

# Public Health Informatics Fellowship Program

## January 16, 2015

# Ontologies and Public Health



*Olivier Bodenreider*

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



U.S. National Library of Medicine



# 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	Hanged & made away themselves	7	Leprotic	14
Colicure	3	Head smote & Moxle fallen	14	Scurvy	109
Cancer, Gangrene and Fillula	56	jaundies	120	Shingles and Swine pox	2
Canker, and Thrush	121	Imposume	227	Sores, Ulcers, broken and healed	82
Childbed	623	Kill'd by severall accidentes	46	Lambs	82
Christomes and Infants	1258	Sings 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	332
Consumption and Tiflick	4808	Liverg-town	21	Stone and Stranguy	28
Convulsion and Morice	1052	Meagrom and Headach	1	Suckt	1100
Distacted	3	Mealles	7	Teeth and Worms	1014
Droove and Terpany	1476	Mothered and Shot	9	Worming	51
Drunkard	3	Overjaud & Starved	45	Vunn	7
Chilmes	5114				
Chilmes of Females	4853				
Chilmes of Males	9567				
		Buried			
		Males	48569		
		Females	48537	Of the Plague	68598
		In all	97106		
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

# Need for ontologies

“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.

# Outline

- ◆ Biomedical ontologies – What they are
- ◆ Biomedical ontologies – What they are for  
(Relevance to public health)
- ◆ Related artifacts
  - Value sets
  - Common data elements
  - Terminology integration systems



# Biomedical ontologies

*What they are*

# Overview

## ◆ Structural perspective

[J. Cimino, YBMI 2006]

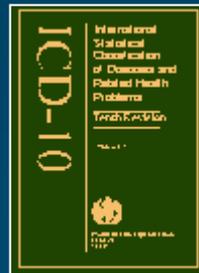
- What are they (vs. what are they for)?

## ◆ “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, NCBO BioPortal
- ◆ URL: <http://www.who.int/classifications/icd/en/>

# ICD Characteristics (2)

- ◆ Number of
  - Concepts: ~12,000
  - 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

<b>Chapter</b>	<b>Blocks</b>	<b>Title</b>
<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: 2.50 (Dec. 2014)
- ◆ Type: Controlled terminology\*
- ◆ Domain: Laboratory and clinical observations
- ◆ Developer: Regenstrief Institute
- ◆ Funding: NLM
- ◆ Availability
  - Publicly available: Yes
  - Repositories: UMLS, NCBO BioPortal
- ◆ URL: [www.regenstrief.org/loinc/loinc.htm](http://www.regenstrief.org/loinc/loinc.htm)



# LOINC Characteristics (2)

- ◆ Number of
  - Concepts: ~70,000 active codes
  - Terms: 1 per concept (“long common name”)
- ◆ 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: (2 annual releases)
  - International release: July 31, 2014
  - U.S. edition: September 2014
- ◆ Type: Reference terminology / ontology
- ◆ Domain: Clinical medicine
- ◆ Developer: IHTSDO
- ◆ Funding: IHTSDO
- ◆ Availability
  - Publicly available: Yes\* (in member countries)
  - Repositories: UMLS, NCBO BioPortal
- ◆ URL: <http://www.ihtsdo.org/>



# SNOMED CT Characteristics (2)

- ◆ Number of
  - Concepts: ~300,000 active concepts
  - Terms: ~1.2M active “descriptions”
- ◆ Major organizing principles:
  - Utility for clinical medicine (e.g., definitional knowledge)
  - Model of meaning (incomplete)
  - Rich set of associative relationships
  - Small proportion of defined concepts (many primitives)
- ◆ Formalism: Description logics



# 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 Terminology

Parents

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

☰ Appendectomy (procedure) ★

SCTID: 80146002

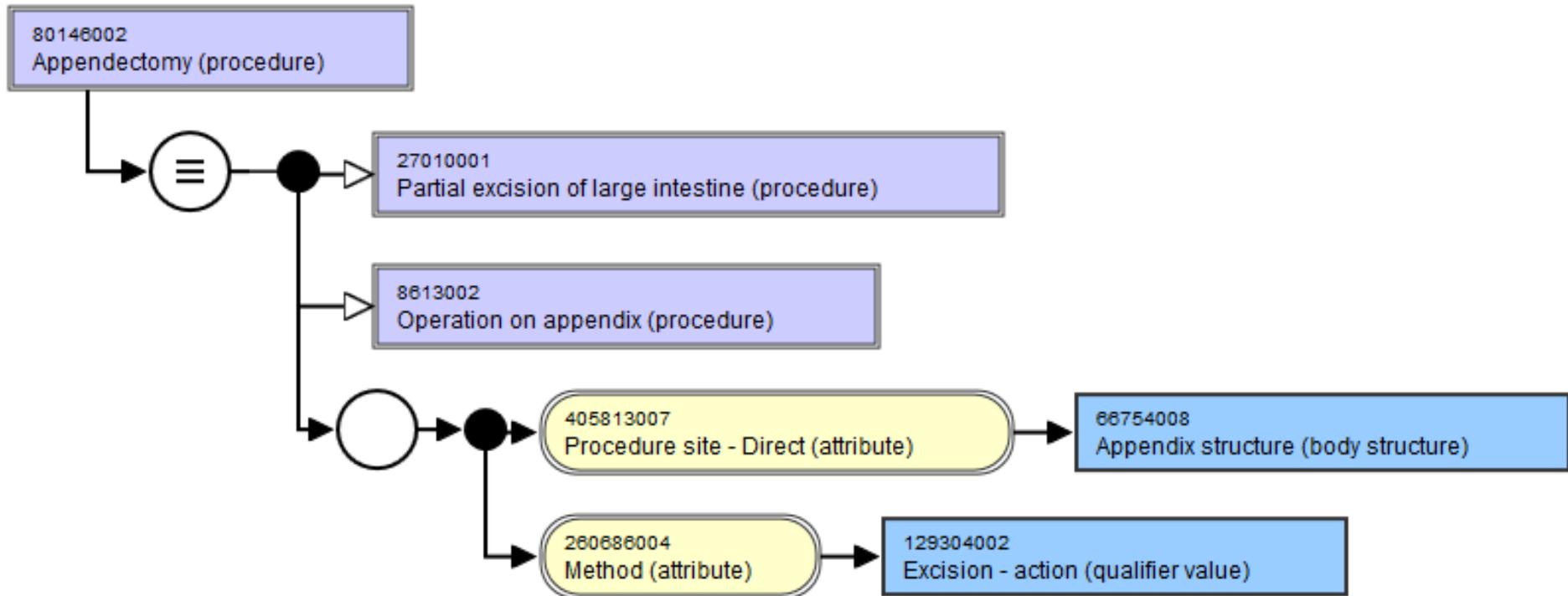
- Appendectomy (procedure)
- Appendectomy
- Appendicectomy
- Excision of appendix

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 Logical definitions



RxNorm

# RxNorm Characteristics (1)

- ◆ Current version: Jan. 2015 (monthly releases)
- ◆ Type: Controlled terminology
- ◆ Domain: Drug names
- ◆ Developer: NLM
- ◆ Funding: NLM
- ◆ Availability
  - Publicly available: Yes\*
  - Repositories: UMLS, NCBO BioPortal
- ◆ URL: <http://www.nlm.nih.gov/research/umls/rxnorm/>

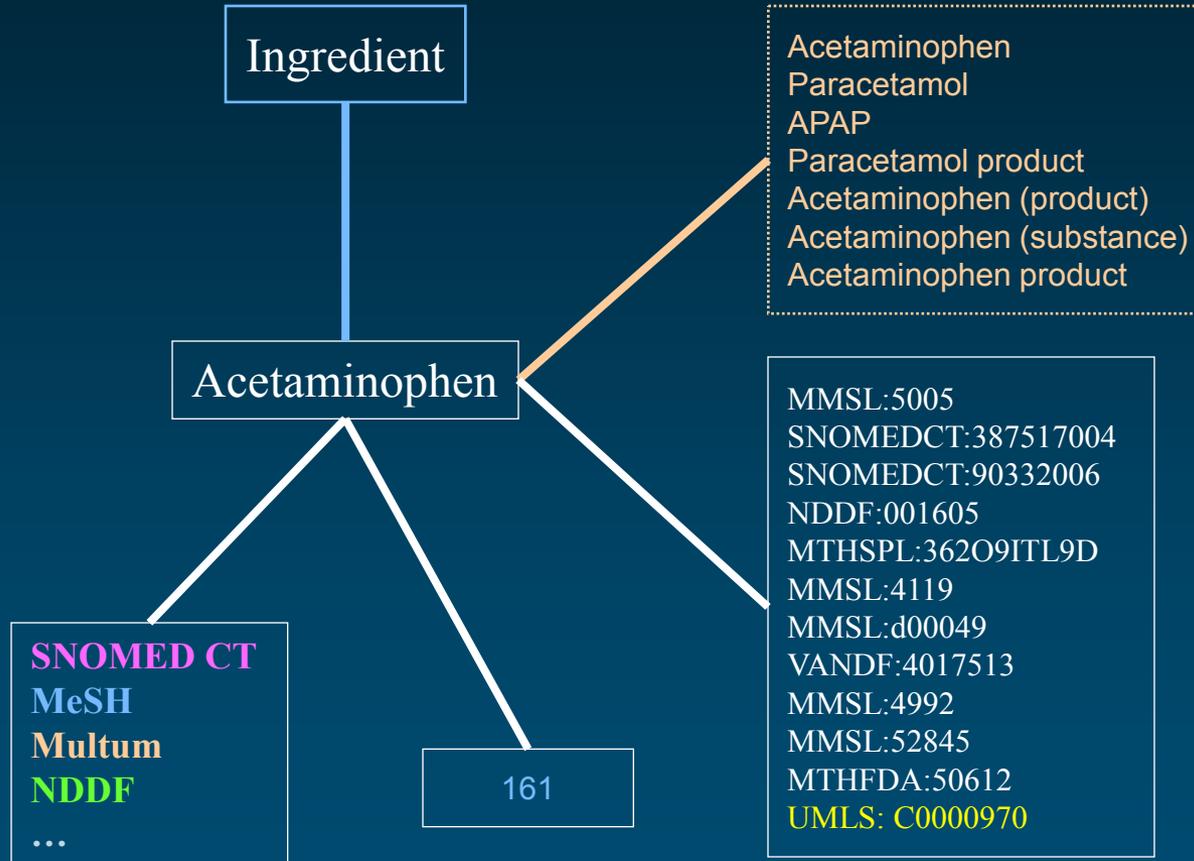


# RxNorm Characteristics (2)

- ◆ Number of
  - Concepts: ~120k
  - Terms: (limited number of synonyms)
- ◆ Major organizing principles:
  - Generic vs. brand
  - Combinations of Ingredient / Form / Strength
  - No hierarchical structure
  - Links to all major US drug information sources
  - No clinical information or drug classes\*
- ◆ Formalism: UMLS RRF format



# RxNorm Concept



# RxNorm Normalized form

**Strength**

4mg/ml

**Ingredient**

Fluoxetine

**Dose form**

Oral Solution

**Strength**

Clinical drug component

**Ingredient**

**Ingredient**

Clinical drug form

**Dose form**

**Strength**

**Ingredient**

Clinical drug

**Dose form**



# Rx Norm Generic vs. Brand

## ◆ Generic

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

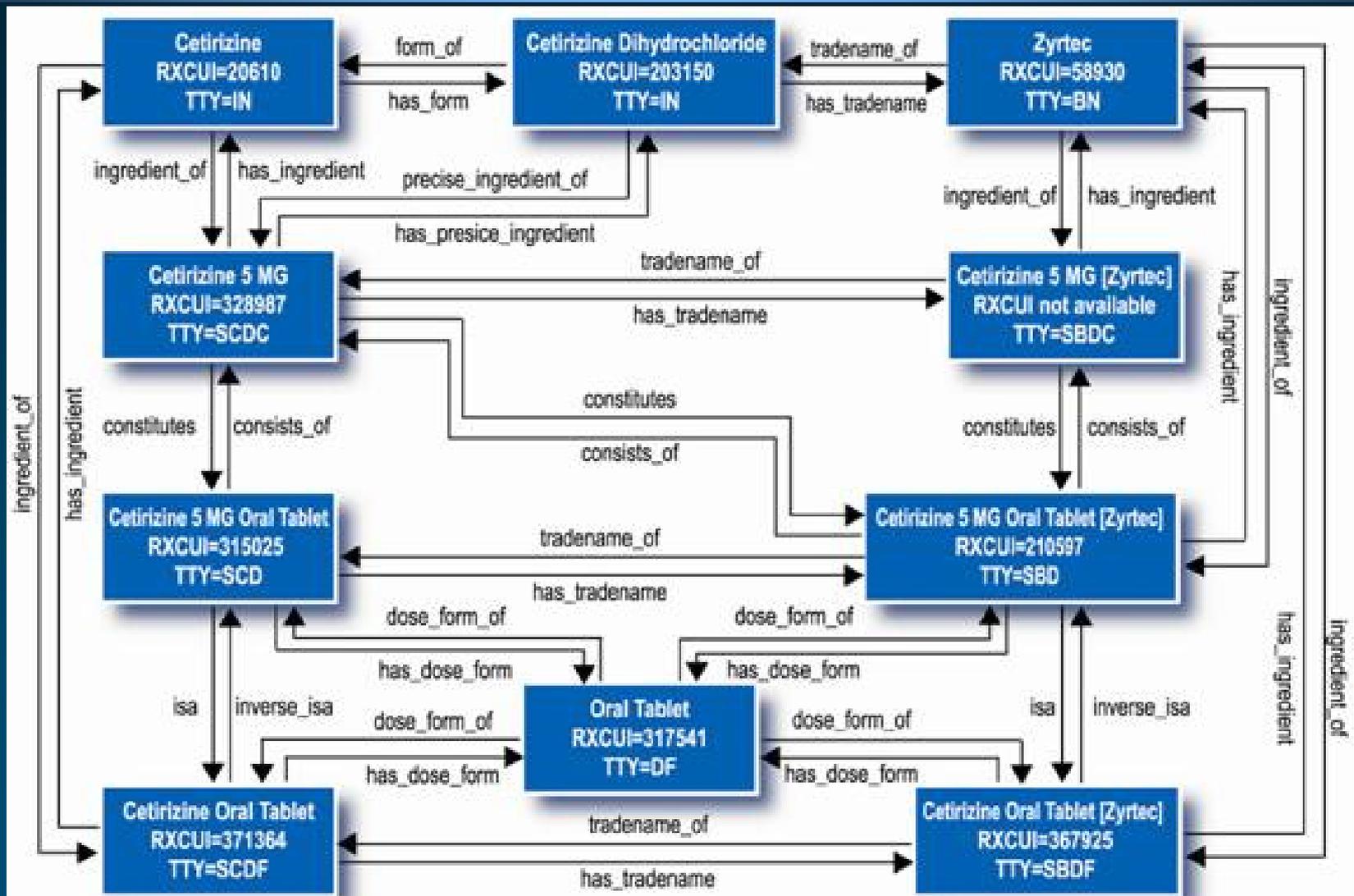
## ◆ Brand

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

*tradename\_of*



# RxNorm Relations among drug entities



# Recap (as of 2008)

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)

# Biomedical ontologies

*What they are for*  
*Relevance to public health*

# 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)

# Needs for public health

- ◆ Support for text mining
- ◆ Controlled vocabulary
- ◆ Aggregation
- ◆ Data integration
- ◆ Reasoning

# Support for text mining

- ◆ Lexical resources
  - Identify mentions in text
    - Lexical variants
- ◆ Terminological resources
  - Identify concepts
    - Synonyms
- ◆ Ontological resources
  - Identify relations, Semantic interpretation
    - Domain knowledge

# Controlled vocabulary

## ◆ Coded information

- Storage
- Processing

## ◆ Standardize

- Definitions
- Usage

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

*The Diseases and Casualties this year.*

Abortive and Stillborne	517	Executed	31	Palfie	30
Aged	1545	Flux and Small Pox	655	Plague	68576
Ajute and Peaver	5257	Found dead in Streets, fields, &c.	20	Plinthe	6
Appoplex and Suddenly	116	French Pox	86	Plurisie	19
Bedrie	12	Frighted	23	Posthead	1
Blasid	5	Gout and Sciatica	27	Quinse	15
Bleeding	16	Grief	26	Rickets	15
Bloody Flux, Scouring & Flux	185	Gripping in the Guts	1238	Riting of the Lightes	157
Burnt and Scalded	8	Hanged & made away themselves	7	Rupture	34
Colicure	3	Headmouldshot & Muskefallen	14	Scurvy	107
Cancer, Gangrene and Fiftula	56	Jaundies	120	Shingles and Swize pox	2
Canker, and Thrush	127	Imposhume	227	Sinco, Ulcers, broken and heilled	2
Childbed	625	Kill'd by severall accidents	46	Limbs	22
Cholikes and Infants	1258	Sings Evill	86	Spleen	14
Cold and Cough	65	Leproute	2	Spotted Fever and Purple	1529
Collick and Winde	124	Lechargy	14	Stopp'g in the stomack	312
Consumption and Tibbick	4808	Livergrowne	21	Stone and Stranguy	28
Convulsion and Merick	1652	Meagrom and Headach	11	Sucket	221
Distracted	1	Mealles	7	Teeth and Wornes	1614
Drownd and Tanquy	1478	Murthered and Shot	9	Urnning	51
Drownd	50	Overjaed & Starved	45	Vvann	7
♂ Males	5114	♂ Males	4858	♀ Of the Plague	68195
♀ Females	4853	♀ Females	48717		
In all	9967	In all	9729		
Increased in the Burials in the 130 Parishes and at the Pest-houses this year	79002				
Decreased of the Plague in the 130 Parishes and at the Pest-houses this year	88570				

# Data aggregation

- ◆ Granularity mismatch
  - Data recorded
  - Data needed for making decisions
- ◆ Aggregation along hierarchies
  - Subsumption hierarchies (isa)
  - Ad hoc linearizations (e.g., ICD for mortality / morbidity)

# Data integration

- ◆ Datasets annotated in reference to multiple ontologies
- ◆ Establish correspondence between equivalent concepts across ontologies (and datasets)
- ◆ Role of terminology integration systems
  - UMLS, RxNorm
  - NCBO ontology services

# Reasoning

- ◆ Description Logics
- ◆ Reasoning services (DL classifiers)
  - Instance classification

# Related artifacts

*(1) Value sets*

# Value sets

- ◆ Sets of standard codes from a given code system used for cohort identification
- ◆ Use cases
  - Clinical quality measures
- ◆ Repositories
  - NLM Value Set Authority Center
    - <https://vsac.nlm.nih.gov/>
  - CDC PHIN VADS
    - [https://phinvads.cdc.gov/vads/BrowseValueSets\\_browse.action](https://phinvads.cdc.gov/vads/BrowseValueSets_browse.action)



Welcome Search Value Sets Download Help

Search the NLM Value Set Repository

Query: gestational diabetes Search

Apply Filters Clear Filters

Narrow search results by selecting from pull-down menus below:

CMS eMeasure (NQF Number)

Select

Quality Data Model Category

Select

Steward

Select

Meaningful Use Measures

Select

Code System

Select

Search Results Value Set Details

Value Set Information Available Updates: Approved By Steward Versions: MU2 EP Update 2013-06-14 Export Value Set Results (Excel)

<b>Metadata</b>	<b>Name:</b> Gestational Diabetes	<b>OID:</b> 2.16.840.1.113883.3.464.1003.103.11.1012
<b>Member Of</b>	<b>Type:</b> Extensional	<b>Definition ID:</b> 20130614
	<b>Steward:</b> National Committee for Quality Assurance	<b>Program:</b> CMS, MU2 EP Update 2013-06-14 using this value set

Value Set Members Expansion ID: 20121025

Expanded Code List

View Toggle Clear Page 1 of 1 20 View 1 - 11 of 11

Code	Descriptor	Code System	Version	Code System OID
<a href="#">11687002</a>	Gestational diabetes mellitus (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">420491007</a>	Gestational diabetes mellitus, class H (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">420738003</a>	Gestational diabetes mellitus, class T (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">420989005</a>	Gestational diabetes mellitus, class R (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">421223006</a>	Gestational diabetes mellitus, class F (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">421389009</a>	Gestational diabetes mellitus, class C (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">421443003</a>	Gestational diabetes mellitus, class D (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">422155003</a>	Gestational diabetes mellitus, class B (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">46894009</a>	Gestational diabetes mellitus, class A>2< (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">71546005</a>	Gestational diabetes mellitus, class B>1< (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9
<a href="#">75022004</a>	Gestational diabetes mellitus, class A>1< (disorder)	SNOMEDCT	2013-01	2.16.840.1.113883.6.9

View Page 1 of 1 20 View 1 - 11 of 11

[Search Home](#)[Search Results](#)[RCMT](#)[Quick Search](#)[Subscribe](#)

VADS Navigation Menu

### Value Set Information

**Value Set Code** PHVS\_GestationalDiabetes\_NCHS

**Value Set Name** Gestational Diabetes (NCHS)

**Value Set OID** 1.3.6.1.4.1.19376.1.7.3.1.1.13.8.137

**Value Set Description** To Reflect Risk Factors of Gestational Diabetes

[Download Value Set](#)[Subscribe](#)

### Version History

◀ Version 1 ▶

(Current)

[Value Set Concepts](#) | [Value Set Details](#)5 Value Set Concepts found ◀◀ 1 ▶▶[Select All](#)[Clear All](#)[Download Selected](#)[Download All](#)

Concept Code	Sequence	Concept Name	Preferred Concept Name	Code System	Value Set	
<input type="checkbox"/> 11687002	1	Gestational diabetes mellitus	Gestational diabetes mellitus	SNOMED-CT	Gestational Diabetes (NCHS)	<a href="#">Details</a>
<input type="checkbox"/> 46894009	2	Gestational diabetes mellitus- class A>2<	Gestational diabetes mellitus- class A>2<	SNOMED-CT	Gestational Diabetes (NCHS)	<a href="#">Details</a>
<input type="checkbox"/> 71546005	3	Gestational diabetes mellitus- class B>1<	Gestational diabetes mellitus- class B>1<	SNOMED-CT	Gestational Diabetes (NCHS)	<a href="#">Details</a>
<input type="checkbox"/> 75022004	4	Gestational diabetes mellitus- class A>1<	Gestational diabetes mellitus- class A>1<	SNOMED-CT	Gestational Diabetes (NCHS)	<a href="#">Details</a>
<input type="checkbox"/> 76751001	5	diabetes mellitus in mother complicating pregnancy, childbirth AND/OR puerperium	diabetes mellitus in mother complicating pregnancy, childbirth AND/OR puerperium	SNOMED-CT	Gestational Diabetes (NCHS)	<a href="#">Details</a>

# Related artifacts

*(2) Common data elements*

# Common data elements

- ◆ Information model vs. terminology
  - Binding between the two
- ◆ Metadata for biomedical datasets
  - Standard annotations for data integration
- ◆ Current efforts at NIH
  - <http://cde.nih.gov/>
  - Big Data to Knowledge (BD2K)



# Common data elements at NIH



- Databases
- Find, Read, Learn
- Explore NLM
- Research at NLM
- NLM for You

Contact NLM



## NIH Common Data Element (CDE) Resource Portal

[Home](#) | [Resource Summaries](#) | [Glossary](#)

Home

### Home

NIH encourages the use of common data elements (CDEs) in clinical research, patient registries, and other human subject research in order to improve data quality and opportunities for comparison and combination of data from multiple studies and with electronic health records. This portal provides access to NIH-supported CDE initiatives and other tools and resources that can assist investigators developing protocols for data collection. [What is a CDE?](#)

### NIH CDE Initiatives

Collections of CDEs that have been identified for use in particular NIH-supported research projects or registries after a formal evaluation and selection processes.

### NIH CDE Tools and Resources

Databases and repositories of data elements and case report forms that may assist investigators in identifying and selecting data elements for use in their projects.

Summary Table

Subject Areas

Summary Table

Subject Areas

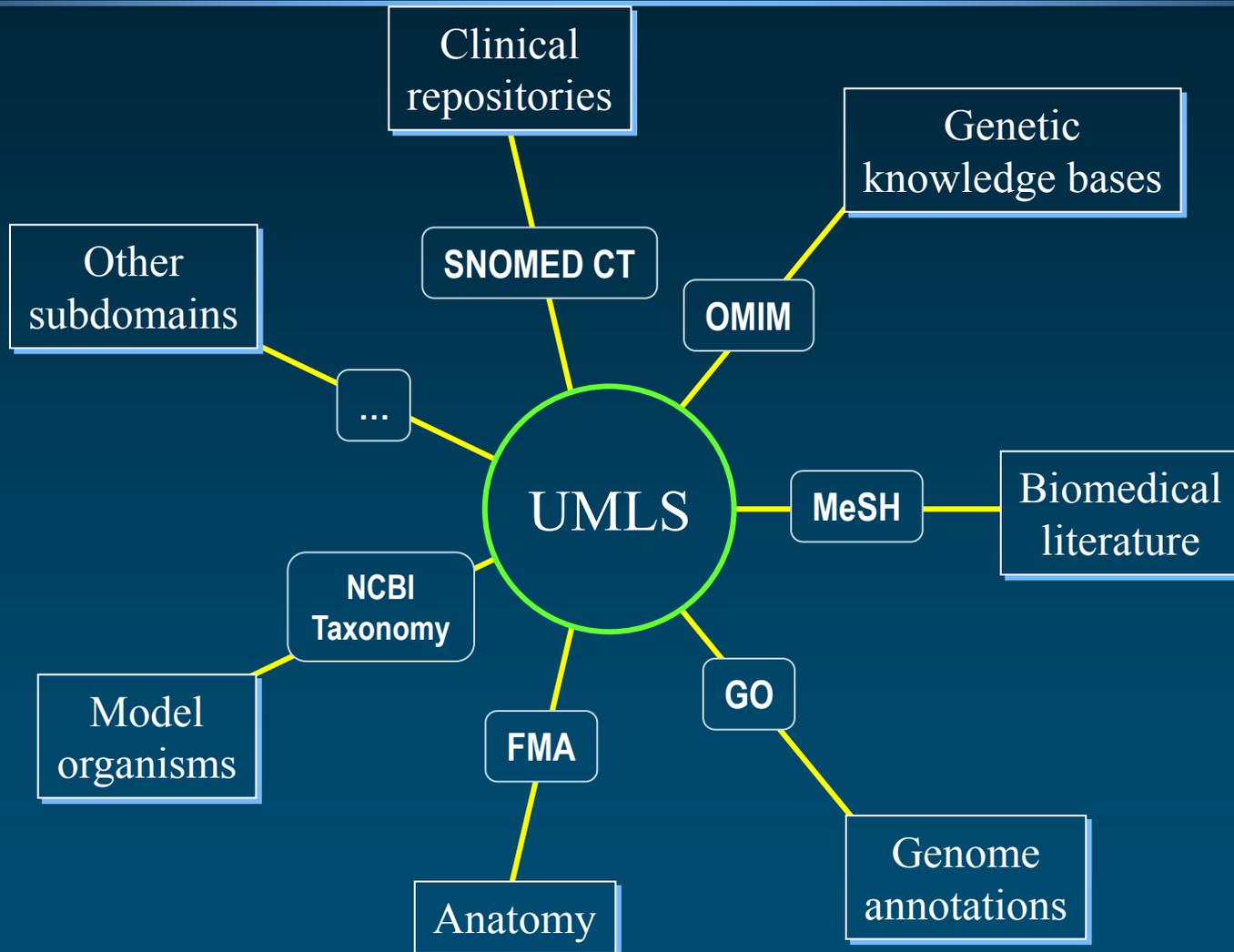
# NIH CDE Initiatives (partial list)

<a href="#">Quality of Life Outcomes in Neurological Disorders</a>	<a href="#">Neuro-QOL</a>	A core set of quality-of-life questions that address chronic neurologic disorders, plus sets of supplemental questions specific to targeted diseases or subgroups of patients. <a href="#">More...</a>	Neurological disorders. <a href="#">More...</a>	500	<a href="#">NINDS</a>
<a href="#">NIDA Substance Abuse Electronic Health Record Data Elements</a>	<a href="#">NIDA EHR</a>	A set of brief screening and initial assessment tools for substance use disorders (SUDs) for use in general medical settings. <a href="#">More...</a>	Substance Use Disorders. <a href="#">More...</a>	45	<a href="#">NIDA</a>
<a href="#">NIH Toolbox for Assessment of Neurological and Behavioral Function</a>	<a href="#">NIH Toolbox</a>	An integrated set of tools for measuring cognitive, emotional, motor and sensory function. <a href="#">More...</a>	Cognitive, emotional, motor, and sensory function. <a href="#">More...</a>	4 batteries of tests, each with 5-24 tests	<a href="#">NIH</a>
<a href="#">NINDS Common Data Elements</a>	<a href="#">NINDS CDEs</a>	A core set of data elements for use in NINDS-funded studies, including core and supplementary sets of data elements for use in disease-specific studies. <a href="#">More...</a>	Neurological disorders. <a href="#">More...</a>	7,000+ variables, 100s of instruments	<a href="#">NINDS</a>
<a href="#">Consensus Measures for Phenotypes and eXposures</a>	<a href="#">PhenX</a>	Standard measures related to complex diseases, phenotypic traits and environmental exposures for inclusion in genome-wide association studies (GWAS) and other large-scale genomic and epidemiologic research efforts. <a href="#">More...</a>	Genome-wide association studies. <a href="#">More...</a>	15,000+ variables, 428 protocols	<a href="#">NHGRI</a>
<a href="#">Patient Reported Outcomes Measurement Information System</a>	<a href="#">PROMIS</a>	A system of item banks measuring patient-reported health status for various domains of physical, mental, and social health across clinical populations (i.e. not disease-specific). <a href="#">More...</a>	Physical, mental, and social health. <a href="#">More...</a>	50 item banks	<a href="#">NIAMS</a>

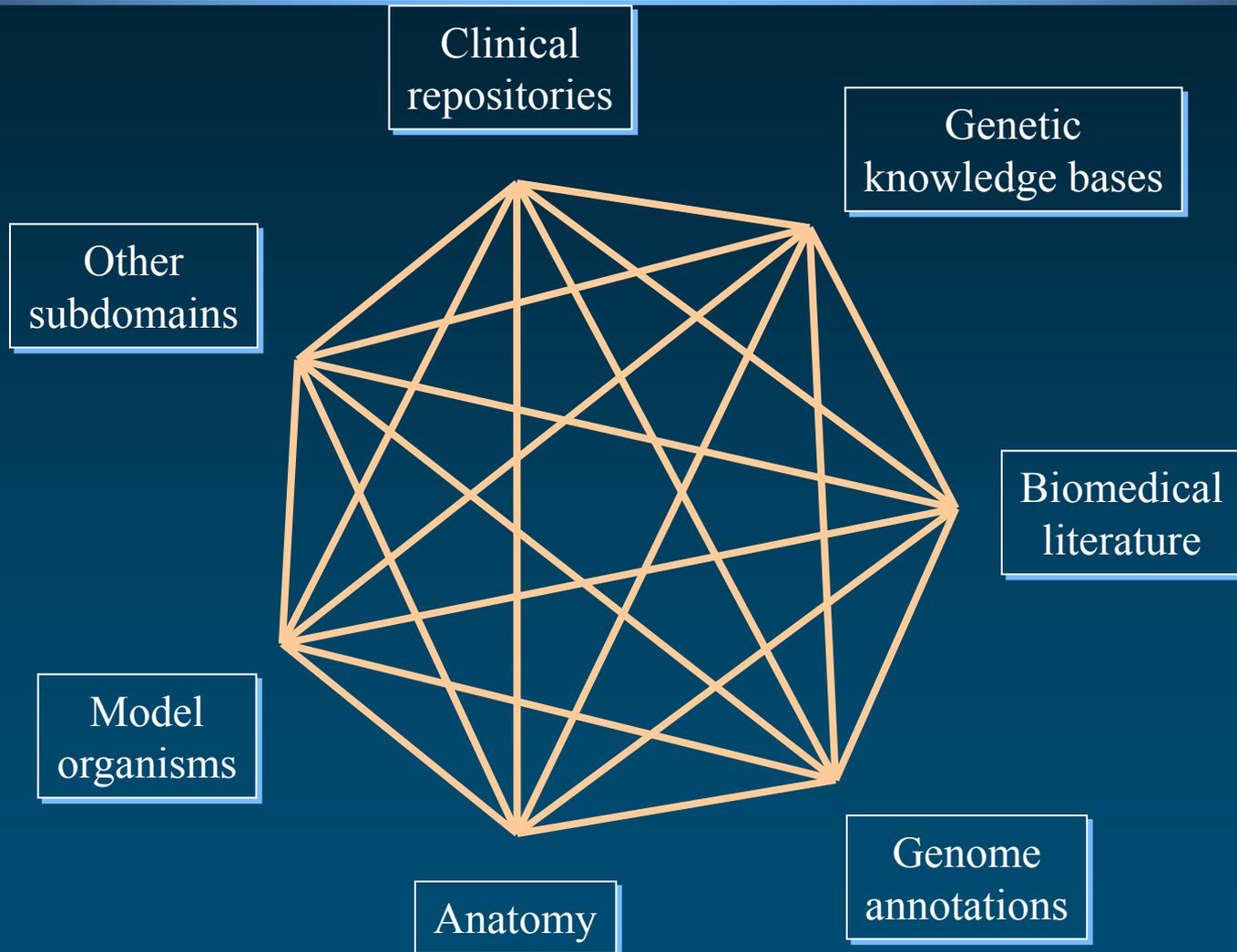
## Related artifacts

*(3) Terminology integration systems*

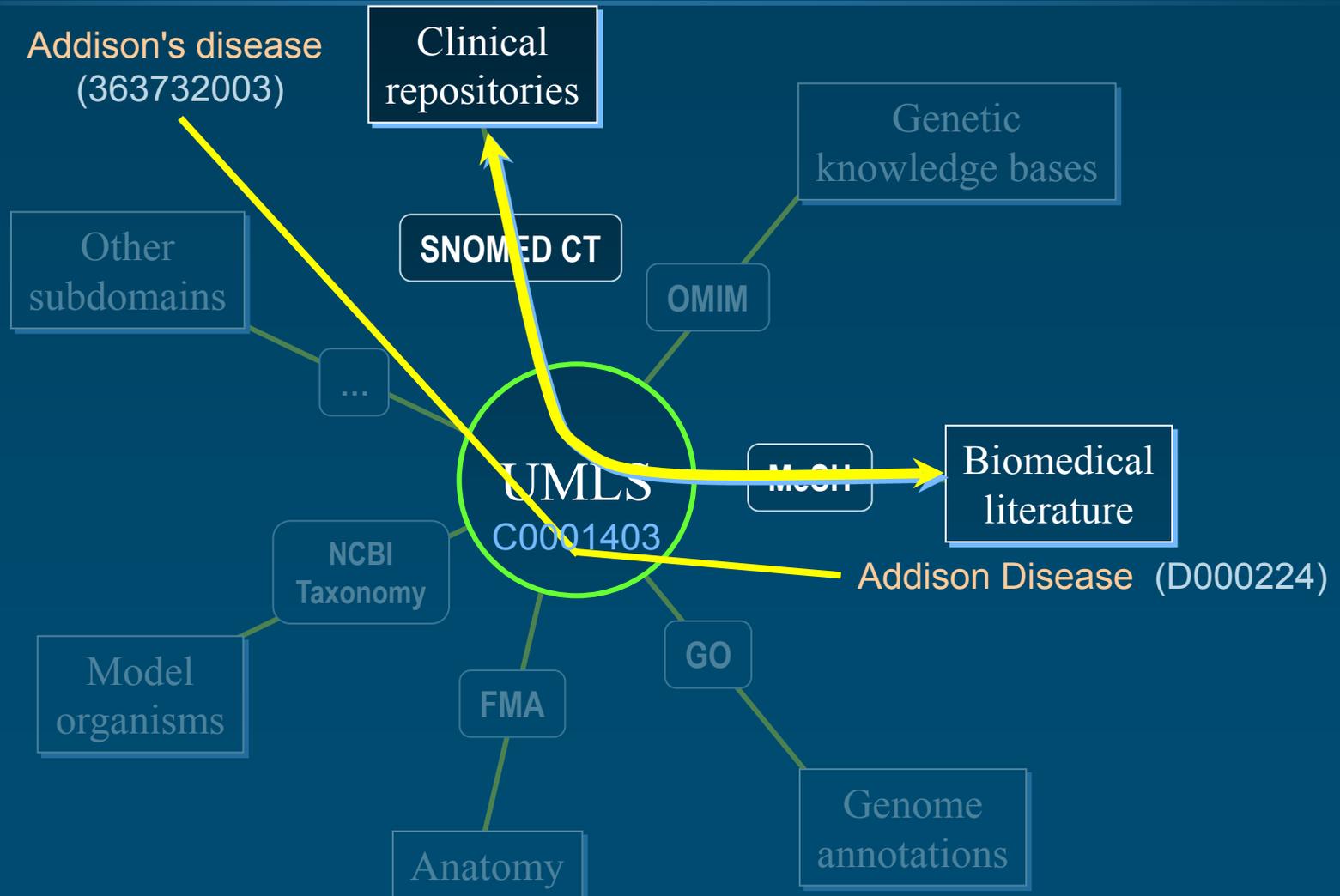
# Integrating subdomains



# Integrating subdomains



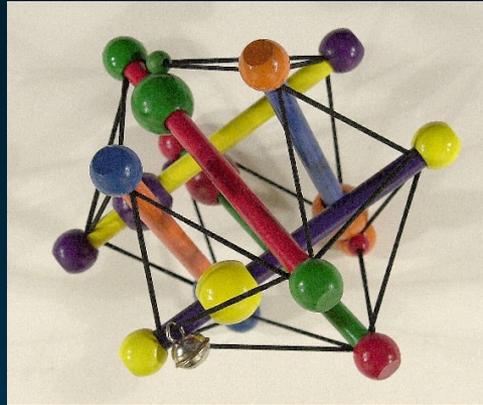
# Trans-namespace integration



# Terminology integration systems

- ◆ Unified Medical Language System (UMLS)
  - National Library of Medicine (NLM)
  - <https://uts.nlm.nih.gov/>
  
- ◆ BioPortal
  - National Center for Biomedical Ontology (NCBO)
  - <http://bioportal.bioontology.org/>





# 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

# References Ontologies

- ◆ 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.