



The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9-11, 2008

Biomedical Ontology in Practice



Olivier Bodenreider

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

Objectives

- ◆ Learn about biomedical ontologies
 - History
 - Design principles, formalisms and tools
 - What are they?
 - What are they used for?
- ◆ Work with biomedical ontologies
 - Search
 - Analyze
 - Extend
 - Use for data integration



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration



References Bio-ontology courses

- ◆ Barry Smith, U. Buffalo / NCBO
 - http://ontology.buffalo.edu/smith/Ontology_Course.html
- ◆ Stefan Schulz, U. Freiburg, Germany / KR-MED
2008 tutorial
 - <http://www.kr-med.org/2008/index.html>



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9, 2008 – Session #1

Introduction to Biomedical Ontologies



Olivier Bodenreider

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

Outline

- ◆ Historical perspective
- ◆ Introduction to biomedical terminologies through an example
- ◆ Biomedical terms as names for biomedical classes
- ◆ Terminological relations as a surrogate for ontological relations



Historical perspective

Why biomedical terminologies?

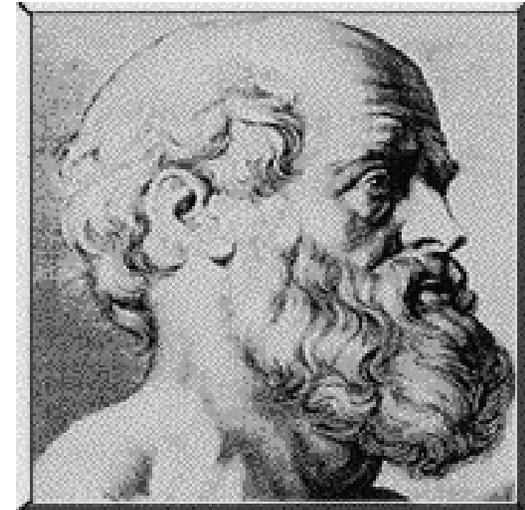
- ◆ To support a theory of diseases
- ◆ To classify diseases
- ◆ To support epidemiology
- ◆ To index and retrieve information
- ◆ To serve as a reference



To support a theory of diseases

◆ Hippocrates

- Dismisses superstition
- Four humors
 - Blood
 - Phlegm
 - Yellow bile
 - Black bile



◆ Thomas Sydenham (1624-1689)

- *Medical observations on the history and cure of acute diseases* (1676)



Figure 36 Thomas Sydenham (1624-1689)



To classify diseases (and plants)

◆ Carolus Linnaeus (1707-1778)

- *Genera Plantarum* (1737)
- *Genera Morborum* (1763)



◆ François Boissier de La Croix a.k.a. F. B. de Sauvages (1706-1767)

- *Methodus Foliorum* (1751)
- *Nosologia Methodica* (1763/68)

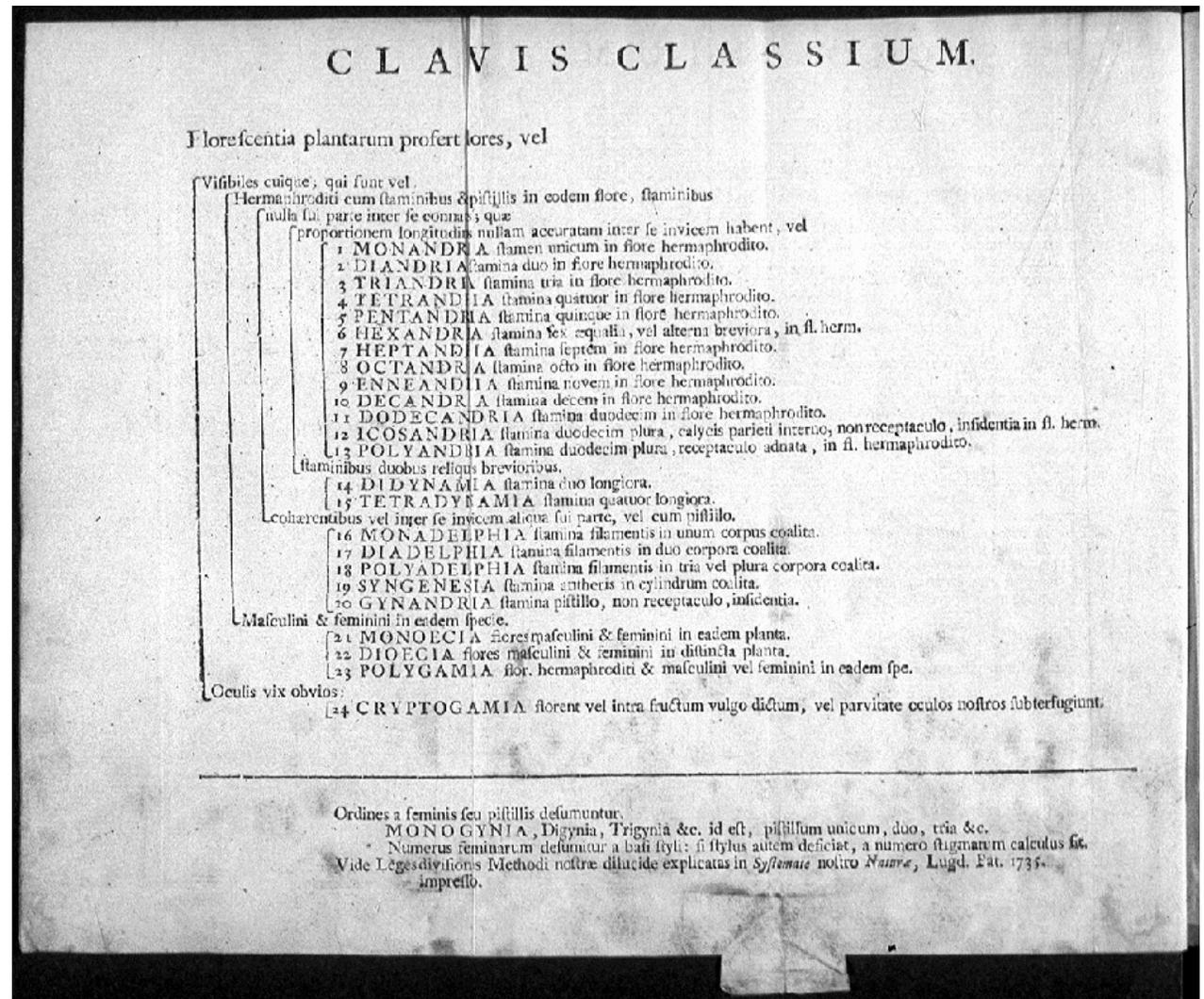
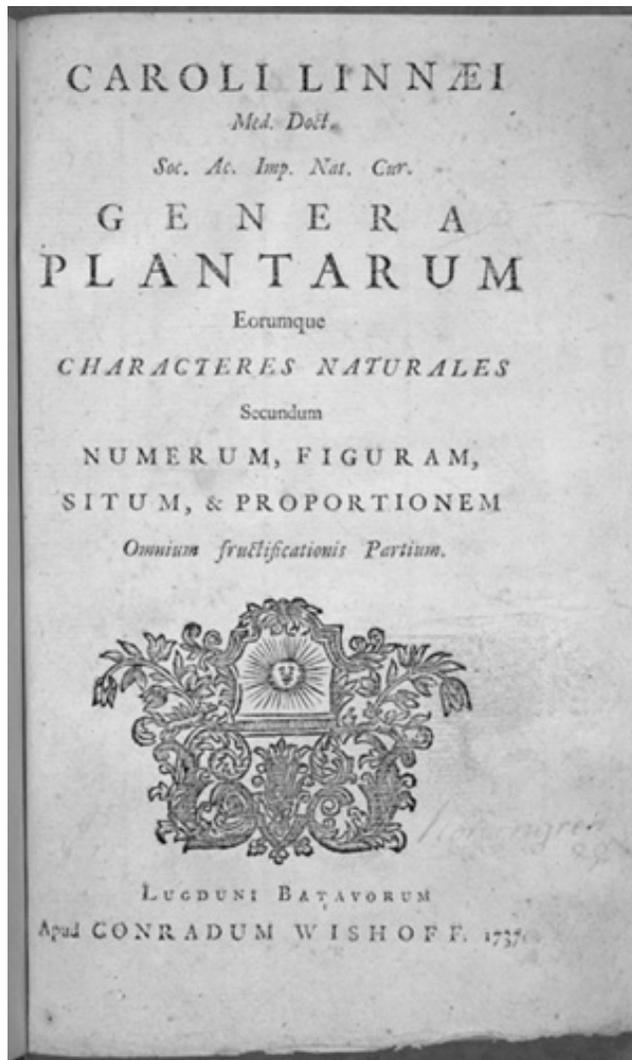


◆ William Cullen (1710-1790)

- *Synopsis Nosologiae Methodicae* (1785)



From plants...



... to diseases

◆ Four categories (W. Cullen)

- Fevers
- Nervous disorders
- Cachexias
- Local diseases

“The distinction of the genera of diseases, the distinction of the species of each, and often even that of the varieties, I hold to be a necessary foundation of every plan of physic, whether dogmatical or empirical.”

– William Cullen, Edinburgh, 1785

Synopsis Nosologia Methodicae

(Cited by Chris Chute)



To support epidemiology

◆ John Graunt (1620-1674)

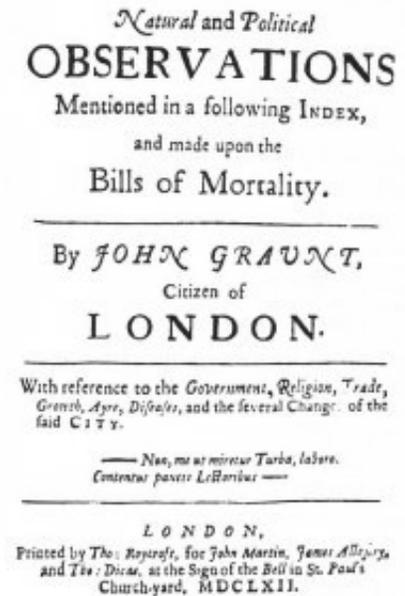
- Analyzes the vital statistics of the citizens of London

◆ William Farr (1807-1883)

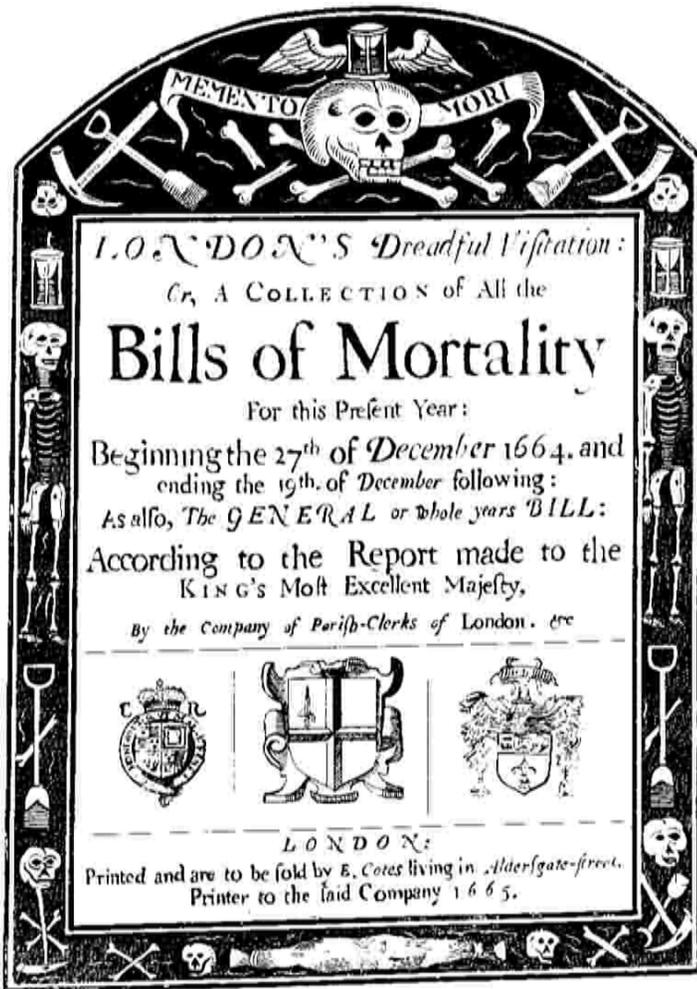
- Medical statistician
- Improves Cullen's classification
- Contributes to creating ICD

◆ Jacques Berthillon (1851-1922)

- Chief of the statistical services (Paris)
- Classification of causes of death (161 rubrics)

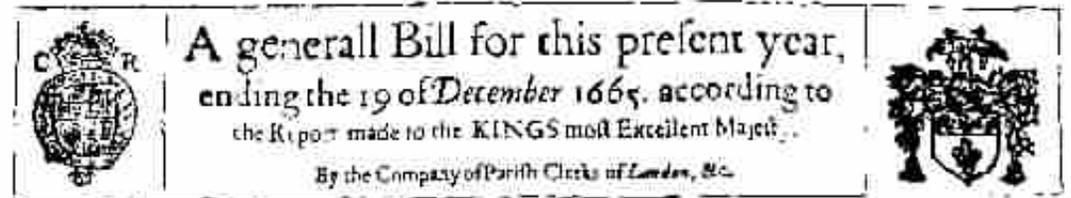


London Bills of Mortality



LONDON'S Dreadful Visitation:
Or, A COLLECTION of All the
Bills of Mortality
For this Present Year:
Beginning the 27th of December 1664, and
ending the 19th. of December following:
As also, The GENERAL or whole years BILL:
According to the Report made to the
KING's Most Excellent Majesty,
By the Company of Parish-Clerks of London. etc.

LONDON:
Printed and are to be sold by E. Cotes living in Aldersgate-street.
Printer to the said Company 1665.



A general Bill for this present year,
ending the 19 of December 1665, according to
the Report made to the KING's most Excellent Majesty.
By the Company of Parish Clerks of London, etc.

The Diseases and Casualties, this year.

A Abortive and Stillborne — 517	Executed — 21	Pallie — 30
Aged — 1545	Flux and Small Pox — 655	Plague — 68598
Aque and Peaver — 5257	Found dead in Streets, fields, &c. — 2	Plaster — 6
Apoplexy and Suddenly — 116	French Pox — 86	Plurisie — 19
Bedric — 10	Frighted — 23	Poysoning — 1
Blasind — 8	Gout and Sciatica — 27	Quintic — 35
Bleeding — 16	Grief — 28	Rickets — 157
Bloody Flux, Scouring & Flux — 187	Griping in the Guts — 228	Killing of the Lights — 157
Burnt and Scalded — 8	Hanged & made away themselves — 7	L. p. pure — 14
Calenture — 2	Headmole shot & Mole fallen — 14	Scurvy — 157
Cancer, Gangrene and Fiftula — 56	Jaundies — 120	Shingles and Swine pox — 2
Canker, and Thrush — 121	Imposiume — 227	Sores, Ulcers, broken and healed — 82
Childhood — 625	Kild by severall accidents — 28	Limbs — 82
Chrixtimes and Infants — 1258	Kings Evill — 28	Spleen — 14
Cold and Cough — 65	Leptotic — 2	Spotted Fever and Purples — 1929
Collick and Winde — 134	Lethargy — 14	Stoppings of the Stomack — 334
Consumption and Tiflick — 4838	Livergown — 21	Stone and Strangury — 8
Convulsion and Morice — 1058	Measum and Headach — 8	Sydet — 121
Disseited — 5	Measles — 7	Teeth and Worms — 2014
Droove and Tempany — 1478	Mothered and Shot — 9	Vomiting — 51
Drowned — 5	Overjaid & Starved — 45	Vinn — 7

♂ Males — 5114	Buried {	♂ Males — 48569	} Of the Plague — 68598
♀ Females — 4853		♀ Females — 48717	
♂ In all — 9967		♂ In all — 97286	

Increased in the Burials in the 130 Parishes and at the Pest-houses this year — 79009
Decreased of the Plague in the 130 Parishes and at the Pest-houses this year — 68598



Limitations of existing classifications

“The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases. The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay.”

– William Farr

First annual report.

London, Registrar General of England and Wales, 1839, p. 99.



To index and retrieve information

◆ Biomedical literature

- MEDLINE (15M citations from 4600 journals)
- Manually indexed
- Medical Subject Headings (MeSH)

◆ Genome

- Model organism databases (Fly, Mouse, Yeast, ...)
- Manually / semi-automatically curated
- Gene Ontology



MEDLINE and MeSH

□ 1: J Hist Neurosci. 2004 Mar;13(1):91-101

[Related Articles](#), [Links](#)

MetaPress

Black bile and psychomotor retardation: shades of melancholia in Dante's Inferno.

Widmer DA.

Memorial Sloan-Kettering Cancer Center, New York, NY 10017, USA. widmerd@mskcc.org

The history of melancholy depression is rich with images of movement retardation and mental dysfunction. The recent restoration of psychomotor symptoms to the diagnostic terminology of affective disorder is not novel to the students of medieval melancholia. The move back to the biology of this psychomotor dysfunction with the technical advances in brain imaging in recent years only echoes centuries-old writings on the centrality of movement changes in the depressive condition. The Inferno, the first cantica of Dante Alighieri's *Commedia*, has a wonderful abundance of allusions to the importance of psychomotor symptoms in describing the depressed individual. Slowed steps, garbled speech, frozen tears, these and many other images keep the physical manifestations of psychomotor suffering in the forefront of the reader's mind. Considering Medieval and Renaissance writings on melancholy suffering, it is fitting that Dante shows a bodily illness reflected in the hellish torments visited on the damned. From the souls of the sullen to those of the violent, the panorama of psychomotor symptoms plays a prominent role in the poem as well as in the medical and literary prose of succeeding centuries.

MeSH Terms:

- ◆ Depressive Disorder/history*
- ◆ History of Medicine, Medieval
- ◆ Human
- ◆ Italy
- ◆ Literature, Medieval/history*
- ◆ Medicine in Literature*
- ◆ Poetry/history*
- ◆ Psychomotor Disorders/history*

PubMed

National
Library
of Medicine 

Mouse Genome Database and GO

Entrez Gene

1: **Nf2 neurofibromatosis 2** [*Mus musculus*]
 GeneID: 18016 Locus tag: [MGI:97307](#)



► General gene information

GeneOntology
 Provided by MGI

Function	Evidence
cytoskeletal protein binding	IEA
protein binding	IPI PubMed
structural molecule activity	IEA
Process	
intercellular junction assembly and/or maintenance	IMP PubMed
negative regulation of cell cycle	IEA
negative regulation of protein kinase activity	IDA PubMed
regulation of cell proliferation	IMP PubMed
Component	
adherens junction	IMP PubMed
cytoplasm	IEA
cytoskeleton	IEA
membrane	IEA



To serve as a reference

◆ Reference terminology/ontology

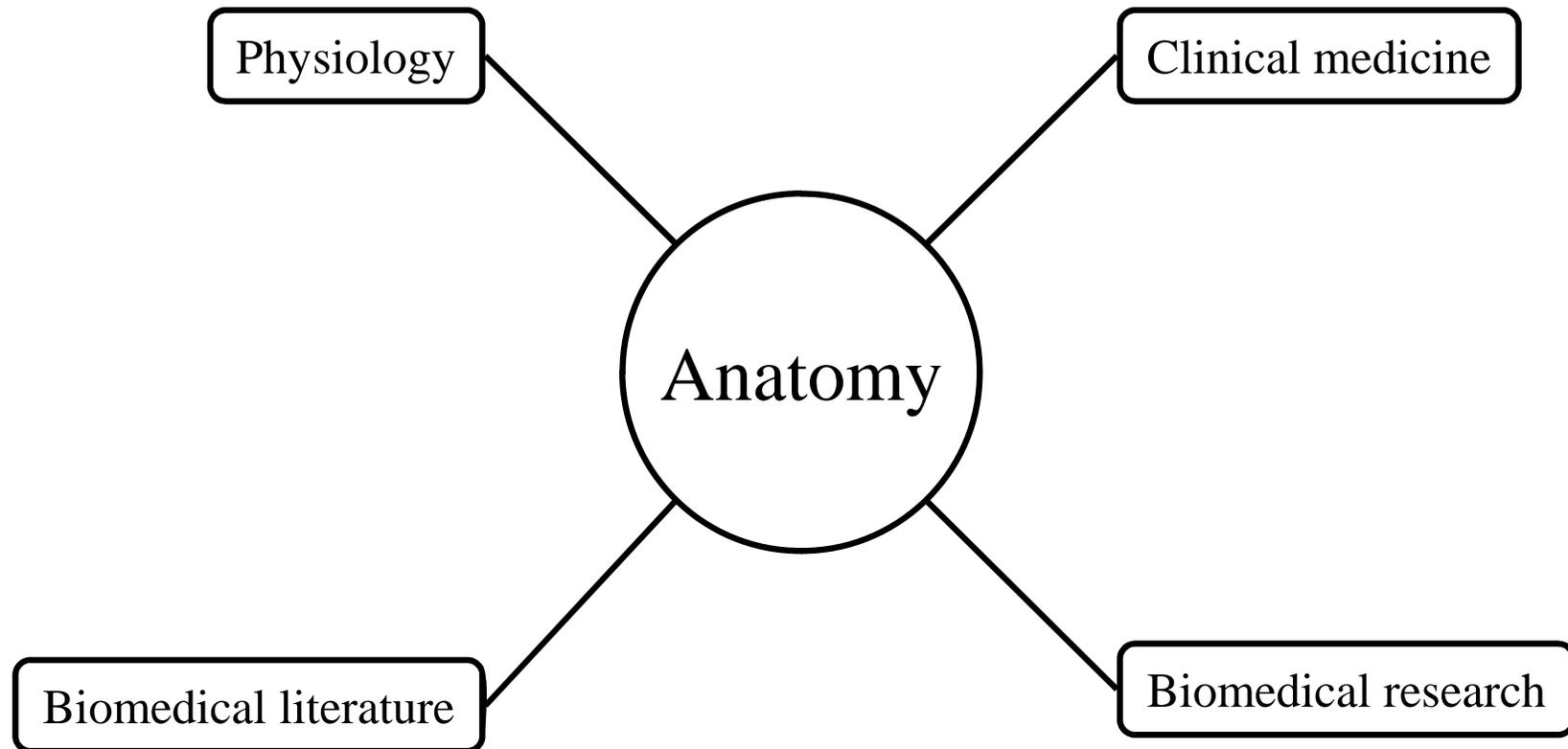
- Universally needed
- Developed independently of any purposes
- Reusable by many applications

◆ Examples

- VA National Drug File (NDF)
- Foundational Model of Anatomy (FMA)
- SNOMED CT



Anatomy in Biomedicine

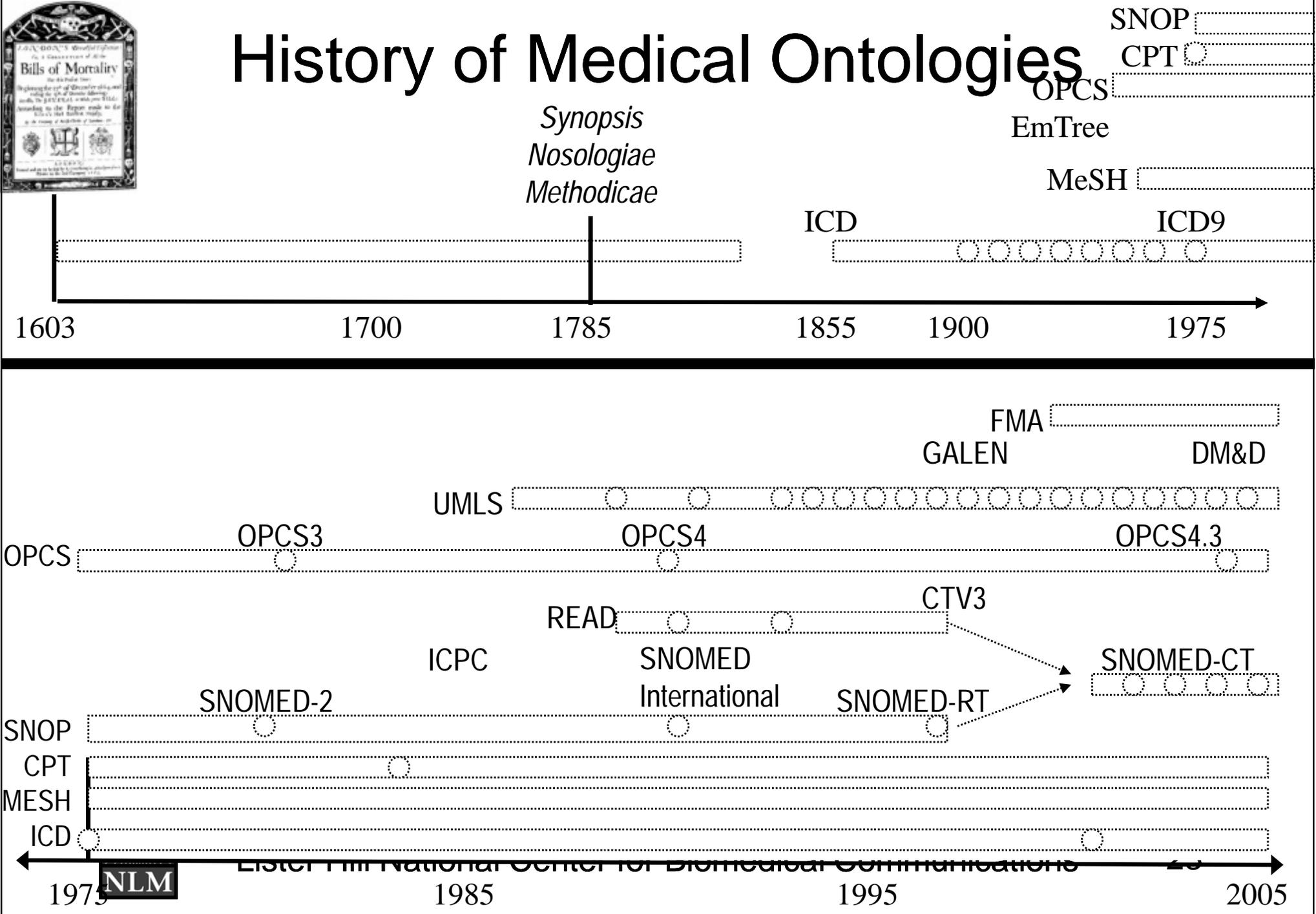
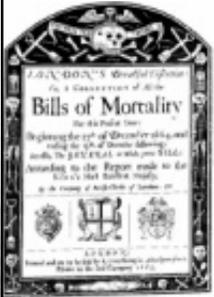


Administrative terminologies

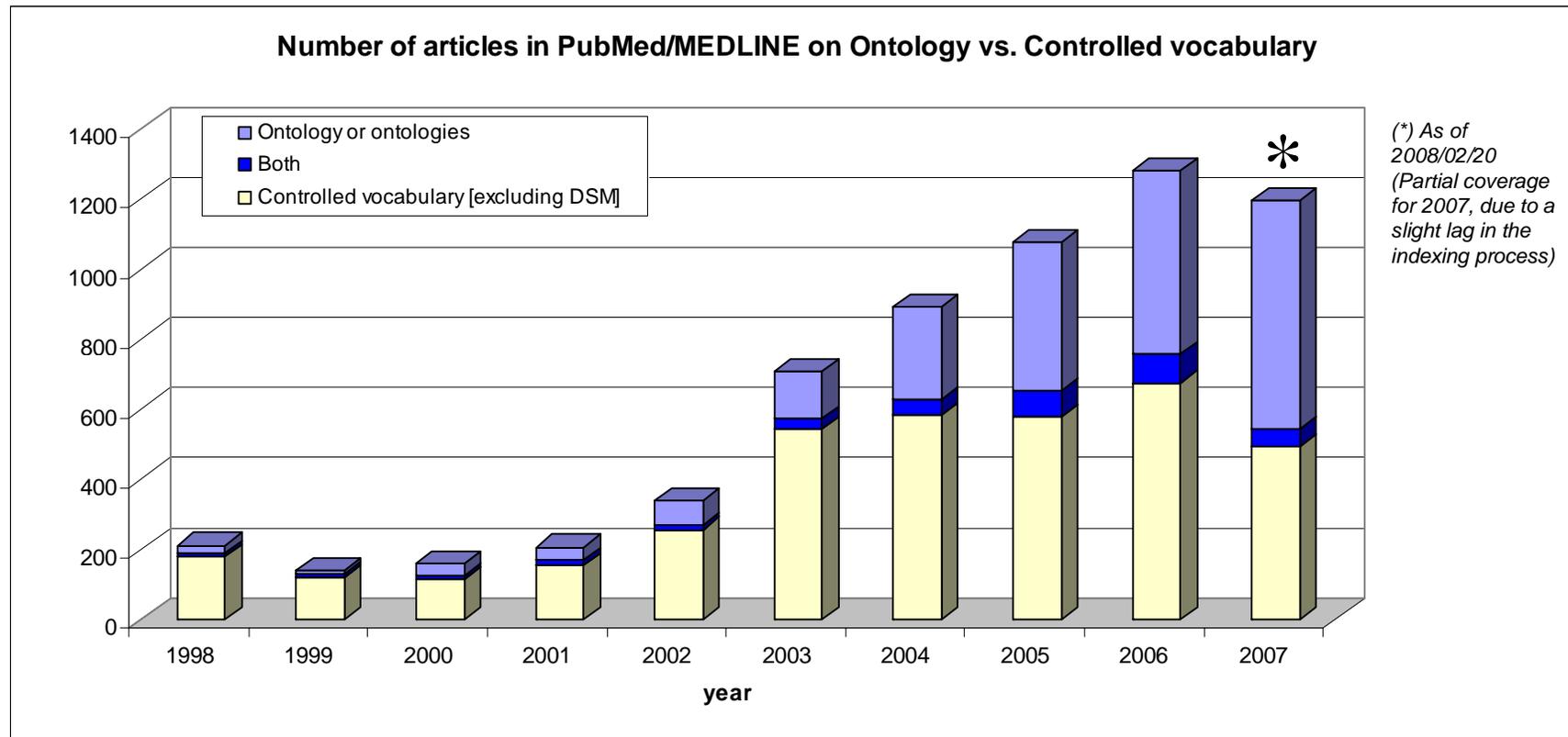
- ◆ Coding patient records
 - International Classification of Primary Care (ICPC)
 - SNOMED
 - Read Codes
- ◆ Reporting claims to health insurance companies
 - Current Procedural Terminology (CPT)
 - International Classification of Diseases (ICD-9 CM)
 - Healthcare Common Procedure Coding System (HCPCS)



History of Medical Ontologies



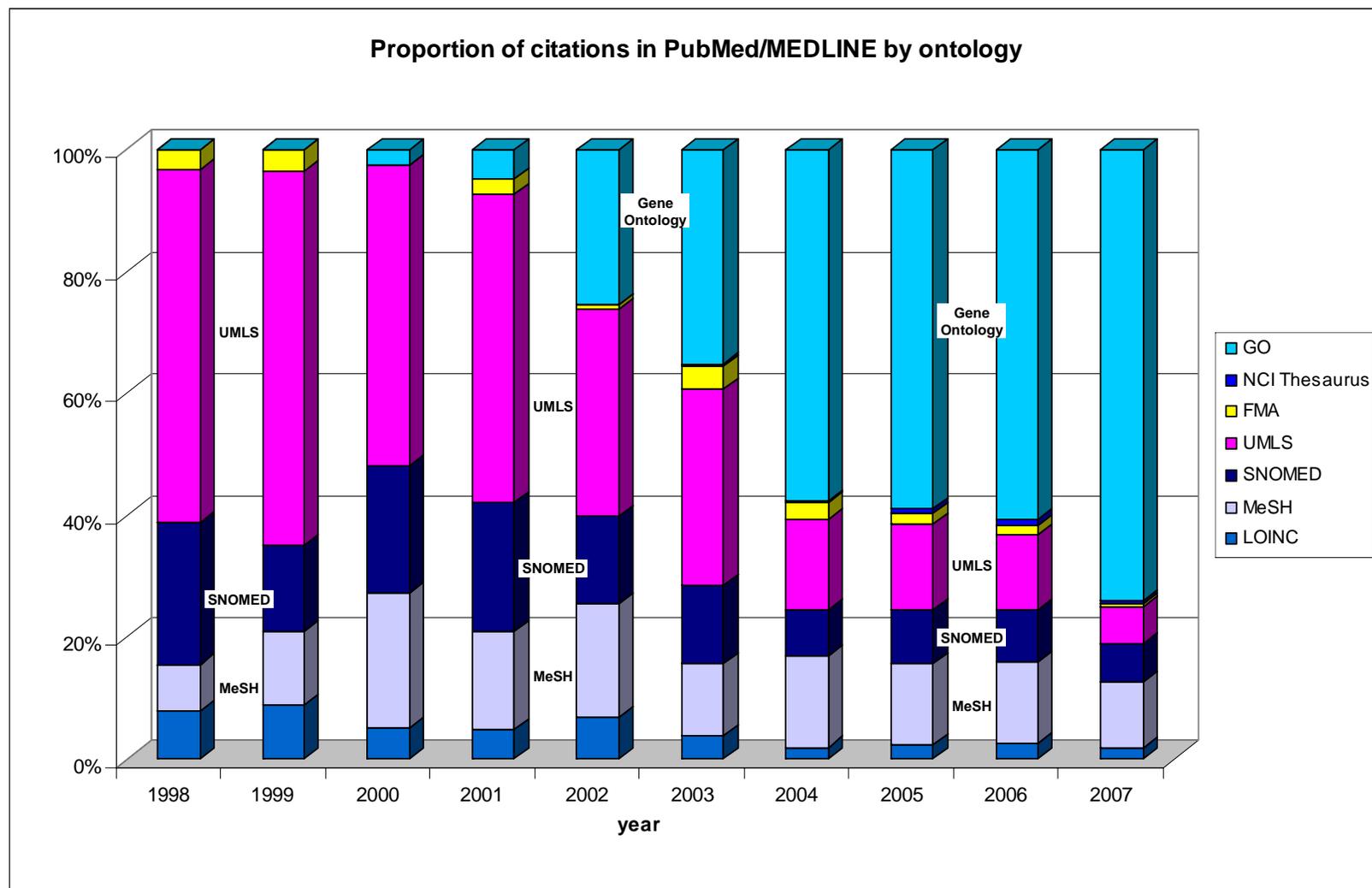
Biomedical ontology in PubMed



[Bodenreider, YBMI 2008]



Biomedical ontologies in PubMed

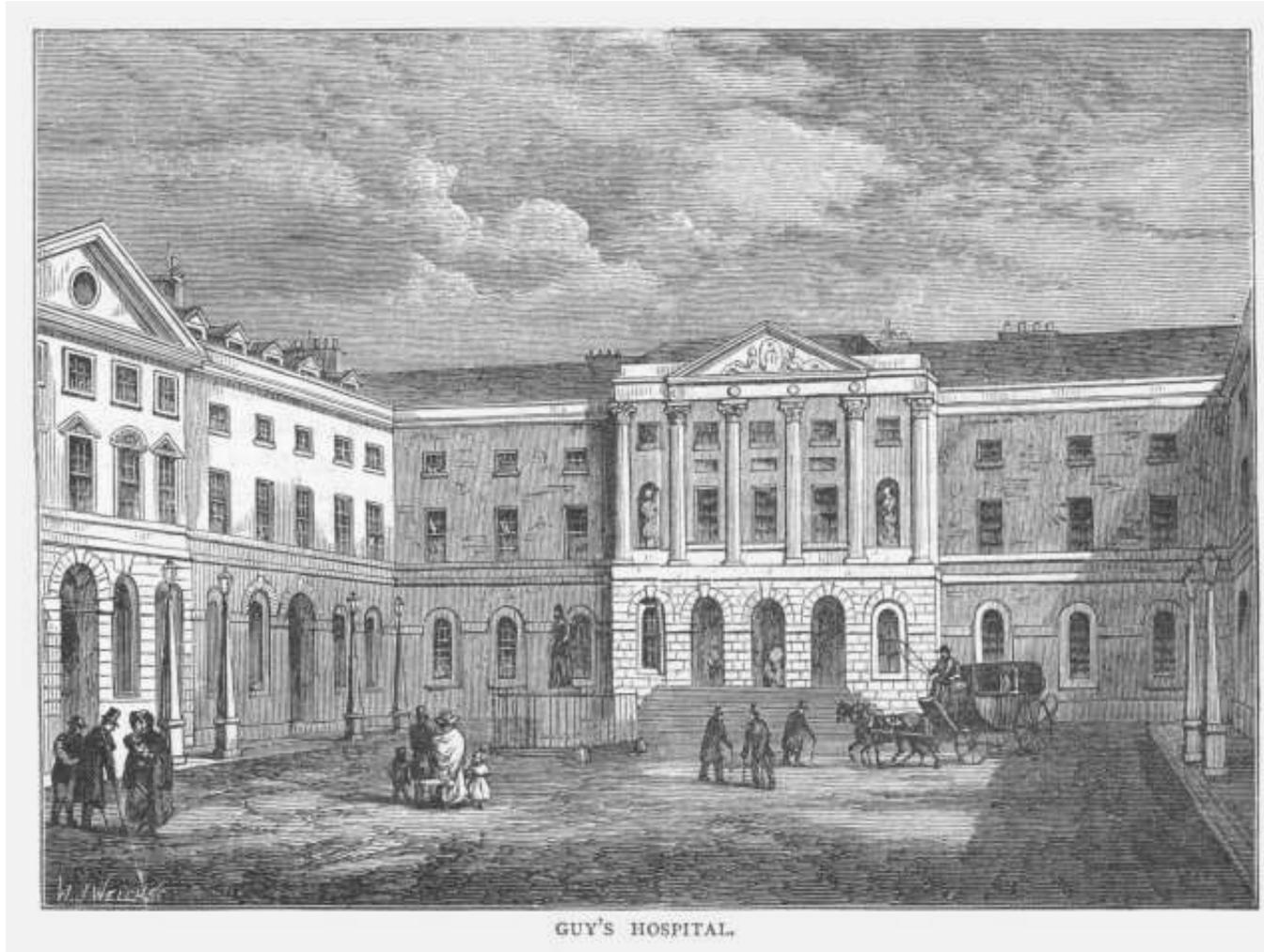


[Bodenreider, YBMI 2008]



Introduction to biomedical terminologies through an example

Guy's Hospital, London

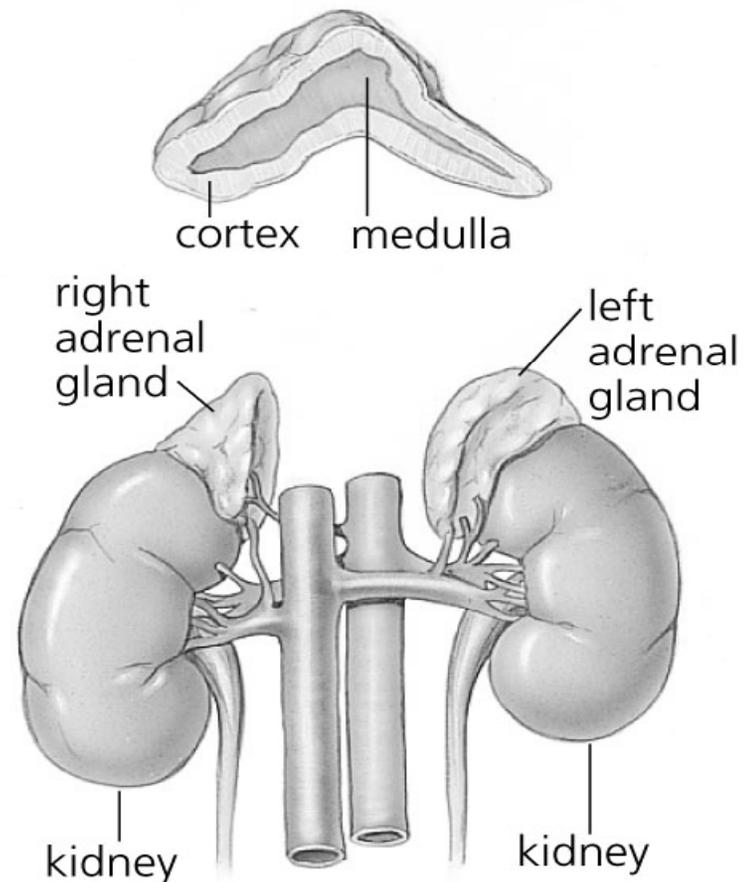


Thomas Addison (1795-1860)



Addison's disease

- ◆ Addison's disease is a rare endocrine disorder
- ◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- ◆ For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism



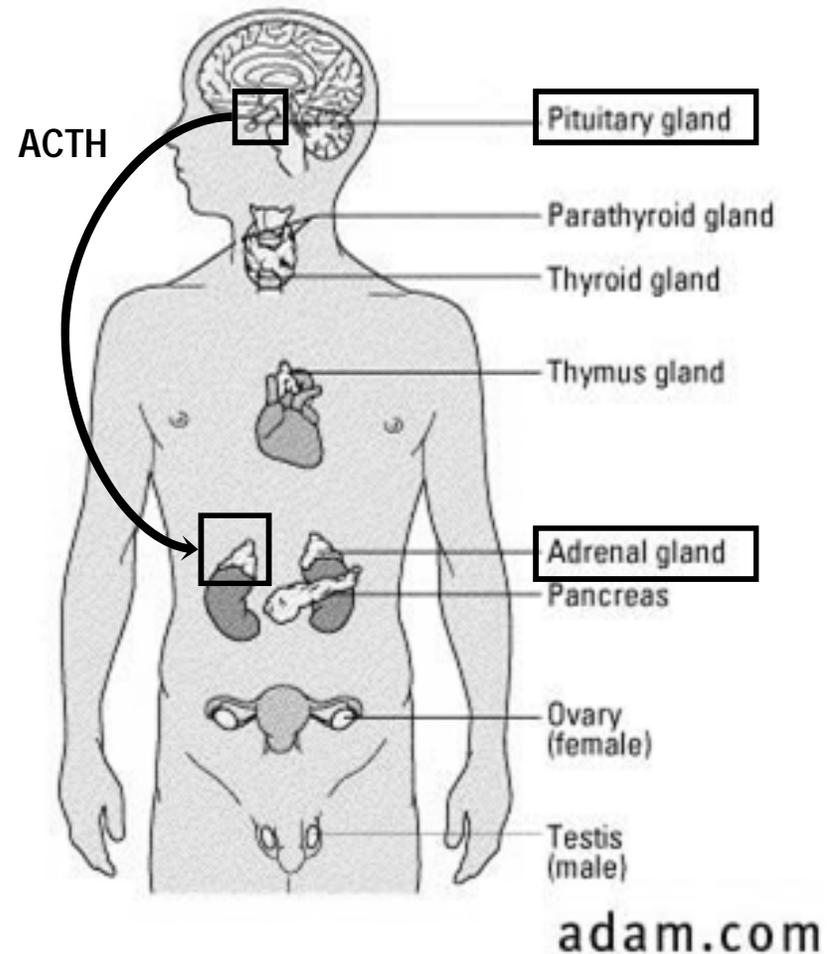
Adrenal insufficiency Clinical variants

◆ Primary / Secondary

- Primary: lesion of the adrenal glands themselves
- Secondary: inadequate secretion of ACTH by the pituitary gland

◆ Acute / Chronic

◆ Isolated / Polyendocrine deficiency syndrome



Addison's disease: Symptoms

- ◆ Fatigue
- ◆ Weakness
- ◆ Low blood pressure
- ◆ Pigmentation of the skin (exposed and non-exposed parts of the body)
- ◆ ...



AD in medical vocabularies

◆ Synonyms: different terms

- Addisonian syndrome
 - Bronzed disease
 - Addison melanoderma
 - Asthenia pigmentosa
 - Primary adrenal deficiency
 - Primary adrenal insufficiency
 - Primary adrenocortical insufficiency
 - Chronic adrenocortical insufficiency
-] eponym
-] symptoms
-] clinical variants

◆ Contexts: different hierarchies



Internal Classification of Diseases



CHAPTER 4

Endocrine, nutritional and metabolic diseases (E00-E90)

Disorders of other endocrine glands (E20-E35)

E27 Other disorders of adrenal gland

E27.0 Other adrenocortical overactivity

Overproduction of ACTH, not associated with Cushing's disease

Premature adrenarche

Excludes 1: Cushing's syndrome (E24.-)

E27.1 Primary adrenocortical insufficiency

Addison's disease

Adrenocortical insufficiency NOS

Autoimmune adrenalitis

Excludes 1: Addison only phenotype adrenoleukodystrophy (E71.428)

amyloidosis (E85)

tuberculous Addison's disease (A18.7)

Waterhouse-Friderichsen syndrome (A39.1)

E27.2 Addisonian crisis

Adrenal crisis

Adrenocortical crisis

E27.3 Drug-induced adrenocortical insufficiency

Code first (T36-T50) to identify drug

E27.4 Other and unspecified adrenocortical insufficiency

Medical Subject Headings



MeSH Tree Structures

Endocrine Diseases [C19]

Adrenal Gland Diseases [C19.053]

Adrenal Gland Hypofunction [C19.053.264]

- ▶ Addison's Disease [C19.053.264.263]
- Adrenoleukodystrophy [C19.053.264.270]
- Hypoadosteronism [C19.053.264.480]

Immunologic Diseases [C20]

Autoimmune Diseases [C20.111]

- ▶ Addison's Disease [C20.111.163]
- Anemia, Hemolytic, Autoimmune [C20.111.175]
- Anti-Glomerular Basement Membrane Disease [C20.111.190]
- Antiphospholipid Syndrome [C20.111.197]
- Arthritis, Rheumatoid [C20.111.199] +
- Autoimmune Diseases of the Nervous System [C20.111.258] +



SNOMED CT



Hierarchy Subtype hierarchy

- 386584007 adrenal cortical hypofunction
 - 363732003 Addison's disease
 - 237760008 Addison's disease with adrenoleucodystrophy
 - 76715008 Addison's disease due to autoimmunity
 - 186270000 tuberculous Addison's disease
 - 11244009 polyglandular autoimmune syndrome, type 1

Addison's disease - Definition
Concept Status: **Current**

- Descriptions
 - Addison's disease (disorder)
 - Addison's disease
 - enfermedad de Addison
 - enfermedad de Addison (trastorno)
- Definition: *Primitive*
 - is a
 - adrenal cortical hypofunction
 - finding site
 - adrenal cortex structure
- Qualifiers
 - severity
 - severities
 - episodicity
 - episodicities
 - clinical course
 - courses
- Codes
 - Original SnomedId : DB-70620
 - Read Code (Ctv3Id) : C1541



Biomedical terms as names for biomedical classes

Terms reflecting valid classes

- Pulmonary anthrax
- BRCA1 protein
- Coronary artery
- Coronary artery bypass
- ...
 - Non-insulin dependent diabetes mellitus
 - Non-Hodgkin lymphoma
 - Non-steroidal anti-inflammatory drugs
 - Non-opioid analgesics
 - Non-invasive medical procedure



Issues

- ◆ Multiple terms for a class
- ◆ Multiple classes for a term
- ◆ Presence of non-ontological features in terms
- ◆ Composite terms



Multiple terms for a class

◆ Synonymy

- Left coronary artery
- LCA
- Arteria coronaria sinistra

- Addison's disease
- Primary adrenocortical insufficiency

◆ “Clinical synonymy” (vs. identity)

- Abdominal swelling
- Swollen abdomen

- Posttransfusion hepatitis
- Posttransfusion viral hepatitis

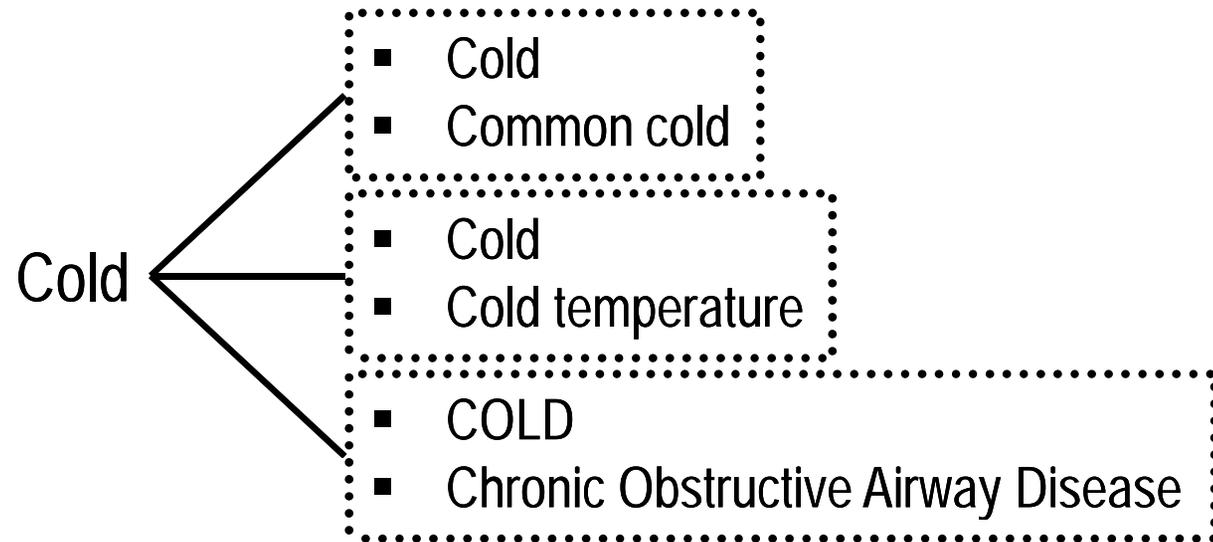
- Addison's disease
- Primary adrenocortical insufficiency

vs. Waterhouse-Friderichsen Syndrome

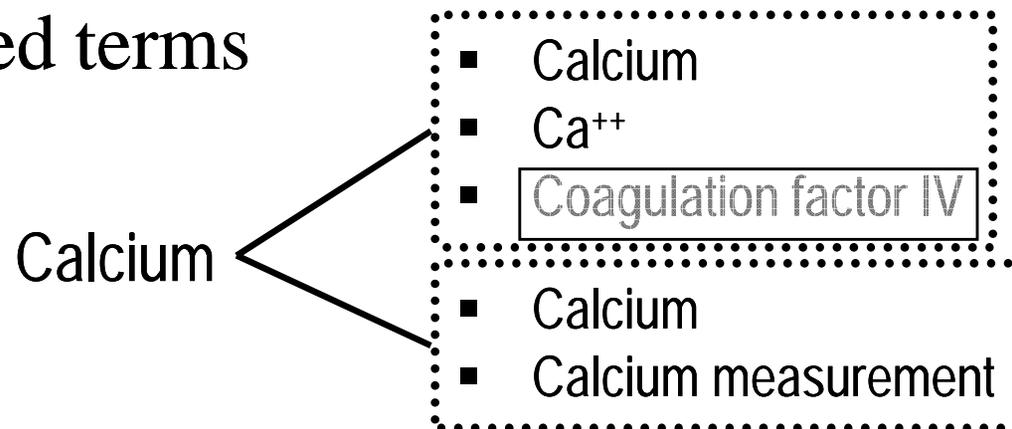


Multiple classes for a term

◆ Polysemy



◆ Truncated terms



Non-ontological features in terms

◆ Epistemological features

- Gallbladder calculus without mention of cholecystitis
- Diarrhea of presumed infectious origin
- Replacement of unspecified heart valve
- ...



Ontology vs. Epistemology

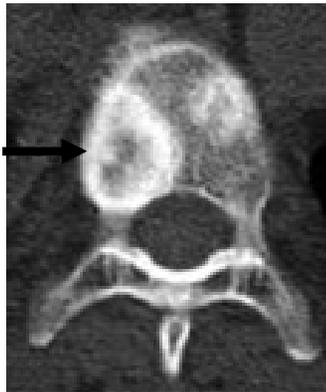
◆ Ontology

- Invariants in reality
 - Classes (universals)
 - Relations between them
- Theory of reality

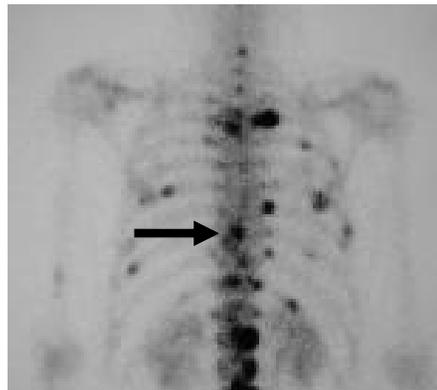
◆ Epistemology

- Knowledge about such entities
- Perception of reality

Bone metastasis



Bone metastasis
diagnosed by CT scan

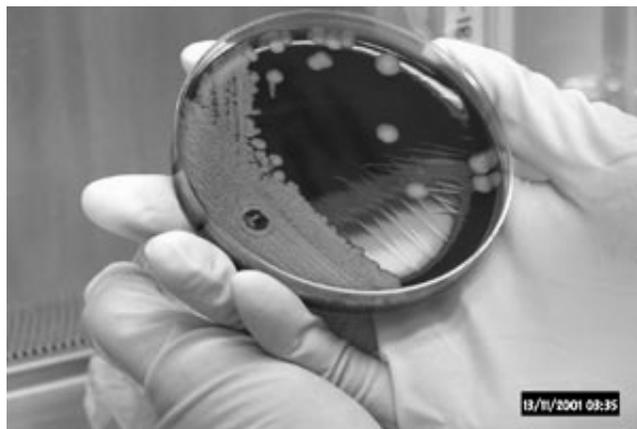


Bone metastasis
diagnosed by Tc99m bone scintiscan

Composite terms

◆ Sentence-like terms

- Several classes and their relations
- May contain epistemological features
- Tuberculosis of adrenal glands, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture



More composite terms

- Nontraffic accident involving being accidentally pushed from motor vehicle, except off-road motor vehicle, while in motion, not on public highway, driver of motor vehicle injured
- Determine whether the elder patient and caretaker have a functional social support network to assist the patient in performing activities of daily living and in obtaining health care, transportation, therapy, medications, community resource information, financial advice, and assistance with personal problems
- Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (eg, nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (eg, lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different

Terminological relations as a
surrogate for ontological relations

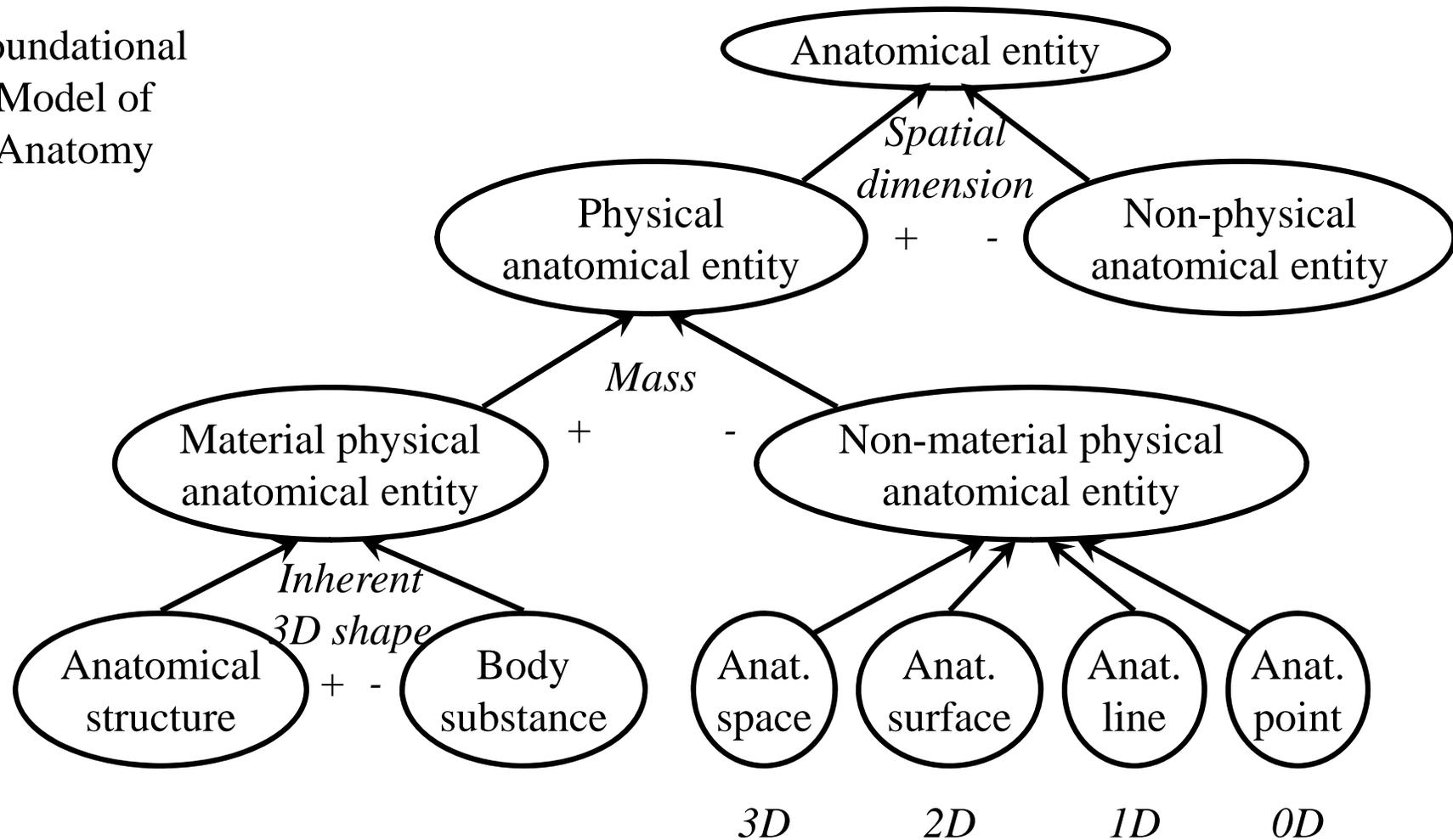
Issues

- ◆ Lack of explicit classificatory principle
- ◆ Underspecification of the relations
- ◆ Thesaurus relations
- ◆ Limited depth in hierarchies “by design”



Explicit classificatory principle

Foundational
Model of
Anatomy



No explicit classificatory principle



3. ☐ Diseases [C]

- ◇ Bacterial Infections and Mycoses [C01] +
- ◇ Virus Diseases [C02] +
- ◇ Parasitic Diseases [C03] +
- ◇ Neoplasms [C04] +
- ◇ Musculoskeletal Diseases [C05] +
- ◇ Digestive System Diseases [C06] +
- ◇ Stomatognathic Diseases [C07] +
- ◇ Respiratory Tract Diseases [C08] +
- ◇ Otorhinolaryngologic Diseases [C09] +
- ◇ Nervous System Diseases [C10] +
- ◇ Eye Diseases [C11] +
- ◇ Urologic and Male Genital Diseases [C12] +
- ◇ Female Genital Diseases and Pregnancy Complications [C13] +
- ◇ Cardiovascular Diseases [C14] +
- ◇ Hemic and Lymphatic Diseases [C15] +
- ◇ Neonatal Diseases and Abnormalities [C16] +
- ◇ Skin and Connective Tissue Diseases [C17] +
- ◇ Nutritional and Metabolic Diseases [C18] +
- ◇ Endocrine Diseases [C19] +
- ◇ Immunologic Diseases [C20] +
- ◇ Disorders of Environmental Origin [C21] +
- ◇ Animal Diseases [C22] +
- ◇ Pathological Conditions, Signs and Symptoms [C23] +

agent/cause

location

stage in life



1. Certain infectious and parasitic diseases
2. Neoplasms
3. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4. Endocrine, nutritional, and metabolic diseases
5. Mental and behavioral disorders
6. Diseases of nervous system
7. Diseases of the eye and adnexa
8. Diseases of the ear and mastoid process
9. Diseases of circulatory system
10. Diseases of respiratory system
11. Diseases of digestive system
12. Diseases of the skin and subcutaneous tissue
13. Diseases of the musculoskeletal system and connective tissue
14. Diseases of the genitourinary system
15. Pregnancy, childbirth, and the puerperium
16. Certain conditions originating in the newborn (perinatal) period
17. Congenital malformations, deformations and chromosomal abnormalities
18. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19. Injury, poisoning and certain other consequences of external causes
20. External causes of morbidity
21. Factors influencing health status and contact with health service



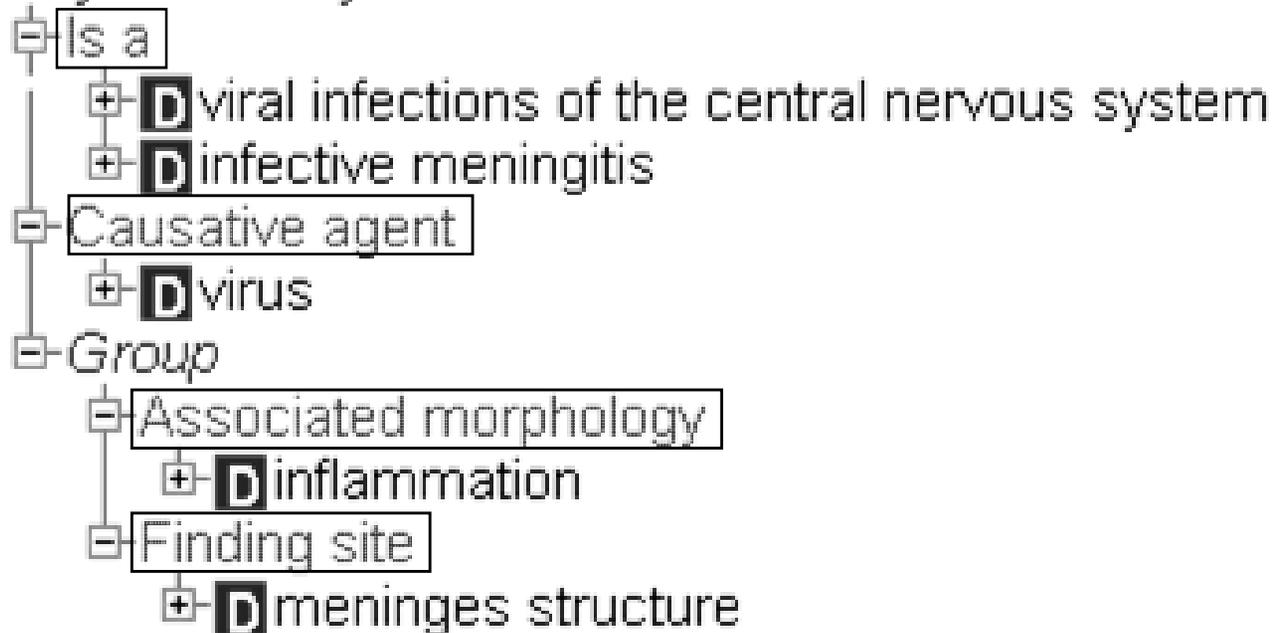
- Attribute
- Body structure
- Clinical finding
- Context-dependent categories
- Environments and geographical locations
- Events
- Observable entity
- Organism
- Pharmaceutical / biologic product
- Physical force
- Physical object
- Procedure
- Qualifier value
- Social context
- Special concept
- Specimen
- Staging and scales
- Substance



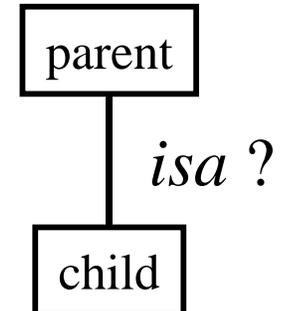
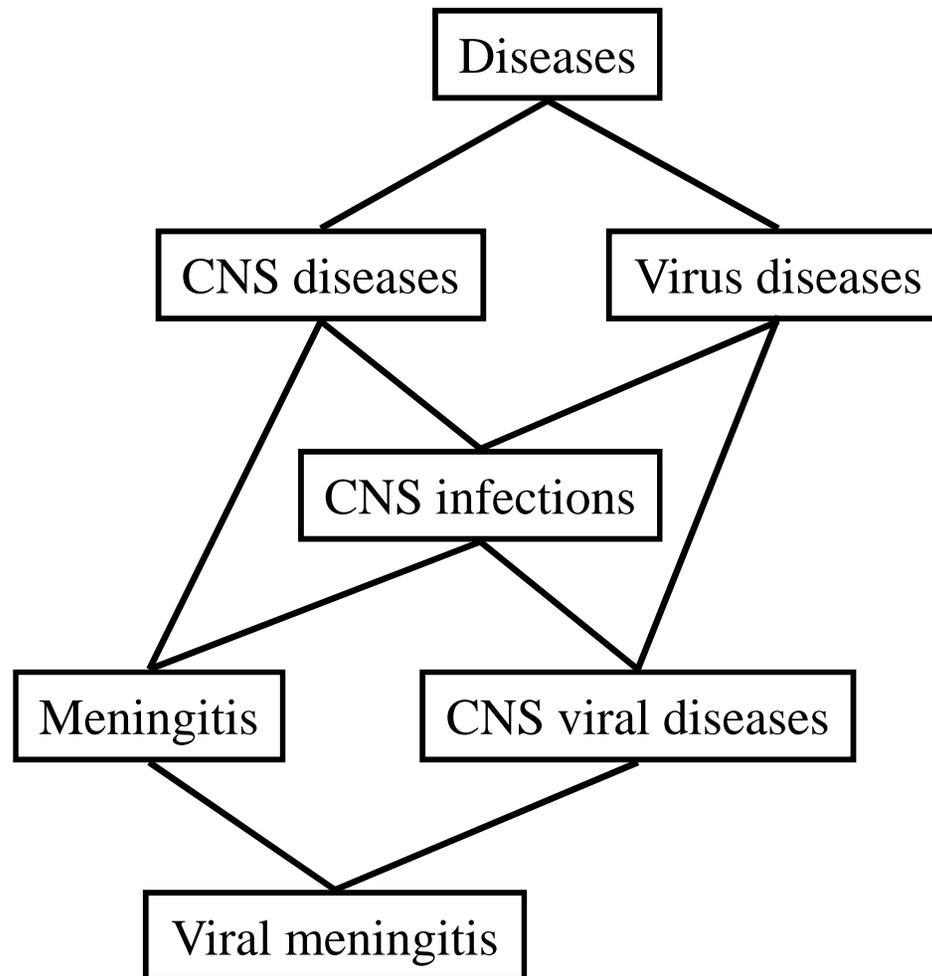
Fully specified relations

Viral meningitis in SNOMED CT

Fully defined by ...



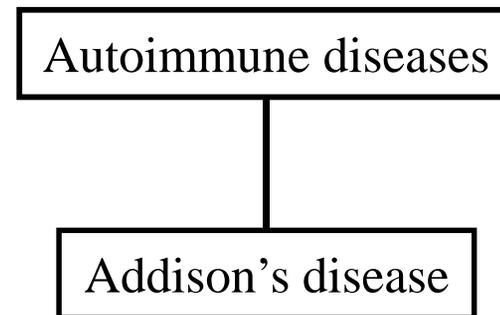
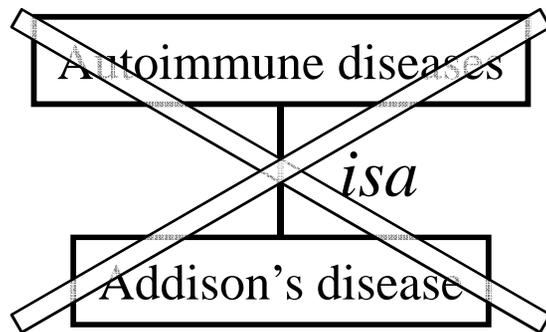
Underspecification of the relations



Thesaurus relations

◆ Addison's disease

- Due to auto-immunity in 80% of the cases
- Other causes include tuberculosis



Relations used to create hierarchical structures
vs. hierarchical relations



Endocrine Diseases [C19]



Adrenal Gland Diseases [C19.053]

Adrenal Gland Hypofunction [C19.053.264]

- ▶ Addison's Disease [C19.053.264.263]
- Adrenoleukodystrophy [C19.053.264.270]
- Hypoaldosteronism [C19.053.264.480]

Immunologic Diseases [C20]

Autoimmune Diseases [C20.111]

- ▶ Addison's Disease [C20.111.163]
- Anemia, Hemolytic, Autoimmune [C20.111.175]
- Anti-Glomerular Basement Membrane Disease [C20.111.190]
- Antiphospholipid Syndrome [C20.111.197]
- Arthritis, Rheumatoid [C20.111.199] +

Hierarchy

Subtype hierarchy



- └ adrenal cortical hypofunction
 - └ Addison's disease
 - └ Addison's disease due to autoimmunity
 - └ Addison's disease with adrenoleucodystrophy
 - └ polyglandular autoimmune syndrome, type 1
 - └ tuberculous Addison's disease

Accidents in MeSH

Environment and Public Health [G03]

Public Health [G03.850]

▶ Accidents [G03.850.110]

Accident Prevention [G03.850.110.060] +

Accidental Falls [G03.850.110.085]

Accidents, Aviation [G03.850.110.185]

Accidents, Home [G03.850.110.205]

Accidents, Occupational [G03.850.110.250] +

Accidents, Radiation [G03.850.110.285]

Accidents, Traffic [G03.850.110.320]

Drowning [G03.850.110.500] +



Limited depth in hierarchies “by design”

- ◆ Term identifier (code) used to record the position in the hierarchy
 - Limited number of digits available
 - May hide part of the structure
- ◆ Terminologies: ICD, SNOMED, ...

E84 Cystic fibrosis

Includes: mucoviscidosis

E84.0 Cystic fibrosis with pulmonary manifestations

Use additional code to identify any infectious organism present, such as:

Pseudomonas (B96.5)

E84.1 Meconium ileus in cystic fibrosis

Excludes1: meconium ileus not due to Cystic fibrosis (P75)

E84.2 Cystic fibrosis with gastrointestinal manifestations

Excludes2: meconium ileus in cystic fibrosis (E84.1)

E84.8 Cystic fibrosis with other manifestations



Cystic fibrosis in ICD

E84 Cystic fibrosis

Includes: mucoviscidosis

E84.0 Cystic fibrosis with pulmonary manifestations

Use additional code to identify any infectious organism present, such as:
Pseudomonas (B96.5)

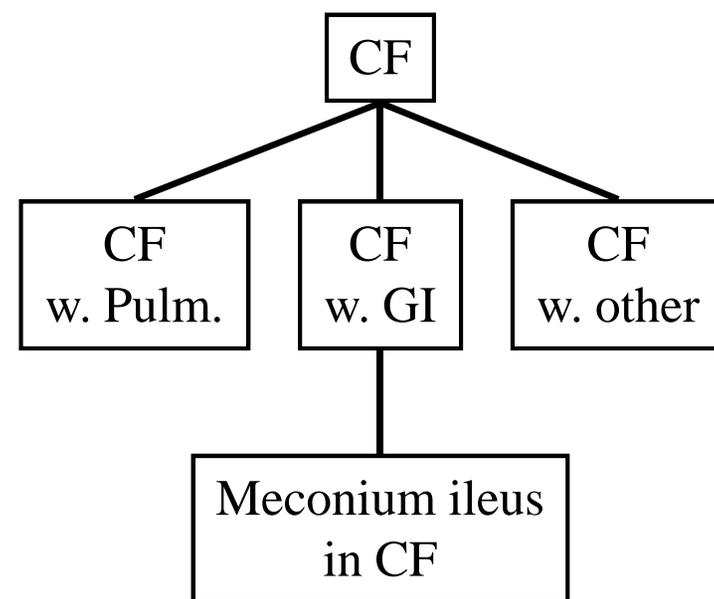
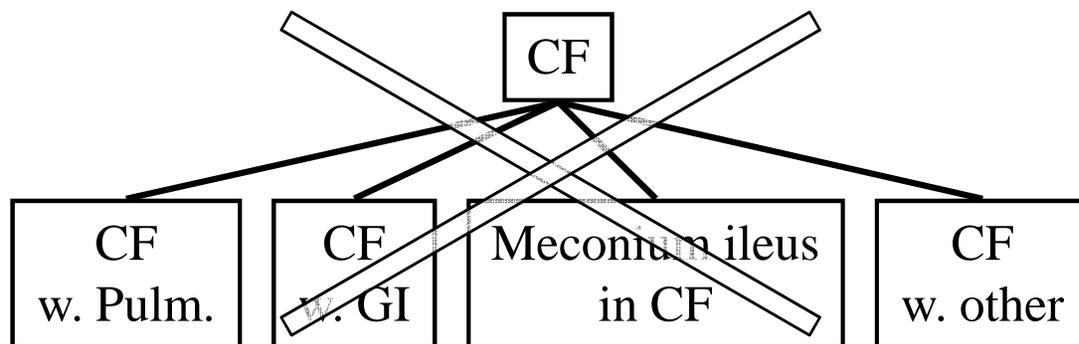
E84.1 Meconium ileus in cystic fibrosis

Excludes1: meconium ileus not due to Cystic fibrosis (P75)

E84.2 Cystic fibrosis with gastrointestinal manifestations

Excludes2: meconium ileus in cystic fibrosis (E84.1)

E84.8 Cystic fibrosis with other manifestations



Conclusions

Conclusions ☹️

◆ Biomedical terms

- reflect some aspects of biomedical reality
 - Although the primary concern of terminology is naming, not reflecting reality
- often convey additional features (e.g., epistemology)

◆ Biomedical terminology tends to offset part of the complexity

- but often reflects utility



Conclusions ☺

- ◆ Biomedical terminologies can help populate biomedical ontologies
- ◆ Resources needed
 - Linguistic analysis of terms
 - Statistical analysis of terms in a corpus / annotation database (dependence relations)
 - Manual curation



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9, 2008 – Session #2

Design Principles, Formalisms and Tools for Biomedical Ontologies



Olivier Bodenreider

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

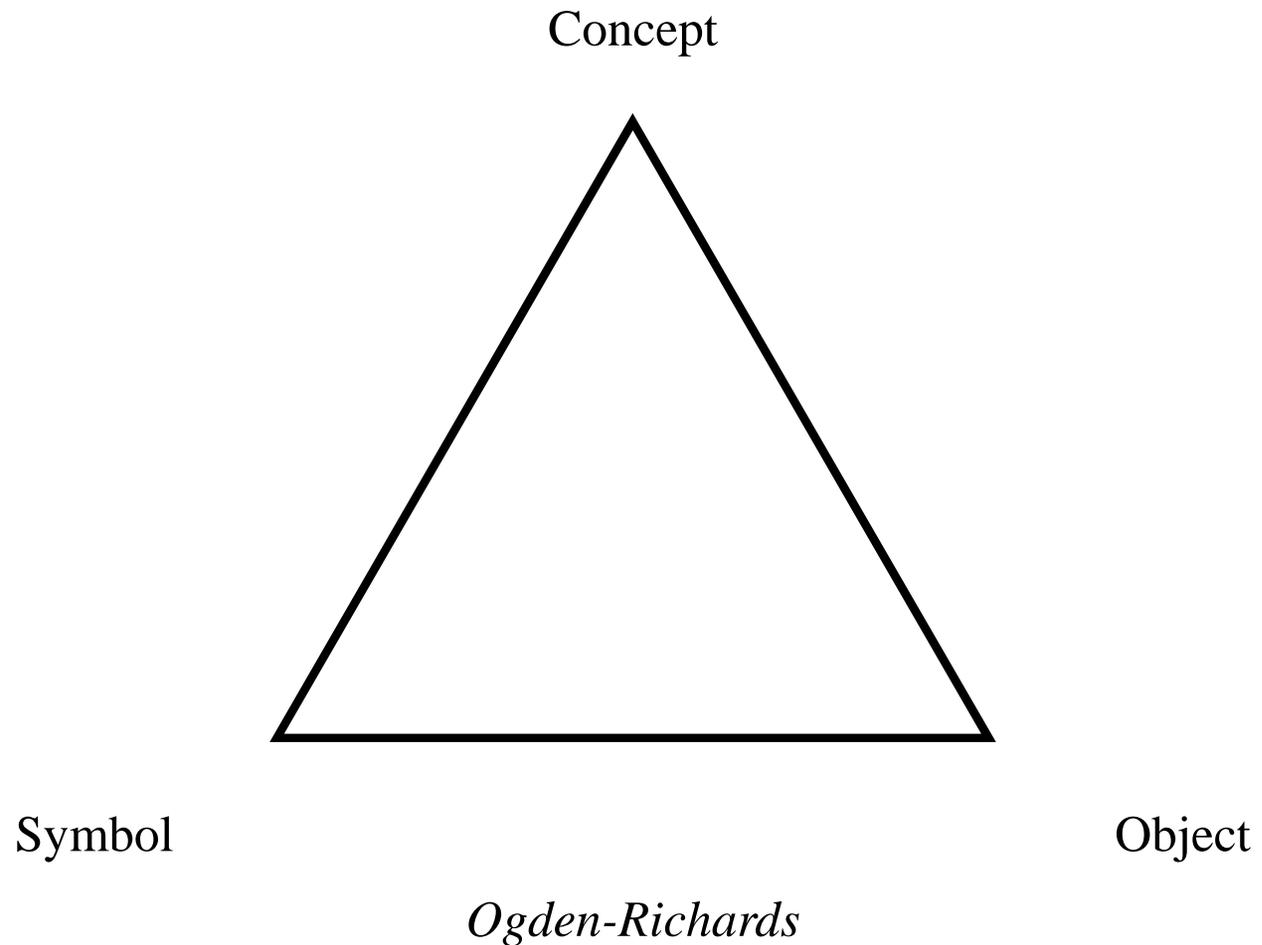
Overview

- ◆ Definitions
 - Ontologies vs. other artifacts
 - Kinds of ontologies
- ◆ Some principles of formal ontology
 - Top-level categories
 - Categories of relationships
- ◆ Formalisms and tools



Definitions

Introduction



Definitions

- ◆ The *What* question
 - Objects in the world
 - With their properties
 - With their relations to other objects
 - Also: events, processes, and states
- ◆ Explicit specification of a conceptualization
 - Support software applications
- ◆ Domain ontology reflects
 - Underlying reality
 - Theory of the domain



Examples of use

- ◆ Natural language processing
- ◆ Access to heterogeneous sources of information
(e.g., Semantic Web)
- ◆ Systems engineering

- ◆ Interoperability



Ontology vs. other artifacts

- ◆ **Ontology**
 - Defining types of things and their relations
- ◆ **Terminology**
 - Naming things in a domain
- ◆ **Thesaurus**
 - Organizing things for a given purpose
- ◆ **Classification**
 - Placing things into (arbitrary) classes
- ◆ **Knowledge bases**
 - Assertional knowledge

[Smith, KR-MED 2006]

[Chute, JAMIA 2000]



(Controlled) Terminology

- ◆ Objective: naming things
- ◆ Example: Current Procedural Terminology (CPT)
- ◆ Shared understanding
 - Agreement on what terms to use
 - Utility-driven (arbitrary)

Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (eg, nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (eg, lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different

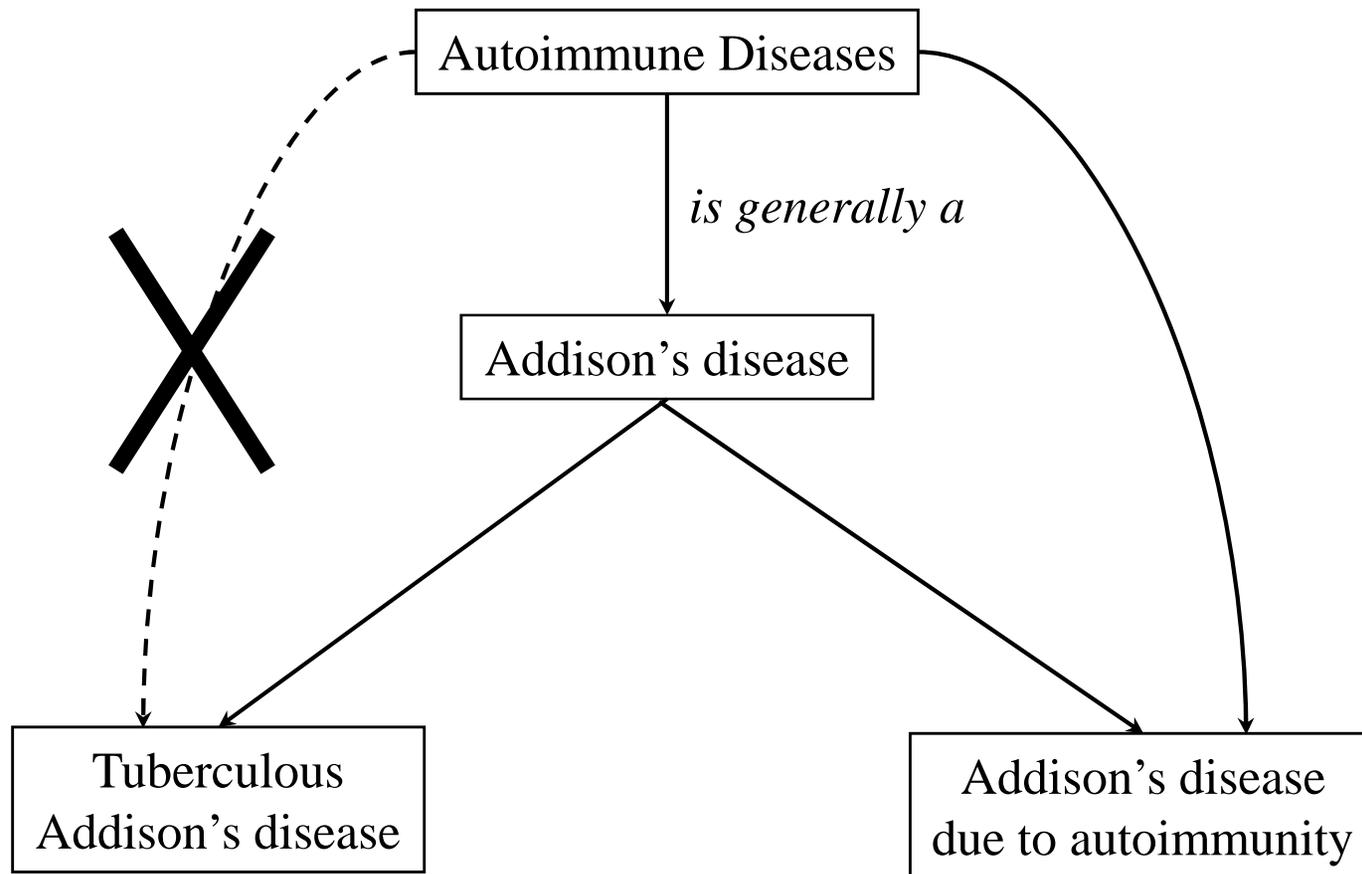


Thesaurus

- ◆ Objective: organize things for a purpose
 - e.g., information retrieval
 - Organization by relatedness
- ◆ Example: Medical Subject Headings (MeSH)
 - Indexing/retrieval of biomedical articles
- ◆ Relations used in hierarchies
vs. hierarchical relations



Thesaurus vs. ontology



Classification

- ◆ Objective: placing things into classes
- ◆ Characteristics
 - Single inheritance (tree)
 - Idiosyncratic inclusion/exclusion criteria

E10

Insulin-dependent diabetes mellitus

[See before E10 for subdivisions]

Includes: diabetes (mellitus):

- brittle
- juvenile-onset
- ketosis-prone
- type I

Excludes: diabetes mellitus (in):

- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)
- glycosuria:
 - NOS (R81)
 - renal (E74.8)
- impaired glucose tolerance (R73.0)
- postsurgical hypoinsulinaemia (E89.1)

Classification

◆ Characteristics (continued)

- Everything must be classified, including
 - When there is no specific slot (NEC)
 - When there is insufficient information (NOS)

E84

Cystic fibrosis

Includes: mucoviscidosis

E84.0

Cystic fibrosis with pulmonary manifestations

E84.1

Cystic fibrosis with intestinal manifestations

Meconium ileus+ (P75*)

Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)

E84.8

Cystic fibrosis with other manifestations

Cystic fibrosis with combined manifestations

E84.9

Cystic fibrosis, unspecified



Knowledge Bases

- ◆ Objective: represent knowledge needed for a given application
- ◆ Example: drug knowledge bases
- ◆ Assertional knowledge
 - Vs. definitional knowledge in ontologies
 - Often probabilistic
- ◆ Examples of assertions
 - Indications of a drug
 - Signs and symptoms of a disease



Fuzzy borders

- ◆ Some ontologies also collect names
 - FMA
- ◆ Some terminologies also provide formal definitions
 - SNOMED CT
- ◆ Some terminologies/ontologies include both definitional and assertional knowledge
 - SNOMED CT



Types of resources

◆ Lexical resources

- Collections of lexical items
- Additional information
 - Part of speech
 - Spelling variants
- Useful for entity recognition
- UMLS SPECIALIST
Lexicon, WordNet

◆ Ontological resources

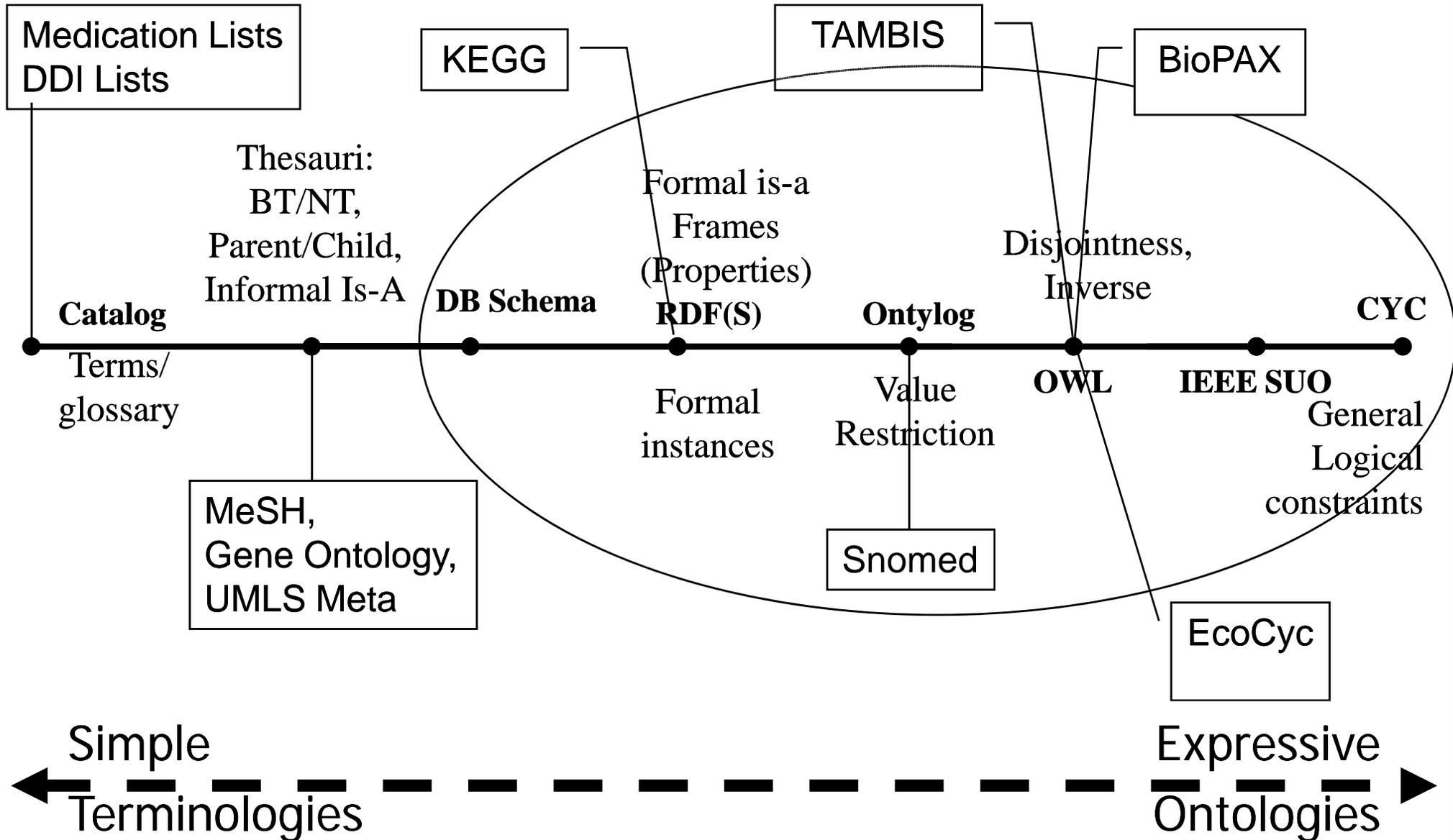
- Collections of
 - kinds of entities
(substances, qualities, processes)
 - relations among them
- Useful for relation extraction
- UMLS Semantic Network, BioTop



◆ Terminological resources

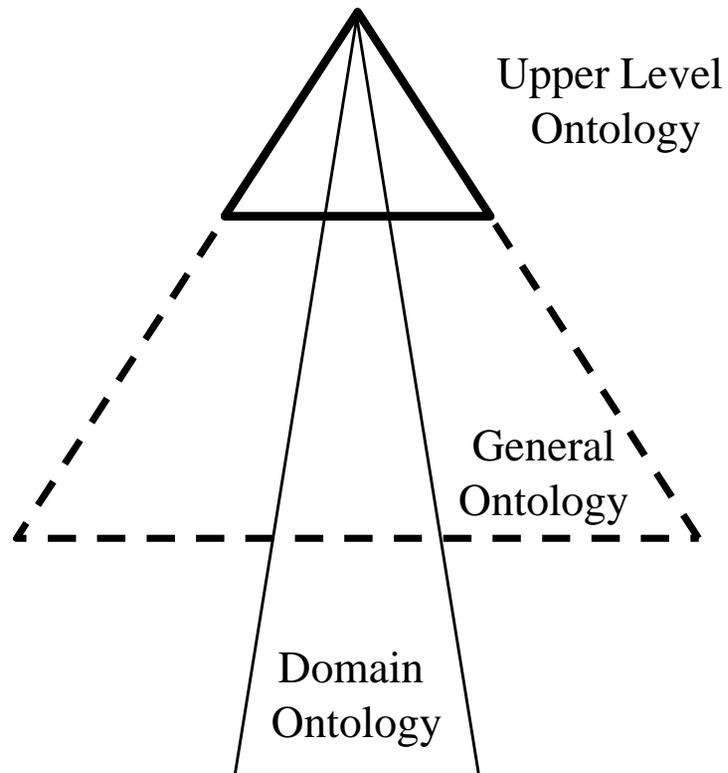
- Collections lexical items + identifiers
 - Useful for entity resolution
 - UMLS Metathesaurus

The Knowledge Semantics Continuum



Ontology Dimensions based on McGuinness and Finin

Kinds of ontologies



Application ontologies



Ontology-related issues

- ◆ Creation
- ◆ Merging
- ◆ Alignment
- ◆ Validation



Formal Ontological Principles

Formal ontological distinctions

- ◆ Universal vs. individual
- ◆ Continuant vs. occurent
- ◆ Independent vs. dependent



Universal vs. Individual

◆ Universal = *category*

◆ Synonyms

- Kind, Type, (Class)

◆ Examples

- eyeball
- blood pressure
- conference

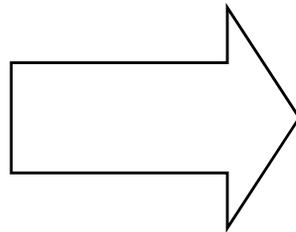
◆ Individual = *instance*

◆ Synonyms

- Particular, Token

◆ Examples

- my right eyeball
- my blood pressure (132/79)
- AMIA Annual Symposium
2003



instantiation



Continuant vs. Occurrent

- ◆ Continuant = *Continues to exist through time*
 - ◆ Synonyms
 - Substance
 - ◆ Examples
 - tennis racquet
 - mitochondrion
 - insulin production
- ◆ Occurrent = *Unfolds through time*
 - ◆ Synonyms
 - Process
 - ◆ Examples
 - tennis tournament
 - metabolism
 - producing insulin



Independent vs. Dependent

◆ Independent = *Can exist without support from other entities*

◆ Examples

- virus
- molecule
- plant

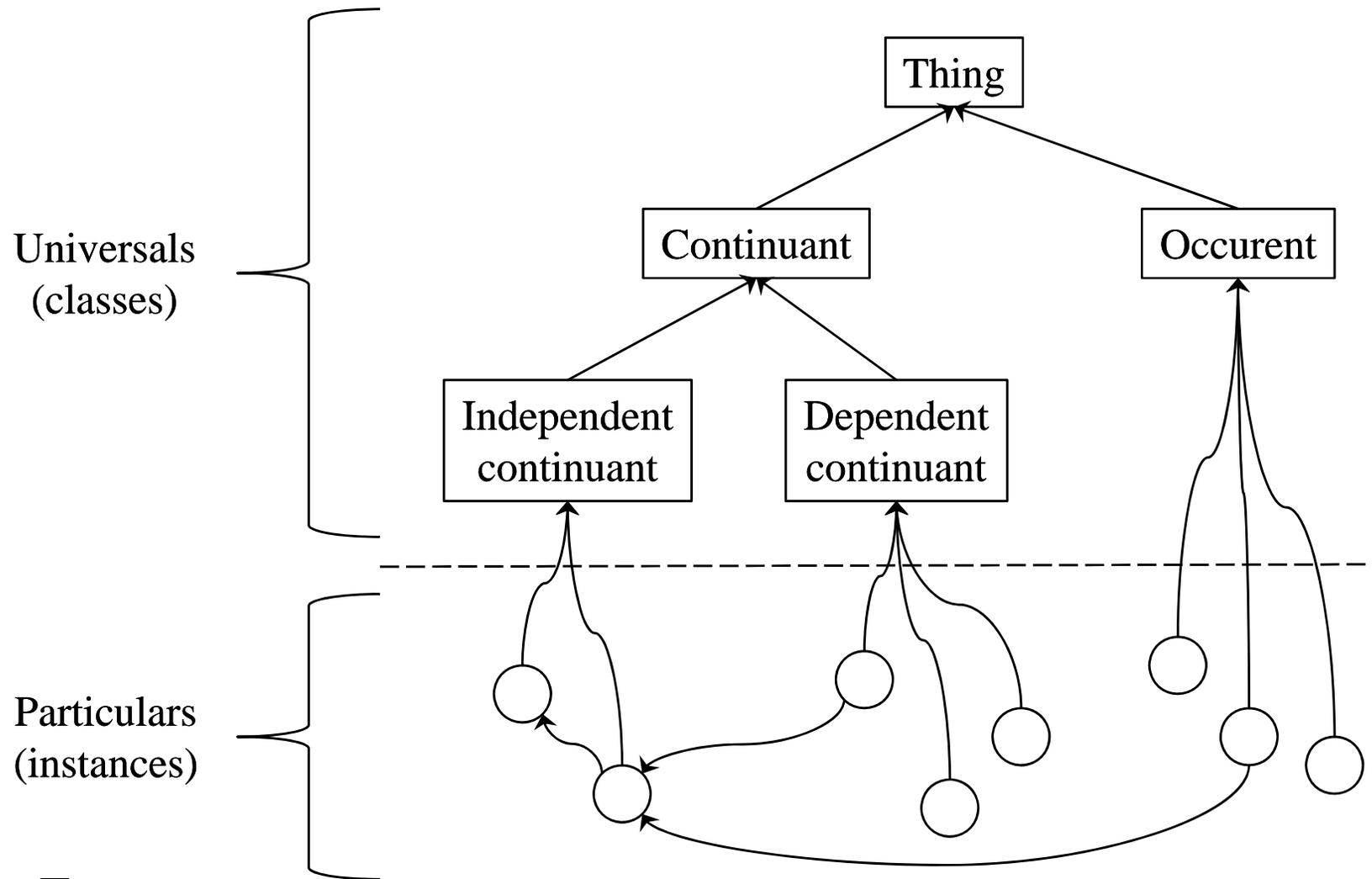
◆ Dependent = *Require support from other entities for its existence*

◆ Examples

- viral infection
- DNA binding
- food



Formal ontology Upper level



Formal ontological distinctions

- ◆ Basic distinctions in many top-level ontologies
 - Generic: BFO, DOLCE
 - Biomedical: BioTop, UMLS Semantic Network
- ◆ Condition the relations between various types of entities
 - Relations
 - Between instances (e.g., *part_of* [at time])
 - Between classes (e.g., *isa*, *part_of* [atemporal])
 - Between one instance and one class (*instance_of*)

[Smith, Genome Biology 2005]



Formal ontology in practice

- ◆ Provides foundational classes and relations
 - Upper level ontologies
 - Relation ontology
- ◆ Provides a framework for analyzing entities and relations



Examples

General ontologies

◆ OpenCyc

- General ontology
- Cycorp, Inc (D. Lenat & al.)
- Over 1M hand-coded assertions
- <http://www.opencyc.org>



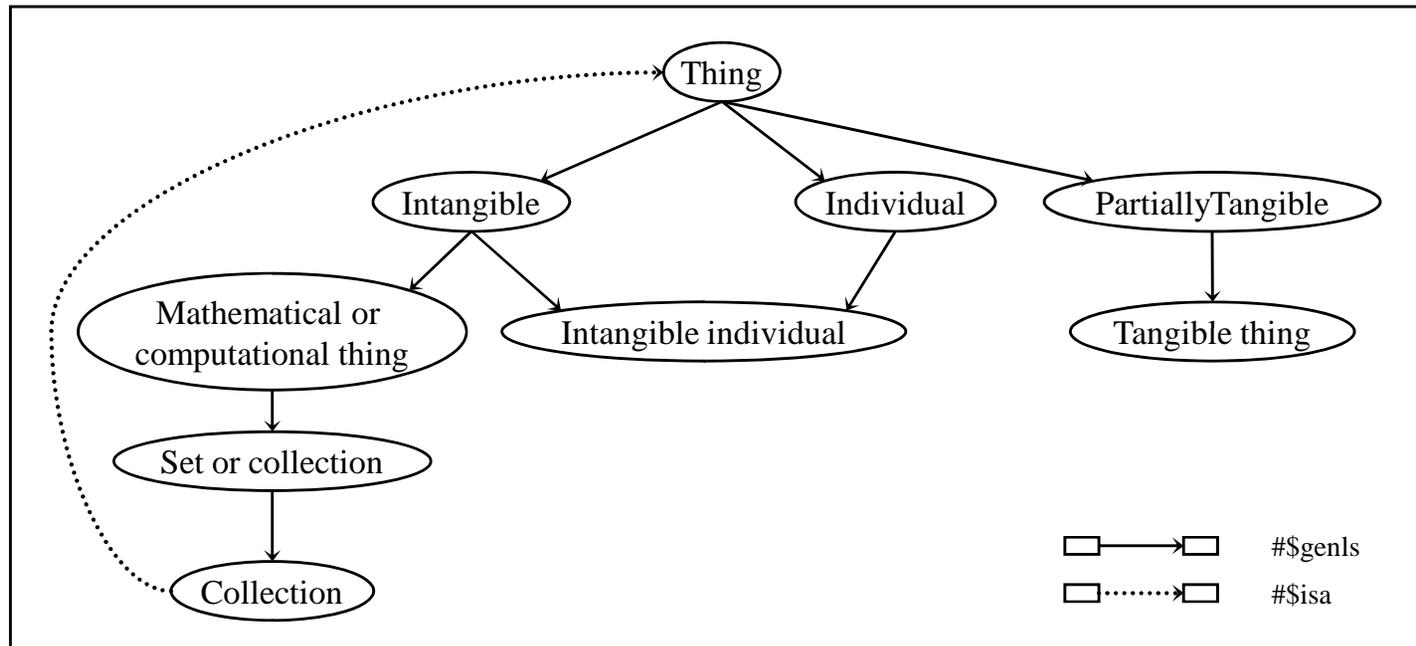
◆ WordNet

- Electronic lexical database
- Princeton University (G. Miller & al.)
- Over 100,000 synsets
- <http://wordnet.princeton.edu/>

WordNet



Top level in OpenCyc



Top level in WordNet

Abstraction
Activity
Entity
Event
Group
Location
Natural phenomenon
Possession
Psychological feature
Shape
State



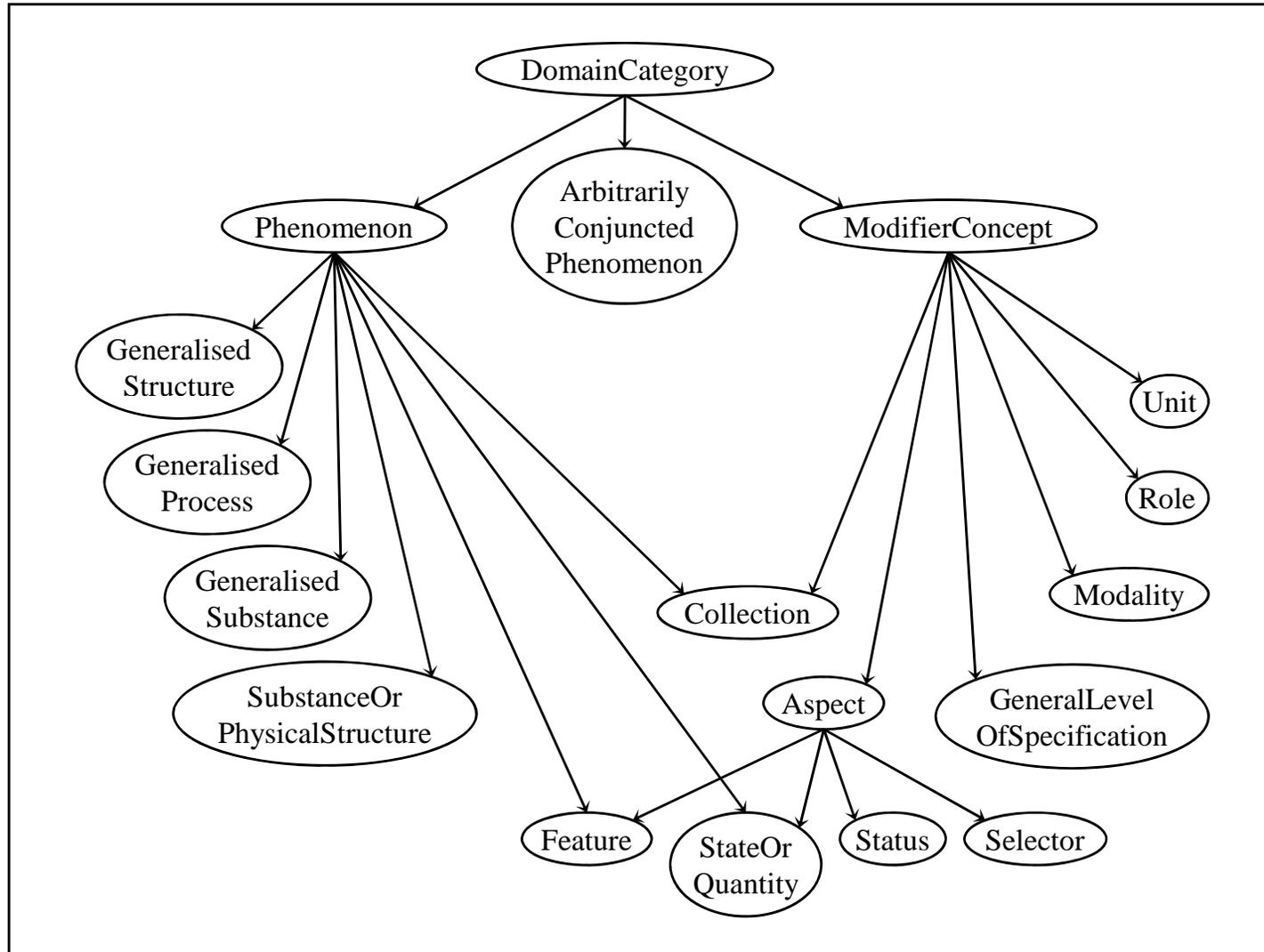
GALEN



- ◆ Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in Medicine
- ◆ European Union project (A. Rector & al.)
- ◆ Over 25,000 concepts (primitives)
- ◆ <http://www.opengalen.org>



Top level in GALEN



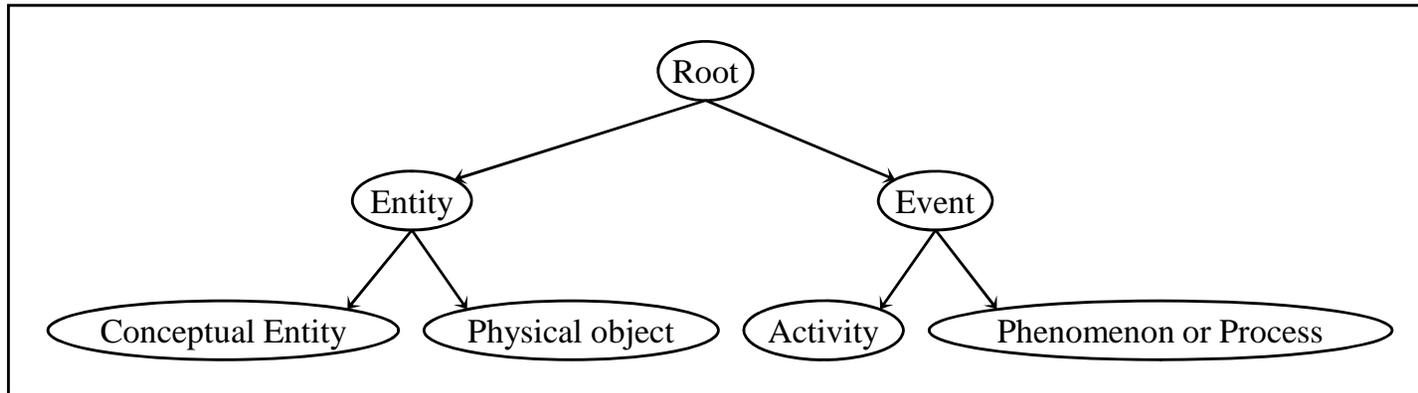
UMLS Semantic Network



- ◆ Definitional knowledge in the biomedical domain
- ◆ NLM (A. McCray & al.)
- ◆ Content
 - 135 semantic types
 - 54 types of relationship
 - 6700 semantic relations
- ◆ <http://semanticnetwork.nlm.nih.gov>



Top level in the Semantic Network



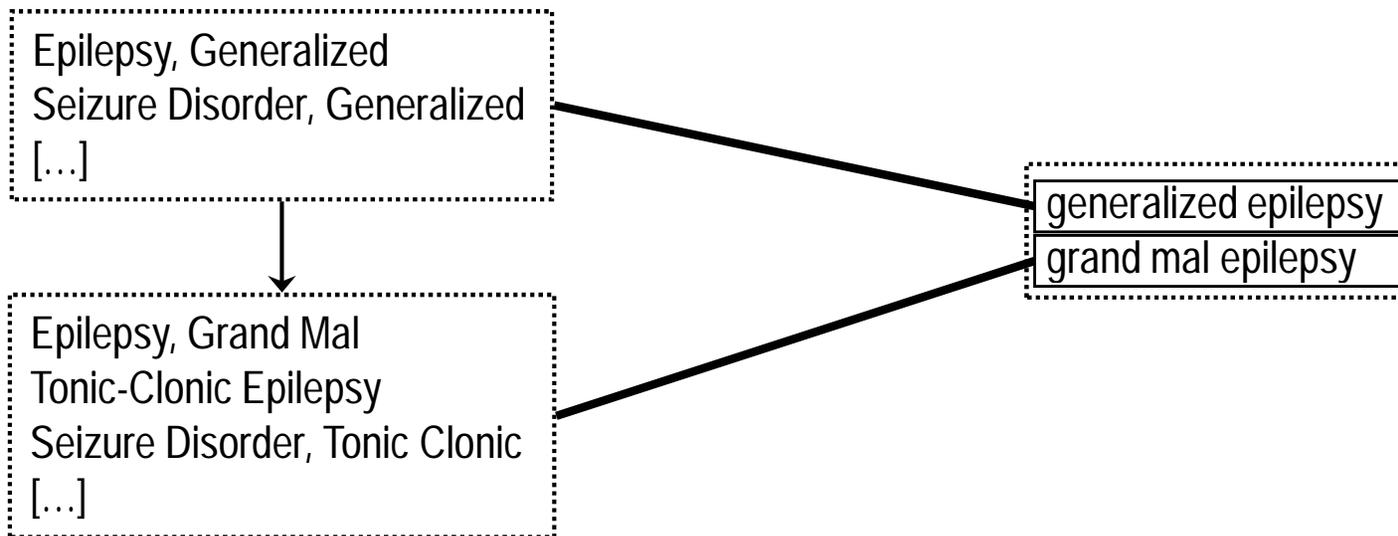
Differences between ontologies

Examples

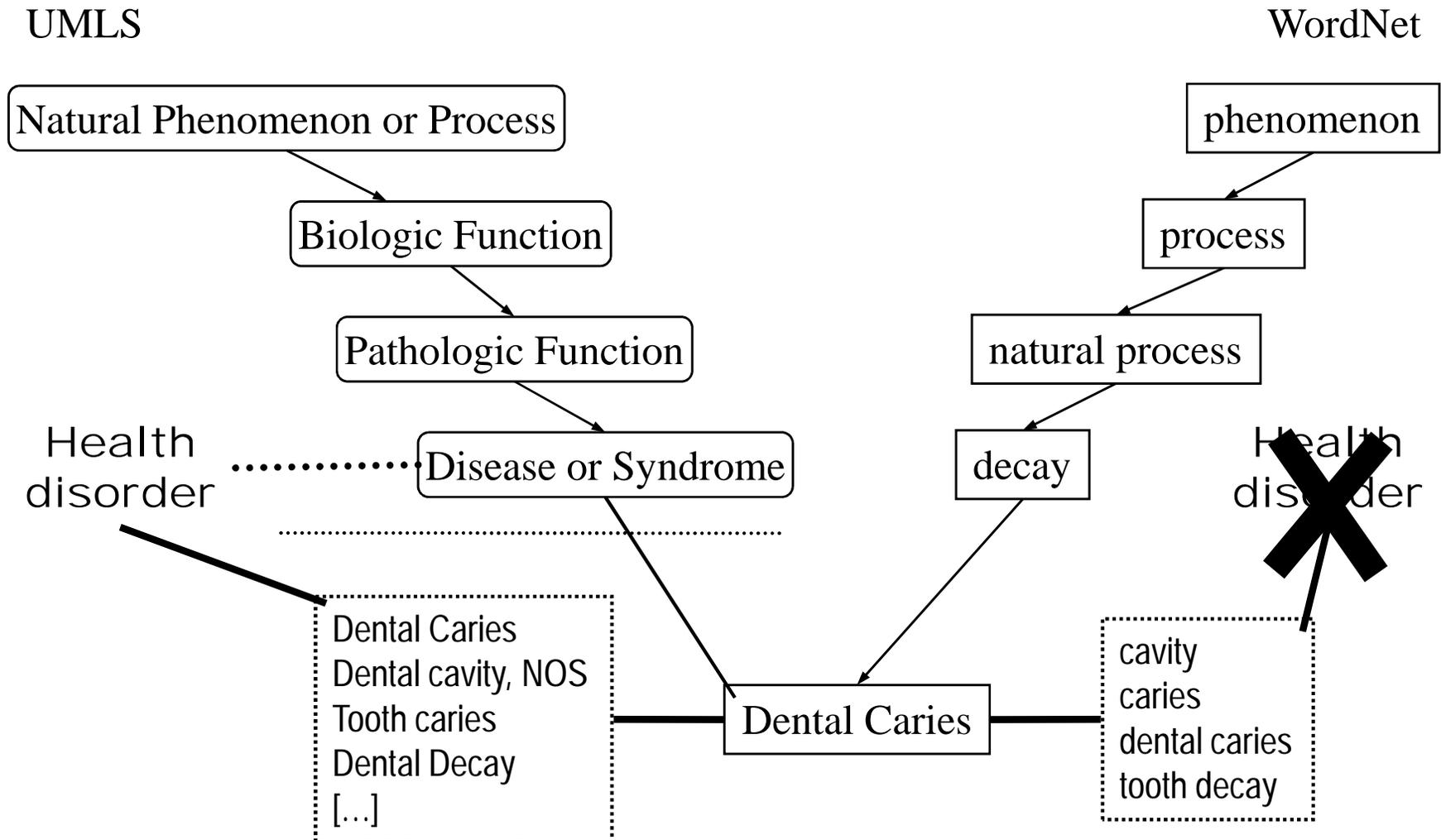
Granularity, plesionymy

UMLS

WordNet



Differing categorization



Formalisms and Tools

Ontology and Formalism

- ◆ Frames
- ◆ Description logics
 - OWL DL
- ◆ First-order logic

- ◆ OBO Format
 - Conversion to OWL DL



Tools for ontology developers

◆ Protégé

- Publicly available
- Frames and DL
- Classifier
- Supports many file formats (import/export)
- Large community of users
- Well supported
- <http://protege.stanford.edu/>



<http://protege.stanford.edu/>

◆ OBO-Edit

- Specific to the biomedical/OBO community
- Simpler than Protégé (“tool for biologists”)
- <http://oboedit.org/>



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies
			- Content and structure
			- Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9, 2008 – Session #3

“High-Impact” Biomedical Ontologies *A Structural Perspective*



Olivier Bodenreider

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

Overview

◆ Structural perspective

[J. Cimino, YBMI 2006]

- What are they (vs. what are they for)?

◆ “High-impact” biomedical ontologies

- International Classification of Diseases (ICD)
- Logical Observation Identifiers, Names and Codes (LOINC)
- SNOMED Clinical Terms
- Foundational Model of Anatomy
- Gene Ontology
- RxNorm
- Medical Subject Headings (MeSH)
- NCI Thesaurus
- Unified Medical Language System (UMLS)



International Classification of Diseases



ICD Characteristics (1)

- ◆ Current version: ICD-10
- ◆ Type: Classification
- ◆ Domain: Disorders
- ◆ Developer: World Health Organization (WHO)
- ◆ Funding: WHO
- ◆ Availability
 - Publicly available: No
 - Repositories: UMLS [ICD9-CM in NCBO BioPortal]
- ◆ URL: <http://www.who.int/classifications/icd/en/>



ICD Characteristics (2)

- ◆ Number of
 - Concepts: 12,318
 - Terms: 1 per concept (tabular)
- ◆ Major organizing principles:
 - Tree (single inheritance hierarchy)
 - No explicit classification criteria
 - Idiosyncratic inclusion/exclusion mechanism
 - .8 slots for Not elsewhere classified (NEC)
 - .9 slots for Not otherwise specified (NOS)
- ◆ Formalism: Proprietary format



ICD Top level

Chapter	Blocks	Title
<u>I</u>	<u>A00-B99</u>	Certain infectious and parasitic diseases
<u>II</u>	<u>C00-D48</u>	Neoplasms
<u>III</u>	<u>D50-D89</u>	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
<u>IV</u>	<u>E00-E90</u>	Endocrine, nutritional and metabolic diseases
<u>V</u>	<u>F00-F99</u>	Mental and behavioural disorders
<u>VI</u>	<u>G00-G99</u>	Diseases of the nervous system
<u>VII</u>	<u>H00-H59</u>	Diseases of the eye and adnexa
<u>VIII</u>	<u>H60-H95</u>	Diseases of the ear and mastoid process
<u>IX</u>	<u>I00-I99</u>	Diseases of the circulatory system
<u>X</u>	<u>J00-J99</u>	Diseases of the respiratory system
<u>XI</u>	<u>K00-K93</u>	Diseases of the digestive system
<u>XII</u>	<u>L00-L99</u>	Diseases of the skin and subcutaneous tissue
<u>XIII</u>	<u>M00-M99</u>	Diseases of the musculoskeletal system and connective tissue
<u>XIV</u>	<u>N00-N99</u>	Diseases of the genitourinary system
<u>XV</u>	<u>O00-O99</u>	Pregnancy, childbirth and the puerperium
<u>XVI</u>	<u>P00-P96</u>	Certain conditions originating in the perinatal period
<u>XVII</u>	<u>Q00-Q99</u>	Congenital malformations, deformations and chromosomal abnormalities
<u>XVIII</u>	<u>R00-R99</u>	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
<u>XIX</u>	<u>S00-T98</u>	Injury, poisoning and certain other consequences of external causes
<u>XX</u>	<u>V01-Y98</u>	External causes of morbidity and mortality
<u>XXI</u>	<u>Z00-Z99</u>	Factors influencing health status and contact with health services
<u>XXII</u>	<u>U00-U99</u>	Codes for special purposes



ICD Example

◆ Idiosyncratic inclusion/exclusion criteria

E10

Insulin-dependent diabetes mellitus

[See before E10 for subdivisions]

Includes: diabetes (mellitus):

- brittle
- juvenile-onset
- ketosis-prone
- type I

Excludes: diabetes mellitus (in):

- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)

glycosuria:

- NOS (R81)
- renal (E74.8)

impaired glucose tolerance (R73.0)

postsurgical hypoinsulinaemia (E89.1)



ICD Example

- ◆ Not elsewhere classified (NEC)
- ◆ Not otherwise specified (NOS)

E84

Cystic fibrosis

Includes: mucoviscidosis

E84.0 Cystic fibrosis with pulmonary manifestations

E84.1 Cystic fibrosis with intestinal manifestations

Meconium ileus+ (P75*)

Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)

E84.8 Cystic fibrosis with other manifestations

Cystic fibrosis with combined manifestations

E84.9 Cystic fibrosis, unspecified



Logical Observation Identifiers, Names and Codes (LOINC)



LOINC Characteristics (1)

- ◆ Current version: 2.22 (Dec. 2007)
- ◆ Type: Controlled terminology*
- ◆ Domain: Laboratory and clinical observations
- ◆ Developer: Regenstrief Institute
- ◆ Funding: NLM
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS
- ◆ URL: www.regenstrief.org/loinc/loinc.htm



LOINC Characteristics (2)

- ◆ Number of
 - Concepts: 50k active codes (2.18)
 - Terms: n/a*
- ◆ Major organizing principles:
 - No hierarchical structure among the main codes
 - 6 axes
 - Component (analyte [+ challenge] [+ adjustments])
 - Property
 - Timing
 - System
 - Scale
 - [Method]
- ◆ Formalism: “DL-like”



LOINC Example

- ◆ *Sodium:SCnc:-Pt:Ser/Plas:Qn*
[the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

Axis	Value
Component	Sodium
Property	SCnc – Substance Concentration (per volume)
Timing	Pt – Point in time (Random)
System	Ser/Plas – Serum or Plasma
Scale	Qn – Quantitative
Method	--



SNOMED Clinical Terms



SNOMED CT Characteristics (1)

- ◆ Current version: January 31, 2008 (2 annual releases)
- ◆ Type: Reference terminology / ontology
- ◆ Domain: Clinical medicine
- ◆ Developer: IHTSDO
- ◆ Funding: IHTSDO
- ◆ Availability
 - Publicly available: Yes* (in member countries)
 - Repositories: UMLS
- ◆ URL: <http://www.ihtsdo.org/>



SNOMED CT Characteristics (2)

◆ Number of

- Concepts: 311,313 active concepts (Jan. 31, 2008)
- Terms: 794,061 active “descriptions”

◆ Major organizing principles:

- Utility for clinical medicine (e.g., assertional + definitional knowledge)
- Model of meaning (incomplete)
- Rich set of associative relationships
- Small proportion of defined concepts (many primitives)

◆ Formalism: Description logics (KRSS)



SNOMED CT Top level

Hierarchy	
Subtype hierarchy	
↳	138875005 SNOMED CT Concept
+	362981000 qualifier value
+	106237007 linkage concept
+	370115009 special concept
+	48176007 social context
+	419891008 record artifact
+	363787002 observable entity
+	308916002 environment or geographical location
+	123038009 specimen
+	254291000 staging and scales
+	123037004 body structure
+	272379006 event
+	78621006 physical force
+	404684003 clinical finding
+	260787004 physical object
+	410607006 organism
+	71388002 procedure
+	373873005 pharmaceutical / biologic product
+	243796009 situation with explicit context
+	105590001 substance



SNOMED CT Example

Hierarchy Subtype hierarchy

- 27010001 partial excision of large intestine
- 8613002 operation on appendix
 - 80146002 **appendectomy**
 - 82730006 incidental appendectomy
 - 49438003 appendectomy with drainage
 - 174036004 emergency appendectomy
 - 174045003 interval appendectomy
 - 6025007 laparoscopic appendectomy
 - 235313004 non-emergency appendectomy
 - 235314005 inversion appendectomy
 - 1299000 excision of appendiceal stump

Definition: Fully defined by ...

- is a
 - D partial excision of large intestine
 - D operation on appendix
- Group
 - method
 - D excision - action
 - procedure site - Direct
 - D appendix structure
- Qualifiers
 - access
 - P surgical access values
 - priority
 - P priorities

appendectomy - Definition

Concept Status: **Current**

Descriptions

- F appendectomy (procedure)
- P appendectomy
- S excision of appendix
- U appendicectomy

Codes

- Original SnomedId : P1-57450
- Read Code (Ctv3Id) : X20Wz



Foundational Model of Anatomy

FMA Characteristics (1)

- ◆ Current version: ? (no fixed release schedule)
- ◆ Type: Ontology
- ◆ Domain: Anatomy (anatomical structures)
- ◆ Developer: U. Washington, Department of Biological Structure
- ◆ Funding: NLM (grants and contract) and others
- ◆ Availability
 - Publicly available: Yes
 - Repositories: [UMLS] / OBO / NCBO BioPortal
- ◆ URL: <http://fma.biostr.washington.edu/>



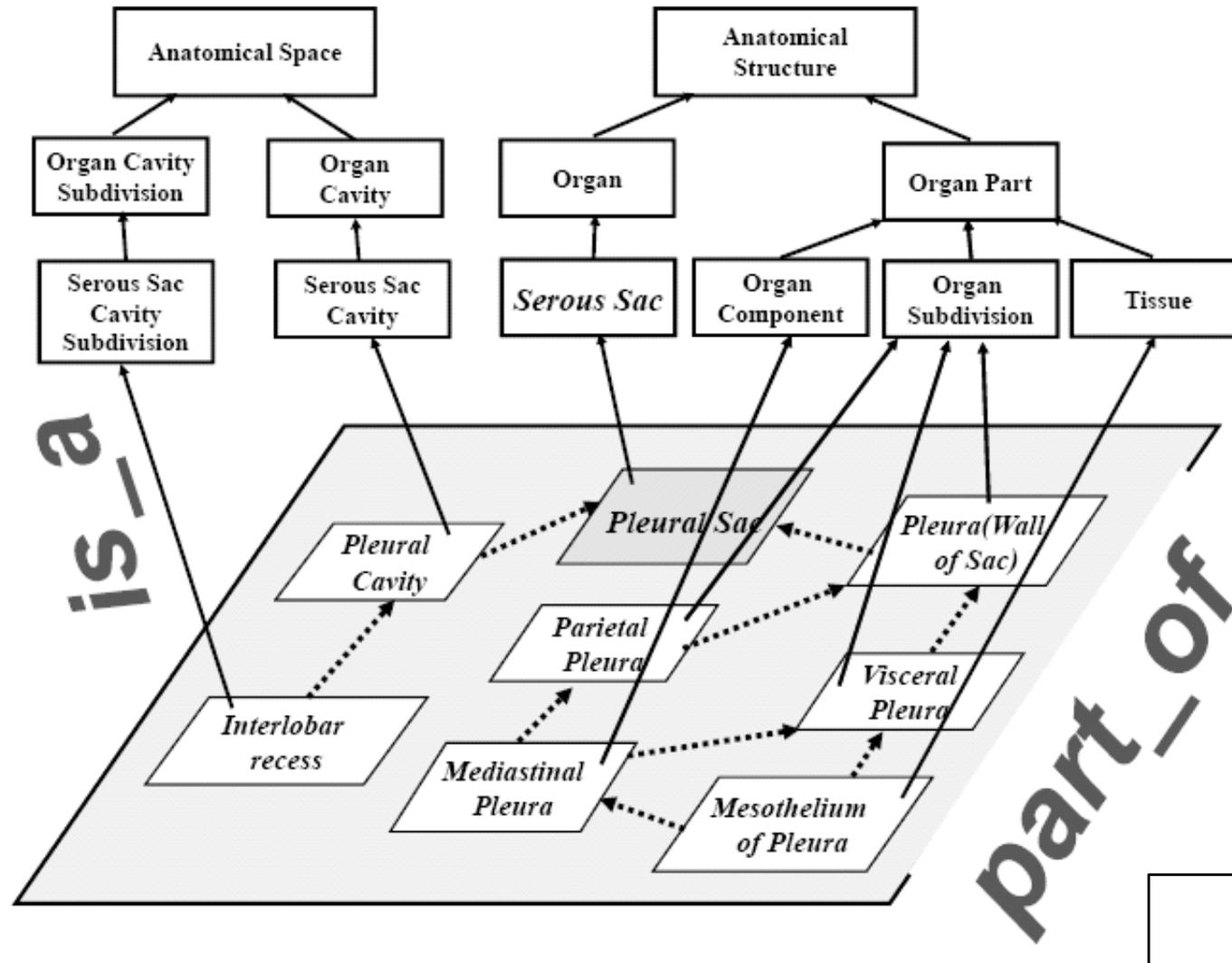
FMA Characteristics (2)

- ◆ Number of
 - Concepts: ~72k
 - Terms: ~1.5 / concept
- ◆ Major organizing principles:
 - Explicit classificatory criteria
 - Distinct *isa* and *part_of* hierarchies
 - Additional spatial relations (e.g., adjacency)
 - Multiple levels of granularity (organism to sub-cellular)
- ◆ Formalism: Frames (Protégé)
 - Conversion to OWL Full and OWL DL available



FMA Example

(Courtesy of C. Rosse)



Gene Ontology



Gene Ontology Characteristics (1)

- ◆ Current version: n/a (daily/monthly releases)
- ◆ Type: Controlled vocabulary
- ◆ Domain: Molecular biology
- ◆ Developer: GO Consortium
- ◆ Funding: NIH (grants)
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS / OBO / NCBO BioPortal
- ◆ URL: <http://www.geneontology.org/>



Gene Ontology Characteristics (2)

- ◆ Number of
 - Concepts: 22,546 (Jan. 2, 2007)
 - Terms: 2.15 per concept
- ◆ Major organizing principles:
 - 3 major hierarchies
 - Molecular function
 - Cellular component
 - Biological process
 - Relations (within hierarchies): *isa, part_of, regulates*
 - No relations between concepts across hierarchies
- ◆ Formalism: OBO format



Gene Ontology Top level (MF)

- ▣ all : all [250418 gene products] 
- ⊕ **I** GO:0008150 : biological_process [166605 gene products]
- ⊕ **I** GO:0005575 : cellular_component [169814 gene products]
- ▣ **I** **GO:0003674 : molecular_function [168558 gene products]** 
 - ⊕ **I** GO:0016209 : antioxidant activity [566 gene products]
 - ⊕ **I** GO:0015457 : auxiliary transport protein activity [161 gene products]
 - ⊕ **I** GO:0005488 : binding [46697 gene products]
 - ⊕ **I** GO:0003824 : catalytic activity [51856 gene products]
 - ⊕ **I** GO:0030188 : chaperone regulator activity [73 gene products]
 - ▣ **I** GO:0042056 : chemoattractant activity [14 gene products]
 - ▣ **I** GO:0045499 : chemorepellent activity [9 gene products]
 - ⊕ **I** GO:0030234 : enzyme regulator activity [2370 gene products]
 - ⊕ **I** GO:0016530 : metallochaperone activity [47 gene products]
 - ⊕ **I** GO:0060089 : molecular transducer activity [7873 gene products]
 - ⊕ **I** GO:0003774 : motor activity [527 gene products]
 - ▣ **I** GO:0045735 : nutrient reservoir activity [49 gene products]
 - ▣ **I** GO:0031386 : protein tag [18 gene products]
 - ⊕ **I** GO:0005198 : structural molecule activity [4324 gene products]
 - ⊕ **I** GO:0030528 : transcription regulator activity [10429 gene products]
 - ⊕ **I** GO:0045182 : translation regulator activity [893 gene products]
 - ⊕ **I** GO:0005215 : transporter activity [10583 gene products]



Gene Ontology Top level (CC)

- all : all [250418 gene products] 
-  GO:0008150 : biological_process [166605 gene products]
-  **GO:0005575 : cellular_component [169814 gene products]** 
-  GO:0005623 : cell [111086 gene products]
-  GO:0044464 : cell part [111049 gene products]
-  GO:0031975 : envelope [3316 gene products]
-  GO:0031012 : extracellular matrix [573 gene products]
-  GO:0044420 : extracellular matrix part [292 gene products]
-  GO:0005576 : extracellular region [5001 gene products]
-  GO:0044421 : extracellular region part [3365 gene products]
-  GO:0032991 : macromolecular complex [14668 gene products]
-  GO:0031974 : membrane-enclosed lumen [5290 gene products]
-  GO:0043226 : organelle [79653 gene products]
-  GO:0044422 : organelle part [16645 gene products]
-  GO:0055044 : symplast [3 gene products]
-  GO:0045202 : synapse [454 gene products]
-  GO:0044456 : synapse part [210 gene products]
-  GO:0019012 : virion [227 gene products]
-  GO:0044423 : virion part [186 gene products]
-  GO:0003674 : molecular_function [168558 gene products]



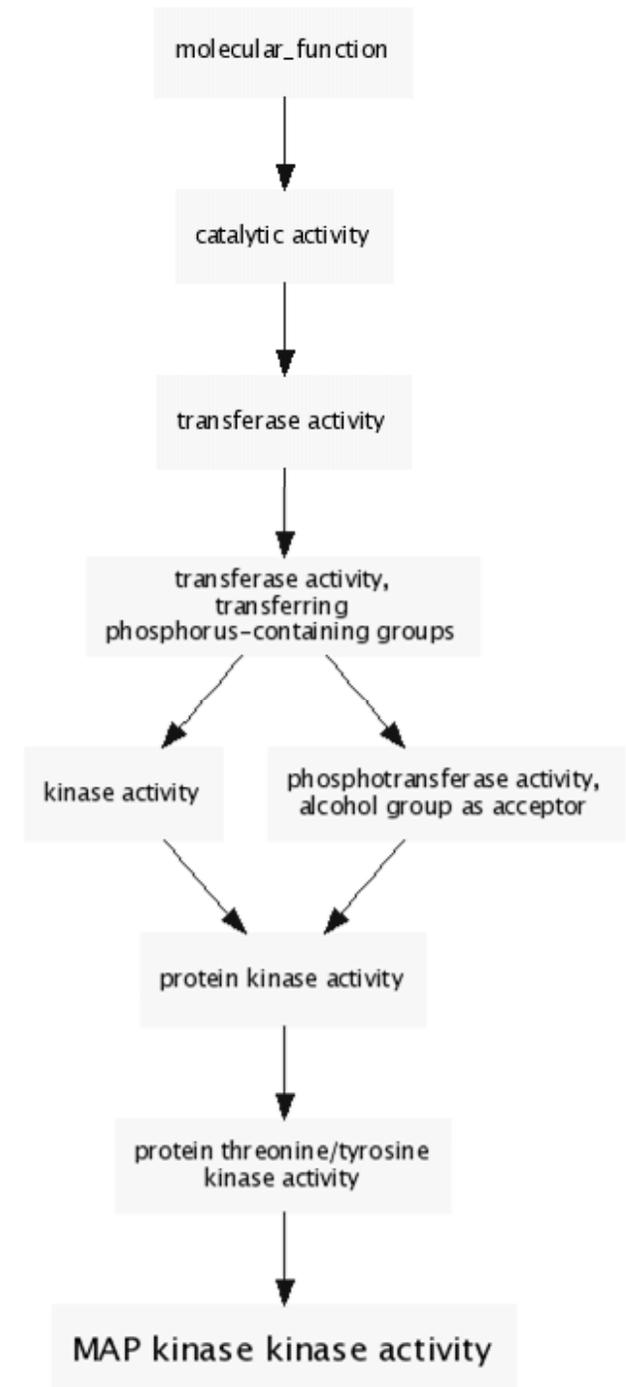
Gene Ontology Top level (BP)

- ▣ all : all [250418 gene products] 
- ▣ **I** GO:0008150 : **biological_process** [166605 gene products] 
 - ▣ **I** GO:0022610 : biological adhesion [1586 gene products]
 - ▣ **I** GO:0065007 : biological regulation [31031 gene products]
 - ▣ **I** GO:0001906 : cell killing [177 gene products]
 - ▣ **I** GO:0009987 : cellular process [79087 gene products]
 - ▣ **I** GO:0032502 : developmental process [19678 gene products]
 - ▣ **I** GO:0051234 : establishment of localization [15270 gene products]
 - ▣ **I** GO:0040007 : growth [4139 gene products]
 - ▣ **I** GO:0002376 : immune system process [2517 gene products]
 - ▣ **I** GO:0051179 : localization [17811 gene products]
 - ▣ **I** GO:0040011 : locomotion [1251 gene products]
 - ▣ **I** GO:0008152 : metabolic process [61127 gene products]
 - ▣ **I** GO:0051704 : multi-organism process [4780 gene products]
 - ▣ **I** GO:0032501 : multicellular organismal process [20567 gene products]
 - ▣ **R** GO:0048519 : negative regulation of biological process [5037 gene products]
 - ▣ **I** GO:0043473 : pigmentation [235 gene products]
 - ▣ **R** GO:0048518 : positive regulation of biological process [6585 gene products]
 - ▣ **R** GO:0050789 : regulation of biological process [28645 gene products]
 - ▣ **I** GO:0000003 : reproduction [6343 gene products]
 - ▣ **I** GO:0022414 : reproductive process [3535 gene products]
 - ▣ **I** GO:0050896 : response to stimulus [16487 gene products]
 - ▣ **I** GO:0048511 : rhythmic process [404 gene products]
 - ▣ **I** GO:0016032 : viral reproduction [536 gene products]



Gene Ontology Ex

- ▣ all : all [250418 gene products]
- ▣ **I** GO:0003674 : molecular_function [168558 gene products]
- ▣ **I** GO:0003824 : catalytic activity [51856 gene products]
- ▣ **I** GO:0016740 : transferase activity [15763 gene products]
- ▣ **I** GO:0016772 : transferase activity, transferring phospho products]
- ▣ **I** GO:0016301 : kinase activity [6093 gene products]
- ▣ **I** GO:0004672 : protein kinase activity [3504 gene products]
- ▣ **I** GO:0004712 : protein serine/threonine/tyrosine kinase activity
- ▣ **I** **GO:0004708 : MAP kinase kinase activity**
- ▣ **I** GO:0016773 : phosphotransferase activity, alcohol group as acceptor
- ▣ **I** GO:0004672 : protein kinase activity [3504 gene products]
- ▣ **I** GO:0004712 : protein serine/threonine/tyrosine kinase activity
- ▣ **I** **GO:0004708 : MAP kinase kinase activity**



RxNorm

RxNorm Characteristics (1)

- ◆ Current version: June 2, 2007 (monthly releases)
- ◆ Type: Controlled terminology
- ◆ Domain: Drug names
- ◆ Developer: NLM
- ◆ Funding: NLM
- ◆ Availability
 - Publicly available: Yes*
 - Repositories: UMLS
- ◆ URL: <http://www.nlm.nih.gov/research/umls/rxnorm/>



RxNorm Characteristics (2)

- ◆ Number of
 - Concepts: 93k
 - Terms: 105k
- ◆ Major organizing principles:
 - Generic vs. brand
 - Combinations of Ingredient / Form / Dose
 - No hierarchical structure
 - Links to all major US drug information sources
 - No clinical information
- ◆ Formalism: UMLS RRF format



RxNorm Normalized form

Strength

4mg/ml

Ingredient

Fluoxetine

Dose form

Oral Solution

Strength

Semantic clinical drug component

Ingredient

Ingredient

Semantic clinical drug form

Dose form

Strength

Semantic clinical drug

Ingredient

Dose form



Rx Norm Generic vs. Brand

◆ Generic

- Ingredient (IN) ←
- Clinical drug form (SCDF) ←
- Clinical drug component (SCDC) ←
- Clinical drug (SCD) ←

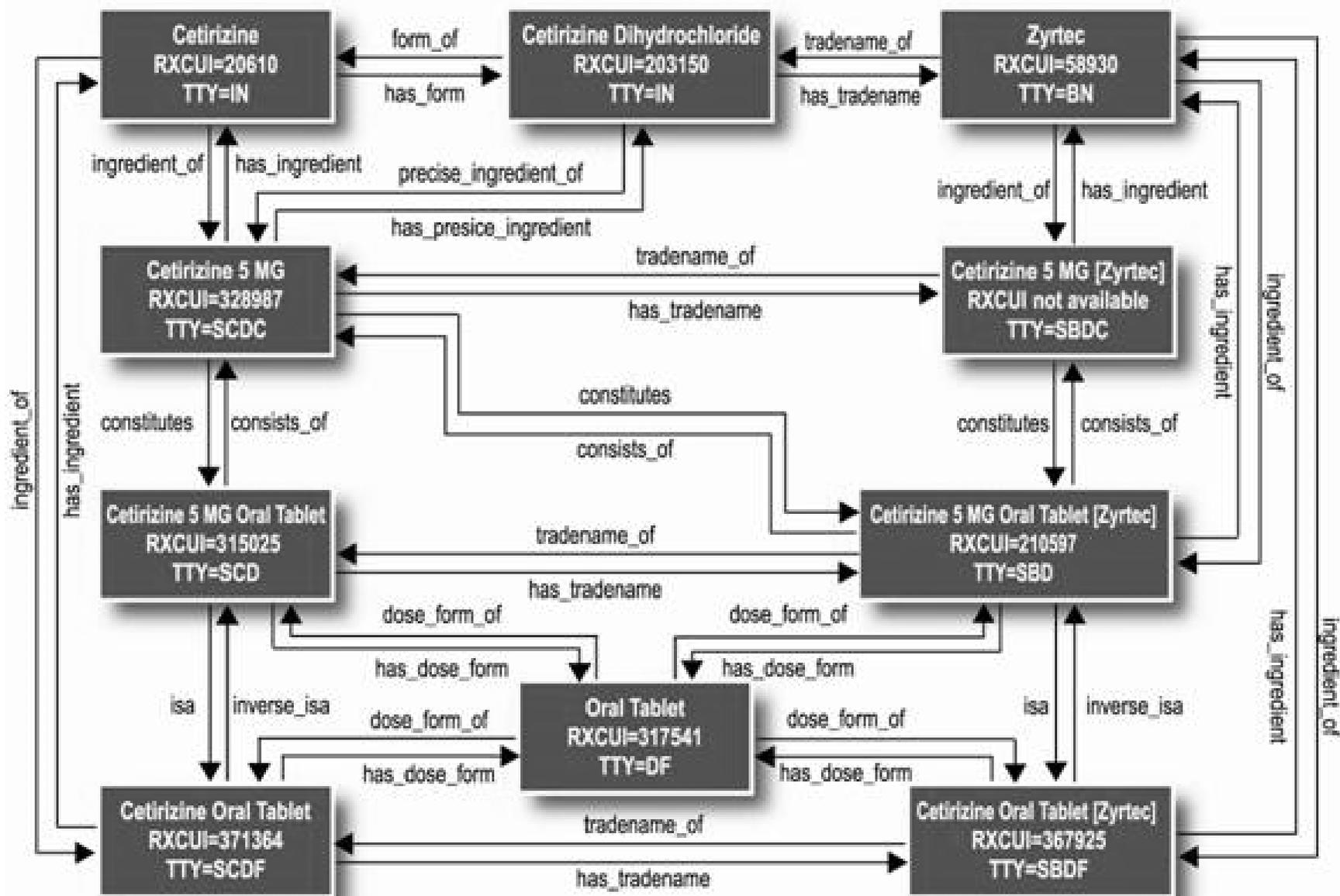
◆ Brand

- Brand name (BN)
- Branded drug form (SBDF)
- Branded drug component (SBDC)
- Branded drug (SBD)

tradename_of



RxNorm Relations among drug entities



Medical Subject Headings (MeSH)



MeSH Characteristics (1)

- ◆ Current version: 2008 (yearly releases)
- ◆ Type: Thesaurus / Controlled vocabulary
- ◆ Domain: Biomedicine
- ◆ Developer: NLM
- ◆ Funding: NLM (Library Operations)
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS / NCBO BioPortal
- ◆ URL: <http://www.nlm.nih.gov/mesh/>



MeSH Characteristics (2)

- ◆ Number of
 - Concepts: 24,767 descriptors (2007)
 - Terms: 7.5 per descriptor
- ◆ Major organizing principles:
 - Descriptor + entry terms
(also: Qualifiers, Supplementary concepts)
 - Thesaurus relations (RB/RN/RO)
- ◆ Formalism: Thesaurus / Proprietary XML DTD



MeSH Top level

1. **+** Anatomy [A]
2. **+** Organisms [B]
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. **+** Natural Sciences [H]
9. **+** Anthropology, Education, Sociology and Social Phenomena [I]
10. **+** Technology, Industry, Agriculture [J]
11. **+** Humanities [K]
12. **+** Information Science [L]
13. **+** Named Groups [M]
14. **+** Health Care [N]
15. **+** Publication Characteristics [V]
16. **+** Geographicals [Z]

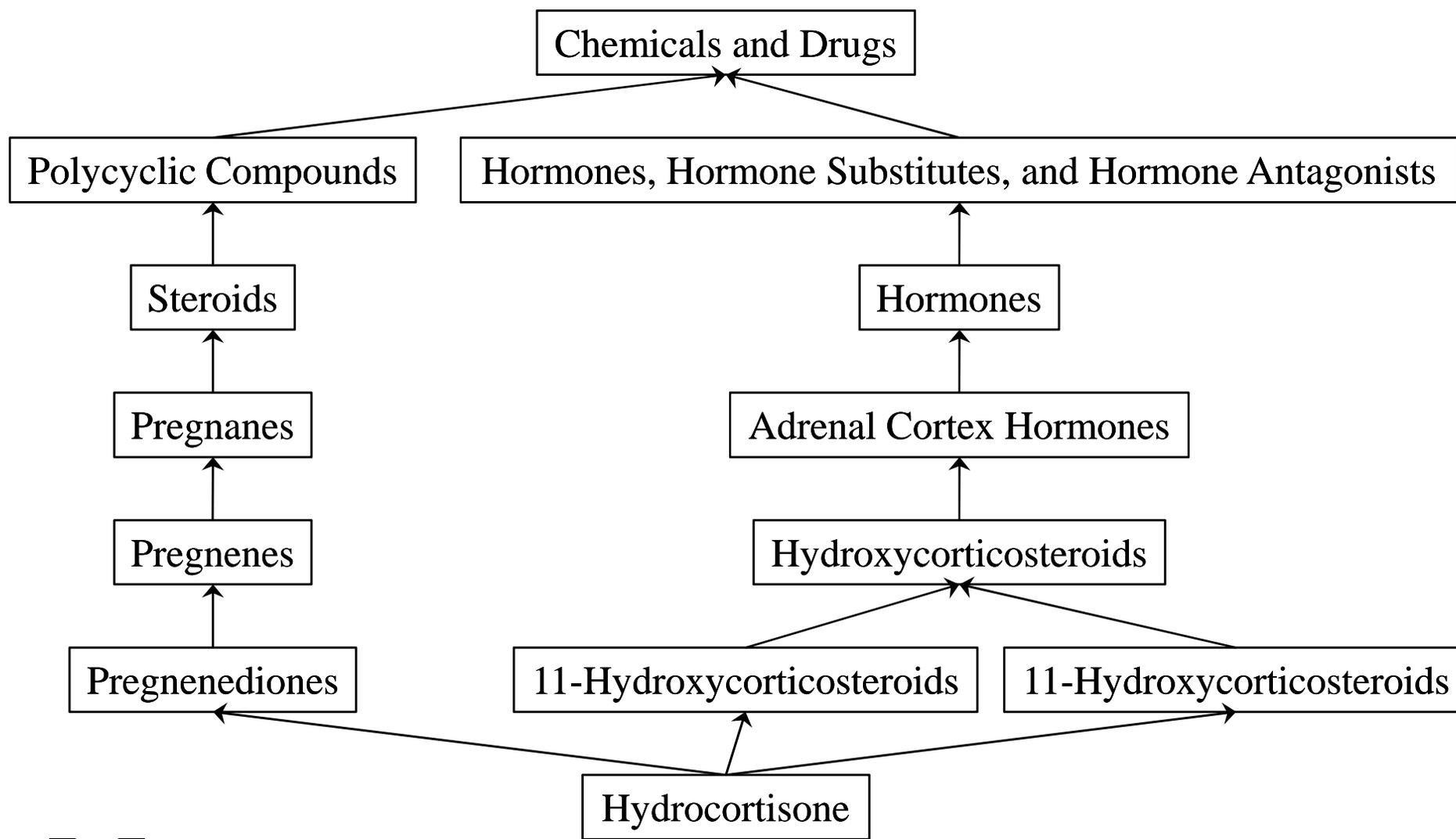


MeSH Example (terms)

MeSH Heading	Hydrocortisone
Tree Number	<u>D04.808.745.745.654.600</u>
Tree Number	<u>D06.472.040.585.353.476</u>
Tree Number	<u>D06.472.040.585.478.392</u>
Scope Note	The main glucocorticoid secreted by the <u>ADRENAL CORTEX</u> . Its synthetic counterpart is used, either as an injection or topically, in the treatment of inflammation, allergy, collagen diseases, asthma, adrenocortical deficiency, shock, and some neoplastic conditions.
Entry Term	11-Epicortisol
Entry Term	Cortifair
Entry Term	Cortisol
Entry Term	Cortril
Entry Term	Epicortisol
Entry Term	Hydrocortisone, (11 alpha)-Isomer
Entry Term	Hydrocortisone, (9 beta,10 alpha,11 alpha)-Isomer



MeSH Example (hierarchies)



NCI Thesaurus



NCI thesaurus Characteristics (1)

- ◆ Current version: 08.04d (~monthly releases)
- ◆ Type: Controlled terminology / ontology
- ◆ Domain: Cancer
- ◆ Developer: NCI Center for Bioinformatics
- ◆ Funding: NCI
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS / OBO / NCBO BioPortal
- ◆ URL: <http://ncitersms.nci.nih.gov/>



NCI thesaurus Characteristics (2)

◆ Number of

- Concepts: 58,868 (2007_05E)
- Terms: 2.68 per concept

◆ Major organizing principles:

- Subsumption hierarchy
- Rich set of associative relationships
- Small proportion of defined concepts (many primitives)
- Links to many external resources

◆ Formalism: OWL Lite



NCI thesaurus Top level

NCI_Thesaurus Taxonomy

- ☰ ⊕ Abnormal Cell
- ☰ ⊕ Activity
- ☰ ⊕ Anatomic Structure, System, or Substance
- ☰ ⊕ Biochemical Pathway
- ☰ ⊕ Biological Process
- ☰ ⊕ Chemotherapy Regimen or Agent Combination
- ☰ ⊕ Conceptual Entity
- ☰ ⊕ Diagnostic, Therapeutic, and Research Equipment
- ☰ ⊕ Diagnostic or Prognostic Factor
- ☰ ⊕ Disease, Disorder or Finding
- ☰ ⊕ Drug, Food, Chemical or Biomedical Material
- ☰ ⊕ Experimental Organism Anatomical Concept
- ☰ ⊕ Experimental Organism Diagnosis
- ☰ ⊕ Gene
- ☰ ⊕ Gene Product
- ☰ ⊕ Molecular Abnormality
- ☰ ⊕ NCI Administrative Concept
- ☰ ⊕ Organism
- ☰ ⊕ Property or Attribute
- ☰ ⊕ Retired Concept



NCI thesaurus Example

Concept Details

URI: http://nciterns.nci.nih.gov:80/NCIBrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&code=C2919
 Version: June 2007 (07.06d)

Prostate Adenocarcinoma

Identifiers:

name	Prostate_Adenocarcinoma
code	C2919

Relationships to other concepts:

Disease_Has_Finding	Invasive Lesion
Disease_Has_Abnormal_Cell	Adenocarcinoma Cell
Disease_Has_Normal_Tissue_Origin	Prostatic Epithelium
Disease_May_Have_Finding	Serum Prostate Specific Antigen Increased
Disease_Has_Finding	Carcinomatous Component Present
Disease_Excludes_Abnormal_Cell	Neoplastic Smooth Muscle Cell
Disease_Excludes_Abnormal_Cell	Malignant Squamous Cell
Disease_Has_Primary_Anatomic_Site	Prostate Gland
Disease_Has_Associated_Anatomic_Site	Male Reproductive System
Disease_Excludes_Abnormal_Cell	Malignant Stromal Cell
Disease_Has_Associated_Anatomic_Site	Prostate Gland
Disease_Has_Normal_Cell_Origin	Epithelial Cell

Information about this concept:

DEFINITION

Synonym with source data
 Synonym with source data
 Synonym with source data

Preferred_Name
 Semantic_Type

Synonym

Synonym

Synonym

Unified Medical Language System Concept Identifier

Superconcepts:

Adenocarcinoma
 Common Carcinoma
 Invasive Prostate Carcinoma

Subconcepts:

Acinar Prostate Adenocarcinoma
 Metastatic Prostatic Adenocarcinoma
 Moderately Differentiated Prostate Adenocarcinoma
 Poorly Differentiated Prostate Adenocarcinoma
 Prostate Adenocarcinoma with Focal Neuroendocrine Differentiation
 Prostate Ductal Adenocarcinoma
 Stage III Prostate Adenocarcinoma
 Stage II Prostate Adenocarcinoma
 Stage I Prostate Adenocarcinoma
 Well Differentiated Prostate Adenocarcinoma



Unified Medical Language System (UMLS)



UMLS Characteristics (1)

- ◆ Current version: 2008AA (2-3 annual releases)
- ◆ Type: Terminology integration system
- ◆ Domain: Biomedicine
- ◆ Developer: NLM
- ◆ Funding: NLM (intramural)
- ◆ Availability
 - Publicly available: Yes* (cost-free license required)
 - Repositories: UMLS
- ◆ URL: <http://umlsks.nlm.nih.gov/>

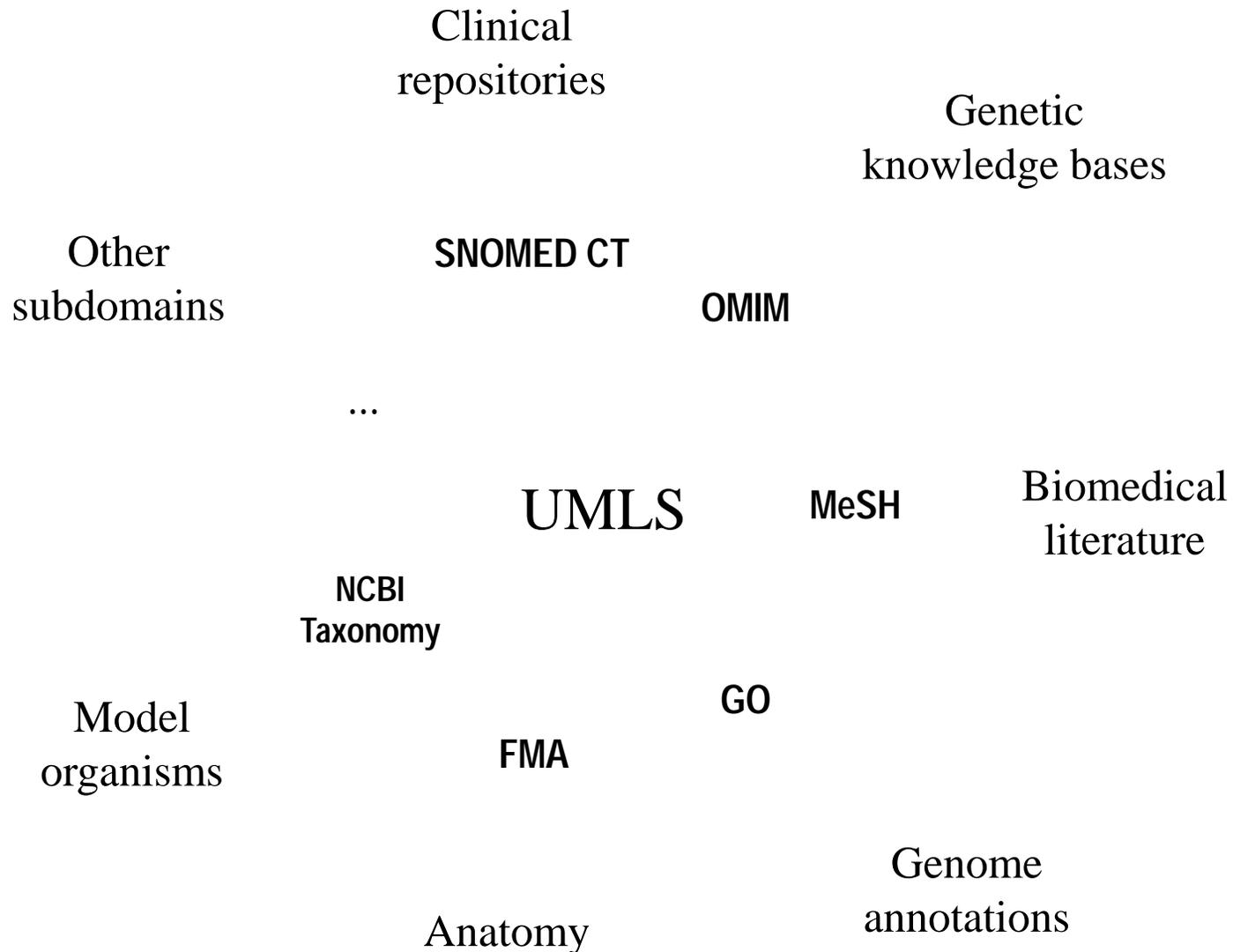


UMLS Characteristics (2)

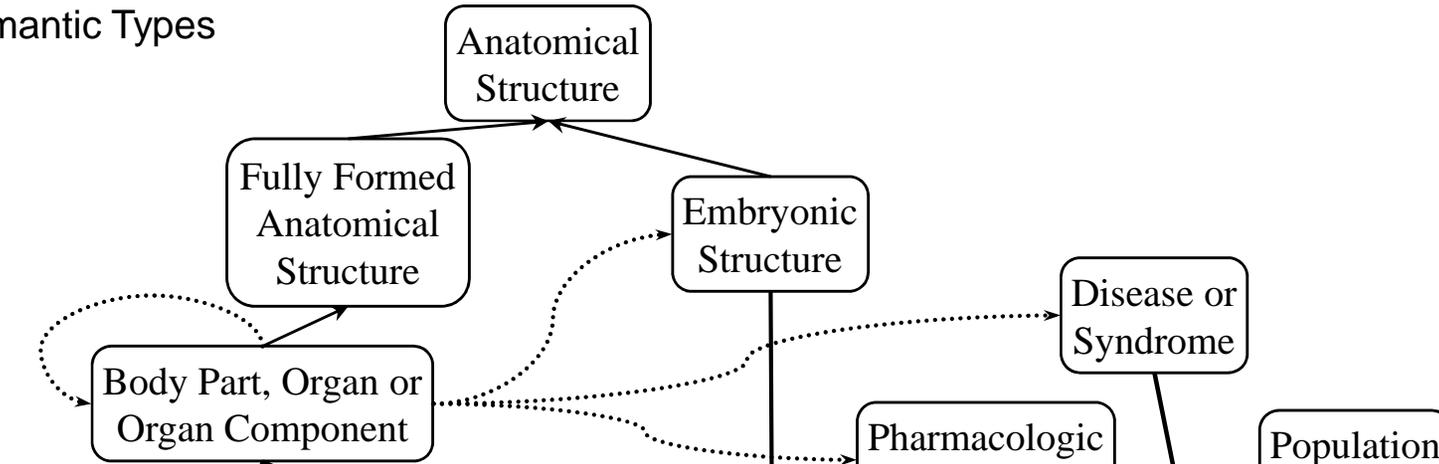
- ◆ Number of
 - Concepts: 1.5M (2008AA)
 - Terms: ~6M
- ◆ Major organizing principles (Metathesaurus):
 - Concept orientation
 - Source transparency
 - Multi-lingual through translation
- ◆ Formalism: Proprietary format (RRF)



UMLS Integrating subdomains



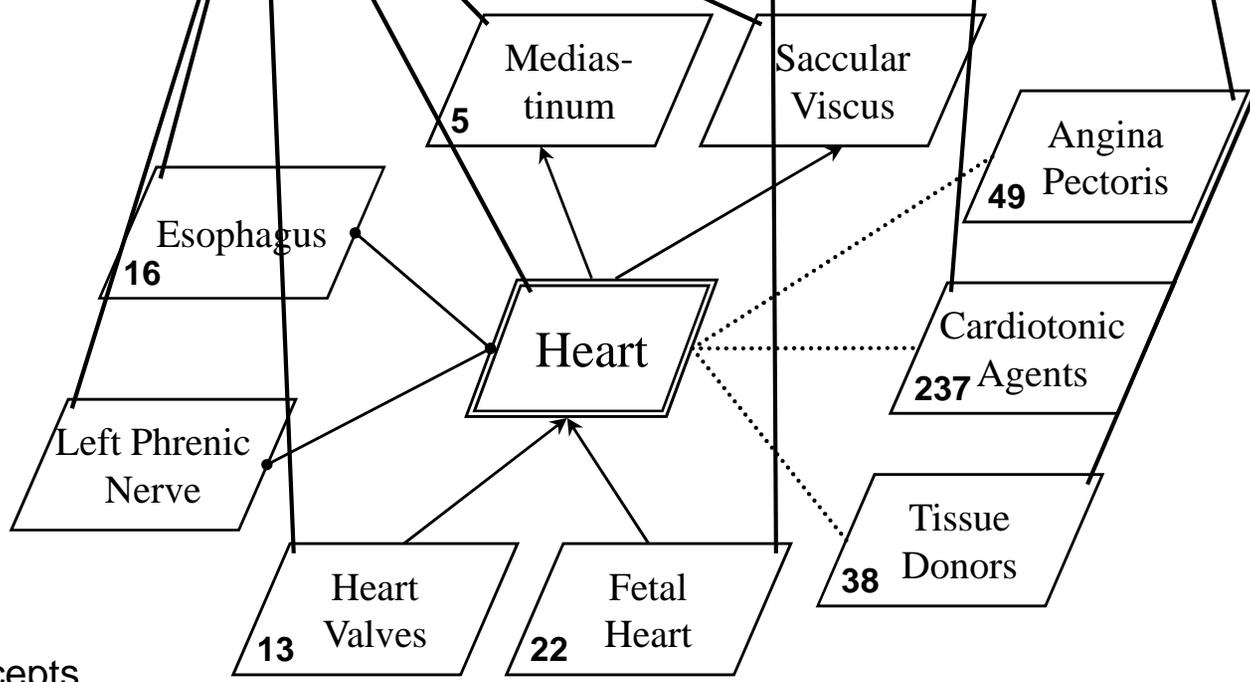
Semantic Types



Semantic Network

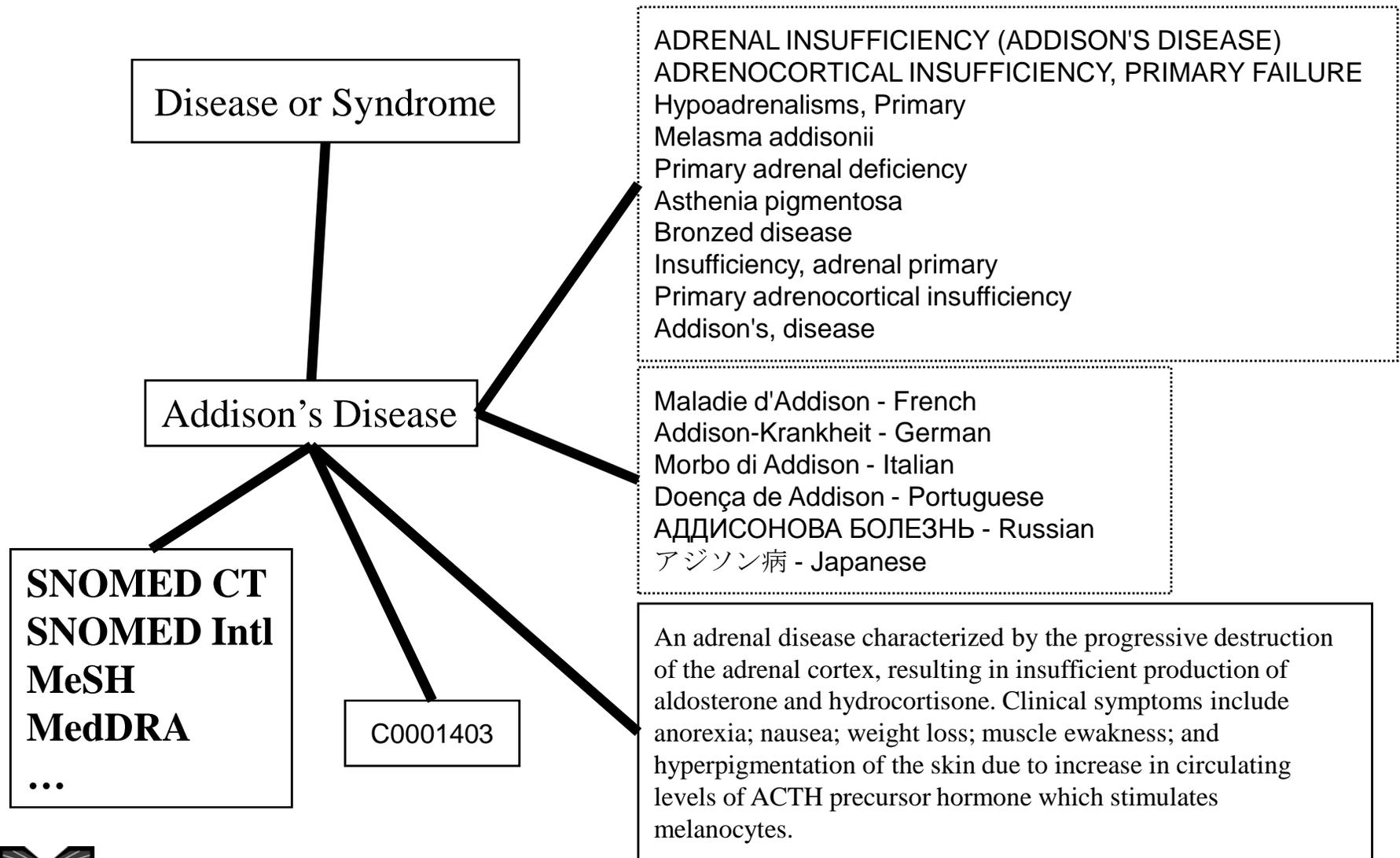


Metathesaurus



Concepts

Addison's Disease: Concept



Metathesaurus Concepts (2007AB)

- ◆ **Concept (~ 1.4M) CUI**
 - Set of synonymous concept names
- ◆ **Term (~ 5.3 M) LUI**
 - Set of normalized names
- ◆ **String (~ 5.9M) SUI**
 - Distinct concept name
- ◆ **Atom (~ 7.2M) AUI**
 - Concept name in a given source

A0066000	Headache	(MeSH)
A0065992	Headache	(ICD-10)
	S0046854	

A0066007	Headaches	(MedDRA)
A12003304	Headaches	(OMIM)
	S0046855	

L0018681

A0540936	Cephalodynia	(MeSH)
	S0475647	

L0380797

C0018681



Recap

Name	Scope	# concepts	Median	Subs. Hier	Version
SNOMED CT	Clinical medicine (patient records)	310,314	2	yes	July 31, 2007
LOINC	Clinical observations and laboratory tests	46,406	3	no	Version 2.21 (no “natural language” names)
FMA	Human anatomical structures	~72,000	?	yes	(not yet in the UMLS)
Gene Ontology	Functional annotation of gene products	22,546	1	yes	Jan. 2, 2007
RxNorm	Standard names for prescription drugs	93,426	1	no	Aug. 31, 2007
NCI Thesaurus	Cancer research, clinical care, public information	58,868	2	yes	2007_05E
ICD-10	Diseases and conditions (health statistics)	12,318	1	no	1998 (tabular)
MeSH	Biomedicine (descriptors for indexing the literature)	24,767	5	no	Aug. 27, 2007
UMLS .	Terminology integration in the life sciences	1,4 M	2	n/a	2007AC (English only)

Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure
			- Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9, 2008 – Session #4

Biomedical Ontologies in Action

A Functional Perspective on Biomedical Ontologies



Olivier Bodenreider

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

Overview

◆ Functional perspective

[Bodenreider, YBMI 2008]

- What are they for (vs. what are they)?

◆ “High-impact” biomedical ontologies

◆ 3 major categories of use

- Knowledge management (indexing and retrieval of data and information, access to information, mapping among ontologies)
- Data integration, exchange and semantic interoperability
- Decision support and reasoning (data selection and aggregation, decision support, natural language processing applications, knowledge discovery).



Knowledge management

Knowledge management

Annotating data and resources

Terminology in ontology

- ◆ **Ontology as a source of vocabulary**
 - List of names for the entities in the ontology (ontology vs. terminology)
- ◆ **Most ontologies have some sort of terminological component**
 - Exceptions: GALEN, LOINC
- ◆ **Not all surface forms represented**
 - Often insufficient for NLP applications
 - Large variation in number of terms per concept across ontologies



Annotating data

◆ Gene Ontology

- Functional annotation of gene products in several dozen model organisms



◆ Various communities use the same controlled vocabularies

◆ Enabling comparisons across model organisms

◆ Annotations

- Assigned manually by curators
- Inferred automatically (e.g., from sequence similarity)



GO Annotations for Aldh2 (mouse)

GO Annotations in Tabular Form (Text View) (GO Graph



Category	Classification Term	Evidence
Molecular Function	<u>aldehyde dehydrogenase (NAD) activity</u>	IEA
Molecular Function	<u>oxidoreductase activity</u>	IEA
Molecular Function	<u>oxidoreductase activity</u>	IEA
Cellular Component	<u>mitochondrion</u>	IDA
Biological Process	<u>metabolic process</u>	IEA
Biological Process	<u>oxidation reduction</u>	IEA

[http:// www.informatics.jax.org/](http://www.informatics.jax.org/)



GO ALD4 in Yeast

GO Annotations

Molecular Function

Manually curated

Biological Process

Manually curated

Cellular Component

Manually curated

High-throughput

All **ALD4** GO evidence and references

*View Computational GO annotations for **ALD4***

- aldehyde dehydrogenase (NAD) activity (IDA, IMP, ISS)
- aldehyde dehydrogenase [NAD(P)+] activity (IDA)

- ethanol metabolic process (IMP)

- mitochondrial nucleoid (IDA)
- mitochondrion (IMP, ISS)
- mitochondrion (IDA)



<http://db.yeastgenome.org/>



GO Annotations for ALDH2 (Human)



Function

GO:0016491	oxidoreductase activity	interpro	EA	IPR015590	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	EA	IPR016160	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	EA	IPR016162	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	EA	IPR016161	UniProt	9606
GO:0016491	oxidoreductase activity	spkw	EA	KW-0560	UniProt	9606
GO:0004029	aldehyde dehydrogenase (NAD) activity	1306115	TAS		PINC	9606
GO:0004030	aldehyde dehydrogenase [NAD(P)+] activity	8903321	TAS		PINC	9606
GO:0009055	electron carrier activity	8903321	TAS		UniProt	9606
GO:0004029	aldehyde dehydrogenase (NAD) activity	enzyme	EA	1.2.1.3	UniProt	9606

<http://www.ebi.ac.uk/GOA/>



Indexing the biomedical literature

◆ MeSH

- Used for indexing and retrieval of the biomedical literature (MEDLINE)



◆ Indexing

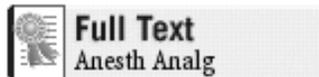
- Performed manually by human indexers
 - With help of semi-automatic systems (suggestions)
e.g., Indexing Initiative at NLM
- Automatic indexing systems



MeSH MEDLINE indexing

□ 1: [Anesth Analg](#). 2008 Jun;106(6):1813-9.

[Related Articles,](#)
[Links](#)



Free cortisol in sepsis and septic shock.

Bendel S, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E, Finnsepsis Study Group.

▶ [Collaborators \(26\)](#)

Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. **METHODS:** In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. **RESULTS:** Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90$, $P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 ± 386 nmol/L vs 793 ± 439 nmol/L, $P = 0.44$). Nonsurvivors had higher calculated serum free (209 ± 151 nmol/L) and total (980 ± 458 nmol/L) cortisol concentrations than survivors (119 ± 111 nmol/L, $P = 0.002$, and 704 ± 383 nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%.

CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.



MeSH MEDLINE indexing

MeSH Terms:

- ◆ Adrenal Cortex Function Tests
- ◆ Adrenal Insufficiency/blood*
- ◆ Adrenal Insufficiency/drug therapy
- ◆ Adrenal Insufficiency/mortality
- ◆ Adult
- ◆ Biological Markers/blood
- ◆ Female
- ◆ Finland/epidemiology
- ◆ Hospital Mortality
- ◆ Humans
- ◆ Hydrocortisone/blood*
- ◆ Hydrocortisone/therapeutic use
- ◆ Kaplan-Meiers Estimate

- ◆ Male
- ◆ Predictive Value of Tests
- ◆ Prospective Studies
- ◆ Sepsis/blood*
- ◆ Sepsis/drug therapy
- ◆ Sepsis/mortality
- ◆ Severity of Illness Index
- ◆ Shock, Septic/blood*
- ◆ Shock, Septic/drug therapy
- ◆ Shock, Septic/mortality
- ◆ Treatment Outcome

Substances:

- ◆ Biological Markers
- ◆ Hydrocortisone



MeSH MEDLINE indexing

□ 1: [Expert Opin Investig Drugs](#). 2008 Apr;17(4):497-509.

[Related Articles,
Links](#)



Replacement therapy for Addison's disease: recent developments.

Lovås K, Husebye ES.

University of Bergen, Institute of Medicine, Section of Endocrinology, 5021 Bergen, Norway.
Kristian.lovås@helse-bergen.no

BACKGROUND: The hormone deficiencies in Addison's disease (primary adrenal insufficiency) are conventionally treated with oral glucocorticoid and mineralocorticoid replacement but the available therapies do not restore the physiological hormone levels and biorhythm. Despite such treatment these patients self-report impaired health-related quality of life (HRQoL) and recent research has indicated increased mortality. **OBJECTIVE/METHODS:** We review the literature and recent developments in replacement therapy. **RESULTS/CONCLUSION:** Patients with Addison's disease require mineralocorticoid replacement, i.e., fludrocortisone 0.05 - 0.20 mg once daily. Starting doses of glucocorticoids should be 15 - 20 mg for hydrocortisone or 20 - 30 mg for cortisone acetate, divided into two or three doses, and preferentially weight-adjusted. There are indications that the synthetic glucocorticoids have undesirable metabolic long-term effects, which make them less suitable as first-line treatment. Timed-release hydrocortisone tablets and continuous subcutaneous hydrocortisone infusion are promising new treatment modalities. Studies of replacement with the adrenal androgen dehydroepiandrosterone (DHEA) in adrenal failure have shown inconsistent benefit on HRQoL. DHEA, or possibly testosterone replacement is likely to be beneficial for selected groups of patients with Addison's disease but this remains to be shown. We here give our opinion of the best treatment and future direction of research in this area.

MeSH MEDLINE indexing

MeSH Terms:

- ◆ Addison Disease/blood
- ◆ Addison Disease/drug therapy*
- ◆ Androgens/administration & dosage*
- ◆ Androgens/therapeutic use
- ◆ Dosage Forms
- ◆ Drug Administration Routes
- ◆ Drug Administration Schedule
- ◆ Glucocorticoids/administration & dosage*
- ◆ Glucocorticoids/adverse effects
- ◆ Glucocorticoids/blood
- ◆ Glucocorticoids/deficiency
- ◆ Hormone Replacement Therapy*
- ◆ Humans
- ◆ Mineralocorticoids/administration & dosage*
- ◆ Mineralocorticoids/adverse effects
- ◆ Mineralocorticoids/blood
- ◆ Mineralocorticoids/deficiency
- ◆ Quality of Life
- ◆ Treatment Outcome

Substances:

- ◆ Androgens
- ◆ Dosage Forms
- ◆ Glucocorticoids
- ◆ Mineralocorticoids



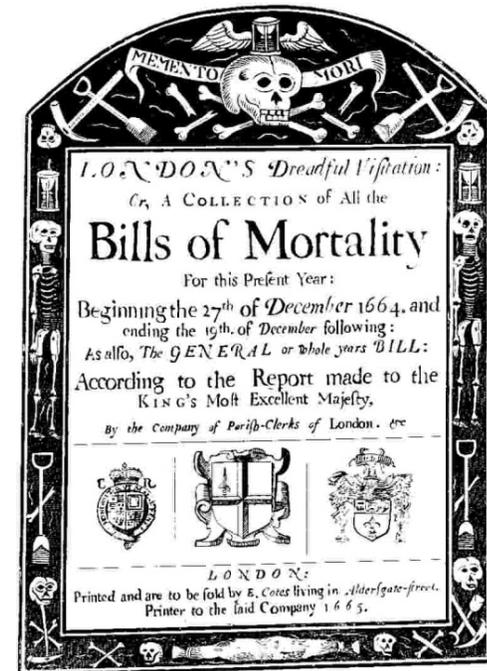
ICD9-CM Coding clinical data

◆ ICD9-CM

- Used for coding clinical data
e.g., for billing purposes

◆ Other uses of ICD

- Morbidity and mortality reporting worldwide



Knowledge management

Accessing biomedical information

Resources for biomedical search engines

- ◆ Synonyms
- ◆ Hierarchical relations
- ◆ High-level categorization
- ◆ Co-occurrence information
- ◆ Translation



MeSH “synonyms” MEDLINE retrieval

◆ MeSH entry terms

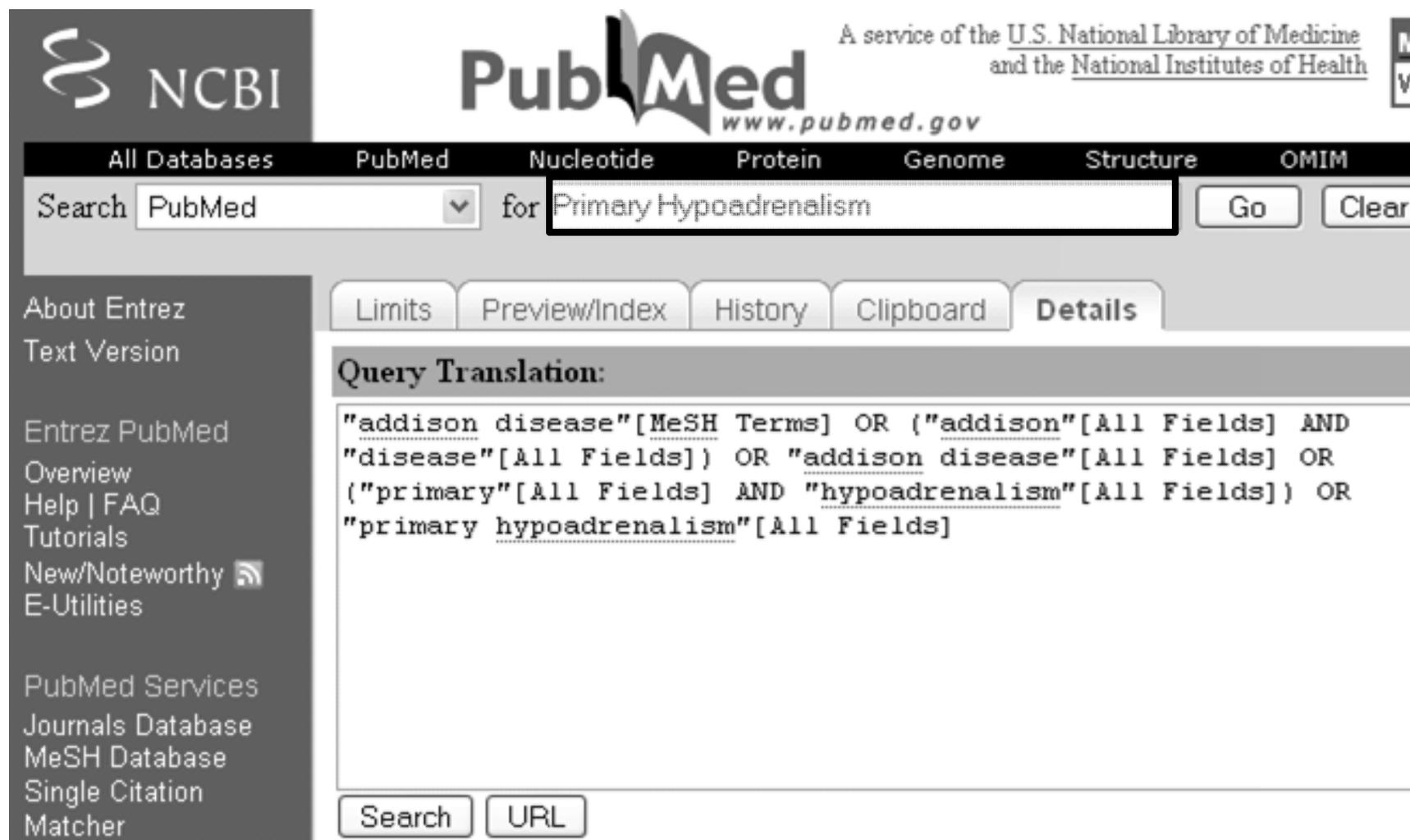
- Used as equivalent terms for retrieval purposes
- Not always synonymous

◆ Increase recall without hurting precision

MeSH Heading	Addison Disease
Entry Term	Addison's Disease
Entry Term	Primary Adrenal Insufficiency
Entry Term	Primary Adrenocortical Insufficiency



MeSH “synonyms” MEDLINE retrieval



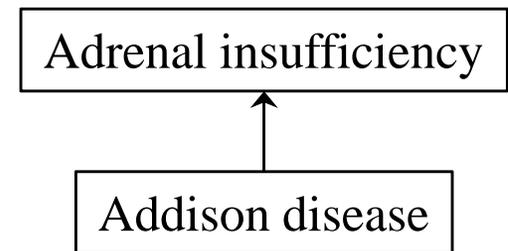
The screenshot displays the PubMed search interface. At the top, the NCBI logo is on the left, and the PubMed logo with the URL www.pubmed.gov is in the center. To the right, it states "A service of the U.S. National Library of Medicine and the National Institutes of Health". Below this is a navigation bar with tabs for "All Databases", "PubMed", "Nucleotide", "Protein", "Genome", "Structure", and "OMIM". The search bar shows "PubMed" selected in a dropdown menu, followed by "for Primary Hypoadrenalism". "Go" and "Clear" buttons are to the right. Below the search bar are tabs for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Query Translation" section shows the following text: `"addison disease"[MeSH Terms] OR ("addison"[All Fields] AND "disease"[All Fields]) OR "addison disease"[All Fields] OR ("primary"[All Fields] AND "hypoadrenalism"[All Fields]) OR "primary hypoadrenalism"[All Fields]`. At the bottom of this section are "Search" and "URL" buttons. On the left side, there is a vertical menu with links: "About Entrez", "Text Version", "Entrez PubMed", "Overview", "Help | FAQ", "Tutorials", "New/Noteworthy", "E-Utilities", "PubMed Services", "Journals Database", "MeSH Database", "Single Citation", and "Matcher".



MeSH hierarchies MEDLINE retrieval

◆ MeSH “explosion”

- Search for a given MeSH term and all its descendants
- A search on Adrenal insufficiency also retrieves articles indexed with Addison disease



Search PubMed [MeSH Terms] [Advanced Search \(beta\)](#)
[Save Search](#)

Display Summary Show 20 Sort By Send to

All: 8994 Review: 1069 

Items 1 - 20 of 8994

Page of 450 Next

1: [Bendel S, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E; Finnsepsis Study Group.](#) Related Articles, Links

 Free cortisol in sepsis and septic shock.
Anesth Analg. 2008 Jun;106(6):1813-9.
PMID: 18499615 [PubMed - indexed for MEDLINE]

2: [Luboshitzky R, Qupti G.](#) Related Articles, Links

 Corticosteroids for septic shock.
N Engl J Med. 2008 May 8;358(19):2069; author reply 2070-1. No abstract available.
PMID: 18467975 [PubMed - indexed for MEDLINE]



12: [Løvås K, Husebye ES.](#) Related Articles, Links

 Replacement therapy for Addison's disease: recent developments.
Expert Opin Investig Drugs. 2008 Apr;17(4):497-509. Review.
PMID: 18363515 [PubMed - indexed for MEDLINE]

Co-indexing

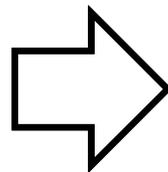
gopubmed

Searching is now sorted!

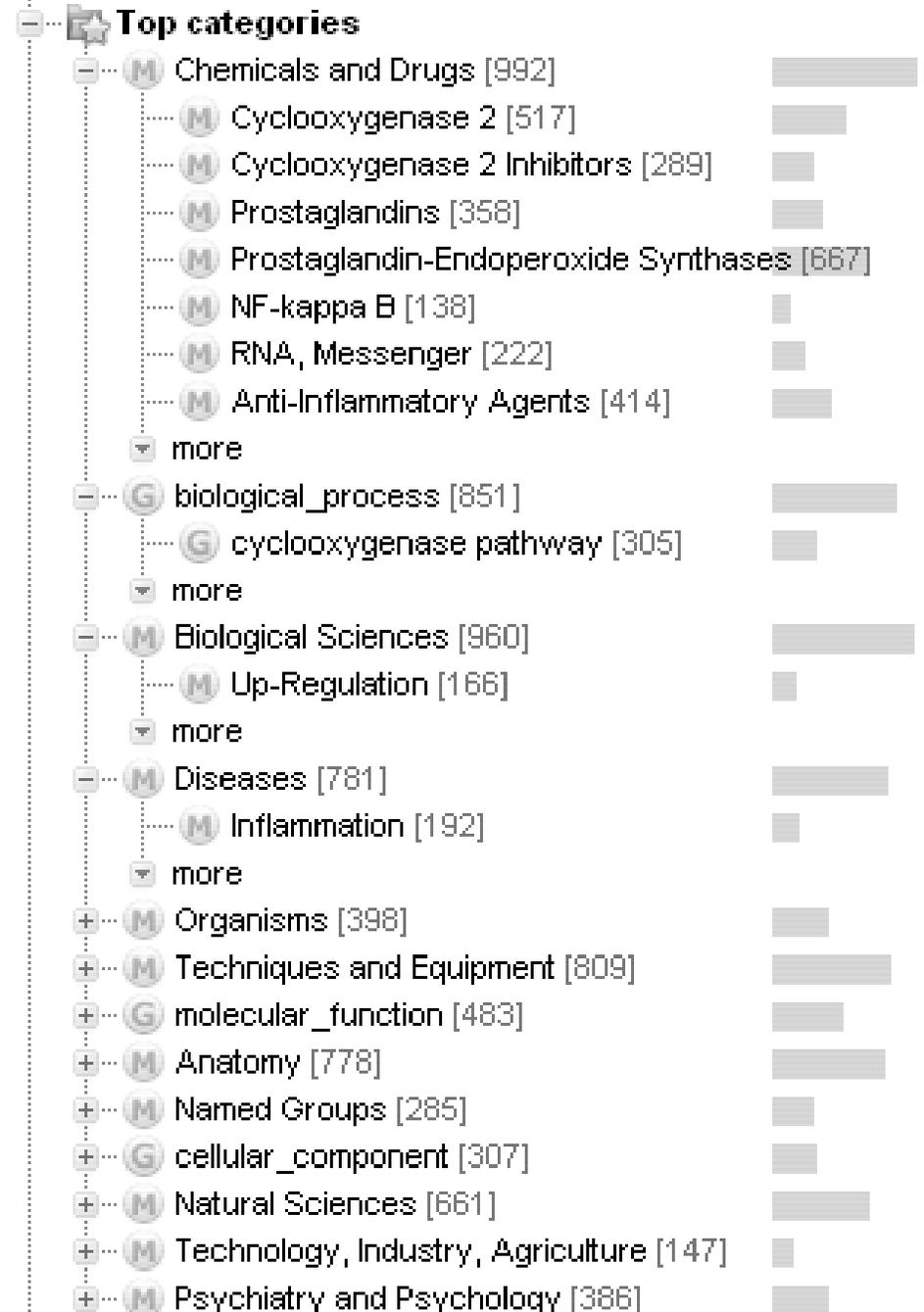
<http://www.gopubmed.com/>



COX-2



what



Lister Hill National Center for E

Knowledge management

Mapping across biomedical ontologies

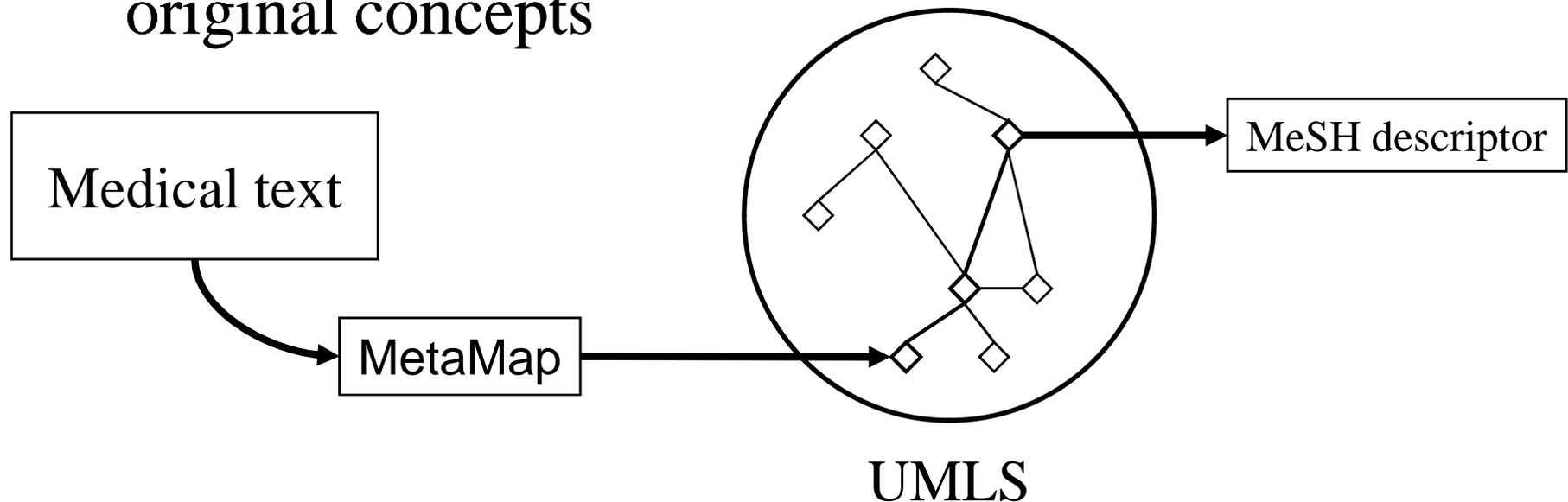
Reusing information

- ◆ Clinical information coded with SNOMED CT
 - Mapped to ICD9-CM and CPT for billing purposes
 - Mapped to ICD-O for epidemiology purposes
- ◆ Existing mapping tables created by terminology developers as an incentive to use SNOMED CT



Reusing tools

- ◆ For noun phrases extracted from medical texts, map to UMLS concepts [Aronson & al., *AMIA*, 2000]
- ◆ Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts

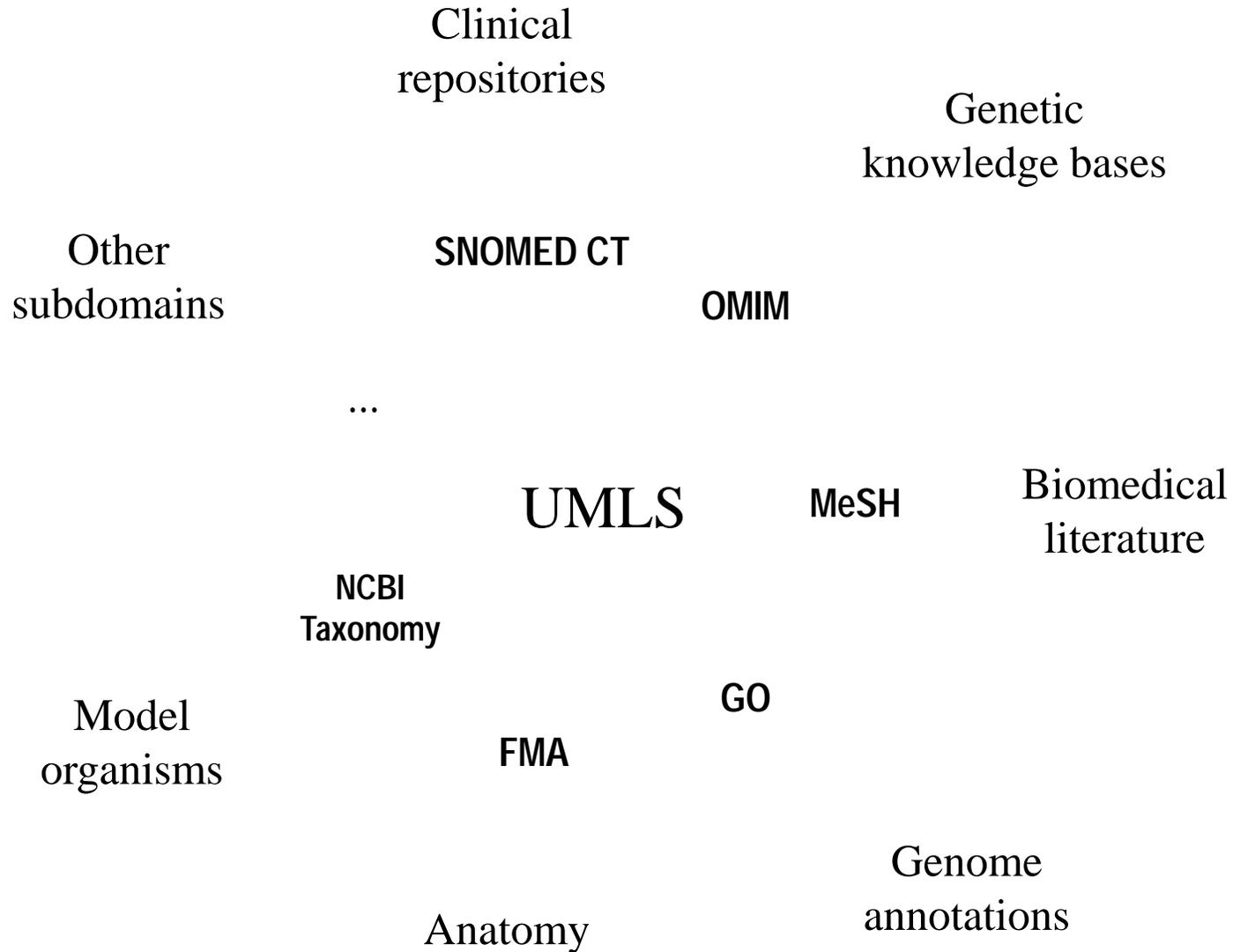


Terminology integration systems

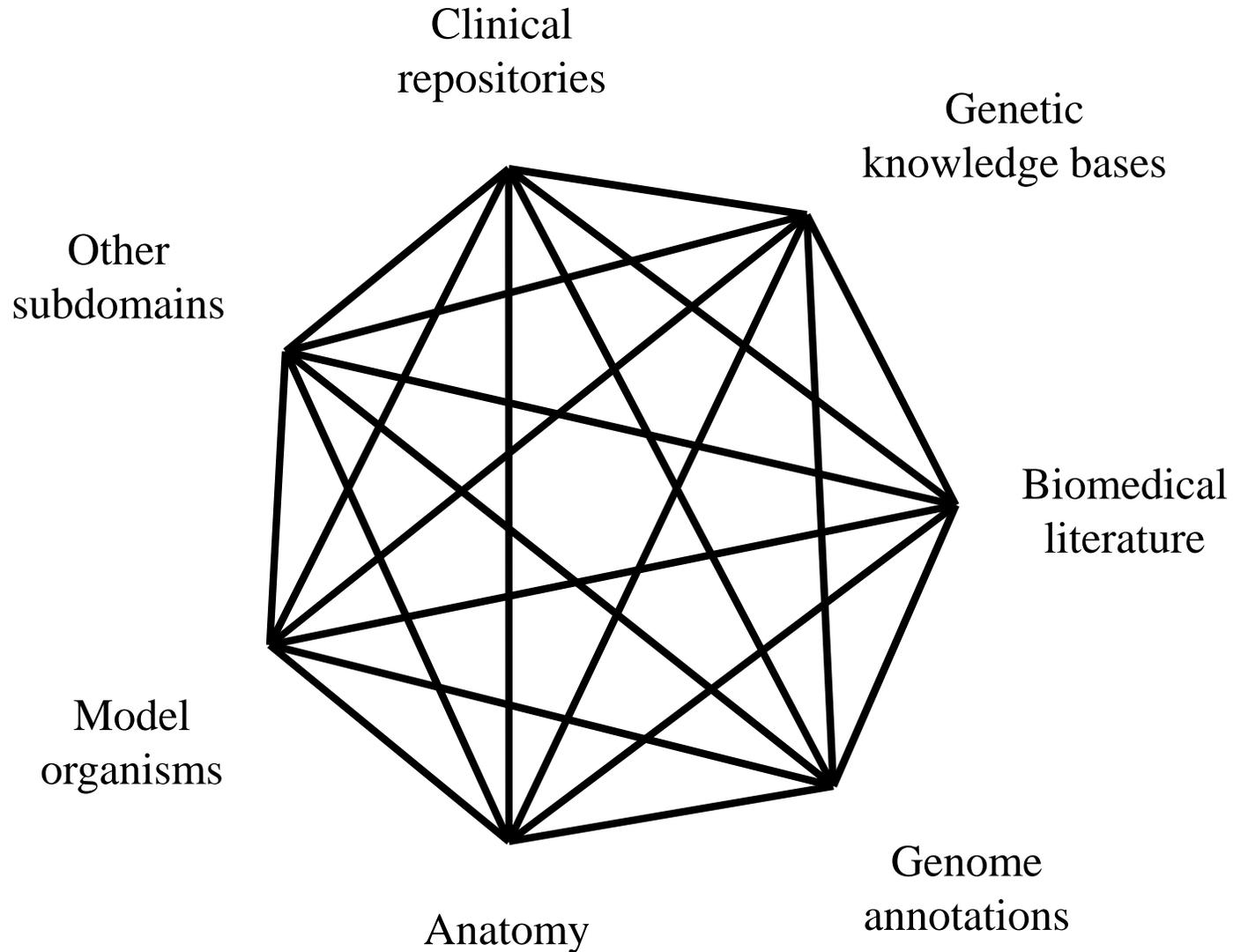
- ◆ Terminology integration systems (UMLS, RxNorm) help bridge across vocabularies
- ◆ Uses
 - Information integration
 - Ontology alignment
 - Medication reconciliation



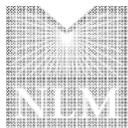
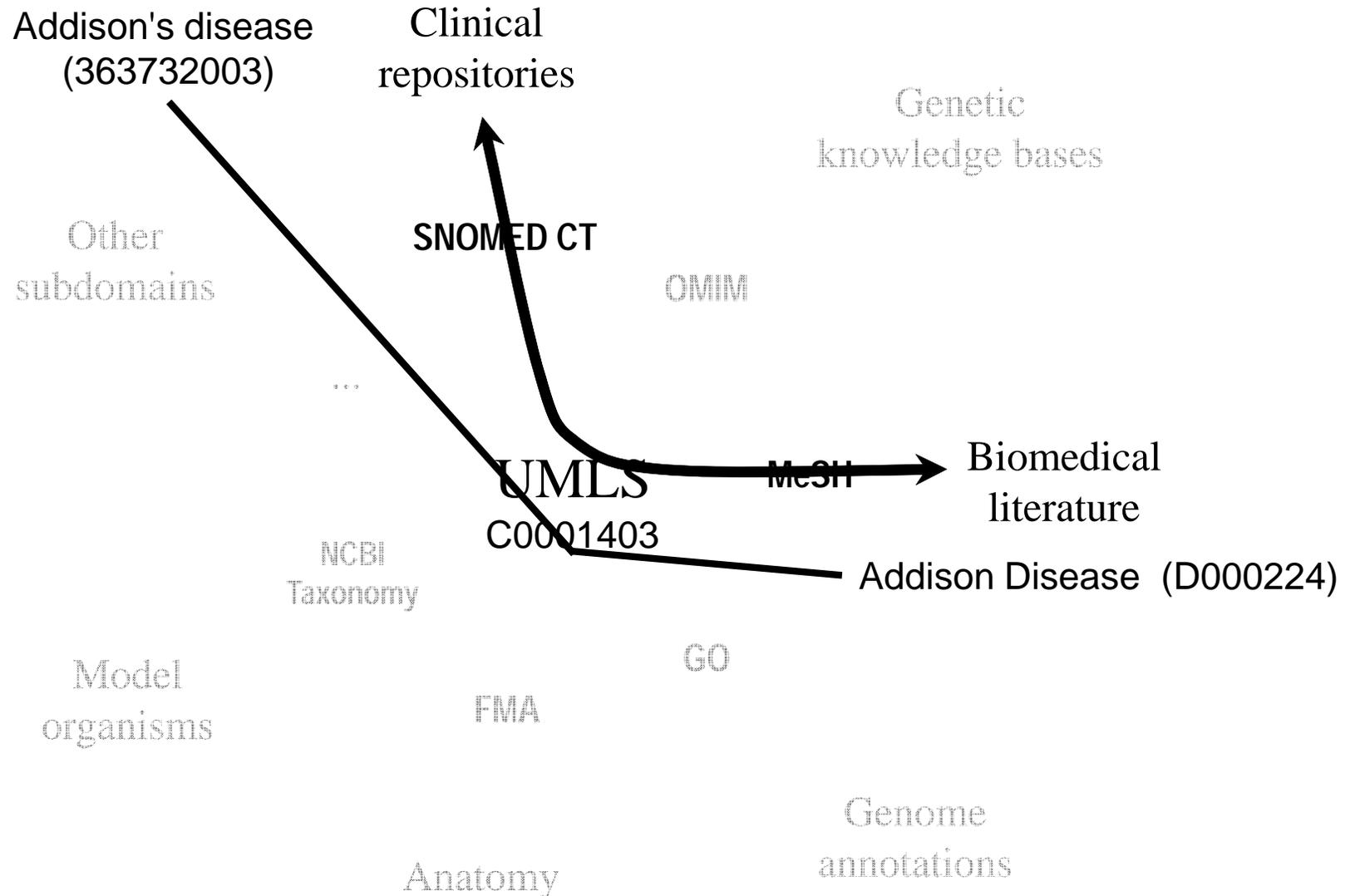
Integrating subdomains



Integrating subdomains



Trans-namespace integration



Data integration, exchange and semantic interoperability

Data integration, exchange and semantic interoperability

*Information exchange
and semantic operability*

“Standards”

- ◆ Ontologies help standardize patients data
 - Facilitate the exchange of data across institutions
 - Help connect “islands of data” (silos)

- ◆ LOINC
 - Exchange of laboratory data
 - In conjunction with HL7 messaging



Semantic interoperability projects BRIDG

◆ Biomedical Research Integrated Domain Group

- Information model for clinical research
- Interoperability between clinical trials information systems
- Ontologies provide value sets to the information model



Semantic interoperability projects CDA

- ◆ Clinical Document Architecture (CDA R2)
 - Formal representation of clinical statements
 - Clinical observations
 - Medication administration
 - Adverse events
 - Associate an information model (HL7 RIM) with terminologies (LOINC, SNOMED CT, RxNorm)



Semantic interoperability projects caCORE

◆ Cancer Common Ontologic Representation Environment

- Infrastructure developed to support an interoperable biomedical information system for cancer research
- Uses the NCI Thesaurus as a component



Data integration, exchange and semantic interoperability

Information and data integration

Approaches to data integration

◆ Warehousing

- Sources to be integrated are transformed into a common format and converted to a common vocabulary
- Normalization through ontologies (e.g., GO annotations)

◆ Mediation

- Local schema (of the sources)
- Global schema (in reference to which the queries are made)
- Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)

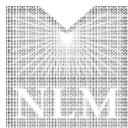
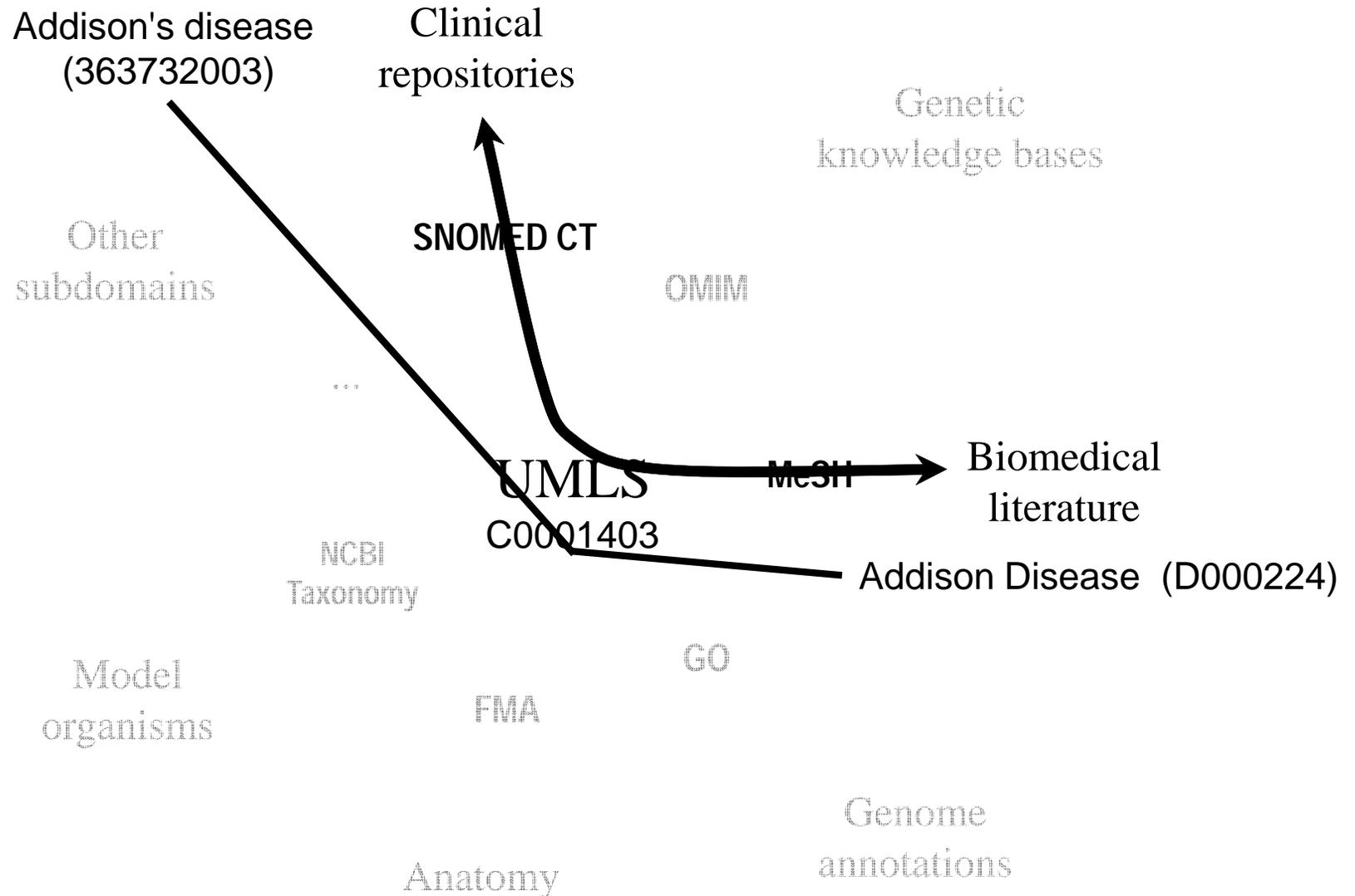


Ontologies and integration

- ◆ Terminology integration systems help bridge across terminologies and the domains they represent
- ◆ Mappings across ontologies enable the integration of namespaces in the Semantic Web



Trans-namespace integration



Decision support and reasoning

Data selection

- ◆ The structure of biomedical ontologies helps define groups of values from a high-level value
 - Vs. enumerating all possible values
- ◆ Useful for data selection in clinical studies
- ◆ ICD is used pervasively for this purpose
 - E.g., Study on supraventricular tachycardia (SVT), based on 2 high-level ICD codes
- ◆ Similarity with the definition of value sets for use in the information model



Data aggregation

- ◆ Ontologies help partition/aggregate data in data analysis
 - Clinical studies: Study a variable in groups of patients corresponding to the top level categories in ICD
 - Biology studies: Functional characterization of gene expression signatures with high-level concepts from the Gene Ontology
 - Recent trend: co-clustering



Decision support

◆ Clinical decision support

- Ontologies help normalize the vocabulary and increase the recall of rules
- Ontologies provide some domain knowledge and make it possible to create high-level rules (e.g., for a class of drugs rather than for each drug in the class)

◆ Other forms of decision support

- Based on automatic reasoning services for OWL ontologies (e.g., grading gliomas with NCIt)



Natural language processing applications

- ◆ Ontologies provide background domain knowledge for NLP applications
 - Question answering
 - Document summarization
 - Literature-based discovery
- ◆ The UMLS is often used, but other specific resources have been developed



Knowledge discovery

- ◆ By standardizing the vocabulary in a given domain, ontologies are enabling resources for knowledge discovery through data mining
- ◆ Less frequently, the structure of the ontology is leveraged by data mining algorithms
- ◆ Example of available datasets
 - ICD-coded clinical data (in conjunction with non-clinical information, e.g., environmental data)
 - Annotation of gene products to the GO (function prediction)



Barriers to usability of biomedical ontologies

Availability

- ◆ Many ontologies are freely available
- ◆ The UMLS is freely available for research purposes
 - Cost-free license required
- ◆ Licensing issues can be tricky
 - SNOMED CT is freely available in member countries of the IHTSDO
- ◆ Being freely available
 - Is a requirement for the Open Biomedical Ontologies (OBO)
 - Is a de facto prerequisite for Semantic Web applications



Discoverability

- ◆ Ontology repositories
 - UMLS: 143 source vocabularies
(biased towards healthcare applications)
 - NCBO BioPortal: ~100 ontologies
(biased towards biological applications)
 - Limited overlap between the two repositories
- ◆ Need for discovery services



Formalism

◆ Several major formalism

- Web Ontology Language (OWL) – NCI Thesaurus
- OBO format – most OBO ontologies
- UMLS Rich Release Format (RRF) – UMLS, RxNorm

◆ Conversion mechanisms

- OBO to OWL
- LexGrid (import/export to LexGrid internal format)



Ontology integration

- ◆ *Post hoc* integration , form the bottom up
 - UMLS approach
 - Integrates ontologies “as is”, including legacy ontologies
 - Facilitates the integration of the corresponding datasets
- ◆ Coordinated development of ontologies
 - OBO Foundry approach
 - Ensures consistency *ab initio*
 - Excludes legacy ontologies



Quality

- ◆ Quality assurance in ontologies is still imperfectly defined
 - Difficult to define outside a use case or application
- ◆ Several approaches to evaluating quality
 - Collaboratively, by users (Web 2.0 approach)
 - Marginal notes enabled by BioPortal
 - Centrally, by experts
 - OBO Foundry approach
- ◆ Important factors besides quality
 - Governance
 - Installed base / Community of practice



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 10, 2008 – Session #1

Interfaces to Biomedical Ontologies



Olivier Bodenreider

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

Overview

◆ Graphical interfaces

- UMLS Knowledge Source Server
- NCBO BioPortal
- NCI Thesaurus
- MeSH browser
- Foundational Model of Anatomy (FMA) Explorer
- Gene Ontology AmiGO
- ICD-10 online
- RxNav (RxNorm)
- [...]

◆ Application Programming Interfaces



Graphical interfaces

UMLS Knowledge Source Server

Home Help Log Off

UMLS Knowledge Source Server (UMLSKS)

Home Metathesaurus SPECIALIST Lexicon Semantic Network UMLS and Source View Tree Browser

Hello, Olivier Bodenreider

Choose a Section: Metathesaurus

Metathesaurus:RRF View

- Contexts
- Relations
- Raw View
- Co-occurrence Info
- Lexical View

Downloads

Metathesaurus Search

Enter term

Input type: Term CUI Code

Term: Addison's disease Release: 2008AA Index: Normalized String

OK

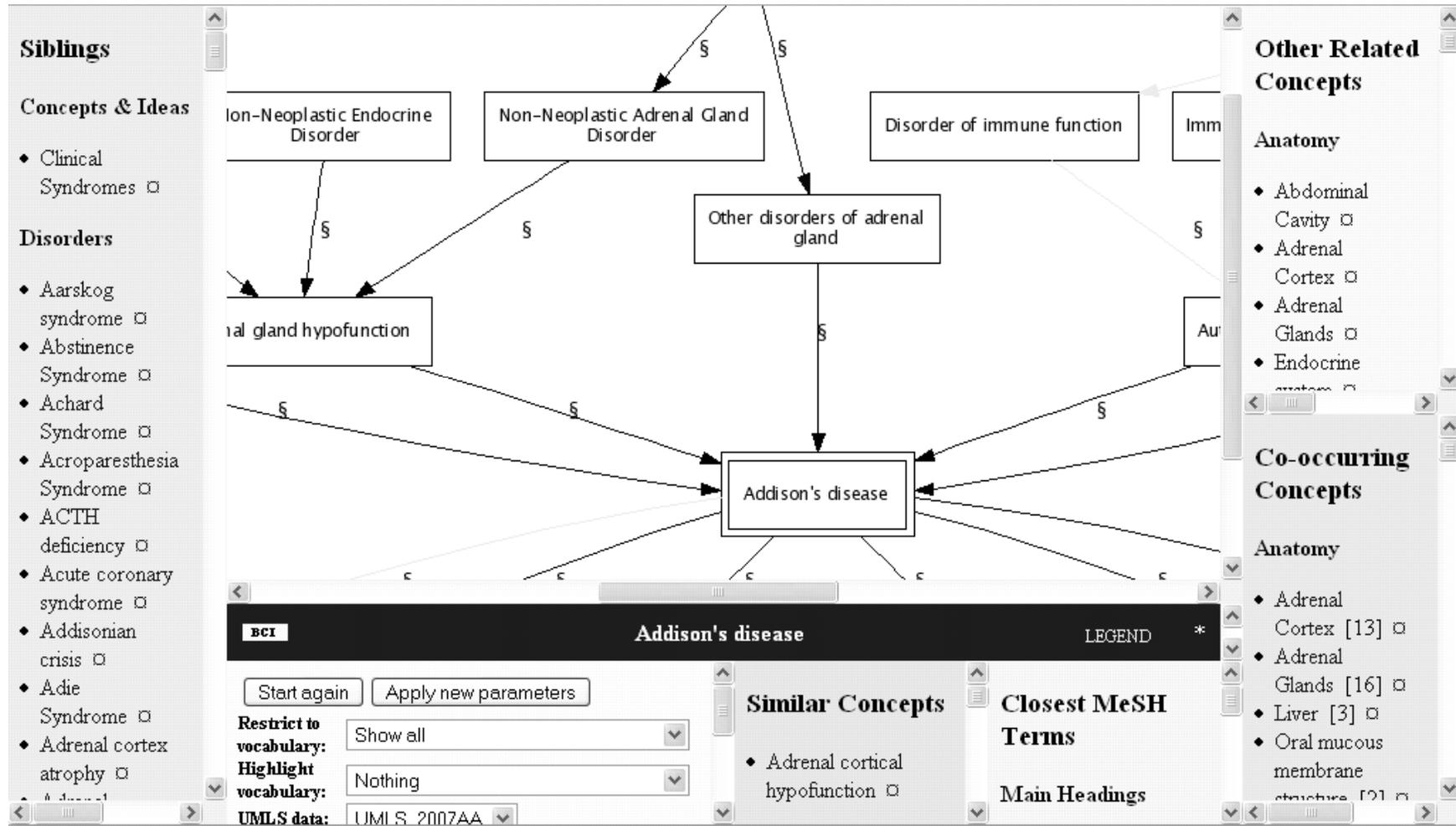
Sources: All sources

Please enter a Term, CUI or Code.

<http://umlsks.nlm.nih.gov/>



UMLS Semantic Navigator



NCBO BioPortal

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY



Version 1.0

[Browse](#) [Search](#) [Home](#) [Sign In](#) [Register](#) [Help](#)

Ontologies

[List View](#) [Category View](#)

[Submit Ontology](#) [Pending Submissions](#) [Download](#) [Visualize](#) [Search](#)

Name	Format	Current Version	Content Location	Action
African Traditional Medicine	OBO	1.0.1	NCBO Library	Download Visualize Search
Amino Acid	OWL Full	1.2	NCBO Library	Download Visualize Search
Amphibian gross anatomy	OBO	1.7	NCBO Library	Download Visualize Search
Animal natural history and life history	Protege	See Remote Site	Remote	Download Visualize Search
Basic Vertebrate Anatomy	OWL Full	1.1	NCBO Library	Download Visualize Search
Biological imaging methods	OBO	1.1	NCBO Library	Download Visualize Search
Biological process	OBO	1.208	NCBO Library	Download Visualize Search
Biomedical Resource Ontology	OWL Lite	1.1	NCBO Library	Download Visualize Search
BIRNLex	OWL DL	1.3.1	NCBO Library	Download Visualize Search

Done

<http://www.bioontology.org/tools/portal/bioportal.html>



NCI Thesaurus (EVS Server)

The screenshot displays the NCI Thesaurus (EVS Server) web interface. At the top, the National Cancer Institute logo and name are visible, along with the U.S. National Institutes of Health website address. The main navigation bar includes links for HELP, RESULTS, CUSTOMIZE, ABOUT, BROWSE HIERARCHY, and LOGOUT. The current page is titled "Concept Details" for "Prostate Adenocarcinoma".

Search Interface: The left sidebar contains a search box with "prostate adenocarcinoma" entered and a "Go!" button. Below the search box, it shows "Concepts visited (during this session): Prostate Adenocarcinoma". A "QUICK LINKS" section includes links for EVS HOME, NCICB HOME, NCI HOME, and KNOWN ISSUES.

Concept Details: The main content area shows the concept name "Prostate Adenocarcinoma" with a "Printable Page", "History", and "Graph" link. Below this, the "Identifiers" section lists the name "Prostate_Adenocarcinoma" and the code "C2919". The "Relationships to other concepts:" section lists several relationships, each with a list icon and a concept icon:

- Disease_Has_Normal_Tissue_Origin: Prostatic Epithelium
- Disease_Has_Abnormal_Cell: Adenocarcinoma Cell
- Disease_Has_Associated_Anatomic_Site: Prostate Gland
- Disease_Excludes_Abnormal_Cell: Neoplastic Smooth Muscle Cell
- Disease_Has_Finding: Invasive Lesion
- Disease_Has_Primary_Anatomic_Site: Prostate Gland

The status bar at the bottom left of the browser window shows "Done".

<http://nciterms.nci.nih.gov/NCIBrowser/SearchConcept.do>



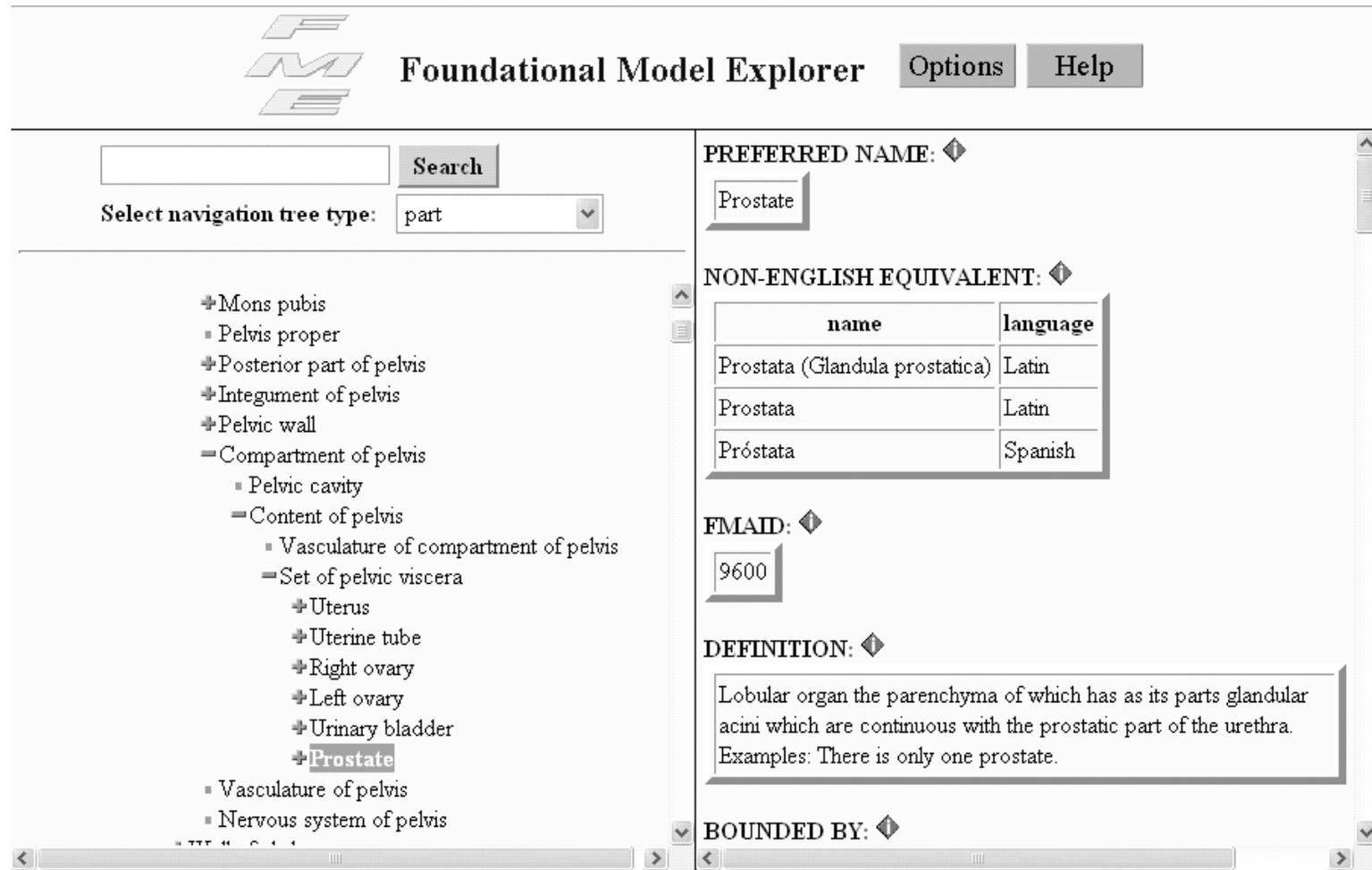
MeSH Browser

MeSH Heading	Prostatic Neoplasms
Tree Number	C04.588.945.440.770
Tree Number	C12.294.260.750
Tree Number	C12.294.565.625
Tree Number	C12.758.409.750
Annotation	coordinate IM with histological type of neoplasm (IM); note PROSTATIC ADENOMA see PROSTATIC HYPERPLASIA is also available
Scope Note	Tumors or cancer of the PROSTATE .
Entry Term	Cancer of Prostate
Entry Term	Cancer of the Prostate
Entry Term	Neoplasms, Prostate
Entry Term	Neoplasms, Prostatic
Entry Term	Prostate Cancer
Entry Term	Prostate Neoplasms
Entry Term	Prostatic Cancer
See Also	Prostate-Specific Antigen
See Also	Prostatic Hyperplasia
Allowable Qualifiers	BL BS CF CH CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SC SE SU TH UL UR US VE VI
Entry Version	PROSTATIC NEOPL
Date of Entry	19990101
Unique ID	D011471

<http://www.nlm.nih.gov/mesh/MBrowser.html>



Foundational Model of Anatomy



Foundational Model Explorer Options Help

Search

Select navigation tree type: part

- Mons pubis
 - Pelvis proper
- Posterior part of pelvis
- Integument of pelvis
- Pelvic wall
- ▬ Compartment of pelvis
 - Pelvic cavity
 - ▬ Content of pelvis
 - Vasculature of compartment of pelvis
 - ▬ Set of pelvic viscera
 - Uterus
 - Uterine tube
 - Right ovary
 - Left ovary
 - Urinary bladder
 - **Prostate**
 - Vasculature of pelvis
 - Nervous system of pelvis

PREFERRED NAME: Prostate

NON-ENGLISH EQUIVALENT:

name	language
Prostata (Glandula prostatica)	Latin
Prostata	Latin
Próstata	Spanish

FMAID: 9600

DEFINITION:

Lobular organ the parenchyma of which has as its parts glandular acini which are continuous with the prostatic part of the urethra.
Examples: There is only one prostate.

BOUNDED BY:

<http://sig.biostr.washington.edu/projects/fm/FME/>



Gene Ontology AmiGO

AmiGO

Term Information

Accession GO:0008375
Ontology molecular function
Synonyms exact: GlcNAc transferase activity
Definition
Comment
Subset

Term Lineage

[Switch to viewing term parents, siblings and children](#)

▼ Filter tree view ?

Filter Gene Product Counts

Data source

All
CGD
dictyBase
FlyBase

Species

All
Anaplasma phagocy...
Arabidopsis thaliana
Bacillus anthraci...

View Options

Tree view Full Compact

Set filters

Remove all filters

- ▣ all : all [250427 gene products]
- ▣ **GO:0003674 : molecular_function** [168568 gene products]
 - ▣ **GO:0003824 : catalytic activity** [51855 gene products]
 - ▣ **GO:0016740 : transferase activity** [15762 gene products]
 - ▣ **GO:0016757 : transferase activity, transferring glycosyl groups** [2274 gene products]
 - ▣ **GO:0016758 : transferase activity, transferring hexosyl groups** [1073 gene products]
 - ▣ **GO:0008375 : acetylglucosaminyltransferase activity** [131 gene products]
 - ▣ **GO:0008194 : UDP-glycosyltransferase activity** [598 gene products]
 - ▣ **GO:0008375 : acetylglucosaminyltransferase activity** [131 gene products]

Actions...

Last action: Reset the tree

Graphical View

View in tree browser

Download...

OBO

RDF/XML

GraphViz dot

[Back to top](#)



Lister Hill National Center for Biomedical Communications

<http://www.geneontology.org/>

219

SNOMED CT University of Sydney

Viral meningitis (disorder)

CONCEPT						
Concept ID	Fully Specified Name	Concept Status	CTV3ID	SNOMED ID	Is Primitive	
58170007	Viral meningitis (disorder)	Current (0)	Xa9B5	DE-30020	Fully defined (0)	
DESCRIPTIONS and SYNONYMS						
Description ID	Term	Description Status	Description Type	Language Code	Initial Capital Status	
1	96672018	Viral meningitis	Current (0)	Preferred (1)	en	Capitalization meaningless (0)
2	96673011	Aseptic meningitis, viral	Current (0)	Synonym (2)	en	Capitalization meaningless (0)
3	96674017	Abacterial meningitis	Current (0)	Synonym (2)	en	Capitalization meaningless (0)
4	96677012	Aseptic meningitis	Non-Current (1)	Synonym (2)	en	Capitalization meaningless (0)
5	96676015	Viral meningitis NOS	Non-Current (1)	Synonym (2)	en	Capitalization meaningless (0)
6	96675016	Viral meningitis, NEC	Non-Current (1)	Synonym (2)	en	Capitalization meaningless (0)
PARENTS						
Concept ID	FSN for Parent Concept (This Concept IS A)					
1	302810003  Viral infections of the central nervous system (disorder)					
2	312216007  Infective meningitis (disorder)					

<http://www.it.usyd.edu.au/~hitru/sct/A1.cgi>



SNOMED CT CliniClue

Hierarchy	Subtype hierarchy
<input type="checkbox"/>	302810003 viral infections of the central nervous system
<input type="checkbox"/>	312216007 infective meningitis
<input checked="" type="checkbox"/>	58170007 viral meningitis
<input type="checkbox"/>	+ 398136003 lymphocytic choriomeningitis
<input type="checkbox"/>	+ 10491005 Herpes zoster with meningitis
<input type="checkbox"/>	+ 315019000 HIV infection with aseptic meningitis
<input type="checkbox"/>	+ 111850006 adenoviral meningitis
<input type="checkbox"/>	+ 404234000 St. Louis meningitis
<input type="checkbox"/>	+ 404233006 West Nile meningitis
<input type="checkbox"/>	+ 28192008 enterovirus meningitis
<input type="checkbox"/>	+ 23291008 Herpes simplex meningitis
<input type="checkbox"/>	+ 44201003 mumps meningitis
<input type="checkbox"/>	+ 404243009 Keystone virus meningitis
<input type="checkbox"/>	+ 404240007 la Crosse meningitis
<input type="checkbox"/>	+ 230146001 post measles meningitis
<input type="checkbox"/>	+ 404239005 California serogroup viral meningitis
<input type="checkbox"/>	+ 13225007 rubella meningoencephalitis
<input type="checkbox"/>	+ 186561002 measles complicated by meningitis

www.cliniclue.com/

viral meningitis - Definition
Concept Status: Current
<input type="checkbox"/> Descriptions
<input type="checkbox"/> F viral meningitis (disorder)
<input type="checkbox"/> P viral meningitis
<input type="checkbox"/> S abacterial meningitis
<input type="checkbox"/> S aseptic meningitis, viral
<input type="checkbox"/> U meningitis abacteriana
<input type="checkbox"/> U meningitis aséptica, viral
<input type="checkbox"/> U meningitis viral
<input type="checkbox"/> U meningitis viral (trastorno)
<input type="checkbox"/> U meningitis virósica
<input type="checkbox"/> Definition: Fully defined by ...
<input type="checkbox"/> is a
<input type="checkbox"/> + D viral infections of the central nervous system
<input type="checkbox"/> + D infective meningitis
<input type="checkbox"/> causative agent
<input type="checkbox"/> + D virus
<input type="checkbox"/> Group
<input type="checkbox"/> associated morphology
<input type="checkbox"/> + D inflammation
<input type="checkbox"/> finding site
<input type="checkbox"/> + D meninges structure
<input type="checkbox"/> Qualifiers
<input type="checkbox"/> severity
<input type="checkbox"/> + P severities
<input type="checkbox"/> episodicity
<input type="checkbox"/> + P episodicities
<input type="checkbox"/> clinical course
<input type="checkbox"/> + P courses
<input type="checkbox"/> Codes
<input type="checkbox"/> Original SnomedId : DE-30020
<input type="checkbox"/> Read Code (Ctv3Id) : Xa9B5



Lister Hill National Center for Biomedical Research and Education

SNOMED CT SNOB

The screenshot displays the SNOB (SNOMED CT Nomenclature Browser) interface. On the left, a tree view shows the hierarchy: 'viral meningitis' (Active concepts) > 'CLINICAL FINDING (1)' > 'Viral meningitis'. The main pane shows details for concept 58170007: 'Viral meningitis'. The interface includes a search bar at the top with 'viral meningitis' and a 'More >>>' button. Below the search bar is a dropdown for 'Active concepts'. The main pane has a header with '96672018', 'PREF', 'Current (0)', 'en', and '20070131'. The content is organized into sections: 'Fully Specified Name(s)', 'Synonym(s)', 'Definition', 'Qualifiers', and 'Concept maps to'. A table at the bottom left shows 'Current (0)' as 'composite' and 'DE-30020' as 'Xa9B5'. The 'Definition' section contains the text: 'this concept Causative agent Virus', 'Group #1', 'this concept Associated morphology Inflammation', and 'this concept Finding site Meninges structure'. The 'Qualifiers' section contains: 'this concept Clinical course Courses', 'this concept Episodicity Episodicities', and 'this concept Severity Severities'. The 'Concept maps to' section is currently empty.

Current (0)	composite
DE-30020	Xa9B5
20070131	
20070131	

<http://snob.eggbird.eu/>



ICD-10



ICD
Version 2007

List of Chapters

Chapter Introduction

List of Blocks

Previous Block

Next Block

Chapter IV

Endocrine, nutritional and metabolic diseases
(E00-E90)

Disorders of other endocrine glands
(E20-E35)

Excludes: galactorrhoea ([N64.3](#))
gynaecomastia ([N62](#))

E20 Hypoparathyroidism

Excludes: Di George's syndrome ([D82.1](#))
postprocedural hypoparathyroidism ([E89.2](#))
tetany NOS ([R29.0](#))
transitory neonatal hypoparathyroidism ([P71.4](#))

E20.0 Idiopathic hypoparathyroidism
E20.1 Pseudohypoparathyroidism
E20.8 Other hypoparathyroidism
E20.9 Hypoparathyroidism, unspecified
Parathyroid tetany

Search ICD-10

Full search

OK Help

Move to ICD code:

 OK

Application prepared by: WHO & DIMDI (German Institute of Medical Documentation and Information)

<http://www.who.int/classifications/apps/icd/icd10online/>



RxNav (RxNorm)

<http://mor.nlm.nih.gov/download/rxnav/>

RxNorm Navigator 08AA_080602F

Terminology

Search

RxNav Search By: String Help Enter Search String: zyrtec Search

Browse

The screenshot shows a search for 'zyrtec' in the RxNorm Navigator. The search results are displayed in a network diagram with the following components and relationships:

- Ingredient** (1 element: Cetirizine) is related to **Ingredient Variant** (1 element: Cetirizine Dihydrochloride) via **form_of** and **has_form**.
- Ingredient Variant** is related to **Brand Name** (1 element: Zyrtec) via **precise_ingr_of** and **has_precise_ingr**.
- Ingredient** is related to **Clinical Drug Component** (3 elements: Cetirizine 1 MG/ML, Cetirizine 10 MG, Cetirizine 5 MG) via **ingredient_of** and **has_ingredient**.
- Brand Name** is related to **Branded Drug Component** (3 elements: Cetirizine 1 MG/ML [Zyrtec], Cetirizine 10 MG [Zyrtec], Cetirizine 5 MG [Zyrtec]) via **ingredient_of** and **has_ingredient**.
- Clinical Drug Component** is related to **Clinical Drug** (5 elements: Cetirizine 1 MG/ML Oral Solution, Cetirizine 10 MG Chewable Tablet, Cetirizine 10 MG Oral Solution, Cetirizine 5 MG Chewable Tablet, Cetirizine 5 MG Oral Solution) via **constitutes** and **consists_of**.
- Branded Drug Component** is related to **Branded Drug** (5 elements: Zyrtec 1 MG/ML Oral Solution, Zyrtec 10 MG Chewable Tablet, Zyrtec 10 MG Oral Solution, Zyrtec 5 MG Chewable Tablet, Zyrtec 5 MG Oral Solution) via **constitutes** and **consists_of**.
- Clinical Drug** is related to **Dose Form** (3 elements: Chewable Tablet, Oral Solution, Oral Tablet) via **is inverse isa** and **dose_form_of** / **has_dose_form**.
- Branded Drug** is related to **Branded Drug Form** (3 elements: Cetirizine Chewable Tablet [Zyrtec], Cetirizine Oral Solution [Zyrtec], Cetirizine Oral Tablet [Zyrtec]) via **is inverse isa** and **dose_form_of** / **has_dose_form**.
- Clinical Drug Form** (3 elements: Cetirizine Chewable Tablet, Cetirizine Oral Solution, Cetirizine Oral Tablet) is related to **Branded Drug Form** via **tradename_of** and **has_tradename**.

Retrieval Status or Detailed View of an RxNorm Entry (RxNorm Concept Unique Identifier (RXCU) | UMLS Concept Unique Identifier (UMLSCUI): 2008AA | RxNorm Sy Retrieved "Zyrtec" for String "zyrtec".

Application Programming Interfaces

Application Programming Interface

- ◆ Expose resources in such a way that they can be integrated in programs
 - Programming “against” a resource
- ◆ Standard protocols for communication
 - Web services (SOAP, REST)
- ◆ Standard libraries for programming
- ◆ Focus on content, not message



UMLSKS Web Service API

- ◆ UMLSKS <http://umlsks.nlm.nih.gov/>
 - Developer's Guide > Webservice Operations
- ◆ WSDL available
- ◆ API give access to all 3 knowledge sources
- ◆ Licensing issues
 - Granting ticket and Single-use tickets



UMLSKS Web Service API Example

```
ConceptIdGroup findCUIByNormString  
    (ConceptIdNormStringRequest request);
```



Argument: ConceptIdNormStringRequest

This class contains the arguments that further restrict the behavior of the call.

```
setCasTicket (String s)  
    - Single-use ticket returned by the AuthorizationPort webservice  
setRelease (String s)  
    - UMLS release of interest  
setSearchString (String s)  
    - input search string  
setSABs (String[] array)  
    - set of source abbreviations to search  
setLanguage (String s)  
    - language restriction  
setIncludeSuppressibles (boolean b)  
    - true if suppressible strings are included in the search  
setCVF (long l)  
    - Bit flag for the content view to search
```



Return: ConceptIdGroup



Other APIs to terminology systems

- ◆ NCBO BioPortal

http://www.bioontology.org/docs/bioportal/development/web_services.html

- ◆ OLS - Ontology Lookup Service

<http://www.ebi.ac.uk/ontology-lookup/WSDLDocumentation.do>

- ◆ RxNorm

<http://mor.nlm.nih.gov/download/rxnav/RxNormAPI.html>



Applications based on WS APIs

◆ UMLSKS API

- UMLSKS

<http://umlsks.nlm.nih.gov/>

◆ RxNorm API

- RxNav

<http://mor.nlm.nih.gov/download/rxnav/rxnav.jnlp>

- MyMedicationList

<http://mml.nlm.nih.gov/MyMedicationList.jnlp>



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 10, 2008 – Session #2

Searching and Analyzing Biomedical Concepts



Olivier Bodenreider

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

Exercise 1

- ◆ What are the Clinical Drug Components for Zyrtec? (RxNav)



Exercise 2

- ◆ What are the parts of the Aorta? (FMA)



Exercise 3

- ◆ What are the parents of Hodgkin's disease in SNOMED CT?
 - Try SNOMEDCTID: 118599009
- ◆ What is its associated morphology?



Exercise 4

- ◆ What are the various meanings of IL-2? (UMLS)



Exercise 5

- ◆ What are the pharmacologic actions of Zyrtec?
(MeSH)



Exercise 6

- ◆ What are some synonyms for Schwannoma? (NCI Thesaurus)



Solutions

Exercise 1

- ◆ What are the Clinical Drug Components for Zyrtec? (RxNav)

The screenshot shows the RxNorm Navigator interface. The search bar contains 'zyrt' and the search button is pressed. The browse view displays a network of drug components and their relationships:

- Ingredient** (1 element): Cetirizine
- Ingredient Variant** (1 element): Cetirizine Dihydrochloride
- Brand Name** (1 element): Zyrtec
- Clinical Drug Component** (3 elements): Cetirizine 1 MG/ML, Cetirizine 10 MG, Cetirizine 5 MG
- Branded Drug Component** (3 elements): Cetirizine 1 MG/ML [Zyrtec], Cetirizine 10 MG [Zyrtec], Cetirizine 5 MG [Zyrtec]

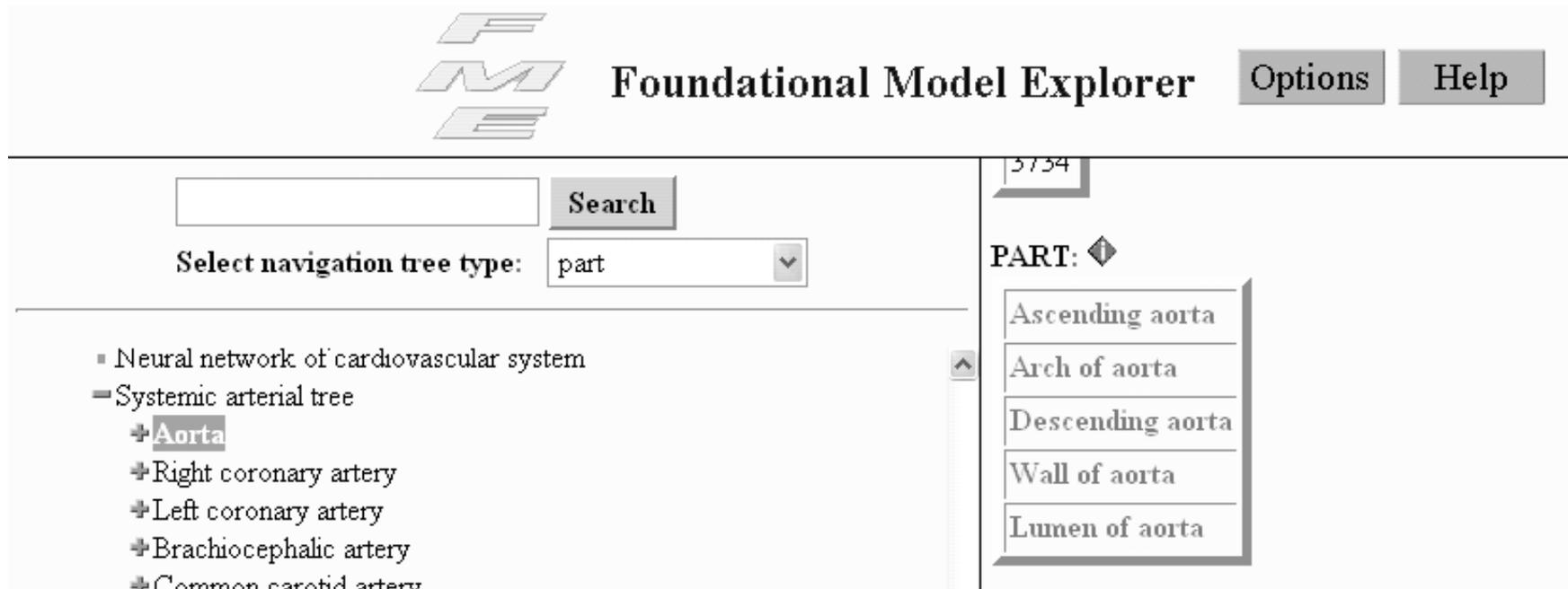
Relationships shown:

- Ingredient** is *form_of* **Ingredient Variant** and *has_form* **Ingredient Variant**.
- Ingredient Variant** is *precise_ingr_of* **Brand Name** and *has_precise_ingr* **Brand Name**.
- Ingredient Variant** is *precise_ingredient_of* **Clinical Drug Component** and *has_precise_ingredient* **Clinical Drug Component**.
- Brand Name** is *ingredient_of* **Branded Drug Component** and *has_ingredient* **Branded Drug Component**.
- Clinical Drug Component** is *tradename_of* **Branded Drug Component** and *has_tradename* **Branded Drug Component**.
- Clinical Drug Component** *constitutes* **Branded Drug Component** and *consists of* **Branded Drug Component**.
- Branded Drug Component** *constitutes* **Clinical Drug Component** and *consists of* **Clinical Drug Component**.



Exercise 2

◆ What are the parts of the Aorta? (FMA)



The screenshot displays the Foundational Model Explorer (FMA) interface. At the top, the logo 'FMA' is shown next to the text 'Foundational Model Explorer'. To the right are 'Options' and 'Help' buttons. Below the header, there is a search bar with a 'Search' button and a dropdown menu for 'Select navigation tree type:' set to 'part'. The main content area shows a hierarchical tree structure on the left and a detailed view of the selected 'Aorta' on the right. The tree structure includes:

- Neural network of cardiovascular system
- Systemic arterial tree
 - + **Aorta**
 - + Right coronary artery
 - + Left coronary artery
 - + Brachiocephalic artery
 - + Common carotid artery

The detailed view on the right, titled 'PART: ⚙', lists the following components of the aorta:

- Ascending aorta
- Arch of aorta
- Descending aorta
- Wall of aorta
- Lumen of aorta

Exercise 3

- ◆ What are the parents of Hodgkin's disease in SNOMED DT?
 - Try SNOMEDCTID: 118599009
- ◆ What is its associated morphology?

CONCEPT					
Concept ID	Fully Specified Name	Concept Status	CTV3ID	SNOMED ID	Is Primitive
118599009	Hodgkin's disease (disorder)	Current (0)	B61..	DC-F1000	Fully defined (0)

PARENTS	
Concept ID	FSN for Parent Concept (This Concept IS A)
1 118600007	⌘ Malignant lymphoma (disorder)

ATTRIBUTES			
Concept ID	FSN for Target Concept	Relationship Type	Values
1 288526004	⌘ Episodicities (qualifier value)	⌘ Episodicity (attribute)	{New episode; Ongoing episode; Old episode; Undefined episodicity; Other episode RCGP; First episode}
2 128930002	⌘ Hodgkin lymphoma - category (morphologic abnormality)	⌘ Associated morphology (attribute)	{Hodgkin lymphoma, lymphocyte-rich; Hodgkin lymphoma, no ICD-O subtype; Hodgkin lymphoma, lymphocyte depletion, diffuse fibrosis; Hodgkin lymphoma, nodular sclerosis, cellular phase; Hodgkin lymphoma, mixed cellularity; Hodgkin lymphoma, nodular sclerosis, grade 2; Hodgkin lymphoma, nodular sclerosis, grade 1; Hodgkin sarcoma [obs]; Hodgkin lymphoma, nodular sclerosis; Hodgkin lymphoma, nodular lymphocyte predominance; Hodgkin lymphoma, lymphocyte depletion, reticular; Hodgkin granuloma [obs]; Hodgkin lymphoma, lymphocyte depletion}

Exercise 4

- ◆ What are the various meanings of IL-2? (UMLS)

Metathesaurus Search

Enter term

Input type: Term CUI Code

Term: IL-2 **Release:** 2008AA **Index:** Normalized String

Sources: All sources

[IL2 gene \[C0879590\]](#)
[Interleukin-2 \[C0021756\]](#)
[Recombinant Interleukin-2 \[C1522405\]](#)
[interleukin-2 binding \[C1149229\]](#)



Exercise 5

- ◆ What are the pharmacologic actions of Zyrtec?
(MeSH)

Pharm. Action	<u>Anti-Allergic Agents</u>
Pharm. Action	<u>Histamine H1 Antagonists, Non-Sedating</u>



Exercise 6

◆ What are some synonyms for Schwannoma? (NCI Thesaurus)

Preferred_Name	Schwannoma
Semantic_Type	Neoplastic Process
Synonym	Neurilemmoma
Synonym	Neurinoma
Synonym	Schwannoma
Synonym	schwannoma



What to look for

- ◆ Search modalities
 - Spelling correction / auto-completion / normalization
 - Word combinations (AND/OR)
- ◆ Visualization
 - Graph vs. (forest of) trees
- ◆ Navigation
- ◆ What properties are displayed



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 10, 2008 – Session #3 / June 11, 2008 – Session #1

Contrasting and Critiquing Biomedical Ontologies



Olivier Bodenreider

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

Exercise #1

◆ Hodgkin's disease

- NCI Thesaurus
- SNOMED CT



Exercise #2

◆ Prostate

- FMA
- SNOMED CT



Exercise #3

◆ Cetirizine

- MeSH
- SNOMED CT



Solutions

Solutions

Exercise #1

Exercise #1

◆ Hodgkin's disease

- NCI Thesaurus
 - Using the NCI browser (EVS)
<http://nciterms.nci.nih.gov/>
- SNOMED CT
 - Using the online browser from U. Sydney
<http://www.cs.usyd.edu.au/~hitru/sct/A3.cgi>



Hodgkin's disease in NCI (1)

URI: http://ncitterms.nci.nih.gov:80/NCIBrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&code=C9357
Version: April 2008 (08.04d)

Hodgkin Lymphoma

Identifiers:

name	Hodgkin_s_Lymphoma
code	C9357

Relationships to other concepts:

Disease_Has_Primary_Anatomic_Site	Hematopoietic and Lymphatic System
Disease_Has_Normal_Tissue_Origin	Lymphoid Tissue
Disease_Excludes_Normal_Cell_Origin	Myeloid Cell
Disease_Excludes_Normal_Cell_Origin	Plasma Cell
Disease_Has_Abnormal_Cell	Reed-Sternberg Cell
Disease_Has_Associated_Anatomic_Site	Hematopoietic and Lymphatic System
Disease_Has_Normal_Cell_Origin	Mature Lymphocyte
Disease_Has_Primary_Anatomic_Site	Lymphatic System

Superconcepts

		Common Hematopoietic Neoplasm
		Lymphoma



Hodgkin's disease in NCIt (1)

Information about this concept:

ALT_DEFINITION

NCI-GLOSS|A malignant disease of the lymphatic system that is characterized by painless enlargement of lymph nodes, the spleen, or other lymphatic tissue. It is sometimes accompanied by symptoms such as fever, weight loss, fatigue, and night sweats.

DEFINITION

NCI|A lymphoma, previously known as Hodgkin's disease, characterized by the presence of Reed-Sternberg cells. There are two distinct subtypes: nodular lymphocyte predominant Hodgkin lymphoma and classical Hodgkin lymphoma. Hodgkin lymphoma has a bimodal age distribution, and involves primarily lymph nodes. Current therapy for Hodgkin lymphoma has resulted in an excellent outcome and cure for the majority of patients.

ICD-O-3_Code

9650/3

Preferred_Name

Hodgkin Lymphoma

Semantic_Type

Neoplastic Process

Synonym

HL

Synonym

Hodgkin Lymphoma

Synonym

Hodgkin's Disease

Synonym

Hodgkin's Lymphoma

Synonym

Hodgkin's disease

Unified Medical Language System Concept Identifier

C0019829

Comments on Hodgkin's disease in NCIIt (1)

◆ Search term: “Hodgkin’s disease”

- Not found, although “Hodgkin’s disease” is listed as a synonym
- Search on “hodgkin”, select “Hodgkin lymphoma”

◆ Parent classes

- Common hematopoietic neoplasm
 - Not an ontological category
 - Would be better represented through an associative relation (e.g., along the lines of “*has_prevalence* high prevalence”)
 - *Isa* overloading



Comments on Hodgkin's disease in NCIt (2)

◆ Associative relations

- For cancers, anatomy and morphology are foundational relations
- Here
 - Anatomy : *Disease_Has_Primary_Anatomic_Site*
Hematopoietic and Lymphatic System
 - Morphology: not directly represented
(indirectly through *Disease_Has_Normal_Cell_Origin* Mature Lymphocyte)



Hodgkin's disease in SNOMED CT (1)

CONCEPT		
Concept ID	Fully Specified Name	Concept Status
118599009	Hodgkin's disease (disorder)	Current (0)

PARENTS		
	Concept ID	FSN for Parent Concept (This Concept
1	118600007	⚠ Malignant lymphoma (disorder)

ATTRIBUTES			
	Concept ID	FSN for Target Concept	Relationship Type
1	288526004	⚠ Episodicities (qualifier value)	⚠ Episodicity (attribute)
2	128930002	⚠ Hodgkin lymphoma - category (morphologic abnormality)	⚠ Associated morphology (attribute)



Hodgkin's disease in SNOMED CT (2)

Hodgkin's disease (disorder)

CONCEPT			
Concept ID	Fully Specified Name	Concept Status	CTV3ID
118599009	Hodgkin's disease (disorder)	Current (0)	B61..

DESCRIPTIONS and SYNONYMS				
	Description ID	Term	Description Status	Description Type
1	177017015	Hodgkin's disease (clinical)	Current (0)	Preferred (1)
2	1220409010	Malignant Hodgkin's lymphoma	Current (0)	Synonym (2)
3	1220408019	HD - Hodgkin's disease	Current (0)	Synonym (2)

Comments on Hodgkin's disease in SNOMED CT (1)

◆ Search term: “Hodgkin's disease”

- Not found, although “Hodgkin's disease” is listed as a synonym
 - Search result: “Hodgkin lymphoma, nodular sclerosis, grade 1 (morphologic abnormality)”
- Search on “lymphoma”, navigate down from “Malignant lymphoma”
- “hodgkin's disease” is ambiguous
 - Hodgkin lymphoma, no ICD-O subtype (morphologic abnormality)
 - Hodgkin's disease (disorder)
- “Malignant lymphoma, Hodgkin's”
 - NB: lymphoma is always malignant

◆ Parent classes

- Malignant lymphoma (clinical) [OK]



Comments on Hodgkin's disease in SNOMED CT (2)

◆ Associative relations

- For cancers, anatomy and morphology are foundational relations
- Here
 - Anatomy : not directly represented
(indirectly through descendant concepts, e.g., Hodgkin's disease of intrathoracic lymph nodes)
 - Morphology: *Associated morphology* Hodgkin lymphoma - category



Hodgkin's disease NCI vs. SNOMED CT (1)

◆ Shared synonyms: NCI 1/2, SNOMED CT 1/3

- Hodgkin's disease

◆ Shared relations

- *Isa*

- NCI: Lymphoma

- Definition: “malignant (clonal) proliferation of B-lymphocytes or T-lymphocytes which involves the lymph nodes, bone marrow and/or extranodal sites. This category includes Non-Hodgkin lymphomas and Hodgkin lymphomas.”

- SNOMED CT: Malignant lymphoma

- Same UMLS concept (CUI: C0024299)



Hodgkin's disease NCI vs. SNOMED CT (2)

◆ Shared relations: Associative relations

- Anatomy
 - In NCI, but not in SNOMED CT
- Morphology
 - In SNOMED CT, but not in NCI
 - Only indirectly, though cell type
- Cell type
 - Only in NCI



Solutions

Exercise #2

Exercise #2

◆ Prostate

- FMA

- Using the Foundational Model Explorer

- <http://sig.biostr.washington.edu/projects/fm/FME/>

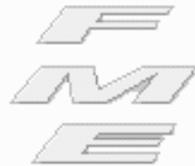
- SNOMED CT

- Using the online browser from U. Sydney

- <http://www.cs.usyd.edu.au/~hitru/sct/A3.cgi>



Prostate in FMA (1)



Foundational Model Explorer

Options

Help

Search

Select navigation tree type:

subclass



- ☐ Organ system subdivision
- ☐ Organ
 - ☐ Solid organ
 - ☐ Parenchymatous organ
 - ☐ Lobular organ
 - ☐ Lung
 - ☐ Liver
 - ☐ Pancreas
 - ☐ **Prostate**
 - ☐ Testis

Intraprostatic part of left ejaculatory duct

Prostatic stroma

Neural network of prostate

Vasculature of prostate

PART OF: ◆

Genital system

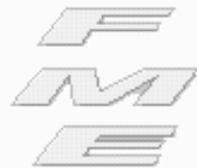
Content of male pelvis

Set of male pelvic viscera

Set of pelvic viscera



Prostate in FMA (2)



Foundational Model Explorer

Options

Help

Search

Select navigation tree type:

subclass



- ↳ Organ system subdivision
- Organ
 - Solid organ
 - Parenchymatous organ
 - Lobular organ
 - ↳ Lung
 - Liver
 - Pancreas
 - **Prostate**
 - ↳ Testis
 - ↳ Salivary gland

PREFERRED NAME:

Prostate

NON-ENGLISH EQUIVALENT:

name	language
Prostata (Glandula prostatica)	Latin
Prostata	Latin
Próstata	Spanish

FMAID:

9600



Comments on Prostate in FMA

- ◆ No synonyms in English
 - Latin and Spanish synonyms
- ◆ Hierarchies
 - *Isa*: Lobular organ
 - *Part_of*: Set of pelvic viscera
- ◆ Associative relations
 - *Lymphatic drainage*
 - No spatial relations



Prostate in SNOMED CT (1)

435 results found for prostate:

Previous

Next

#	Concept ID	Fully Specified Name	Preferred Terms and Synonyms
1	9713002	Prostatitis (disorder)	Inflammation of prostate- Prostatitis [PT]- Prostatitis, NOS
2	11441004	Prostatism (disorder)	Prostatism [PT]- Prostatism, NOS
3	41216001	Prostatic structure (body structure)	Prostatic structure [PT]- Prostate- Prostate, NOS
4	181422007	Entire prostate (body structure)	Entire prostate [PT]- Prostate

CONCEPT

Concept ID	Fully Specified Name
181422007	Entire prostate (body structure)

DESCRIPTIONS and SYNONYMS

Description ID	Term
1 280451017	Entire prostate
2 280452012	Prostate

PARENTS

Concept ID	FSN for Parent Concept (This Concept IS A)
1 310536002	Male internal genital organ (body structure)
2 41216001	Prostatic structure (body structure)
3 300443000	Entire male genital organ (body structure)



Prostate in SNOMED CT (2)

ATTRIBUTES			
	Concept ID	FSN for Target Concept	Relationship Type
1	118760003	Entire viscus (body structure)	Part of (attribute)
2	245461005	Entire urinary tract (body structure)	Part of (attribute)
3	362265004	Entire male internal genitalia (body structure)	Part of (attribute)
4	362267007	Entire lower male genitourinary tract (body structure)	Part of (attribute)
5	362717004	Entire minor pelvis (body structure)	Part of (attribute)
6	362206001	Entire lower genitourinary tract (body structure)	Part of (attribute)
7	361340001	Entire male genital system (body structure)	Part of (attribute)
8	302553009	Entire abdomen (body structure)	Part of (attribute)



Comments on Prostate in SNOMED CT

- ◆ “Ambiguous” term
 - Entire prostate
 - Prostatic structure
- ◆ Structure-Entire-Part representation of anatomical entities in SNOMED CT
 - Reification of *part_of*
 - Enables mereological inference through *isa* hierarchy
 - Not intuitive



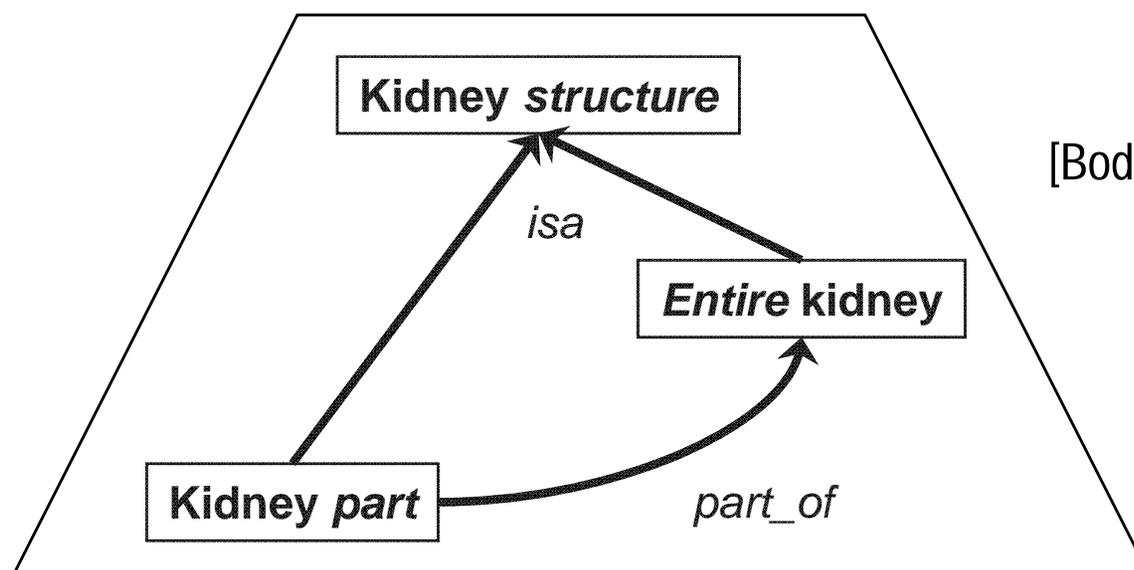
Structure-Entire-Part (SEP) triples

- ◆ S – The entity or any of its parts
- ◆ E – The entire anatomical entity
- ◆ P – Any parts of the anatomical entity

X structure

Entire X

X part



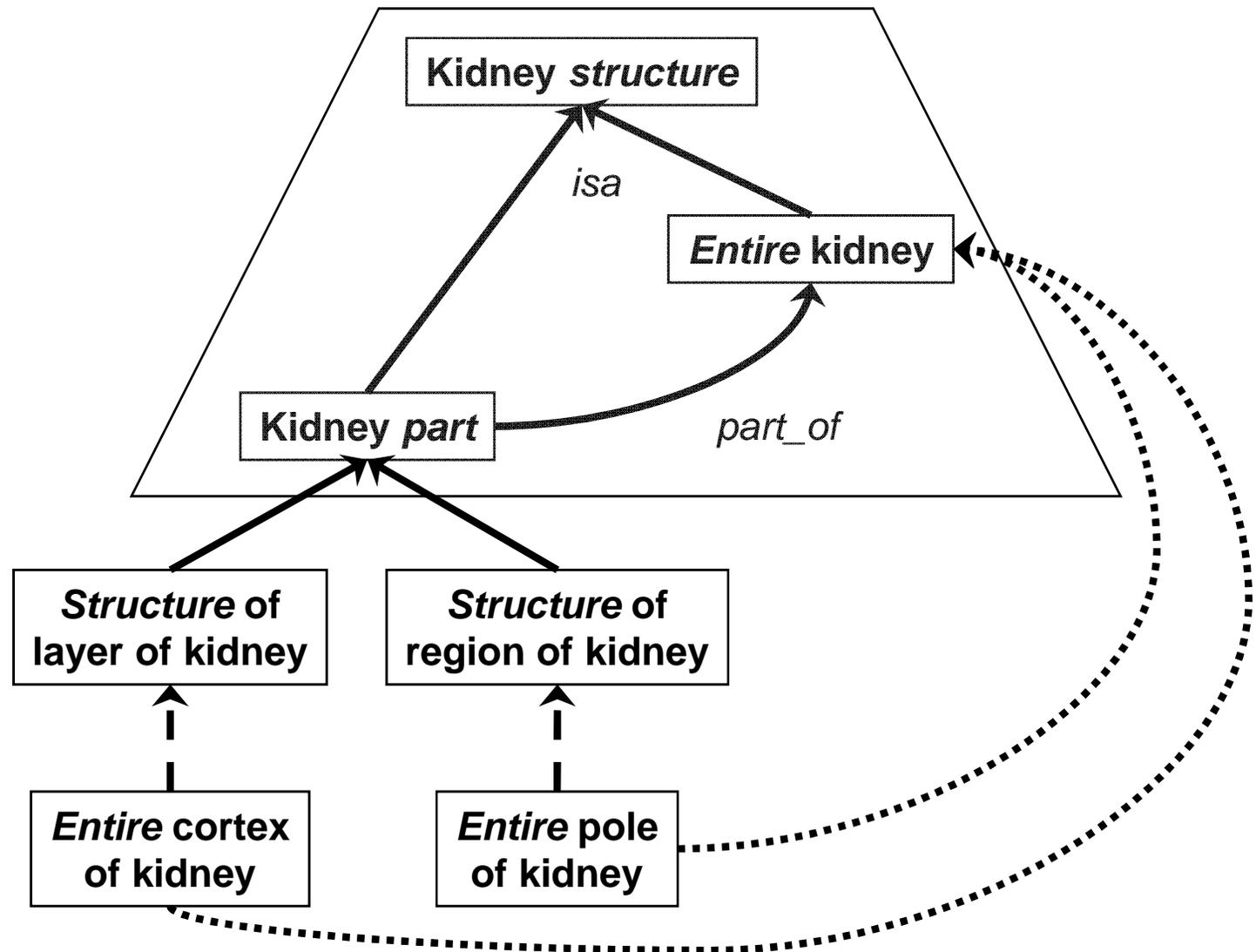
[Schulz & al., 1997]

[Schulz & al., 1998]

[Bodenreider et al, 2006]



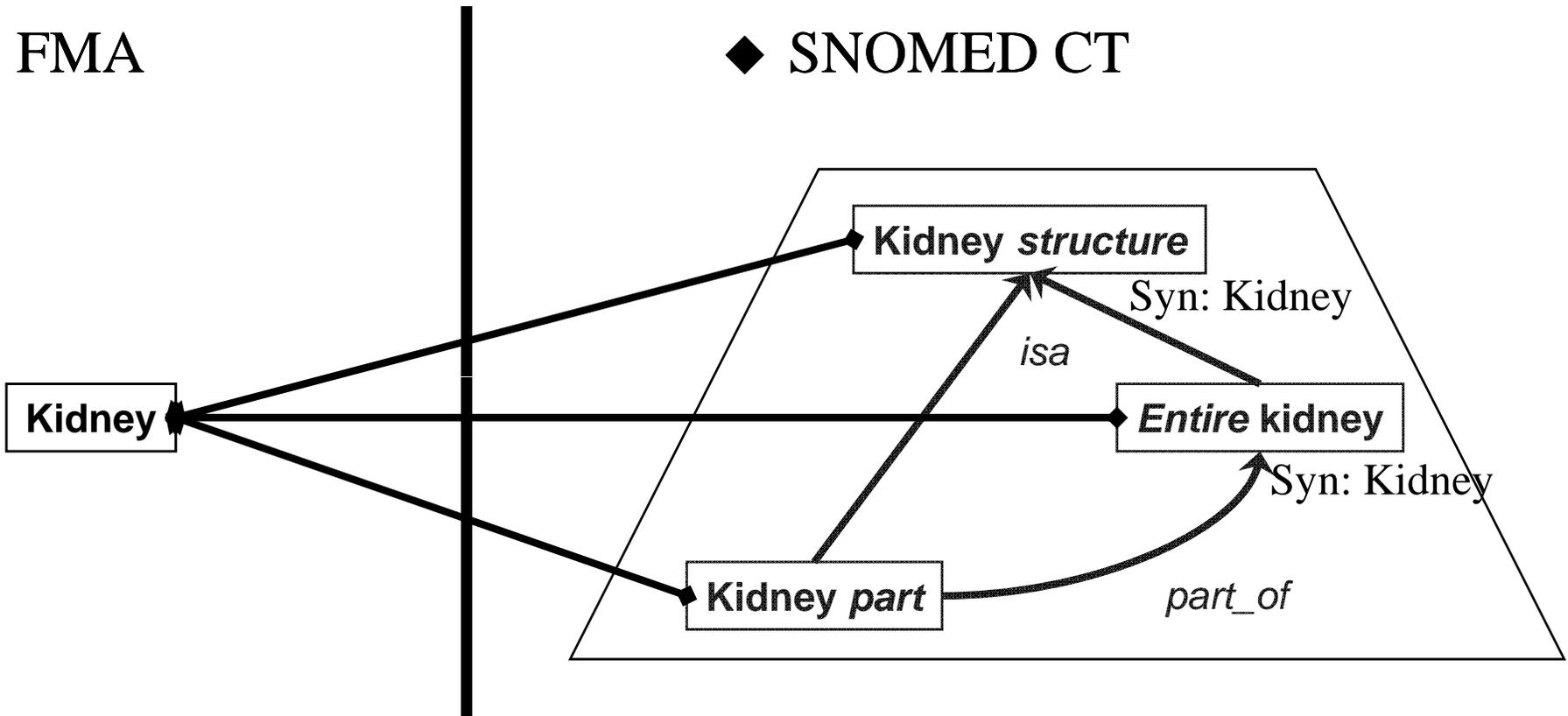
Mereological inference through *isa*



FMA mapping goes to *Entire*

◆ FMA

◆ SNOMED CT



Prostate FMA vs. SNOMED CT

- ◆ Shared synonyms: FMA 1/1, SNOMED CT 1/2
 - Prostate
- ◆ Shared relations
 - *Isa*: no
 - FMA
 - Lobular organ
 - SNOMED CT
 - Prostatic structure
 - Male internal genital organ
 - Entire male genital organ



Prostate FMA vs. SNOMED CT

◆ Shared relations

- *Part of*: almost

- FMA

- Genital system

- Content of *male* pelvis

- Set of male pelvic viscera

- Set of pelvic viscera

- SNOMED CT

- Entire minor pelvis

- Entire *male* genital system

- ...



Solutions

Exercise #3

Exercise #3

◆ Cetirizine

- MeSH

- Using the MeSH browser

<http://www.nlm.nih.gov/mesh/MBrowser.html>

- SNOMED CT

- Using the online browser from U. Sydney

<http://www.cs.usyd.edu.au/~hitru/sct/A3.cgi>



Cetirizine in MeSH (1)

Entry Term	(2-(4-((4-Chlorophenyl)phenylmethyl)-1-piperazinyl)ethoxy)acetic Acid
Entry Term	Alerisin
Entry Term	Aliud Brand of Cetirizine Dihydrochloride
Entry Term	Alpharma Brand of Cetirizine Dihydrochloride
Entry Term	AWD.pharma Brand of Cetirizine Dihydrochloride
Entry Term	Azupharma Brand of Cetirizine Dihydrochloride
Entry Term	Basics Brand of Cetirizine Dihydrochloride
Entry Term	Cetalerg



Entry Term	Voltric
Entry Term	Wolff Brand of Cetirizine Dihydrochloride
Entry Term	Wörwag Brand of Cetirizine Dihydrochloride
Entry Term	Zetir
Entry Term	Zirtek
Entry Term	Zyrtec

Cetirizine in MeSH (2)

Heterocyclic Compounds [D03]

Heterocyclic Compounds, 1-Ring [D03.383]

Piperazines [D03.383.606]

Hydroxyzine [D03.383.606.515]

▶ Cetirizine [D03.383.606.515.200]

Pharm. Action	<u>Anti-Allergic Agents</u>
----------------------	-----------------------------

Pharm. Action	<u>Histamine H1 Antagonists, Non-Sedating</u>
----------------------	---



Comments on Cetirizine in MeSH

◆ 45 entry terms

- Various generic and brand names
- Chemical formula
- Code (P-071)

◆ Hierarchy

- *Isa*: Piperazines [chemistry]

◆ Pharmacologic action

- Anti-Allergic Agents
- Histamine H1 Antagonists, Non-Sedating



Cetirizine in SNOMED CT (1)

15 results found for cetirizine:

Previous Next

#	Concept ID	Fully Specified Name	Preferred Terms and Synonyms
1	108655000	Cetirizine (product)	Cetirizine [PT]
2	372523007	Cetirizine (substance)	Cetirizine [PT]

Cetirizine (substance)

CONCEPT						
Concept ID	Fully Specified Name	Concept Status	CTV3ID	SNOMED ID	Is Primitive	
372523007	Cetirizine (substance)	Current (0)	XUVwU	F-61523	Primitive (1)	
DESCRIPTIONS and SYNONYMS						
Description ID	Term	Description Status	Description Type	Language Code	Initial Capital Status	
1	1211057019	Cetirizine	Current (0)	Preferred (1)	en	Capitalization meaningless (0)
PARENTS						
Concept ID	FSN for Parent Concept (This Concept IS A)					
1	372624008	Non-sedating antihistamine (substance)				
CHILDREN						
Concept ID	FSN for Child Concept					
1	108656004	Cetirizine hydrochloride (substance)				

Cetirizine in SNOMED CT (2)

CONCEPT				
Concept ID	Fully Specified Name	Concept Status	CTV3ID	
108655000	Cetirizine (product)	Current (0)	x01Dq	
DESCRIPTIONS and SYNONYMS				
Description ID	Term	Description Status	Description Type	
1	173189012	Cetirizine	Current (0)	Preferred (1)
PARENTS				
Concept ID	FSN for Parent Concept (This Concept IS A)			
1	349956006	⌚ Non-sedating antihistamine (product)		
ATTRIBUTES				
Concept ID	FSN for Target Concept	Relationship Type		
1	372523007	⌚ Cetirizine (substance)	⌚ Has active ingredient (attribute) {Cetirizine hydroc	
CHILDREN				
Concept ID	FSN for Child Concept			
1	320818006	⌚ Cetirizine dihydrochloride 10mg tablet (product)		
2	320820009	⌚ Cetirizine dihydrochloride 1 mg/1 mL s/f liquid (product)		
3	371746005	⌚ Cetirizine dihydrochloride 5mg tablet (product)		
4	375571002	⌚ Cetirizine hydrochloride 5mg tablet (product)		
5	375572009	⌚ Cetirizine hydrochloride 10mg tablet (product)		
6	375573004	⌚ Cetirizine hydrochloride 5mg/5 mL syrup (product)		
7	400462001	⌚ Cetirizine hydrochloride+pseudoephedrine hydrochloride (product)		
8	409491005	⌚ Cetirizine hydrochloride 5mg chewable tablet (product)		
9	409492003	⌚ Cetirizine hydrochloride 10mg chewable tablet (product)		

Comments on Cetirizine in SNOMED CT

- ◆ Ambiguous term
 - Cetirizine (product)
 - Cetirizine (substance)
- ◆ Hierarchy
 - *Isa*: Non-sedating antihistamine (substance) [pharmacologic action]
- ◆ No associative relations

Hierarchy	Subtype hierarchy
138875005	SNOMED CT Concept
+ C 362981000	qualifier value
+ C 106237007	linkage concept
+ C 370115009	special concept
+ C 48176007	social context
+ C 419891008	record artifact
+ C 363787002	observable entity
+ C 308916002	environment or geographical location
+ C 123038009	specimen
+ C 254291000	staging and scales
+ C 123037004	body structure
+ C 272379006	event
+ C 78621006	physical force
+ C 404684003	clinical finding
+ C 260787004	physical object
+ C 410607006	organism
+ C 71388002	procedure
+ C 373873005	pharmaceutical / biologic product
+ C 243796009	situation with explicit context
+ C 105590001	substance

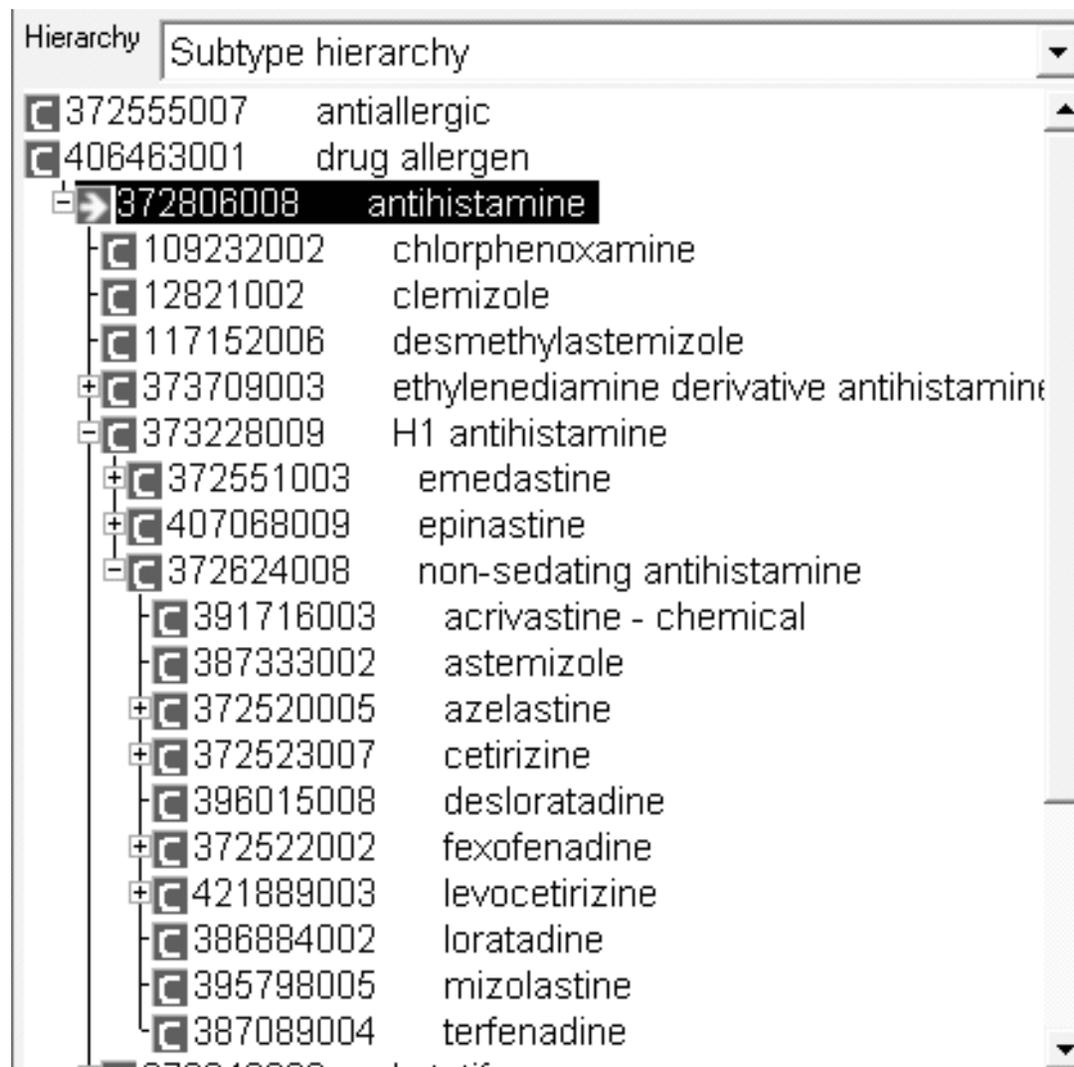


Cetirizine MeSH vs. SNOMED CT (1)

- ◆ Shared synonyms: MeSH 1/45, SNOMED CT 1/1
 - Cetirizine
- ◆ Shared relations: none
 - MeSH:
 - *Isa*: <chemistry>
 - Associative: <pharmacologic action>
 - SNOMED CT
 - *Isa*: < pharmacologic action>
 - Associative: none



Cetirizine MeSH vs. SNOMED CT (2)



Summary

- ◆ Differing representations
 - Not necessarily inconsistent
 - Consistency may be difficult to assess automatically
- ◆ Often due to idiosyncratic representation in one ontology

- ◆ Hindrance to ontology alignment and evaluation methods relying on shared relations



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 9, 2008 – Session #2

Extending Biomedical Ontologies



Olivier Bodenreider

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

Overview

- ◆ Corpus terminology
- ◆ Identify terms in biomedical text (in reference to the UMLS)
- ◆ Identify additional terms
- ◆ Place these terms in UMLS hierarchies

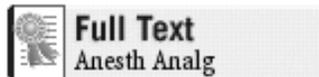
[Bodenreider, ACL 2002]



Tiny corpus One MEDLINE abstract

□ 1: [Anesth Analg](#). 2008 Jun;106(6):1813-9.

[Related Articles,](#)
[Links](#)



PMID: 18499615

Free cortisol in sepsis and septic shock.

Bendel S, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E, Finnsepsis Study Group.

► [Collaborators \(26\)](#)

Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. **METHODS:** In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. **RESULTS:** Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90$, $P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 ± 386 nmol/L vs 793 ± 439 nmol/L, $P = 0.44$). Nonsurvivors had higher calculated serum free (209 ± 151 nmol/L) and total (980 ± 458 nmol/L) cortisol concentrations than survivors (119 ± 111 nmol/L, $P = 0.002$, and 704 ± 383 nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%.

CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.



Identify UMLS concepts with MetaMap

Semantic Knowledge Representation - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://skr.nlm.nih.gov/

Google

Semantic Knowledge Representation

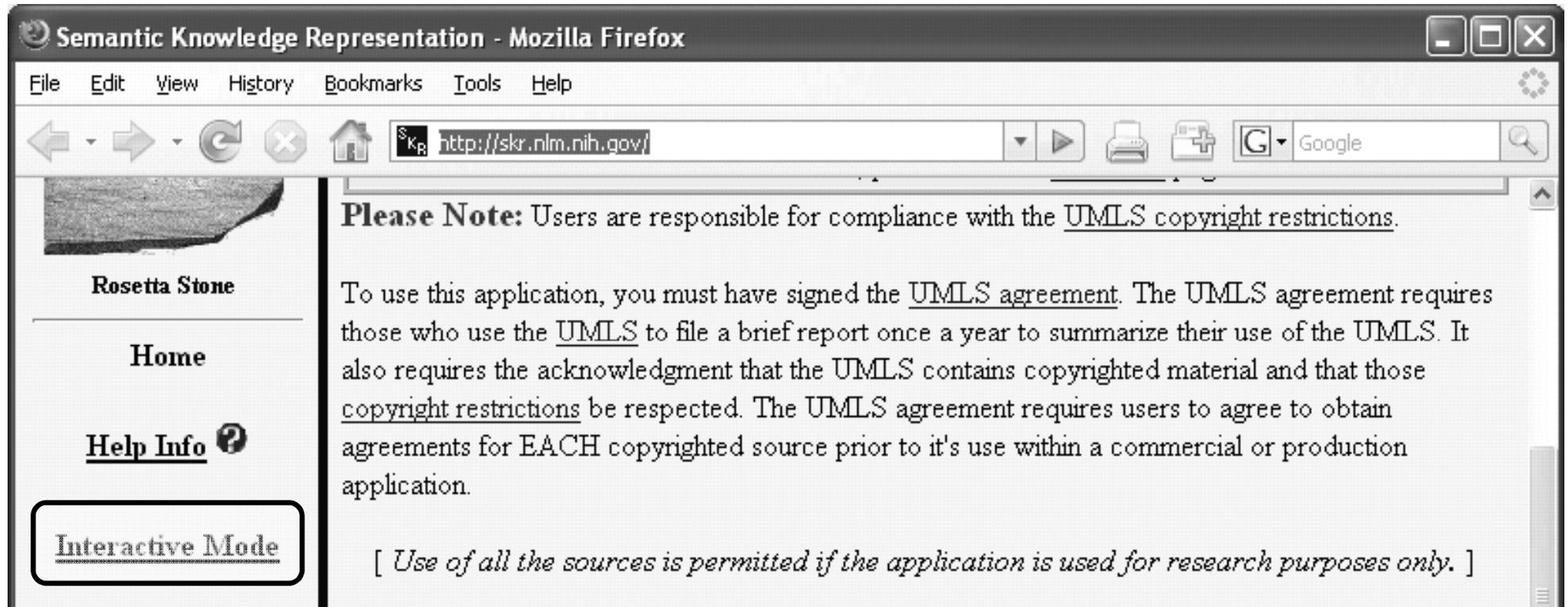
Arbiter
Edgar
SemRep
MetaMap
MMI
UMLS Metathesaurus
SPECIALIST Lexicon
UMLS Semantic Network

Home NLM & LNCRC SKR

<http://skr.nlm.nih.gov/>



Interactive mode



The screenshot shows a Mozilla Firefox browser window titled "Semantic Knowledge Representation - Mozilla Firefox". The address bar contains the URL "http://skr.nlm.nih.gov/". The page content includes a "Please Note" section with the following text:

Please Note: Users are responsible for compliance with the [UMLS copyright restrictions](#).

To use this application, you must have signed the [UMLS agreement](#). The UMLS agreement requires those who use the [UMLS](#) to file a brief report once a year to summarize their use of the UMLS. It also requires the acknowledgment that the UMLS contains copyrighted material and that those [copyright restrictions](#) be respected. The UMLS agreement requires users to agree to obtain agreements for EACH copyrighted source prior to it's use within a commercial or production application.

[*Use of all the sources is permitted if the application is used for research purposes only.*]

On the left side of the browser window, there is a sidebar with the following links: "Rosetta Stone", "Home", "Help Info" (with a question mark icon), and "Interactive Mode" (which is highlighted with a rounded rectangle).



Interactive MetaMap

Semantic Knowledge Representation - Interactive Mode - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://skr.nlm.nih.gov/interactive/index.shtml

Interactive Mode

User: umls1: [NLM](#) » [LHNCBC](#) » [SKR](#) » [Interactive](#)

[Home](#)

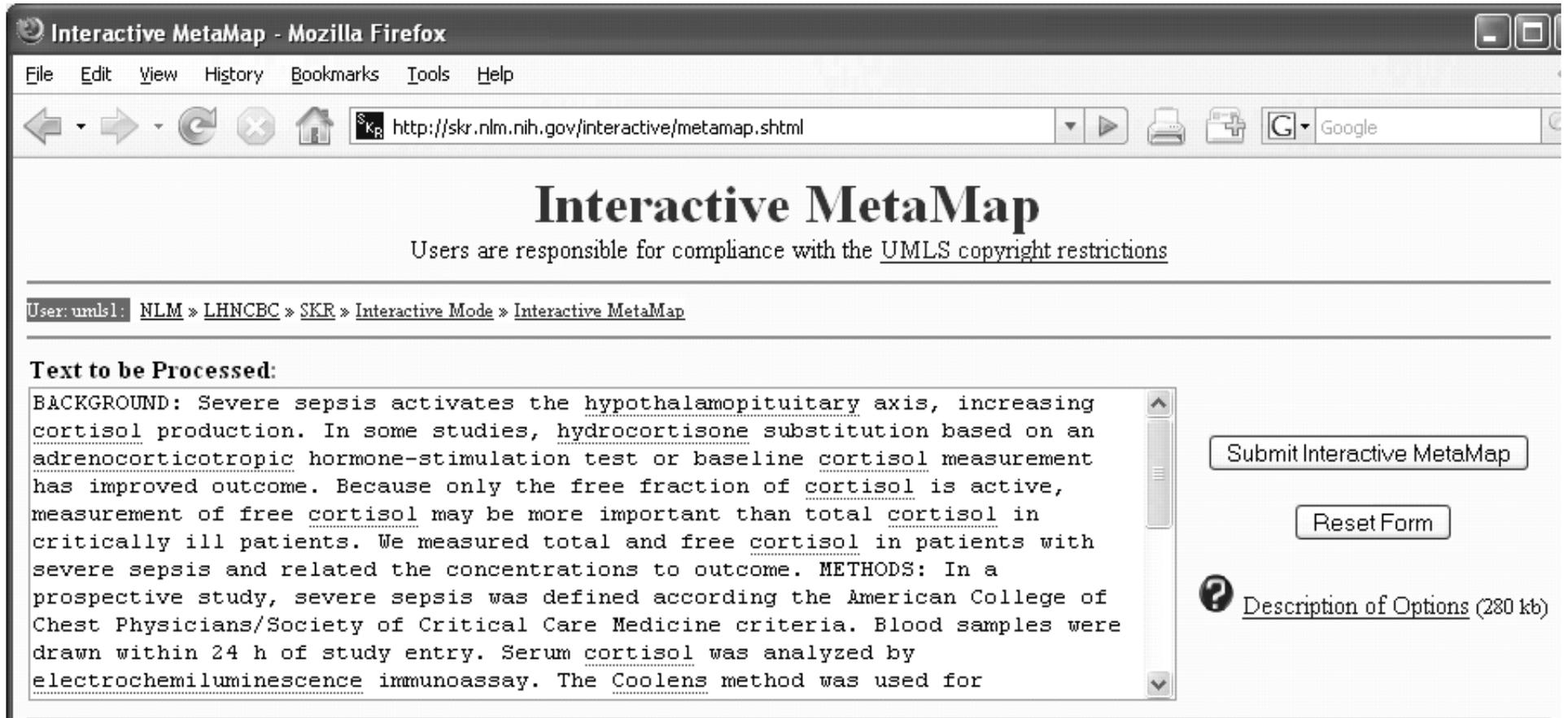
[Interactive MetaMap](#)

Please NOTE:

The Interactive mode is only intended for the testing of the various programs and their options.



Paste abstract



The screenshot shows a Mozilla Firefox browser window titled "Interactive MetaMap - Mozilla Firefox". The address bar contains the URL "http://skr.nlm.nih.gov/interactive/metamap.shtml". The main heading is "Interactive MetaMap" with a sub-heading "Users are responsible for compliance with the [UMLS copyright restrictions](#)". Below this is a breadcrumb trail: "User: umsl: [NLM](#) » [LHNCBC](#) » [SKR](#) » [Interactive Mode](#) » [Interactive MetaMap](#)".

Text to be Processed:

BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotrophic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for

Submit Interactive MetaMap

Reset Form

? [Description of Options \(280 kb\)](#)



Select options

Data Options		
Knowledge Source:	2007 (2007 Full - 2007AA) ▼	Data Model: Strict Model (-A) ▼
<input type="checkbox"/> Data Version (-V): <input type="text"/>		
Output Display Options	Output Display Options (continued)	Behavior Options (continued)
<input type="checkbox"/> Tagger Output (-T)	<input type="checkbox"/> Show Original Phrases (-H)	<input type="checkbox"/> Prefer Multiple Concepts (-Y)
<input type="checkbox"/> Variants (-v)	<input type="checkbox"/> Show Concept's Sources (-G)	<input checked="" type="checkbox"/> Best Mappings Only (-b)
<input checked="" type="checkbox"/> Plain Syntax (-p)	<input type="checkbox"/> Show Acronym/Abbreviations (-j)	<input type="checkbox"/> Truncate Candidates Mapping (-X)
<input type="checkbox"/> Syntax (-x)		<input type="checkbox"/> Use Word Sense Disambiguation (-y)
<input checked="" type="checkbox"/> Candidates (-c)	Behavior Options	Browse Mode Options
<input type="checkbox"/> Number Candidates (-n)	<input checked="" type="checkbox"/> Tag Text (-t)	<input type="checkbox"/> Term Processing (-z)
<input checked="" type="checkbox"/> Semantic Types (-s)	<input type="checkbox"/> No Derivational Variants (-d)	<input type="checkbox"/> Allow Overmatches (-o)
<input checked="" type="checkbox"/> Show CUIs (-I)	<input checked="" type="checkbox"/> Adj/Noun Derivational Variants (-D)	<input type="checkbox"/> Allow Concept Gaps (-g)
<input checked="" type="checkbox"/> Mappings (-m)	<input checked="" type="checkbox"/> No Acronym/Abbreviation Variants (-a)	Misc. Options
<input type="checkbox"/> Show Preferred Names Only (-O)	<input type="checkbox"/> Unique Acronym/Abbrev Variants (-u)	<input type="checkbox"/> Display Phrases Only
<input type="checkbox"/> MMI Output (-M)	<input type="checkbox"/> Ignore Stop Phrases (-K) <i>(System Use)</i>	<input type="checkbox"/> Dynamic Variant Generation (-8)
<input type="checkbox"/> Machine Output (-q) ?	<input checked="" type="checkbox"/> Stop Large N (-l)	
<input type="checkbox"/> Fielded Output (-f) ?	<input type="checkbox"/> Threshold (-r): <input type="text"/>	
<input type="checkbox"/> Formal Tagger Output (-F)	<input type="checkbox"/> Ignore Word Order (-i)	
<input type="checkbox"/> Fielded MMI output (-N)		



Run MetaMap

"Restrict to" or "Exclude" Vocabulary Sources		
<input type="checkbox"/> Restrict to Sources (-R)	<input type="text"/>	<input type="button" value="Edit"/>
<input type="checkbox"/> Exclude Sources (-e)	<input type="text"/>	<input type="button" value="Edit"/>
"Restrict to" or "Exclude" Semantic Types		
<input type="checkbox"/> Restrict to Semantic Type(s) (-J)	<input type="text"/>	<input type="button" value="Edit"/>
<input type="checkbox"/> Exclude Semantic Type(s) (-k)	<input type="text"/>	<input type="button" value="Edit"/>



Output

```
Processing 00000000.tx.1: BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production.
```

```
Phrase: "Severe sepsis"
```

```
>>>> Phrase
```

```
severe sepsis
```

```
<<<<< Phrase
```

```
>>>> Candidates
```

```
Meta Candidates (8):
```

```
1000 C1719672:Severe Sepsis [Disease or Syndrome]
```

```
861 C0036690:Sepsis (Septicemia) [Disease or Syndrome]
```

```
861 C0243026:Sepsis (Systemic infection) [Disease or Syndrome]
```

```
861 C1090821:Sepsis [Invertebrate]
```

```
789 C0333534:Septic [Functional Concept]
```

```
694 C0205082:Severe [Qualitative Concept]
```

```
694 C1519275:SEVERE (Severe Adverse Event) [Finding]
```

```
694 C1561581:Severe (Allergy Severity - Severe) [Finding]
```

```
<<<<< Candidates
```

```
>>>> Mappings
```

```
Meta Mapping (1000):
```

```
1000 C1719672:Severe Sepsis [Disease or Syndrome]
```

```
<<<<< Mappings
```



Suggest term candidates

- ◆ Not recognized by MetaMap at all
- ◆ Partially identified by MetaMap
- ◆ Missing terms in a concept



Suggest placement in UMLS

- ◆ Use a browser
- ◆ Identify close parent
- ◆ Examine its children
- ◆ Assess placement by comparing with potential siblings



Possible new terms (1)

◆ Hypothalamopituitary axis

- Concept exists: C0678897, but missing exact (neoclassical) synonym
 - hypothalamic pituitary axis
 - hypothalamus hypophysis axis
 - hypothalamus-pituitary axis

◆ American College of Chest Physicians

- Similar to other American Colleges (e.g., American College of Physicians ())
- Integrate as a child of Professional Organization or Group (C1522486)
- NB: instance, cannot be a child of ACP



Possible new terms (2)

◆ Free cortisol

- Identified as a substance (C0443476), not a laboratory procedure / test result
 - Cortisol, free measurement (C0236401)

◆ Coolens method

- Missing term / concept
- Method for estimating (not measuring directly) the free fraction of cortisol



Possible new terms (3)

- ◆ Electrochemiluminescence immunoassay
 - Missing concept
 - Create as a child of Chemiluminescence assay (C0201709)
- ◆ Nonsurvivors
 - Survivors exists as a concept (C0206194)
 - Create as a child of Patients (C0030705)



Agenda

Monday, June 9	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
Tuesday, June 10	Interfaces to Biomedical Ontologies	Searching and Analyzing Biomedical Concepts	Contrasting Biomedical Ontologies
Wednesday, June 11	Critical Analysis of Biomedical Ontologies	Extending Biomedical Ontologies	Using Biomedical Ontologies for Data Integration





The University of Utah
Biomedical Informatics

Short course – Summer 2008 Biomedical Ontology in Practice

June 11, 2008 – Session #3

Using Biomedical Ontologies for Data Integration



Olivier Bodenreider

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

Overview

- ◆ Motivation
- ◆ Some practical considerations and issues
 - Integration approaches
 - Concept repositories
 - Using existing mappings
 - Creating mappings through the UMLS
 - Comparing semantic descriptions
- ◆ Thinking outside the integration box



Motivation

Motivation Translational research

- ◆ “Bench to Bedside”
- ◆ Integration of clinical and research activities and results
- ◆ Supported by research programs
 - NIH Roadmap
 - Clinical and Translational Science Awards (CTSA)
- ◆ Requires the effective integration and exchange and of information between
 - Basic research
 - Clinical research



Translational research NIH Roadmap



NIH Roadmap FOR MEDICAL RESEARCH



Re-engineering the Clinical Research Enterprise

- ▶ [Overview](#)
- ▶ [Implementation Group Members](#)
- ▶ [Funding Opportunities](#)
- ▶ [Funded Research](#)
- ▶ [Meetings](#)
- ▶ [Mid-course Reviews](#)

- ▶ [CTSAweb.org](#) [EXIT Disclaimer](#)

TRANSLATIONAL RESEARCH

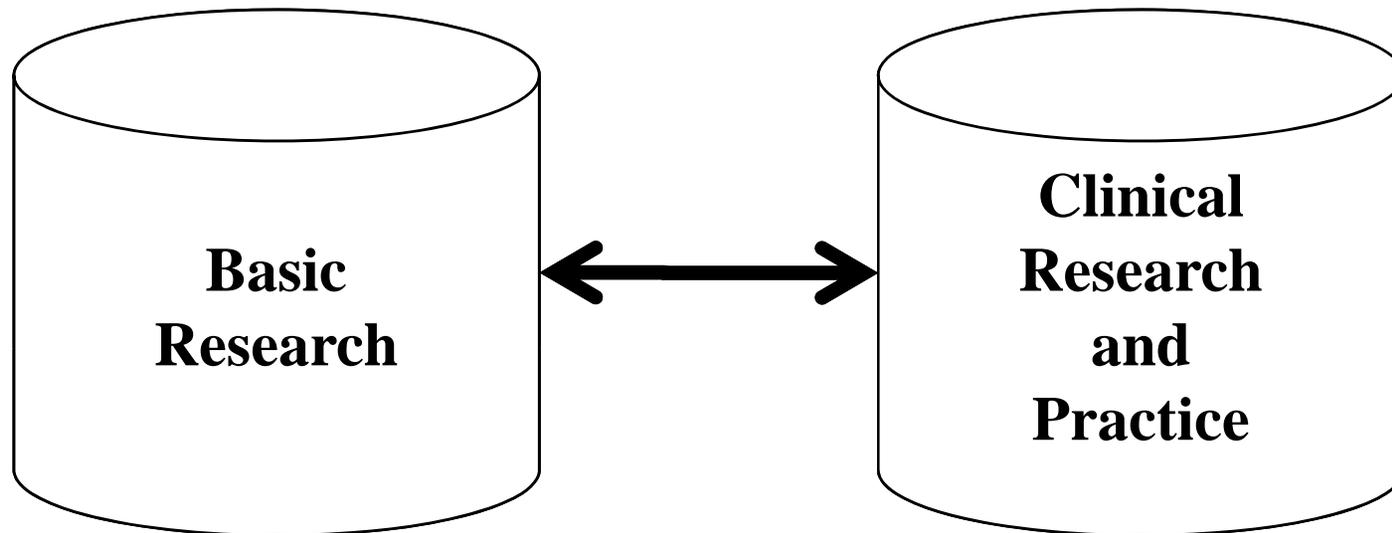
OVERVIEW

To improve human health, scientific discoveries must be translated into practical applications. Such discoveries typically begin at "the bench" with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient's "bedside."

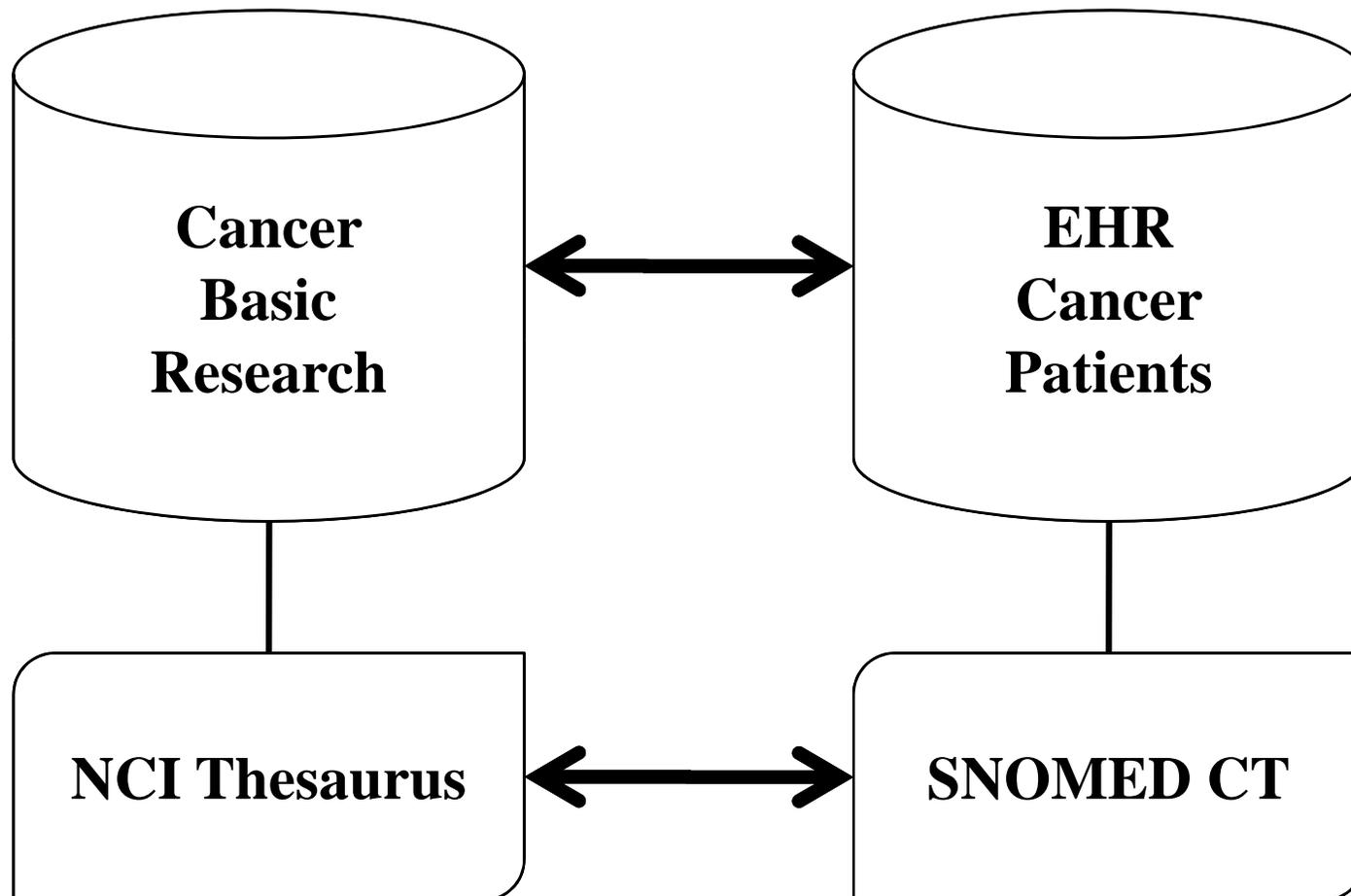
Scientists are increasingly aware that this bench-to-bedside approach to translational research is really a two-way street. Basic scientists provide clinicians with new tools for use in patients and for assessment of their impact, and clinical researchers make novel observations about the nature and progression of disease that often stimulate basic investigations.



Motivation Translational research



Terminology and translational research



Some practical considerations and issues

Integration approaches

Approaches to data integration

◆ Warehousing

- Sources to be integrated are transformed into a common format and converted to a common vocabulary
- Normalization through ontologies (e.g., GO annotations)

◆ Mediation

- Local schema (of the sources)
- Global schema (in reference to which the queries are made)
- Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)



Some practical considerations and issues

Concept repositories

(Integrated) concept repositories

- ◆ Unified Medical Language System

<http://umlsks.nlm.nih.gov>

- ◆ NCBO's BioPortal

<http://www.bioontology.org/tools/portal/bioportal.html>

- ◆ Open Biomedical Ontologies (OBO)

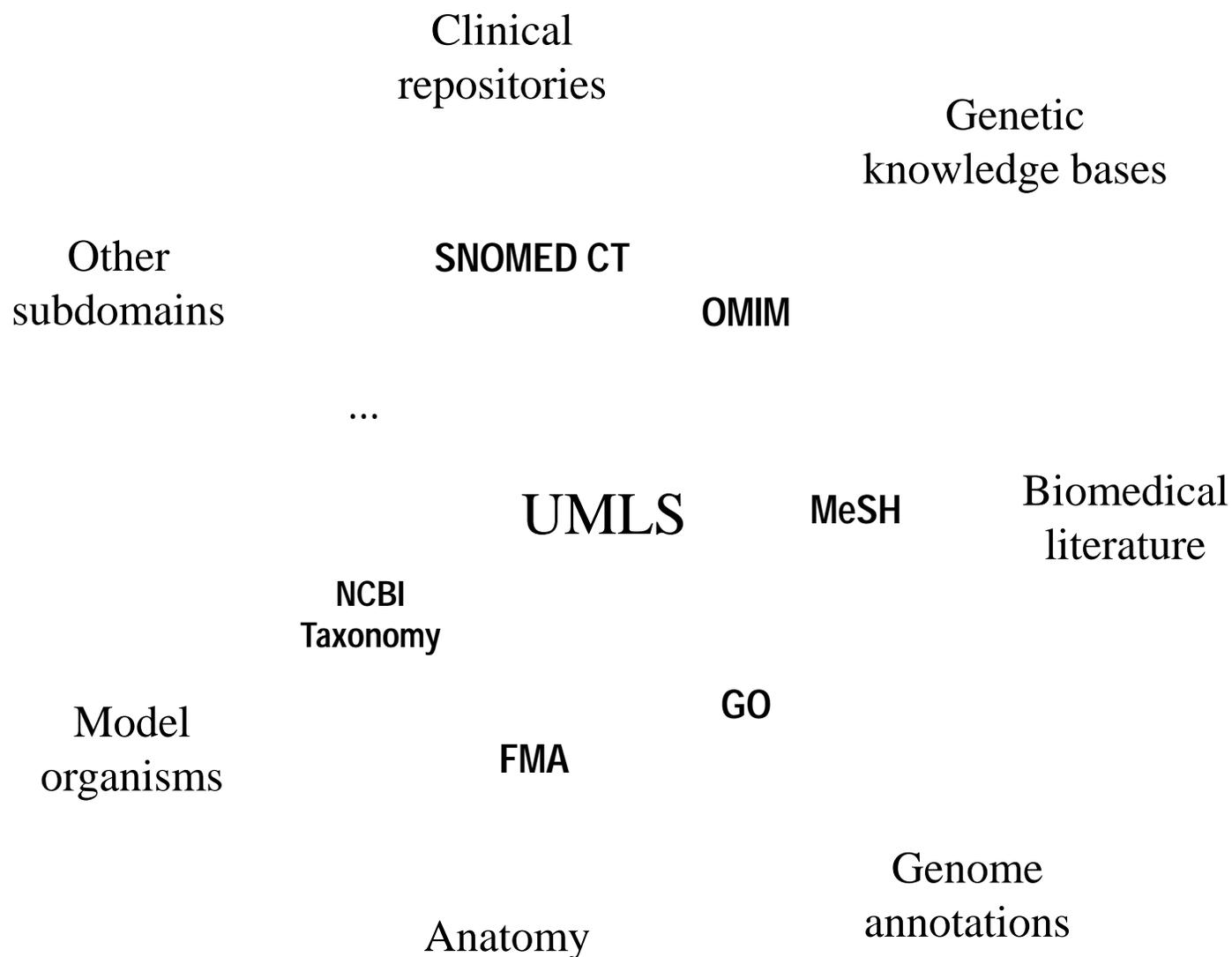
<http://obofoundry.org/>

- ◆ caDSR

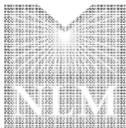
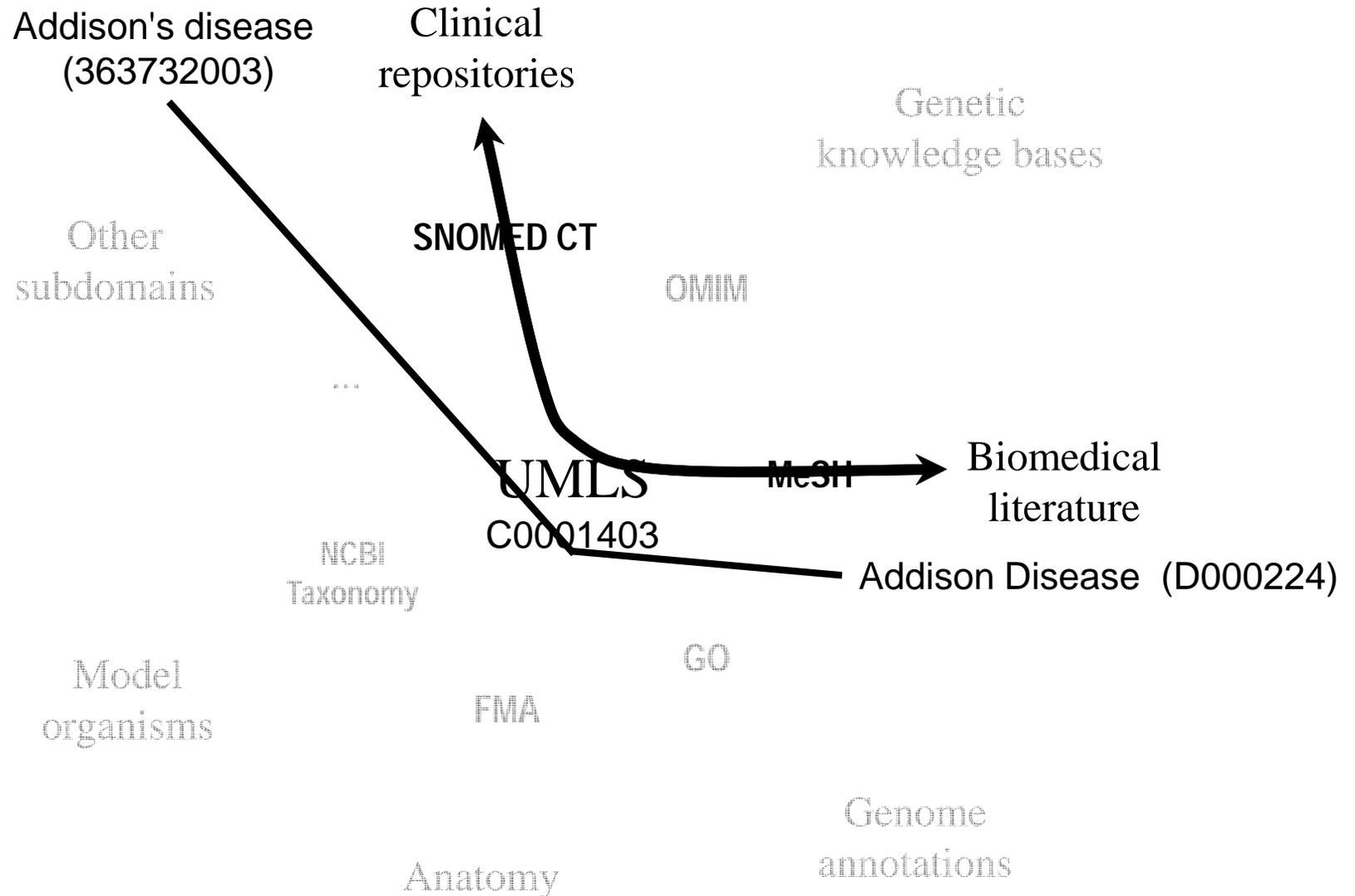
http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr



Integrating subdomains



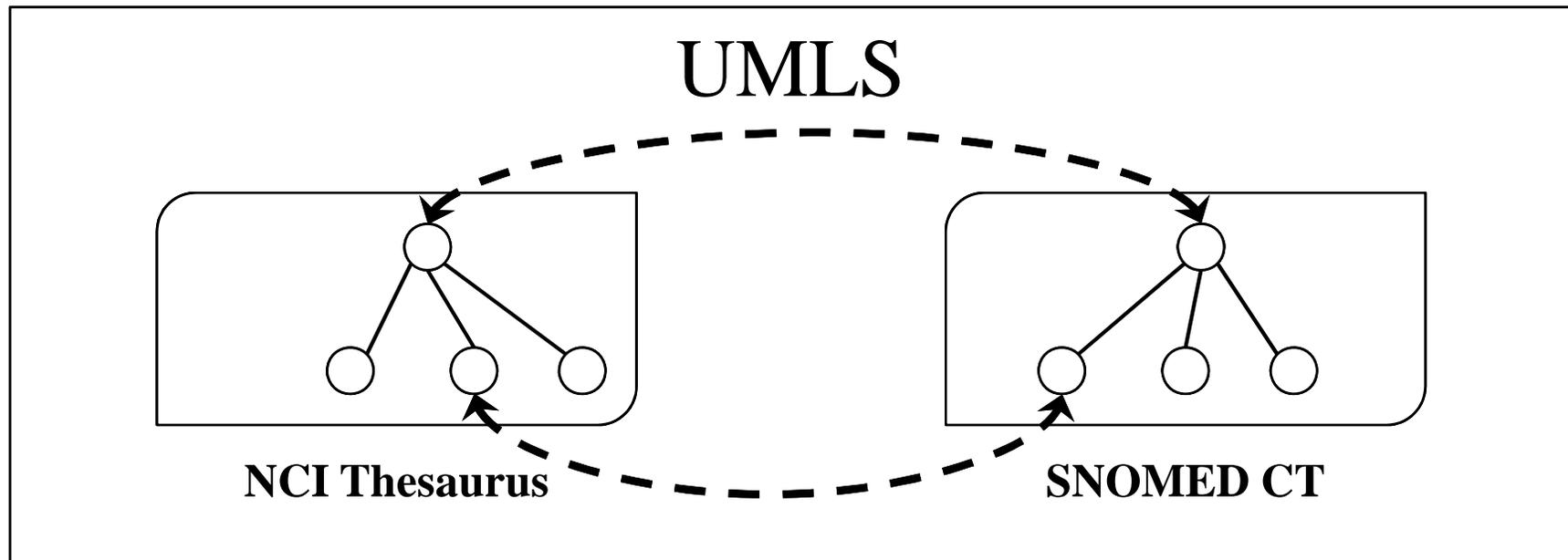
Trans-namespace integration



Some practical considerations and issues

Mappings

Mappings



Mappings

- ◆ Created manually
 - UMLS
- ◆ Created automatically
 - BioPortal

- ◆ Key to enabling semantic interoperability
- ◆ Enabling resource for the Semantic Web



Quality of mappings

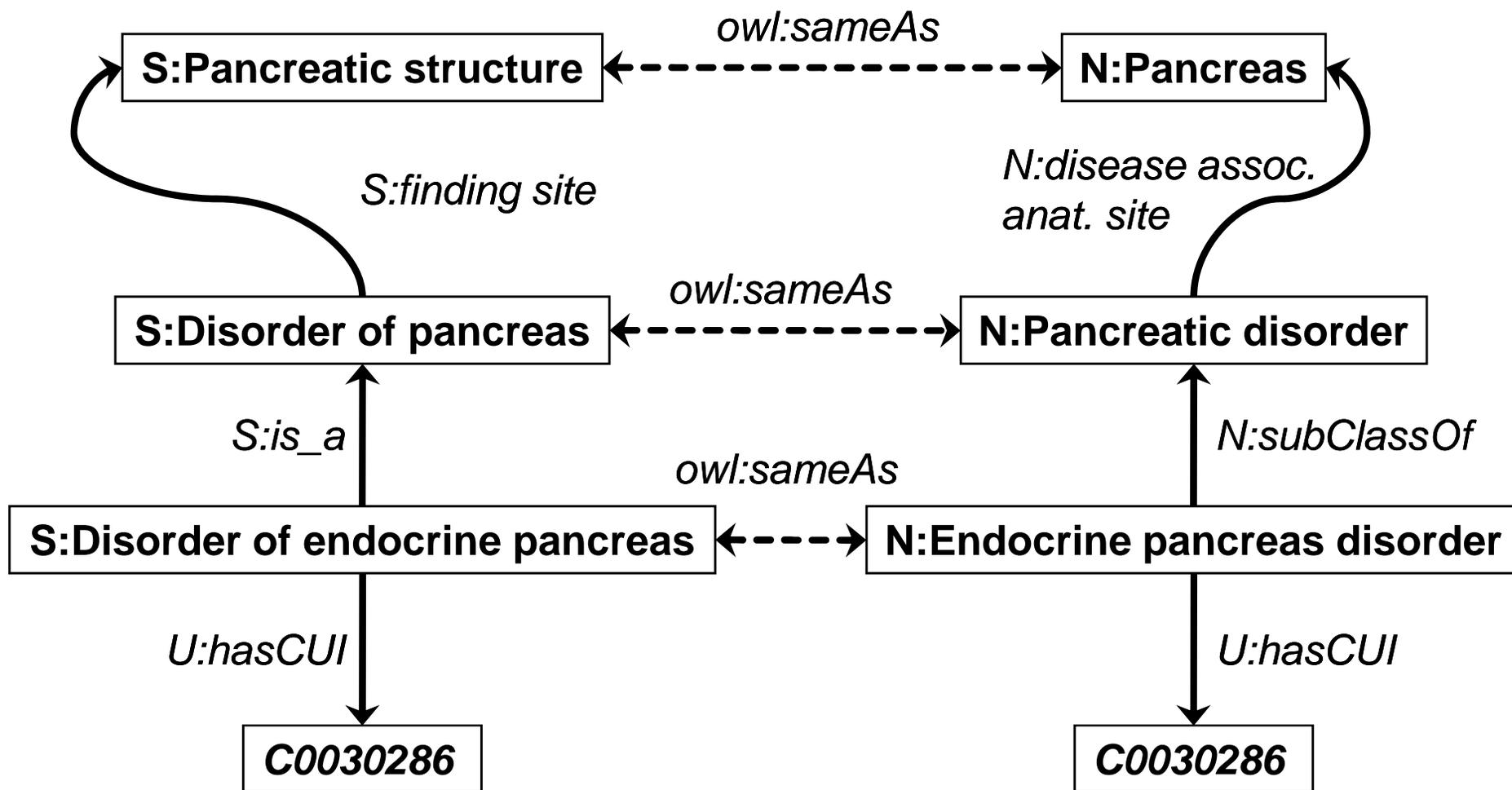
- ◆ Created for a purpose
 - Reusability issues
- ◆ Generally unidirectional
 - Mapping from ontology 1 to ontology 2
 - Not necessarily reversible



Some practical considerations and issues

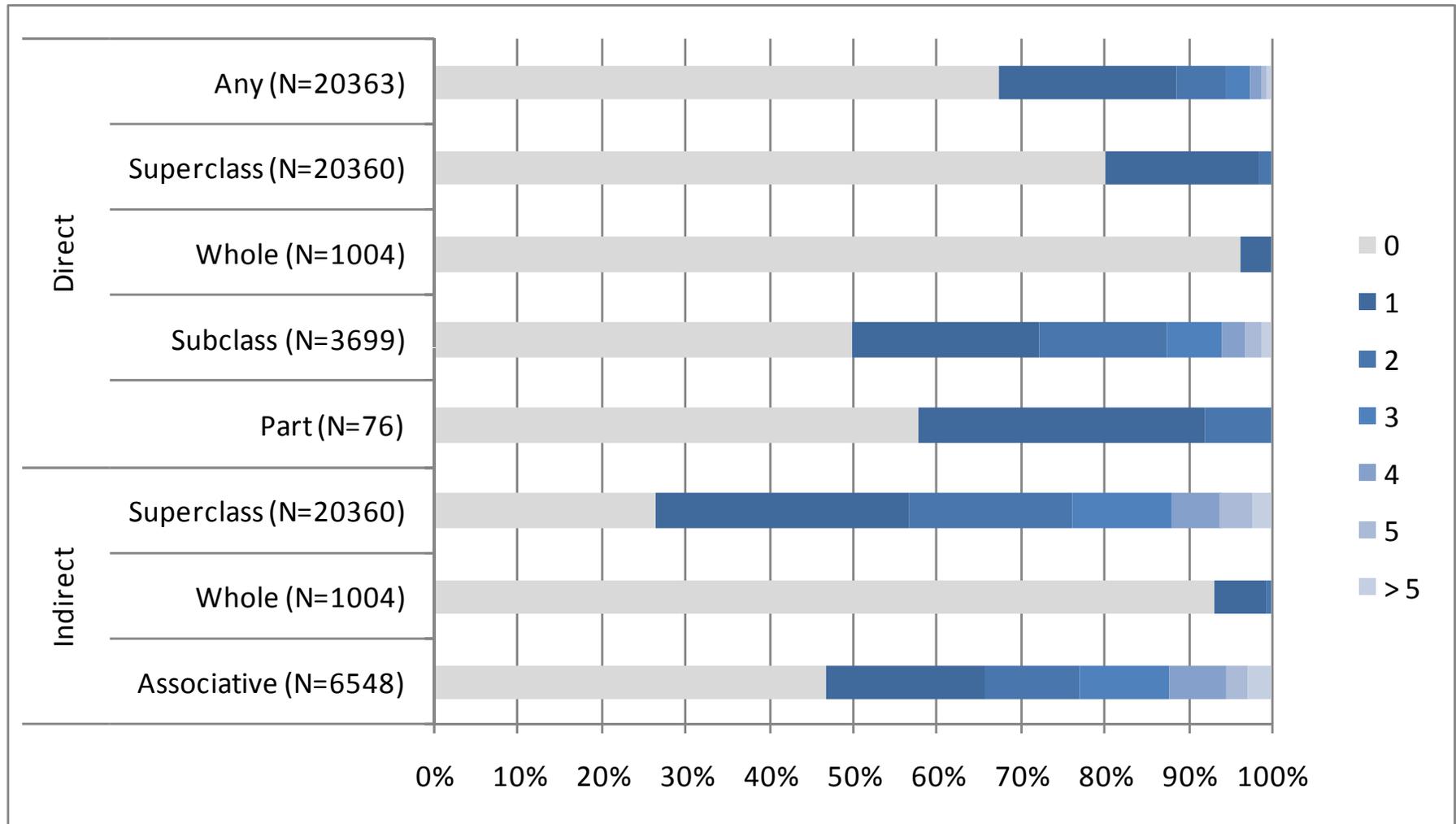
Comparing semantic descriptions

Semantic descriptions Consistent



SNOMED CT/NCI Thesaurus

Limited consistency



Comparing formal definitions

- ◆ Relatively small proportion of relata in common between equivalent concepts from NCI and SNOMED CT
- ◆ Large number of primitive concepts in NCI and SNOMED CT (70-80%)
- ◆ Insufficient for effectively comparing definitions
 - Could not be used for validating the mapping provided by the UMLS

[Bodenreider, KRMed 2008]



Exercises

Exercise #1

- ◆ Check the equivalence (shared relata) between these 2 concepts:
 - NCI Thesaurus: N:Endocrine pancreas disorder
 - SNOMED CT: S:Disorder of endocrine pancreas



Exercise #2

- ◆ Find a correspondence in SNOMED CT for the LOINC term: *Sodium:SCnc:-Pt:Ser/Plas:Qn*
[the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

Axis	Value
Component	Sodium
Property	SCnc – Substance Concentration (per volume)
Timing	Pt – Point in time (Random)
System	Ser/Plas – Serum or Plasma
Scale	Qn – Quantitative
Method	--



Comments on exercise #2

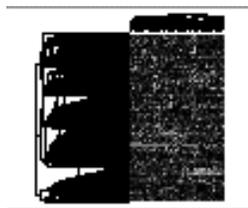
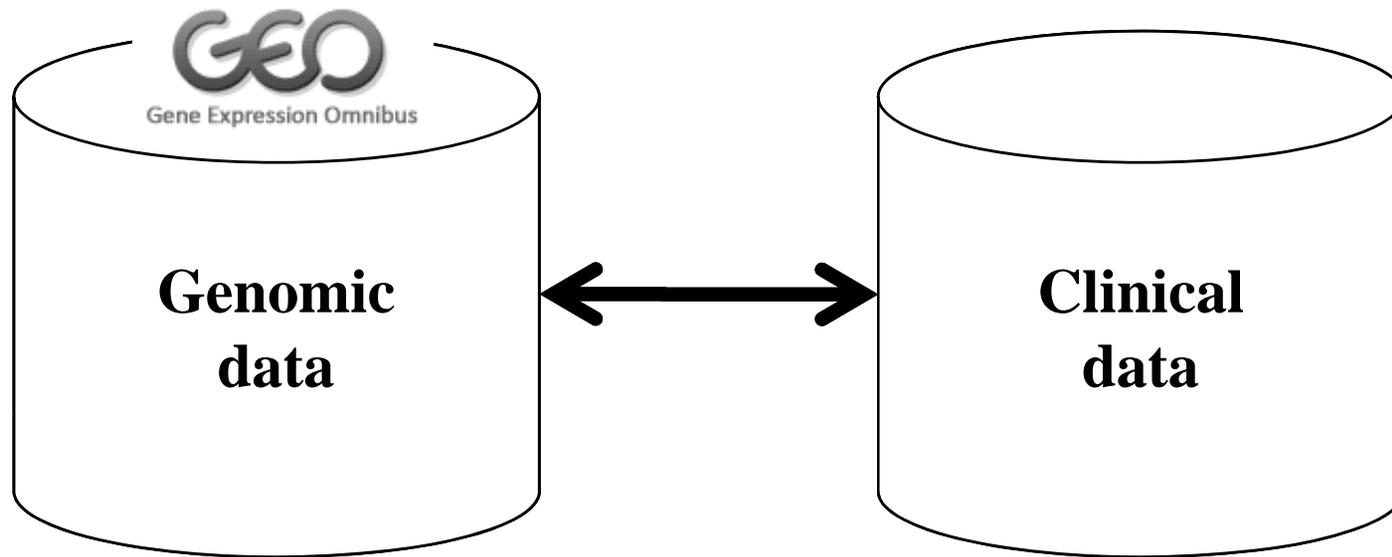
- ◆ Difficult in the absence of a search mechanism on the values of the relations
- ◆ Large number of underspecified descriptions in SNOMED CT
- ◆ 2 separate concepts for plasma and serum concentrations of sodium in SNOMED CT
- ◆ Property, time and scale not represented in SNOMED CT



Thinking outside the integration box

The Butte approach

Integrating genomic and clinical data



GDS Summary	
Accession:	GDS2639 View Expression (GEO profiles)
Title:	Aging and cognitive impairment: hippocampus
DataSet type:	gene expression array-based (RNA, in situ oligonucleotide)
Summary:	Analysis of hippocampus from aged learning-impaired animals on the last day of training in the Morris water maze (MWM) or 21 days post-training. The MWM task is a dorsal hippocampal-dependent task. Results provide insight into the molecular basis of aging-related cognitive impairment.
Platform:	GPL341: Affymetrix GeneChip Rat Expression Set 230 Array RA230A
Chaffin:	Rowe WB, Blalock EM, Chen KC, Kadish I et al. Hippocampal expression analyses reveal selective association of immediate-early, neuroenergetic, and myelinogenic pathways with cognitive impairment in aged rats. <i>J Neurosci</i> 2007 Mar 21;27(12):3098-110. PMID: 17376971
Sample organism:	Rattus norvegicus
Platform organism:	Rattus norvegicus
Feature count:	15923
Value type:	count
Series:	G285666
Series published:	03/15/2007
Last GDS update:	04/27/2007

Coded discharge summaries

Laboratory data

Upregulated genes

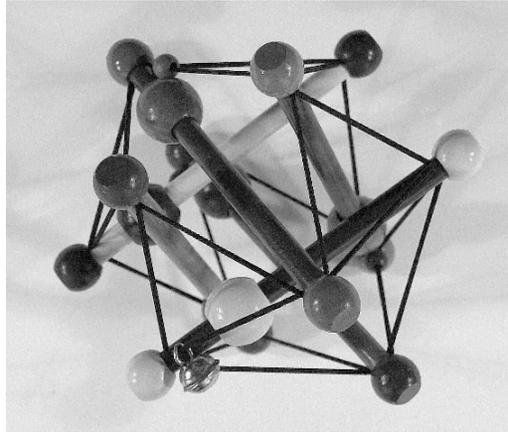
Diseases (extracted from text)



References

- ◆ Dudley J, Butte AJ "Enabling integrative genomic analysis of high-impact human diseases through text mining." *Pac Symp Biocomput* 2008; 580-91
- ◆ Chen DP, Weber SC, Constantinou PS, Ferris TA, Lowe HJ, Butte AJ "Novel integration of hospital electronic medical records and gene expression measurements to identify genetic markers of maturation." *Pac Symp Biocomput* 2008; 243-54
- ◆ Butte AJ, "Medicine. The ultimate model organism." *Science* 2008; 320: 5874: 325-7





Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

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