

The Second International Symposium  
on Languages in Biology and Medicine

**LBM2007**

December 6-7, 2007



Singapore - December 6, 2007

# Terminological systems in biomedicine

*From terminology integration  
to information integration*



*Olivier Bodenreider*

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

# Outline

- ◆ Information integration in biomedicine
  - Use case: Oncology
  - Some issues: naming, normalization, mapping
- ◆ Terminology integration in biomedicine
  - Bottom-up
    - Unified Medical Language System*
  - Top-down
    - OBO Foundry ontologies*
- ◆ Applications
  - Biomedical Semantic Web*



# Information integration in oncology

# Information sources and terminologies

## ◆ Multiple terminologies for oncology

- International Classification of Diseases-Oncology (ICD-O-3)
  - Cancer registries
  - Epidemiology, Public health
- SNOMED CT
  - Patient records
  - Clinical care
- NCI Thesaurus
  - Annotation of research data



# SNOMED CT



CliniClue 2006: SNOMED CT(International 0707Int[Release]) [Registered user: olivier@nlm.nih.gov]

File Edit Subsets Restrict Language Layout Tools Help

Concept Id: 399490008 **Adenocarcinoma of prostate**  
Description Id: 1778899017 clinical finding

Words - any order  
Find: prostate adenocarcinoma  
adenocarcinoma of prostate

Hierarchy: Subtype hierarchy  
254900004 carcinoma of prostate  
423746001 adenocarcinoma of pelvis  
399490008 adenocarcinoma of prostate  
278060005 endometrioid carcinoma of prostate

adenocarcinoma of prostate - Definition  
Concept Status: **Current**  
Descriptions  
F adenocarcinoma of prostate (disorder)  
P adenocarcinoma of prostate  
Definition: Fully defined by ...  
is a  
D carcinoma of prostate  
D adenocarcinoma of pelvis  
Group  
Associated morphology  
D malignant adenomatous neoplasm - category  
finding site  
D prostatic structure  
Group  
Associated morphology  
D carcinoma  
finding site  
D prostatic structure  
Qualifiers  
episodicity  
P episodicities  
Codes  
Original SnomedId : D7-F046E  
Read Code (Ctv3Id) : XUYqi

<http://www.clinical-info.co.uk/>

# NCI Thesaurus



## Concept Details

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

### Prostate Adenocarcinoma

#### Identifiers:

name	Prostate_Adenocarcinoma
code	C2919

#### Relationships to other concepts:

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

#### Information about this concept:

##### DEFINITION

Synonym with source data

Synonym with source data

Synonym with source data

Preferred\_Name

Semantic\_Type

Synonym

Synonym

Synonym

Unified Medical Language System Concept Identifier

#### Superconcepts:

- Adenocarcinoma
- Common Carcinoma
- Invasive Prostate Carcinoma

#### Subconcepts:

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

#### Quick Search **Advanced Search**

Max Results: 25

prostate adenocarcinoma



# ICD-O-3



## ◆ Morphology

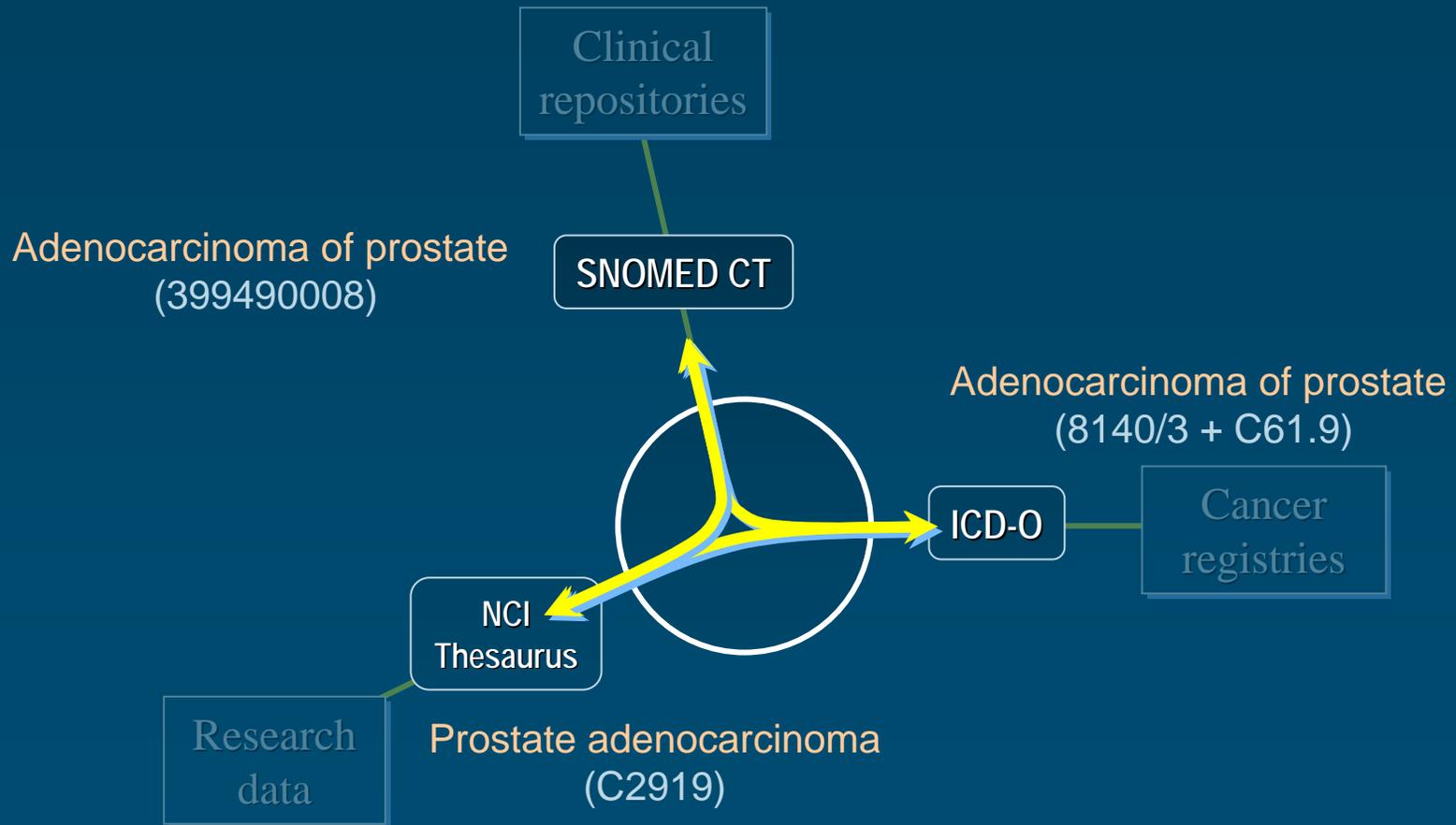
- [...]
- 814-838 Adenomas and adenocarcinomas
  - 8140/3 Adenocarcinoma, NOS

## ◆ Anatomy

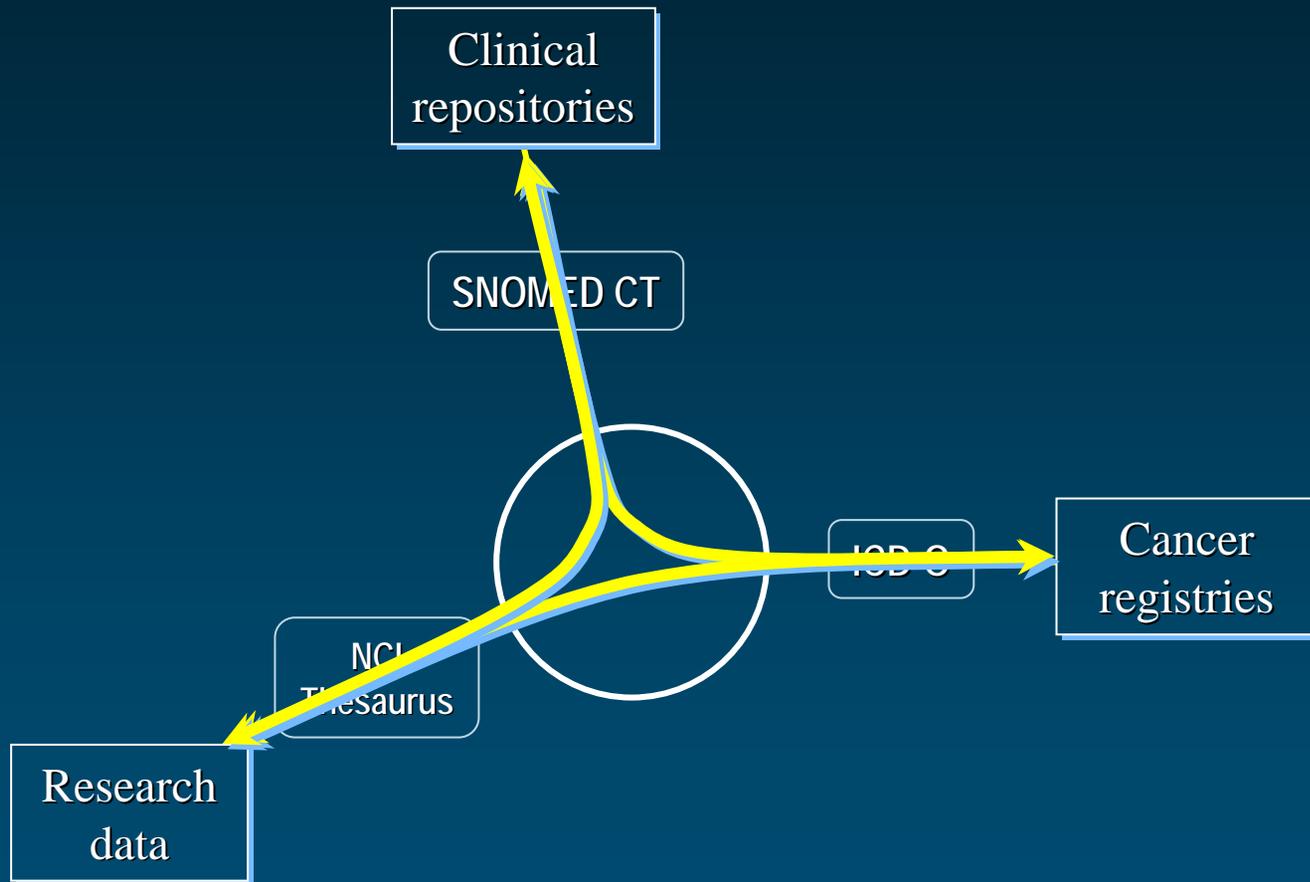
- [...]
- C60-C63 Male genital organs
  - C61 Prostate gland
    - C61.9 Prostate, NOS  
Prostate gland

Adenocarcinoma  
of prostate

# Integrating terminologies



# Integrating subdomains



# Information integration in biomedicine

*Some issues: naming, normalization, mapping*

# 1

# Naming

- ◆ Many biomedical entities have several names (synonymy)
  - Drug names
  - Gene names
  - Disease names
  - ...
- ◆ A given name may refer to several different entities (polysemy)
  - Nail (body part)
  - Nail (medical device)

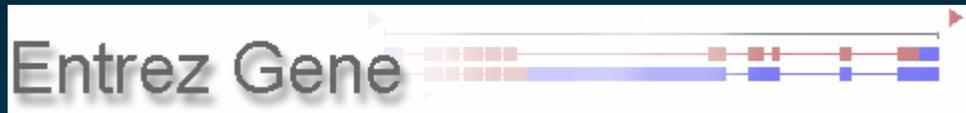
# Brand names for paracetamol (acetaminophen)

[http://en.wikipedia.org/wiki/List\\_of\\_paracetamol\\_brand\\_names](http://en.wikipedia.org/wiki/List_of_paracetamol_brand_names)

Brand name	Countries
<b>Acamol</b>	Israel
<b>Atamel</b>	Venezuela
<b>Adol</b>	Oman
<b>Aldolor</b>	Israel
<b>Alvedon</b>	Sweden
<b>APAP</b>	Poland
<b>Benuron</b>	Portugal, Germany
<b>Biogesic</b>	Philippines
<b>Buscapina</b>	Argentina
<b>Cemol</b>	Thailand
<b>Crocina</b>	India
<b>Dafalgan</b>	Belgium, France, Portugal, Russia, Ukraine
<b>Daleron</b>	Slovenia
<b>Depon</b>	Greece
<b>Dexamol</b>	Israel
<b>Dolex</b>	Colombia
<b>Doliprane</b>	France, Portugal, Russia, Ukraine
<b>Efferalgan</b>	France, Italy, Portugal, Russia, Spain, Ukraine
<b>FeverAll</b>	United States
<b>Gelocatil</b>	Spain
<b>Gripin</b>	Turkey
<b>Lekadol</b>	Croatia, Slovenia
<b>Metacin</b>	India

<b>Pamol</b>	Denmark, Finland, France
<b>Panado</b>	South Africa
<b>Panadol</b>	Australia, Azerbaijan, Central America, Egypt, Finland, Greece, Hong Kong, Hungary, Indonesia, Ireland, Kenya, Lebanon, Macedonia, Malaysia, Malta, Netherlands, New Zealand, Nigeria, Pakistan, Poland, Portugal, Romania, Russia, Saudi Arabia, Singapore, Sri Lanka, Switzerland, Taiwan, Ukraine, Estonia, United Kingdom
<b>Panamax</b>	Australia, United Kingdom
<b>Panodil</b>	Denmark, Iceland, Sweden
<b>Paracet</b>	Norway
<b>Paralen</b>	Czech Republic, Slovakia
<b>Paramed</b>	Botswana, South Africa, Zimbabwe
<b>Paramol</b>	Israel, Taiwan
<b>Perdolan</b>	Belgium
<b>Perfalgan</b>	Germany
<b>Pinex</b>	Denmark, Iceland, Norway
<b>Plicet</b>	Croatia
<b>Reliv</b>	Sweden
<b>Rokamol</b>	Israel
<b>Sara</b>	Thailand
<b>Tachipirina</b>	Italy
<b>Tylenol</b>	Brazil, Canada, Japan, South Korea, Thailand, United States
<b>Tempra</b>	Philippines

# Names for dystrophin



<http://www.ncbi.nlm.nih.gov/sites/entrez>

DMD

[Order cDNA clone](#), [Links](#)

**Official Symbol** DMD and **Name:** dystrophin (muscular dystrophy, Duchenne and Becker types) [*Homo sapiens*]

**Other Aliases:** GS1-19024.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272

**Other Designations:** Duchenne muscular dystrophy protein; dystrophin

**Chromosome:** X, **Location:** Xp21.2

**Annotation:** Chromosome X, NC\_000023.9 (33267646..31047265, complement)

**MIM:** 300377

**GeneID:** 1756



# Names for renal cell carcinoma

Details of 'clear cell carcinoma of kidney' Distributed Relationships

ConceptStatus **Current**

*Descriptions*

- F clear cell carcinoma of kidney (disorder)
- P clear cell carcinoma of kidney
- S adenocarcinoma of kidney
- S carcinoma of kidney
- S Grawitz tumor
- S renal cell adenocarcinoma
- S renal cell carcinoma

Fully defined by...

- Is a
  - malignant tumor of kidney parenchyma
  - primary malignant neoplasm of kidney
  - primary malignant neoplasm of retroperitoneum
- Group
  - Associated morphology
    - clear cell adenocarcinoma
  - Finding site
    - structure of parenchyma of kidney
- Laterality
  - side
  - side

*Qualifiers*

*Legacy codes*

- SNOMED: D7-F011C
- CTV3ID: X78Yx



renal cell adenocarcinoma

Search renal cell adenocarcinoma Words - any order Related search

Hierarchy for 'clear cell carcinoma of kidney' Subtype hierarchy

- malignant tumor of kidney parenchyma
- primary malignant neoplasm of kidney
- primary malignant neoplasm of retroperitoneum
- clear cell carcinoma of kidney

Details of 'clear cell carcinoma of kidney' Distributed Relationships

ConceptStatus **Current**

*Descriptions*

- F clear cell carcinoma of kidney (disorder)
- P clear cell carcinoma of kidney
- S adenocarcinoma of kidney
- S carcinoma of kidney
- S Grawitz tumor
- S renal cell adenocarcinoma
- S renal cell carcinoma

Fully defined by...

- Is a
  - malignant tumor of kidney parenchyma
  - primary malignant neoplasm of kidney
  - primary malignant neoplasm of retroperitoneum
- Group
  - Associated morphology
    - clear cell adenocarcinoma
  - Finding site
    - structure of parenchyma of kidney
- Laterality
  - side
  - side

*Qualifiers*

*Legacy codes*

- SNOMED: D7-F011C
- CTV3ID: X78Yx

<http://www.clinical-info.co.uk/>



# Entity recognition

- ◆ Identifying biomedical entities in text
  - Names entity recognition
  - Tagging “mentions”
  - Semantic annotation
- ◆ Supported by terminology
  - Collects the names used in the domain
  - Often incompletely
- ◆ Example: BioCreative
  - 1A – Gene name identification
  - 2GM – Gene mention tagging



## 2

# Normalization

- ◆ Biomedical entities are identified by unique identifiers in various terminology systems
- ◆ Resolve names into identifiers (in a given namespace)
- ◆ Supported (in part) by terminology resources
- ◆ Example: BioCreative
  - 1B and 2GN – Gene Normalization



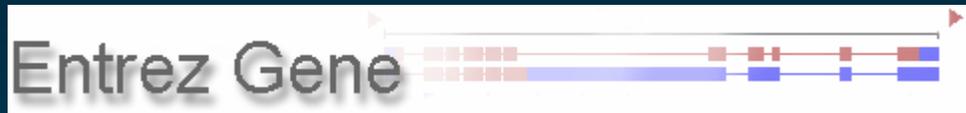
# Identifier for paracetamol (acetaminophen)

Master Drug Data Base. Medi-Span	5005	Acetaminophen
FDA National Drug Code Directory	50612	PARACETAMOL
FDA Structured Product Labels	36209ITL9D	ACETAMINOPHEN
First DataBank NDDF Plus	001605	Acetaminophen
SNOMED Clinical Terms	90332006	Acetaminophen (product)
SNOMED Clinical Terms	387517004	Acetaminophen (substance)
VA National Drug File	4017513	ACETAMINOPHEN

Source: RxNorm database (5/3/2007)



# Identifier for dystrophin



<http://www.ncbi.nlm.nih.gov/sites/entrez>

DMD

[Order cDNA clone](#), [Links](#)

**Official Symbol DMD and Name:** dystrophin (muscular dystrophy, Duchenne and Becker types) [*Homo sapiens*]

**Other Aliases:** GS1-19024.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272

**Other Designations:** Duchenne muscular dystrophy protein; dystrophin

**Chromosome:** X, **Location:** Xp21.2

**Annotation:** Chromosome X, NC\_000023.9 (33267646..31047265, complement)

**MIM:** 300377

**GeneID:** 1756



# Identifier for renal cell carcinoma

Details of 'clear cell carcinoma of kidney' Distributed Relationships

ConceptStatus **Current**

Descriptions

- F clear cell carcinoma of kidney (disorder)
- P clear cell carcinoma of kidney
- S adenocarcinoma of kidney
- S carcinoma of kidney
- S Grawitz tumor
- S renal cell adenocarcinoma
- S renal cell carcinoma

Fully defined by...

- Is a
  - maligant tumor of kidney parenchyma
  - primary malignant neoplasm of kidney
  - primary malignant neoplasm of retroperitoneum
- Group
  - Associated morphology
    - clear cell adenocarcinoma
  - Finding site
    - structure of parenchyma of kidney
- Laterality
  - side
  - side

Qualifiers

Legacy codes

- SNOMED: D7-F011C
- CTV3ID: X78Yx



renal cell adenocarcinoma

ConceptId: 254915003

Description Id: 379803017

clinical finding

renal cell adenocarcinoma

maligant tumor of kidney parenchyma

primary malignant neoplasm of kidney

primary malignant neoplasm of retroperitoneum

clear cell carcinoma of kidney

ConceptId: 254915003 renal cell adenocarcinoma

Description Id: 379803017

clinical finding

<http://www.clinical-info.co.uk/>



### 3

## Mapping / Integration

- ◆ Identify equivalent entities across systems (across namespaces)
  - Shared identifiers
  - Existing mappings (e.g., SNOMED CT to ICD-9-CM)
  - Ontology alignment techniques (lexical + structural)
- ◆ Align equivalent entities
  - Pairwise: mapping
  - More broadly: integration
- ◆ Forms the basis for information integration in the Semantic Web (mashups)

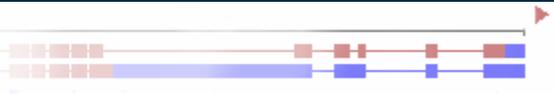
# Identifier for paracetamol (acetaminophen)

Master Drug Data Base. Medi-Span	5005	Acetaminophen
FDA National Drug Code Directory	50612	PARACETAMOL
FDA Structured Product Labels	36209ITL9D	ACETAMINOPHEN
First DataBank NDDF Plus	001605	Acetaminophen
SNOMED Clinical Terms	90332006	Acetaminophen (product)
SNOMED Clinical Terms	387517004	Acetaminophen (substance)
VA National Drug File	4017513	ACETAMINOPHEN
RxNorm	161	Acetaminophen



# Identifier for dystrophin

Entrez Gene



<http://www.ncbi.nlm.nih.gov/sites/entrez>

DMD

Order cDNA clone, Links

**Official Symbol DMD and Name:** dystrophin (muscular dystrophy, Duchenne and Becker types) [*Homo sapiens*]

**Other Aliases:** GS1-19024.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272

**Other Designations:** Duchenne muscular dystrophy protein; dystrophin

**Chromosome:** X, **Location:** Xp21.2

**Annotation:** Chromosome X, NC\_000023.9 (33267646..31047265, complement)

**MIM:** 300377

**GeneID:** 1756



# Identifier for renal cell carcinoma

Details of 'clear cell carcinoma of kidney' Distributed Relationships

ConceptStatus **Current**

Descriptions

F	clear cell carcinoma of kidney (disorder)	645875019
P	clear cell carcinoma of kidney	379798014
S	adenocarcinoma of kidney	379801015
S	carcinoma of kidney	379800019
S	Grawitz tumor	379797016
S	renal cell adenocarcinoma	379803017
S	renal cell carcinoma	379802010

Fully defined by...

- Is a
  - maligant tumor of kidney parenchyma
  - primary malignant neoplasm of kidney
  - primary malignant neoplasm of retroperitoneum
- Group
  - Associated morphology
    - clear cell adenocarcinoma
  - Finding site
    - structure of parenchyma of kidney
- Laterality
  - side
  - side

Qualifiers

Legacy codes

- SNOMED: D7-F011C
- CTV3ID: X78Yx



Details of 'renal cell adenocarcinoma' Distributed Relationships

ConceptId: 254915003

Description Id: 379803017

renal cell adenocarcinoma

clinical finding

ConceptId: 254915003

Description Id: 379803017

renal cell adenocarcinoma

clinical finding

<http://www.clinical-info.co.uk/>



# Bottom-up terminology integration

## *Unified Medical Language System*



# Motivation

- ◆ Started in 1986
- ◆ National Library of Medicine
- ◆ “Long-term R&D project”

«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- The first is **the variety of ways the same concepts are expressed** in different machine-readable sources and by different people.
- The second is the **distribution** of useful information among many disparate databases and systems.»



# Source Vocabularies

(2007AC)

- ◆ 141 source vocabularies
  - 17 languages
- ◆ Broad coverage of biomedicine
  - 6.1M names
  - 1.5M concepts
  - 8M relations
- ◆ Common presentation



# Biomedical terminologies in UMLS

## ◆ General vocabularies

- anatomy (UWDA, Neuronames)
- drugs (RxNorm, First DataBank, Micromedex, ...)
- medical devices (UMD, SPN)

## ◆ Several perspectives

- clinical terms (SNOMED CT)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- data exchange terminologies (HL7, LOINC)

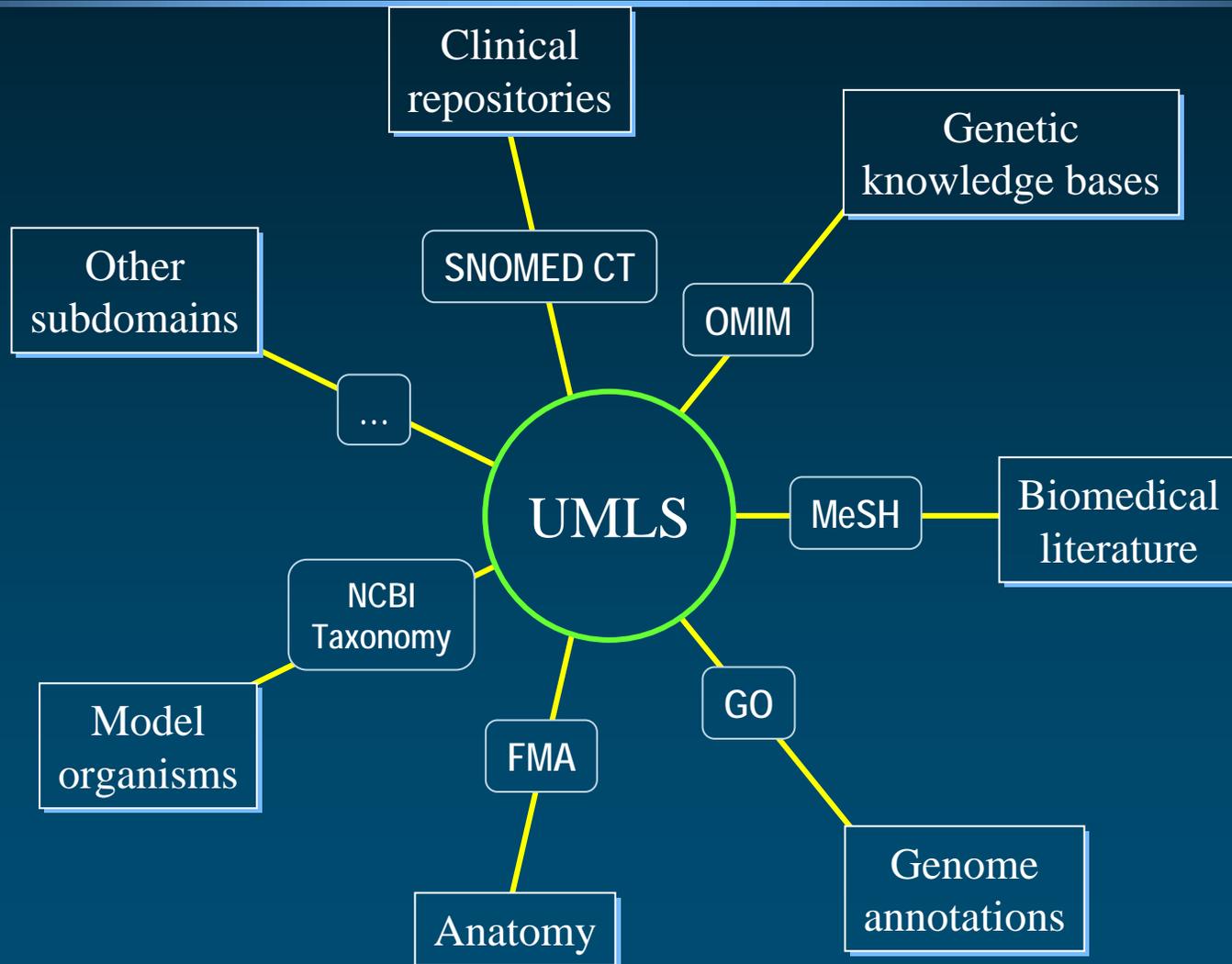


# Biomedical terminologies in UMLS

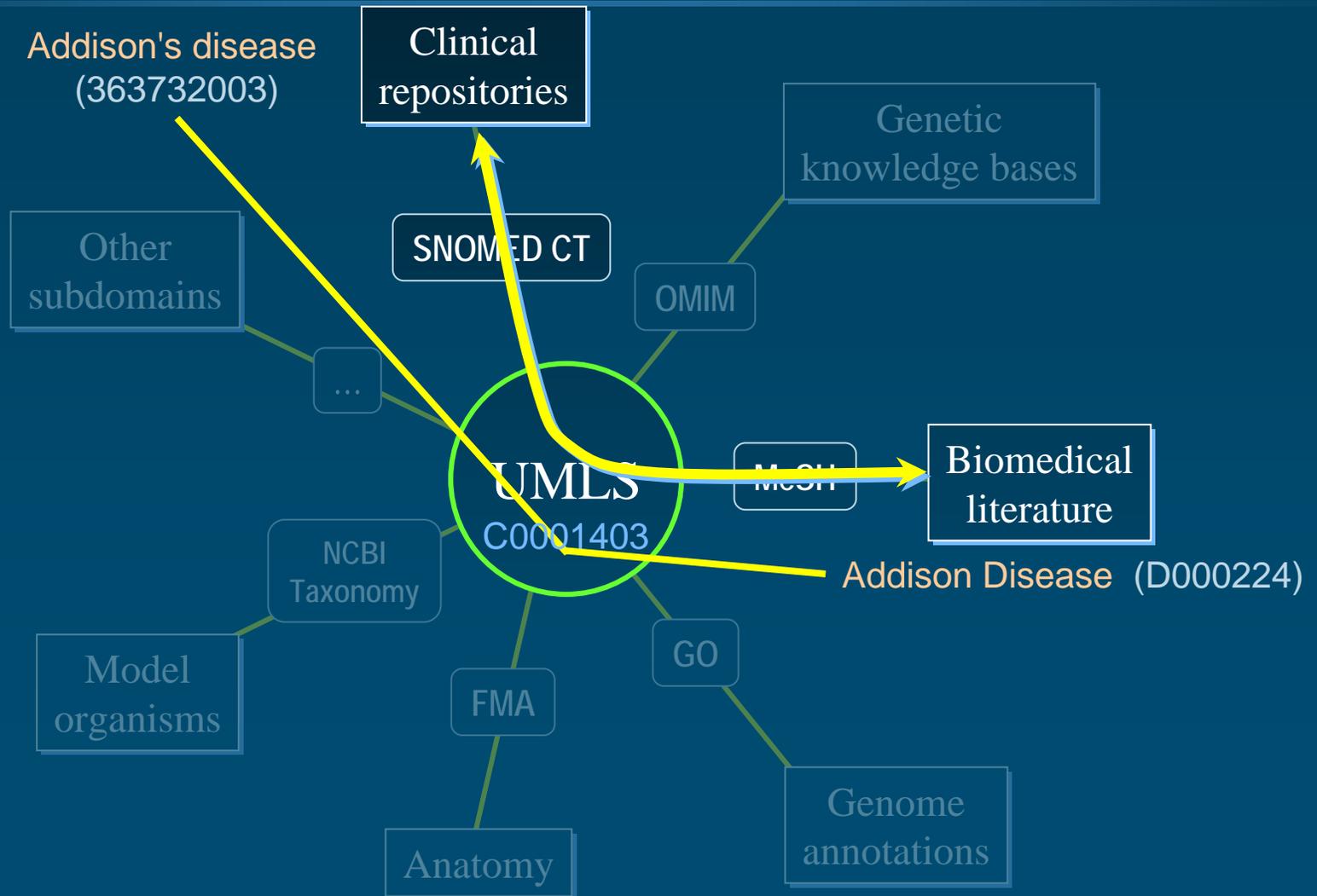
- ◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (NCI Thesaurus, PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART, MedDRA)
  - primary care (ICPC)
  - genomics (Gene Ontology, HUGO, OMIM)
- ◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)



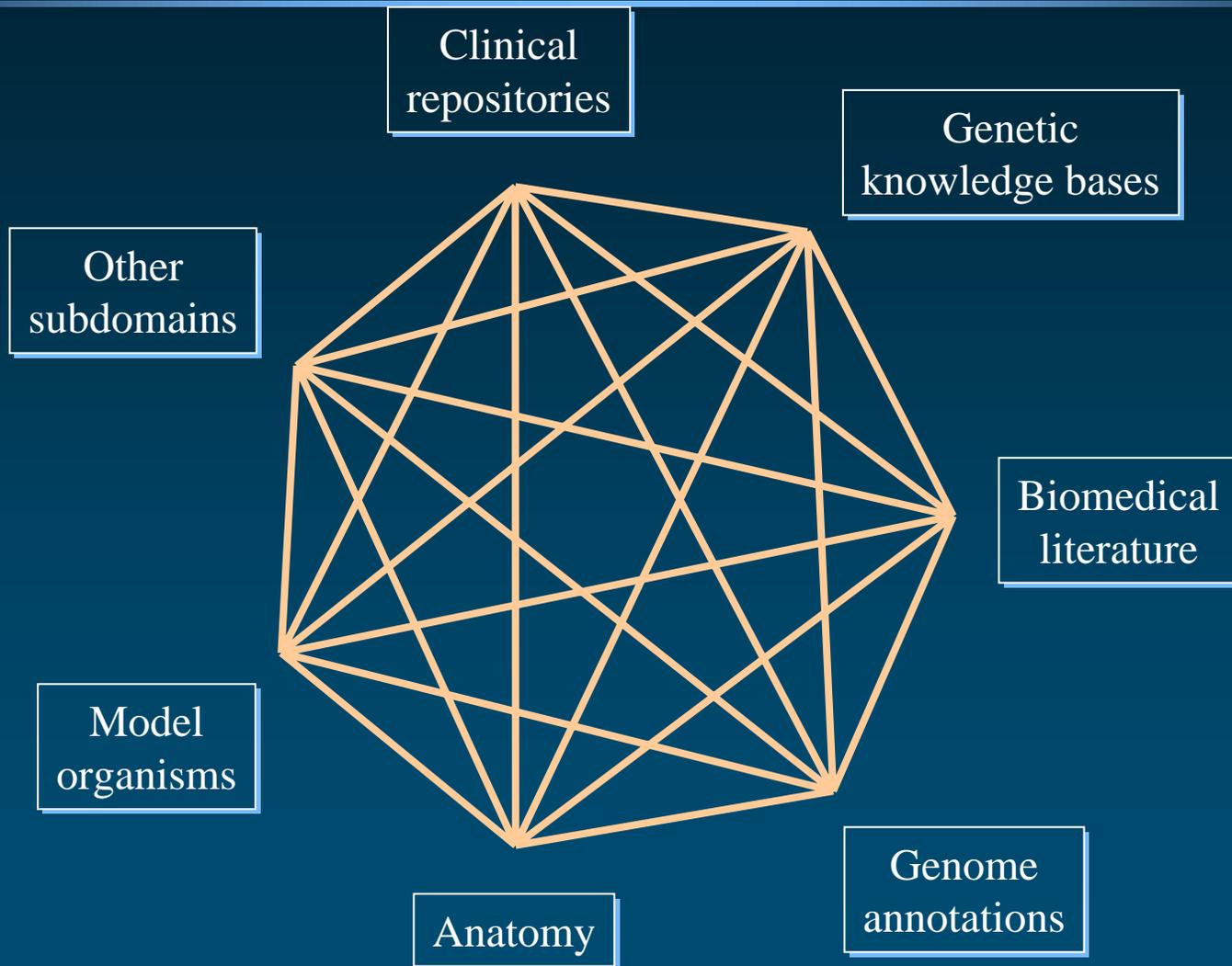
# Integrating subdomains



# Trans-namespace integration



# Integrating subdomains



# Top-down terminology integration

*OBO Foundry ontologies*

# Open Biological Ontologies



- ◆ Extended family of the Gene Ontology (GO)
- ◆ Collaborative development
  - <http://obo.sourceforge.net/>
- ◆ National Center for Biomedical Ontology
  - <http://bioontology.org/>
- ◆ OBO Foundry
  - <http://obofoundry.org/>
  - Promote best practices in ontology development
  - 10 inclusion criteria



# Open Biological Ontologies (OBO)

Domain	Prefix	File	Format
<a href="#">Biological imaging methods</a>	FBbi	<a href="#">image.obo</a>	obo
<a href="#">Biological process</a>	GO	<a href="#">gene ontology.obo</a> 	obo
<a href="#">BRENDA tissue / enzyme source</a>	BTO	<a href="#">BrendaTissue.obo</a>	obo
<a href="#">C. elegans development</a>	WBls	<a href="#">worm development.obo</a>	obo
<a href="#">C. elegans gross anatomy</a>	WBbt		obo
<a href="#">C. elegans phenotype</a>	WBPhenotype	<a href="#">phenotype ontology_obo.cgi</a>	obo
<a href="#">Cell type</a>	CL	<a href="#">cell.obo</a> 	obo
<a href="#">Cellular component</a>	GO	<a href="#">gene ontology.obo</a> 	obo
<a href="#">Cereal plant development</a>	GRO	<a href="#">cereals development.obo</a>	obo
<a href="#">Cereal plant trait</a>	TO	<a href="#">plant trait.obo</a> 	obo
<a href="#">Chemical entities of biological interest</a>	CHEBI	<a href="#">chebi.obo</a>	obo
<a href="#">Common Anatomy Reference Ontology</a>	CARO	<a href="#">caro.obo</a> 	obo
<a href="#">Dictyostelium discoideum anatomy</a>	DDANAT	<a href="#">dictyostelium anatomy.obo</a> 	obo
<a href="#">Drosophila development</a>	FBdv	<a href="#">fly development.obo</a>	obo
<a href="#">Drosophila gross anatomy</a>	FBbt	<a href="#">fly anatomy.obo</a> 	obo
<a href="#">Environment Ontology</a>	ENVO	<a href="#">envo.obo</a> 	obo

<http://obo.sourceforge.net/>



# Integrating subdomains

RELATION TO TIME GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	<b>Organism</b> (NCBI Taxonomy?)	<b>Anatomical Entity</b> (FMA, CARO)	<b>Organ Function</b> (FMP, CPRO)	<b>Phenotypic Quality</b> (PaTO)	<b>Biological Process</b> (GO)
CELL AND CELLULAR COMPONENT	<b>Cell</b> (CL)	<b>Cellular Component</b> (FMA, GO)	<b>Cellular Function</b> (GO)		
MOLECULE	<b>Molecule</b> (ChEBI, SO, RnaO, PrO)		<b>Molecular Function</b> (GO)		Molecular Process (GO)

(Barry Smith)



# OBO ontologies Examples

- ◆ Gene Ontology
- ◆ Cell types
- ◆ Sequence Ontology
- ◆ ChEBI
- ◆ Foundational Model of Anatomy
- ◆ PATO – phenotypic qualities
- ◆ Relationship types
- ◆ Ontology for Biomedical Investigations

# Applications

*Biomedical Semantic Web*

# W3C Health Care and Life Sciences IG



## W3C Semantic Web Health Care and Life Sciences Interest Group

The Semantic Web Health Care and Life Sciences Interest Group is designed to improve collaboration, research and development, and innovation adoption in the health care and life science industries. Aiding decision-making in clinical research, Semantic Web technologies will bridge many forms of biological and medical information across institutions.

**Contents:** [Mission and Scope](#) | [Membership and Joining](#) | [Charter /History](#) | [Resources](#) | [Presentations](#) | [Articles](#) | [New and Events](#) | [Conferences](#) | [Task Forces](#)

**Nearby:** [Discussion archive](#) | [HCLS WIKI](#) | [Applications and Demonstrations](#) | [OWL](#) | [RDF Data Access](#) | [Rules](#) | [Semantic Web Best Practices and Deployment](#)

### Introduction

Both Life Science Research and Health Care are areas undergoing phenomenal growth, holding much promise for our future as long as we can manage and apply the new knowledge gained without driving up costs. Key to their success is the implementation of new informatics models that will unite many forms of biological and medical information across all institutions, through the encoding of meaning into the data and their interpretations. By focusing on the semantics of information, researchers will have more access to the knowledge required to effectively find cures to diseases, while doctors will have better tools for individualized clinical management of patients.

### Mission and Scope

The Semantic Web for Health Care and Life Sciences Interest Group (HCLSIG) is chartered to develop and support the use of Semantic Web technologies and practices to improve collaboration, research and development, and innovation adoption in the of Health Care and Life Science domains. Success in these domains depends on a foundation of semantically rich system, process and information interoperability. ([more](#)).

### News and Events

- [Last Call: SPARQL Query Language for RDF 2007-03-27](#): : Comments are due by 18 April. ([Permalink](#))
- [HCLS demo](#), planned for [WWW2007 in Banff](#). To help participate in the demo, please contact [Alan Ruttenberg](#).
- [FIRST INTERNATIONAL WORKSHOP ON HEALTH CARE AND LIFE SCIENCES DATA INTEGRATION FOR THE SEMANTIC WEB](#), May 8, [WWW2007 in Banff](#).
- [Eric Prud'hommeaux](#), new W3C staff contact for HCLS.
- [GRDDL links Microformats and Semantic Web: Working Draft](#)  
`xmlns="http://www.w3.org/2000/svg"` ([Permalink](#))

<http://www.w3.org/2001/sw/hcls/>

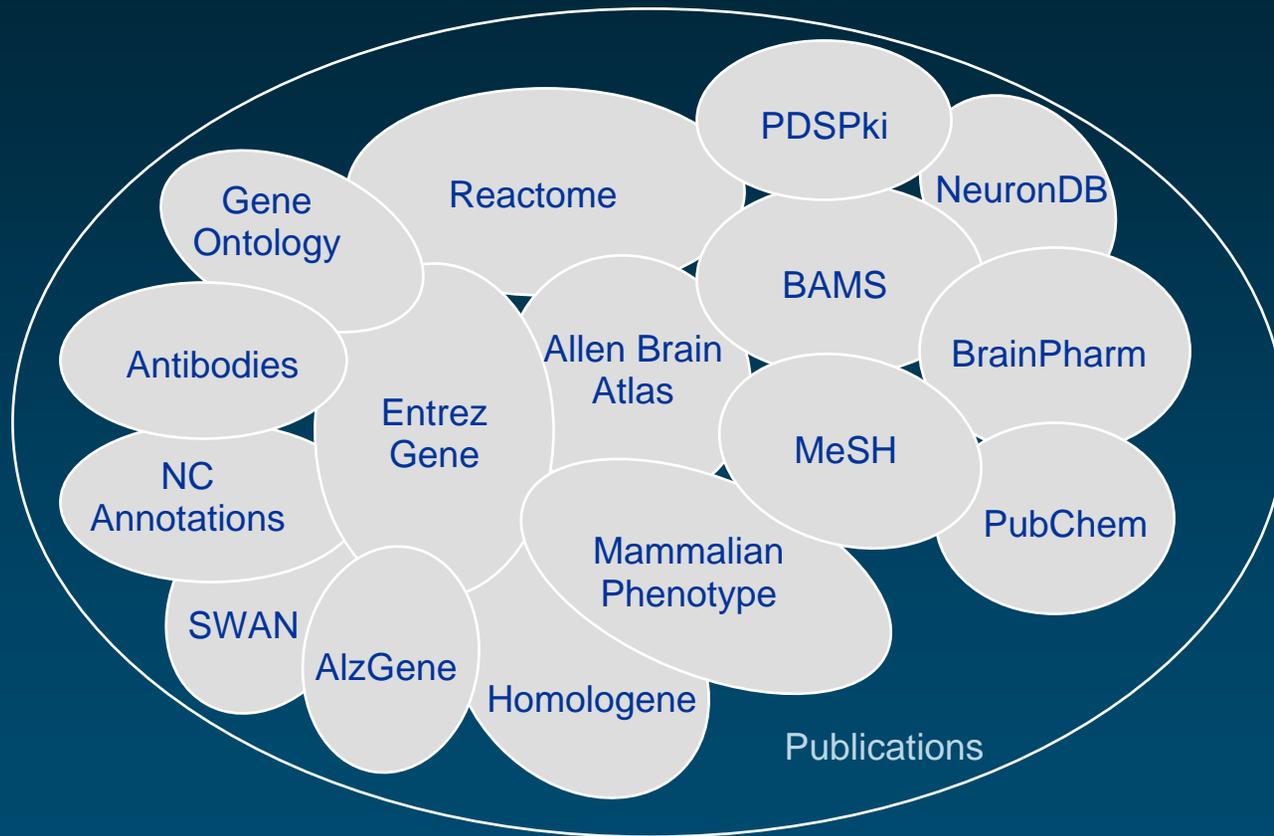


# Biomedical Semantic Web

- ◆ Integration
  - Data/Information
  - E.g., translational research
- ◆ Hypothesis generation
- ◆ Knowledge discovery
  
- ◆ Clinical data
  - Aggregation, sharing, exchange
  - Support for clinical decision



# HCLS mashup of biomedical sources



[http://esw.w3.org/topic/HCLS/HCLSIG\\_DemoHomePage\\_HCLSIG\\_Demo](http://esw.w3.org/topic/HCLS/HCLSIG_DemoHomePage_HCLSIG_Demo)



# Shared identifiers Example

## Entrez Gene

**CH25H**

Order cDNA clone, Links

Official Symbol CH25H and Name: cholesterol 25-hydroxylase [*Homo sapiens*]

Other Aliases: C25H

Chromosome: 10; Location: 10q23

Annotation: Chromosome 10, NC\_000010.9 (90957050..90955509, complement)

MIM: 604551

GeneID: **9023**

### Pathways

Reactome Event: Lipid and lipoprotein metabolism  
73923

### Homology

Mouse, Rat  
[Map Viewer](#)

### GeneOntology

#### Function

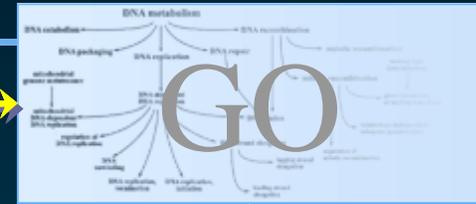
iron ion binding  
metal ion binding  
steroid hydroxylase activity

#### Process

cholesterol metabolic process  
lipid metabolic process  
metabolic process  
sterol biosynthetic process

#### Component

endoplasmic reticulum  
integral to membrane  
membrane  
membrane fraction



## Cholesterol 25-hydroxylase [cytosol]

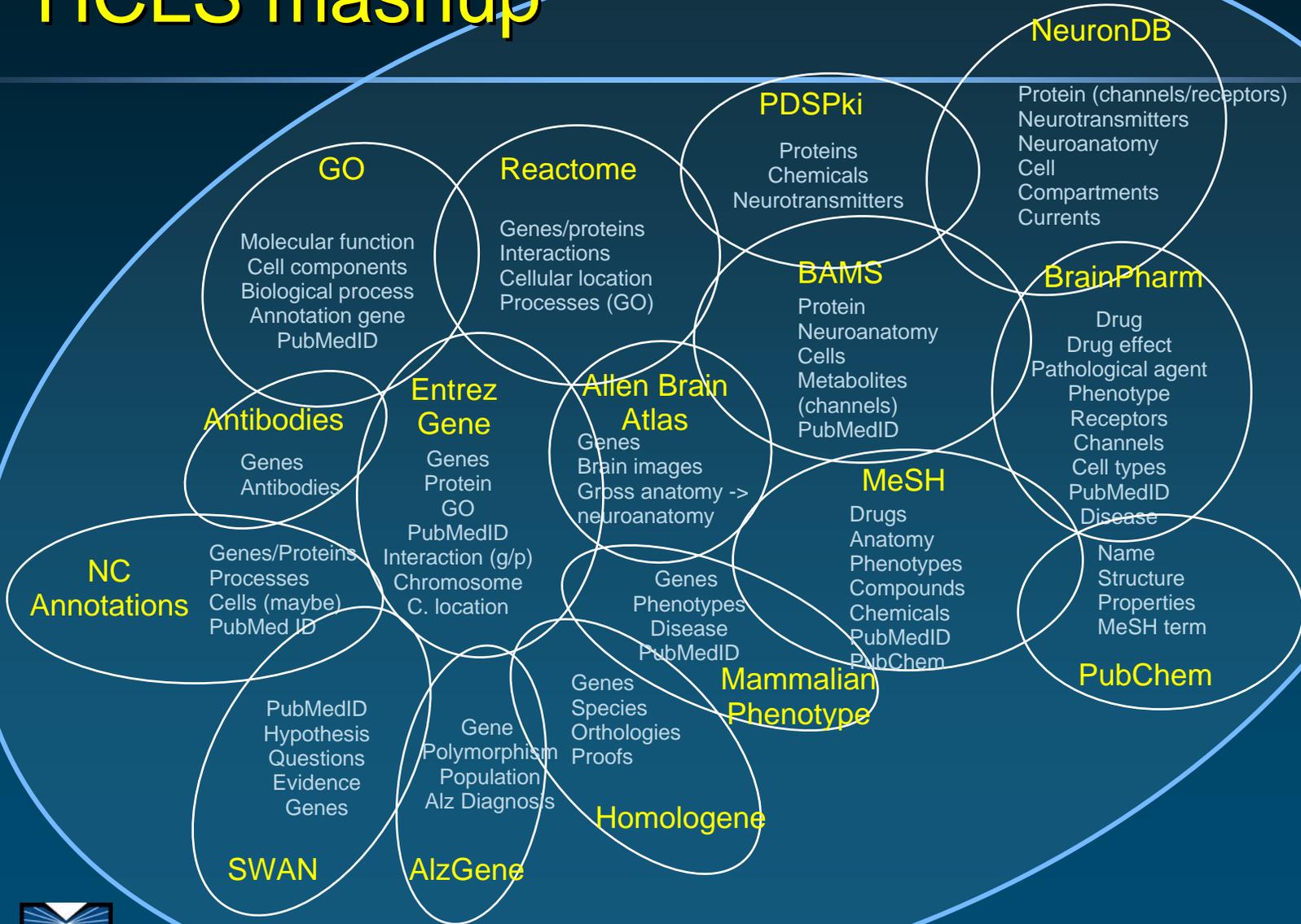


### + Details

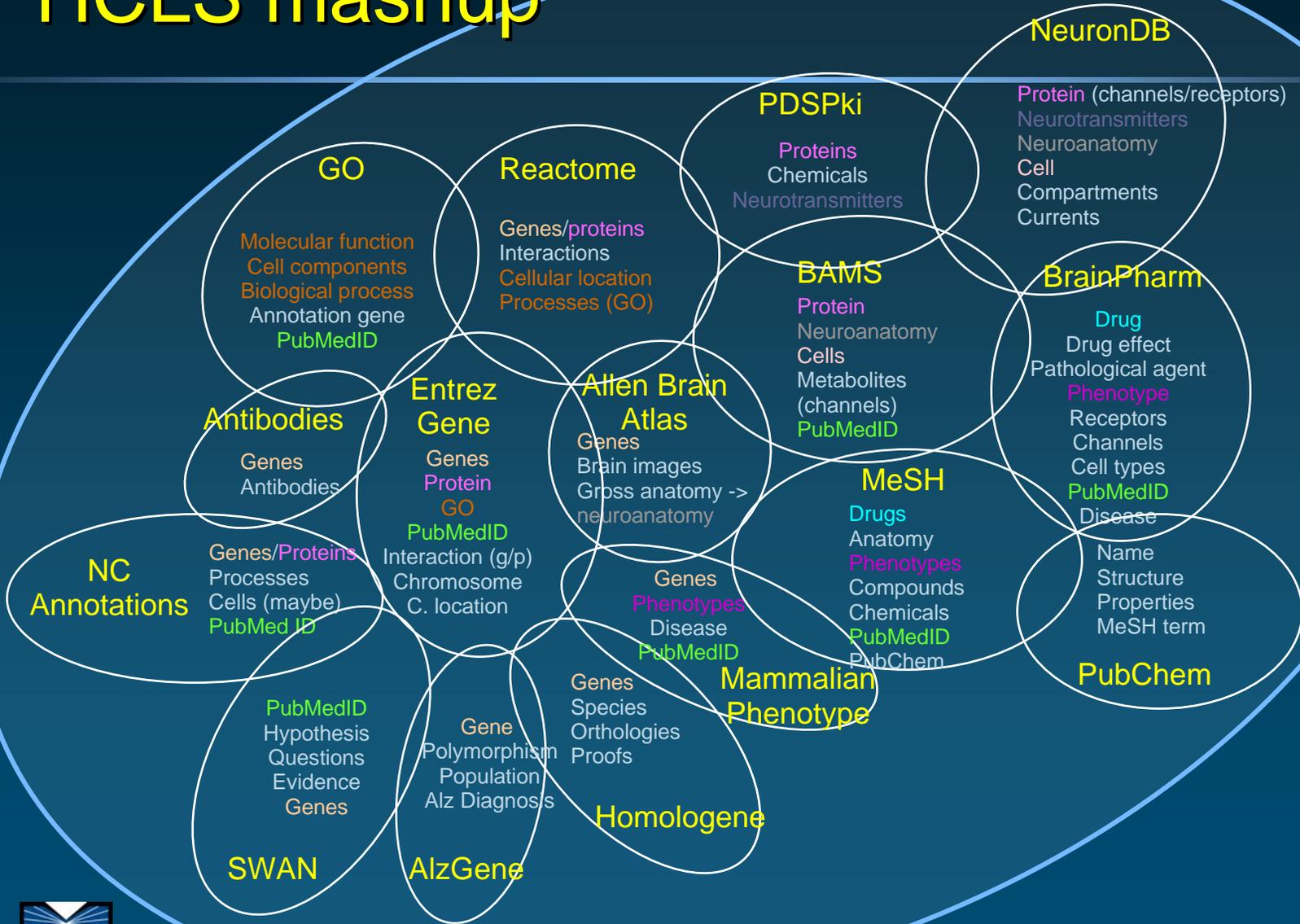
<b>Name</b>	Cholesterol 25-hydroxylase CH25H_HUMAN CH25H
<b>Stable identifier</b>	REACT_10656.1
<b>Link to corresponding entries in other databases</b>	ENSEMBL:ENSG00000138135 <b>Entrez Gene:9023</b> HapMap:NM_003956 KEGG Gene:9023 MIM:604551 RefSeq:NM_003956 RefSeq:NP_003947 UCSC:O95992 UniProt:O95992
<b>Other identifiers related to this sequence</b>	CH25H_HUMAN, ENSG00000138135, ENST00000371852, ENSP00000360918, ENST00000260706, ENSP00000260706, 206932_at, 3236_at, 45019_at, g4502498_3p_at, A_14_P139081, A_23_P86470, CCDS7400, GE6210, AF059212, AF059214, AL513533, BC017843, BC072430, EntrezGene:9023, GI_31542304-S, LMN_8057, IPI00022560, MIM:604551, OTTHUMT0000049291, AAC97481, AAC97483, CAI13519, AAH17843, AAH72430, NM_003956, NP_003947, Hs.47357, Hs.597033, O95992, CH25H_HUMAN, IPR006088
<b>Reference entity</b>	UniProt:O95992 Cholesterol 25-hydroxylase
<b>Coordinates in the reference sequence</b>	..
<b>Cellular compartment</b>	cytosol <b>GO</b>
<b>Organism</b>	Homo sapiens
<b>Component of</b>	CH25H (Fe2+ cofactor) [endoplasmic reticulum membrane]
<b>Participates in processes</b>	<a href="#">Lipid and lipoprotein metabolism</a> <ul style="list-style-type: none"> <li>└ Steroid metabolism           <ul style="list-style-type: none"> <li>└ Metabolism of bile acids and bile salts               <ul style="list-style-type: none"> <li>└ Synthesis of bile acids and bile salts                   <ul style="list-style-type: none"> <li>└ Cholesterol is hydroxylated to 25-hydroxycholesterol [Homo sapiens]</li> </ul> </li> </ul> </li> </ul> </li> </ul>



# HCLS mashup



# HCLS mashup



# Some unresolved issues

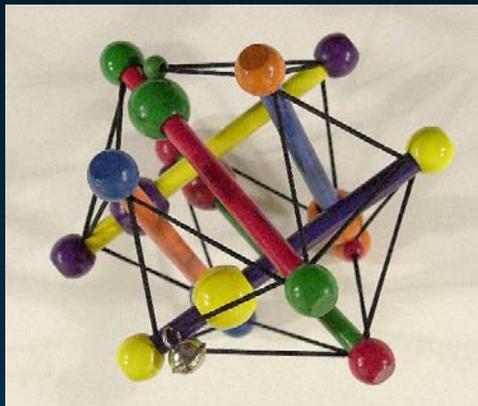
- ◆ Format
  - RDF/S, OWL, SKOS vs. OBO, RRF, etc.
  - Converters
- ◆ Permanent identification of biomedical entities
  - Syntax: URI vs. LSID
  - Semantic: Trans-namespace identification
- ◆ Availability, openness
- ◆ Governance, trust

# Summary

- ◆ Terminologies/Ontologies provide
  - Lists of entities
  - Names for entities
  - Identifiers for entities
- ◆ Additionally
  - Information model for integration
  - Trans-namespace resolution
  - Support for inference

# Future directions

- ◆ Information integration
  - Knowledge extracted from text
  - Knowledge in structured knowledge bases
- ◆ Ontologies for relations
  - In complement to ontologies for entities
  - To support reasoning



# 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