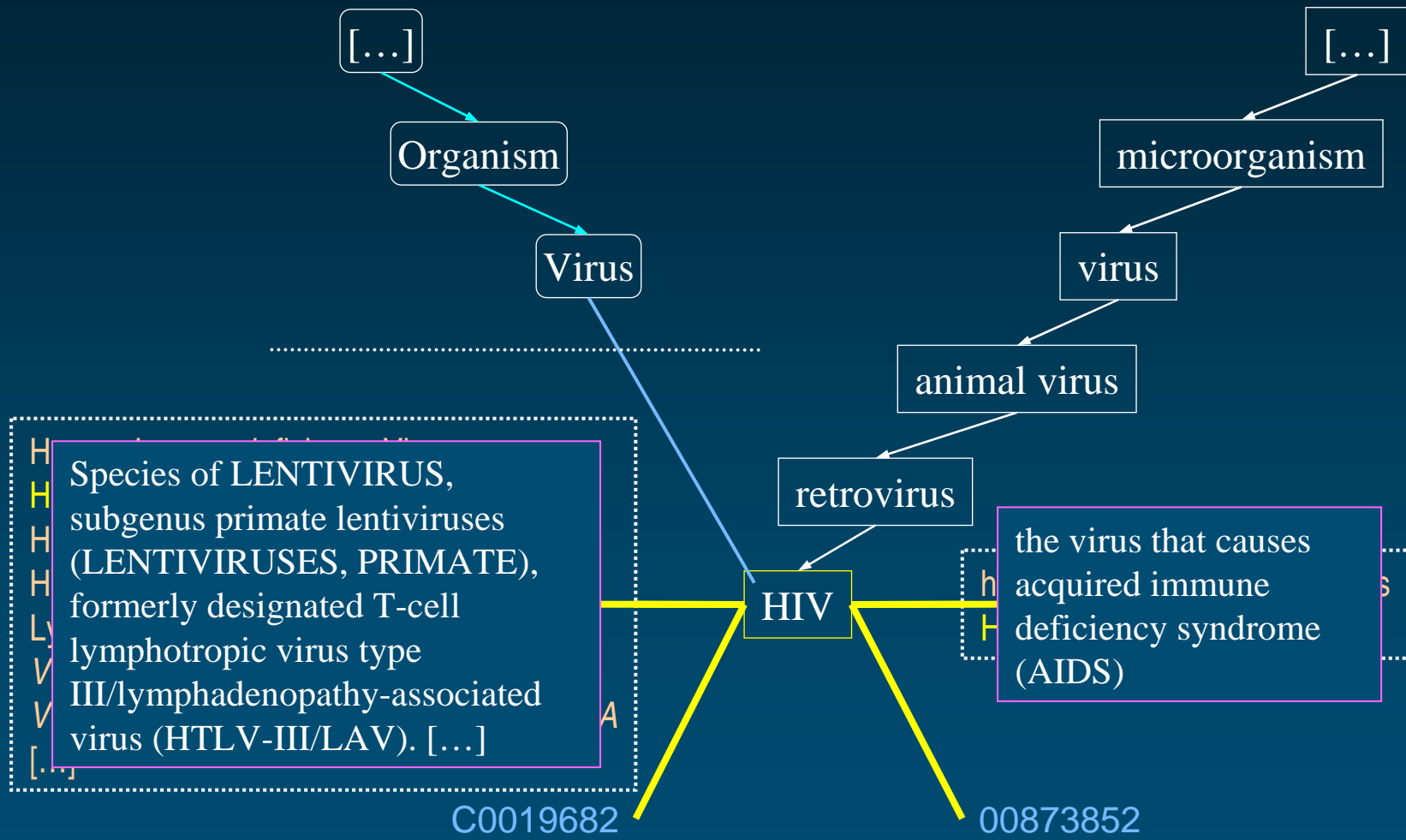 - Definition(s)
- Semantic classes
  - Categorization

WordNet (1.6)

- Domain
  - General world
- Terms
  - Canonical form only
  - English terms only
- Synsets
  - Clusters of synonymous terms
  - Definition
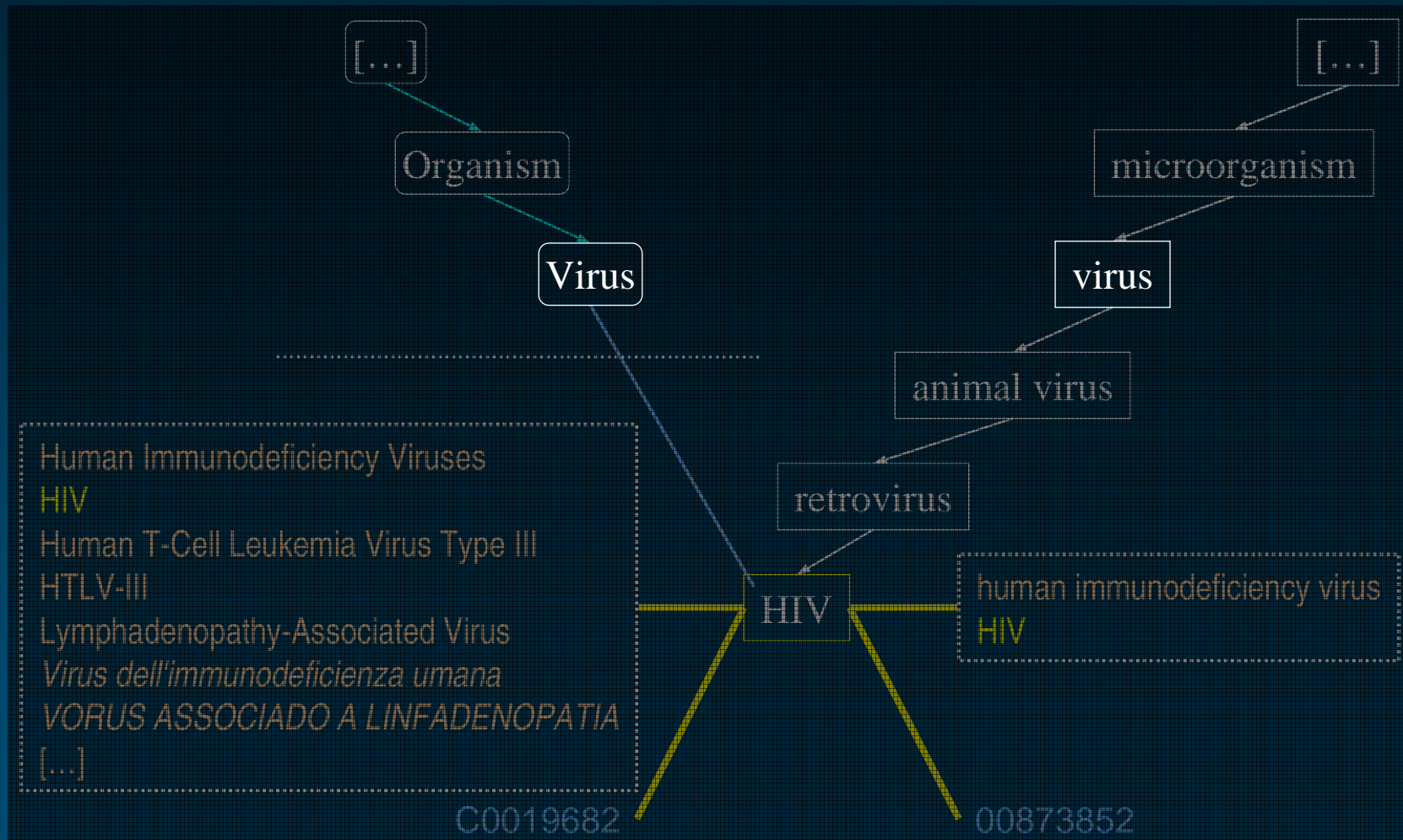- Semantic classes
  - Hyponymy

UMLS   Example   WordNet

[…]

Organism

Virus

[…]

microorganism

virus

animal virus

retrovirus

HIV

Species of LENTIVIRUS, subgenus primate lentiviruses (LENTIVIRUSES, PRIMATE), formerly designated T-cell lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV). […]

the virus that causes acquired immune deficiency syndrome (AIDS)

C0019682

00873852

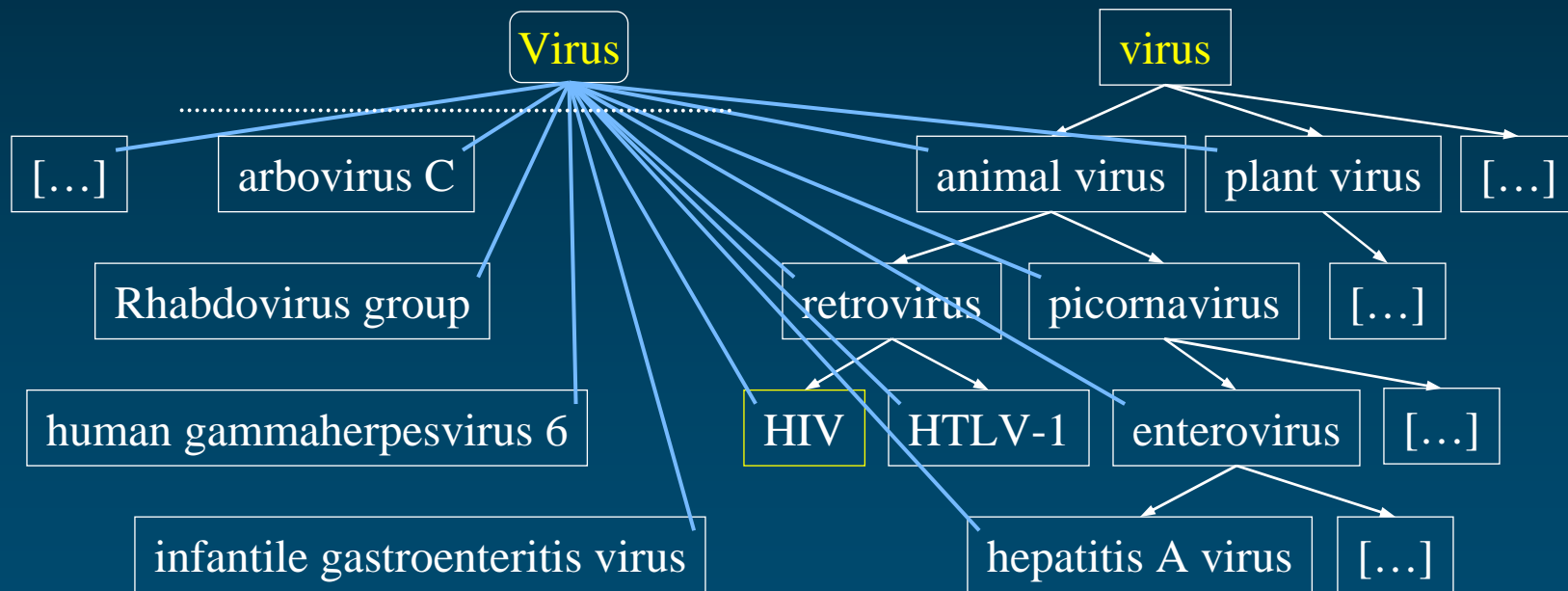# UMLS   Semantic class   WordNet

2 semantic classes

◆ **Animal**

- General class
- UMLS: concepts assigned to the semantic type Animal (or any of its subtypes)
- WordNet: synset Animal and all its hyponyms

◆ **Health Disorder**

- Medical domain
- UMLS: concepts assigned to these semantic types
- WordNet: these synsets and all their hyponyms

- *Anatomical Abnormality*
- *Congenital Abnormality*
- *Acquired Abnormality*
- *Finding*
- *Sign or Symptom*
- *Pathologic Function*

- *Disease or Syndrome*
- *Mental or Behavioral Dysfunction*
- *Neoplastic Process*
- *Cell or Molecular Dysfunction*
- *Experimental Model of Disease*
- *Injury or Poisoning*

- *Symptom*
- *Ill Health*
- *Disorder (sense 1)*
- *Mental retardation*
- *Mental Illness*
- *Defect (sense 1)*
- *Abnormalcy*

Semantic classes

	Animal	Health Disorder
UMLS concepts	11,634	143,991
WordNet synsets	3,984	1,379

Mapping WordNet terms to the UMLS

- ◆ **What?**
  - ● Each term of each synset for a given class
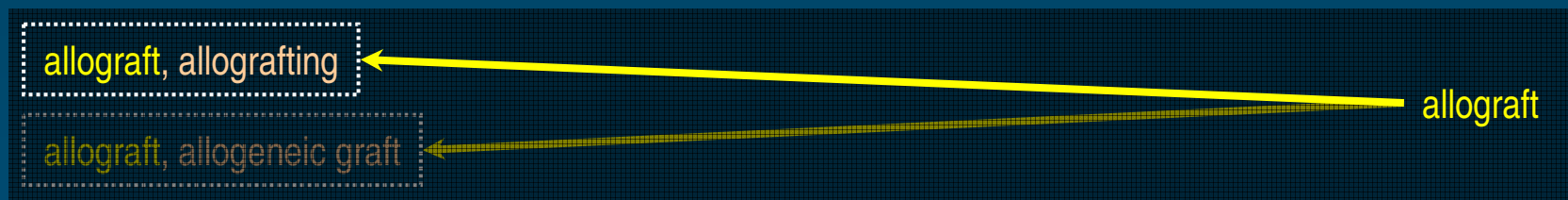- ◆ **How?**
  - ● **ks** function (Knowledge Source Server)
    - ■ Exact match
    - ■ After normalization
- ◆ **Disambiguation issues**

UMLS   WordNet

allograft, allografting

allograft, allogeneic graft   allograft

Comparison 3 levels

- ◆ Terms
  - ● Does the term T from S1 also belong to S2?
- ◆ Concepts
  - ● How do terms for concept C in S1 overlap with terms for concept C' in S2?
- ◆ Semantic classes
  - ● How do concepts for class K in S1 overlap with concepts for class K' in S2?

Results **Animal**

	From WordNet	Found in UMLS
Synsets	3,984	51%
Terms	7,961	36%

Same class: 94%

	From UMLS	Found in WordNet
Concepts	11,634	19%

Same class: 73%

Results **Health Disorder**

	From WordNet	Found in UMLS
Synsets	1,379	83%
Terms	2,194	77%

Same class: 97%

	From UMLS	Found in WordNet
Concepts	143,991	2%

Same class: 48%

Specific terms

◆ **UMLS**

- Specialized terms
- Terminology-specific terms

◆ **WordNet**

- Lay synonyms

UMLS

WordNet

Infectious Mononucleosis
Glandular Fever
Pfeiffer's disease
MONONUCLEOSIS
Monocytic angina
Gammaherpesviral mononucleosis
Infectious mononucleosis, unspecified
Infective mononucleosis
[…]

Infectious
Mononucleosis

infectious mononucleosis
glandular fever
kissing disease

Specific concepts

- UMLS
  - **Animal**
    - Angiostrongylus
    - Angiostrongylus cantonensis (Rodent lungworm)
    - Acanthamoeba
  - **Health disorder**
    - Many domain-specific concepts

- WordNet
  - **Animal**
    - Kitty
    - Unicorn
    - Cotton ballworm
    - Mickey Mouse
  - **Health disorder**
    - Plant diseases
    - Astraphobia
    - Crick
    - Sword cut

Granularity, plesionymy

UMLS   WordNet

Epilepsy, Generalized
Seizure Disorder, Generalized
[…]

generalized epilepsy
grand mal epilepsy

Epilepsy, Grand Mal
Tonic-Clonic Epilepsy
Seizure Disorder, Tonic Clonic
[…]

Differing categorization

UMLS   WordNet

| Natural Phenomenon or Process | | phenomenon |

| Biologic Function | | process |

| Pathologic Function | | natural process |

Health disorder ············ | Disease or Syndrome | | decay | **Health disorder**

Dental Caries
Dental cavity, NOS
Tooth caries
Dental Decay
[…]

| Dental Caries |

cavity
caries
dental caries
tooth decay

31
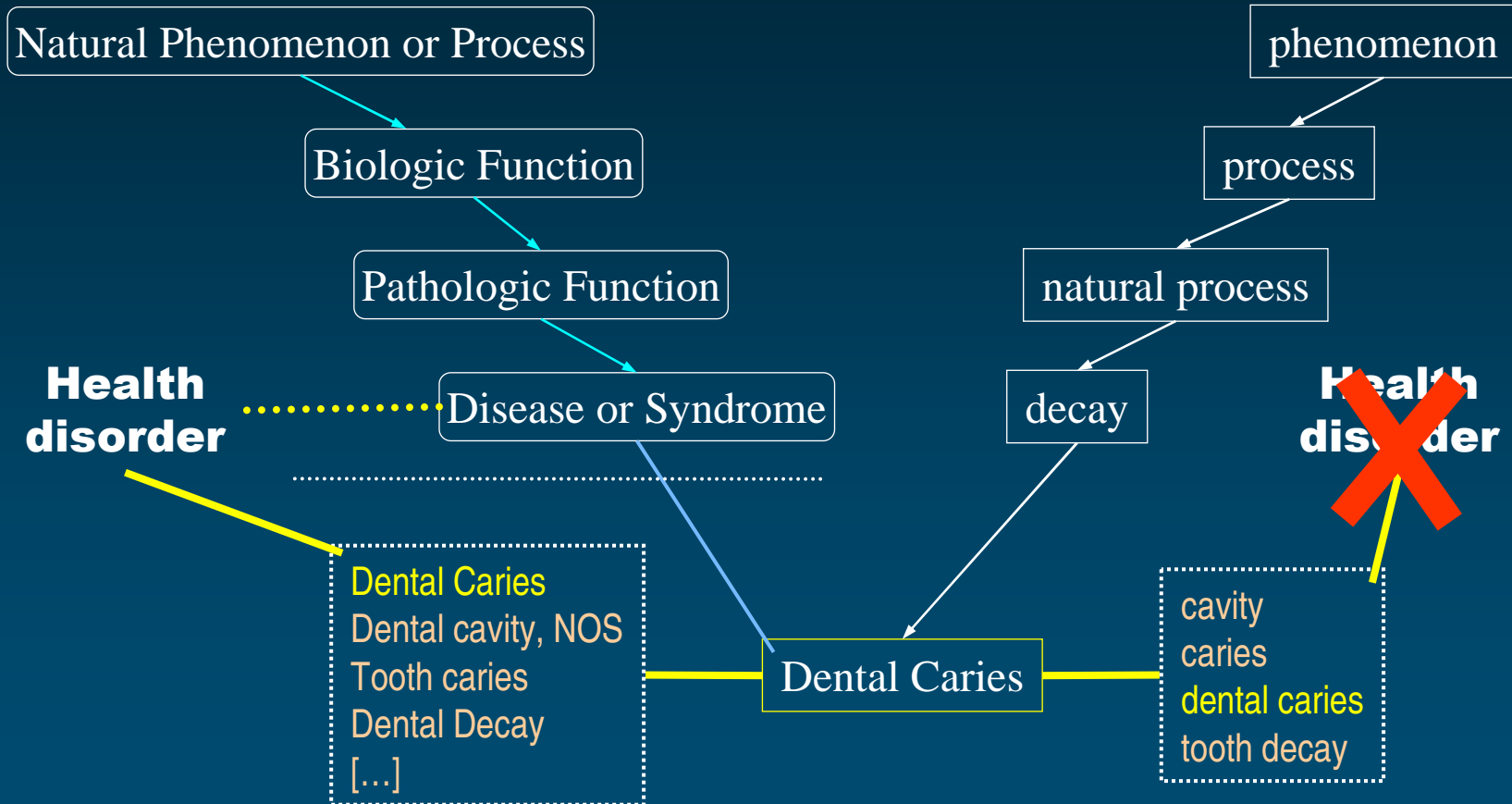
Representation of the biomedical domain
In different systems
Blood

Representation of Blood

- **In general ontologies**
  - Cyc Knowledge Representation, common-sense
  - WordNet, NLP oriented
- **In domain ontologies**
  - GALEN
  - UMLS
  - SNOMED RT
- **In a specific ontology : Digital Anatomist**
- **In application ontologies : MENELAS**

Representation of Blood in Cyc

ExistingStuffType

TangibleThing

#$isa

#$genls
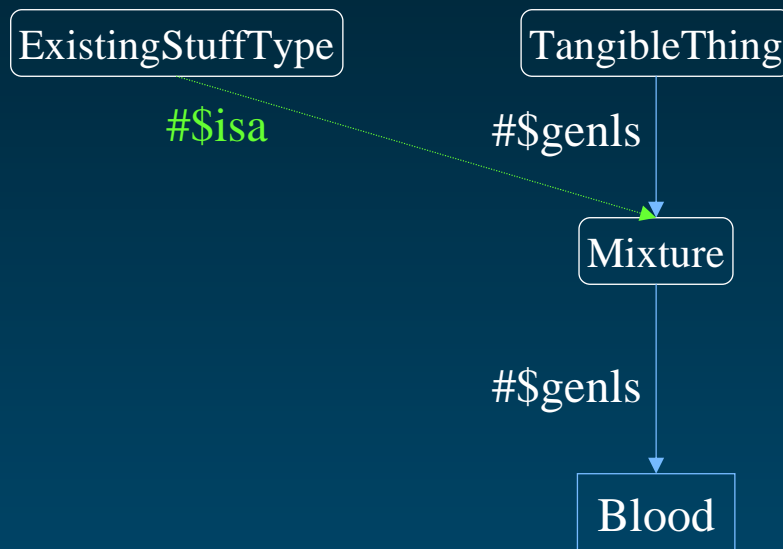
Mixture

A tangible stuff composed of two or more different constituents which have been mixed. These constituents do not form chemical bonds, and later the mixture may be resolved by some separation event. A mixture has a composition but not a structure

#$genls

Blood

As well as mud, air and carbonate beverage

The function Separation-Event can apply to it.

Representation of Blood in WordNet

Entity
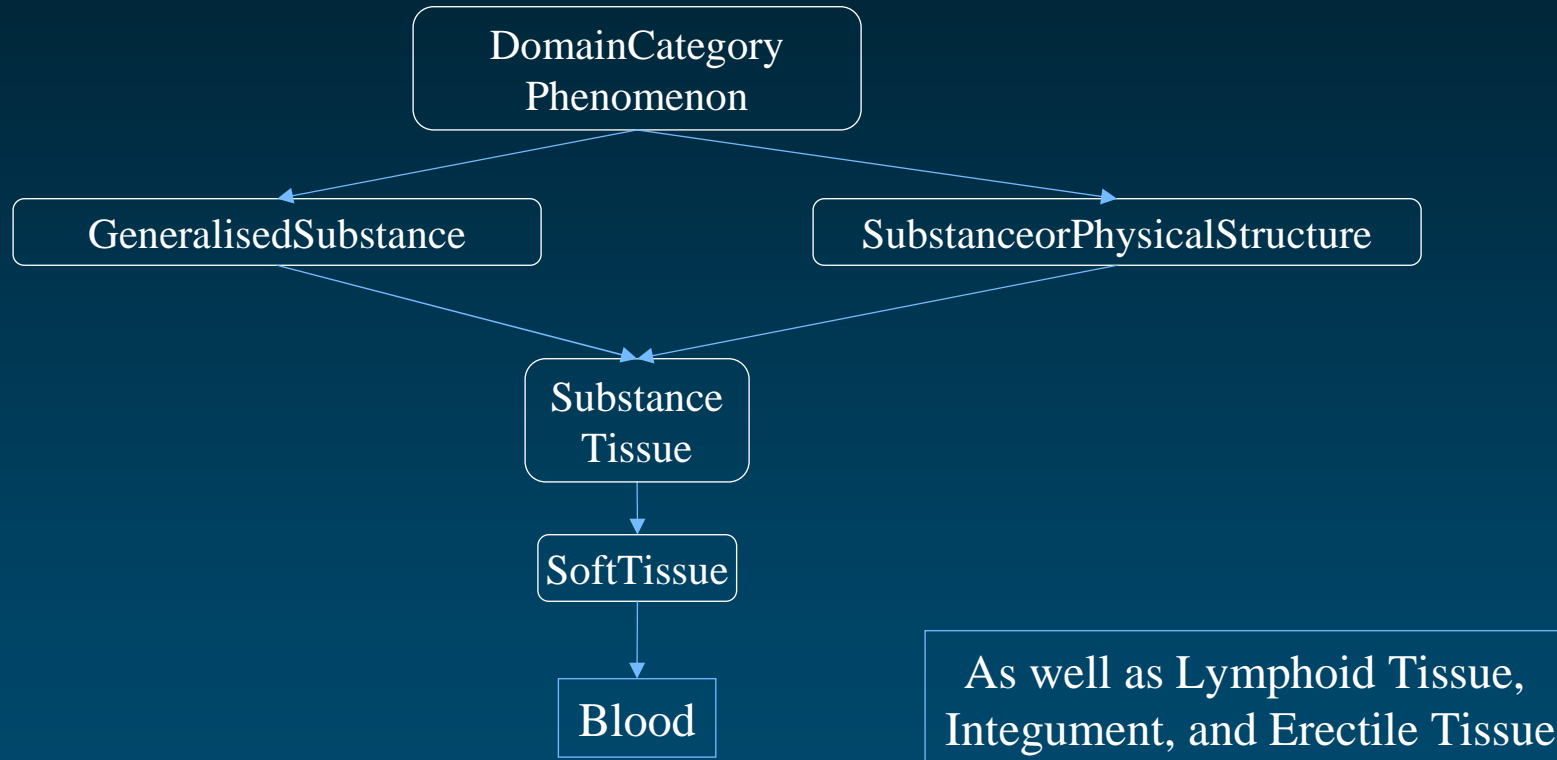Physical Object
Substance
Body Substance
Body Fluid

Humor

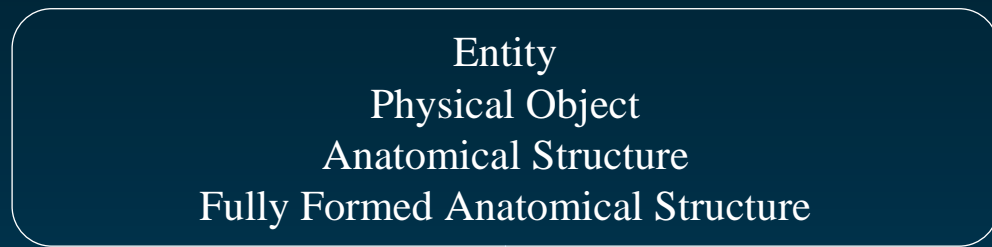*the four fluids in the body whose balance was believed to determine our emotional and physical state*

Blood

As well as phlegm, yellow and black bile

35

Representation of Blood in GALEN

Blood has two states, LiquidBlood and CoagulatedBlood

Representation of Blood in the UMLS

Entity
Physical Object
Anatomical Structure
Fully Formed Anatomical Structure

Tissue

An aggregation of similarly specialized cells and the associated intercellular substance. Tissues are relatively non-localized in comparison to body parts, organs or organ components
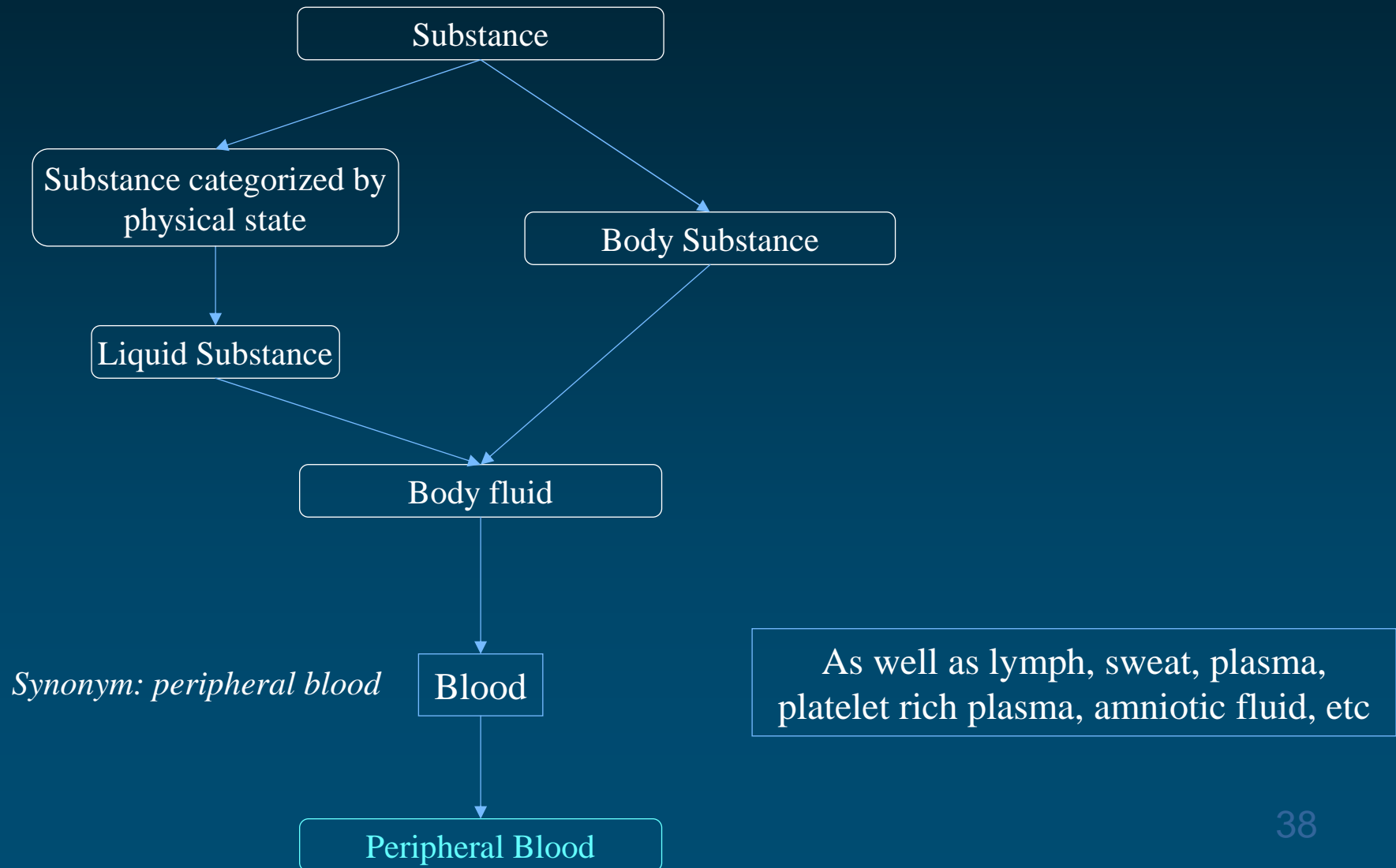
Body Fluid   Soft Tissue   Body Substance

Blood

37

Representation of Blood in SNOMED

Representation of Blood in D. Anatomist
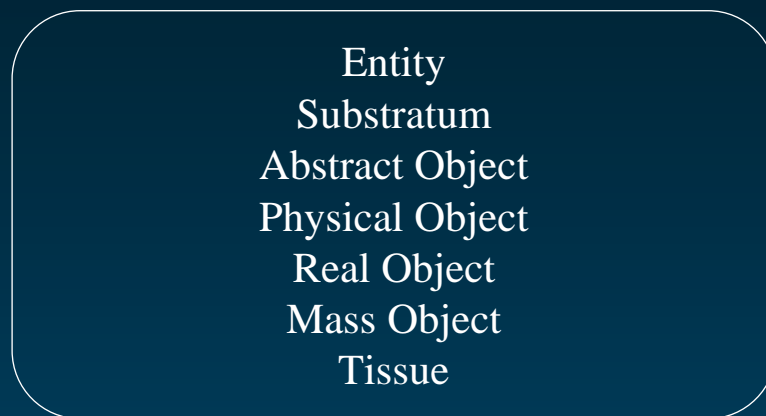
Anatomical Entity
Physical Anatomical Entity

a physical anatomical entity and a substance in gaseous, liquid, semisolid or solid state, with or without the admixture of cells, which is produced by anatomical structures or derived from inhaled and ingested substances that become modified by anatomical structures as they pass into or through the body

Body Substance

As well as saliva, semen, growth hormone, inhaled air, feces, lymph

Blood

Tissue is an Organ Part.

39

Representation of Blood in MENELAS

Entity
Substratum
Abstract Object
Physical Object
Real Object
Mass Object
Tissue

↓

Body Fluid

↓

Blood

Model body_fluid(_x) is
[body_fluid: _x]--(attr)-->[viscosity]

As well as Lymph

Mass Objects are constituted of Countable Objects

40

From an example to discussion about…

- ◆ Knowledge and representation of knowledge
  - ● Within the biomedical domain (core concepts)
    - ■ Tissue in UWDA ("Tissue is an organ part…"), and the SN ("…Tissues are relatively non-localized in comparison to body parts, organs or organ components")
  - ● Expert knowledge vs general
    - ■ Humors as microtheories
  - ● Upper level categories
    - ■ Mixtures in Cyc, Mass objects (non countable) in MENELAS
  - ● Level of Knowledge to be represented in a DO
    - ■ Coagulated Blood, Liquid Blood in GALEN