Clinical Ontology in Practice

June 15-17, 2010

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

◆ Learn about clinical ontologies
  ● History
  ● Design principles, formalisms and tools
  ● What are they?
  ● What are they used for?

◆ Work with clinical ontologies
  ● Search, browse, navigate, query with application programming interfaces
  ● Analyze, compare
  ● Specific clinical uses (e.g., decisions support, natural language processing, medication reconciliation, e-prescription)
  ● Specific issues (e.g., mapping across ontologies, ontologies and information models)
# Agenda

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

Lister Hill National Center for Biomedical Communications
References  Review articles

◆ Bodenreider O, Stevens R.
  Brief Bioinform. 2006 Sep;7(3):256-74.

◆ Cimino JJ, Zhu X.
  The practical impact of ontologies on biomedical informatics.

◆ Bodenreider O.
  Biomedical ontologies in action: role in knowledge management, data integration and decision support.
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
Introduction to Biomedical Ontologies

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

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

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

- **Hippocrates**
  - Dismisses superstition
  - Four humors
    - Blood
    - Phlegm
    - Yellow bile
    - Black bile

- **Thomas Sydenham (1624-1689)**
  - *Medical observations on the history and cure of acute diseases* (1676)
To classify diseases (and plants)

◆ Carolus Linnaeus (1707-1778)
  ● *Genera Plantarum* (1737)
  ● *Genera Morborum* (1763)

◆ François Boissier de La Croix
  a.k.a. F. B. de Sauvages (1706-1767)
  ● *Methodus Foliorum* (1751)
  ● *Nosologia Methodica* (1763/68)

◆ William Cullen (1710-1790)
  ● *Synopsis Nosologiae Methodicae* (1785)
From plants...
… to diseases

◆ Four categories (W. Cullen)
  ● Fevers
  ● Nervous disorders
  ● Cachexias
  ● Local diseases

“The distinction of the genera of diseases, the distinction of the species of each, and often even that of the varieties, I hold to be a necessary foundation of every plan of physic, whether dogmatical or empirical.”
– William Cullen, Edinburgh, 1785
Synopsis Nosologia Methodicae

(Cited by Chris Chute)
To support epidemiology

◆ John Graunt (1620-1674)
  ● Analyzes the vital statistics of the citizens of London

◆ William Farr (1807-1883)
  ● Medical statistician
  ● Improves Cullen’s classification
  ● Contributes to creating ICD

◆ Jacques Berthillon (1851-1922)
  ● Chief of the statistical services (Paris)
  ● Classification of causes of death (161 rubrics)
London Bills of Mortality

A generall Bill for this present year, ending the 19 of December 1664, according to the Report made to the KING's Most Excellent Majesty
By the Company of Parish Clerks of London, &c.

The diseases and Casualties this year:

<table>
<thead>
<tr>
<th>Disease</th>
<th>Aged</th>
<th>Flighted</th>
<th>Famine</th>
<th>Flood</th>
<th>Hunger</th>
<th>Scouring &amp; Flux</th>
<th>Spleen</th>
<th>Tumours &amp; Swellings</th>
<th>Wounds with Blows</th>
<th>Witches, Plagues &amp;c.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortion and Stillborn</td>
<td>617</td>
<td>1,235</td>
<td>8</td>
<td>1</td>
<td>3,548</td>
<td>12</td>
<td>17</td>
<td>11</td>
<td>8</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Aged</td>
<td>1,546</td>
<td>2,057</td>
<td>20</td>
<td>15</td>
<td>4,288</td>
<td>35</td>
<td>30</td>
<td>24</td>
<td>15</td>
<td>47</td>
<td>59</td>
</tr>
<tr>
<td>Aged and Fever</td>
<td>821</td>
<td>1,351</td>
<td>32</td>
<td>12</td>
<td>2,940</td>
<td>40</td>
<td>34</td>
<td>19</td>
<td>16</td>
<td>27</td>
<td>44</td>
</tr>
<tr>
<td>Appendicitis &amp; Sudden Death</td>
<td>12</td>
<td>45</td>
<td>11</td>
<td>11</td>
<td>116</td>
<td>16</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Famine</td>
<td>10</td>
<td>25</td>
<td>12</td>
<td>9</td>
<td>40</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Flighted</td>
<td>12</td>
<td>30</td>
<td>13</td>
<td>9</td>
<td>1,290</td>
<td>17</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Famine</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>15</td>
<td>70</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Flood</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Hunger</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Scouring &amp; Flux</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Spleen</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Tumours &amp; Swellings</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Wounds with Blows</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Witches, Plagues &amp;c.</td>
<td>8</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Lester 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.
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
Black bile and psychomotor retardation: shades of melancholia in Dante's Inferno.

Widmer DA.

Memorial Sloan-Kettering Cancer Center, New York, NY 10017, USA. widmerd@mskcc.org

The history of melancholy depression is rich with images of movement retardation and mental dysfunction. The recent restoration of psychomotor symptoms to the diagnostic terminology of affective disorder is not novel to the students of medieval melancholia. The move back to the biology of this psychomotor dysfunction with the technical advances in brain imaging in recent years only echoes centuries-old writings on the centrality of movement changes in the depressive condition. The Inferno, the first cantica of Dante Alighieri's Commedia, has a wonderful abundance of allusions to the importance of psychomotor symptoms in describing the depressed individual. Slowed steps, garbled speech, frozen tears, these and many other images keep the physical manifestations of psychomotor suffering in the forefront of the reader's mind. Considering Medieval and Renaissance writings on melancholy suffering, it is fitting that Dante shows a bodily illness reflected in the hellish torments visited on the damned. From the souls of the sullen to those of the violent, the panorama of psychomotor symptoms plays a prominent role in the poem as well as in the medical and literary prose of succeeding centuries.

MeSH Terms:
- Depressive Disorder/history*
- History of Medicine, Medieval
- Human
- Italy
- Literature, Medieval/history*
- Medicine in Literature*
- Poetry/history*
- Psychomotor Disorders/history*
Mouse Genome Database and GO

□ 1: Nf2  neurofibromatosis 2  [Mus musculus]
GeneID: 18016  Locus tag: MGI:97307

Entrez Gene

Gene Ontology

Function
- cytoskeletal protein binding
- protein binding
- structural molecule activity

Process
- intercellular junction assembly and/or maintenance
- negative regulation of cell cycle
- negative regulation of protein kinase activity
- regulation of cell proliferation

Component
- adherens junction
- cytoplasm
- cytoskeleton
- membrane

Evidence
IEA
IPI  PubMed
IEA
IMP  PubMed
IEA
IDA  PubMed
IMP  PubMed
IMP  PubMed
IEA
IEA
IEA

Lister Hill National Center for Biomedical Communications
To serve as a reference

◆ Reference terminology/ontology
  ● Universally needed
  ● Developed independently of any purposes
  ● Reusable by many applications

◆ Examples
  ● VA National Drug File (NDF)
  ● Foundational Model of Anatomy (FMA)
  ● SNOMED CT
Anatomy in Biomedicine

- Physiology
- Clinical medicine
- Biomedical literature
- Biomedical research
Administrative terminologies

◆ Coding patient records
  ● International Classification of Primary Care (ICPC)
  ● SNOMED
  ● Read Codes

◆ Reporting claims to health insurance companies
  ● International Classification of Diseases (ICD-9 CM)
  ● Healthcare Common Procedure Coding System (HCPCS)
Biomedical ontology in PubMed

Number of articles in PubMed/MEDLINE on Ontology vs. Controlled vocabulary

- Ontology or ontologies
- Both
- Controlled vocabulary (excluding DSM)

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

[Bodenreider, YBMI 2008]
Biomedical ontologies in PubMed

Proportion of citations in PubMed/MEDLINE by ontology

[Graph showing the proportion of citations in PubMed/MEDLINE by ontology from 1998 to 2007, with categories including GO, NCI Thesaurus, FMA, UMLS, SNOMED, MeSH, and LOINC.

Bodenreider, YBMI 2008]
Introduction to biomedical terminologies through an example
Guy’s Hospital, London
Thomas Addison (1795-1860)
Addison's disease

- Addison's disease is a rare endocrine disorder.
- Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol.
- For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism.
Adrenal insufficiency  Clinical variants

◆ Primary / Secondary
  ● Primary: lesion of the adrenal glands themselves
  ● Secondary: inadequate secretion of ACTH by the pituitary gland

◆ Acute / Chronic

◆ Isolated / Polyendocrine deficiency syndrome
Addison’s disease: Symptoms

- Fatigue
- Weakness
- Low blood pressure
- Pigmentation of the skin (exposed and non-exposed parts of the body)
- …
AD in medical vocabularies

◆ Synonyms: different terms
  ● Addisonian syndrome
  ● Bronzed disease
  ● Addison melanoderma
  ● Asthenia pigmentosa
  ● Primary adrenal deficiency
  ● Primary adrenal insufficiency
  ● Primary adrenocortical insufficiency
  ● Chronic adrenocortical insufficiency

◆ Contexts: different hierarchies
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 adenocortical overactivity
   Overproduction of ACTH, not associated with Cushing’s disease
   Premature adrenarche
   Excludes1: Cushing’s syndrome (E24.-)

E27.1 Primary adenocortical insufficiency
   Addison’s disease
   Adrenocortical insufficiency NOS
   Autoimmune adrenals
   Excludes1: Addison only phenotype adenoleukodystrophy (E71.428)
   amyloidosis (E85)
   tuberculous Addison’s disease (A18.7)
   Waterhouse-Friderichsen syndrome (A39.1)

E27.2 Addisonian crisis
   Adrenal crisis
   Adrenocortical crisis

E27.3 Drug-induced adenocortical insufficiency
   Code first (T36-T50) to identify drug

E27.4 Other and unspecified adenocortical insufficiency
Medical Subject Headings

MeSH Tree Structures

**Endocrine Diseases [C19]**

- Adrenal Gland Diseases [C19.053]
  - Adrenal Gland Hypofunction [C19.053.264]
    - Addison's Disease [C19.053.264.263]
    - Adrenoleukodystrophy [C19.053.264.270]
    - Hypoaldosteronism [C19.053.264.480]

**Immunologic Diseases [C20]**

- Autoimmune Diseases [C20.111]
  - Addison's Disease [C20.111.163]
    - Anemia, Hemolytic, Autoimmune [C20.111.175]
    - Anti-Glomerular Basement Membrane Disease [C20.111.190]
    - Antiphospholipid Syndrome [C20.111.197]
    - Arthritis, Rheumatoid [C20.111.199]
      - Autoimmune Diseases of the Nervous System [C20.111.258]
### SNOMED CT

**Hierarchy**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>386584007</td>
<td>adrenal cortical hypofunction</td>
</tr>
<tr>
<td>383732003</td>
<td>Addison's disease</td>
</tr>
<tr>
<td>237780009</td>
<td>Addison's disease with adrenoleucodystrophy</td>
</tr>
<tr>
<td>76715009</td>
<td>Addison's disease due to autoimmunity</td>
</tr>
<tr>
<td>186270000</td>
<td>tuberculous Addison's disease</td>
</tr>
<tr>
<td>11244009</td>
<td>polyglandular autoimmune syndrome, type 1</td>
</tr>
</tbody>
</table>

**Addison's disease - Definition**

- **Concept Status**: Current
- **Descriptions**
  - Addison's disease (disorder)
  - Addison's disease
  - enfermedad de Addison
  - enfermedad de Addison (trastorno)

**Definition: Primitive**

- is a
  - adrenal cortical hypofunction
  - finding site
  - adrenal cortex structure

**Qualifiers**

- severity
  - severities
  - episodicity
  - episodicties
  - clinical course
  - courses

**Codes**

- Original SnomedId: DB-70620
- Read Code (Ctv3Id): C1541
Biomedical terms as names for biomedical classes
Terms reflecting valid classes

- Pulmonary anthrax
- BRCA1 protein
- Coronary artery
- Coronary artery bypass
- ...
Issues

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

◆ Synonymy
  - Left coronary artery
  - LCA
  - Arteria coronaria sinistra
  - Addison’s disease
  - Primary adrenocortical insufficiency

◆ “Clinical synonymy” (vs. identity)
  - Abdominal swelling
  - Swollen abdomen
  - Posttransfusion hepatitis
  - Posttransfusion viral hepatitis
  - Addison’s disease
  - Primary adrenocortical insufficiency
  - vs. Waterhouse-Friderichsen Syndrome
Multiple classes for a term

- **Polysemy**
  - Cold
  - Common cold
  - Cold
  - Cold temperature
  - COLD
  - Chronic Obstructive Airway Disease

- **Truncated terms**
  - Calcium
  - Ca++
  - Coagulation factor IV
  - Calcium
  - Calcium measurement
Non-ontological features in terms

● Epistemological features

  ▪ Gallbladder calculus without mention of cholecystitis
  ▪ Diarrhea of presumed infectious origin
  ▪ Replacement of unspecified heart valve
  ▪ ...

Ontology vs. Epistemology

- **Ontology**
  - Invariants in reality
    - Classes (universals)
    - Relations between them
  - Theory of reality

- **Epistemology**
  - Knowledge about such entities
  - Perception of reality

Bone metastasis

*diagnosed by CT scan*

Bone metastasis

*diagnosed by Tc99m bone scintiscan*
Composite terms

◆ Sentence-like terms
  ● Several classes and their relations
  ● May contain epistemological features

  ▪ Tuberculosis of adrenal glands, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture
More composite terms

- Nontraffic accident involving being accidentally pushed from motor vehicle, except off-road motor vehicle, while in motion, not on public highway, driver of motor vehicle injured

- Determine whether the elder patient and caretaker have a functional social support network to assist the patient in performing activities of daily living and in obtaining health care, transportation, therapy, medications, community resource information, financial advice, and assistance with personal problems

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

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

Foundational Model of Anatomy

- Physical anatomical entity
  - Material physical anatomical entity
    - Anatomical structure
    - Inherent 3D shape
    - Body substance
  - Non-material physical anatomical entity
    - Inherent 3D shape
    - Mass

- Non-physical anatomical entity
  - Spatial dimension
  - 2D
  - 1D
  - 0D

Lister Hill National Center for Biomedical Communications
No explicit classificatory principle

3. **Diseases**

- Bacterial Infections and Mycoses [C01] +
- Virus Diseases [C02] +
- Parasitic Diseases [C03] +
- Neoplasms [C04] +
- Musculoskeletal Diseases [C05] +
- Digestive System Diseases [C06] +
- Stomatognathic Diseases [C07] +
- Respiratory Tract Diseases [C08] +
- Otorhinolaryngologic Diseases [C09] +
- Nervous System Diseases [C10] +
- Eye Diseases [C11] +
- Urologic and Male Genital Diseases [C12] +
- Female Genital Diseases and Pregnancy Complications [C13] +
- Cardiovascular Diseases [C14] +
- Hemic and Lymphatic Diseases [C15] +
- Neonatal Diseases and Abnormalities [C16] +
- Skin and Connective Tissue Diseases [C17] +
- Nutritional and Metabolic Diseases [C18] +
- Endocrine Diseases [C19] +
- Immunologic Diseases [C20] +
- Disorders of Environmental Origin [C21] +
- Animal Diseases [C22] +
- Pathological Conditions, Signs and Symptoms [C23] +
1. Certain infectious and parasitic diseases
2. Neoplasms
3. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4. Endocrine, nutritional, and metabolic diseases
5. Mental and behavioral disorders
6. Diseases of nervous system
7. Diseases of the eye and adnexa
8. Diseases of the ear and mastoid process
9. Diseases of circulatory system
10. Diseases of respiratory system
11. Diseases of digestive system
12. Diseases of the skin and subcutaneous tissue
13. Diseases of the musculoskeletal system and connective tissue
14. Diseases of the genitourinary system
15. Pregnancy, childbirth, and the puerperium
16. Certain conditions originating in the newborn (perinatal) period
17. Congenital malformations, deformations and chromosomal abnormalities
18. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19. Injury, poisoning and certain other consequences of external causes
20. External causes of morbidity
21. Factors influencing health status and contact with health service
• Attribute
• Body structure
• Clinical finding
• Context-dependent categories
• Environments and geographical locations
• Events
• Observable entity
• Organism
• Pharmaceutical / biologic product
• Physical force
• Physical object
• Procedure
• Qualifier value
• Social context
• Special concept
• Specimen
• Staging and scales
• Substance
Fully specified relations

Viral meningitis in SNOMED CT

Fully defined by ...

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

Diseases

- CNS diseases
- Virus diseases

CNS infections

- Meningitis
- CNS viral diseases

Viral meningitis

Diagram:

- Diseases
  - CNS diseases
  - Virus diseases
- CNS infections
  - Meningitis
  - CNS viral diseases
  - Viral meningitis

Questions:

- What is the relationship between CNS diseases and Virus diseases?
- What are the subcategories of CNS infections?
- Is there a specific condition under CNS viral diseases?
- Can we determine if Viral meningitis is a child of CNS infections?
Thesaurus relations

- Addison’s disease
  - Due to auto-immunity in 80% of the cases
  - Other causes include tuberculosis

Relations used to create hierarchical structures vs. hierarchical relations
Endocrine Diseases [C19]

Adrenal Gland Diseases [C19.053]
- Adrenal Gland Hypofunction [C19.053.264]
  - Addison's Disease [C19.053.264.263]
- Adrenoleukodystrophy [C19.053.264.270]
- Hypoaldosteronism [C19.053.264.480]

Immunologic Diseases [C20]

Autoimmune Diseases [C20.111]
- Addison's Disease [C20.111.163]
  - Anemia, Hemolytic, Autoimmune [C20.111.175]
  - Anti-Glomerular Basement Membrane Disease [C20.111.190]
  - Antiphospholipid Syndrome [C20.111.197]
  - Arthritis, Rheumatoid [C20.111.199] +
Accidents in MeSH

Environment and Public Health [G03]
  Public Health [G03.850]
  Accidents [G03.850.110]
    Accident Prevention [G03.850.110.060] +
    Accidental Falls [G03.850.110.085]
    Accidents, Aviation [G03.850.110.185]
    Accidents, Home [G03.850.110.205]
    Accidents, Occupational [G03.850.110.250] +
    Accidents, Radiation [G03.850.110.285]
    Accidents, Traffic [G03.850.110.320]
    Drowning [G03.850.110.500] +
Limited depth in hierarchies “by design”

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

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

E84  Cystic fibrosis
    Includes: mucoviscidosis
    E84.0  Cystic fibrosis with pulmonary manifestations
           Use additional code to identify any infectious organism present, such as:
           Pseudomonas (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
Conclusions
Conclusions 😞

◆ Biomedical terms
  ● reflect some aspects of biomedical reality
    ■ Although the primary concern of terminology is naming, not reflecting reality
  ● often convey additional features (e.g., epistemology)

◆ Biomedical terminology tends to offset part of the complexity
  ● but often reflects utility
Conclusions 😊

- Biomedical terminologies can help populate biomedical ontologies

- Resources needed
  - Linguistic analysis of terms
  - Statistical analysis of terms in a corpus / annotation database (dependence relations)
  - Manual curation
Design Principles, Formalisms and Tools for Biomedical Ontologies

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

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

Concept

Symbol

Object

*Ogden-Richards*
Definitions

◆ The *What* question
  - Objects in the world
    - With their properties
    - With their relations to other objects
  - Also: events, processes, and states

◆ Explicit specification of a conceptualization
  - Support software applications

◆ Domain ontology reflects
  - Underlying reality
  - Theory of the domain
Examples of use

- Natural language processing
- Access to heterogeneous sources of information (e.g., Semantic Web)
- Systems engineering
- Interoperability
Ontology vs. other artifacts

◆ Ontology
  ● Defining types of things and their relations

◆ Terminology
  ● Naming things in a domain

◆ Thesaurus
  ● Organizing things for a given purpose

◆ Classification
  ● Placing things into (arbitrary) classes

◆ Knowledge bases
  ● Assertional knowledge

[Smith, KR-MED 2006]
[Chute, JAMIA 2000]
(Controlled) Terminology

◆ Objective: naming things
◆ Shared understanding
  ● Agreement on what terms to use
  ● Utility-driven (arbitrary)

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

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

Autoimmune Diseases

Addison’s disease

Tuberculous Addison’s disease

Addison’s disease due to autoimmunity

is generally a
Classification

◆ Objective: placing things into classes

◆ Characteristics
  • Single inheritance (tree)
  • Idiosyncratic inclusion/exclusion criteria

---

Insulin-dependent diabetes mellitus

[See before E10 for subdivisions.]

Includes: diabetes (mellitus):
  • brittle
  • juvenile-onset
  • ketosis-prone
  • type 1

Excludes: diabetes mellitus (in):
  • malnutrition-related (E12-)
  • neonatal (P70.2)
  • pregnancy, childbirth and the puerperium (O24-)
  glycocuria:
  • NOS (R81)
  • renal (E74.8)
  impaired glucose tolerance (R73.0)
  postsurgical hypoinsulinaemia (E89.1)
Classification

◆ Characteristics (continued)

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

E84  Cystic fibrosis
  Includes: mucoviscidosis
E84.0 Cystic fibrosis with pulmonary manifestations
E84.1 Cystic fibrosis with intestinal manifestations
  Meconium ileus+ (P75*)
  Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)
E84.8 Cystic fibrosis with other manifestations
  Cystic fibrosis with combined manifestations
E84.9 Cystic fibrosis, unspecified
Knowledge Bases

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

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

◆ Lexical resources
  • Collections of lexical items
  • Additional information
    - Part of speech
    - Spelling variants
  • Useful for entity recognition
  • UMLS SPECIALIST Lexicon, WordNet

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

◆ Terminological resources
  • Collections lexical items + identifiers
    • Useful for entity resolution
    • UMLS Metathesaurus
The Knowledge Semantics Continuum

Ontology Dimensions based on McGuinness and Finin

Simple Terminologies

Expressive Ontologies

Medication Lists
DDI Lists

Catalog
Terms/glossary

Thesauri:
BT/NT,
Parent/Child,
Informal Is-A

KeGG

DB Schema

Formal is-a
Frames
(Properties)

RDF(S)

Formal instances

Ontylog

Value Restriction

Snomed

Disjointness,
Inverse

TAMBIS

BioPAX

BioPAX

CyC

MeSH,
Gene Ontology,
UMLS Meta

IEEE SUO

General Logical
constraints

EcoCyc
Kinds of ontologies

- Upper Level Ontology
- General Ontology
- Domain Ontology
- Application ontologies
Ontology-related issues

- Creation
- Merging
- Alignment
- Validation
Formal Ontological Principles
Formal ontological distinctions

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

- **Universal** = *category*
- **Synonyms**
  - Kind, Type, (Class)
- **Examples**
  - eyeball
  - blood pressure
  - conference

- **Individual** = *instance*
- **Synonyms**
  - Particular, Token
- **Examples**
  - my right eyeball
  - my blood pressure (132/79)
  - AMIA Annual Symposium 2003

*instantiation*
Continuant vs. Occurrent

- **Continuant** = *Continues to exist through time*
- **Synonyms**
  - Substance
- **Examples**
  - tennis racquet
  - mitochondrion
  - insulin production

- **Occurrent** = *Unfolds through time*
- **Synonyms**
  - Process
- **Examples**
  - tennis tournament
  - metabolism
  - producing insulin
Independent vs. Dependent

- Independent = *Can exist without support from other entities*
  - Examples
    - virus
    - molecule
    - plant

- Dependent = *Require support from other entities for its existence*
  - Examples
    - viral infection
    - DNA binding
    - food
Formal ontology Upper level

Universals
(classes)

Particulars
(instances)

Independent continuant
Dependent continuant

Continuant

Occurent

Thing
Formal ontological distinctions

◆ Basic distinctions in many top-level ontologies
  ● Generic: BFO, DOLCE
  ● Biomedical: BioTop, UMLS Semantic Network

◆ Condition the relations between various types of entities
  ● Relations
    ■ Between instances (e.g., part_of [at time])
    ■ Between classes (e.g., isa, part_of [atemporal])
    ■ Between one instance and one class (instance_of)

[Smith, Genome Biology 2005]
Formal ontology in practice

- Provides foundational classes and relations
  - Upper level ontologies
  - Relation ontology
- Provides a framework for analyzing entities and relations
Examples
General ontologies

◆ OpenCyc
  ● General ontology
  ● Cycorp, Inc (D. Lenat & al.)
  ● Over 1M hand-coded assertions
  ● http://www.opencyc.org

◆ WordNet
  ● Electronic lexical database
  ● Princeton University (G. Miller & al.)
  ● Over 100,000 synsets
  ● http://wordnet.princeton.edu/
Top level in OpenCyc

- **Thing**
  - **Intangible**
    - **Mathematical or computational thing**
      - **Set or collection**
        - **Collection**
  - **Individual**
  - **PartiallyTangible**
    - **Tangible thing**

Symbols:
- -$\text{genls}$
- -$\text{isa}$
Top level in WordNet

- Abstraction
- Activity
- Entity
- Event
- Group
- Location
- Natural phenomenon
- Possession
- Psychological feature
- Shape
- State
◆ Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in Medicine
◆ European Union project (A. Rector & al.)
◆ Over 25,000 concepts (primitives)
◆ http://www.opengalen.org
Top level in GALEN

DomainCategory

Phenomenon

Arbitrarily Conjuncted Phenomenon

ModifierConcept

Generalised Structure

Generalised Process

Generalised Substance

SubstanceOr PhysicalStructure

Collection

Aspect

Feature

StateOr Quantity

Status

Selector

Unit

Role

Modality

GeneralLevel OfSpecification
UMLS Semantic Network

- Definitional knowledge in the biomedical domain
- NLM (A. McCray & al.)

Content
- 133 semantic types
- 54 types of relationship
- 6700 semantic relations

Top level in the Semantic Network

```
Root

Entity
  - Conceptual Entity
  - Physical object

Event
  - Activity
  - Phenomenon or Process
```
Differences between ontologies

Examples
Granularity, plesionymy

UMLS

Epilepsy, Generalized
Seizure Disorder, Generalized
[...]

Epilepsy, Grand Mal
Tonic-Clonic Epilepsy
Seizure Disorder, Tonic Clonic
[...]

WordNet

generalized epilepsy
grand mal epilepsy
Differing categorization

UMLS

Natural Phenomenon or Process
  Biologic Function
  Pathologic Function
  Disease or Syndrome
    Dental Caries
    Dental cavity, NOS
    Tooth caries
    Dental Decay
    [...] 

WordNet

phenomenon
  process
  natural process
  decay
  cavity caries
dental caries
tooth decay

Dental Caries

Health disorder

Lister Hill National Center for Biomedical Communications
Formalisms and Tools
Ontology and Formalism

- Frames
- Description logics
  - OWL DL
- First-order logic

- OBO Format
  - Conversion to OWL DL
Tools for ontology developers

◆ Protégé
  ● Publicly available
  ● Frames and DL
  ● Classifier
  ● Supports many file formats (import/export)
  ● Large community of users
  ● Well supported

◆ OBO-Edit
  ● Specific to the biomedical/OBO community
  ● Simpler than Protégé ("tool for biologists")
  ● [http://oboedit.org/](http://oboedit.org/)

Lister Hill National Center for Biomedical Communications
"High-Impact" Biomedical Ontologies

A Structural Perspective

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

◆ Structural perspective [J. Cimino, YBMI 2006]
  - What are they (vs. what are they for)?
◆ “High-impact” biomedical ontologies
  - International Classification of Diseases (ICD)
  - Logical Observation Identifiers, Names and Codes (LOINC)
  - SNOMED Clinical Terms
  - Foundational Model of Anatomy
  - Gene Ontology
  - RxNorm
  - Medical Subject Headings (MeSH)
  - NCI Thesaurus
  - Unified Medical Language System (UMLS)
International Classification of Diseases
ICD Characteristics (1)

- Current version: ICD-10
- Type: Classification
- Domain: Disorders
- Developer: World Health Organization (WHO)
- Funding: WHO
- Availability
  - Publicly available: No
  - Repositories: UMLS  [ICD9-CM in NCBO BioPortal]
- URL: http://www.who.int/classifications/icd/en/
ICD Characteristics (2)

◆ Number of
  ● Concepts: 12,318
  ● Terms: 1 per concept (tabular)

◆ Major organizing principles:
  ● Tree (single inheritance hierarchy)
  ● No explicit classification criteria
    ■ Idiosyncratic inclusion/exclusion mechanism
  ● .8 slots for Not elsewhere classified (NEC)
  ● .9 slots for Not otherwise specified (NOS)

◆ Formalism: Proprietary format
## ICD Top level

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

◆ Idiosyncratic inclusion/exclusion criteria

**E10**

**Insulin-dependent diabetes mellitus**

[See before E10 for subdivisions.]

**Includes:** diabetes (mellitus):
- brittle
- juvenile-onset
- ketosis-prone
- type I

**Excludes:** diabetes mellitus (in):
- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)
glycosuria:
- NOS (R81.)
- renal (E74.8)
impaired glucose tolerance (R73.0)
postsurgical hypoinsulinaemia (E89.1)
ICD Example

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

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E84</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td></td>
<td><em>Includes</em>: mucoviscidosis</td>
</tr>
<tr>
<td>E84.0</td>
<td>Cystic fibrosis with pulmonary manifestations</td>
</tr>
<tr>
<td>E84.1</td>
<td>Cystic fibrosis with intestinal manifestations</td>
</tr>
<tr>
<td></td>
<td>Meconium ileus+ (P75*)</td>
</tr>
<tr>
<td></td>
<td><em>Excludes</em>: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)</td>
</tr>
<tr>
<td>E84.8</td>
<td>Cystic fibrosis with other manifestations</td>
</tr>
<tr>
<td>E84.9</td>
<td>Cystic fibrosis, unspecified</td>
</tr>
</tbody>
</table>
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
LOINC Characteristics (2)

◆ Number of
  ● Concepts: 50k active codes (2.18)
    (2 annual releases)
  ● Terms: n/a*

◆ Major organizing principles:
  ● No hierarchical structure among the main codes
  ● 6 axes
    ■ Component (analyte [+ challenge] [+ adjustments])
    ■ Property
    ■ Timing
    ■ System
    ■ Scale
    ■ [Method]

◆ Formalism: “DL-like”
LOINC Example

- Sodium:SCnc:-Pt:Ser/Plas:Qn
  [the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

<table>
<thead>
<tr>
<th>Axis</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>SCnc – Substance Concentration (per volume)</td>
</tr>
<tr>
<td>Timing</td>
<td>Pt – Point in time (Random)</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas – Serum or Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn – Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>--</td>
</tr>
</tbody>
</table>
SNOMED Clinical Terms
SNOMED CT Characteristics (1)

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

- Number of
  - Concepts: ~310,000 active concepts (Jan. 31, 2010)
  - Terms: ~800,000 active “descriptions”

- Major organizing principles:
  - Utility for clinical medicine (e.g., assertional + definitional knowledge)
  - Model of meaning (incomplete)
  - Rich set of associative relationships
  - Small proportion of defined concepts (many primitives)

- Formalism: Description logics (KRSS)
### SNOMED CT Top level

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Subtype hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>138875005</td>
<td>SNOMED CT Concept</td>
</tr>
<tr>
<td>362981000</td>
<td>qualifier value</td>
</tr>
<tr>
<td>106237007</td>
<td>linkage concept</td>
</tr>
<tr>
<td>370115009</td>
<td>special concept</td>
</tr>
<tr>
<td>48176007</td>
<td>social context</td>
</tr>
<tr>
<td>419891008</td>
<td>record artifact</td>
</tr>
<tr>
<td>363787002</td>
<td>observable entity</td>
</tr>
<tr>
<td>308916002</td>
<td>environment or geographical location</td>
</tr>
<tr>
<td>123038009</td>
<td>specimen</td>
</tr>
<tr>
<td>254291000</td>
<td>staging and scales</td>
</tr>
<tr>
<td>123037004</td>
<td>body structure</td>
</tr>
<tr>
<td>272379006</td>
<td>event</td>
</tr>
<tr>
<td>78621006</td>
<td>physical force</td>
</tr>
<tr>
<td>404684003</td>
<td>clinical finding</td>
</tr>
<tr>
<td>260787004</td>
<td>physical object</td>
</tr>
<tr>
<td>410607006</td>
<td>organism</td>
</tr>
<tr>
<td>71388002</td>
<td>procedure</td>
</tr>
<tr>
<td>373873005</td>
<td>pharmaceutical / biologic product</td>
</tr>
<tr>
<td>243796009</td>
<td>situation with explicit context</td>
</tr>
<tr>
<td>105590001</td>
<td>substance</td>
</tr>
</tbody>
</table>
SNOMED CT Example
Foundational Model of Anatomy
FMA Characteristics (1)

◆ Current version: ? (no fixed release schedule)
◆ Type: Ontology
◆ Domain: Anatomy (anatomical structures)
◆ Developer: U. Washington, Department of Biological Structure
◆ Funding: NLM (grants and contract) and others
◆ Availability
  ● Publicly available: Yes
  ● Repositories: [UMLS] / OBO / NCBO BioPortal
◆ URL: http://fma.biostr.washington.edu/
FMA Characteristics (2)

◆ Number of
  ● Concepts: ~72k
  ● Terms: ~1.5 / concept

◆ Major organizing principles:
  ● Explicit classificatory criteria
  ● Distinct isa and part_of hierarchies
  ● Additional spatial relations (e.g., adjacency)
  ● Multiple levels of granularity (organism to sub-cellular)

◆ Formalism: Frames (Protégé)
  ● Conversion to OWL Full and OWL DL available
FMA Example

(Courtesy of C. Rosse)
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/
Gene Ontology  Characteristics (2)

- **Number of**
  - Concepts: 27,800 (July 22, 2009)
  - Terms: 2.15 per concept

- **Major organizing principles:**
  - 3 major hierarchies
    - Molecular function
    - Cellular component
    - Biological process
  - Relations (within hierarchies): *isa, part_of, regulates*
  - No relations between concepts across hierarchies

- **Formalism: OBO format**
Gene Ontology  Top level (MF)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
  - GO:0005575 : cellular_component [169814 gene products]
  - GO:0003674 : molecular_function [168550 gene products]
    - GO:0016209 : antioxidant activity [565 gene products]
    - GO:0015457 : auxiliary transport protein activity [161 gene products]
    - GO:0005488 : binding [46697 gene products]
    - GO:0003824 : catalytic activity [51856 gene products]
    - GO:0030188 : chaperone regulator activity [73 gene products]
    - GO:0042056 : chemoattractant activity [14 gene products]
    - GO:0045499 : chemorepellent activity [9 gene products]
    - GO:0030234 : enzyme regulator activity [2370 gene products]
    - GO:0016530 : metallochaperone activity [47 gene products]
    - GO:0060009 : molecular transducer activity [7873 gene products]
    - GO:0037774 : 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 [10429 gene products]
    - GO:0045182 : translation regulator activity [893 gene products]
    - GO:0005215 : transporter activity [10583 gene products]
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:0044464 : 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:0043225 : organelle [79653 gene products]
    - GO:0044422 : organelle part [16645 gene products]
    - GO:0055044 : symplast [3 gene products]
    - GO:0045202 : synapse [454 gene products]
    - GO:0044456 : synapse part [210 gene products]
    - GO:0019012 : virion [227 gene products]
    - GO:0044423 : virion part [186 gene products]
    - GO:0003674 : molecular_function [168558 gene products]
Gene Ontology  Top level (BP)

- all : all  [250418 gene products]
  - GO:0008150 : biological_process  [166605 gene products]
    - GO:0022610 : biological adhesion  [1585 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  [5343 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]
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
    - 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 kinase activity
      - GO:0016773: phosphotransferase activity, alcohol group as acceptor
  - 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
RxNorm
RxNorm Characteristics (1)

- Current version: June 7, 2010 (monthly releases)
- Type: Controlled terminology
- Domain: Drug names
- Developer: NLM
- Funding: NLM
- Availability
  - Publicly available: Yes*
  - Repositories: UMLS
- URL: http://www.nlm.nih.gov/research/umls/rxnorm/
RxNorm Characteristics (2)

◆ Number of
  • Concepts: 166k
  • Terms: ~1 term per concept

◆ Major organizing principles:
  • Generic vs. brand
  • Combinations of Ingredient / Form / Dose
  • No hierarchical structure
  • Links to all major US drug information sources
  • No clinical information

◆ Formalism: UMLS RRF format
RxNorm  Normalized form

- **Strength**: 4mg/ml
- **Ingredient**: Fluoxetine
- **Dose form**: Oral Solution

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug

Lister Hill National Center for Biomedical Communications
Rx Norm  Generic vs. Brand

◆ Generic
  ● Ingredient
    (IN)
  ● Clinical drug form
    (SCDF)
  ● Clinical drug component
    (SCDC)
  ● Clinical drug
    (SCD)

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

\textit{tradename}\_of
RxNorm Relations among drug entities
Medical Subject Headings (MeSH)
MeSH Characteristics (1)

- Current version: 2010 (yearly releases)
- Type: Thesaurus / Controlled vocabulary
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (Library Operations)
- Availability
  - Publicly available: Yes
  - Repositories: UMLS / NCBO BioPortal
- URL: http://www.nlm.nih.gov/mesh/
MeSH Characteristics (2)

◆ Number of
  ● Terms: 7.5 per descriptor

◆ Major organizing principles:
  ● Descriptor + entry terms
    (also: Qualifiers, Supplementary concepts)
  ● Thesaurus relations (RB/RN/RO)

◆ Formalism: Thesaurus / Proprietary XML DTD
MeSH  Top level

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

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Hydrocortisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>D04.808.745.745.654.600</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.353.476</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.478.392</td>
</tr>
<tr>
<td>Scope Note</td>
<td>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.</td>
</tr>
<tr>
<td>Entry Term</td>
<td>11-Epicortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortifair</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortal</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Epicortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (11 alpha)-Isomer</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (9 beta,10 alpha,11 alpha)-Isomer</td>
</tr>
</tbody>
</table>
MeSH Example (hierarchies)

- Polycyclic Compounds
  - Steroids
    - Pregnanes
      - Pregnenes
        - Pregnenediones
  - Hormones, Hormone Substitutes, and Hormone Antagonists
    - Hormones
      - Adrenal Cortex Hormones
        - Hydroxycorticosteroids
          - 11-Hydroxycorticosteroids
  - Hormones
    - Hydroxycorticosteroids
      - Hydrocortisone
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/
NCI thesaurus Characteristics (2)

◆ Number of
  ● Concepts: ~60,000
  ● Terms: 2.68 per concept

◆ Major organizing principles:
  ● Subsumption hierarchy
  ● Rich set of associative relationships
  ● Small proportion of defined concepts (many primitives)
  ● Links to many external resources

◆ Formalism: OWL Lite
NCI thesaurus  Top level

NCI_Thesaurus Taxonomy

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
- Biochemical Pathway
- Biological Process
- Chemotherapy Regimen or Agent Combination
- Conceptual Entity
- Diagnostic, Therapeutic, and Research Equipment
- Diagnostic or Prognostic Factor
- Disease, Disorder or Finding
- Drug, Food, Chemical or Biomedical Material
- Experimental Organism Anatomical Concept
- Experimental Organism Diagnosis
- Gene
- Gene Product
- Molecular Abnormality
- NCI Administrative Concept
- Organism
- Property or Attribute
- Retired Concept
NCI thesaurus Example
Unified Medical Language System (UMLS)
UMLS Characteristics (1)

- Current version: 2010AA (2 annual releases)
- Type: Terminology integration system
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (intramural)

**Availability**
- Publicly available: Yes* (cost-free license required)
- Repositories: UMLS

UMLS Characteristics (2)

◆ Number of
  - Concepts: 2.2M (2010AA)
  - Terms: ~10M

◆ Major organizing principles (Metathesaurus):
  - Concept orientation
  - Source transparency
  - Multi-lingual through translation

◆ Formalism: Proprietary format (RRF)
UMLS Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

UMLS

MeSH

Biomedical literature

NCBI Taxonomy

Model organisms

UMLS

FMA

Genome annotations

Anatomy

Lister Hill National Center for Biomedical Communications
Addison’s Disease: Concept

An adrenal disease characterized by the progressive destruction of the adrenal cortex, resulting in insufficient production of aldosterone and hydrocortisone. Clinical symptoms include anorexia; nausea; weight loss; muscle weakness; and hyperpigmentation of the skin due to increase in circulating levels of ACTH precursor hormone which stimulates melanocytes.
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

<table>
<thead>
<tr>
<th>CUI</th>
<th>Concept</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0066000</td>
<td>Headache</td>
<td>MeSH</td>
</tr>
<tr>
<td>A0065992</td>
<td>Headache</td>
<td>ICD-10</td>
</tr>
<tr>
<td>S0046854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0066007</td>
<td>Headaches</td>
<td>MedDRA</td>
</tr>
<tr>
<td>A12003304</td>
<td>Headaches</td>
<td>OMIM</td>
</tr>
<tr>
<td>S0046855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0018681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0540936</td>
<td>Cephalodynia</td>
<td>MeSH</td>
</tr>
<tr>
<td>S0475647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0380797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0018681</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Recap

<table>
<thead>
<tr>
<th>Name</th>
<th>Scope</th>
<th># concepts</th>
<th>Median</th>
<th>Subs. Hier</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED CT</td>
<td>Clinical medicine (patient records)</td>
<td>310,314</td>
<td>2</td>
<td>yes</td>
<td>July 31, 2007</td>
</tr>
<tr>
<td>LOINC</td>
<td>Clinical observations and laboratory tests</td>
<td>46,406</td>
<td>3</td>
<td>no</td>
<td>Version 2.21</td>
</tr>
<tr>
<td>FMA</td>
<td>Human anatomical structures</td>
<td>~72,000</td>
<td>?</td>
<td>yes</td>
<td>(not yet in the UMLS)</td>
</tr>
<tr>
<td>RxNorm</td>
<td>Standard names for prescription drugs</td>
<td>93,426</td>
<td>1</td>
<td>no</td>
<td>Aug. 31, 2007</td>
</tr>
<tr>
<td>NCI Thesaurus</td>
<td>Cancer research, clinical care, public information</td>
<td>58,868</td>
<td>2</td>
<td>yes</td>
<td>2007_05E</td>
</tr>
<tr>
<td>ICD-10</td>
<td>Diseases and conditions (health statistics)</td>
<td>12,318</td>
<td>1</td>
<td>no</td>
<td>1998 (tabular)</td>
</tr>
<tr>
<td>MeSH</td>
<td>Biomedicine (descriptors for indexing the literature)</td>
<td>24,767</td>
<td>5</td>
<td>no</td>
<td>Aug. 27, 2007</td>
</tr>
<tr>
<td>UMLS .</td>
<td>Terminology integration in the life sciences</td>
<td>1,4 M</td>
<td>2</td>
<td>n/a</td>
<td>2007AC (English only)</td>
</tr>
</tbody>
</table>
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
  ● 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).

[Bodenreider, YBMI 2008]
Knowledge management
Knowledge management

Annotating data and resources
Terminology in ontology

◆ Ontology as a source of vocabulary
  ● List of names for the entities in the ontology (ontology vs. terminology)

◆ Most ontologies have some sort of terminological component
  ● Exceptions: GALEN, LOINC

◆ Not all surface forms represented
  ● Often insufficient for NLP applications
  ● Large variation in number of terms per concept across ontologies
Annotating data

- Gene Ontology
  - Functional annotation of gene products in several dozen model organisms
- Various communities use the same controlled vocabularies
- Enabling comparisons across model organisms
- Annotations
  - Assigned manually by curators
  - Inferred automatically (e.g., from sequence similarity)
GO Annotations for Aldh2 (mouse)

<table>
<thead>
<tr>
<th>Category</th>
<th>Classification Term</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Function</td>
<td>aldehyde dehydrogenase (NAD) activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td>oxidoreductase activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td>oxidoreductase activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Cellular Component</td>
<td>mitochondrion</td>
<td>IDA</td>
</tr>
<tr>
<td>Biological Process</td>
<td>metabolic process</td>
<td>IEA</td>
</tr>
<tr>
<td>Biological Process</td>
<td>oxidation reduction</td>
<td>IEA</td>
</tr>
</tbody>
</table>

http://www.informatics.jax.org/
GO ALD4 in Yeast

**GO Annotations**

**Molecular Function**
- Manually curated

**Biological Process**
- Manually curated

**Cellular Component**
- Manually curated
- High-throughput

*All ALD4 GO evidence and references*

View Computational GO annotations for ALD4

- aldehyde dehydrogenase (NAD) activity (IDA, IMP, ISS)
- aldehyde dehydrogenase [NAD(P)+] activity (IDA)
- ethanol metabolic process (IMP)
- mitochondrial nucleoid (IDA)
- mitochondrion (IMP, ISS)
- mitochondrion (IDA)

http://db.yeastgenome.org/
## GO Annotations for ALDH2 (Human)

| Function                                      | Source  | Identifier   | Reference  | ID   | Source  
|-----------------------------------------------|---------|--------------|------------|------|---------
| GO:0016491 oxidoreductase activity            | interpro| IEA          | IPR015590  | UniProt 9606 |
| GO:0016491 oxidoreductase activity            | interpro| IEA          | IPR016160  | UniProt 9606 |
| GO:0016491 oxidoreductase activity            | interpro| IEA          | IPR016162  | UniProt 9606 |
| GO:0016491 oxidoreductase activity            | spkw    | IEA          | KVV-0560   | UniProt 9606 |
| GO:0004029 aldehyde dehydrogenase (NAD)       |         |              | 1306115    | PINC 9606 |
| activity                                     |         |              |            |      |         |
| GO:0004030 aldehyde dehydrogenase (NAD(P)+) activity | | TAS | 8903321 | PINC 9606 |
| GO:0009055 electron carrier activity          |         |              | 8903321    | UniProt 9606 |
| GO:0004029 aldehyde dehydrogenase (NAD)       |         |              | enzyme     | 1.2.1.3 | UniProt 9606 |
| activity                                     |         |              |            |      |         |

http://www.ebi.ac.uk/GOA/
Indexing the biomedical literature

◆ MeSH
  ● Used for indexing and retrieval of the biomedical literature (MEDLINE)

◆ Indexing
  ● Performed manually by human indexers
    ■ With help of semi-automatic systems (suggestions)
      e.g., Indexing Initiative at NLM
  ● Automatic indexing systems
Free cortisol in sepsis and septic shock.

Bendel S, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E; Finnish Sepsis Study Group

Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@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 the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. RESULTS: Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol (r = 0.90, P < 0.001). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 +/- 386 nmol/L vs 793 +/- 439 nmol/L, P = 0.44). Non-survivors had higher calculated serum free (209 +/- 151 nmol/L) and total (980 +/- 458 nmol/L) cortisol concentrations than survivors (119 +/- 111 nmol/L, P = 0.002, and 704 +/- 383 nmol/L, P = 0.002). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%. CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.
MeSH MEDLINE indexing

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

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

Substances:
- Biological Markers
- Hydrocortisone
Replacement therapy for Addison's disease: recent developments.

Lovás K, Husebye ES.

University of Bergen, Institute of Medicine, Section of Endocrinology, 5021 Bergen, Norway.
Kristian.lovas@helse-bergen.no

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

MeSH Terms:
- Addison Disease, blood
- Addison Disease, drug therapy*
- Androgens/administration & dosage*
- Androgens/therapeutic use
- Dosage Forms
- Drug Administration Routes
- Drug Administration Schedule
- Glucocorticoids/administration & dosage*
- Glucocorticoids/adverse effects
- Glucocorticoids/blood
- Glucocorticoids/deficiency
- Hormone Replacement Therapy*
- Humans
- Mineralocorticoids/administration & dosage*
- Mineralocorticoids/adverse effects
- Mineralocorticoids/blood
- Mineralocorticoids/deficiency
- Quality of Life
- Treatment Outcome

Substances:
- Androgens
- Dosage Forms
- Glucocorticoids
- Mineralocorticoids
ICD9-CM Coding clinical data

- ICD9-CM
  - Used for coding clinical data
e.g., for billing purposes

- Other uses of ICD
  - Morbidity and mortality reporting worldwide
Knowledge management

Accessing biomedical information
Resources for biomedical search engines

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

Lister Hill National Center for Biomedical Communications
MeSH “synonyms” MEDLINE retrieval

- MeSH entry terms
  - Used as equivalent terms for retrieval purposes
  - Not always synonymous

- Increase recall without hurting precision

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Addison Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Term</td>
<td>Addison's Disease</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenal Insufficiency</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenocortical Insufficiency</td>
</tr>
</tbody>
</table>
MeSH “synonyms” MEDLINE retrieval

```
"addison disease"[MeSH Terms] OR ("addison"[All Fields] AND "disease"[All Fields]) OR "addison disease"[All Fields] OR ("primary"[All Fields] AND "hypoadrenalism"[All Fields]) OR "primary hypoadrenalism"[All Fields]
```
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
Free cortisol in sepsis and septic shock.
PMID: 18499615 [PubMed - indexed for MEDLINE]

2: Luboshitzky R, Qupi G.
Corticosteroids for septic shock.
PMID: 18467975 [PubMed - indexed for MEDLINE]

Related Articles, Links

12: Løvås K, Husebye ES.
Replacement therapy for Addison's disease: recent developments.
PMID: 18363515 [PubMed - indexed for MEDLINE]
Co-indexing

Searching is now sorted!

http://www.gopubmed.com/

cox-2

Lister Hill National Center for E
Knowledge management

Mapping across biomedical ontologies
Reusing information

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

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

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

Clinical repositories

Genetic knowledge bases

Other subdomains

Biomedical literature

Model organisms

Genome annotations

Anatomy

Lister Hill National Center for Biomedical Communications
Trans-namespace integration

Genetic knowledge bases

OMIM

Other subdomains

... SNOMED CT

Clinical repositories

Addison's disease (363732003)

UMLS

NCBI Taxonomy

Model organisms

FMA

GO

Anatomy

Genome annotations

Biomedical literature

Addison Disease (D000224)

Lister Hill National Center for Biomedical Communications
Data integration, exchange and semantic interoperability
Data integration, exchange and semantic interoperability

Information exchange and semantic operability
“Standards”

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

◆ LOINC
  ● Exchange of laboratory data
  ● In conjunction with HL7 messaging
Semantic interoperability projects  BRIDG

◆ Biomedical Research Integrated Domain Group
  ◆ Information model for clinical research
  ◆ Interoperability between clinical trials information systems
  ◆ Ontologies provide value sets to the information model
Semantic interoperability projects  CDA

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

◆ Cancer Common Ontologic Representation Environment
  ● Infrastructure developed to support an interoperable biomedical information system for cancer research
  ● Uses the NCI Thesaurus as a component
Data integration, exchange and semantic interoperability

Information and data integration
Approaches to data integration

◆ Warehousing
  ● Sources to be integrated are transformed into a common format and converted to a common vocabulary
  ● Normalization through ontologies (e.g., GO annotations)

◆ Mediation
  ● Local schema (of the sources)
  ● Global schema (in reference to which the queries are made)
  ● Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)

Lister Hill National Center for Biomedical Communications
Ontologies and integration

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

Addison's disease (363732003) → Clinical repositories

Other subdomains

SNOMED CT

OMIM

Genetic knowledge bases

GO

UMLS C0001403

Mesh

Biomedical literature

Addison Disease (D000224)

Model organisms

NCBI Taxonomy

FMA

Annotations

Anatomy

Lister Hill National Center for Biomedical Communications
Decision support and reasoning
Data selection

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

◆ Similarity with the definition of value sets for use in the information model
Data aggregation

◆ Ontologies help partition/aggregate data in data analysis
  ● Clinical studies: Study a variable in groups of patients corresponding to the top level categories in ICD
  ● Biology studies: Functional characterization of gene expression signatures with high-level concepts from the Gene Ontology
    ■ Recent trend: co-clustering
Decision support

◆ Clinical decision support
  ● Ontologies help normalize the vocabulary and increase the recall of rules
  ● Ontologies provide some domain knowledge and make it possible to create high-level rules (e.g., for a class of drugs rather than for each drug in the class)

◆ Other forms of decision support
  ● Based on automatic reasoning services for OWL ontologies (e.g., grading gliomas with NCIt)
Natural language processing applications

- Ontologies provide background domain knowledge for NLP applications
  - Question answering
  - Document summarization
  - Literature-based discovery
- The UMLS is often used, but other specific resources have been developed
Knowledge discovery

- By standardizing the vocabulary in a given domain, ontologies are enabling resources for knowledge discovery through data mining.
- Less frequently, the structure of the ontology is leveraged by data mining algorithms.
- Example of available datasets:
  - ICD-coded clinical data (in conjunction with non-clinical information, e.g., environmental data)
  - Annotation of gene products to the GO (function prediction)
Barriers to usability of biomedical ontologies
Availability

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

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

◆ Need for discovery services
Formalism

◆ Several major formalism
  ● Web Ontology Language (OWL) – NCI Thesaurus
  ● OBO format – most OBO ontologies
  ● UMLS Rich Release Format (RRF) – UMLS, RxNorm

◆ Conversion mechanisms
  ● OBO to OWL
  ● LexGrid (import/export to LexGrid internal format)
Ontology integration

◆ Post hoc integration, form the bottom up
  ● UMLS approach
  ● Integrates ontologies “as is”, including legacy ontologies
  ● Facilitates the integration of the corresponding datasets
  ● Current harmonization efforts (e.g., IHTSDO)

◆ Coordinated development of ontologies
  ● OBO Foundry approach
  ● Ensures consistency ab initio
  ● Excludes legacy ontologies
Quality

- Quality assurance in ontologies is still imperfectly defined
  - Difficult to define outside a use case or application

- Several approaches to evaluating quality
  - Collaboratively, by users (Web 2.0 approach)
    - Marginal notes enabled by BioPortal
  - Centrally, by experts
    - OBO Foundry approach

- Important factors besides quality
  - Governance
  - Installed base / Community of practice
Exploring Clinical Ontologies

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)
UMLS UMLSKS

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

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

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

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

◆ Search concepts
  ● Appendectomy (80146002)
  ● Simvastatin (387584000)
  ● Addison's disease (363732003)

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

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

◆ BioPortal
  ● Graphical interface
    ■ Search for Lithium, then navigate down the tree
  ● web services
    http://www.bioontology.org/wiki/index.php/NCBO_REST_services
    ■ 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&pagemenu=1
LOINC  NCI Term Browser

◆ NCI Term Browser
  ● Search for Lithium, then navigate through the Relationships tab
  ● Search by code

◆ Search concept
  ● Substance concentration of lithium in urine (quantitative)
  ● Lithium:Substance Concentration:Point in time:Urine:Quantitative
  ● 25463-1
RxNorm  RxNav

◆ RxNav
http://umlsks.nlm.nih.gov/  
(launch the browser)

◆ Search by string (default): zyrtec, clopidogrel
  ● Restrict the graph to one particular clinical drug: double-click on Cetirizine 10 MG Oral Tablet
  ● RxCUI is displayed in the information bar in the bottom when clicking on a drug entity (e.g., RxCUI for Cetirizine 10 MG Oral Tablet = 309130)
  ● Right-click on Cetirizine 10 MG Oral Tablet
  ● View NDCs to open a window with the list of NDCs for this drug
  ● View Drug Label → link out to DailyMed
Search by ID (select ID in the drop-down “Search by” menu)
- NDC, with search string 00781168401 (one of the NDC from the list obtained from Cetirizine 10 MG Oral Tablet)
- SNOMED ID, with search string 1039008
  - Returns: 103|C0000618||6-Mercaptopurine

Packs: Search for z-pak
- Packs displayed with double diamonds in the clinical drug / generic pack and branded drug / branded pack boxes
RxNorm SOAP API

◆ RxNorm SOAP API (demo client)

◆ Functions
  • getRxNormVersion()
  • getIdTypes()
  • findRxcuiById(00904582941, 309130) → 309130
  • getAllRelatedInfo(309130)

◆ Documentation
  http://rxnav.nlm.nih.gov/RxNormAPI.html
RxNorm REST API

◆ Test resources

◆ Documentation
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, …)
June 16, 2010 – Discussion Sessions

Issues and Challenges
Related to Clinical Ontologies

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

- NLP / indexing
- PHR / consumer health information
- Decision support (drugs)
- Decision support (other)
- Medication reconciliation
- E-prescribing
- CPOE
- Problem list
- Terminology services

- Value sets
- Terminology management (versioning)
- Mapping / integration
- Meaningful use
- Health information exchange
- Clinical documentation
Questions

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

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

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