



Short course – Summer 2010
Clinical Ontology in Practice
June 15-17, 2010

Clinical Ontology in Practice



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

Objectives

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



Lister Hill National Center for Biomedical Communications 2

Agenda

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



Lister Hill National Center for Biomedical Communications 3

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.



Lister Hill National Center for Biomedical Communications 4

References Bio-ontology courses

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



Lister Hill National Center for Biomedical Communications 5



Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov



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

 Short course – Summer 2010
Clinical Ontology in Practice

June 15, 2010 – Session #1

Introduction to Biomedical Ontologies

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

Outline

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

 Lister Hill National Center for Biomedical Communications 8

Historical perspective

Why biomedical terminologies?

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

 Lister Hill National Center for Biomedical Communications 10

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)




 Lister Hill National Center for Biomedical Communications

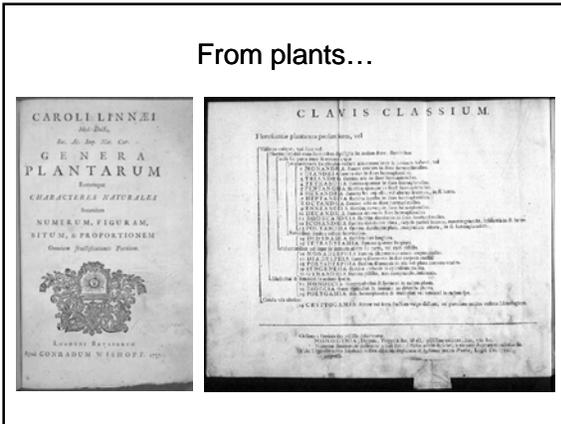
To classify diseases (and plants)

- ◆ 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)





 Lister Hill National Center for Biomedical Communications 12



... 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 Methodica

(Cited by Chris Chute)

Lister Hill National Center for Biomedical Communications
14

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)

Lister Hill National Center for Biomedical Communications
15

London Bills of Mortality

Lister Hill National Center for Biomedical Communications
16

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.

Lister Hill National Center for Biomedical Communications
17

To index and retrieve information

- ◆ Biomedical literature
 - MEDLINE (15M citations from 4600 journals)
 - Manually indexed
 - Medical Subject Headings (MeSH)
- ◆ Genome
 - Model organism databases (Fly, Mouse, Yeast, ...)
 - Manually / semi-automatically curated
 - Gene Ontology

Lister Hill National Center for Biomedical Communications
18

MEDLINE and MeSH

1. Hist. Neurosci. 2004 Mar; 13(3):91-101

MeSH Press
Black bile and psychomotor retardation: shades of melancholia in Dante's Inferno.
Widnes DA.
Memorial Sloan-Kettering Cancer Center, New York, NY 10017, USA. widnes4@mskcc.org

The faculty of melancholy depression is said with aspects of ancient melancholia and animal dysfunction. The recent evolution of psychomotor symptoms in the depressive terminology of affective disorder was traced to the studies of medicine in medieval times. The more back to the biology of this psychomotor dysfunction with the technical advances in brain imaging in recent years only echoes centuries-old writings on the centrality of movement changes in the depressive condition. The Inferno, the first cantica of Dante Alighieri's *Commedia*, has a wonderful description of hellfire to the symptoms of psychomotor symptoms as described in the first cantica. *Commedia*, medieval speech, Italian terms, those and more other aspects from the physical manifestations of psychomotor suffering in the forefront of the reader's mind. Considering *Medieval* and Renaissance writings on melancholy suffering, it is fitting that Dante shows a bodily illness reflected in the hellish torments visited on the damned. From the words of the poet to those of the scientist, the symptoms of psychomotor symptoms play a prominent role in the process as well as in the medical and literary aspects of melancholy condition.

MeSH Terms
 • Depressive Disorder/hist*
 • History of Medicine, Medieval
 • Human
 • Italy
 • Literature, Medieval/hist*
 • Medicine in Literature*
 • Psychiatry/hist*
 • Psychomotor Disorder/hist*

PubMed National Library of Medicine

Mouse Genome Database and GO

Entrez Gene
1. NF2 neurofibromin 2 [Homo sapiens]
GeneID: 18014 Location: MGI:927302

General gene information
Gene Ontology Provided by MGI

Function	Evidence
cytoskeletal protein binding	IEA
cytoskeleton binding	IEA
proteasome binding	IEA
proteasome binding	IEA
Protein	IEA
intracellular junction assembly and/or maintenance	IEA
regulator of cell cycle	IEA
regulator of protein kinase activity	IEA
regulator of cell proliferation	IEA
Regulation	IEA
cellular junction	IEA
cytoskeleton	IEA
cytoskeleton	IEA
cytoskeleton	IEA

Lister Hill National Center for Biomedical Communications 20

To serve as a reference

- ◆ Reference terminology/ontology
 - Universally needed
 - Developed independently of any purposes
 - Reusable by many applications
- ◆ Examples
 - VA National Drug File (NDF)
 - Foundational Model of Anatomy (FMA)
 - SNOMED CT

Lister Hill National Center for Biomedical Communications 21

Anatomy in Biomedicine

Lister Hill National Center for Biomedical Communications 22

Administrative terminologies

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

Lister Hill National Center for Biomedical Communications 23

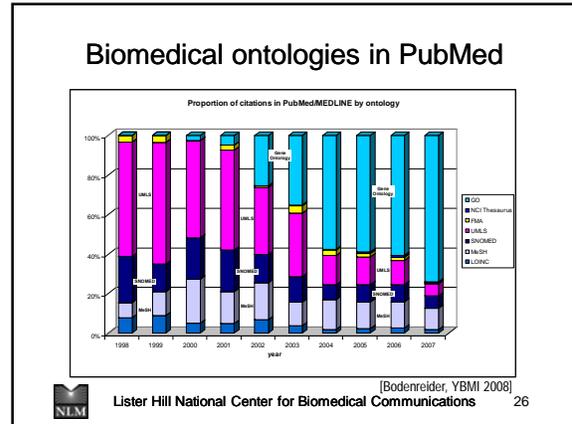
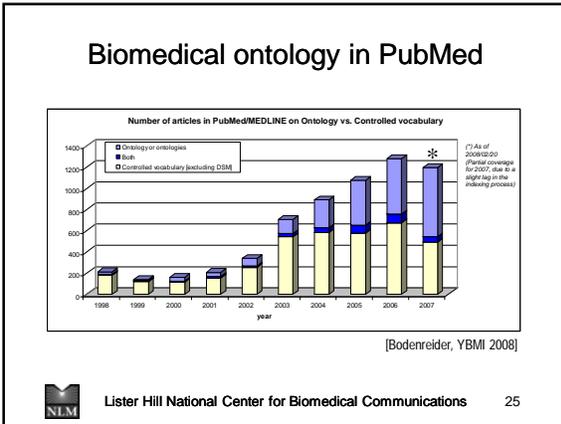
History of Medical Ontologies

[Bodenreider, BIB 2006] (courtesy of J. Rogers)

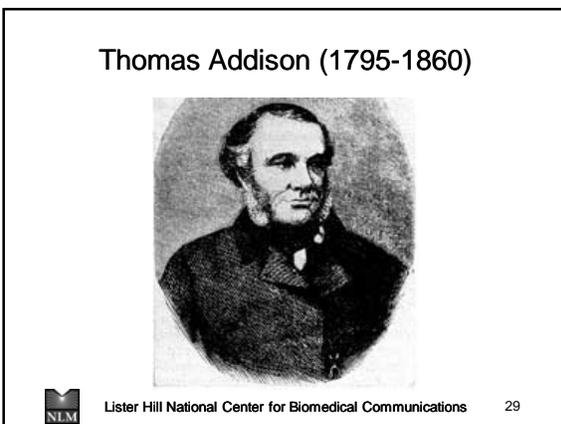
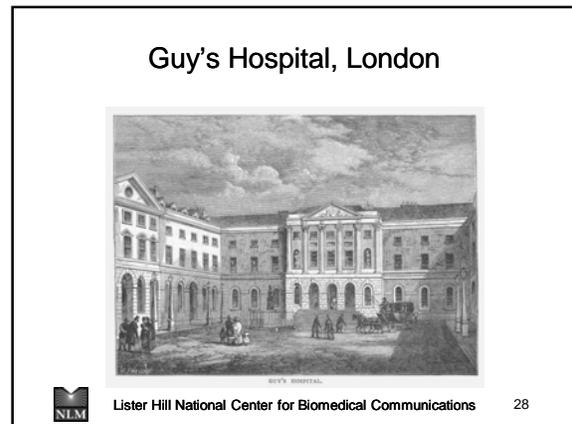
1603 1700 1785 1855 1900 1975

1975 1985 1995 2005

Lister Hill National Center for Biomedical Communications



Introduction to biomedical terminologies through an example



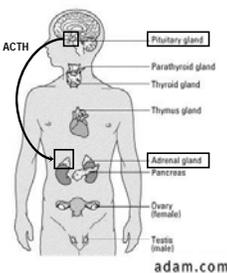
Addison's disease

- ◆ 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

NLM Lister Hill National Center for Biomedical Communications 30

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



adam.com

 Lister Hill National Center for Biomedical Communications 31

Addison's disease: Symptoms

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

 Lister Hill National Center for Biomedical Communications 32

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

 Lister Hill National Center for Biomedical Communications 33

Internal Classification of Diseases

CHAPTER 4
Endocrine, nutritional and metabolic diseases (E00-E90)

Disorders of other endocrine glands (E20-E35)

E27 Other disorders of adrenal gland

E27.0 Other adrenocortical insufficiency
Overproduction of ACTH, not associated with Cushing's disease
Primary adrenocortical insufficiency
Excludes1: Cushing's syndrome (E24.-)

E27.1 Primary adrenocortical insufficiency
EXCLUDES (PRIMARY DEFICIENCY)
Adrenocortical insufficiency-NOS
Autoimmune adrenalitis
Excludes1: Addison's only phenotype adrenoleukodystrophy (E27.1.428)
amyloidosis (E85)
tuberculosis: Addison's disease (A18.-)
Waterhouse-Friderichsen syndrome (A39.-)

E27.2 Addisonian crisis
Adrenocortical crisis

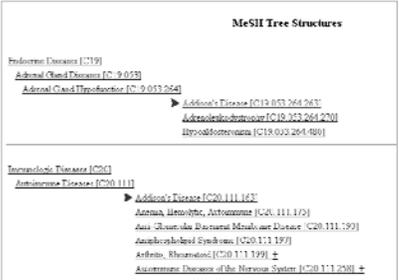
E27.3 Drug-induced adrenocortical insufficiency
Code first (E26-E29) to identify drug

E27.4 Other and unspecified adrenocortical insufficiency

Diagnosis ICD-10-CM Tabular Page 180 June 2003

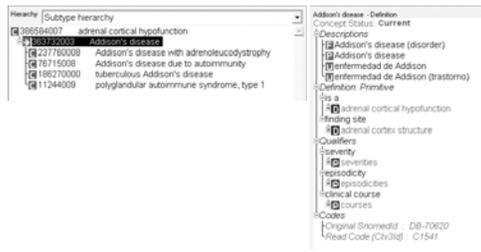


Medical Subject Headings



 Lister Hill National Center for Biomedical Communications 35

SNOMED CT



 Lister Hill National Center for Biomedical Communications 36

Biomedical terms as names for biomedical classes

Terms reflecting valid classes

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

 Lister Hill National Center for Biomedical Communications 38

Issues

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

 Lister Hill National Center for Biomedical Communications 39

Multiple terms for a class

- ◆ Synonymy

<ul style="list-style-type: none"> ▪ Left coronary artery ▪ LCA ▪ Arteria coronaria sinistra 	<ul style="list-style-type: none"> ▪ Addison's disease ▪ Primary adrenocortical insufficiency
---	---
- ◆ "Clinical synonymy" (vs. identity)

<ul style="list-style-type: none"> ▪ Abdominal swelling ▪ Swollen abdomen 	<ul style="list-style-type: none"> ▪ Addison's disease ▪ Primary adrenocortical insufficiency
<ul style="list-style-type: none"> ▪ Posttransfusion hepatitis ▪ Posttransfusion viral hepatitis 	
<small>vs. Waterhouse-Friderichsen Syndrome</small>	

 Lister Hill National Center for Biomedical Communications 40

Multiple classes for a term

- ◆ Polysemy

Cold	}	<ul style="list-style-type: none"> ▪ Cold ▪ Common cold
	}	<ul style="list-style-type: none"> ▪ Cold ▪ Cold temperature
	}	<ul style="list-style-type: none"> ▪ COLD ▪ Chronic Obstructive Airway Disease
- ◆ Truncated terms

Calcium	}	<ul style="list-style-type: none"> ▪ Calcium ▪ Ca⁺⁺ ▪ Coagulation factor IV
	}	<ul style="list-style-type: none"> ▪ Calcium ▪ Calcium measurement

 Lister Hill National Center for Biomedical Communications 41

Non-ontological features in terms

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

 Lister Hill National Center for Biomedical Communications 42

Ontology vs. Epistemology

- ◆ **Ontology**
 - Invariants in reality
 - Classes (universals)
 - Relations between them
 - Theory of reality
- ◆ **Epistemology**
 - Knowledge about such entities
 - Perception of reality

Bone metastasis

Bone metastasis diagnosed by CT scan

Bone metastasis diagnosed by Tc-99m bone scintiscan

43

Composite terms

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

Lister Hill National Center for Biomedical Communications 44

More composite terms

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

45

Terminological relations as a surrogate for ontological relations

Issues

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

Lister Hill National Center for Biomedical Communications 47

Explicit classificatory principle

Foundational Model of Anatomy

Lister Hill National Center for Biomedical Communications 48

No explicit classificatory principle

3. Diseases [C]

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

agent/cause

location

stage in life

NLM Lister Hill National Center for Biomedical Communications 49

- Certain infectious and parasitic diseases
- Neoplasms
- Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- Endocrine, nutritional, and metabolic diseases
- Mental and behavioral disorders
- Diseases of nervous system
- Diseases of the eye and adnexa
- Diseases of the ear and mastoid process
- Diseases of circulatory system
- Diseases of respiratory system
- Diseases of digestive system
- Diseases of the skin and subcutaneous tissue
- Diseases of the musculoskeletal system and connective tissue
- Diseases of the genitourinary system
- Pregnancy, childbirth, and the puerperium
- Certain conditions originating in the newborn (perinatal) period
- Congenital malformations, deformations and chromosomal abnormalities
- Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- Injury, poisoning and certain other consequences of external causes
- External causes of morbidity
- Factors influencing health status and contact with health service

ICD-10

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

NLM Lister Hill National Center for Biomedical Communications 51

Fully specified relations

Viral meningitis in SNOMED CT

Fully defined by ...

- is a
 - viral infections of the central nervous system
 - infective meningitis
- Causative agent
 - virus
- Group
 - Associated morphology
 - inflammation
 - Finding site
 - meninges structure

NLM Lister Hill National Center for Biomedical Communications 52

Underspecification of the relations

parent

isa ?

child

NLM Lister Hill National Center for Biomedical Communications 53

Thesaurus relations

◆ Addison's disease

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

Relations used to create hierarchical structures vs. hierarchical relations

NLM Lister Hill National Center for Biomedical Communications 54

The screenshot shows two hierarchical views of MeSH terms. The top view is for Endocrine Diseases (C19), with Adrenal Gland Diseases (C19.023) expanded to show Adrenal Gland Hypofunction (C19.053.264), which includes Addison's Disease (C19.053.264.263), Adrenoleucodystrophy (C19.053.264.270), and Hypoadosteronism (C19.053.264.480). The bottom view is for Immunologic Diseases (C20), with Autoimmune Diseases (C20.111) expanded to show Addison's Disease (C20.111.163), which includes Atensia, Hemolytic, Autoimmune (C20.111.175), Anti-Clonalular Basement Membrane Disease (C20.111.190), Autoimmunologic Syndrome (C20.111.197), and Arthritis, Rheumatoid (C20.111.199). A 'Hierarchy' window at the bottom shows a 'Subtype hierarchy' for 'adrenal cortical hypofunction', listing 'Addison's disease', 'Addison's disease due to autoimmunity', 'Addison's disease with adrenoleucodystrophy', 'polyglandular autoimmune syndrome, type 1', and 'tuberculous Addison's disease'.

Accidents in MeSH

Environment and Public Health [G03]
Public Health [G03.850]
Accidents [G03.850.110]

- Accident Prevention [G03.850.110.060] +
- Accidental Falls [G03.850.110.085]
- Accidents, Aviation [G03.850.110.185]
- Accidents, Home [G03.850.110.205]
- Accidents, Occupational [G03.850.110.250] +
- Accidents, Radiation [G03.850.110.285]
- Accidents, Traffic [G03.850.110.320]
- Drowning [G03.850.110.500] +

Lister Hill National Center for Biomedical Communications 56

Limited depth in hierarchies "by design"

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

E84 Cystic fibrosis
Includes: mucoviscidosis
E84.0 Cystic fibrosis with pulmonary manifestations
Use additional code to identify any infectious organism present, such as: Pseudomonas (B96.5)
E84.1 Meconium ileus in cystic fibrosis
Excludes1: meconium ileus not due to Cystic fibrosis (P75)
E84.2 Cystic fibrosis with gastrointestinal manifestations
Excludes2: meconium ileus in cystic fibrosis (E84.1)
E84.8 Cystic fibrosis with other manifestations

Lister Hill National Center for Biomedical Communications 57

Cystic fibrosis in ICD

E84 Cystic fibrosis
Includes: mucoviscidosis
E84.0 Cystic fibrosis with pulmonary manifestations
Use additional code to identify any infectious organism present, such as: Pseudomonas (B96.5)
E84.1 Meconium ileus in cystic fibrosis
Excludes1: meconium ileus not due to Cystic fibrosis (P75)
E84.2 Cystic fibrosis with gastrointestinal manifestations
Excludes2: meconium ileus in cystic fibrosis (E84.1)
E84.8 Cystic fibrosis with other manifestations

The diagram illustrates two hierarchical structures for Cystic Fibrosis (CF). The left structure shows a root 'CF' branching into 'CF w. Pulm.', 'CF w. GI', 'Meconium ileus in CF', and 'CF w. other'. The right structure shows a root 'CF' branching into 'CF w. Pulm.', 'CF w. GI', and 'CF w. other', with 'Meconium ileus in CF' as a sub-term under 'CF w. other'.

Lister Hill National Center for Biomedical Communications 58

Conclusions

Conclusions ☹

- ◆ Biomedical terms
 - reflect some aspects of biomedical reality
 - Although the primary concern of terminology is naming, not reflecting reality
 - often convey additional features (e.g., epistemology)
- ◆ Biomedical terminology tends to offset part of the complexity
 - but often reflects utility

Lister Hill National Center for Biomedical Communications 60

Conclusions ☺

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



Lister Hill National Center for Biomedical Communications 61



Short course – Summer 2010
Clinical Ontology in Practice

June 15, 2010 – Session #2

Design Principles, Formalisms and Tools for Biomedical Ontologies



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

Overview

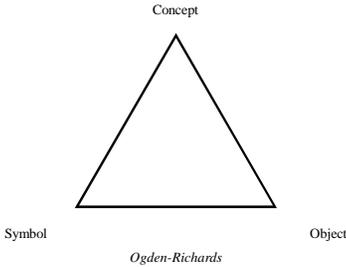
- ◆ Definitions
 - Ontologies vs. other artifacts
 - Kinds of ontologies
- ◆ Some principles of formal ontology
 - Top-level categories
 - Categories of relationships
- ◆ Formalisms and tools



Lister Hill National Center for Biomedical Communications 63

Definitions

Introduction



Concept

Symbol Object

Ogden-Richards



Lister Hill National Center for Biomedical Communications 65

Definitions

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



Lister Hill National Center for Biomedical Communications 66

Examples of use

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

- ◆ Interoperability

 Lister Hill National Center for Biomedical Communications 67

Ontology vs. other artifacts

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

[Smith, KR-MED 2006]
[Chute, JAMIA 2000]

 Lister Hill National Center for Biomedical Communications 68

(Controlled) Terminology

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

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

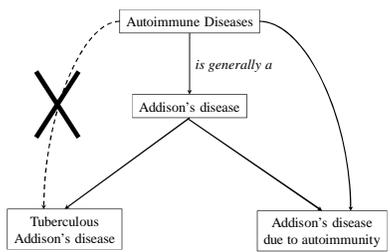
 Lister Hill National Center for Biomedical Communications 69

Thesaurus

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

 Lister Hill National Center for Biomedical Communications 70

Thesaurus vs. ontology



 Lister Hill National Center for Biomedical Communications 71

Classification

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

E10 **Insulin-dependent diabetes mellitus**
[See before E10 for subdivisions.]
Includes: diabetes (mellitus):
- brittle
- juvenile-onset
- latios-prone
- type 1

Excludes: diabetes mellitus (in):
- malnutrition-related (E12.-)
- neonatal (E70.2)
- pregnancy, childbirth and the puerperium (O24.-)
glycosuria:
- NOS (R91)
- renal (E24.8)
impaired glucose tolerance (E73.0)
postsurgical hypoinsulinaemia (E89.1)

 Lister Hill National Center for Biomedical Communications 72

Classification

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

E84 Cystic fibrosis
Includes: mucoviscidosis

E84.0 Cystic fibrosis with pulmonary manifestations

E84.1 Cystic fibrosis with intestinal manifestations
Meconium ileus+ (P75*)
Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)

E84.8 Cystic fibrosis with other manifestations
Cystic fibrosis with combined manifestations

E84.9 Cystic fibrosis, unspecified

 Lister Hill National Center for Biomedical Communications 73

Knowledge Bases

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

 Lister Hill National Center for Biomedical Communications 74

Fuzzy borders

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

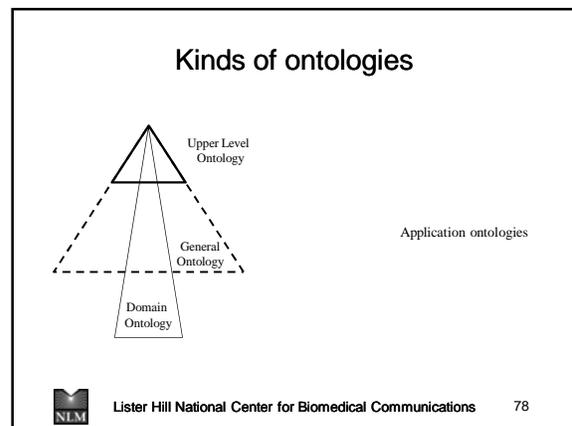
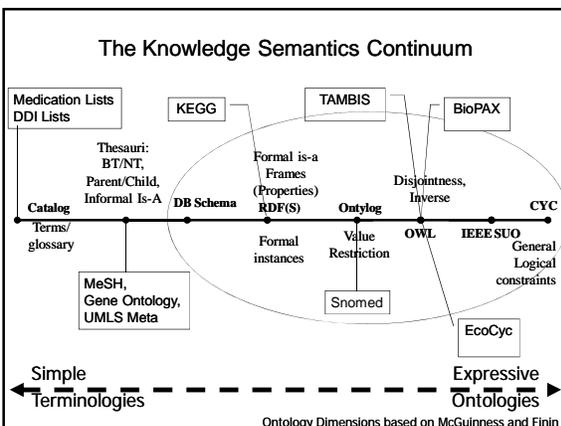
 Lister Hill National Center for Biomedical Communications 75

Types of resources

- ◆ Lexical resources
 - Collections of lexical items
 - Additional information
 - Part of speech
 - Spelling variants
 - Useful for entity recognition
 - UMLS SPECIALIST Lexicon, WordNet
- ◆ Ontological resources
 - Collections of
 - kinds of entities (substances, qualities, processes)
 - relations among them
 - Useful for relation extraction
 - UMLS Semantic Network, BioTop
- ◆ Terminological resources
 - Collections lexical items + identifiers
 - Useful for entity resolution
 - UMLS Metathesaurus



 Lister Hill National Center for Biomedical Communications 74



Ontology-related issues

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



Lister Hill National Center for Biomedical Communications 79

Formal Ontological Principles

Formal ontological distinctions

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



Lister Hill National Center for Biomedical Communications 81

Universal vs. Individual

◆ Universal = <i>category</i>	◆ Individual = <i>instance</i>
◆ Synonyms <ul style="list-style-type: none">• Kind, Type, (Class)	◆ Synonyms <ul style="list-style-type: none">• Particular, Token
◆ Examples <ul style="list-style-type: none">• eyeball• blood pressure• conference	◆ Examples <ul style="list-style-type: none">• my right eyeball• my blood pressure (132/79)• AMIA Annual Symposium 2003


instantiation



Lister Hill National Center for Biomedical Communications 82

Continuant vs. Occurrent

◆ Continuant = <i>Continues to exist through time</i>	◆ Occurrent = <i>Unfolds through time</i>
◆ Synonyms <ul style="list-style-type: none">• Substance	◆ Synonyms <ul style="list-style-type: none">• Process
◆ Examples <ul style="list-style-type: none">• tennis racquet• mitochondrion• insulin production	◆ Examples <ul style="list-style-type: none">• tennis tournament• metabolism• producing insulin



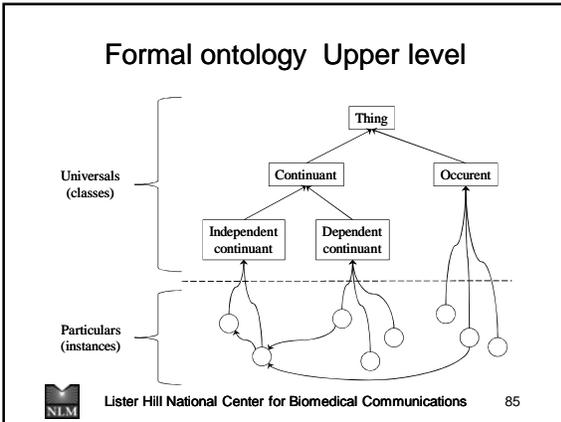
Lister Hill National Center for Biomedical Communications 83

Independent vs. Dependent

◆ Independent = <i>Can exist without support from other entities</i>	◆ Dependent = <i>Require support from other entities for its existence</i>
◆ Examples <ul style="list-style-type: none">• virus• molecule• plant	◆ Examples <ul style="list-style-type: none">• viral infection• DNA binding• food



Lister Hill National Center for Biomedical Communications 84

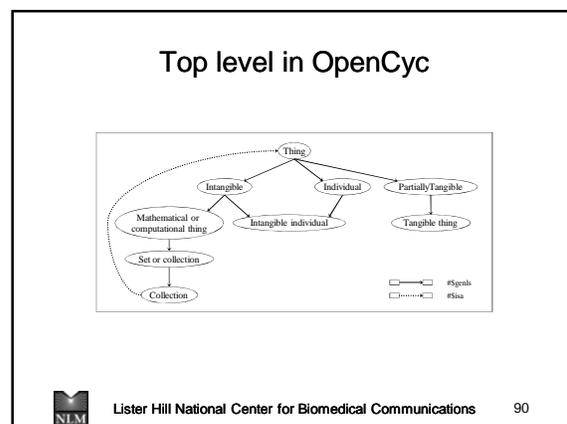


- ### Formal ontological distinctions
- ◆ Basic distinctions in many top-level ontologies
 - Generic: BFO, DOLCE
 - Biomedical: BioTop, UMLS Semantic Network
 - ◆ Condition the relations between various types of entities
 - Relations
 - Between instances (e.g., *part_of* [at time])
 - Between classes (e.g., *isa*, *part_of* [atemporal])
 - Between one instance and one class (*instance_of*)
- [Smith, Genome Biology 2005]
- NLM Lister Hill National Center for Biomedical Communications 86

- ### Formal ontology in practice
- ◆ Provides foundational classes and relations
 - Upper level ontologies
 - Relation ontology
 - ◆ Provides a framework for analyzing entities and relations
- NLM Lister Hill National Center for Biomedical Communications 87

Examples

- ### General ontologies
- ◆ OpenCyc
 - General ontology
 - Cycorp, Inc (D. Lenat & al.)
 - Over 1M hand-coded assertions
 - <http://www.opencyc.org>
 - ◆ WordNet
 - Electronic lexical database
 - Princeton University (G. Miller & al.)
 - Over 100,000 synsets
 - <http://wordnet.princeton.edu/>
- NLM Lister Hill National Center for Biomedical Communications 89



Top level in WordNet

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

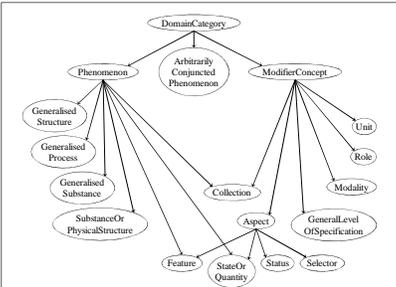
 Lister Hill National Center for Biomedical Communications 91

GALEN

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

 Lister Hill National Center for Biomedical Communications 92

Top level in GALEN



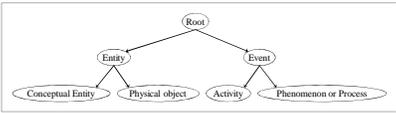
 Lister Hill National Center for Biomedical Communications 93

UMLS Semantic Network

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

 Lister Hill National Center for Biomedical Communications 94

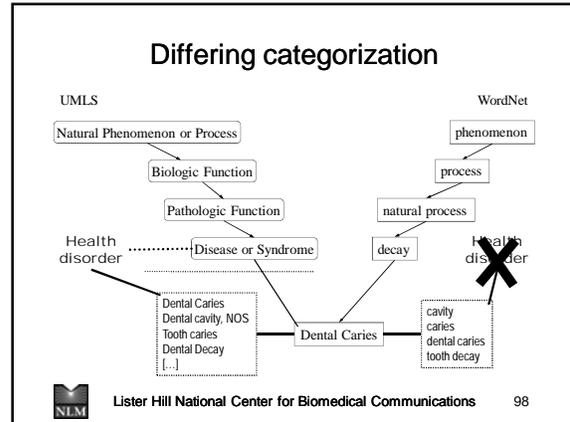
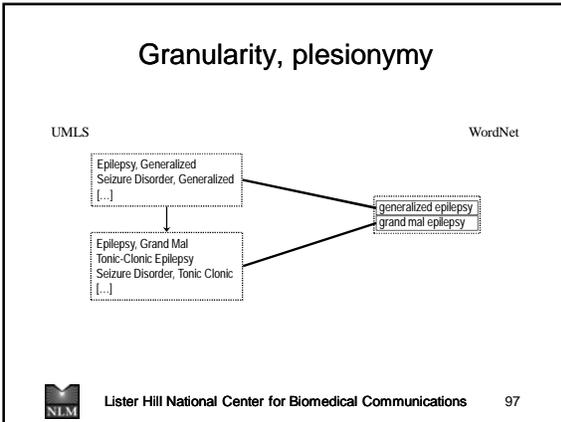
Top level in the Semantic Network



 Lister Hill National Center for Biomedical Communications 95

Differences between ontologies

Examples



Formalisms and Tools

- ## Ontology and Formalism
- ◆ Frames
 - ◆ Description logics
 - OWL DL
 - ◆ First-order logic
 - ◆ OBO Format
 - Conversion to OWL DL
- Lister Hill National Center for Biomedical Communications 100

- ## Tools for ontology developers
- ◆ Protégé
 - Publicly available
 - Frames and DL
 - Classifier
 - Supports many file formats (import/export)
 - Large community of users
 - Well supported
 - <http://protege.stanford.edu/>
 - ◆ OBO-Edit
 - Specific to the biomedical/OBO community
 - Simpler than Protégé (“tool for biologists”)
 - <http://oboedit.org/>
- <http://protege.stanford.edu/>
- The OBO Ontology Editor
- Lister Hill National Center for Biomedical Communications 101

Short course – Summer 2010
Clinical Ontology in Practice

June 15, 2010 – Session #3

“High-Impact” Biomedical Ontologies

A Structural Perspective

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

Overview

- ◆ Structural perspective [J. Cimino, YBMI 2006]
 - What are they (vs. what are they for)?
- ◆ “High-impact” biomedical ontologies
 - International Classification of Diseases (ICD)
 - Logical Observation Identifiers, Names and Codes (LOINC)
 - SNOMED Clinical Terms
 - Foundational Model of Anatomy
 - Gene Ontology
 - RxNorm
 - Medical Subject Headings (MeSH)
 - NCI Thesaurus
 - Unified Medical Language System (UMLS)

 Lister Hill National Center for Biomedical Communications 103

International Classification of Diseases



ICD Characteristics (1)

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

 Lister Hill National Center for Biomedical Communications 105

ICD Characteristics (2)

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

 Lister Hill National Center for Biomedical Communications 106

ICD Top level

Chapter	Blocks	Title
I	A00-A99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	I60-I99	Diseases of the ear and mastoid process
IX	J00-J99	Diseases of the circulatory system
X	K00-K93	Diseases of the respiratory system
XI	L00-L99	Diseases of the skin and subcutaneous tissue
XII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIII	N00-N99	Diseases of the genitourinary system
XIV	O00-O99	Pregnancy, childbirth and the puerperium
XV	P00-P96	Certain conditions originating in the perinatal period
XVI	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XVIII	S00-T98	Injury, poisoning and certain other consequences of external causes
XIX	V01-Y98	External causes of morbidity and mortality
XX	Z00-Z99	Factors influencing health status and contact with health services
XXI	U00-U99	Codes for special purposes

 Lister Hill National Center for Biomedical Communications 107

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 1

Excludes: diabetes mellitus (m):

- malnutrition-related (E12.-)
- neonatal (E70.2)
- pregnancy, childbirth and the puerperium (O24.-)
- glycosuria
- NOS (R01)
- renal (E74.8)
- impaired glucose tolerance (E73.0)
- postsurgical hypoinsulinaemia (E89.1)

 Lister Hill National Center for Biomedical Communications 108

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



Lister Hill National Center for Biomedical Communications 109

Logical Observation Identifiers, Names and Codes (LOINC)



LOINC Characteristics (1)

- ◆ Current version: 2.30 (Feb. 2010)
- ◆ Type: Controlled terminology*
- ◆ Domain: Laboratory and clinical observations
- ◆ Developer: Regenstrief Institute
- ◆ Funding: NLM
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS
- ◆ URL: www.regenstrief.org/loinc/loinc.htm



Lister Hill National Center for Biomedical Communications 111

LOINC Characteristics (2)

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



Lister Hill National Center for Biomedical Communications 112

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	--



Lister Hill National Center for Biomedical Communications 113

SNOMED Clinical Terms



SNOMED CT Characteristics (1)

- ◆ Current version: January 31, 2010 (2 annual releases)
- ◆ Type: Reference terminology / ontology
- ◆ Domain: Clinical medicine
- ◆ Developer: IHTSDO
- ◆ Funding: IHTSDO
- ◆ Availability
 - Publicly available: Yes* (in member countries)
 - Repositories: UMLS
- ◆ URL: <http://www.ihtsdo.org/>



Lister Hill National Center for Biomedical Communications 115

SNOMED CT Characteristics (2)

- ◆ Number of
 - Concepts: ~310,000 active concepts (Jan. 31, 2010)
 - Terms: ~800,000 active “descriptions”
- ◆ Major organizing principles:
 - Utility for clinical medicine (e.g., assertional + definitional knowledge)
 - Model of meaning (incomplete)
 - Rich set of associative relationships
 - Small proportion of defined concepts (many primitives)
- ◆ Formalism: Description logics (KRSS)



Lister Hill National Center for Biomedical Communications 116

SNOMED CT Top level

Hierarchy	Subtype hierarchy
100072005	SNOMED CT Concept
362361000	qualifier value
108237007	linkage concept
370115009	special concept
48176007	social context
419891008	record artifact
363787002	observable entity
308918002	environment or geographical location
123038009	specimen
254281000	staging and scales
123037004	body structure
272379006	event
78621006	physical force
404884003	clinical finding
260787004	physical object
410607006	organism
71388002	procedure
373873005	pharmaceutical / biologic product
243789009	situation with explicit context
105580001	substance



Lister Hill National Center for Biomedical Communications 117

SNOMED CT Example

Hierarchy	Subtype hierarchy	Definition Fully defined by ...
27010001	partial excision of large intestine	partial excision of large intestine
8813002	operation on appendix	operation on appendix
39415002	appendectomy	excision - action
82730006	incidental appendectomy	method
49438003	appendectomy with drainage	excision - action
174036004	emergency appendectomy	procedure site - Direct
174045003	interval appendectomy	appendix structure
6025007	laparoscopic appendectomy	Qualifiers
235113004	non-emergency appendectomy	access
235314005	inversion appendectomy	surgical access values
1298000	excision of appendiceal stump	priority
	appendectomy - Definition	Code
	Concept Status: Current	Original SnomedId : P1-57450
	Descriptions	Read Code (Civ3Id) : X20Wz
	appendectomy (procedure)	
	appendectomy	
	excision of appendix	
	appendectomy	



Lister Hill National Center for Biomedical Communications 118

Foundational Model of Anatomy

FMA Characteristics (1)

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



Lister Hill National Center for Biomedical Communications 120

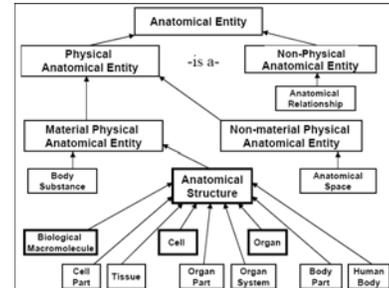
FMA Characteristics (2)

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

 Lister Hill National Center for Biomedical Communications 121

FMA Top level

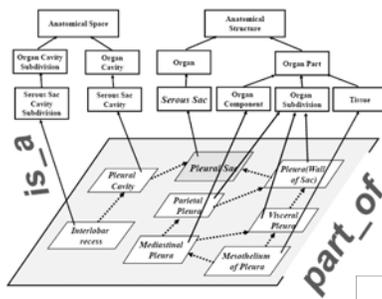
(Courtesy of C. Rosse)



 Lister Hill National Center for Biomedical Communications 122

FMA Example

(Courtesy of C. Rosse)



 123

Gene Ontology



Gene Ontology Characteristics (1)

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

 Lister Hill National Center for Biomedical Communications 125

Gene Ontology Characteristics (2)

- ◆ Number of
 - Concepts: 27,800 (July 22, 2009)
 - Terms: 2.15 per concept
- ◆ Major organizing principles:
 - 3 major hierarchies
 - Molecular function
 - Cellular component
 - Biological process
 - Relations (within hierarchies): *isa*, *part_of*, *regulates*
 - No relations between concepts across hierarchies
- ◆ Formalism: OBO format

 Lister Hill National Center for Biomedical Communications 126

Gene Ontology Top level (MF)

```

all: all [250418 gene products] &
  GO:0008150: biological_process [166605 gene products]
  GO:0005575: cellular_component [169814 gene products]
  GO:0003674: molecular_function [168558 gene products] &
    GO:0016209: abundant activity [566 gene products]
    GO:0015457: auxiliary transport protein activity [161 gene products]
    GO:0005486: binding [46697 gene products]
    GO:0003824: catalytic activity [51856 gene products]
    GO:0030188: chaperone regulator activity [73 gene products]
    GO:0042056: chemottractant activity [14 gene products]
    GO:0045499: chemorepellent activity [9 gene products]
    GO:0030234: enzyme regulator activity [2370 gene products]
    GO:0016530: metalloprotease activity [47 gene products]
    GO:0060089: molecular transducer activity [7873 gene products]
    GO:0003774: motor activity [527 gene products]
    GO:0045735: nutrient reservoir activity [49 gene products]
    GO:0031386: protein tag [18 gene products]
    GO:0005198: structural molecule activity [4324 gene products]
    GO:0030528: transcription regulator activity [110429 gene products]
    GO:0045182: translation regulator activity [893 gene products]
    GO:0005215: transporter activity [10593 gene products]
  
```

 Lister Hill National Center for Biomedical Communications 127

Gene Ontology Top level (CC)

```

all: all [250418 gene products] &
  GO:0008150: biological_process [166605 gene products]
  GO:0005575: cellular_component [169814 gene products] &
    GO:0005623: cell [111086 gene products]
    GO:0044664: cell part [111049 gene products]
    GO:0031975: envelope [3316 gene products]
    GO:0031012: extracellular matrix [573 gene products]
    GO:0044420: extracellular matrix part [292 gene products]
    GO:0005576: extracellular region [5001 gene products]
    GO:0044421: extracellular region part [3365 gene products]
    GO:0032991: macromolecular complex [14668 gene products]
    GO:0031974: membrane-enclosed lumen [5290 gene products]
    GO:0043226: organelle [79653 gene products]
    GO:0044422: organelle part [16645 gene products]
    GO:0055044: synplast [3 gene products]
    GO:0045202: synapse [454 gene products]
    GO:0044486: synapse part [210 gene products]
    GO:0019012: vision [227 gene products]
    GO:0044423: vision part [186 gene products]
    GO:0003674: molecular_function [168558 gene products]
  
```

 Lister Hill National Center for Biomedical Communications 128

Gene Ontology Top level (BP)

```

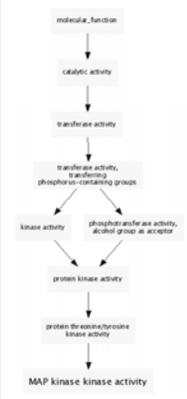
all: all [250418 gene products] &
  GO:0008150: biological_process [166605 gene products] &
    GO:0022610: biological adhesion [1586 gene products]
    GO:0065007: biological regulation [31031 gene products]
    GO:0001906: cell killing [177 gene products]
    GO:0009987: cellular process [79087 gene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

 Lister Hill National Center for Biomedical Communications 129

Gene Ontology Example

```

all: all [250418 gene products]
  GO:0003674: molecular_function [168558 gene products]
  GO:0003824: catalytic activity [51856 gene products]
  GO:0016740: transferase activity [15763 gene products]
  GO:0016772: transferase activity, transferring phosphorus-containing groups [products]
  GO:0016301: kinase activity [6093 gene products]
  GO:0004672: protein kinase activity [3504 gene products]
  GO:0004712: protein serine/threonine/tyrosine kinase activity
  GO:0004708: MAP kinase activity
  GO:0016773: phosphotransferase activity, alcohol
  GO:0004672: protein kinase activity [3504 gene products]
  GO:0004712: protein serine/threonine/tyrosine kinase activity
  GO:0004708: MAP kinase kinase activity
  
```



 Lister Hill National Center for Biomedical Communications 130

RxNorm

RxNorm Characteristics (1)

- ◆ Current version: June 7, 2010 (monthly releases)
- ◆ Type: Controlled terminology
- ◆ Domain: Drug names
- ◆ Developer: NLM
- ◆ Funding: NLM
- ◆ Availability
 - Publicly available: Yes*
 - Repositories: UMLS
- ◆ URL: <http://www.nlm.nih.gov/research/umls/rxnorm/>

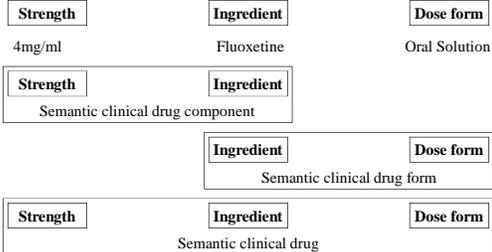
 Lister Hill National Center for Biomedical Communications 132

RxNorm Characteristics (2)

- ◆ Number of
 - Concepts: 166k
 - Terms: ~1 term per concept
- ◆ Major organizing principles:
 - Generic vs. brand
 - Combinations of Ingredient / Form / Dose
 - No hierarchical structure
 - Links to all major US drug information sources
 - No clinical information
- ◆ Formalism: UMLS RRF format

 Lister Hill National Center for Biomedical Communications 133

RxNorm Normalized form



The diagram illustrates the normalization process in three levels:

- Level 1:** Individual components: Strength (4mg/ml), Ingredient (Fluoxetine), and Dose form (Oral Solution).
- Level 2:** A box labeled "Semantic clinical drug component" containing Strength and Ingredient.
- Level 3:** A box labeled "Semantic clinical drug form" containing Ingredient and Dose form.
- Level 4:** A box labeled "Semantic clinical drug" containing Strength, Ingredient, and Dose form.

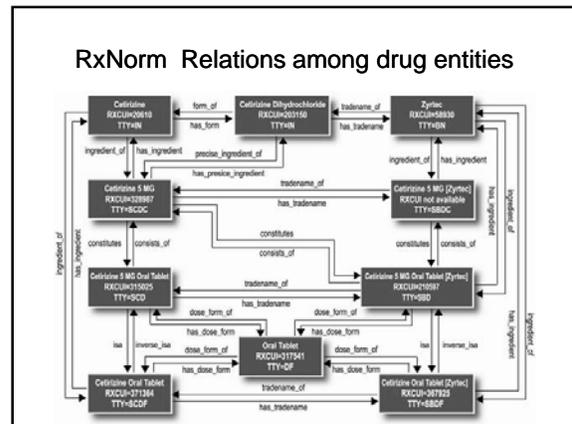
 Lister Hill National Center for Biomedical Communications 134

Rx Norm Generic vs. Brand

◆ Generic	◆ Brand
• Ingredient (IN)	• Brand name (BN)
• Clinical drug form (SCDF)	• Branded drug form (SBDF)
• Clinical drug component (SCDC)	• Branded drug component (SBDC)
• Clinical drug (SCD)	• Branded drug (SBD)

tradename_of

 Lister Hill National Center for Biomedical Communications 135



Medical Subject Headings (MeSH)



MeSH Characteristics (1)

- ◆ Current version: 2010 (yearly releases)
- ◆ Type: Thesaurus / Controlled vocabulary
- ◆ Domain: Biomedicine
- ◆ Developer: NLM
- ◆ Funding: NLM (Library Operations)
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS / NCBO BioPortal
- ◆ URL: <http://www.nlm.nih.gov/mesh/>

 Lister Hill National Center for Biomedical Communications 138

MeSH Characteristics (2)

- ◆ Number of
 - Concepts: 25,588 descriptors (2010)
 - Terms: 7.5 per descriptor
- ◆ Major organizing principles:
 - Descriptor + entry terms
(also: Qualifiers, Supplementary concepts)
 - Thesaurus relations (RB/RN/RO)
- ◆ Formalism: Thesaurus / Proprietary XML DTD



Lister Hill National Center for Biomedical Communications 139

MeSH Top level

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographicals [Z]



Lister Hill National Center for Biomedical Communications 140

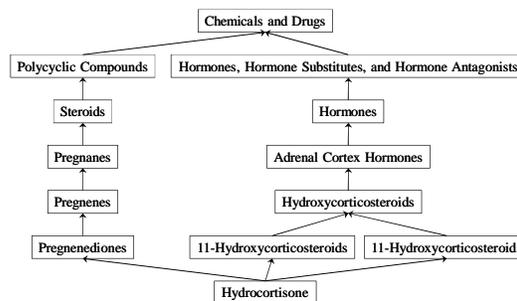
MeSH Example (terms)

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



Lister Hill National Center for Biomedical Communications 141

MeSH Example (hierarchies)



Lister Hill National Center for Biomedical Communications 142

NCI Thesaurus

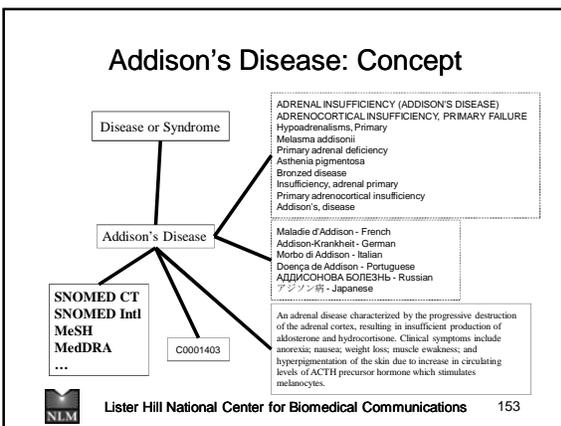
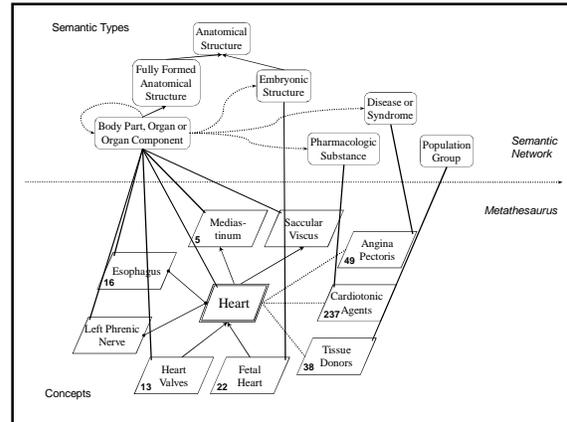
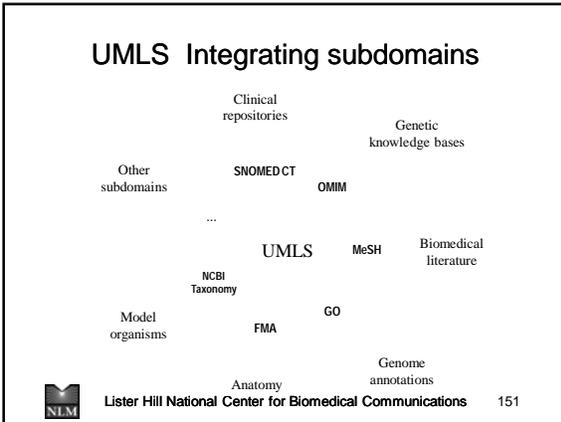


NCI thesaurus Characteristics (1)

- ◆ Current version: 10.05d (~monthly releases)
- ◆ Type: Controlled terminology / ontology
- ◆ Domain: Cancer
- ◆ Developer: NCI Center for Bioinformatics
- ◆ Funding: NCI
- ◆ Availability
 - Publicly available: Yes
 - Repositories: UMLS / OBO / NCBO BioPortal
- ◆ URL: <http://nciterms.nci.nih.gov/>



Lister Hill National Center for Biomedical Communications 144



Metathesaurus Concepts (2010AA)

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

A0066000	Headache (MeSH)
A0065992	Headache (ICD-10) S0046854
A0066007	Headaches (MedDRA)
A12003304	Headaches (OMIM) S0046855
L0018681	
A0540936	Cephalodynia (MeSH) S0475647
L0380797	
C0018681	

Lister Hill National Center for Biomedical Communications 154

Recap

Name	Scope	# concepts	Median	Subs. Hier	Version
SNOMED CT	Clinical medicine (patient records)	310,314	2	yes	July 31, 2007
LOINC	Clinical observations and laboratory tests	46,406	3	no	Version 2.21 (no "natural language" names)
FMA	Human anatomical structures	~72,000	?	yes	(not yet in the UMLS)
Gene Ontology	Functional annotation of gene products	22,546	1	yes	Jan. 2, 2007
ReNorm	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)

Lister Hill National Center for Biomedical Communications [Bodenreider, YBMI 2008]

PARTNERS HEALTHCARE Short course – Summer 2010
Clinical Ontology in Practice
June 15, 2010 – Session #4

Biomedical Ontologies in Action

A Functional Perspective on Biomedical Ontologies

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

Overview

- ◆ Functional perspective [Bodenreider, YBMI 2008]
 - What are they for (vs. what are they)?
- ◆ “High-impact” biomedical ontologies
- ◆ 3 major categories of use
 - Knowledge management (indexing and retrieval of data and information, access to information, mapping among ontologies)
 - Data integration, exchange and semantic interoperability
 - Decision support and reasoning (data selection and aggregation, decision support, natural language processing applications, knowledge discovery).

 Lister Hill National Center for Biomedical Communications 157

Knowledge management

Knowledge management

Annotating data and resources

Terminology in ontology

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

 Lister Hill National Center for Biomedical Communications 160

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)

 Lister Hill National Center for Biomedical Communications 161

GO Annotations for Aldh2 (mouse)

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

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

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

 Lister Hill National Center for Biomedical Communications 162

GO ALD4 in Yeast

GO Annotations

Molecular Function
Manually curated

Biological Process
Manually curated

Cellular Component
Manually curated

High-throughput

All **ALD4** GO evidence and references
View **Computational GO annotations for ALD4**

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

• ethanol metabolic process (IMP)

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

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

163

Lister Hill National Center for Biomedical Communications

GO Annotations for ALDH2 (Human)

GO ID	Function	inferred	IEA	IPR015590	UniProt	9606
GO:0016491	oxidoreductase activity	inferred	IEA	IPR016160	UniProt	9606
GO:0016491	oxidoreductase activity	inferred	IEA	IPR016162	UniProt	9606
GO:0016491	oxidoreductase activity	inferred	IEA	IPR016161	UniProt	9606
GO:0016491	oxidoreductase activity	inferred	IEA	IPR016161	UniProt	9606
GO:004029	aldehyde dehydrogenase (NAD) activity	inferred	IEA	IPR015590	UniProt	9606
GO:004030	aldehyde dehydrogenase [NAD(P)+] activity	inferred	IEA	IPR015590	UniProt	9606
GO:000955	electron carrier activity	inferred	IEA	IPR015590	UniProt	9606
GO:004029	aldehyde dehydrogenase (NAD) activity	inferred	IEA	IPR015590	UniProt	9606

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

Lister Hill National Center for Biomedical Communications

Indexing the biomedical literature

- ◆ MeSH
 - Used for indexing and retrieval of the biomedical literature (MEDLINE)
- ◆ Indexing
 - Performed manually by human indexers
 - With help of semi-automatic systems (suggestions) e.g., Indexing Initiative at NLM
 - Automatic indexing systems

165

Lister Hill National Center for Biomedical Communications

MeSH MEDLINE indexing

□ 1: *Anesth Analg*, 2008 Jun;106(6):1813-9

Free cortisol in sepsis and septic shock.

Rendell S, Karlsson S, Penttilä V, Lehto P, Varpala M, Ruokonen E, Fineran-Stabley G.

Department of Intensive Care, Kuopio University Hospital, FL 16222 Kuopio, Finland. Stepana.Rendell@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamic-pituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. **METHODS:** In a prospective study, severe sepsis was defined according to 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 electrochemoluminescence immunoassay. The Coles 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 ($r = 0.50, P < 0.001$). 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, $P = 0.44$). Non-survivors had higher calculated serum free 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.

165

MeSH MEDLINE indexing

MeSH Terms

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

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

Substances:

- Biological Markers
- Hydrocortisone

Lister Hill National Center for Biomedical Communications

MeSH MEDLINE indexing

□ 1: *Endocr Opin Investig Drugs*, 2008 Apr;17(4):497-509

Replacement therapy for Addison's disease: recent developments.

Leivás K, Husebye ES.

University of Bergen, Institute of Medicine, Section of Endocrinology, 5021 Bergen, Norway. Kristian.leivas@uhb.uib.no

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

167

MeSH MEDLINE indexing

MeSH Terms <ul style="list-style-type: none"> • Addison Disease • Addison Disease/blood • Addison Disease/drug therapy* • Addison Disease/administration & dosage* • Androgen/therapeutic use • Dosage Forms • Drug Administration Routes • Drug Administration Schedule • Glucocorticoids/administration & dosage* • Glucocorticoids/adverse effects • Glucocorticoids/blood • Glucocorticoids/deficiency • Hormone Replacement Therapy* • Human • Mineralocorticoids/administration & dosage* • Mineralocorticoids/adverse effects • Mineralocorticoids/blood • Mineralocorticoids/deficiency • Quality of Life • Treatment Outcome 	Substances <ul style="list-style-type: none"> • Androgens • Dosage Forms • Glucocorticoids • Mineralocorticoids
--	--

NLM Communications 169

ICD9-CM Coding clinical data

- ◆ ICD9-CM
 - Used for coding clinical data e.g., for billing purposes
- ◆ Other uses of ICD
 - Morbidity and mortality reporting worldwide



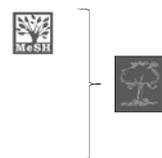
NLM Lister Hill National Center for Biomedical Communications 170

Knowledge management

Accessing biomedical information

Resources for biomedical search engines

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



NLM Lister Hill National Center for Biomedical Communications 172

MeSH “synonyms” MEDLINE retrieval

- ◆ MeSH entry terms
 - Used as equivalent terms for retrieval purposes
 - Not always synonymous
- ◆ Increase recall without hurting precision

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

NLM Lister Hill National Center for Biomedical Communications 173

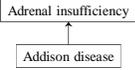
MeSH “synonyms” MEDLINE retrieval



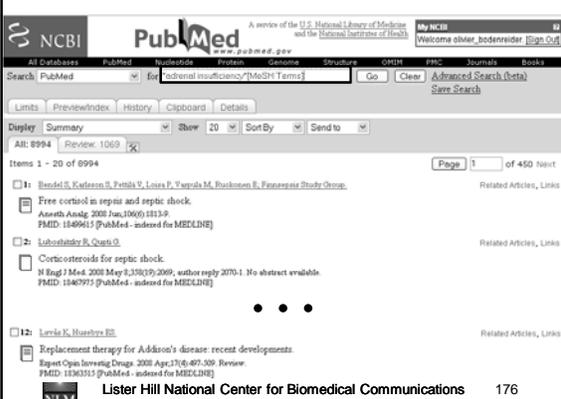
NLM Lister Hill National Center for Biomedical Communications 174

MeSH hierarchies MEDLINE retrieval

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



NLM Lister Hill National Center for Biomedical Communications 175

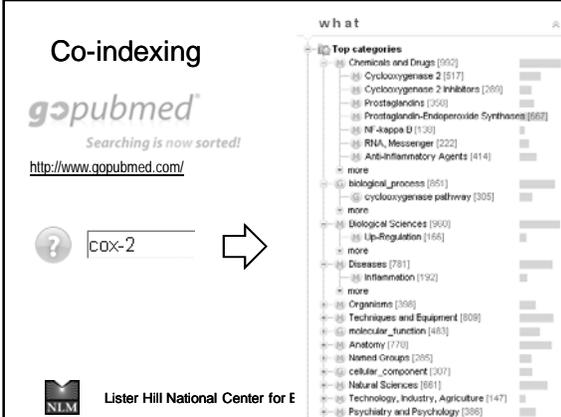


NLM Lister Hill National Center for Biomedical Communications 176

Co-indexing

gpubmed
Searching is now sorted!
<http://www.gpubmed.com/>

COX-2



NLM Lister Hill National Center for E 177

Knowledge management

Mapping across biomedical ontologies

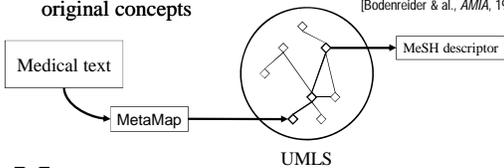
Reusing information

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

NLM Lister Hill National Center for Biomedical Communications 179

Reusing tools

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



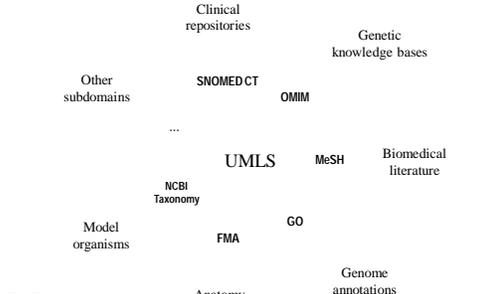
NLM Lister Hill National Center for Biomedical Communications 180

Terminology integration systems

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

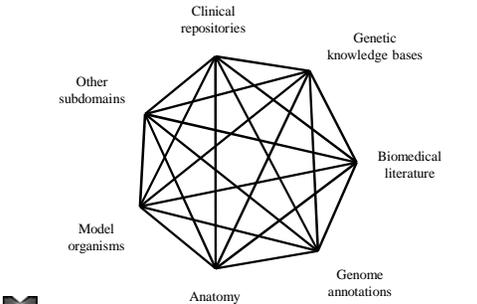
 Lister Hill National Center for Biomedical Communications 181

Integrating subdomains



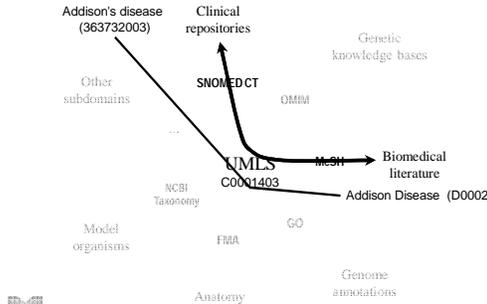
 Lister Hill National Center for Biomedical Communications 182

Integrating subdomains



 Lister Hill National Center for Biomedical Communications 183

Trans-namespace integration



 Lister Hill National Center for Biomedical Communications 184

Data integration, exchange and semantic interoperability

Data integration, exchange and semantic interoperability

Information exchange and semantic operability

“Standards”

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

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



Lister Hill National Center for Biomedical Communications 187

Semantic interoperability projects BRIDG

- ◆ **Biomedical Research Integrated Domain Group**
 - Information model for clinical research
 - Interoperability between clinical trials information systems
 - Ontologies provide value sets to the information model



Lister Hill National Center for Biomedical Communications 188

Semantic interoperability projects CDA

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



Lister Hill National Center for Biomedical Communications 189

Semantic interoperability projects caCORE

- ◆ **Cancer Common Ontologic Representation Environment**
 - Infrastructure developed to support an interoperable biomedical information system for cancer research
 - Uses the NCI Thesaurus as a component



Lister Hill National Center for Biomedical Communications 190

Data integration, exchange and semantic interoperability

Information and data integration

Approaches to data integration

<ul style="list-style-type: none">◆ Warehousing<ul style="list-style-type: none">● Sources to be integrated are transformed into a common format and converted to a common vocabulary● Normalization through ontologies (e.g., GO annotations)	<ul style="list-style-type: none">◆ Mediation<ul style="list-style-type: none">● Local schema (of the sources)● Global schema (in reference to which the queries are made)● Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)
--	---



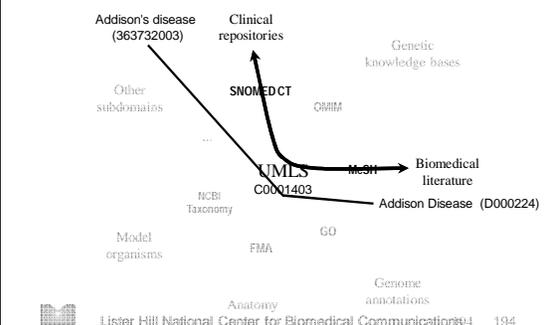
Lister Hill National Center for Biomedical Communications 192

Ontologies and integration

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

 Lister Hill National Center for Biomedical Communications 193

Trans-namespace integration



 Lister Hill National Center for Biomedical Communications 194

Decision support and reasoning

Data selection

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

 Lister Hill National Center for Biomedical Communications 196

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

 Lister Hill National Center for Biomedical Communications 197

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 NCI)

 Lister Hill National Center for Biomedical Communications 198

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



Lister Hill National Center for Biomedical Communications 199

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)



Lister Hill National Center for Biomedical Communications 200

Barriers to usability of biomedical ontologies

Availability

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



Lister Hill National Center for Biomedical Communications 202

Discoverability

- ◆ Ontology repositories
 - UMLS: 156 source vocabularies (biased towards healthcare applications)
 - NCBO BioPortal: ~200 ontologies (biased towards biological applications)
 - Some overlap between the two repositories
- ◆ Need for discovery services



Lister Hill National Center for Biomedical Communications 203

Formalism

- ◆ Several major formalism
 - Web Ontology Language (OWL) – NCI Thesaurus
 - OBO format – most OBO ontologies
 - UMLS Rich Release Format (RRF) – UMLS, RxNorm
- ◆ Conversion mechanisms
 - OBO to OWL
 - LexGrid (import/export to LexGrid internal format)



Lister Hill National Center for Biomedical Communications 204

Ontology integration

- ◆ *Post hoc* integration, form the bottom up
 - UMLS approach
 - Integrates ontologies “as is”, including legacy ontologies
 - Facilitates the integration of the corresponding datasets
 - Current harmonization efforts (e.g., IHTSDO)
- ◆ Coordinated development of ontologies
 - OBO Foundry approach
 - Ensures consistency *ab initio*
 - Excludes legacy ontologies



Lister Hill National Center for Biomedical Communications 205

Quality

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



Lister Hill National Center for Biomedical Communications 206



Short course – Summer 2010
Clinical Ontology in Practice

June 16, 2010 – “Hands-on” Sessions

Exploring Clinical Ontologies



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

UMLS UMLSKS

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



Lister Hill National Center for Biomedical Communications 208

UMLS UMLSKS

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



Lister Hill National Center for Biomedical Communications 209

UMLS Semantic Navigator

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



Lister Hill National Center for Biomedical Communications 210

UMLS SKS API

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



Lister Hill National Center for Biomedical Communications 211

SNOMED CT

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



Lister Hill National Center for Biomedical Communications 212

SNOMED CT

- ◆ Search concepts
 - Appendectomy (80146002)
 - Simvastatin (387584000)
 - Addison's disease (363732003)
- ◆ Notes
 - No post-coordination services in standard browsers
 - Some standalone browsers offer additional services (CliniClue, SNOB)
 - Search on Addison's disease in The SNOMED CT Browser © does not return any results



Lister Hill National Center for Biomedical Communications 213

LOINC

- ◆ Multiple web-based browsers available
 - RELMA (specific to LOINC) web version of a standalone application <http://loinc.org/relma>
NB: Citrix ICA Client required
 - BioPortal (LOINC 2.26) <http://www.bioontology.org/BioPortal>
 - NCI Term Browser (LOINC 2.24) <http://nciterms.nci.nih.gov/>



Lister Hill National Center for Biomedical Communications 214

LOINC BioPortal

- ◆ BioPortal
 - Graphical interface
 - Search for Lithium, then navigate down the tree
 - web services http://www.bioontology.org/wiki/index.php/NCBO_REST_services
 - Ontology Id: 1350
 - Get ID for latest version
 - <http://rest.bioontology.org/bioportal/virtual/ontology/1350>
 - Returns: 40400
 - Get the "first" 50 terms
 - <http://rest.bioontology.org/bioportal/concepts/40400/all?pagesize=50&pagenum=1>



Lister Hill National Center for Biomedical Communications 215

LOINC NCI Term Browser

- ◆ NCI Term Browser
 - Search for Lithium, then navigate through the Relationships tab
 - Search by code
- ◆ Search concept
 - Substance concentration of lithium in urine (quantitative)
 - Lithium:Substance Concentration:Point in time:Urine:Quantitative
 - 25463-1



Lister Hill National Center for Biomedical Communications 216

RxNorm RxNav

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



Lister Hill National Center for Biomedical Communications 217

RxNorm RxNav

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



Lister Hill National Center for Biomedical Communications 218

RxNorm SOAP API

- ◆ RxNorm SOAP API (demo client)
http://mor.nlm.nih.gov/perl/rxnav_api_demo.pl
- ◆ Functions
 - `getRxNormVersion()`
 - `getIdTypes()`
 - `findRxcuiById(00904582941, 309130) → 309130`
 - `getAllRelatedInfo(309130)`
- ◆ Documentation
<http://rxnav.nlm.nih.gov/RxNormAPI.html>



Lister Hill National Center for Biomedical Communications 219

RxNorm REST API

- ◆ Test resources
 - <http://rxnav.nlm.nih.gov/REST/spellingsuggestions?name=sulfametoaxol>
 - <http://rxnav.nlm.nih.gov/REST/rxcui/10180>
 - <http://rxnav.nlm.nih.gov/REST/rxcui/309130/properties>
 - <http://rxnav.nlm.nih.gov/REST/rxcui/309130/ndcs>
 - http://rxnav.nlm.nih.gov/REST/rxcui/151399/related?rela=tradename_of
- ◆ Documentation
http://rxnav.nlm.nih.gov/RxNorm_RESTful_UserGuide.pdf



Lister Hill National Center for Biomedical Communications 220

NDF-RT

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



Lister Hill National Center for Biomedical Communications 221



Short course – Summer 2010
Clinical Ontology in Practice

June 16, 2010 – Discussion Sessions

Issues and Challenges Related to Clinical Ontologies



Olivier Bodenreider

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

Topics

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



Lister Hill National Center for Biomedical Communications 223

Questions

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



Lister Hill National Center for Biomedical Communications 224



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

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