



# Short course – Summer 2010 Clinical Ontology in Practice

June 15-17, 2010

## Clinical Ontology in Practice



*Olivier Bodenreider*

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

# Objectives

- ◆ Learn about clinical ontologies
  - History
  - Design principles, formalisms and tools
  - What are they?
  - What are they used for?
- ◆ Work with clinical ontologies
  - Search, browse, navigate, query with application programming interfaces
  - Analyze, compare
  - Specific clinical uses (e.g., decisions support, natural language processing, medication reconciliation, e-prescription)
  - Specific issues (e.g., mapping across ontologies, ontologies and information models)



# Agenda

<b>Tuesday, June 15</b> <i>(lecture)</i>	Introduction to Biomedical Ontologies	Design Principles, Formalisms and Tools for Biomedical Ontologies	Biomedical Ontologies - Content and structure - Function
<b>Wednesday, June 16</b> <i>(hands-on)</i>	UMLS	SNOMED CT  LOINC	RxNorm  NDF-RT
<b>Thursday, June 17</b> <i>(discussion)</i>	Decision support  Medication reconciliation	E-prescribing  Natural language processing	Mapping across ontologies  Value sets



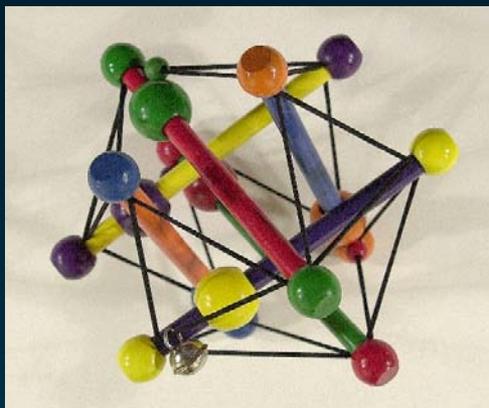
# References Review articles

- ◆ Bodenreider O, Stevens R.  
Bio-ontologies: current trends and future directions.  
Brief Bioinform. 2006 Sep;7(3):256-74.
- ◆ Cimino JJ, Zhu X.  
The practical impact of ontologies on biomedical informatics.  
Yearb Med Inform. 2006:124-35.
- ◆ Bodenreider O.  
Biomedical ontologies in action: role in knowledge management, data integration and decision support.  
Yearb Med Inform. 2008:67-79.



# References Bio-ontology courses

- ◆ Barry Smith, U. Buffalo / NCBO
  - [http://ontology.buffalo.edu/smith/Ontology\\_Course.html](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>



# Medical Ontology Research

Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)

Web: [mor.nlm.nih.gov](http://mor.nlm.nih.gov)



*Olivier Bodenreider*

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



# Short course – Summer 2010 Clinical Ontology in Practice

June 15, 2010 – 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)*

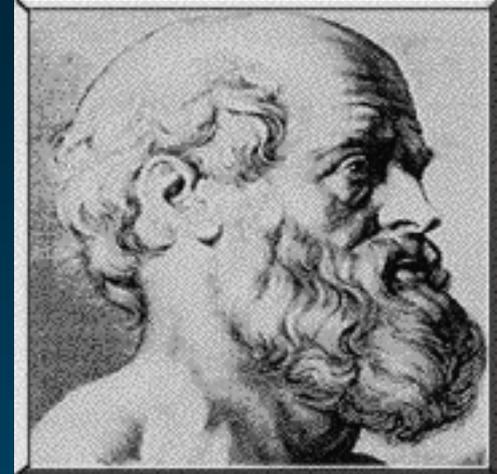
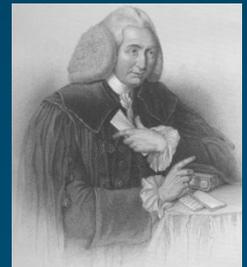
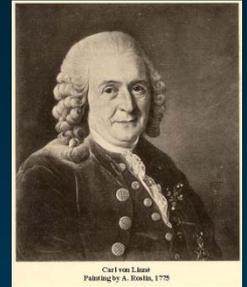


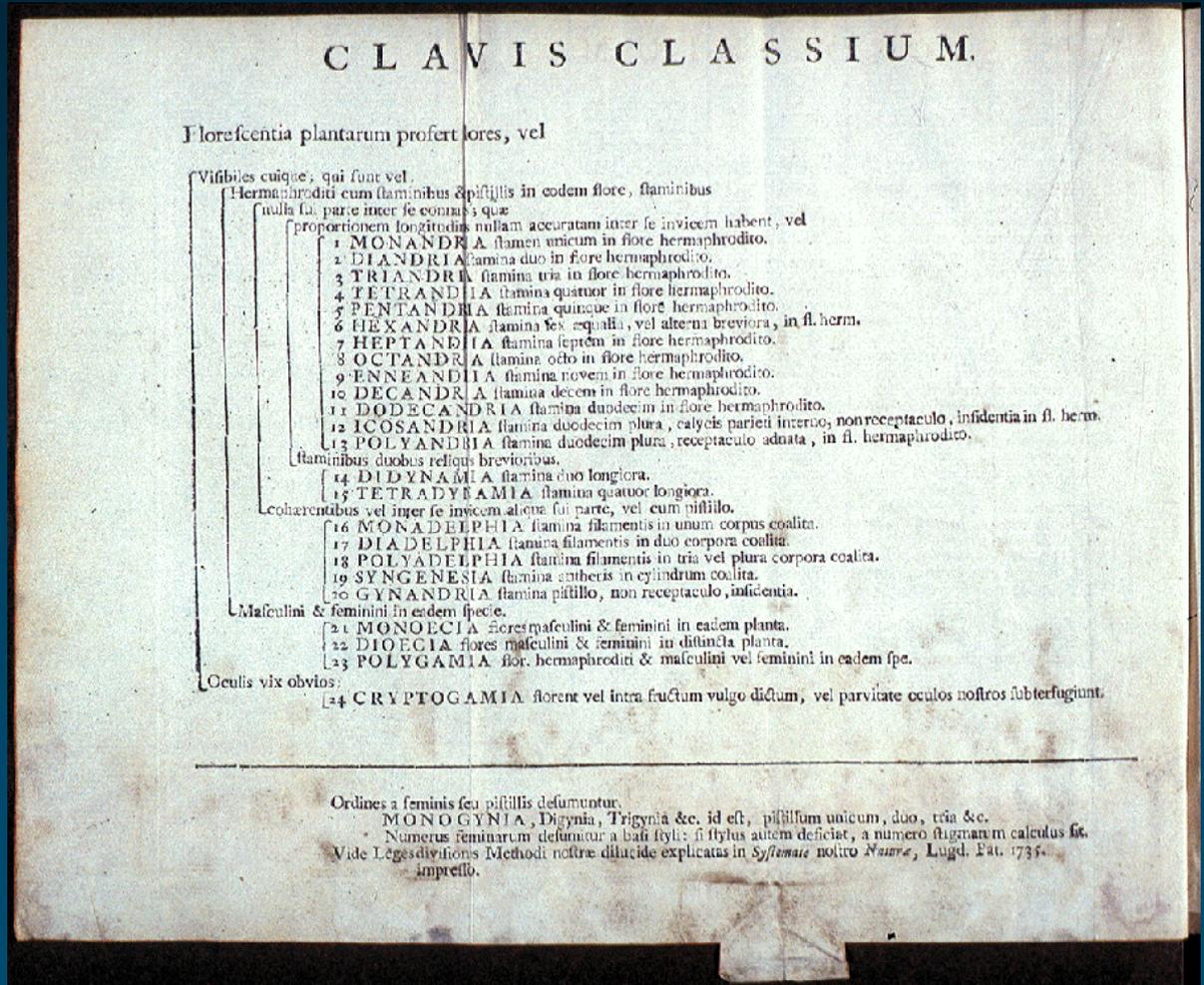
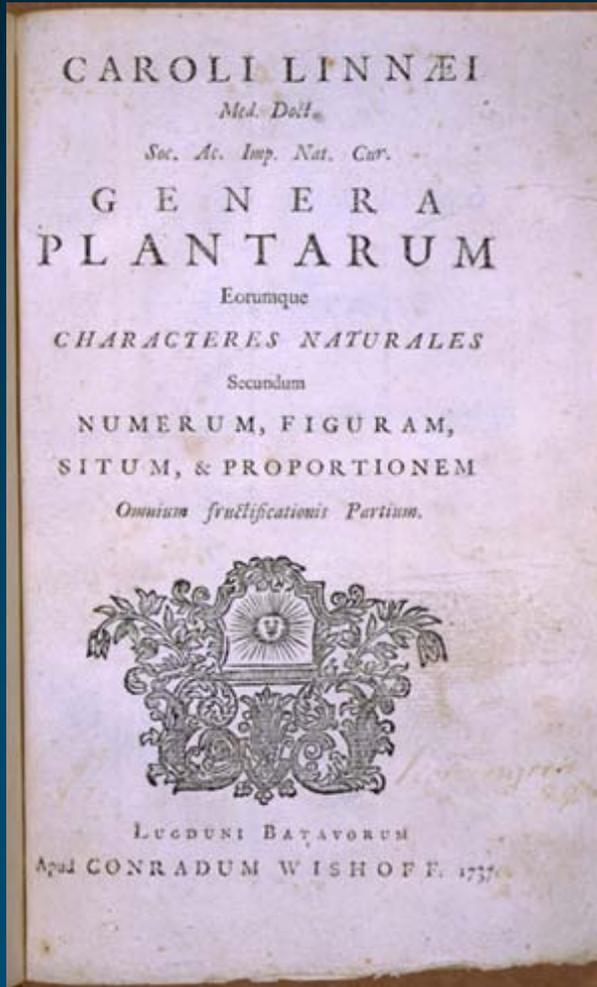
Figure 36 Thomas Sydenham (1624-1689)

# 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

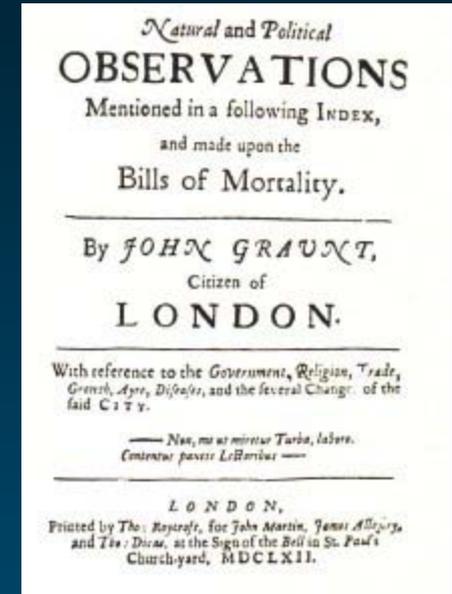
*Synopsis Nosologia Methodicae*

(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 27<sup>th</sup> of December 1664. and  
ending the 19<sup>th</sup> 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. &c

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, &c.

*The Diseases and Casualties this year.*

<b>A</b> Bortive and Stillborne	517	Executed	21	Palfie	30
Aged	1545	Flux and Small Pox	655	Plague	68598
Aque and Peaver	5257	Found dead in Streets, fields, &c.	2	Plasmod	6
Apoplex and Suddenly	116	French Pox	86	Pluritic	19
Bedric	10	Frighted	23	Posioned	4
Blind	1	Gout and Sciatica	27	Quinse	35
Bleeding	16	Grief	46	Rickets	137
Bloody Flux, Scouring & Flux	185	Griping in the Guts	228	Killing of the Lights	397
Burnt and Scalded	8	Hang'd & made away themselves	7	Leprotic	14
Colicure	3	Headmole shot & Moxie fallen	14	Scurvy	109
Cancer, Gangrene and Fiftula	56	jaundies	120	Shingles and Swine pox	2
Canker, and Thrush	121	Imposiume	227	Sores, Ulcers, broken and healed	338
Childbed	625	Kill'd by severall accidentes	46	Lambs	82
Christomes and Infants	1258	Kings Evill	28	Spleen	14
Cold and Cough	62	Leptotic	2	Spotted Fever and Purples	1929
Collick and Winde	124	Lechary	14	Scopping of the stomack	338
Consumption and Tiflick	4808	Liverg-town	21	Stone and Stranguy	28
Convulsion and Morice	1052	Meagrom and Headach	1	Sucket	1100
Distacted	3	Mealles	7	Teeth and Worms	1014
Droove and Terpany	1476	Mothered and Shot	9	Worming	51
Drwaed	3	Overjaed & Starved	45	Vunn	7
		Males	48569		
		Females	48737	Of the Plague	68598
		In all	97306		

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. 88598

# 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@mskccc.org](mailto:widmerd@mskccc.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\*



# Mouse Genome Database and GO

Entrez Gene 

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

► **General gene information** 

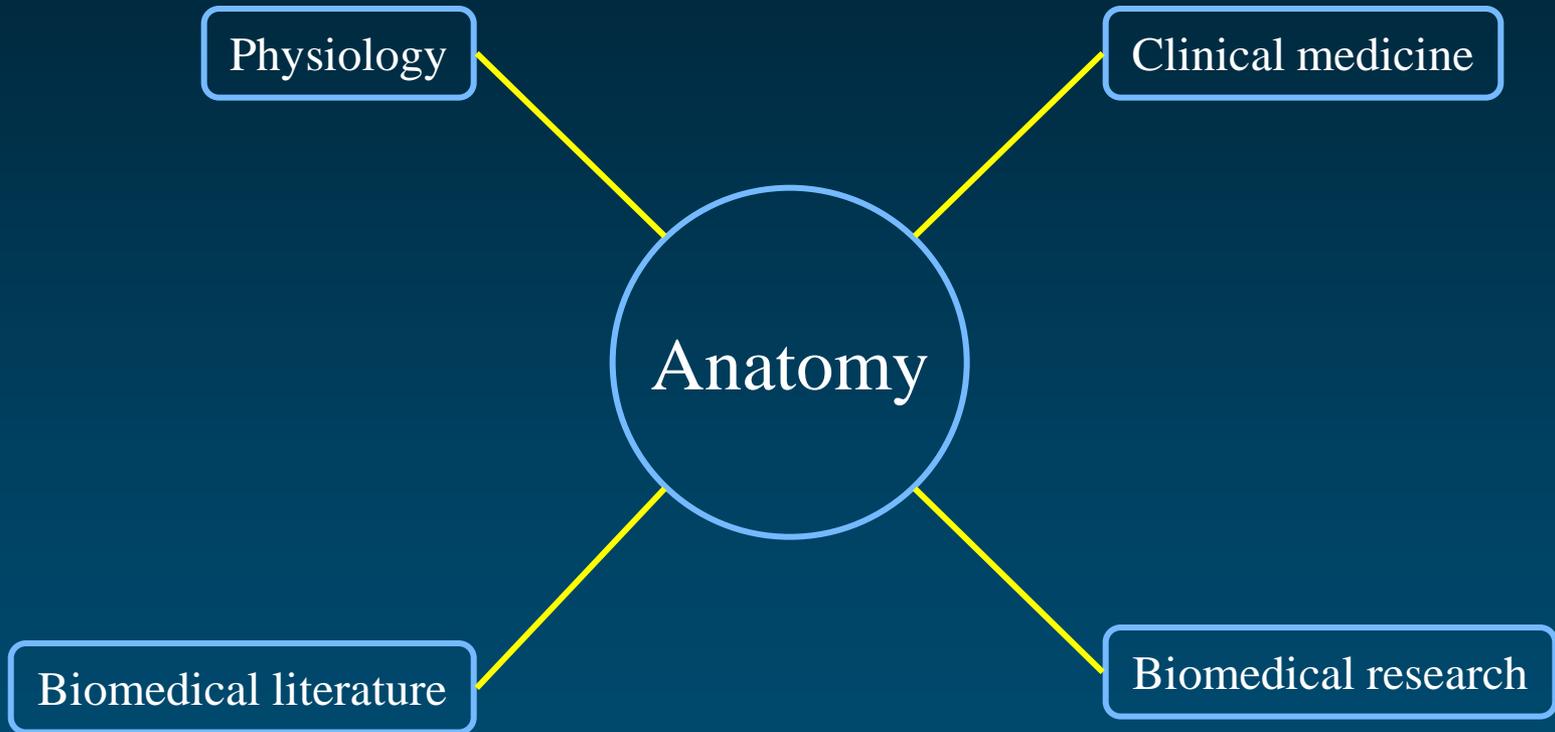
**GeneOntology**  
Provided by [MGI](#)

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

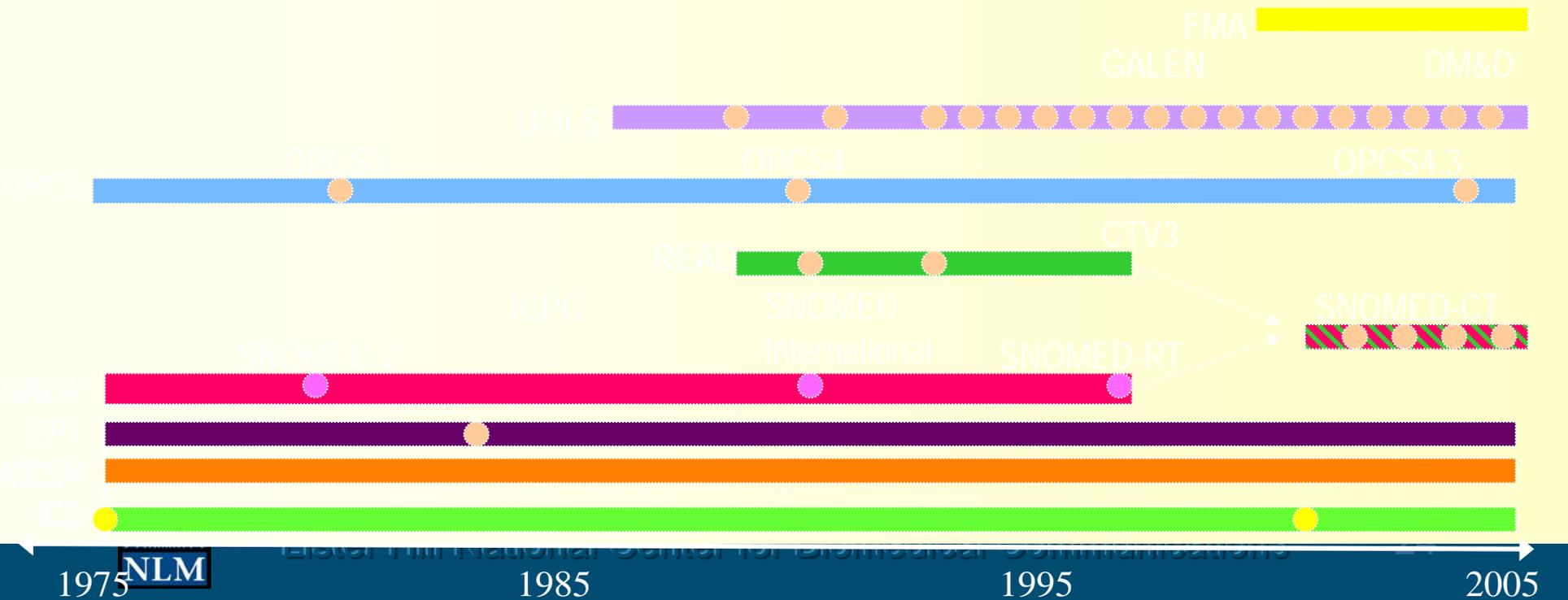
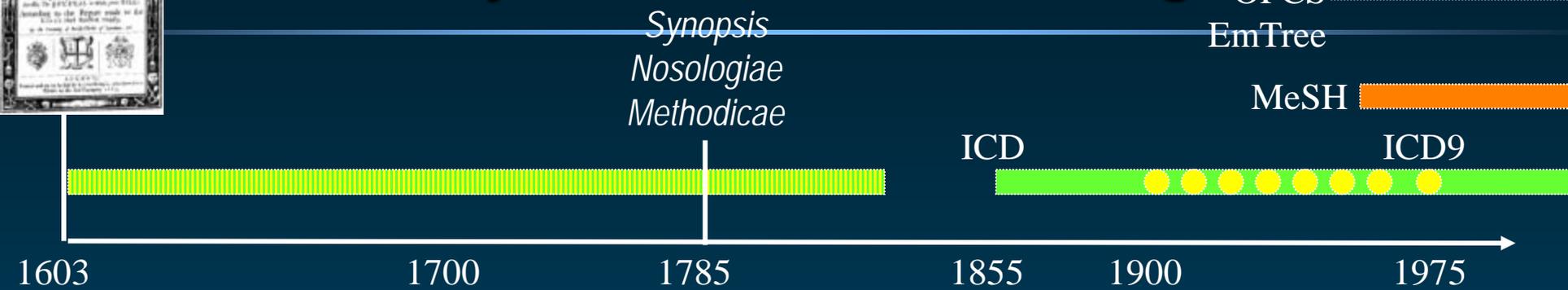
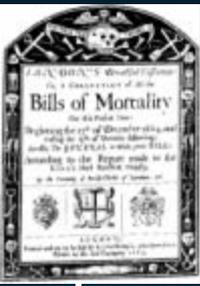
# Anatomy in Biomedicine



# 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

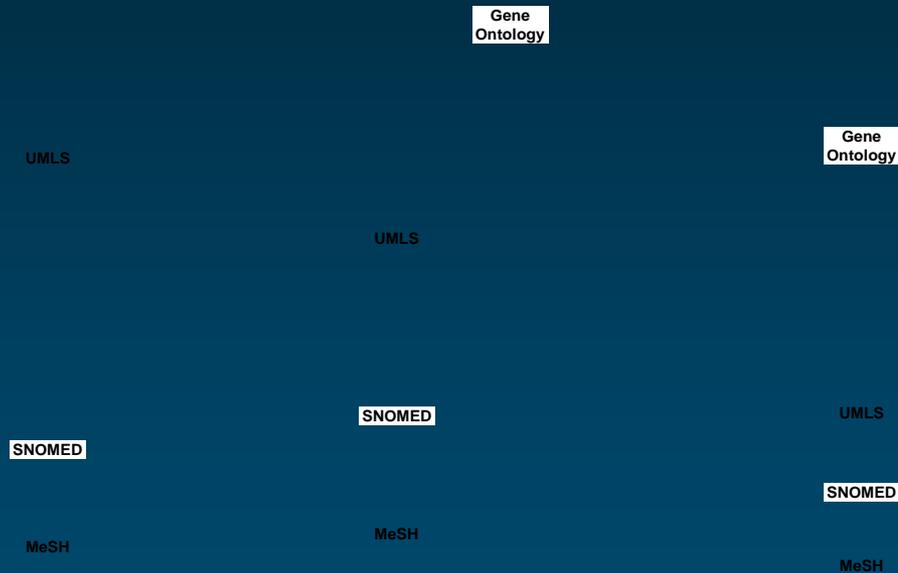


*(\*) As of  
2008/02/20  
(Partial coverage  
for 2007, due to a  
slight lag in the  
indexing process)*

[Bodenreider, YBMI 2008]

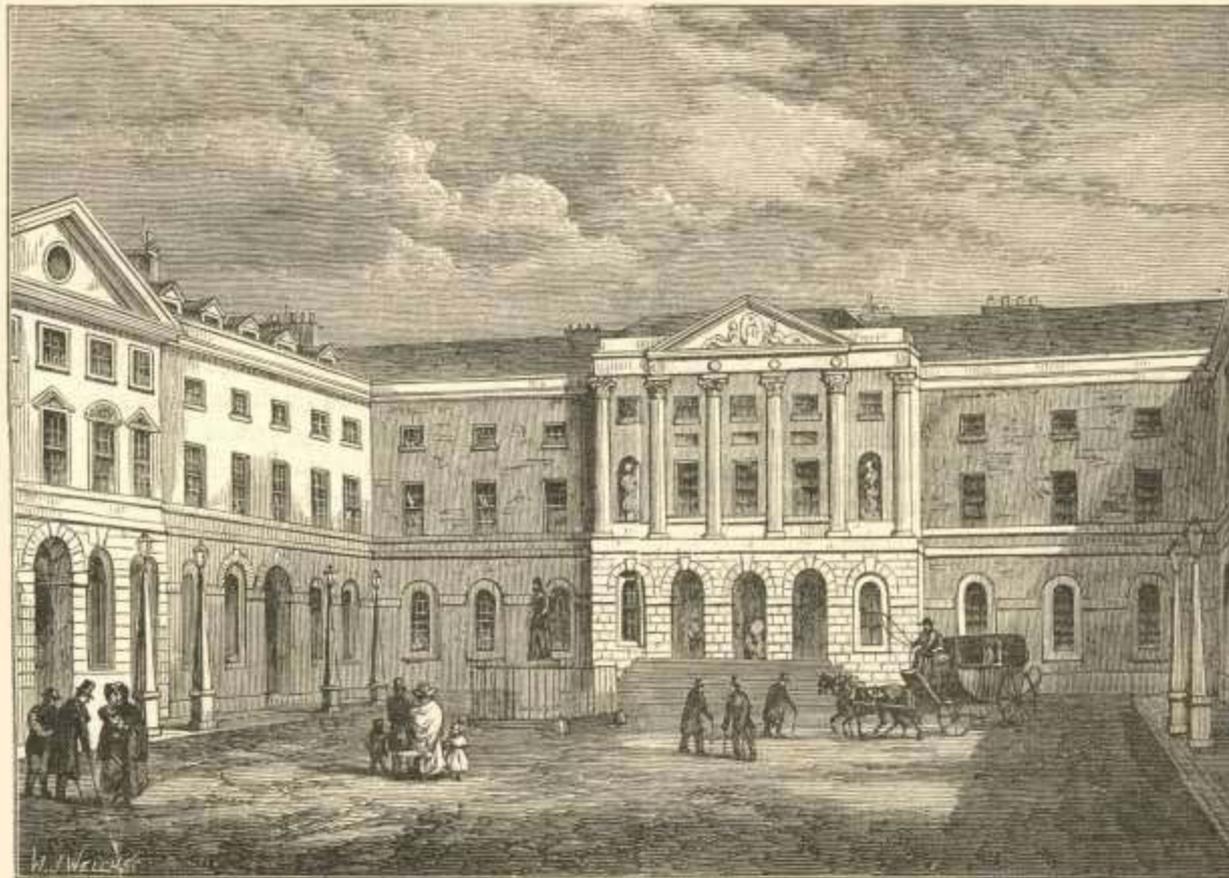


# Biomedical ontologies in PubMed



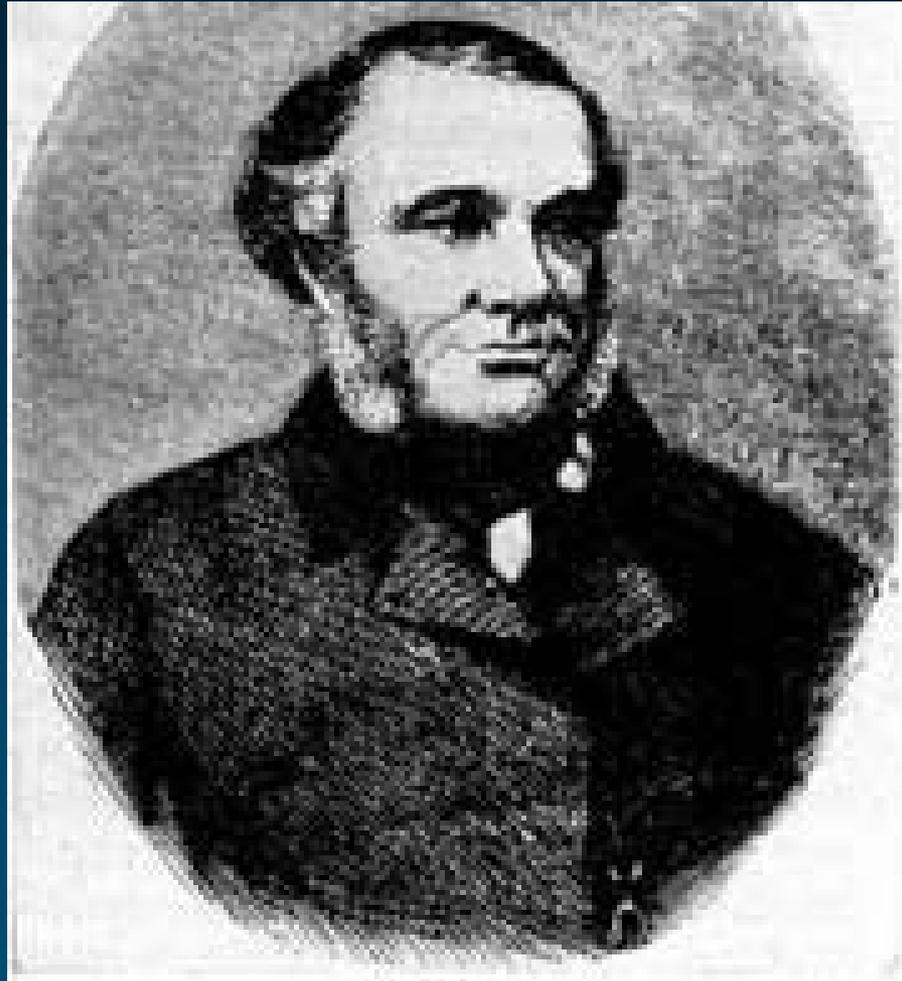
# Introduction to biomedical terminologies through an example

# Guy's Hospital, London



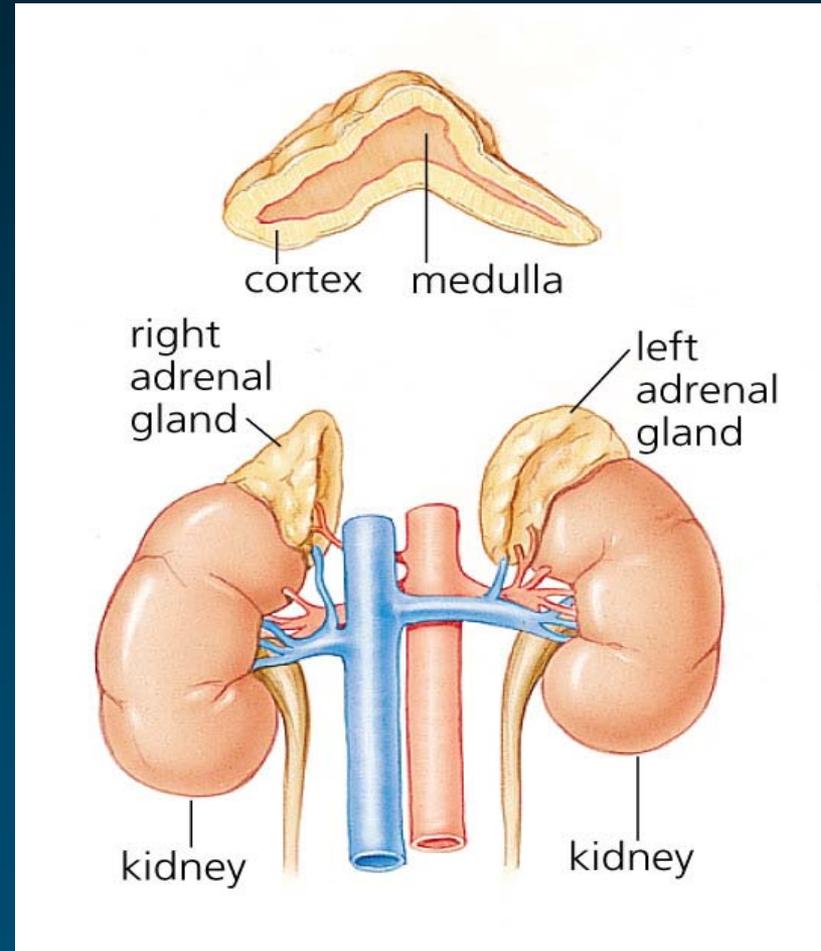
GUY'S HOSPITAL.

# Thomas Addison (1795-1860)



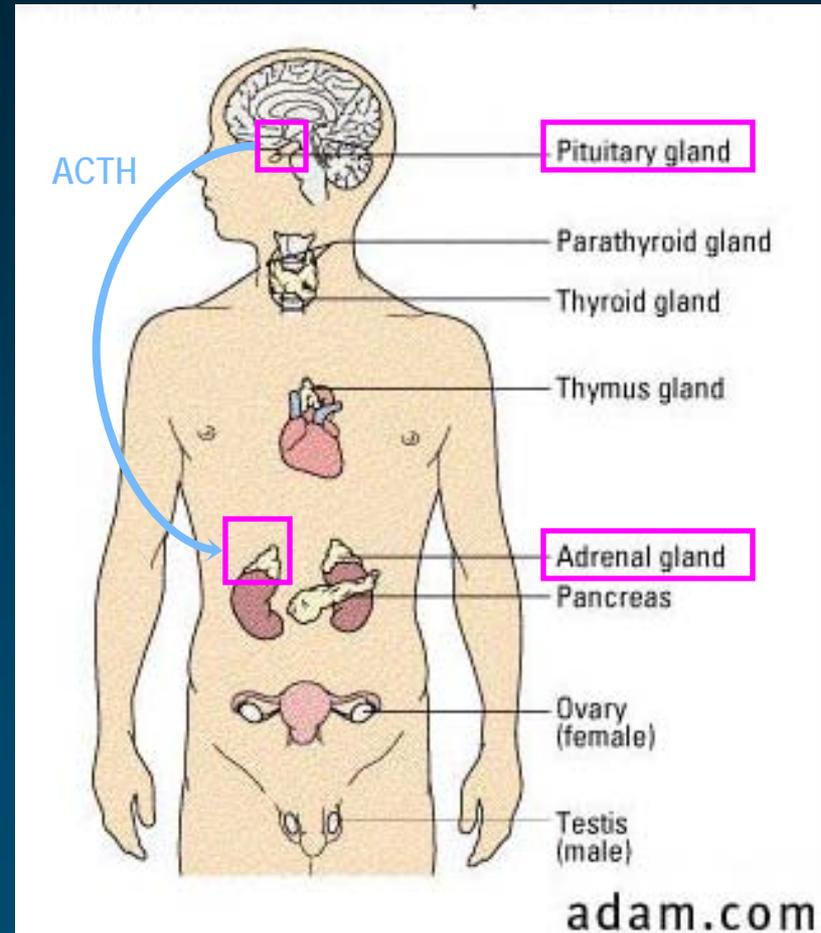
# Addison's disease

- ◆ 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

Excludes1: Cushing's syndrome (E24.-)

###### **E27.1 Primary adrenocortical insufficiency**

###### **Addison's disease**

Adrenocortical insufficiency NOS

Autoimmune adrenalitis

Excludes1: 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 \[C19.053.264.270\]](#)

[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 \[C20.111.199\] +](#)

[Autoimmune Diseases of the Nervous System \[C20.111.258\] +](#)



# SNOMED CT



Hierarchy Subtype hierarchy

- 386584007 adrenal cortical hypofunction
  - 383732003 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
  - F Addison's disease (disorder)
  - P Addison's disease
  - U enfermedad de Addison
  - U enfermedad de Addison (trastorno)
- Definition: Primitive
  - is a
    - D adrenal cortical hypofunction
  - finding site
    - D adrenal cortex structure
- Qualifiers
  - severity
    - p severities
  - episodicity
    - p episodicities
  - clinical course
    - p 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

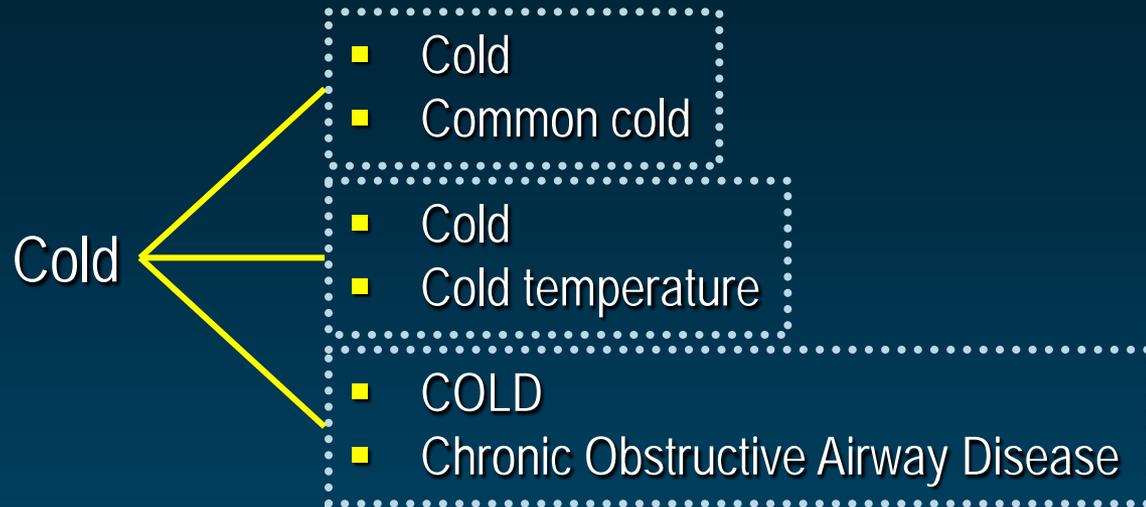
- Posttransfusion hepatitis
- Posttransfusion viral hepatitis

- Addison's disease
- Primary adrenocortical insufficiency

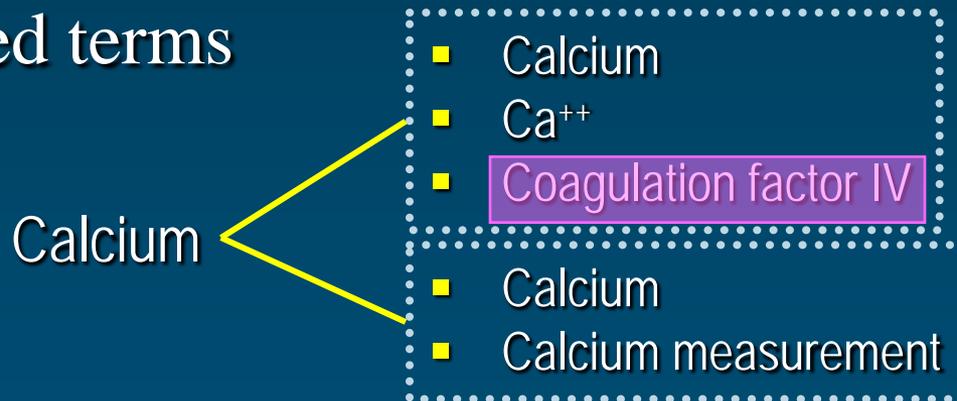
vs. **Waterhouse-Friderichsen Syndrome**

# 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

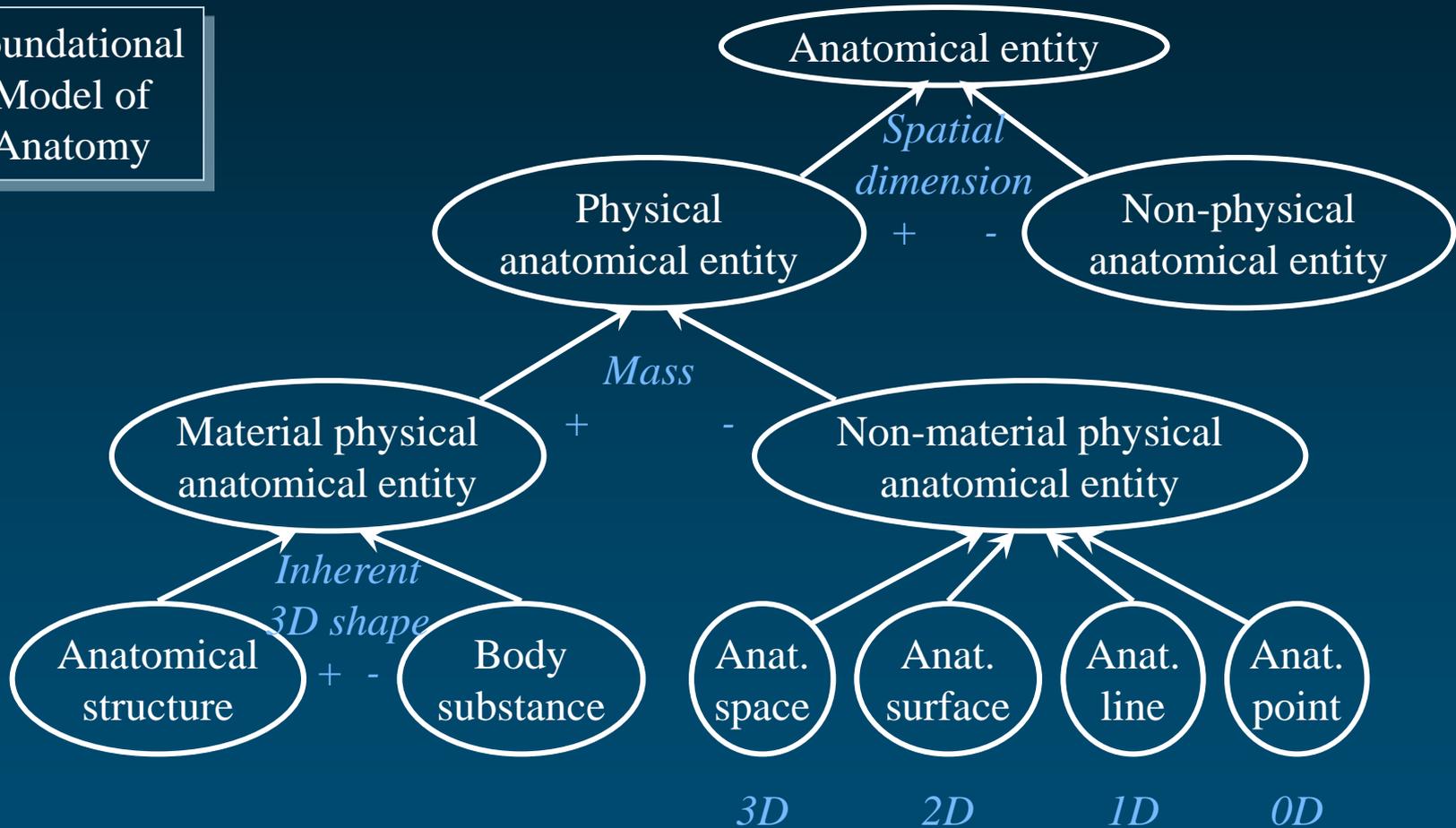
Terminological relations as a  
surrogate for ontological relations

# 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



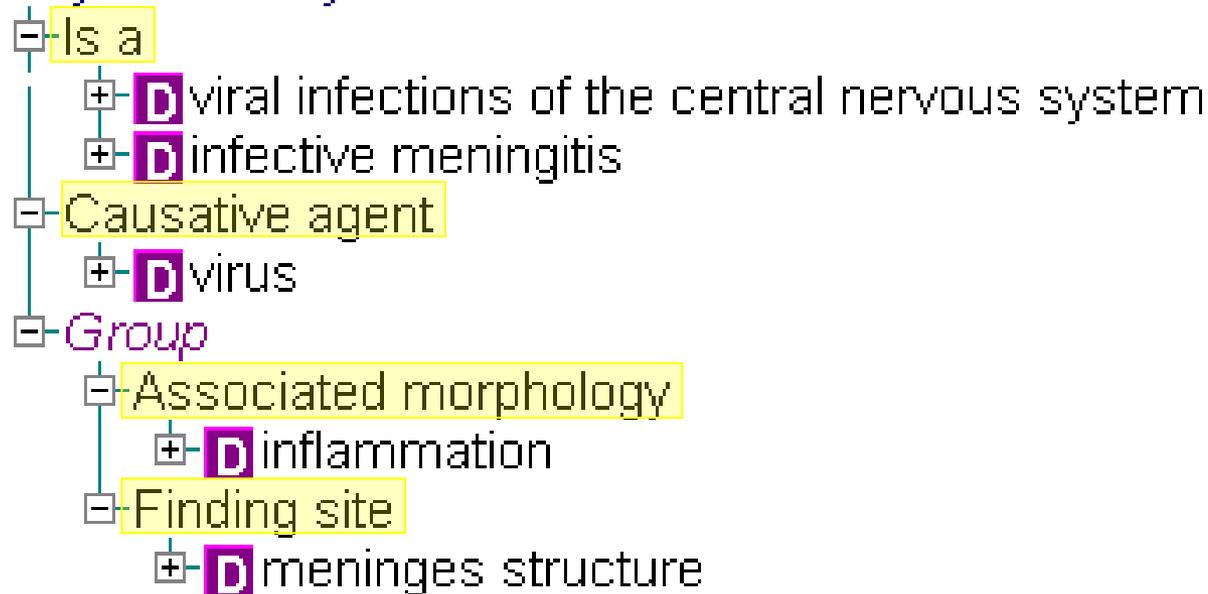
- Attribute
- Body structure
- Clinical finding
- Context-dependent categories
- Environments and geographical locations
- Events
- Observable entity
- Organism
- Pharmaceutical / biologic product
- Physical force
- Physical object
- Procedure
- Qualifier value
- Social context
- Special concept
- Specimen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



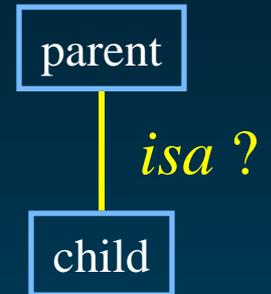
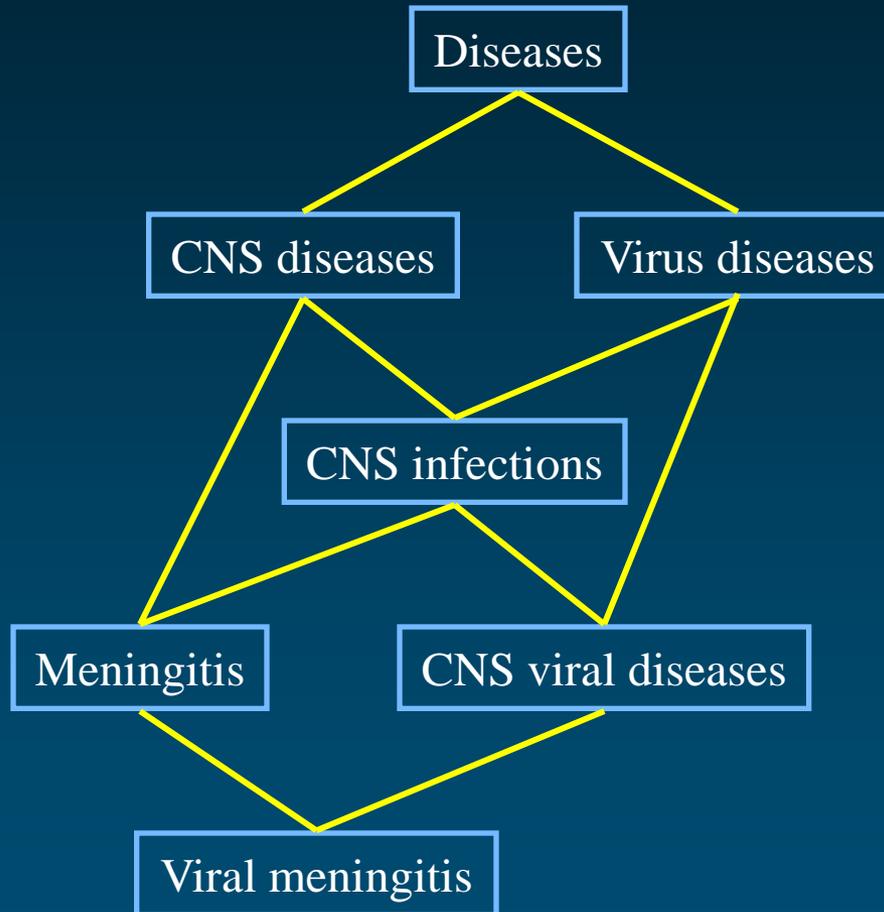
# Fully specified relations

## Viral meningitis in SNOMED CT

*Fully defined by ...*



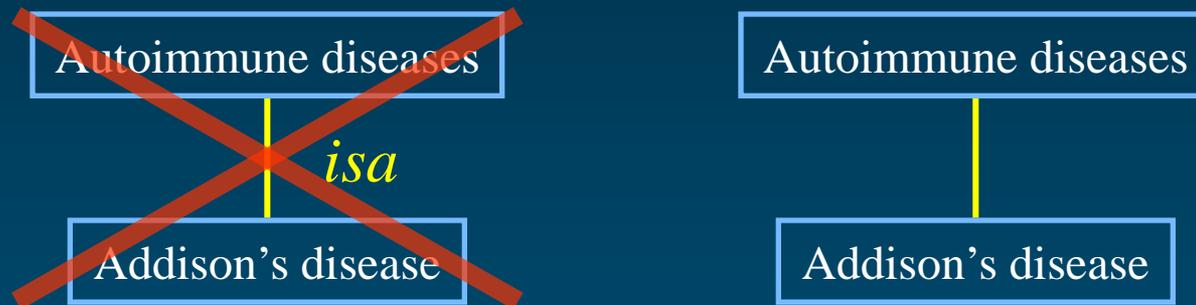
# Underspecification of the relations



# 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 adrenoleukodystrophy 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 (R96.5)

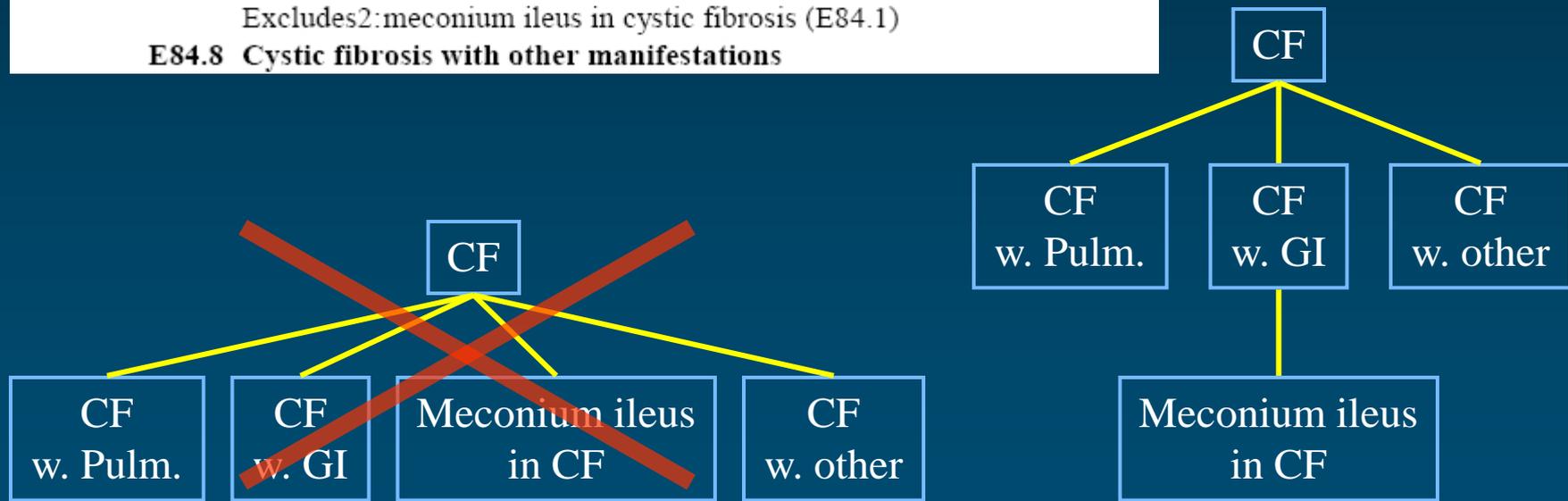
### E84.1 Meconium ileus in cystic fibrosis

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



# Short course – Summer 2010 Clinical Ontology in Practice

June 15, 2010 – Session #2

## Design Principles, Formalisms and Tools for Biomedical Ontologies



*Olivier Bodenreider*

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

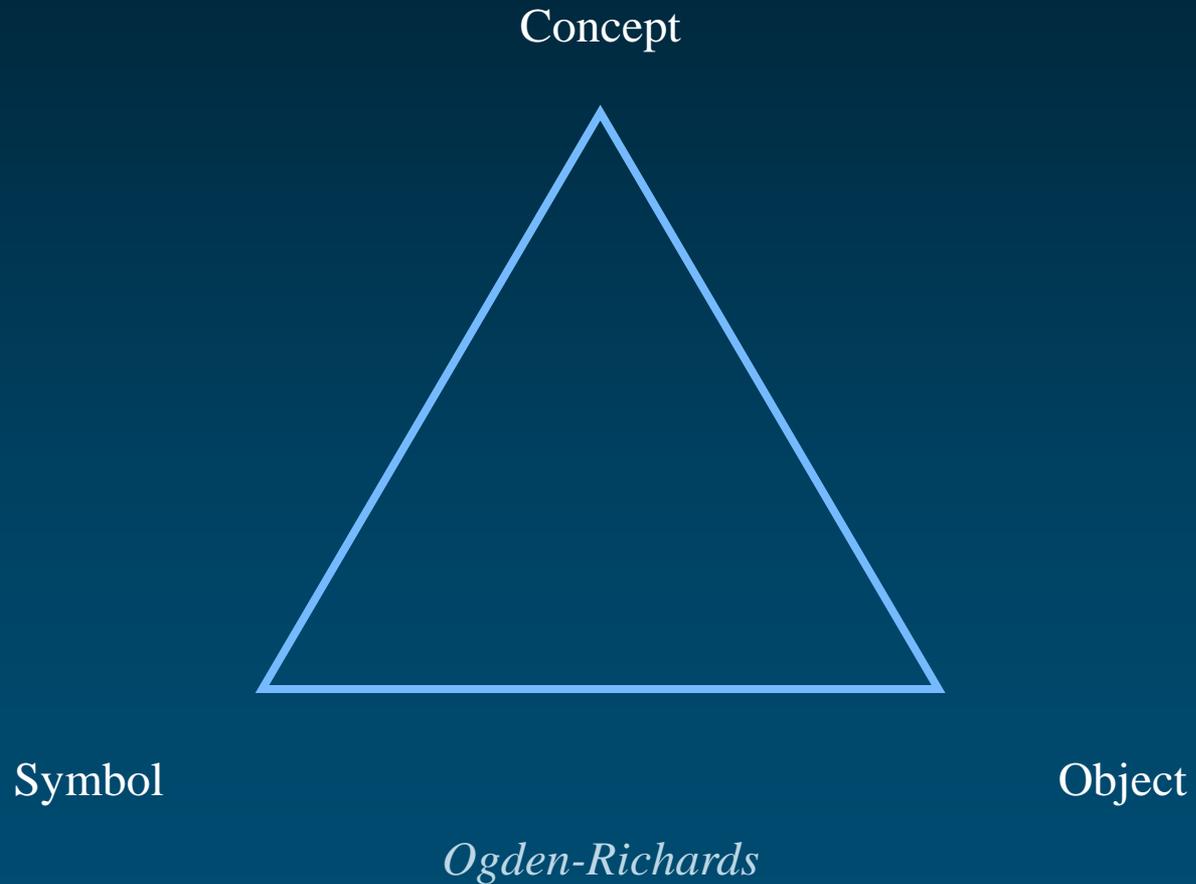
# Overview

- ◆ 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



# 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

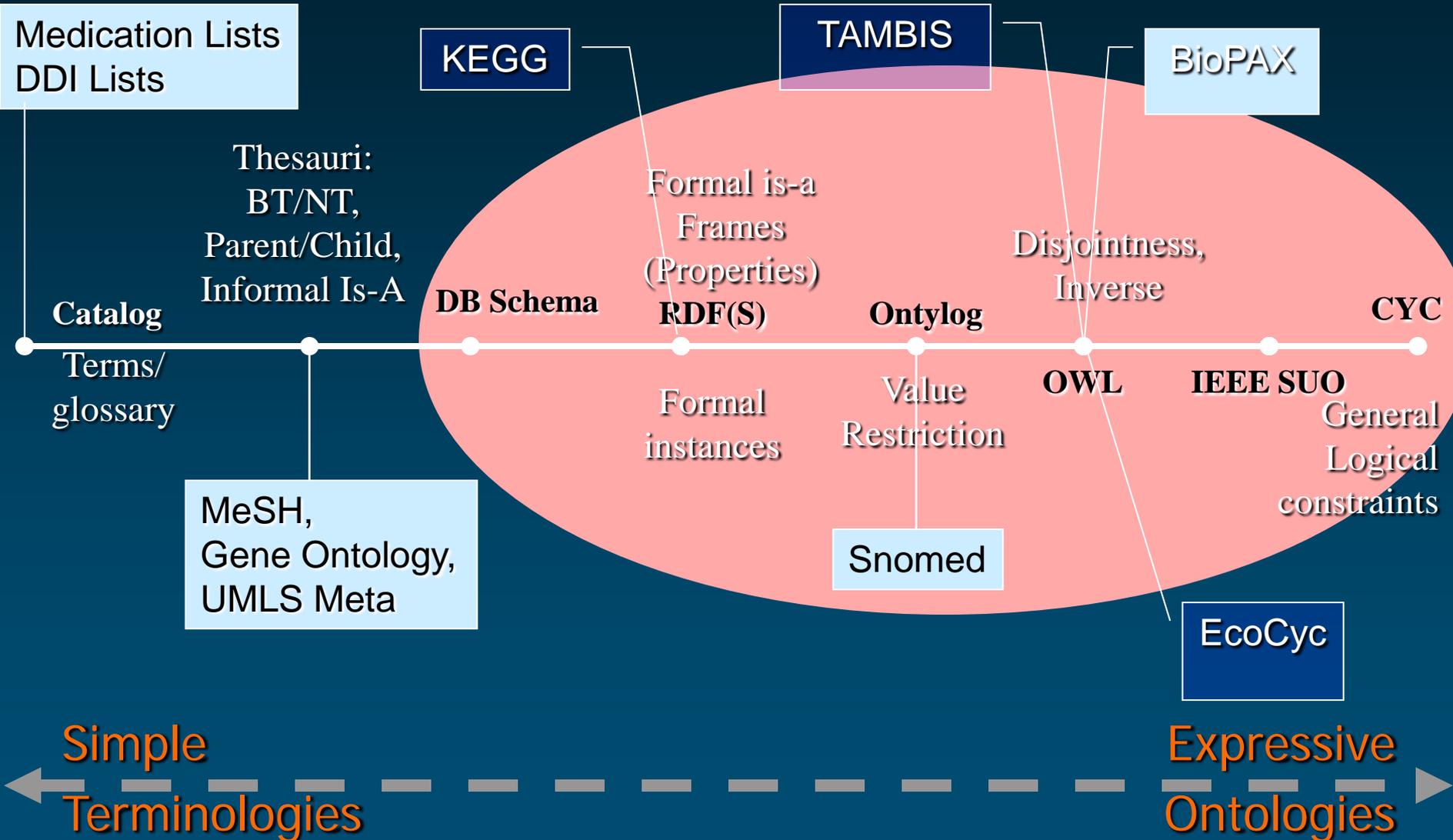
- Collections of
  - kinds of entities (substances, qualities, processes)
  - relations among them
- Useful for **relation extraction**
- UMLS Semantic Network, BioTop



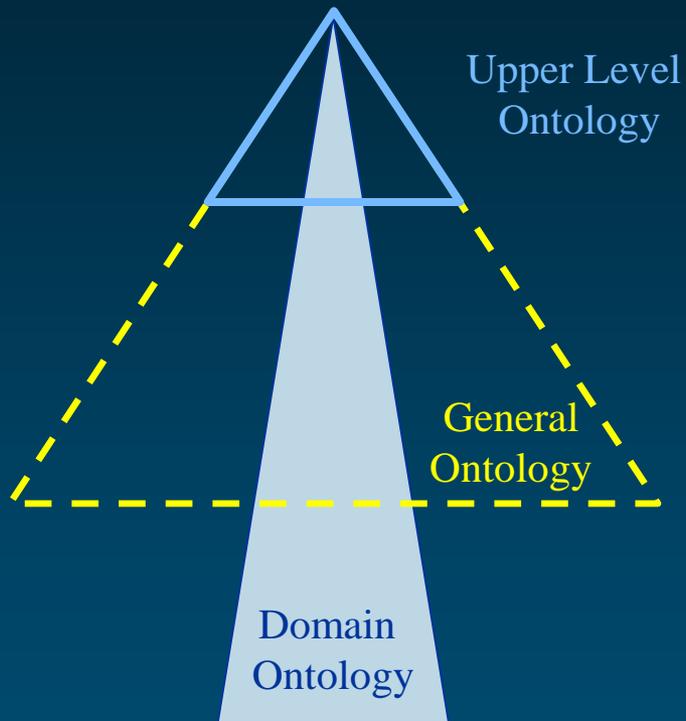
## ◆ Terminological resources

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

# The Knowledge Semantics Continuum



# 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*

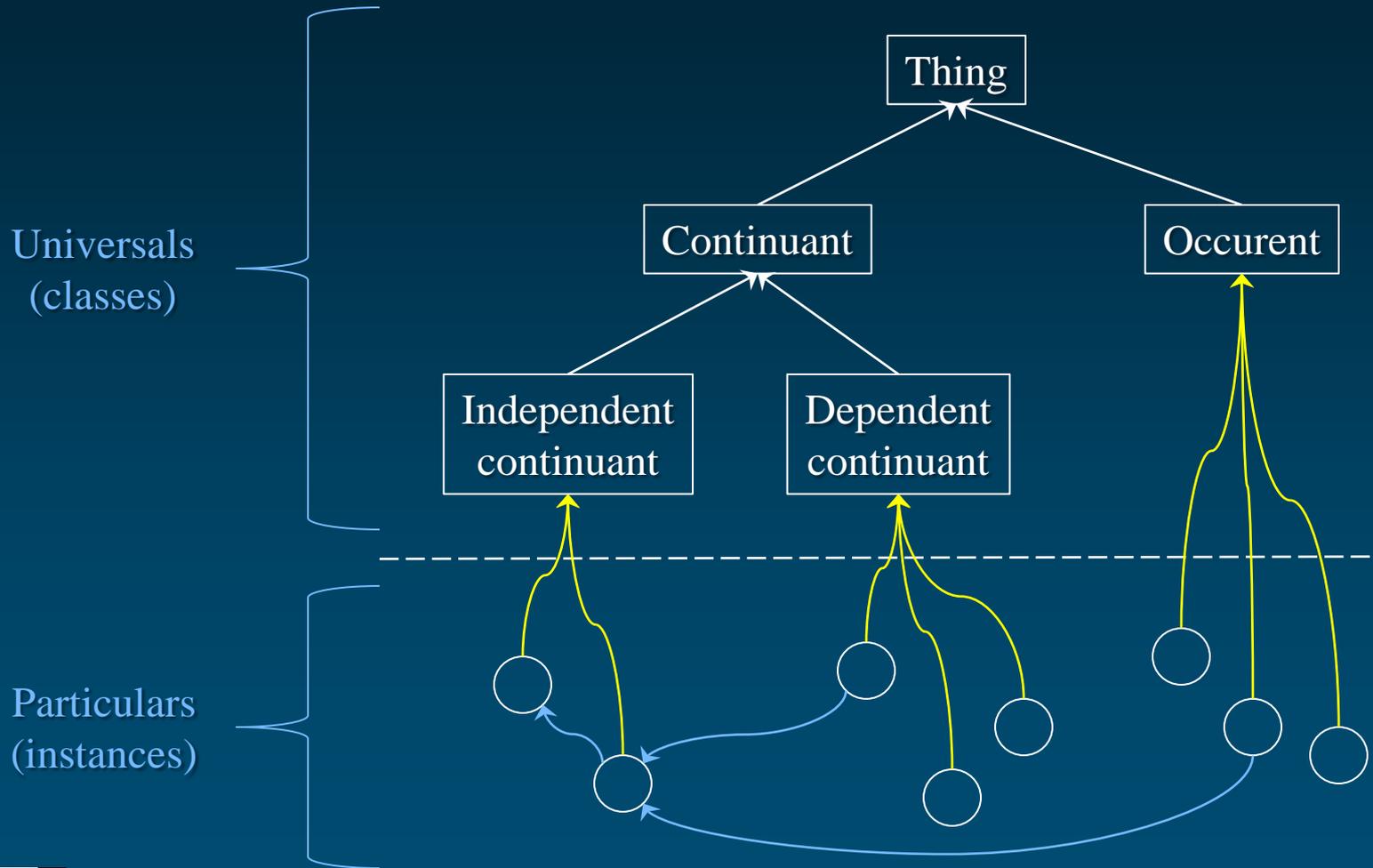
# Continuant vs. Occurrent

- ◆ 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>

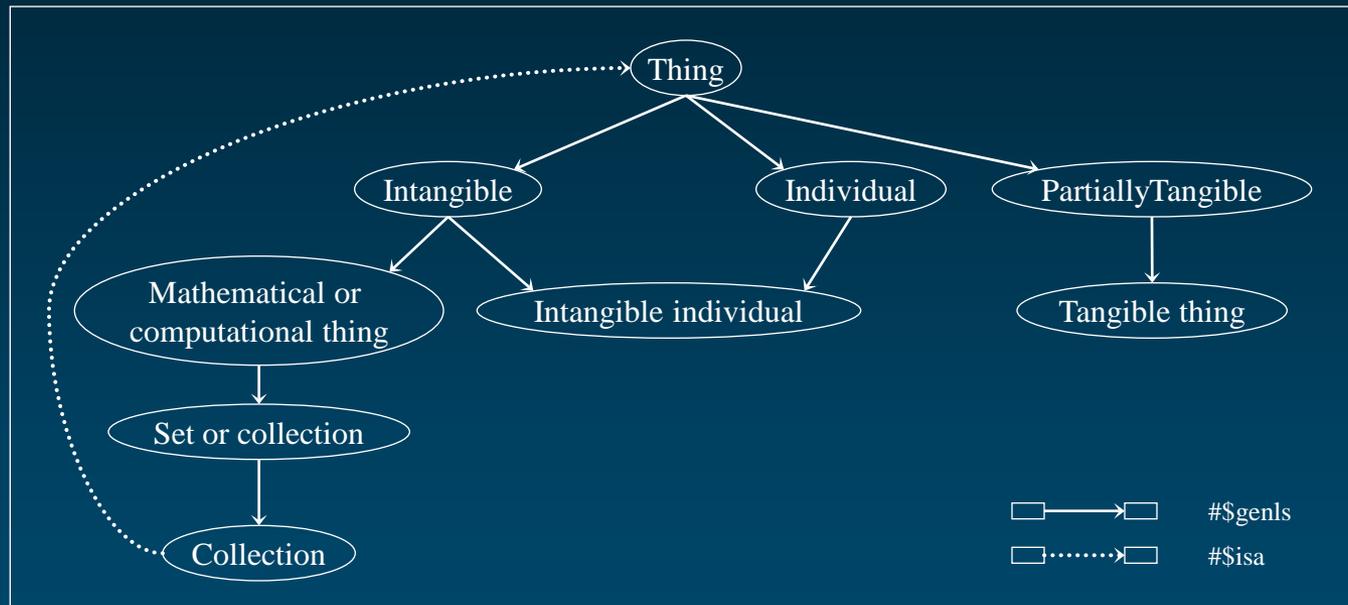


## ◆ WordNet

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



# Top level in OpenCyc



# Top level in WordNet

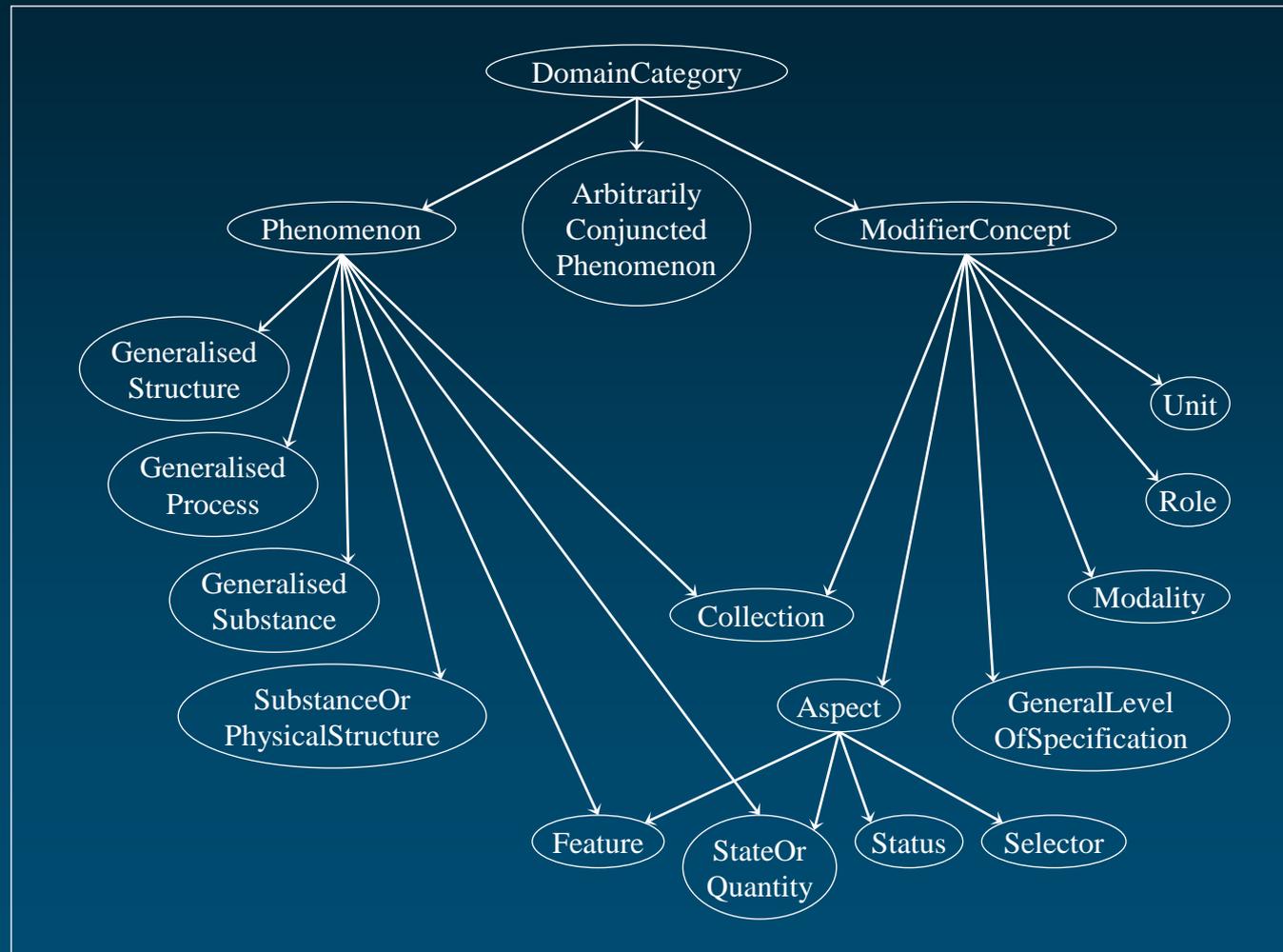
Abstraction  
Activity  
Entity  
Event  
Group  
Location  
Natural phenomenon  
Possession  
Psychological feature  
Shape  
State

# 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



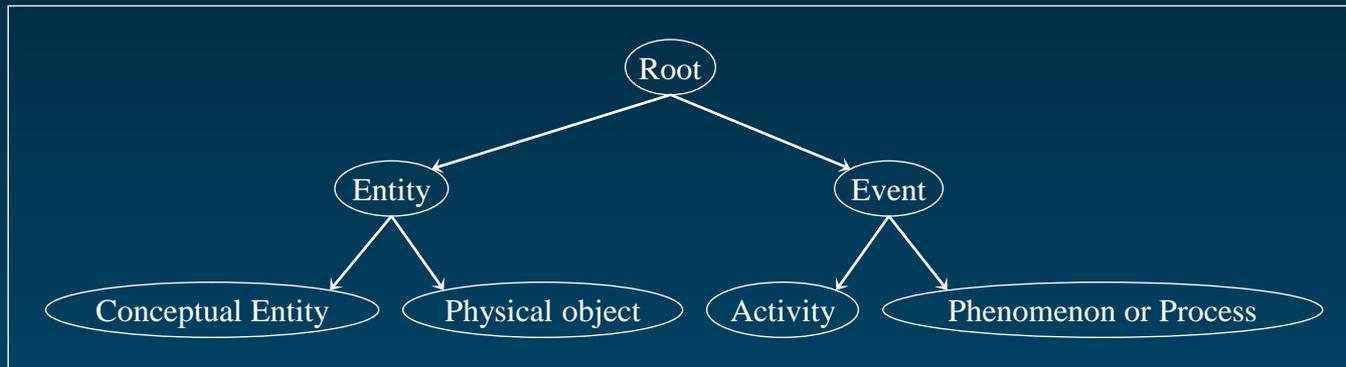
# UMLS Semantic Network



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



# Top level in the Semantic Network



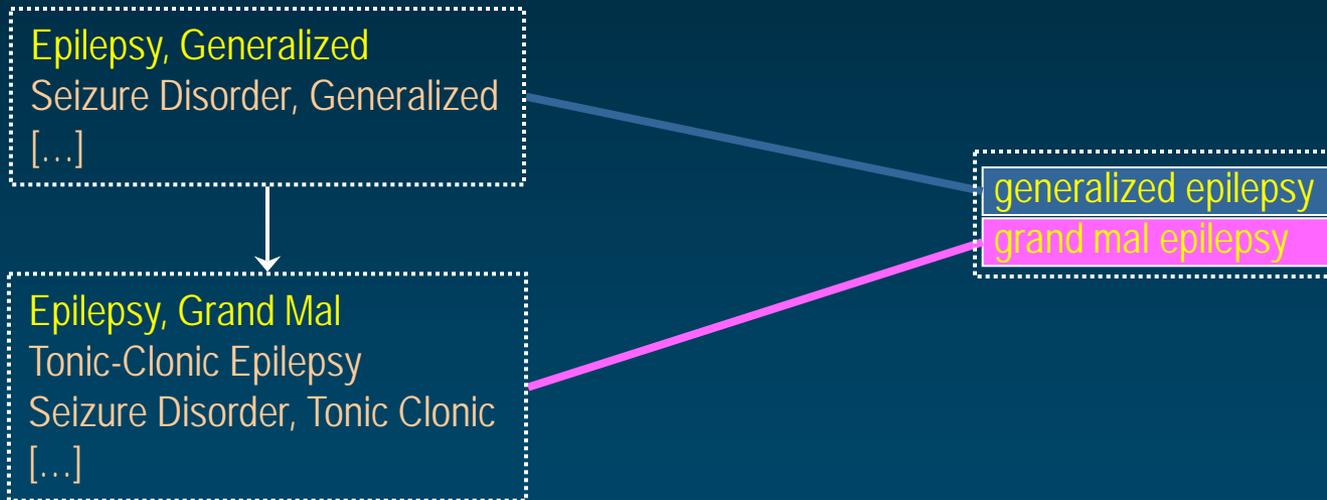
# Differences between ontologies

*Examples*

# Granularity, plesionymy

UMLS

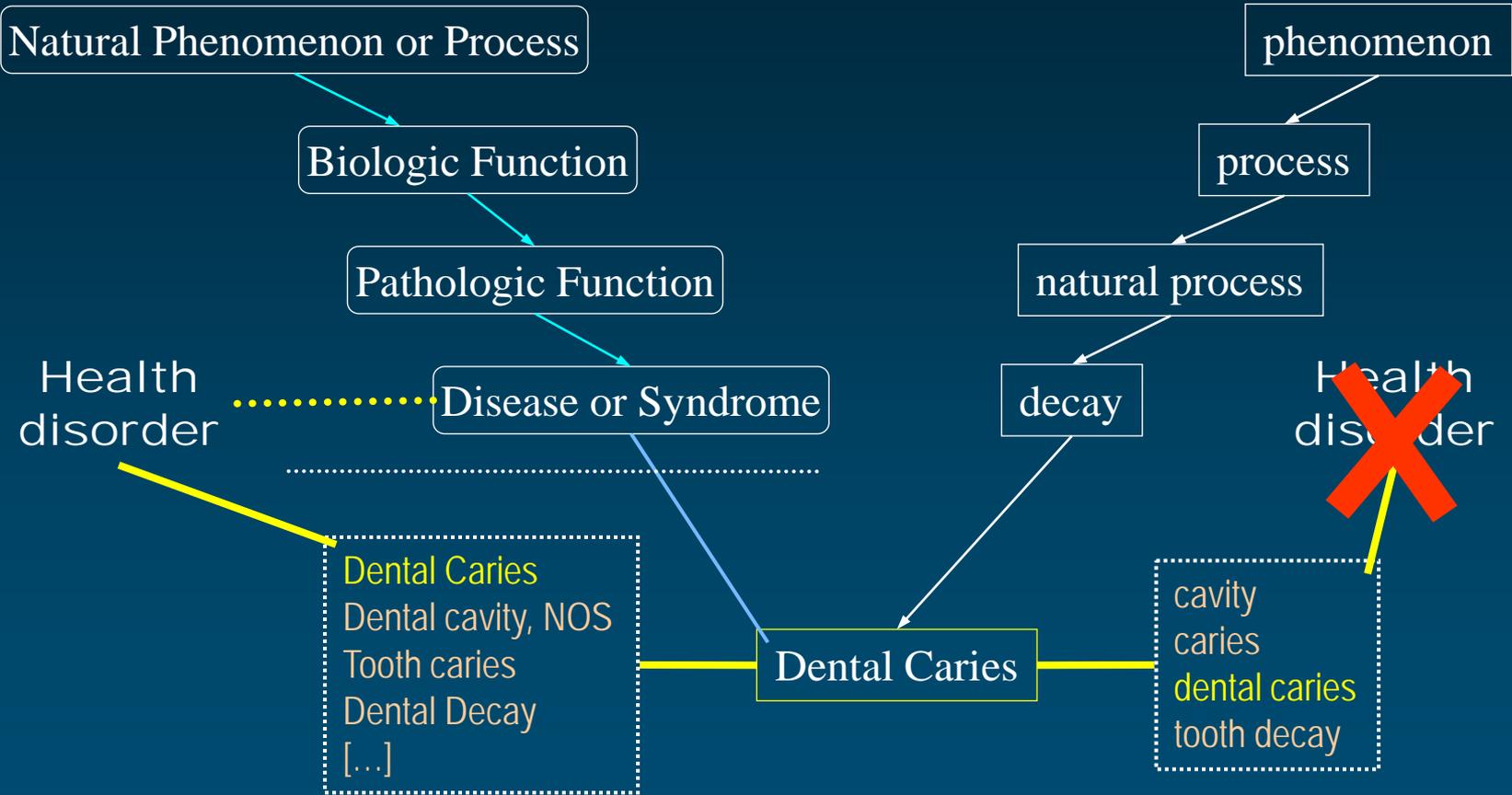
WordNet



# Differing categorization

UMLS

WordNet



# 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/>





# Short course – Summer 2010 Clinical Ontology in Practice

June 15, 2010 – 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]

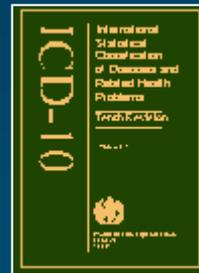
- What are they (vs. what are they for)?

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

*Logical Observation Identifiers Names and Codes*

# LOINC Characteristics (1)

- ◆ Current version: 2.30 (Feb. 2010)
- ◆ 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](http://www.regenstrief.org/loinc/loinc.htm)



# LOINC Characteristics (2)

- ◆ Number of
  - Concepts: 50k active codes (2.18)  
(2 annual releases)
  - 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, 2010 (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: ~310,000 active concepts (Jan. 31, 2010)
  - Terms: ~800,000 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
8025007	laparoscopic appendectomy
235313004	non-emergency appendectomy
235314005	inversion appendectomy
1299000	excision of appendiceal stump

Definition: Fully defined by ...

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

appendectomy - Definition

Concept Status: **Current**

Descriptions

- appendectomy (procedure)
- appendectomy
- excision of appendix
- 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/>

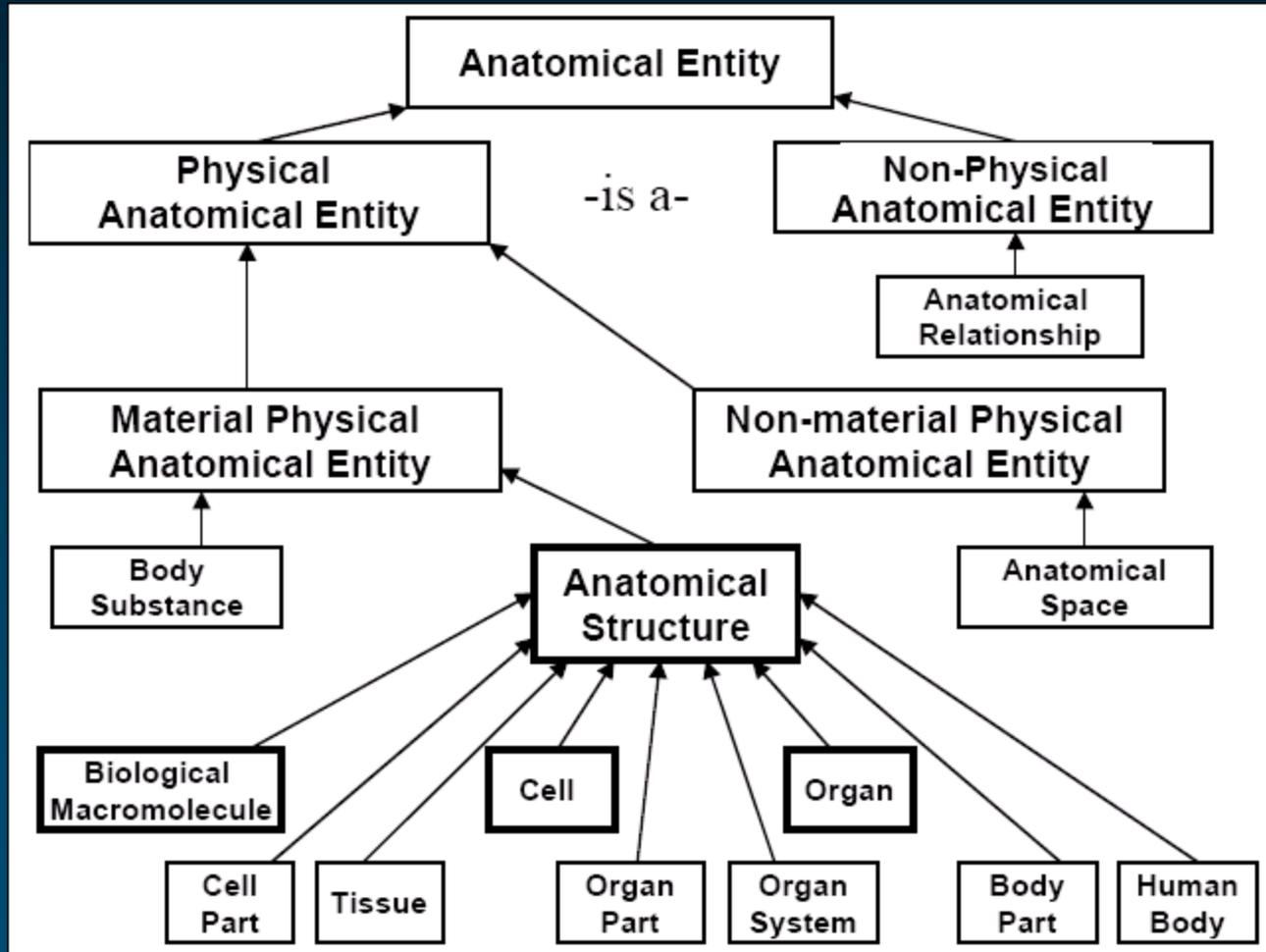


# FMA Characteristics (2)

- ◆ 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

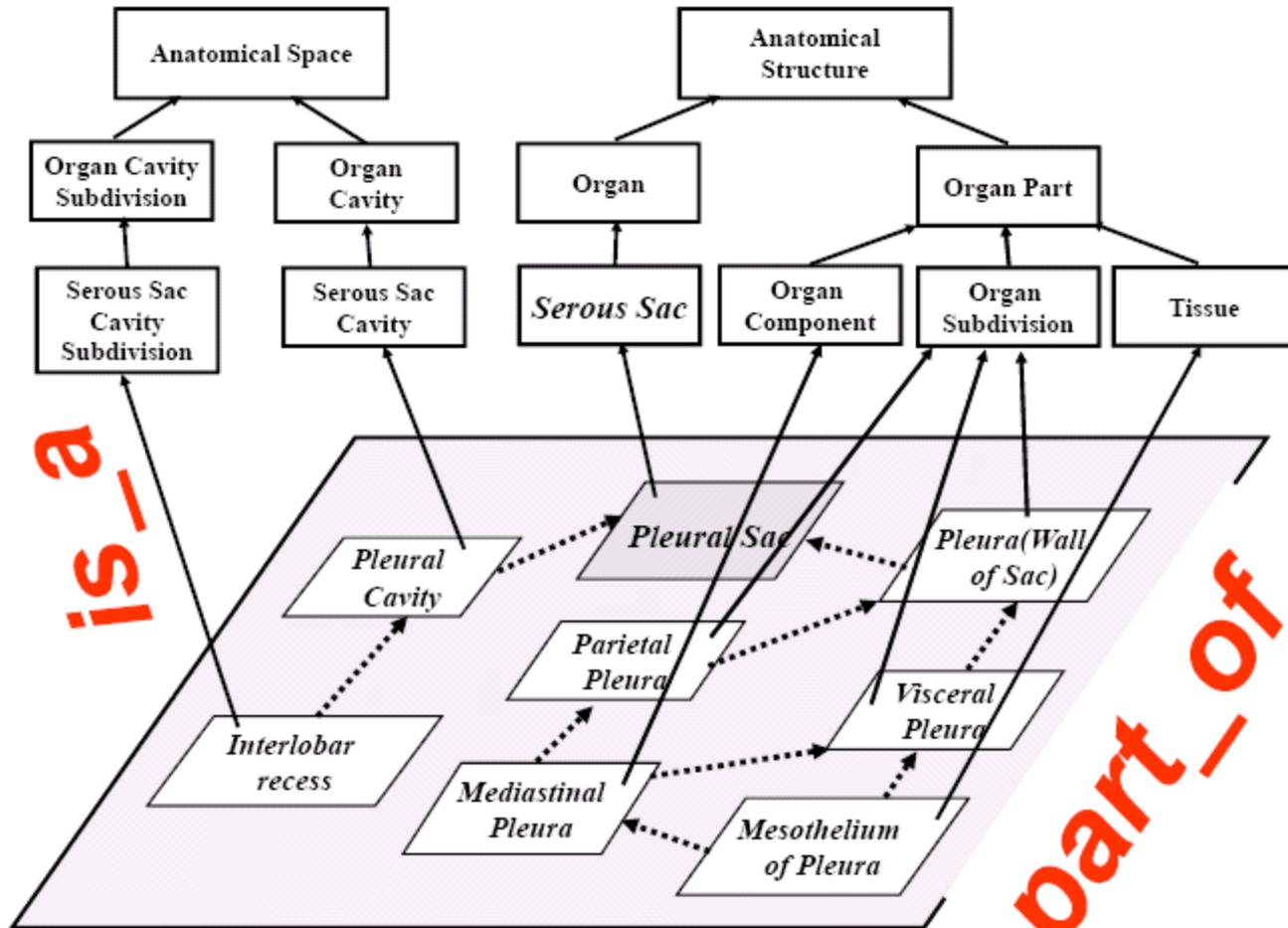
# FMA Top level

(Courtesy of C. Rosse)



# FMA Example

(Courtesy of C. Rosse)



is\_a

part\_of



# 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: 27,800 (July 22, 2009)
  - 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] [E](#)
- ▣ **I** GO:0008150 : biological\_process [166605 gene products]
- ▣ **I** GO:0005575 : cellular\_component [169814 gene products]
- ▣ **I** **GO:0003674 : molecular\_function [168558 gene products]** [E](#)
  - ▣ **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)

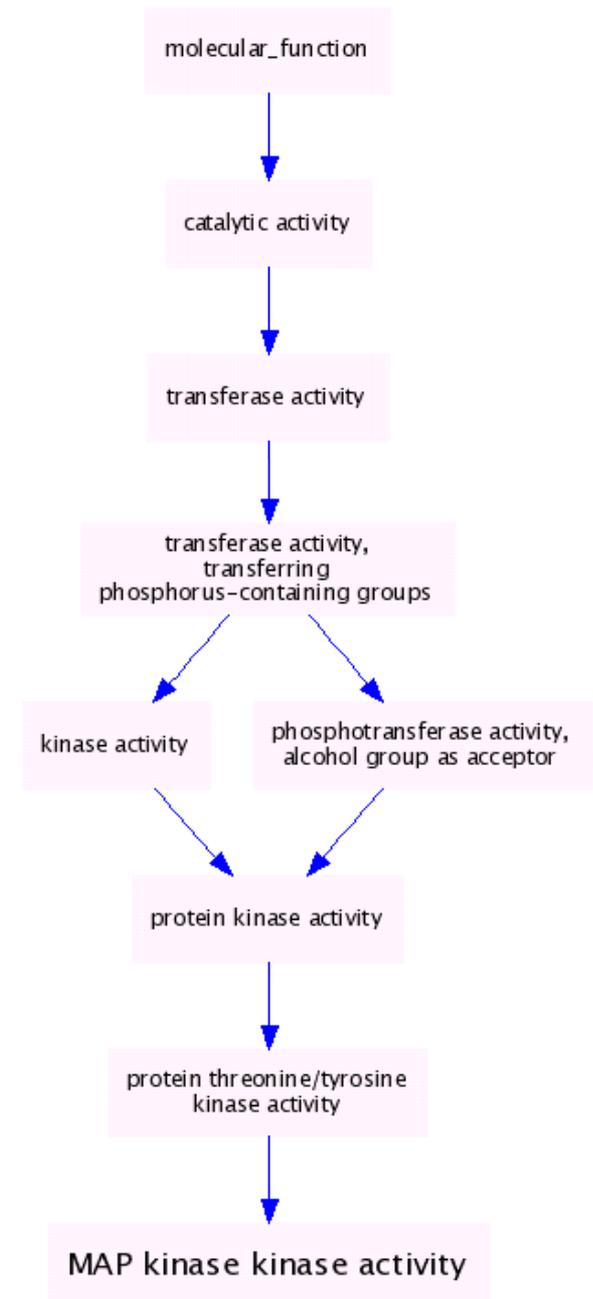
- ▣ all : all [250418 gene products] [E](#)
- ▣ **I** GO:0008150 : biological\_process [166605 gene products]
- ▣ **I** **GO:0005575 : cellular\_component [169814 gene products]** [E](#)
  - ▣ **I** GO:0005623 : cell [111086 gene products]
  - ▣ **I** GO:0044464 : cell part [111049 gene products]
  - ▣ **I** GO:0031975 : envelope [3316 gene products]
  - ▣ **I** GO:0031012 : extracellular matrix [573 gene products]
  - ▣ **I** GO:0044420 : extracellular matrix part [292 gene products]
  - ▣ **I** GO:0005576 : extracellular region [5001 gene products]
  - ▣ **I** GO:0044421 : extracellular region part [3365 gene products]
  - ▣ **I** GO:0032991 : macromolecular complex [14668 gene products]
  - ▣ **I** GO:0031974 : membrane-enclosed lumen [5290 gene products]
  - ▣ **I** GO:0043226 : organelle [79653 gene products]
  - ▣ **I** GO:0044422 : organelle part [16645 gene products]
  - ▣ **I** GO:0055044 : symplast [3 gene products]
  - ▣ **I** GO:0045202 : synapse [454 gene products]
  - ▣ **I** GO:0044456 : synapse part [210 gene products]
  - ▣ **I** GO:0019012 : virion [227 gene products]
  - ▣ **I** GO:0044423 : virion part [186 gene products]
- ▣ **I** GO:0003674 : molecular\_function [168558 gene products]

# Gene Ontology Top level (BP)

- ▣ all : all [250418 gene products] [E](#)
- ▣ **I** **GO:0008150 : biological\_process** [166605 gene products] [E](#)
  - ▣ **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 phosphorus-containing groups
- ▣ **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 7, 2010 (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: 166k
  - Terms: ~1 term per concept
- ◆ 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) ←

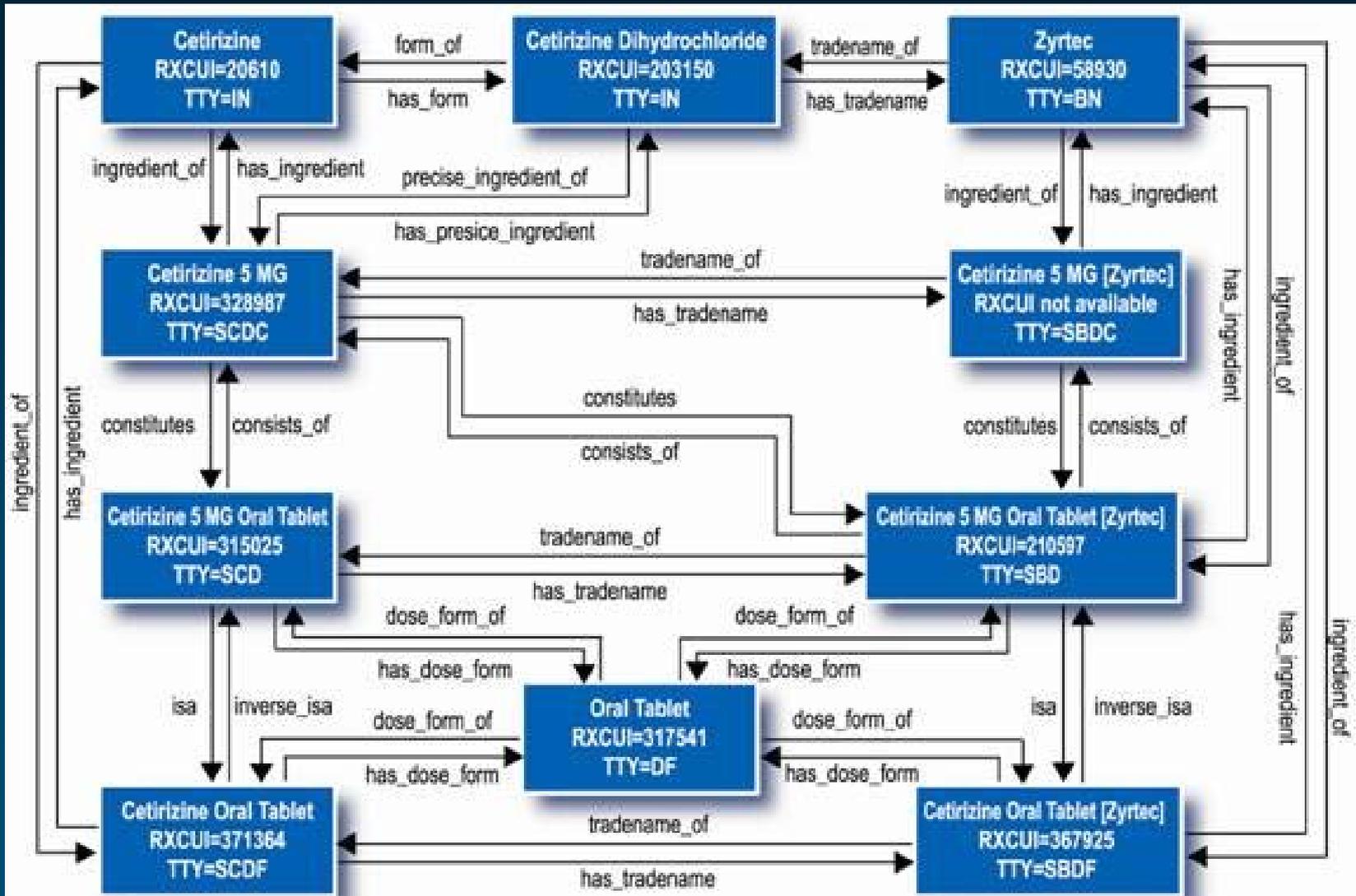
## ◆ Brand

- 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: 2010 (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: 25,588 descriptors (2010)
  - 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

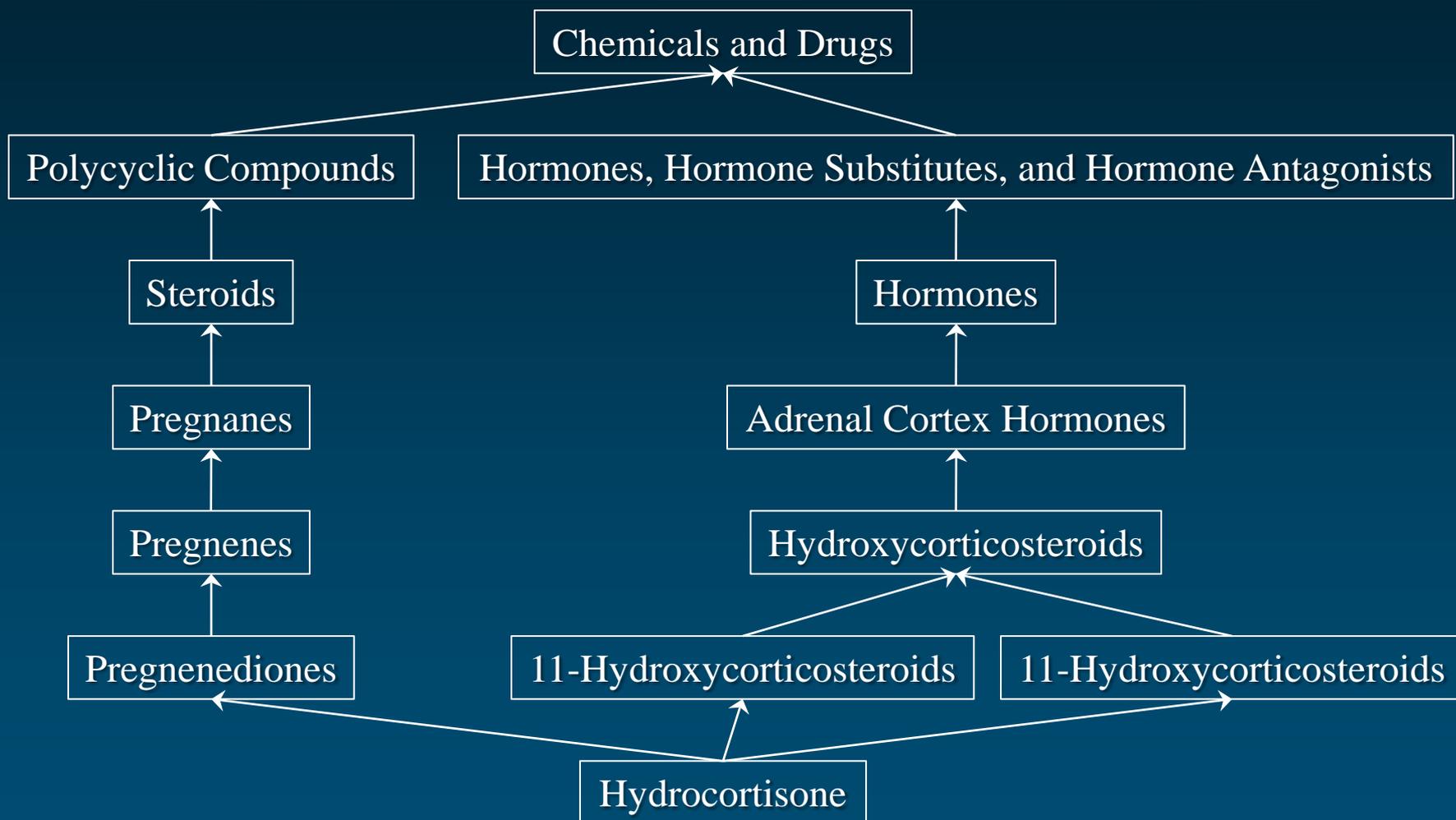
# MeSH Top level

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)

<b>MeSH Heading</b>	Hydrocortisone
<b>Tree Number</b>	<a href="#">D04.808.745.745.654.600</a>
<b>Tree Number</b>	<a href="#">D06.472.040.585.353.476</a>
<b>Tree Number</b>	<a href="#">D06.472.040.585.478.392</a>
<b>Scope Note</b>	The main glucocorticoid secreted by the <a href="#">ADRENAL CORTEX</a> . 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.
<b>Entry Term</b>	11-Epicortisol
<b>Entry Term</b>	Cortifair
<b>Entry Term</b>	Cortisol
<b>Entry Term</b>	Cortril
<b>Entry Term</b>	Epicortisol
<b>Entry Term</b>	Hydrocortisone, (11 alpha)-Isomer
<b>Entry Term</b>	Hydrocortisone, (9 beta,10 alpha,11 alpha)-Isomer

# MeSH Example (hierarchies)



# NCI Thesaurus



# NCI thesaurus Characteristics (1)

- ◆ Current version: 10.05d (~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://nciterns.nci.nih.gov/>



# NCI thesaurus Characteristics (2)

- ◆ Number of
  - Concepts: ~60,000
  - 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://ncit.nsl.ncl.nih.gov:80/NCIBrowser/ConceptReport.jsp?dictionary=NCI\\_Thesaurus&code=C2919](http://ncit.nsl.ncl.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)

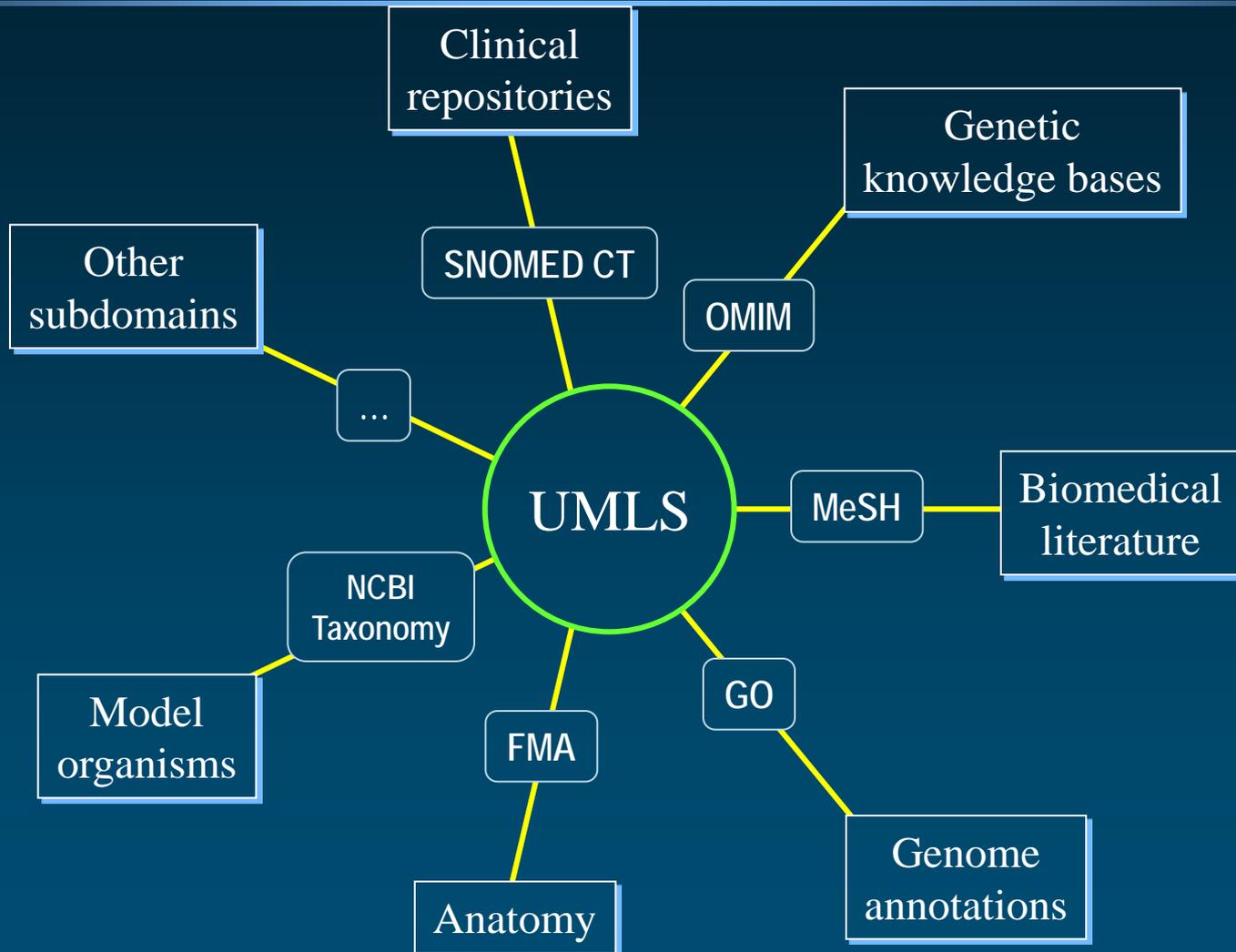
- ◆ Current version: 2010AA (2 annual releases)
- ◆ Type: Terminology integration system
- ◆ Domain: Biomedicine
- ◆ Developer: NLM
- ◆ Funding: NLM (intramural)
- ◆ Availability
  - Publicly available: Yes\* (cost-free license required)
  - Repositories: UMLS
- ◆ URL: <http://umlsks.nlm.nih.gov/>



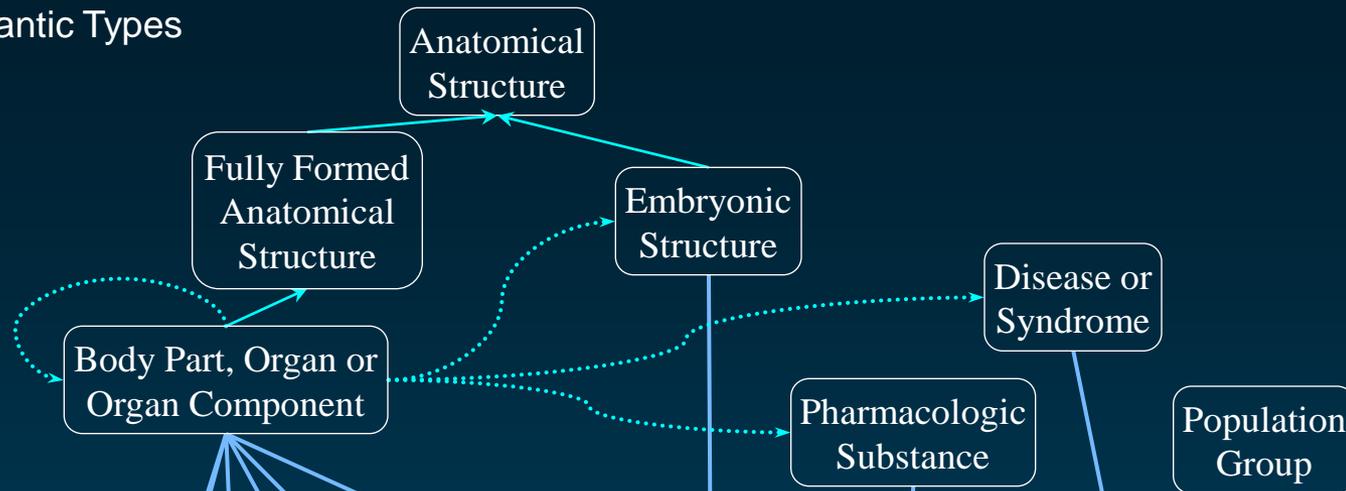
# UMLS Characteristics (2)

- ◆ Number of
  - Concepts: 2.2M (2010AA)
  - Terms: ~10M
- ◆ Major organizing principles (Metathesaurus):
  - Concept orientation
  - Source transparency
  - Multi-lingual through translation
- ◆ Formalism: Proprietary format (RRF)

# UMLS Integrating subdomains

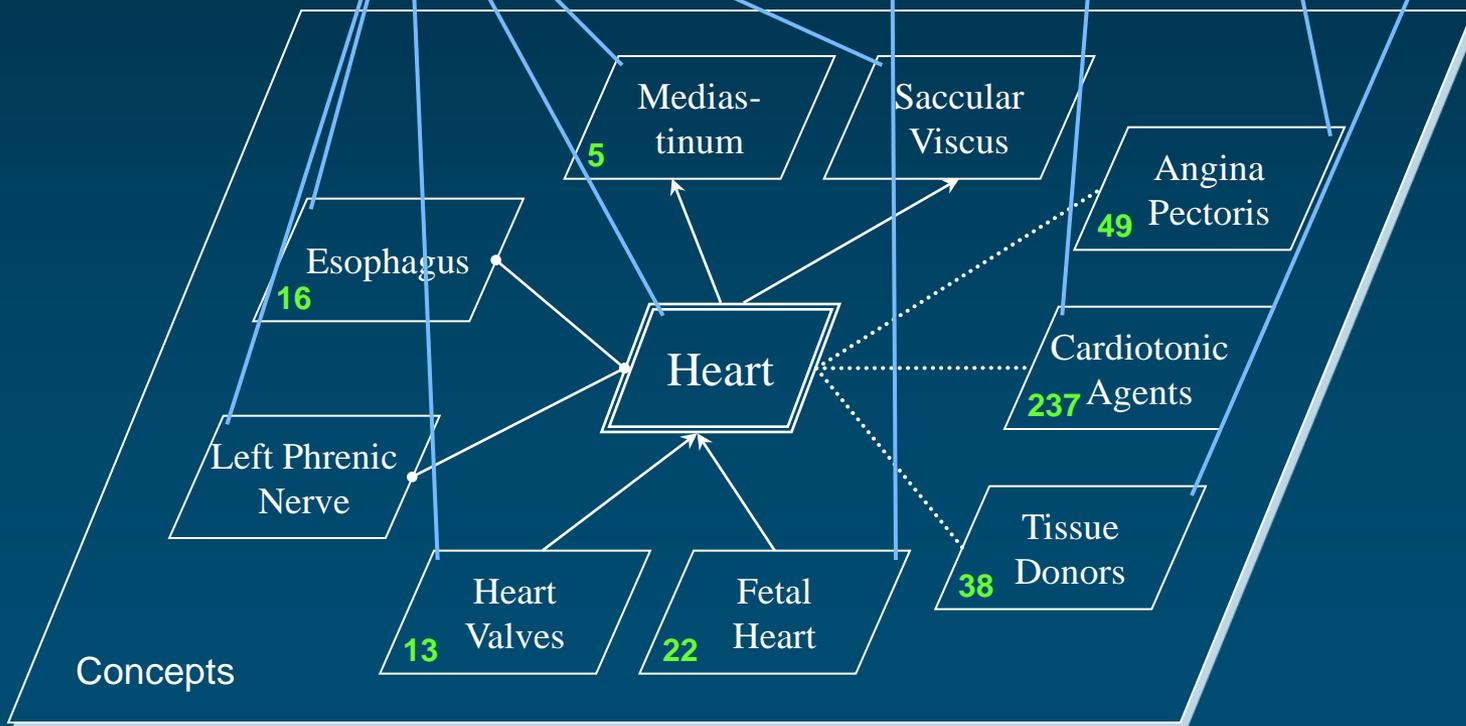


Semantic Types



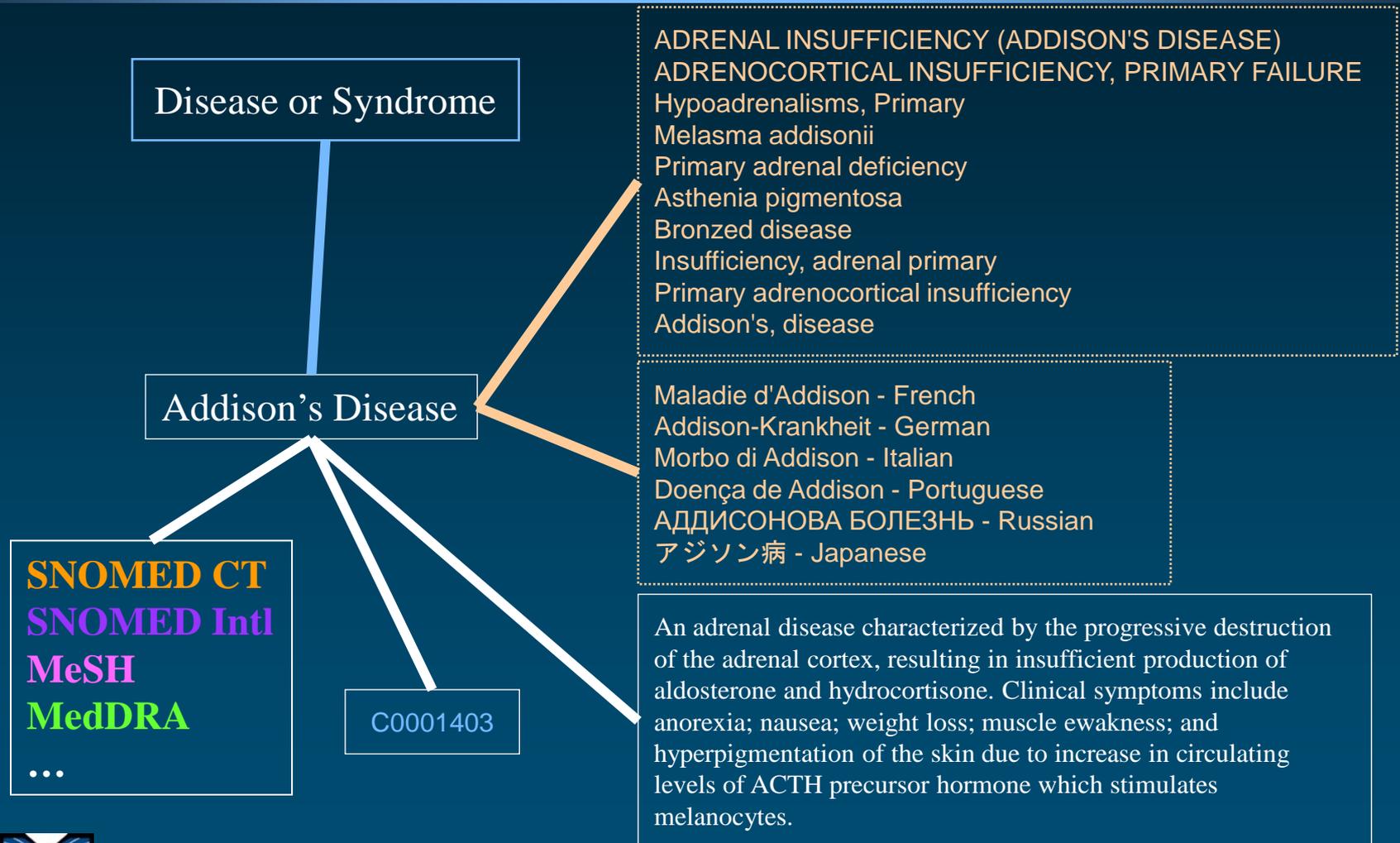
*Semantic Network*

*Metathesaurus*



Concepts

# Addison's Disease: Concept



# Metathesaurus Concepts (2010AA)

- ◆ Concept (~ 2.2M) CUI
  - Set of synonymous concept names
- ◆ Term (~ 7.5M) LUI
  - Set of normalized names
- ◆ String (~ 8.2M) SUI
  - Distinct concept name
- ◆ Atom (~ 10M) AUI
  - Concept name in a given source

A0066000	Headache	(MeSH)
A0065992	Headache	(ICD-10)
<b>S0046854</b>		

A0066007	Headaches	(MedDRA)
A12003304	Headaches	(OMIM)
<b>S0046855</b>		

**L0018681**

A0540936	Cephalodynia	(MeSH)
<b>S0475647</b>		

**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)



# Short course – Summer 2010 Clinical Ontology in Practice

June 15, 2010 – 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	<a href="#">aldehyde dehydrogenase (NAD) activity</a>	IEA
Molecular Function	<a href="#">oxidoreductase activity</a>	IEA
Molecular Function	<a href="#">oxidoreductase activity</a>	IEA
Cellular Component	<a href="#">mitochondrion</a>	IDA
Biological Process	<a href="#">metabolic process</a>	IEA
Biological Process	<a href="#">oxidation reduction</a>	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	IEA	IPR015590	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	IEA	IPR016160	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	IEA	IPR016162	UniProt	9606
GO:0016491	oxidoreductase activity	interpro	IEA	IPR016161	UniProt	9606
GO:0016491	oxidoreductase activity	spkw	IEA	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	IEA	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

□ I: [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](#), [Ruukonen E](#); [Finnsepsis Study Group](#).

► [Collaborators \(26\)](#)

Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. [Stepani.Bendel@kuh.fi](mailto: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 ( $728 \pm 386$  nmol/L vs  $793 \pm 439$  nmol/L,  $P = 0.44$ ). Nonsurvivors had higher calculated serum free ( $209 \pm 151$  nmol/L) and total ( $980 \pm 458$  nmol/L) cortisol concentrations than survivors ( $119 \pm 111$  nmol/L,  $P = 0.002$ , and  $704 \pm 383$  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](mailto: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:

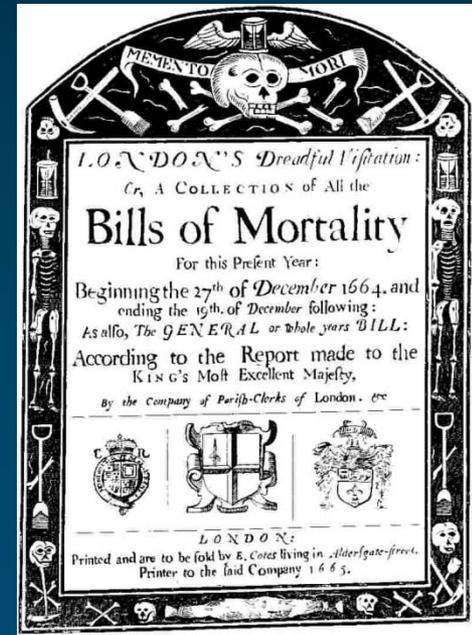
- [Addison Disease/blood](#)
- [Addison Disease/drug therapy\\*](#)
- [Androgens/administration & dosage\\*](#)
- [Androgens/therapeutic use](#)
- [Dosage Forms](#)
- [Drug Administration Routes](#)
- [Drug Administration Schedule](#)
- [Glucocorticoids/administration & dosage\\*](#)
- [Glucocorticoids/adverse 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

<b>MeSH Heading</b>	Addison Disease
<b>Entry Term</b>	Addison's Disease
<b>Entry Term</b>	Primary Adrenal Insufficiency
<b>Entry Term</b>	Primary Adrenocortical Insufficiency

# MeSH “synonyms” MEDLINE retrieval

The screenshot shows the PubMed search interface. At the top, the NCBI logo is on the left, and the PubMed logo with the URL [www.pubmed.gov](http://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 the logo is a navigation bar with tabs for "All Databases", "PubMed", "Nucleotide", "Protein", "Genome", "Structure", and "OMIM". The search bar contains the text "Search PubMed" and a dropdown menu set to "PubMed". The search term "Primary Hypoadrenalism" is entered in the search box. To the right of the search box are "Go" and "Clear" buttons. Below the search bar is a navigation bar with buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Query Translation" section is highlighted in blue and contains 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]
```

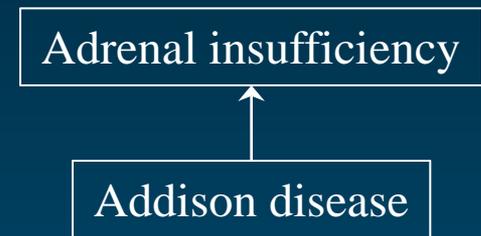
 At the bottom of the query translation section are "Search" and "URL" buttons. On the left side of the page, there is a sidebar with links for "About Entrez", "Text Version", "Entrez PubMed", "Overview", "Help | FAQ", "Tutorials", "New/Noteworthy", "E-Utilities", "PubMed Services", "Journals Database", "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 for **"adrenal insufficiency"[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 1 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

**gpubmed**

*Searching is now sorted!*

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



COX-2



what

## Top categories

- [-] (M) Chemicals and Drugs [992]
  - (M) Cyclooxygenase 2 [517]
  - (M) Cyclooxygenase 2 Inhibitors [289]
  - (M) Prostaglandins [358]
  - (M) Prostaglandin-Endoperoxide Synthases [667]
  - (M) NF-kappa B [138]
  - (M) RNA, Messenger [222]
  - (M) Anti-Inflammatory Agents [414]
  - more
- [-] (G) biological\_process [851]
  - (G) cyclooxygenase pathway [305]
  - more
- [-] (M) Biological Sciences [960]
  - (M) Up-Regulation [166]
  - more
- [-] (M) Diseases [781]
  - (M) Inflammation [192]
  - more
- [+] (M) Organisms [398]
- [+] (M) Techniques and Equipment [809]
- [+] (G) molecular\_function [483]
- [+] (M) Anatomy [778]
- [+] (M) Named Groups [285]
- [+] (G) cellular\_component [307]
- [+] (M) Natural Sciences [661]
- [+] (M) Technology, Industry, Agriculture [147]
- [+] (M) Psychiatry and Psychology [386]



Lister Hill National Center for E

# Knowledge management

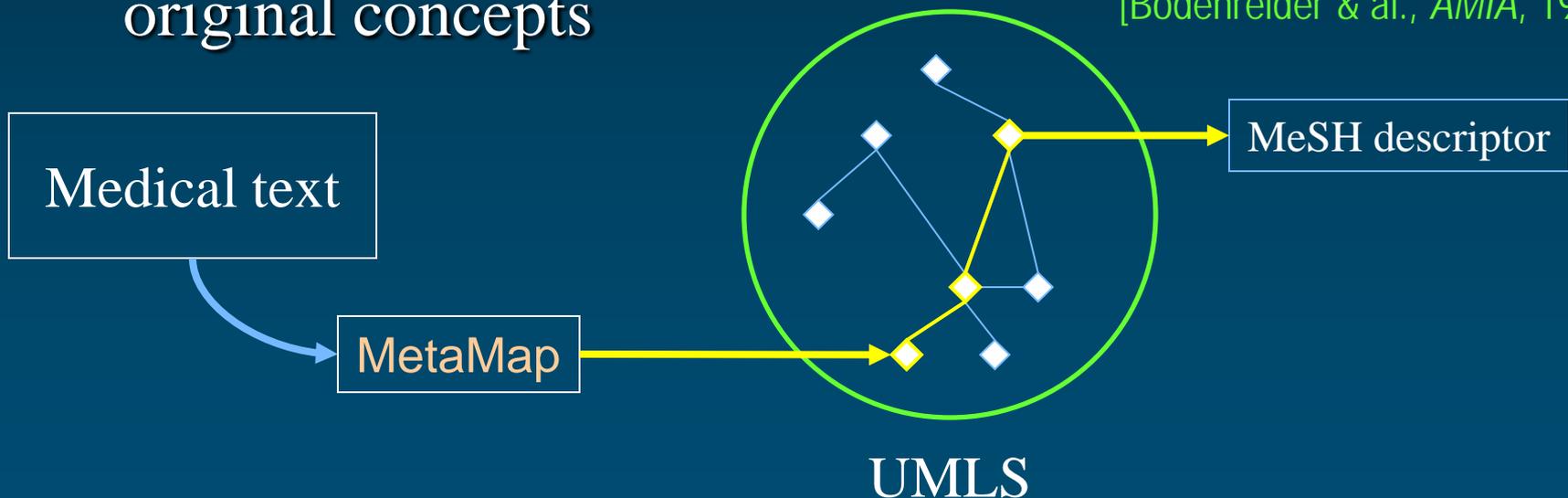
*Mapping across biomedical ontologies*

# 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

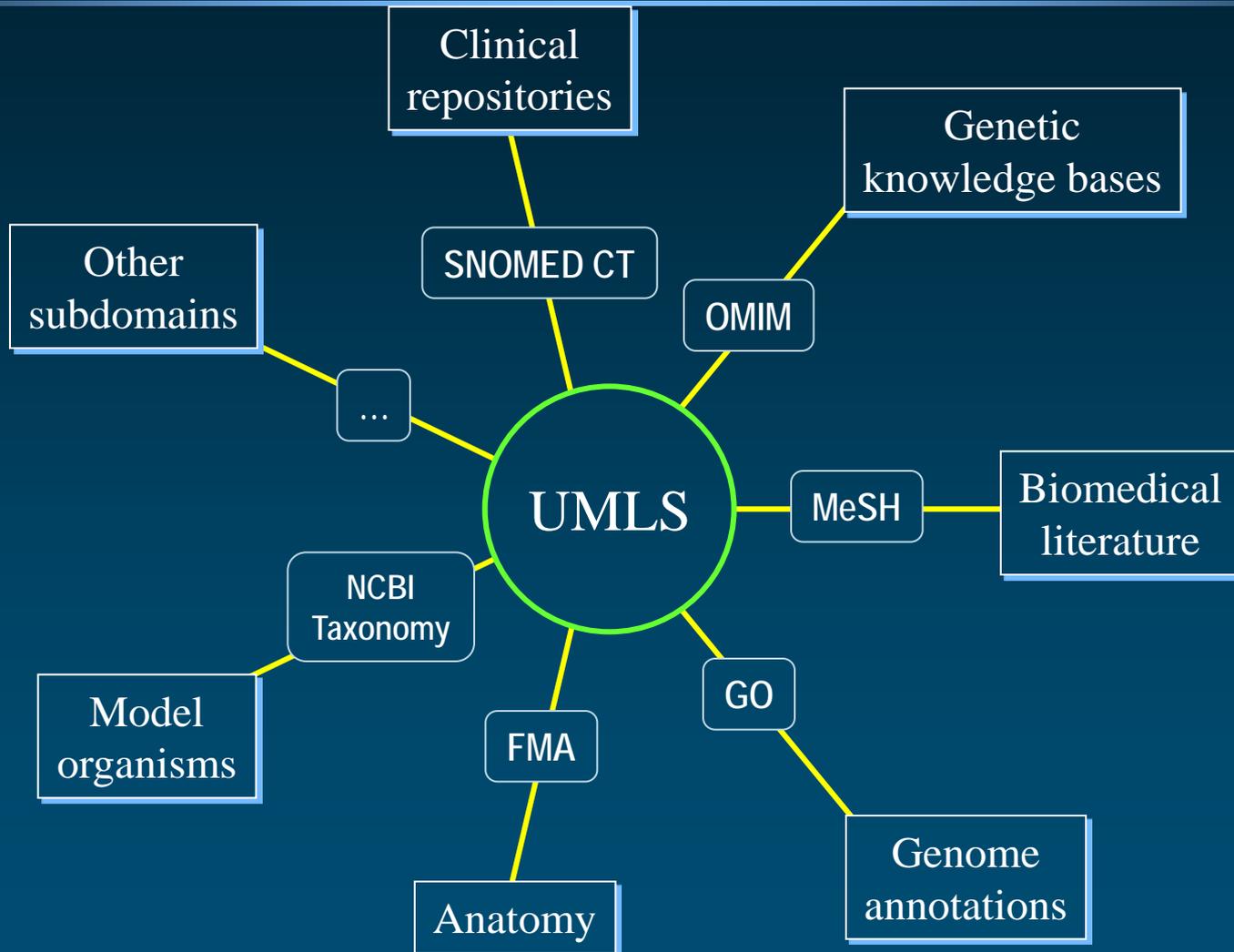
- ◆ For noun phrases extracted from medical texts, map to UMLS concepts (MetaMap) [Aronson & al., *JAMIA*, 2010]
- ◆ Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts [Bodenreider & al., *AMIA*, 1998]



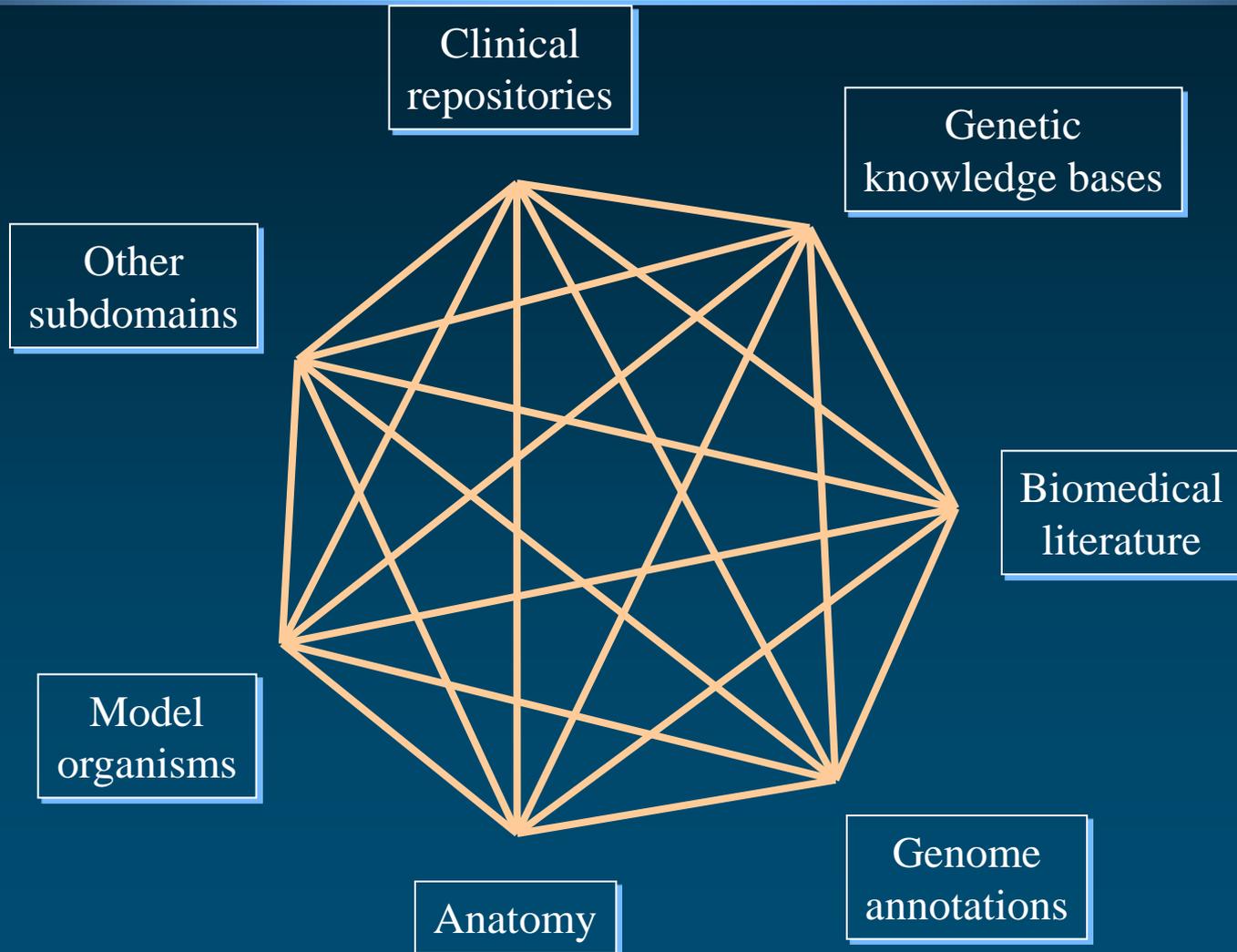
# Terminology integration systems

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

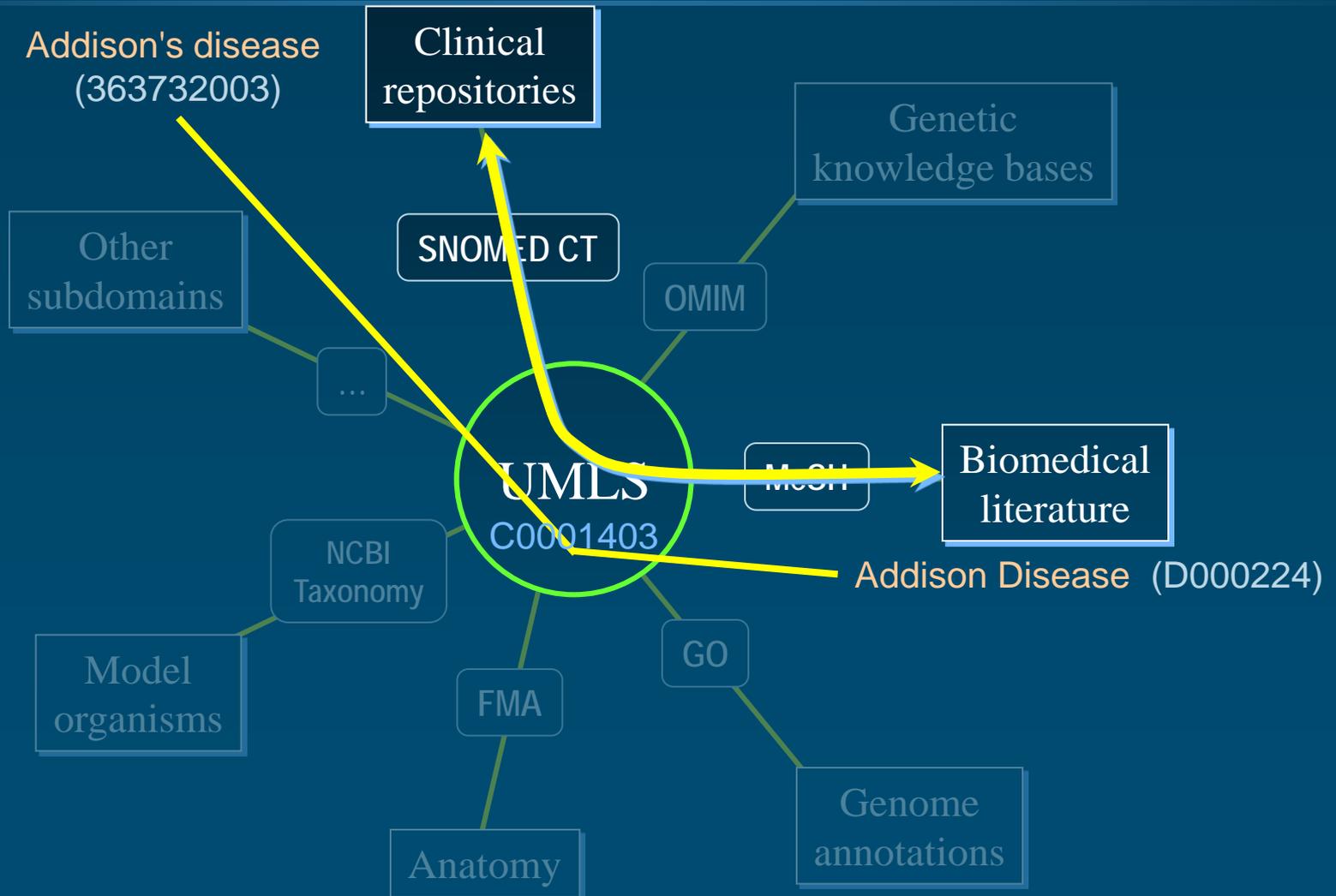
# Integrating subdomains



# 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)

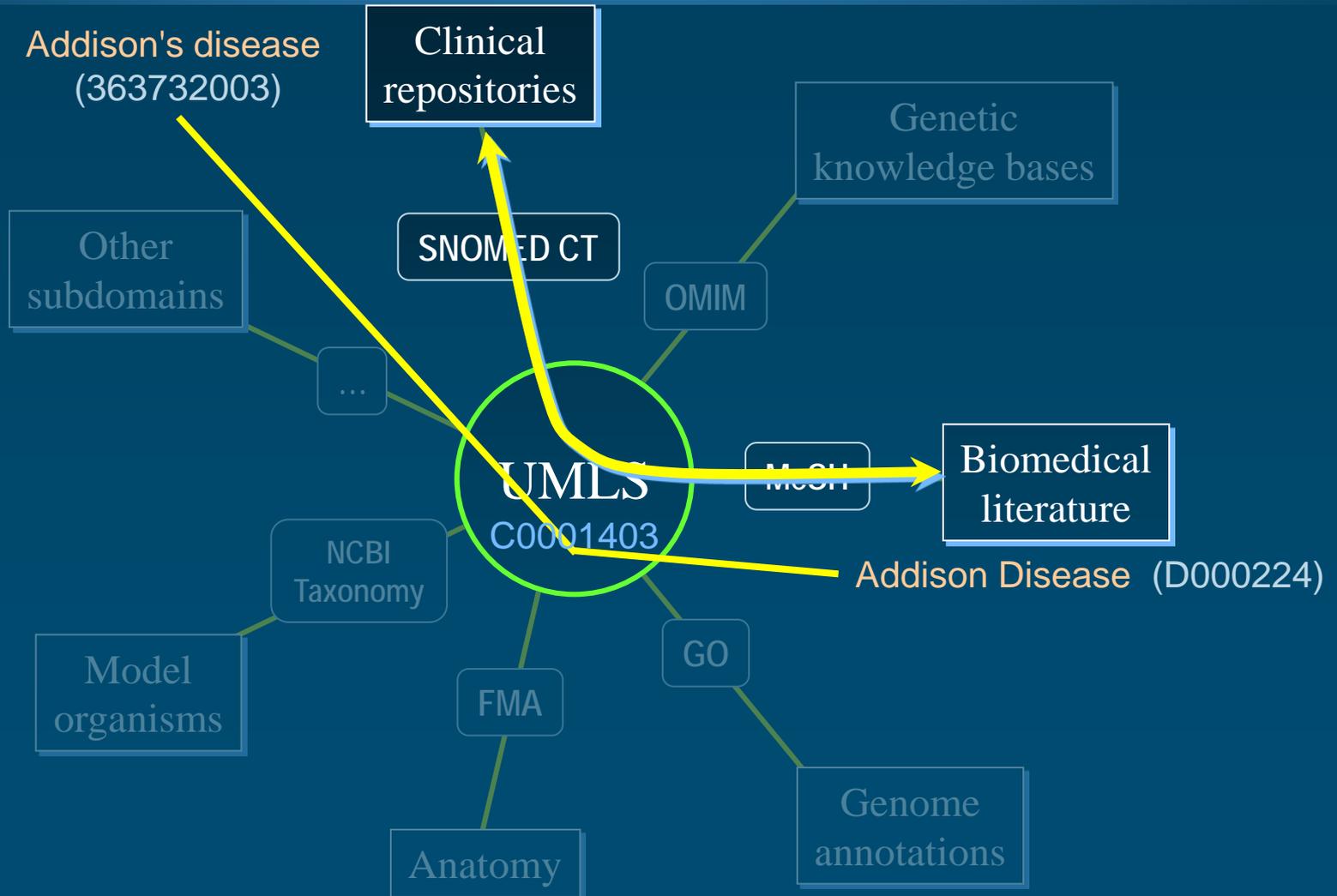
## ◆ Mediation

- 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: 156 source vocabularies  
(biased towards healthcare applications)
  - NCBO BioPortal: ~200 ontologies  
(biased towards biological applications)
  - Some 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
  - **Current harmonization efforts (e.g., IHTSDO)**
- ◆ 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



# Short course – Summer 2010 Clinical Ontology in Practice

June 16, 2010 – “Hands-on” Sessions

## Exploring Clinical Ontologies



*Olivier Bodenreider*

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

# UMLS UMLS KS

- ◆ **UMLS KS** (Knowledge Source Server)  
<http://umlsks.nlm.nih.gov/>
- ◆ Search by term: **appendectomy** (C0003611)
  - (default) RRF view (atom-centric)
  - Lexical View (normalized strings / lexical units)
  - Relations
  - Co-occurrence Info
  - Contexts (paths to root)
- ◆ Search by code
  - R73.0 (Postprocedural hypoinsulinaemia)



# UMLS UMLS KS

## ◆ Notes

- Ambiguity: appendectomy, heart, calcium
- Several kinds of lexical matches (exact, normalized, approximate)

# UMLS Semantic Navigator

- ◆ Available under UMLSKS  
(bottom of left-hand side pane)
- ◆ Search by term:
  - [appendectomy](#) (C0003611)
  - [Addison's disease](#) (C0001403)
- ◆ Concept-centric vs. atom-centric
- ◆ Selection of hierarchical relations (and co-occurrences)
- ◆ Transitive reduction on/off



# UMLSKS API

- ◆ UMLSKS Developer's Guide  
(<http://umlsks.nlm.nih.gov/>)
- ◆ Authentication vs. UMLSKS services
- ◆ SOAP-based (examples and documentation mostly for java, but usable with other environments, e.g., Perl, .NET)



# SNOMED CT

## ◆ Multiple web-based browsers available

- U. Sydney browser (specific to SNOMED CT)  
<http://www.it.usyd.edu.au/~hitru/sct/A1.cgi>
- Virginia Tech browser (specific to SNOMED CT)  
<http://terminology.vetmed.vt.edu/SCT/menu.cfm>
- The SNOMED CT Browser © (specific to SNOMED CT)  
<http://www.medicalclassifications.com/SNOMEDbrowser/>
- BioPortal  
<http://www.bioontology.org/BioPortal>
- NCI Term Browser  
<http://nciterms.nci.nih.gov/>

# SNOMED CT

## ◆ Search concepts

- Appendectomy (80146002)
- Simvastatin (387584000)
- Addison's disease (363732003)

## ◆ Notes

- No post-coordination services in standard browsers
- Some standalone browsers offer additional services (CliniClue, SNOB)
- Search on Addison's disease in The SNOMED CT Browser © does not return any results



# LOINC

## ◆ Multiple web-based browsers available

- RELMA (specific to LOINC)  
web version of a standalone application

<http://loinc.org/relma>

*NB: Citrix ICA Client required*

- BioPortal (LOINC 2.26)  
<http://www.bioontology.org/BioPortal>

- NCI Term Browser (LOINC 2.24)  
<http://nciterms.nci.nih.gov/>



# LOINC BioPortal

## ◆ BioPortal

- Graphical interface

- Search for Lithium, then navigate down the tree

- web services

[http://www.bioontology.org/wiki/index.php/NCBO\\_REST\\_services](http://www.bioontology.org/wiki/index.php/NCBO_REST_services)

- Ontology Id: 1350

- Get ID for latest version

- <http://rest.bioontology.org/bioportal/virtual/ontology/1350>

- Returns: 40400

- Get the “first” 50 terms

- <http://rest.bioontology.org/bioportal/concepts/40400/all?pagesize=50&pagenum=1>



# LOINC NCI Term Browser

## ◆ NCI Term Browser

- Search for Lithium, then navigate through the Relationships tab
- Search by code

## ◆ Search concept

- Substance concentration of lithium in urine (quantitative)
- [Lithium:Substance Concentration:Point in time:Urine:Quantitative](#)
- 25463-1



# RxNorm RxNav

## ◆ RxNav

<http://umlsks.nlm.nih.gov/>  
(launch the browser)

- ◆ Search by string (default): **zyrtec**, **clopidogrel**
  - Restrict the graph to one particular clinical drug: double-click on Cetirizine 10 MG Oral Tablet
  - RxCUI is displayed in the information bar in the bottom when clicking on a drug entity (e.g., RxCUI for Cetirizine 10 MG Oral Tablet = 309130)
  - Right-click on Cetirizine 10 MG Oral Tablet
  - View NDCs to open a window with the list of NDCs for this drug
  - View Drug Label → link out to DailyMed



# RxNorm RxNav

- ◆ Search by ID (select ID in the drop-down “Search by” menu)
  - NDC, with search string **00781168401** (one of the NDC from the list obtained from Cetirizine 10 MG Oral Tablet)
  - SNOMED ID, with search string 1039008
    - Returns: 103|C0000618||6-Mercaptopurine
- ◆ Packs: Search for **z-pak**
  - Packs displayed with double diamonds in the clinical drug / generic pack and branded drug / branded pack boxes

# RxNorm SOAP API

## ◆ RxNorm SOAP API (demo client)

[http://mor.nlm.nih.gov/perl/rxnav\\_api\\_demo.pl](http://mor.nlm.nih.gov/perl/rxnav_api_demo.pl)

## ◆ Functions

- `getRxNormVersion()`
- `getIdTypes()`
- `findRxcuiById(00904582941, 309130) → 309130`
- `getAllRelatedInfo(309130)`

## ◆ Documentation

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



# RxNorm REST API

## ◆ Test resources

- <http://rxnav.nlm.nih.gov/REST/spellingsuggestions?name=sulfametasol>
- <http://rxnav.nlm.nih.gov/REST/rxcui/10180>
- <http://rxnav.nlm.nih.gov/REST/rxcui/309130/properties>
- <http://rxnav.nlm.nih.gov/REST/rxcui/309130/ndcs>
- [http://rxnav.nlm.nih.gov/REST/rxcui/151399/related?rela=tradenname\\_of](http://rxnav.nlm.nih.gov/REST/rxcui/151399/related?rela=tradenname_of)

## ◆ Documentation

[http://rxnav.nlm.nih.gov/RxNorm\\_RESTful\\_UserGuide.pdf](http://rxnav.nlm.nih.gov/RxNorm_RESTful_UserGuide.pdf)



# NDF-RT

- ◆ **RxNav** (pilot version integrating NDF-RT)  
<http://rxnav.nlm.nih.gov/rxnavdemo.jnlp>
- ◆ Search for **clopidogrel** (RxNorm tab)  
other example: **cetirizine**
  - Double-click on clopidogrel 75 MG Oral Tablet
  - Click on the NDF-RT tab
  - Explore the relations to the different categories of entities (Drug, Disease, Dose form, ...)



# Short course – Summer 2010 Clinical Ontology in Practice

June 16, 2010 – Discussion Sessions

## Issues and Challenges Related to Clinical Ontologies



*Olivier Bodenreider*

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

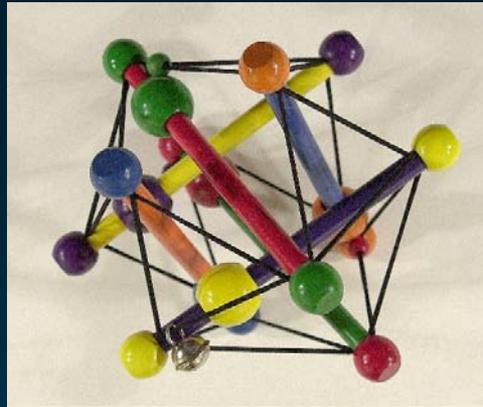
# Topics

- ◆ NLP / indexing
- ◆ PHR / consumer health information
- ◆ Decision support (drugs)
- ◆ Decision support (other)
- ◆ Medication reconciliation
- ◆ E-prescribing
- ◆ CPOE
- ◆ Problem list
- ◆ Terminology services
- ◆ Value sets
- ◆ Terminology management (versioning)
- ◆ Mapping / integration
- ◆ Meaningful use
- ◆ Health information exchange
- ◆ Clinical documentation



# Questions

- ◆ What are some of the issues and challenges related to this topic?
- ◆ Do ontologies contribute to the solution?  
Which ones? Which features?
- ◆ Have you learned anything that is applicable towards this issue?



# Medical Ontology Research

Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)

Web: [mor.nlm.nih.gov](http://mor.nlm.nih.gov)



*Olivier Bodenreider*

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