



# NLM Georgia Biomedical Informatics Course

Spring 2016 Course

## Controlled Vocabularies (aka Biomedical Terminologies/Ontologies)



*Olivier Bodenreider*

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



U.S. National Library of Medicine



# Learning objectives

- ◆ Describe the history of biomedical ontologies
- ◆ Explain how clinical features are reflected in disease names
- ◆ List and describe the main biomedical ontologies used in 21st century healthcare
- ◆ Discuss the purpose of biomedical ontologies in knowledge management, clinical decision support and analytics



# References Review articles

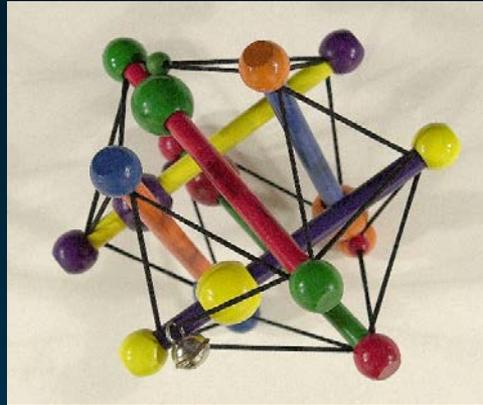
- ◆ Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. *Brief Bioinform.* 2006 Sep;7(3):256-74.
- ◆ Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearb Med Inform.* 2006:124-35.
- ◆ Bodenreider O. Biomedical ontologies in action: role in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



# Additional references

- ◆ Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998 Nov;37(4-5):394-403.
- ◆ Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004 Jan 1;32(Database issue):D267-70.





# Medical Ontology Research

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

Web: <https://mor.nlm.nih.gov>



*Olivier Bodenreider*

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



U.S. National Library of Medicine



# Outline

- ◆ Historical perspective
  - ◆ Introduction to biomedical terminologies through an example
- 
- ◆ “High-Impact” Biomedical Ontologies
    - Structural perspective
- 
- ◆ Biomedical Ontologies “in Action”
    - Functional perspective



*Controlled Vocabularies*  
*Part 1*

Historical perspective

# 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)*

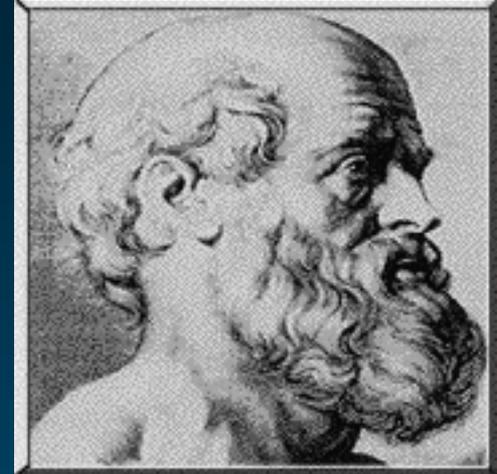
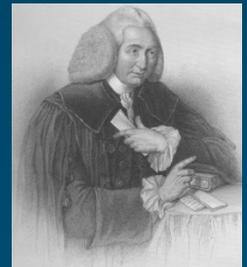
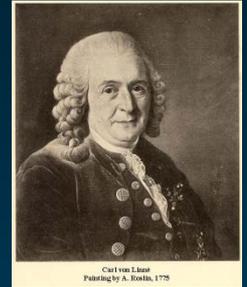


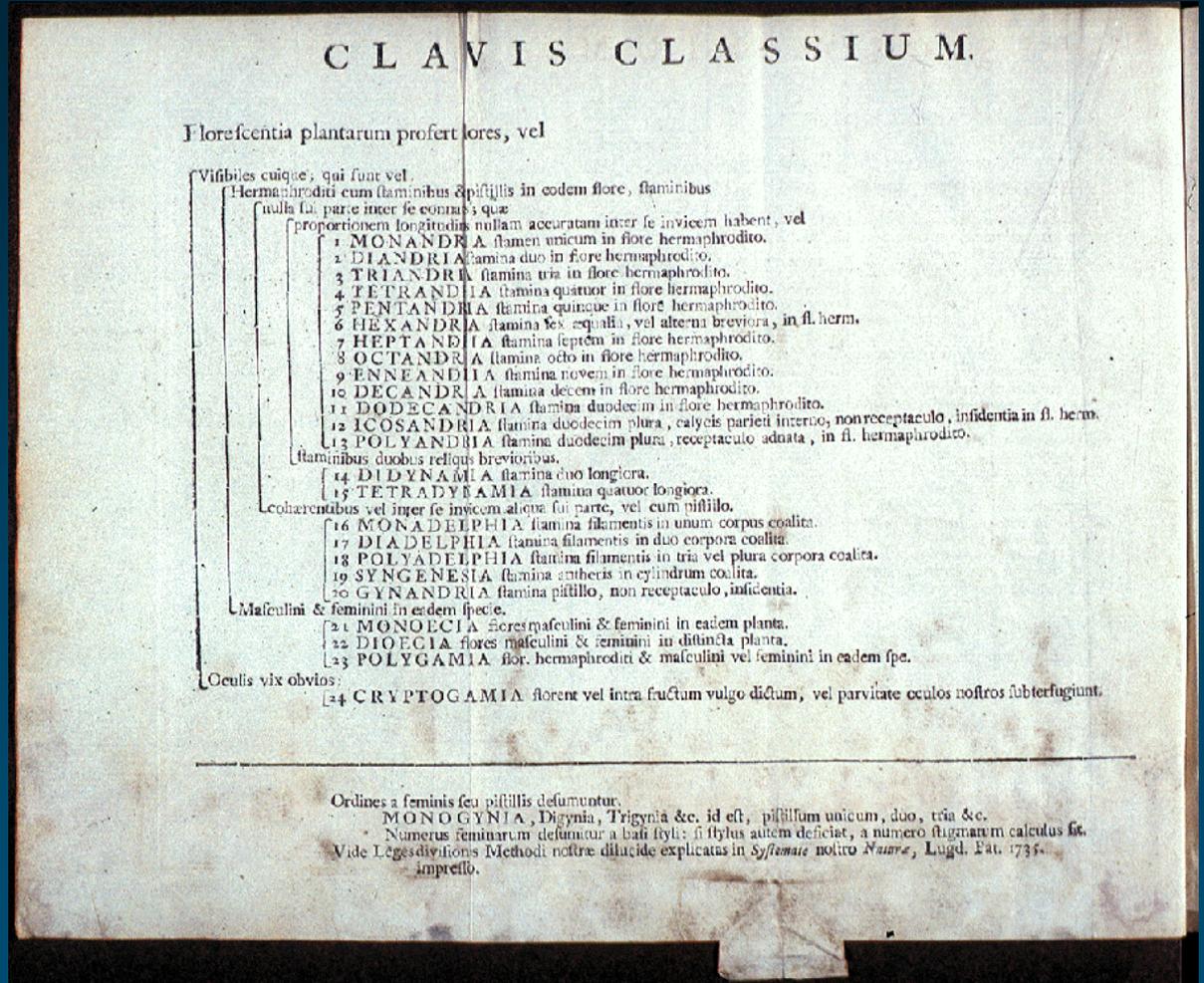
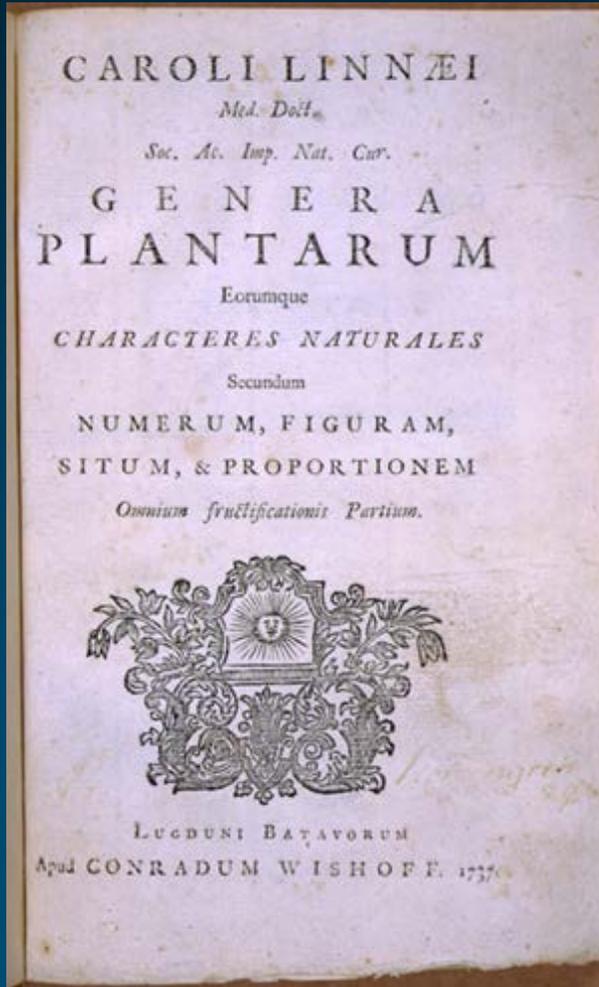
Figure 36 Thomas Sydenham (1624-1689)

# 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)



# 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. Coles living in Aldersgate-street.  
 Printer to the said Company 1665.

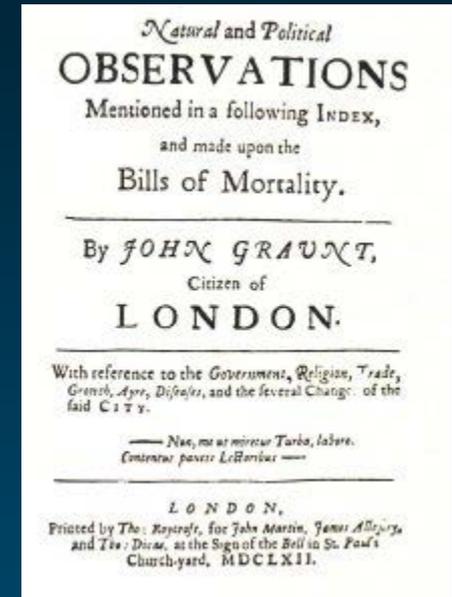
**A general Bill for this present year,**  
 ending the 19 of December 1665. according to  
 the Report made to the KING'S most Excellent Majesty.  
 By the Company of Parish Clerks of London, &c.

*The Diseases and Casualties this year.*

<b>A</b> Abortive 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	Plurisie	19
Bedric	12	Frighted	23	Posioned	2
Bedrid	1	Gout and Sciatica	27	Quinse	35
Bleeding	16	Grief	26	Rickets	137
Bloody Flux, Scouring & Flux	185	Griping in the Guts	228	Killing of the Lights	197
Burnt and Scalded	8	Hanged & made away themselves	7	Lapitate	14
Colicure	3	Headmole shot & Moxie fallen	14	Scurvy	109
Cancer, Gangrene and Fistula	56	jaundies	120	Shingles and Swine pox	2
Canker, and Thrush	121	Leptosume	227	Sores, Ulcers, broken and healed	332
Childbed	623	Killed by severall accidents	46	Lambs	82
Christomes and Infants	1258	Sings Evil	28	Spleen	14
Cold and Cough	62	Leptotic	2	Spotted Fever and Purples	1929
Collick and Winde	124	Lechary	14	Scopping of the Gromack	332
Consumption and Tiflick	4808	Livergown	21	Stone and Stranguy	28
Convulsion and Morice	1052	Meagrom and Headach	1	Suckit	1102
Distacted	3	Mealles	7	Teeth and Worms	1014
Droove and Terpany	1478	Morthered and Shot	9	Worming	51
Drward	3	Overjaid & Starved	45	Vunn	7
Colicures	5114				
Childbed	4853				
In all	9567				
		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					88590

# 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)



# 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.”

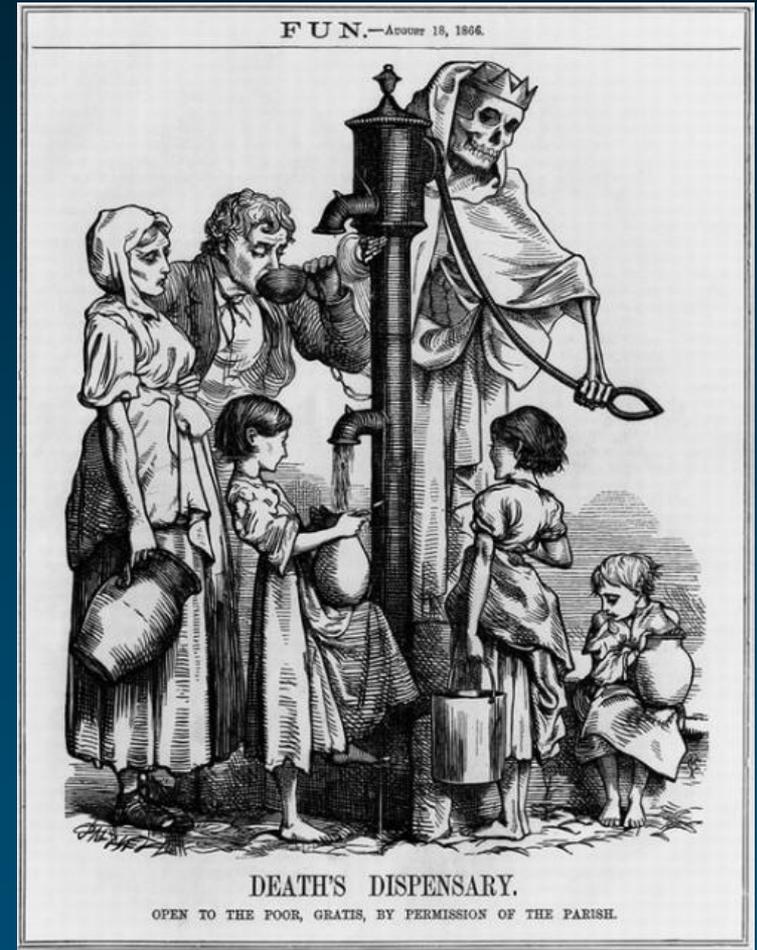
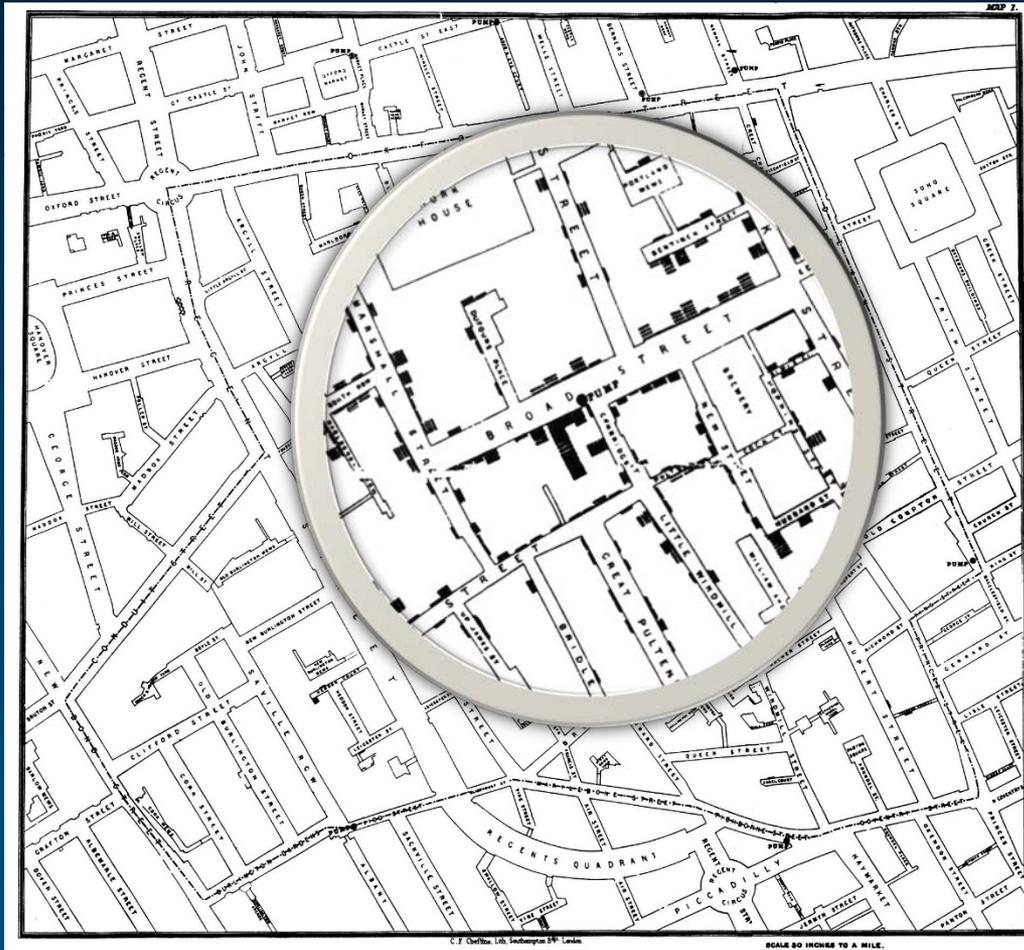
– William Farr

*First annual report.*

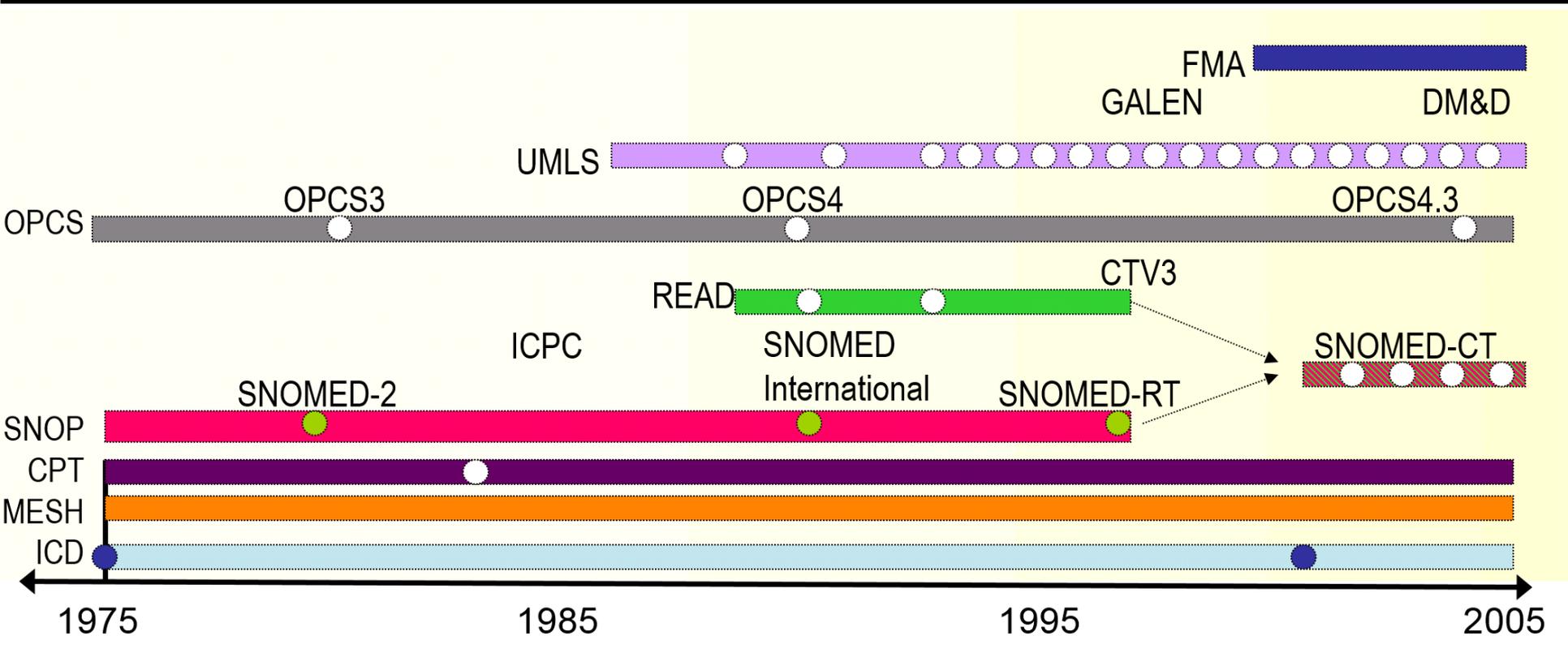
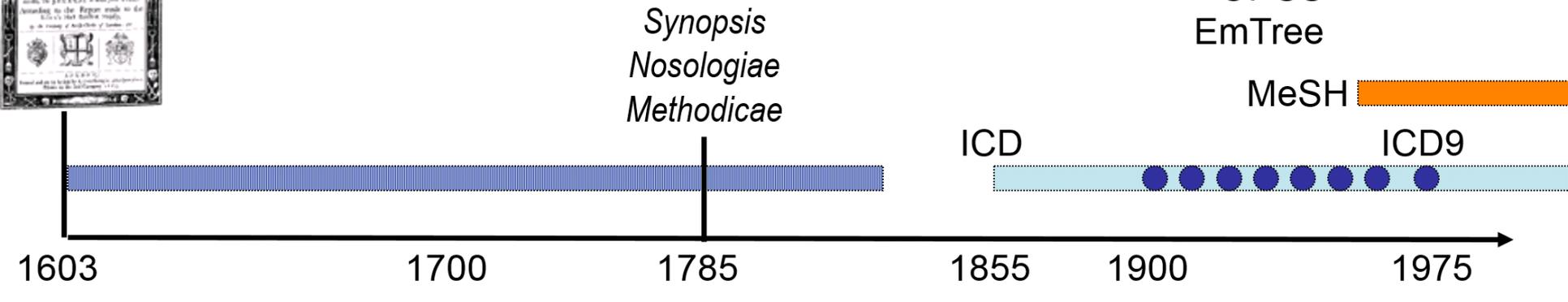
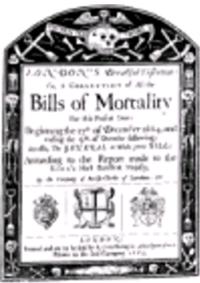
London, Registrar General of England and Wales, 1839, p. 99.



# From “bad air” to “bad water” (John Snow)



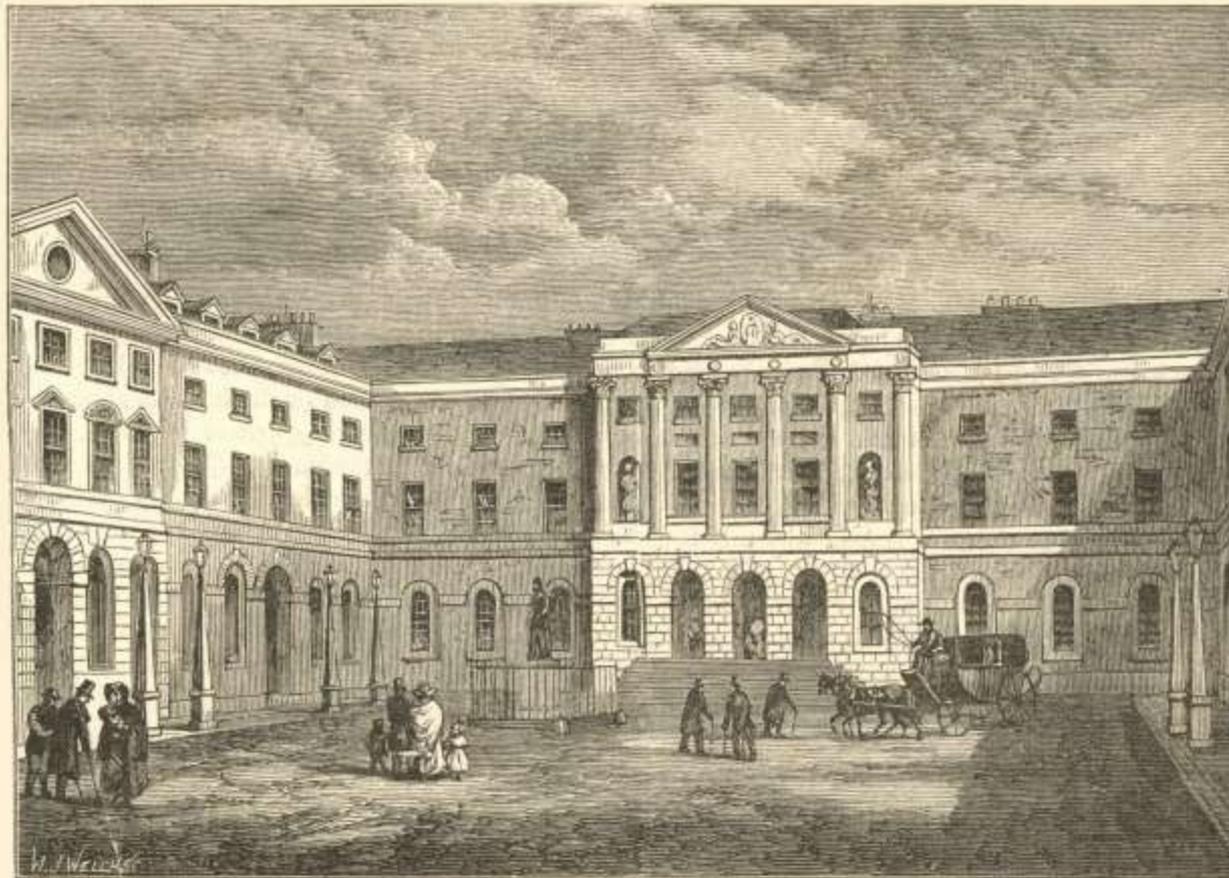
# History of Medical Ontologies



*Controlled Vocabularies*  
*Part 2*

Introduction to biomedical  
terminologies through an example

# Guy's Hospital, London



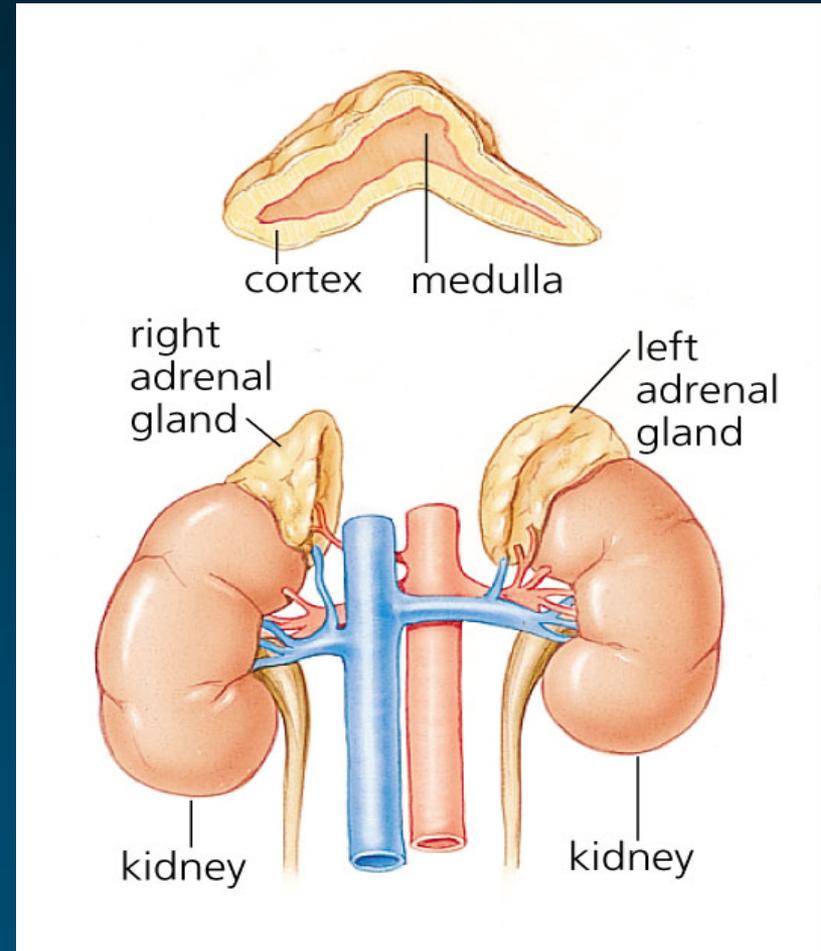
GUY'S HOSPITAL.

# Thomas Addison (1795-1860)



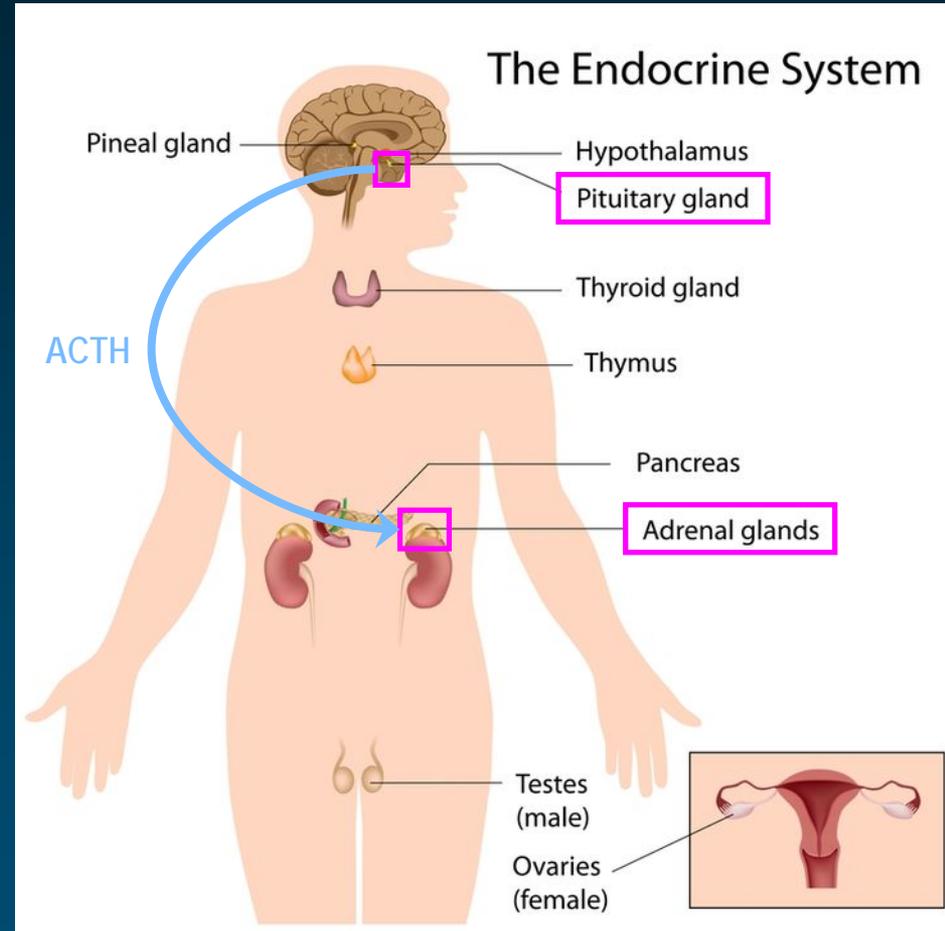
# Addison's disease

- ◆ 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

- ▼ **IV Endocrine, nutritional and metabolic diseases**
  - ▶ E00-E07 Disorders of thyroid gland
  - ▶ E10-E14 Diabetes mellitus
  - ▶ E15-E16 Other disorders of glucose regulation and pancreatic internal secretion
  - ▼ **E20-E35 Disorders of other endocrine glands**
    - ▶ E20 Hypoparathyroidism
    - ▶ E21 Hyperparathyroidism and other disorders of parathyroid gland
    - ▶ E22 Hyperfunction of pituitary gland
    - ▶ E23 Hypofunction and other disorders of pituitary gland
    - ▶ E24 Cushing syndrome
    - ▶ E25 Adrenogenital disorders
    - ▶ E26 Hyperaldosteronism
    - ▼ **E27 Other disorders of adrenal gland**
      - E27.0 Other adrenocortical overactivity
      - E27.1 Primary adrenocortical insufficiency
      - E27.2 Addisonian crisis
      - E27.3 Drug-induced adrenocortical insufficiency
      - E27.4 Other and unspecified adrenocortical insufficiency
      - E27.5 Adrenomedullary hyperfunction
      - E27.8 Other specified disorders of adrenal gland
      - E27.9 Disorder of adrenal gland, unspecified
    - ▶ E28 Ovarian dysfunction
    - ▶ E29 Testicular dysfunction
    - ▶ E30 Disorders of puberty, not elsewhere classified
    - ▶ E31 Polyglandular dysfunction
    - ▶ E32 Diseases of thymus
    - ▶ E34 Other endocrine disorders
    - ▶ E35 Disorders of endocrine glands in diseases classified elsewhere

## **E27 Other disorders of adrenal gland**

### **E27.0 Other adrenocortical overactivity**

Overproduction of ACTH, not associated with Cushing disease  
Premature adrenarache

**Excl.:** Cushing syndrome ([E24.-](#))

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

Addison disease  
Autoimmune adrenalitis

**Excl.:** amyloidosis ([E85.-](#))  
tuberculous Addison disease ([A18.7](#))  
Waterhouse-Friderichsen syndrome ([A39.1](#))

### **E27.2 Addisonian crisis**

Adrenal crisis  
Adrenocortical crisis

### **E27.3 Drug-induced adrenocortical insufficiency**

Use additional external cause code (Chapter XX), if desired, to identify drug.

### **E27.4 Other and unspecified adrenocortical insufficiency**

Adrenal:

- haemorrhage
- infarction

Adrenocortical insufficiency NOS  
Hypoadosteronism

**Excl.:** adrenoleukodystrophy [Addison-Schilder] ([E71.3](#))  
Waterhouse-Friderichsen syndrome ([A39.1](#))

### **E27.5 Adrenomedullary hyperfunction**

Adrenomedullary hyperplasia  
Catecholamine hypersecretion

### **E27.8 Other specified disorders of adrenal gland**

Abnormality of cortisol-binding globulin

### **E27.9 Disorder of adrenal gland, unspecified**



# Medical Subject Headings

## MeSH Tree Structures

[Endocrine System Diseases \[C19\]](#)

[Adrenal Gland Diseases \[C19.053\]](#)

[Adrenal Insufficiency \[C19.053.500\]](#)

▶ [Addison Disease \[C19.053.500.263\]](#)

[Adrenoleukodystrophy \[C19.053.500.270\]](#)

[Hypoaldosteronism \[C19.053.500.480\]](#)

[Waterhouse-Friderichsen Syndrome \[C19.053.500.740\]](#)

---

[Immune System Diseases \[C20\]](#)

[Autoimmune Diseases \[C20.111\]](#)

▶ [Addison Disease \[C20.111.163\]](#)

[Anemia, Hemolytic, Autoimmune \[C20.111.175\]](#)

[Anti-Glomerular Basement Membrane Disease \[C20.111.190\]](#)

[Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis \[C20.111.193\] +](#)

[Antiphospholipid Syndrome \[C20.111.197\]](#)

[Arthritis, Juvenile \[C20.111.198\]](#)

[Arthritis, Rheumatoid \[C20.111.199\] +](#)

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

[...]

# SNOMED CT

Concept Details

**Concept Details**

Summary Details Diagram Expression Re

**Parents**

- ▶ ● Abdominal organ finding (finding)
- ▶ ● Disorder of abdomen (disorder)
- ▶ ● Disorder of endocrine system (disorder)
- ▲ ● Disorder of adrenal gland (disorder)
- ▲ ● Hypoadrenalism (disorder)
- ▲ ● Adrenal hypofunction (disorder)
- ▶ ● Disorder of adrenal gland (disorder)
- ▲ ● Disorder of adrenal cortex (disorder)
- ▲ ● Adrenal cortical hypofunction (disorder)

Disorder of endocrine system

Disorder of adrenal gland

Hypoadrenalism

Adrenal hypofunction

Disorder of adrenal cortex

Adrenal cortical hypofunction

Addison's Disease

● Addison's disease (disorder) ☆ ↗

SCTID: 363732003

363732003 | Addison's disease (disorder) |

Addison disease

Addison's disease

Addison's disease (disorder)

Finding site → Adrenal cortex structure

**Children (4)**

- ● Addison's disease due to autoimmunity (disorder)
- ● Addison's disease with adrenoleucodystrophy (disorder)
- ● Polyglandular autoimmune syndrome, type 1 (disorder)
- ● Tuberculous Addison's disease (disorder)

26

*Controlled Vocabularies*  
*Part 3*

**“High-Impact” Biomedical  
Ontologies**

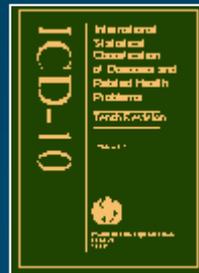
*A Structural Perspective*

# Overview

- ◆ Structural perspective
  - What are they (vs. what are they for)?
- ◆ “High-impact” biomedical ontologies [J. Cimino, YBMI 2006]
  - 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 (2016)
  - Annual updates
- ◆ Type: Classification
- ◆ Domain: Disorders
- ◆ Developer: World Health Organization (WHO)
- ◆ Funding: WHO
- ◆ Publicly available: Yes
- ◆ Used for: Mortality and morbidity statistics worldwide
- ◆ URL: <http://www.who.int/classifications/icd/en/>



# ICD Characteristics (2)

- ◆ Number of
  - Concepts: 12,320 (ICD-10, 2004)
  - 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)
- ◆ Specific coding rules
- ◆ Distribution: Proprietary format



# ICD Top level

## ▼ ICD-10 Version:2016

- ▶ I Certain infectious and parasitic diseases
- ▶ II Neoplasms
- ▶ III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- ▶ IV Endocrine, nutritional and metabolic diseases
- ▶ V Mental and behavioural disorders
- ▶ VI Diseases of the nervous system
- ▶ VII Diseases of the eye and adnexa
- ▶ VIII Diseases of the ear and mastoid process
- ▶ IX Diseases of the circulatory system
- ▶ X Diseases of the respiratory system
- ▶ XI Diseases of the digestive system
- ▶ XII Diseases of the skin and subcutaneous tissue
- ▶ XIII Diseases of the musculoskeletal system and connective tissue
- ▶ XIV Diseases of the genitourinary system
- ▶ XV Pregnancy, childbirth and the puerperium
- ▶ XVI Certain conditions originating in the perinatal period
- ▶ XVII Congenital malformations, deformations and chromosomal abnormalities
- ▶ XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- ▶ XIX Injury, poisoning and certain other consequences of external causes
- ▶ XX External causes of morbidity and mortality
- ▶ XXI Factors influencing health status and contact with health services
- ▶ XXII Codes for special purposes

# ICD Example

## ◆ Idiosyncratic inclusion/exclusion criteria

E10	Type 1 diabetes mellitus
<a href="#">[See before E10 for subdivisions]</a>	
<b>Incl.:</b>	diabetes (mellitus): <ul style="list-style-type: none"><li>• brittle</li><li>• juvenile-onset</li><li>• ketosis-prone</li></ul>
<b>Excl.:</b>	diabetes mellitus (in): <ul style="list-style-type: none"><li>• malnutrition-related (<a href="#">E12.-</a>)</li><li>• neonatal (<a href="#">P70.2</a>)</li><li>• pregnancy, childbirth and the puerperium (<a href="#">O24.-</a>)</li></ul> glycosuria: <ul style="list-style-type: none"><li>• NOS (<a href="#">R81</a>)</li><li>• renal (<a href="#">E74.8</a>)</li></ul> impaired glucose tolerance ( <a href="#">R73.0</a> ) postsurgical hypoinsulinaemia ( <a href="#">E89.1</a> )

# ICD Example

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

<b>E84</b>	<b>Cystic fibrosis</b>
	<i>Incl.:</i> mucoviscidosis
<b>E84.0</b>	<b>Cystic fibrosis with pulmonary manifestations</b>
<b>E84.1</b>	<b>Cystic fibrosis with intestinal manifestations</b>
	Distal intestinal obstruction syndrome
	Meconium ileus in cystic fibrosis† ( <a href="#">P75*</a> )
	<i>Excl.:</i> meconium obstruction (ileus) in cases where cystic fibrosis is known not to be present ( <a href="#">P76.0</a> )
<b>E84.8</b>	<b>Cystic fibrosis with other manifestations</b>
<b>E84.9</b>	<b>Cystic fibrosis, unspecified</b>

# ICD-10-CM

- ◆ Derived from: ICD-10
  - Finer-grained (both clinically and administratively)
- ◆ Type: Classification
  - 92,042 codes (2015)
  - Terms: 1.2 per concept
- ◆ Domain: Disorders
- ◆ Developer: National Center for Health Statistics (NCVHS)
- ◆ Funding: U.S. Government
- ◆ Publicly available: Yes
- ◆ Used for: Billing
- ◆ URL: <http://www.cdc.gov/nchs/icd/icd10cm.htm>



# ICD-10 vs. ICD-10-CM

**E72 Other disorders of amino-acid metabolism**  
*Excl.:* abnormal findings without manifest disease (R7) disorders of:  
 • aromatic amino-acid metabolism (E70.-)  
 • branched-chain amino-acid metabolism (E71.0-E71.2)  
 • fatty-acid metabolism (E71.3)  
 • purine and pyrimidine metabolism (E79.-) gout (M10.-)

**E72.0 Disorders of amino-acid transport**  
 Cystine storage disease† (N29.8\*)  
 Cystinosis  
 Cystinuria  
 Fanconi(-de Toni)(-Debré) syndrome  
 Hartnup disease  
 Lowe syndrome  
*Excl.:* disorders of tryptophan metabolism (E70.8)



**E72 Other disorders of amino-acid metabolism**  
**Excludes1:** disorders of:  
 aromatic amino-acid metabolism (E70.-)  
 branched-chain amino-acid metabolism (E71.0-E71.2)  
 fatty-acid metabolism (E71.3)  
 purine and pyrimidine metabolism (E79.-)  
 gout (M1A.-, M10.-)

**E72.0 Disorders of amino-acid transport**  
**Excludes1:** disorders of tryptophan metabolism (E70.5)

**E72.00 Disorders of amino-acid transport, unspecified**  
**E72.01 Cystinuria**  
**E72.02 Hartnup's disease**  
**E72.03 Lowe's syndrome**  
**Use additional code for associated glaucoma (H42)**  
**E72.04 Cystinosis**  
 Fanconi (-de Toni) (-Debré) syndrome with cystinosis  
**Excludes1:** Fanconi (-de Toni) (-Debré) syndrome with cystinosis  
**E72.09 Other disorders of amino-acid transport**  
 Fanconi (-de Toni) (-Debré) syndrome, unspecified

x6

# ICD-10 vs. ICD-10-CM

**W58** Bitten or struck by crocodile or alligator



## W58 Contact with crocodile or alligator

The appropriate 7th character is to be added to each code from category W58

A - initial encounter

D - subsequent encounter

S - sequela

### W58.0 Contact with alligator

W58.01 Bitten by alligator

W58.02 Struck by alligator

W58.03 Crushed by alligator

W58.09 Other contact with alligator

W58.01A Bitten by alligator, initial encounter

W58.01D Bitten by alligator, subsequent encounter

W58.01S Bitten by alligator, sequela

### W58.1 Contact with crocodile

W58.11 Bitten by crocodile

W58.12 Struck by crocodile

W58.13 Crushed by crocodile

W58.19 Other contact with crocodile

x24



# Logical Observation Identifiers, Names and Codes (LOINC)



# LOINC Characteristics (1)

- ◆ Current version: 2.54 (Dec. 2015)
  - 2 annual releases
- ◆ Type: Controlled terminology\*
- ◆ Domain: Laboratory and clinical observations
- ◆ Developer: Regenstrief Institute
- ◆ Funding: NLM and other sources
- ◆ Publicly available: Yes
- ◆ Used for: information exchange
- ◆ URL: <https://loinc.org/>



# LOINC Characteristics (2)

- ◆ Number of
  - Concepts: 73,958 active codes (2.52, June 2015)
  - Terms: 1 per concept (“long name”)
- ◆ Major organizing principles:
  - No hierarchical structure among the main codes
  - 6 axes
    - Component (analyte [+ challenge] [+ adjustments])
    - Property
    - Timing
    - System
    - Scale
    - [Method]
- ◆ Distribution: proprietary database format



# LOINC Example

- ◆ *Sodium [Moles/volume] in Serum or Plasma*  
[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	--

## 2951-2 Sodium [Moles/volume] in Serum or Plasma

### NAME

Fully-Specified Name:	<b>Component</b>	<b>Property</b>	<b>Time</b>	<b>System</b>	<b>Scale</b>	<b>Method</b>
	Sodium	SCnc	Pt	Ser/Plas	Qn	

### PART DEFINITION/DESCRIPTION(S)

Sodium is an essential nutrient that regulates blood volume, blood pressure, osmotic equilibrium and electrolyte balance. Sodium chloride is the principal source of sodium in the diet, and is used for seasoning and as a preservative. Increased levels of sodium intake can cause hypertension and reportedly leads to 7.6 million premature deaths worldwide. Sodium is also important in neuron function and osmoregulation between cells and the extracellular fluid.

**Copyright:** Text is available under the Creative Commons Attribution/Share-Alike License. See <http://creativecommons.org/licenses/by-sa/3.0/> for details.

Source: Wikipedia, URL: [Sodium \(Wikipedia\)](#)

### BASIC ATTRIBUTES

Class/Type:	CHEM/Lab
CDISC Lab Test:	Y
Common Lab Results Rank:	#5
Common SI Lab Results Rank:	#5
Common Orders Rank:	#107
Last Updated in Version:	2.34
Order vs. Obs.:	Both
Status:	Active

### EXAMPLE UNITS

<b>Unit</b>	<b>Source Type</b>
mmol/L	EXAMPLE UCUM UNITS
mmol/L	REGENSTRIEF
mmol/L	eCHN

### UNITS AND RANGE

<b>Range</b>	<b>Units Type</b>
mmol/L:[136,145]	

# SNOMED Clinical Terms



# SNOMED CT Characteristics (1)

- ◆ Current version: January 31, 2016
  - 2 annual releases
- ◆ Type: Reference terminology / ontology
- ◆ Domain: Clinical medicine
- ◆ Developer: IHTSDO
- ◆ Funding: IHTSDO member countries
- ◆ Publicly available: Yes\*
- ◆ Used for: clinical documentation, information exchange, analytics
- ◆ URL: <http://www.ihtsdo.org/>



# SNOMED CT Characteristics (2)

- ◆ Number of
  - Concepts: 320,912 active concepts (Sept. 2016)
  - Terms: 2.6 per concept (“descriptions”)
- ◆ Major organizing principles:
  - Polyhierarchy
  - Rich set of associative relationships
  - Logical definitions (incomplete: many primitives)
  - Built using description logics (EL++)
- ◆ Distribution: RF2 (proprietary)



# SNOMED CT Top level

- ▼ ● SNOMED CT Concept
  - ▶ ● Body structure (body structure)
  - ▶ ● Clinical finding (finding)
  - ▶ ● Environment or geographical location (environment / location)
  - ▶ ● Event (event)
  - ▶ ● Observable entity (observable entity)
  - ▶ ● Organism (organism)
  - ▶ ● Pharmaceutical / biologic product (product)
  - ▶ ● Physical force (physical force)
  - ▶ ● Physical object (physical object)
  - ▶ ● Procedure (procedure)
  - ▶ ● Qualifier value (qualifier value)
  - ▶ ● Record artifact (record artifact)
  - ▶ ● Situation with explicit context (situation)
  - ▶ ● SNOMED CT Model Component (metadata)
  - ▶ ● Social context (social concept)
  - ▶ ● Special concept (special concept)
  - ▶ ● Specimen (specimen)
  - ▶ ● Staging and scales (staging scale)
  - ▶ ● Substance (substance)

# SNOMED CT Example

## Parents

- ▶ ☰ Operation on appendix (procedure)
- ▶ ☰ Partial excision of large intestine (procedure)

## ☰ Appendectomy (procedure) ☆ ↗

SCTID: 80146002

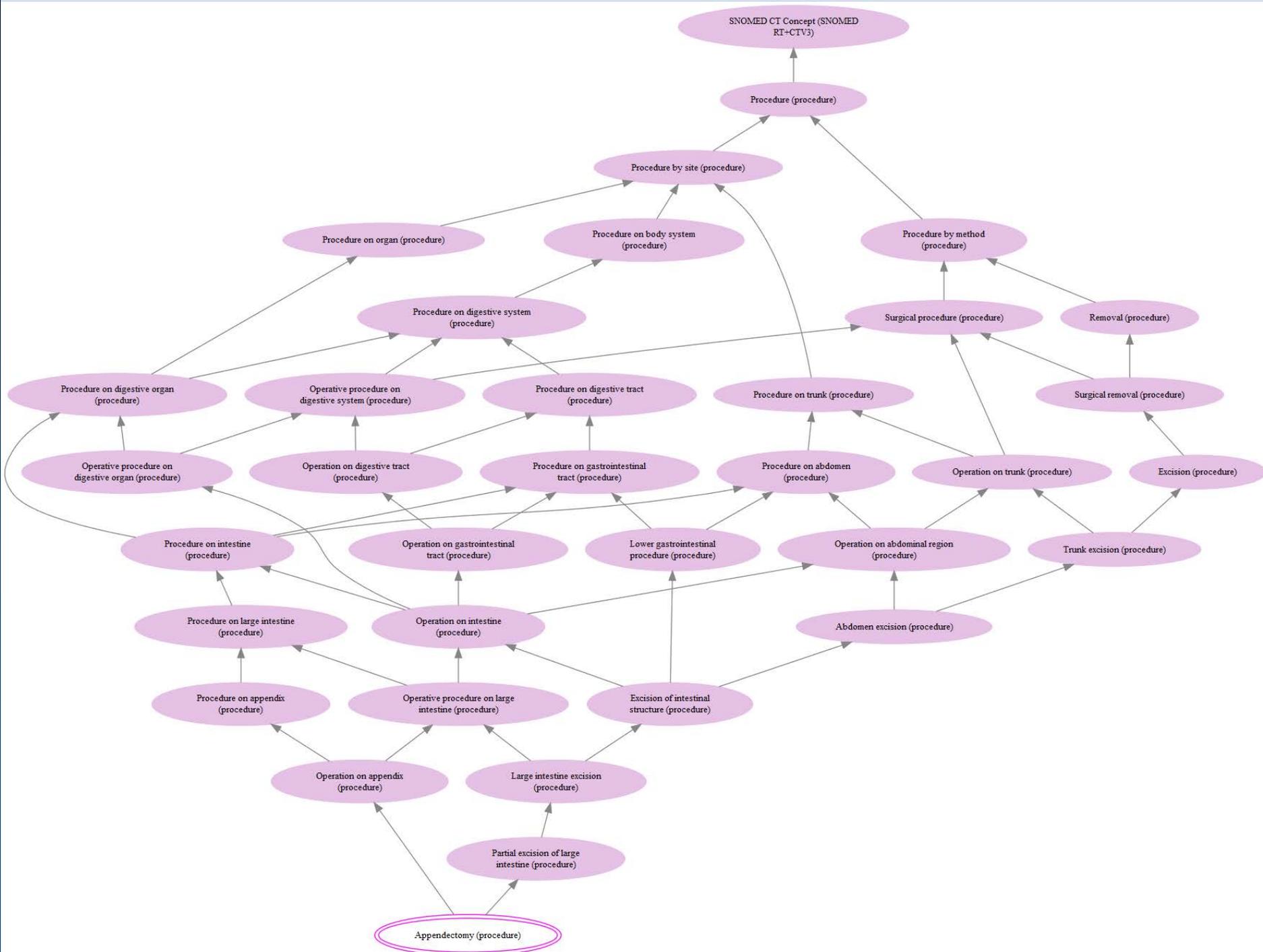
80146002 | Appendectomy (procedure) |

Appendectomy  
Excision of appendix  
Appendicectomy  
Appendectomy (procedure)

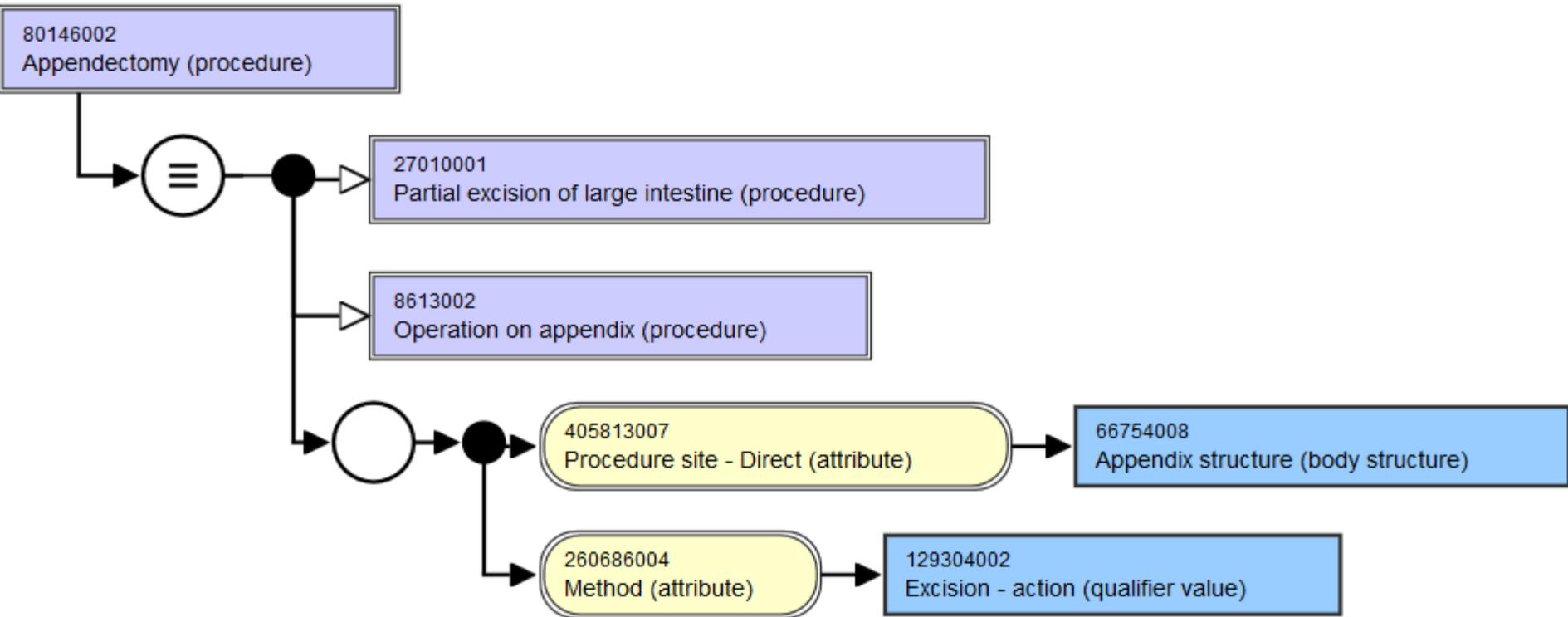
Procedure site - Direct → Appendix structure  
Method → Excision - action

## Children (8)

- ☰ Appendectomy with drainage (procedure)
- ▶ ☰ Emergency appendectomy (procedure)
- ● Excision of appendiceal stump (procedure)
- ● Excision of ruptured appendix by open approach (procedure)
- ● Incidental appendectomy (procedure)
- ● Interval appendectomy (procedure)
- ▶ ☰ Laparoscopic appendectomy (procedure)
- ☰ Non-emergency appendectomy (procedure)



# SNOMED CT Example



RxNorm

# RxNorm Characteristics (1)

- ◆ Current version: April 2016
  - Monthly releases (+weekly updates)
- ◆ Type: Controlled terminology
- ◆ Domain: Drug names
- ◆ Developer: NLM
- ◆ Funding: NLM
- ◆ Publicly available: Yes\*
- ◆ Used for: e-prescribing, information exchange, analytics
- ◆ URL: <http://www.nlm.nih.gov/research/umls/rxnorm/>



# RxNorm Characteristics (2)

- ◆ Number of
  - Concepts: 117,774 (March 2016)
  - Terms: 1.5 per concept
- ◆ Major organizing principles:
  - Generic vs. brand
  - Ingredient + Strength + Dose form
  - No hierarchical structure; rich graph of associative relations
  - Integrates all major US drug information sources
  - No clinical information
- ◆ Distribution: similar to 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**

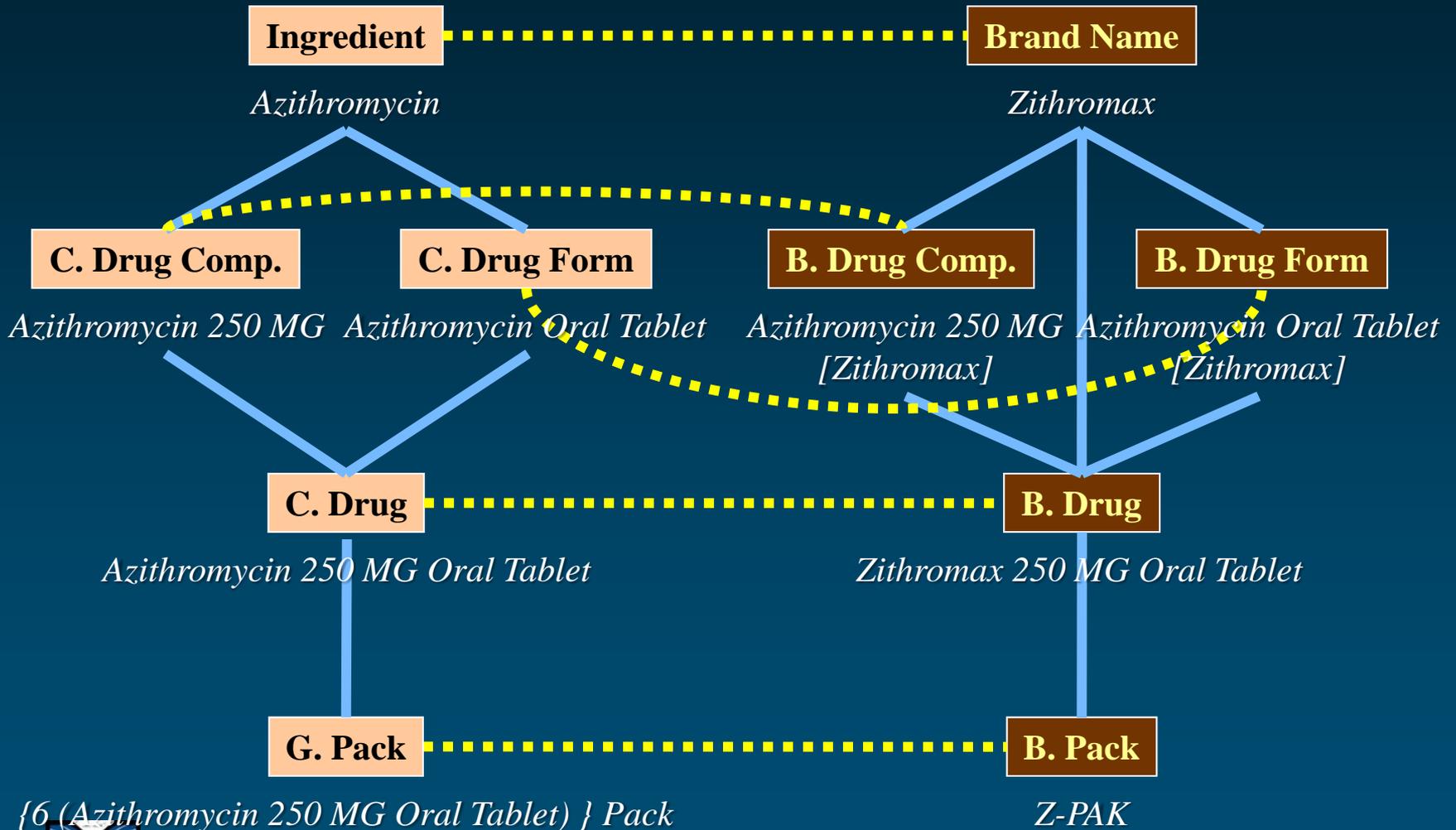
Semantic clinical drug

**Ingredient**

**Dose form**

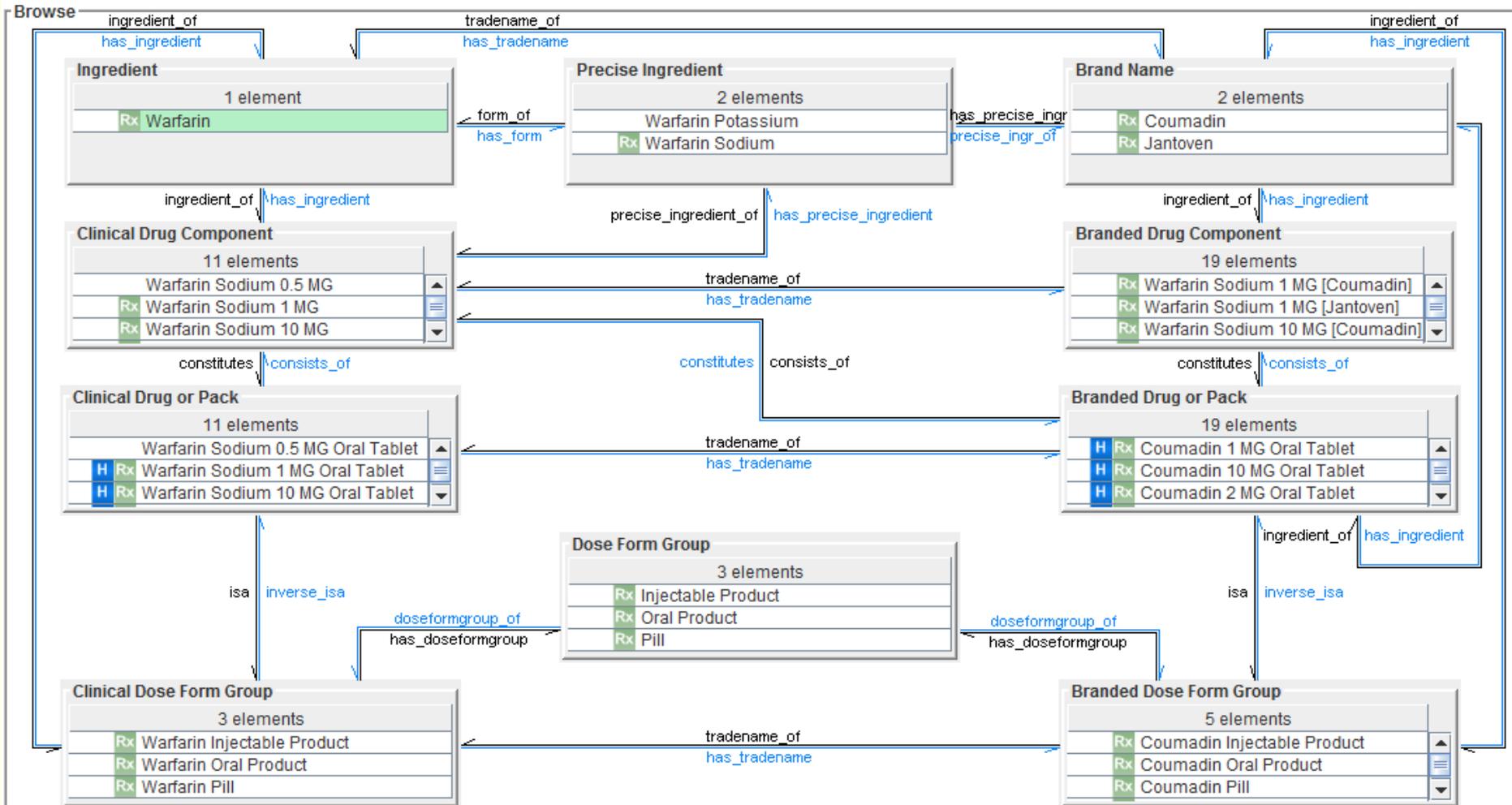


# RxNorm Example



Version: 07-Mar-2016

Search By: String Search String: warfarin Search Select Filter Group View RxNav



Copy

Status

Retrieved "Warfarin" for String "warfarin".

Drug Label

MedlinePlus

Drug Info

*Controlled Vocabularies*  
*Part 4*

**Biomedical Ontologies “in Action”**

*A Functional Perspective*

# 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 analytics** (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
- ◆ 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 across species

## ALDH2 aldehyde dehydrogenase 2 family (mitochondrial) [ *Homo sapiens* (human) ]

Gene ID: 217, updated on 13-Mar-2016

Gene Ontology [Provided by GOA](#)

Function	Evidence Code	Pubs
<a href="#">aldehyde dehydrogenase (NAD) activity</a>	EXP	
<a href="#">aldehyde dehydrogenase (NAD) activity</a>	IDA	<a href="#">PubMed</a>
<a href="#">aldehyde dehydrogenase [NAD(P)+] activity</a>	TAS	<a href="#">PubMed</a>
<a href="#">electron carrier activity</a>	TAS	<a href="#">PubMed</a>

## Aldh2 aldehyde dehydrogenase 2, mitochondrial [ *Mus musculus* (house mouse) ]

Gene ID: 11669, updated on 26-Jan-2016

Gene Ontology [Provided by MGI](#)

Function	Evidence Code	Pubs
<a href="#">NADH binding</a>	ISO	
<a href="#">aldehyde dehydrogenase (NAD) activity</a>	IBA	
<a href="#">aldehyde dehydrogenase (NAD) activity</a>	ISO	
<a href="#">identical protein binding</a>	ISO	
<a href="#">oxidoreductase activity</a>	IEA	
<a href="#">oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor</a>	IEA	
<a href="#">protein binding</a>	IPI	<a href="#">PubMed</a>



# 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
- Specific indexing rules

# MeSH MEDLINE indexing

*Anesth Analg.* 2008 Jun;106(6):1813-9. doi: 10.1213/ane.0b013e318172fdb.

## Free cortisol in sepsis and septic shock.

Bendel S<sup>1</sup>, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E; Finnsepsis Study Group.

⊕ Collaborators (26)

⊕ Author information

### Abstract

**BACKGROUND:** Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotrophic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome.

**METHODS:** In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations.

**RESULTS:** Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ( $r = 0.90$ ,  $P < 0.001$ ). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 +/- 386 nmol/L vs 793 +/- 439 nmol/L,  $P = 0.44$ ). Nonsurvivors had higher calculated serum free (209 +/- 151 nmol/L) and total (980 +/- 458 nmol/L) cortisol concentrations than survivors (119 +/- 111 nmol/L,  $P = 0.002$ , and 704 +/- 383 nmol/L,  $P = 0.002$ ). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%.

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

# MeSH MEDLINE indexing

## MeSH Terms

Adrenal Cortex Function Tests

Adrenal Insufficiency/blood\*

Adrenal Insufficiency/drug therapy

Adrenal Insufficiency/mortality

Adult

Biomarkers/blood

Female

Finland/epidemiology

Hospital Mortality

Humans

Hydrocortisone/blood\*

Hydrocortisone/therapeutic use

Kaplan-Meier 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



# SNOMED CT/ICD Coding clinical data

## ◆ SNOMED CT

- Used for clinical documentation
- E.g., problem lists

## ◆ ICD-10-CM

- Used for coding clinical data for billing purposes
- Other uses of ICD
  - Morbidity and mortality reporting worldwide
- Specific coding rules



# 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 (query expansion)
  - 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
<b>Entry Term</b>	Primary Hypoadrenalism



# MeSH “synonyms” MEDLINE retrieval

NCBI Resources How To Sign in to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed primary hypoadrenalism Search

Create RSS Create alert Advanced Help

**Search details**

```
"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]
```

Search See more...



# 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 its descendant, Addison disease

PubMed.gov

US National Library of Medicine  
National Institutes of Health

PubMed

adrenal insufficiency [mh]



Search

Create RSS Create alert Advanced

Help

- [Adrenal insufficiency in prolonged critical illness.](#)  
145. Wu JY, Hsu SC, Ku SC, Ho CC, Yu CJ, Yang PC.  
Crit Care. 2008;12(3):R65. doi: 10.1186/cc6895. Epub 2008 May 8.  
PMID: 18466605 **Free PMC Article**  
[Similar articles](#)
- [Addison's disease: a rare cause of hypertransaminasaemia.](#)  
146. Ersan O, Demirezer B.  
Dig Dis Sci. 2008 Dec;53(12):3269-71. doi: 10.1007/s10620-008-0297-8. Epub 2008 May 9.  
PMID: 18465237  
[Similar articles](#)

### MeSH Terms

[Adrenal Insufficiency/blood](#)  
[Adrenal Insufficiency/drug therapy](#)  
[Adrenal Insufficiency/mortality\\*](#)

### MeSH Terms

[Addison Disease/blood\\*](#)  
[Addison Disease/complications](#)  
[Addison Disease/diagnosis\\*](#)

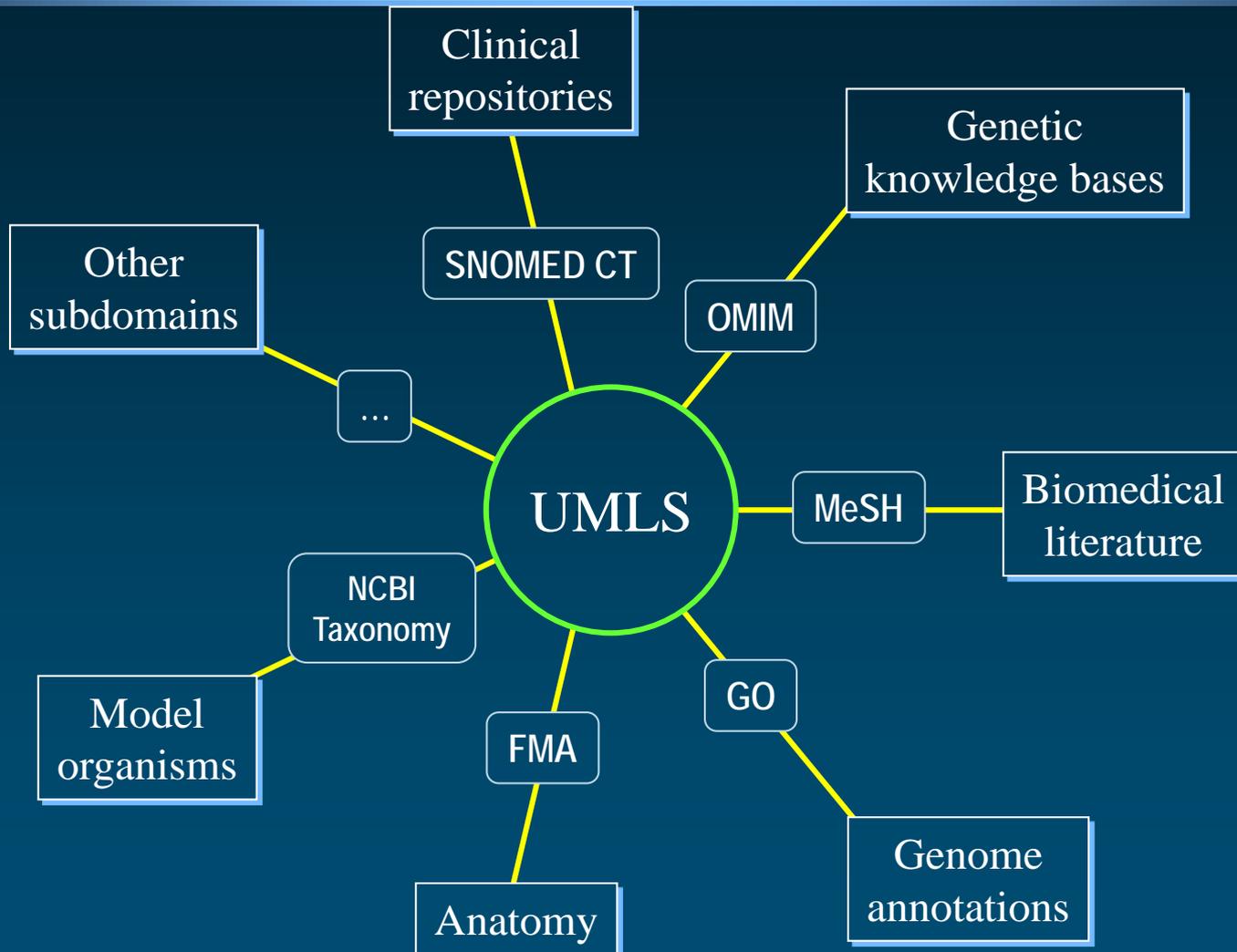
# Knowledge management

*Mapping across biomedical ontologies*

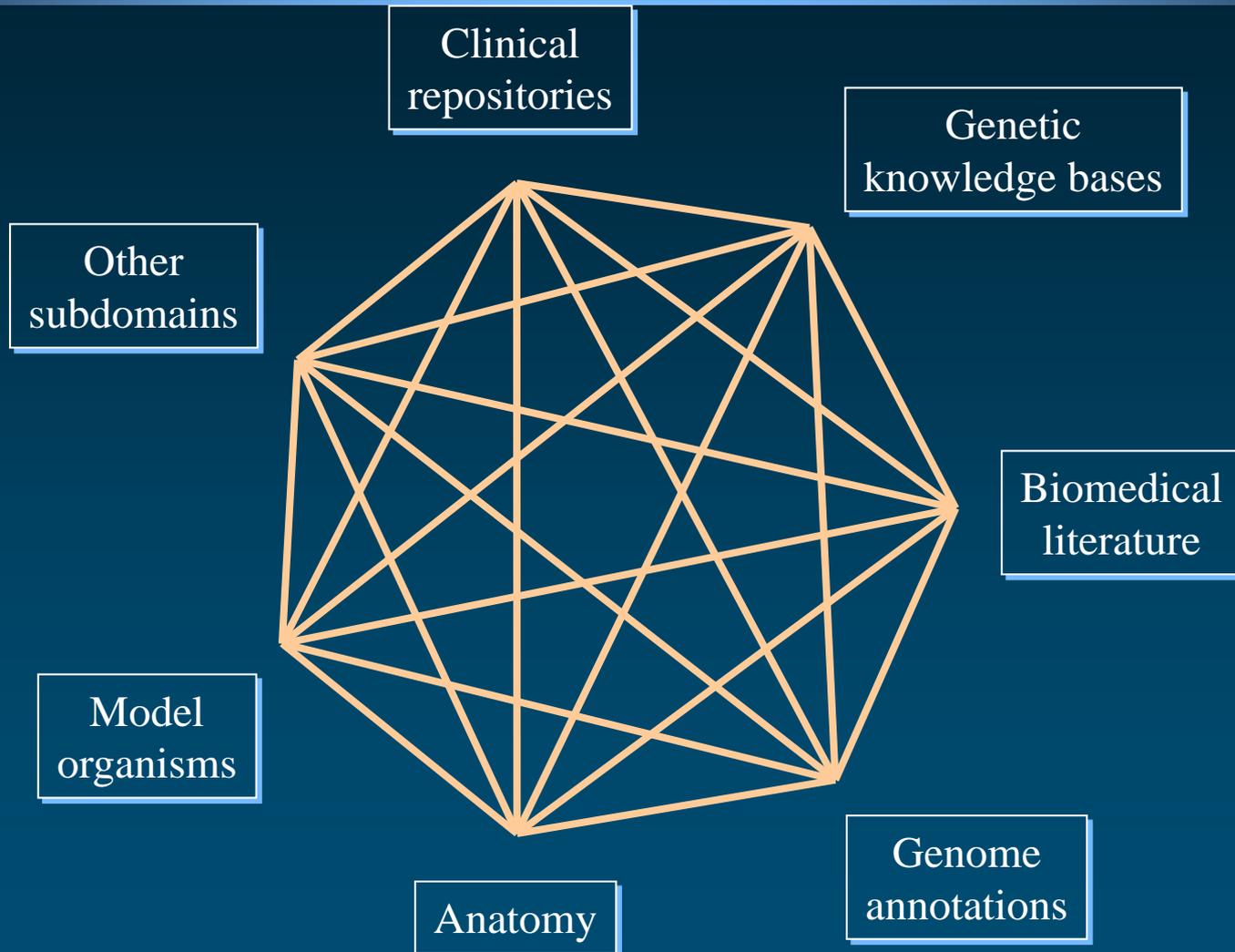
# Terminology integration systems

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

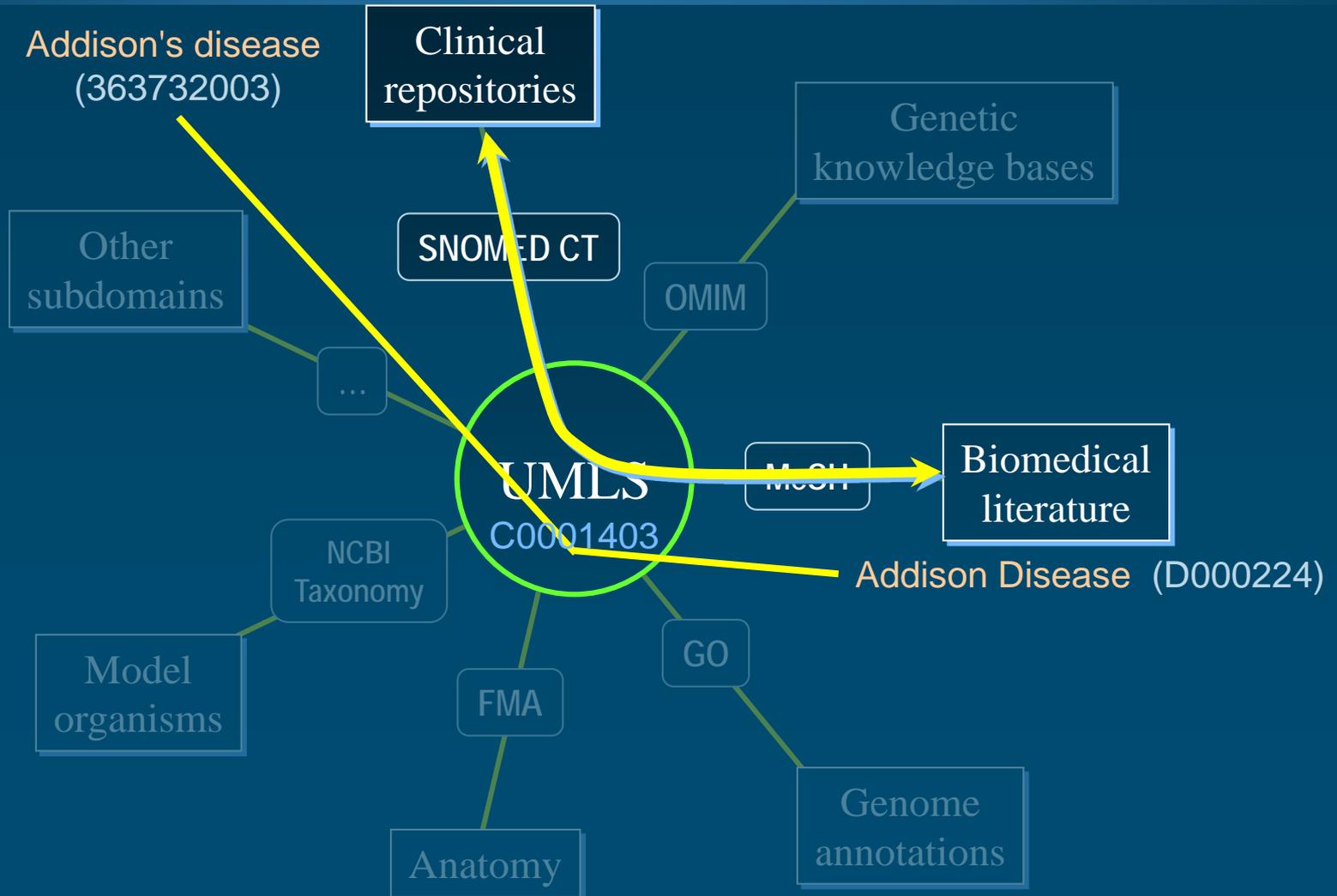
# Integrating subdomains



# Integrating subdomains



# Trans-namespace integration



# UMLS Source Vocabularies

(2015AB)

- ◆ 153 families of source vocabularies
  - Not counting translations
- ◆ 25 languages
- ◆ Broad coverage of biomedicine
  - 9.8M names (normalized)
  - 3.2M concepts
  - ~13M relations among concepts
- ◆ Common presentation



# Metathesaurus Basic organization

## ◆ Concepts

- Synonymous terms are clustered into a concept
- Properties are attached to concepts, e.g.,
  - Unique identifier
  - Definition

## ◆ Relations

- Concepts are related to other concepts
- Properties are attached to relations, e.g.,
  - Type of relationship
  - Source



# Decision support and analytics

# 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)

# *Controlled Vocabularies*

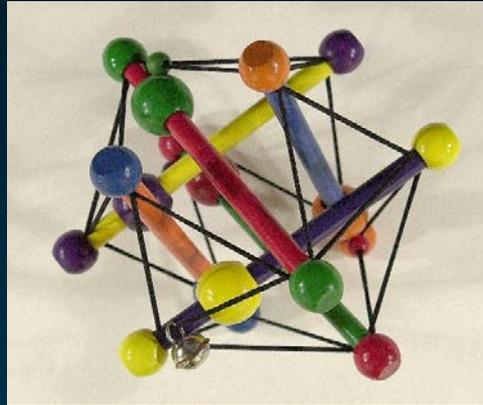
## Summary

# Summary

- ◆ History of biomedical ontologies
- ◆ How clinical features are reflected in disease names
- ◆ Structure of the main clinical ontologies used
  - ICD, SNOMED CT, LOINC, RxNorm
- ◆ Purpose of biomedical ontologies
  - Knowledge management, [health information exchange and semantic interoperability], and clinical decision support and analytics

# Topics not discussed

- ◆ Semantic Web, URIs, Linked Data
- ◆ Ontology creation, Protege
- ◆ Accessing terminology resources (APIs)
- ◆ Ontology repositories
  - [UMLS], NCBO BioPortal, EBI Ontology Lookup Service
- ◆ NLP, named entity recognition, MetaMap
- ◆ Mapping local terms to standard terminologies
- ◆ VSAC, value sets, common data elements
- ◆ OBO ontologies, OBO Foundry
- ◆ Coordinated development of ontologies, harmonization
- ◆ Boundary between terminology and information model
- ◆ [...]



# Medical Ontology Research

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

Web: <https://mor.nlm.nih.gov>



*Olivier Bodenreider*

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



U.S. National Library of Medicine

