Workshop on
Foundations of Clinical Terminologies
and Classifications (FCTC 2006)
Timişoara, Romania, April 8, 2006

Biomedical terminology and beyond

Ontology and terminology services

Olivier Bodenreider

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

Outline

- Why biomedical terminologies?
- Building biomedical terminologies: *Recent experiences*
- Terminology vs. ontology
- Terminology services

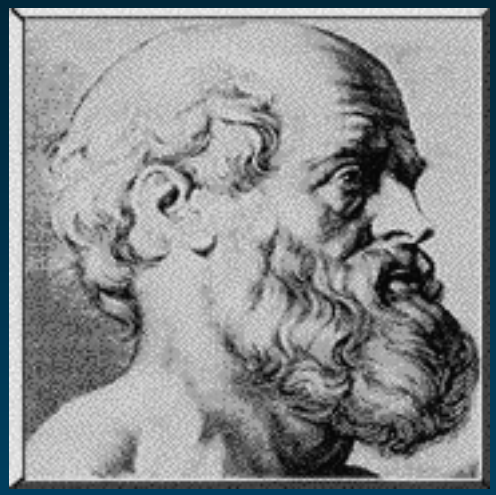
Why biomedical terminologies?

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

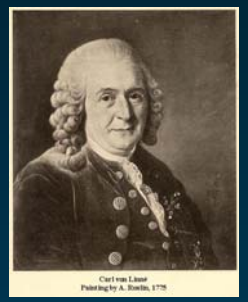
To support a theory of diseases

- 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)

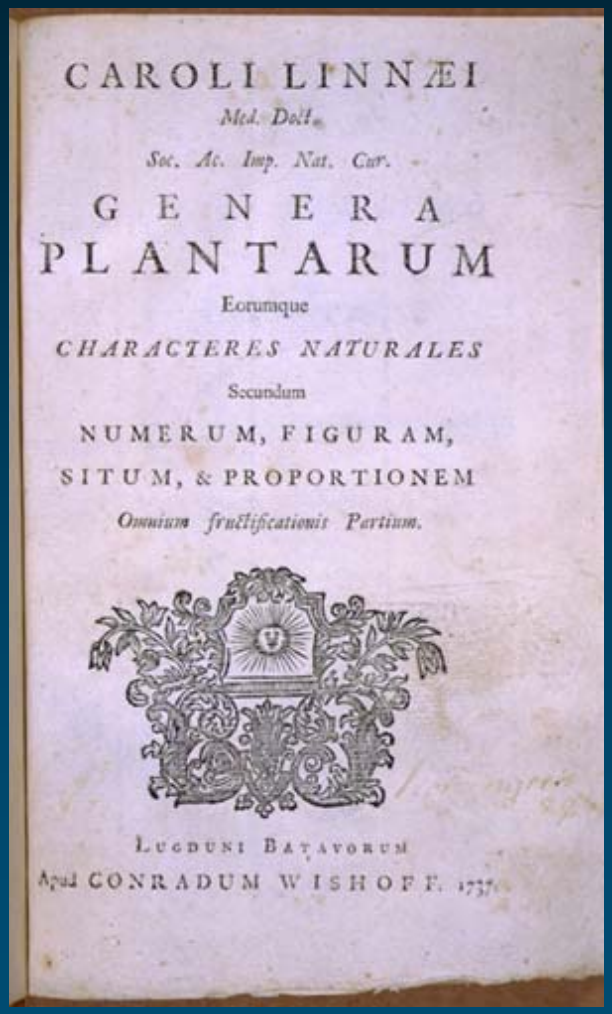
- Carolus Linnaeus (1707-1778)
  - *Genera Plantarum* (1737)
  - *Genera Morborum* (1763)



- 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…



CAROLI LINNÆI

Med. Doct.

Soc. Ac. Imp. Nat. Cur.

GENERA PLANTARUM

Eorumque

CHARACTERES NATURALES

Secundum

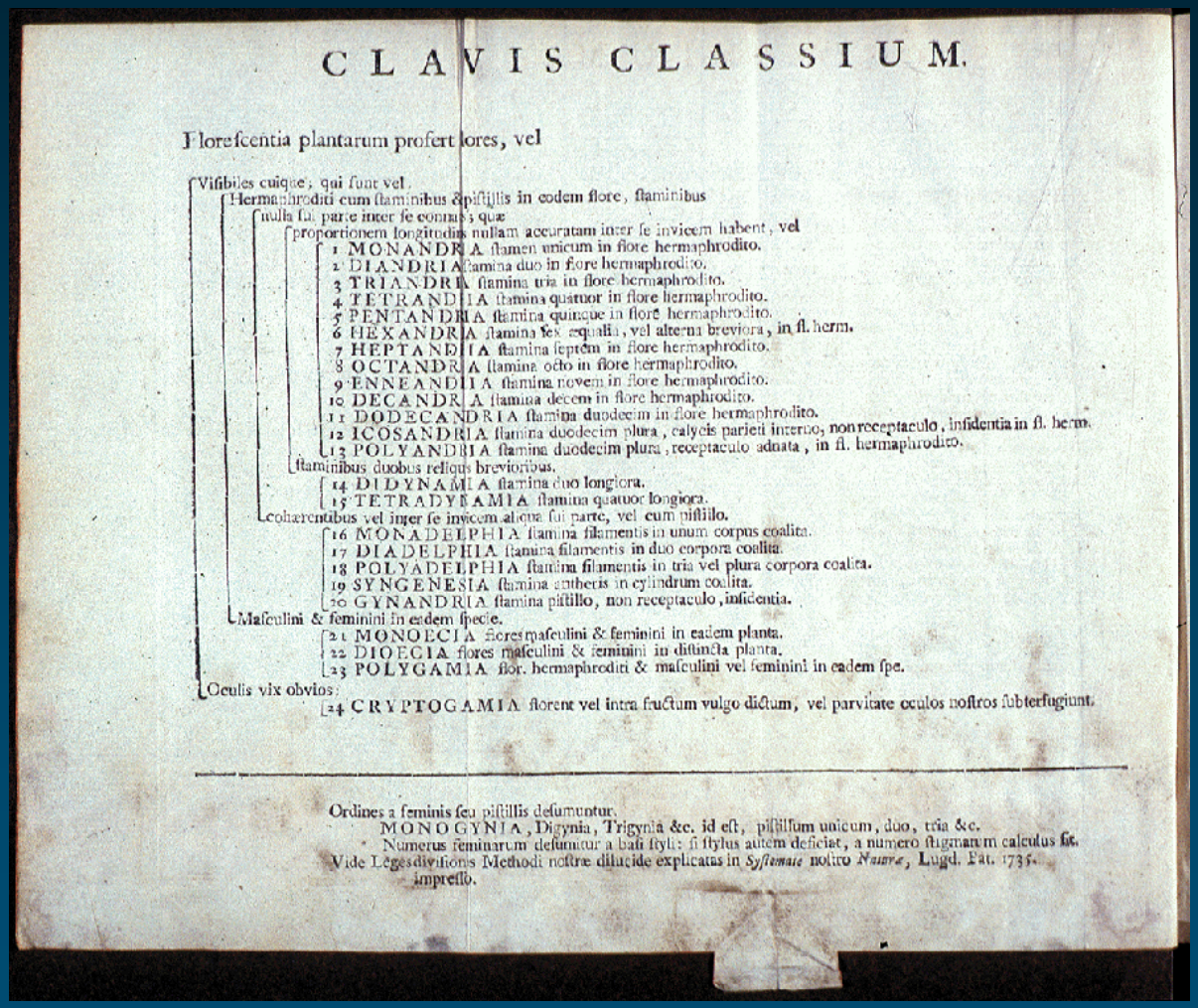
NUMERUM, FIGURAM,

SITUM, & PROPORTIONEM

Omnium fructificationis Partium.

LUGDUNI BATAVORUM

Apud CONRADUM WISHOFF. 1737.



CLAVIS CLASSIUM.

Florescentia plantarum profert flores, vel

- Visibiles cuique, qui sunt vel:
  - Hermaphroditi cum staminibus & pistillis in eodem flore, staminibus
    - nulla sui parte inter se connatis; quæ
      - proportionem longitudinis nullam accuratam inter se invicem habent, vel
        - 1 MONANDRIA stamen unicum in flore hermaphrodito.
        - 2 DIANDRIA stamina duo in flore hermaphrodito.
        - 3 TRIANDRIA stamina tria in flore hermaphrodito.
        - 4 TETRANDRIA stamina quatuor in flore hermaphrodito.
        - 5 PENTANDRIA stamina quinque in flore hermaphrodito.
        - 6 HEXANDRIA stamina sex æqualia, vel alterna breviora, in fl. herm.
        - 7 HEPTANDRIA stamina septem in flore hermaphrodito.
        - 8 OCTANDRIA stamina octo in flore hermaphrodito.
        - 9 ENNEANDRIA stamina novem in flore hermaphrodito.
        - 10 DECANDRIA stamina decem in flore hermaphrodito.
        - 11 DODECANDRIA stamina duodecim in flore hermaphrodito.
        - 12 ICOSANDRIA stamina duodecim plura, calycis parieti interno, non receptaculo, insidentia in fl. herm.
        - 13 POLYANDRIA stamina duodecim plura, receptaculo adnata, in fl. hermaphrodito.
      - staminibus duobus reliquis brevioribus.
        - 14 DIDYNAMIA stamina duo longiora.
        - 15 TETRADYNAMIA stamina quatuor longiora.
    - cohærentibus vel inter se invicem aliqua sui parte, vel cum pistillo.
      - 16 MONADELPHIA stamina filamentis in unum corpus coalita.
      - 17 DIADELPHIA stamina filamentis in duo corpora coalita.
      - 18 POLYADELPHIA stamina filamentis in tria vel plura corpora coalita.
      - 19 SYNGENESIA stamina antheris in cylindrum coalita.
      - 20 GYNANDRIA stamina pistillo, non receptaculo, insidentia.
  - Masculini & feminini in eadem specie.
    - 21 MONOECIA flores masculini & feminini in eadem planta.
    - 22 DIOECIA flores masculini & feminini in distincta planta.
    - 23 POLYGAMIA flor. hermaphroditi & masculini vel feminini in eadem spe.
- Oculis vix obvios:
  - 24 CRYPTOGAMIA florent vel intra fructum vulgo dictum, vel parvitate oculos nostros subterfugiunt.

Ordines a feminis seu pistillis desumuntur.
MONOGYNIA, Digynia, Trigynia &c. id est, pistillum unicum, duo, tria &c.
Numerus feminarum desumitur a basi styli: si stylus autem deficiat, a numero stigmatum calculus fit.
Vide Leges divisionis Methodi nostræ dilucide explicatas in *Systemate* nostro *Naturæ*, Lugd. Bat. 1735. impresso.

… to diseases

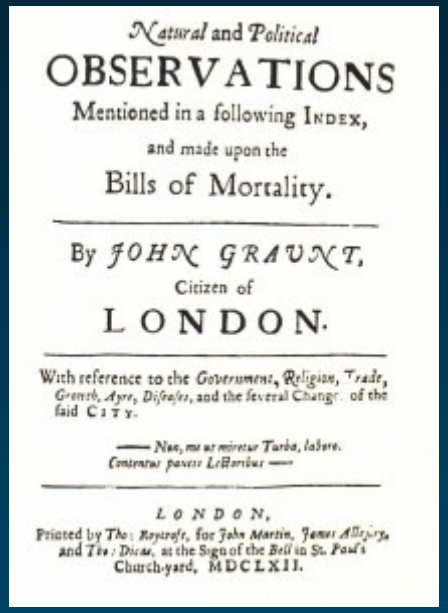
- Four categories (W. Cullen)
  - Fevers
  - Nervous disorders
  - Cachexias
  - Local diseases

> "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)


Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

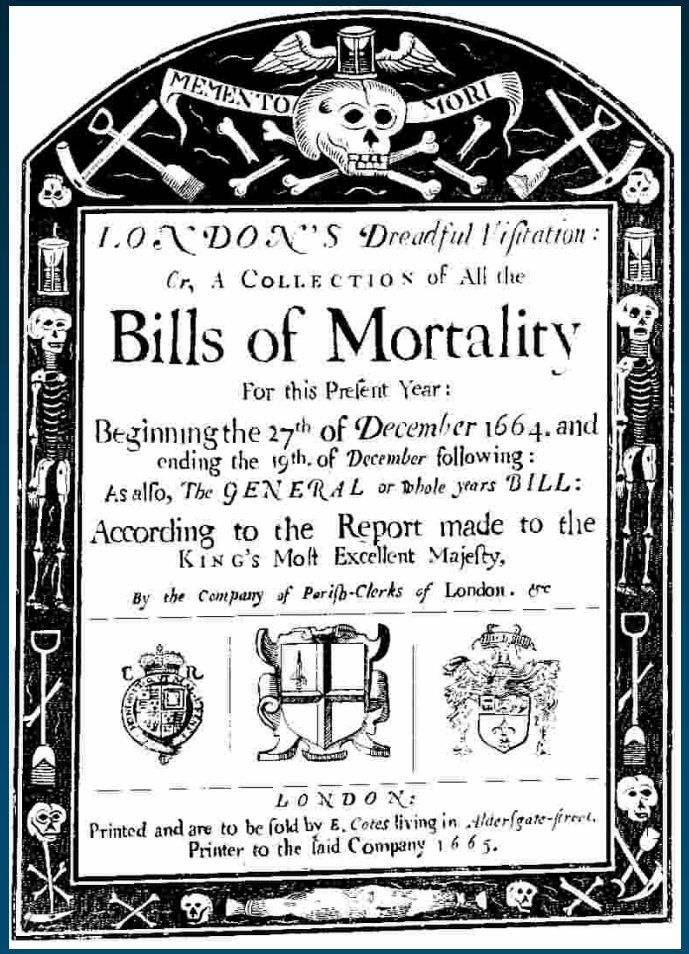
By JOHN GRAUNT,
Citizen of
LONDON.

With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Changes of the said CITY.

—— Non, me ut miretur Turba, labore.
Contentus paucis Lectoribus ——

LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

London Bills of Mortality



MEMENTO MORI

LONDON'S Dreadful Visitation:

Or, A COLLECTION of All the

Bills of Mortality

For this Present Year:

Beginning the 27th of December 1664. and ending the 19th. of December following:

As also, The GENERAL or whole years BILL:

According to the Report made to the KING'S Most Excellent Majesty,

By the Company of Parish-Clerks of London. &c

LONDON:

Printed and are to be sold by E. Cotes living in Aldersgate-street, Printer to the said Company 1665.

A generall Bill for this present year, ending the 19 of December 1665. according to the Report made to the KINGS most Excellent Majesty.

By the Company of Parish Clerks of London, &c.

The Diseases and Casualties this year.

Abortive and Stillborne	617	Executed	21	Palsie	50
Aged	1545	Flox and Small Pox	655	Plague	68596
Ague and Feaver	5257	Found dead in Streets, fields, &c.	2	Plaimet	6
Apoplex and Suddenly	116	French Pox	86	Plurisie	19
Bedric	10	Frighted	23	Poysoned	1
Blasted	5	Gout and Sciatica	27	Quinsie	35
Bleeding	16	Grief	46	Rickets	157
Bloody Flux, Scowring & Flux	185	Griping in the Guts	1238	Rising of the Lights	397
Burnt and Scalded	8	Hanged & made away themselves	7	Rupture	34
Calenture	3	Headmould shot & Mould fallen	14	Scurvy	1[illegible]7
Cancer, Gangrene and Fistula	56	Jaundies	120	Shingles and Swine pox	2
Canker, and Thrush	121	Impostume	227	Sores, Ulcers, broken and bruised	
Childbed	625	Killed by severall accidents	46	Limbs	52
Chrisomes and Infants	1258	Kings Evill	86	Spleen	14
Cold and Cough	65	Leprosie	2	Spotted Feaver and Purples	1929
Collick and Winde	134	Lethargy	14	Stopping of the Stomack	334
Consumption and Tisick	48[illegible]8	Liver-grown	21	Stone and Strangury	58
Convulsion and Mother	2035	Measlom and Headach	12	Suddenly	[illegible]
Dilitated	5	Measles	7	Teeth and Worms	2614
Dropsie and Tympany	1478	Murthered and Shot	9	Vomiting	54
Drowned	53	Overlaid & Starved	45	Vrine	7

Christened: Males	5114	Buried: Males	48568	Of the Plague	68595
Christened: Females	4853	Buried: Females	48737		
Christened: In all	9967	Buried: In all	97305		

Increased in the Burials in the 130 Parishes and at the Pest-house this year 79009

Increased of the Plague in the 130 Parishes and at the Pest-house this year 68590

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.

To index and retrieve information

- Biomedical literature
  - MEDLINE (15M citations from 4600 journals)
  - Manually indexed
  - Medical Subject Headings (MeSH)
- Genome
  - Model organisms (Fly, Mouse, Yeast, …)
  - Manually / semi-automatically annotated
  - Gene Ontology

MEDLINE and MeSH

1: J Hist Neurosci. 2004 Mar;13(1):91-101. Related Articles, Links

MetaPress

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*

PubMed

National Library of Medicine NLM

Mouse Genome Database and GO

Entrez Gene

1: **Nf2** **neurofibromatosis 2** [*Mus musculus*]

GeneID: 18016 Locus tag: MGI:97307

▶ **General gene information**

GeneOntology

Provided by MGI

Function	**Evidence**	
cytoskeletal protein binding	IEA	
protein binding	IPI	PubMed
structural molecule activity	IEA	
Process		
intercellular junction assembly and/or maintenance	IMP	PubMed
negative regulation of cell cycle	IEA	
negative regulation of protein kinase activity	IDA	PubMed
regulation of cell proliferation	IMP	PubMed
Component		
adherens junction	IMP	PubMed
cytoplasm	IEA	
cytoskeleton	IEA	
membrane	IEA	

To serve as a reference

- Reference terminology/ontology
  - Universally needed
  - Developed independently of any purposes
  - Reusable by many applications
- Examples
  - RxNorm
  - Foundational Model of Anatomy (FMA)
  - SNOMED CT
  - ChEBI

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)

Building biomedical terminologies

Recent experiences

Building biomedical terminologies

Recent experiences

- Description logics approach
- Reengineering terminologies with DL
- Reorganizing MeSH
- Gene Ontology
- UMLS SemanticNetwork

Description logics approach

- Pioneered by GALEN
  - Although GALEN itself is not a terminology
- SNOMED CT
  - Although it is distributed as a relational database (terms, relations), not in DL format

- DL is used to support the creation of terminologies
- The goal is not to have terminologies in OWL

Benefits of using a DL approach

- Consistent organization
  - Equivalent classes
  - Automatic classification
  - Error detection through reclassification
  - …
- But DL does nothing for the naming component of terminologies
  - Inconsistent synonyms for anatomical concepts in SNOMED CT (Structure/Entire)

Reengineering terminologies with DL

- Ontologizing terminologies
  - e.g., UMLS
    - Metathesaurus [Hahn, PSB 2003], [Cornet, AMIA 2002] , [Pisanelli, AMIA 1998]
    - Semantic Network [Kashyap, ISWC 2003]
- Migrating to OWL
  - NCI Thesaurus [Golbeck, JWS 2003]
  - Gene Ontology [Wroe, PSB 2003]
  - MeSH [Soualmia , KE-MED 2004]
  - FMA [Golbreich, OWLED 205]

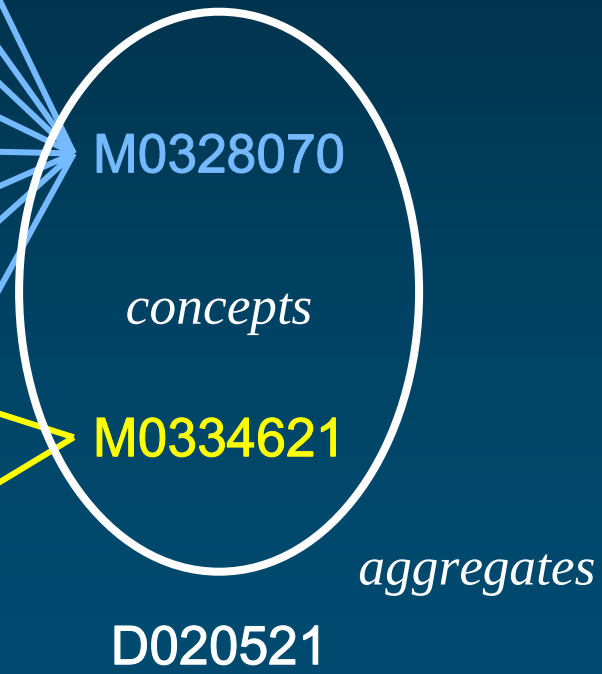
Reengineering with DL Limitations

- No trivial isomorphism
- Never purely a matter or formalism
  - Not every thesaurus relation should become *isa*
  - Necessary and sufficient conditions for anatomical structures?
- Never completely automatic
- Costly in terms of human resources

Terminology + formalism ≠ Formal terminology

Reorganizing MeSH

MeSH Heading	Cerebrovascular Accident
Entry Term	Apoplexy
Entry Term	Cerebral Stroke
Entry Term	Cerebrovascular Apoplexy
Entry Term	Stroke
Entry Term	Vascular Accident, Brain
Entry Term	CVA (Cerebrovascular Accident)
Entry Term	Cerebrovascular Accident, Acute
Entry Term	Cerebrovascular Stroke
Entry Term	Stroke, Acute
Unique ID	D020521



Gene Ontology

- Developed by biologists in the early 2000's
- Extremely popular
  - Genome annotation across model organism databases
- Simplistic
  - No relations across hierarchies
  - Only *isa* and *part_of* relationships
- Being reengineered/ontologized
  - OBOL (formal language for representing lexical relations)
  - National Center for Biomedical Ontology
  - Relations across hierarchies will be added

UMLS Semantic Network

- Weak (some-some) semantics
- Metathesaurus concepts linked to semantic types, but no link between MT and SN relationships

- Being reanalyzed from the perspective of formal ontology
  - e.g., distinction between continuants and occurrents
  - Mapping of relationships between MT and SN

Terminology vs. Ontology

Terminology vs. Ontology

- Types of resources
  - Lexical
  - Terminological
  - Ontological
- Ontology is overloaded
- Terminology is overloaded too
- Formal approaches to terminology

Lexical vs. ontological 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, SNOMED CT

Types of resources revisited

- Lexical and terminological resources
  - Mostly collections of names for biomedical entities
  - Often have some kind or hierarchical organization (e.g., relations)
- Ontological resources
  - Mostly collections of relations among biomedical entities
  - Sometimes also collect names

Unified Medical Language System

- SPECIALIST Lexicon
  - 200,000 lexical items
  - Part of speech and variant information
- Metathesaurus
  - 5M names from over 100 terminologies
  - 1M concepts
  - 16M relations
- Semantic Network
  - 135 high-level categories
  - 7000 relations among them

Lexical resources

Terminological resources

Ontological resources

Ontology is overloaded

- Hype
- Not every ontology built
  - is formal
  - has definitions
  - is consistent
  - …
- Not everything in OWL (resp. Protégé) is an ontology

Terminology is overloaded too Terms

- "Terms" are not necessarily named for biomedical entities

- 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 aspects of the total patient care plan)

Terminology is overloaded too Relations

- Hierarchical structures created to support a task e.g., information retrieval for 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] +

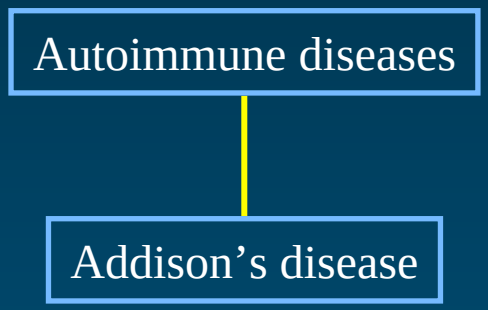
Accidents, Radiation [G03.850.110.285]

Accidents, Traffic [G03.850.110.320]

Drowning [G03.850.110.500] +

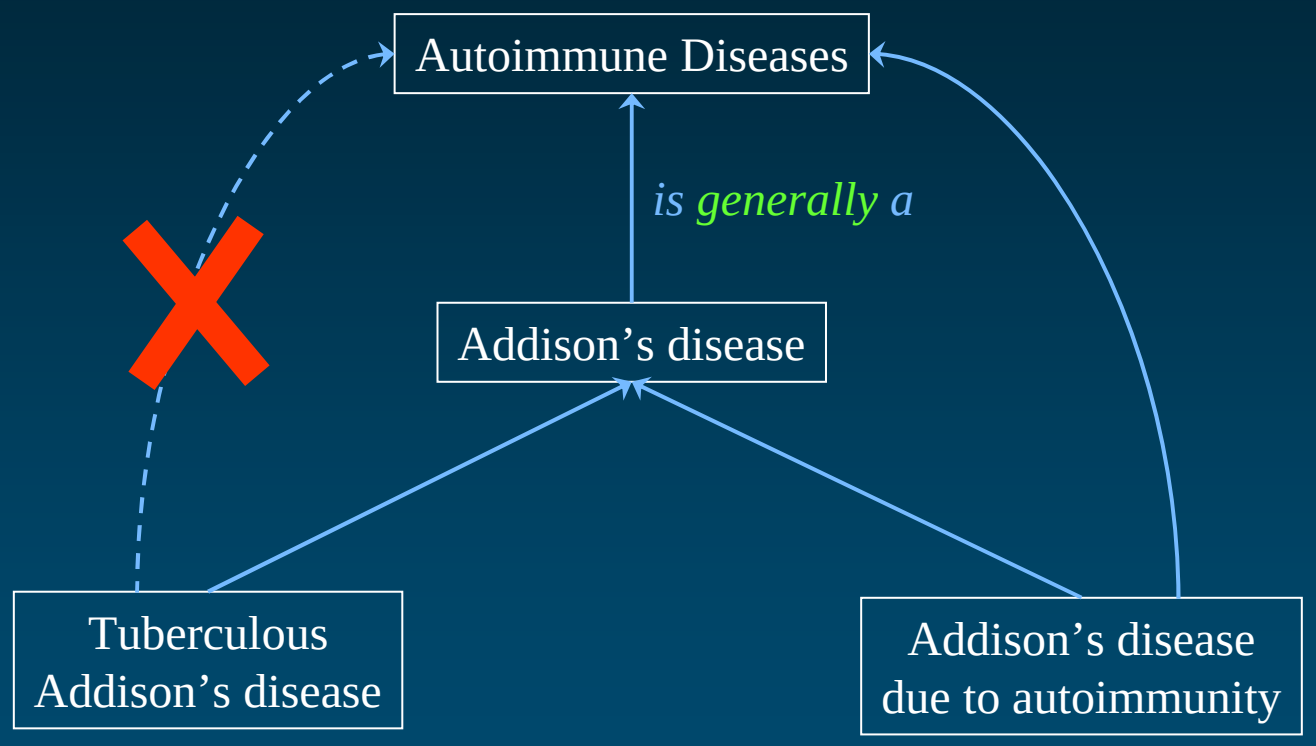
Thesaurus relations

- 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

Not all “isa” relations are transitive!



Terminologies do not necessarily support reasoning

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] +

Hierarchy

Subtype hierarchy

adrenal cortical hypofunction

Addison's disease

Addison's disease due to autoimmunity

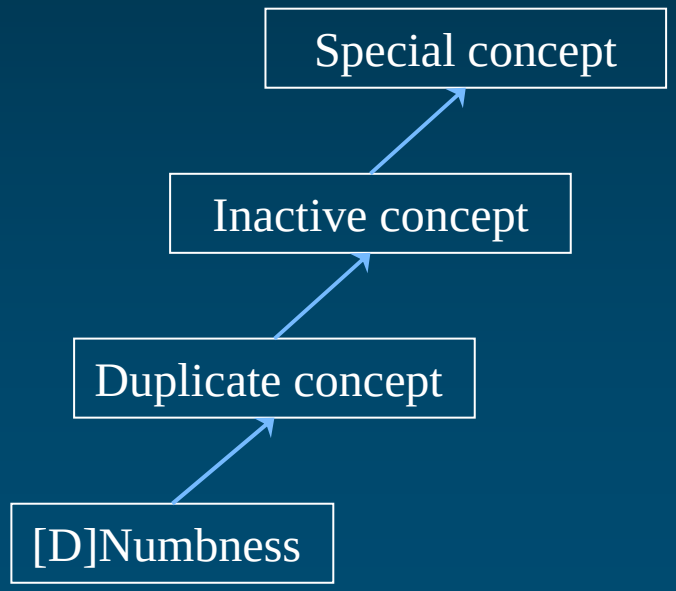
Addison's disease with adrenoleucodystrophy

polyglandular autoimmune syndrome, type 1

tuberculous Addison's disease

Housekeeping relations

- Obsolete terms
  - Maintained in the terminology (permanence principle)
  - Linked to special "housekeeping" concepts



Formal approaches to terminology

- Computational terminology
  - Tasks
    - Identifying terms from text corpora automatically
    - Organizing terms automatically
  - Methods
    - Lexicosyntactic and semantic analysis
    - Machine learning
    - Information science
- Limited interest in biomedicine because of the existence of comprehensive terminologies

Terminology services

Terminology services

- Defining terminology services
- Lexical issues
- Ontological issues

The GALEN terminology server

- *Managing external references*
- *Managing internal representations*
- *Mapping natural language to concepts*
- *Mapping concepts to classification schemes*
- *Management of extrinsic information*

[Rector, Methods 1995]

Chris Chute's desiderata

- *Word normalization*
- *Word completion*
- *Spelling correction*
- *Lexical matching*
- *Term completion*

Lexical resources

- *Target terminology specification*
- *Semantic locality*
- *Term composition*
- *Term decomposition*

Ontological resources

[Chute, AMIA 1999]

Requirements

- Model of the term
  - Lexico-syntactic level (lexical resemblance)
    - Supported by lexicons
      - Word properties
    - Edit distance for spelling correction
    - Rules for normalization (defining inessential features)
  - Semantic level (semantic similarity)
    - Supported by ontologies
      - Concept properties
      - Relations to other concepts
    - Constraints for composition

Requirements (continued)

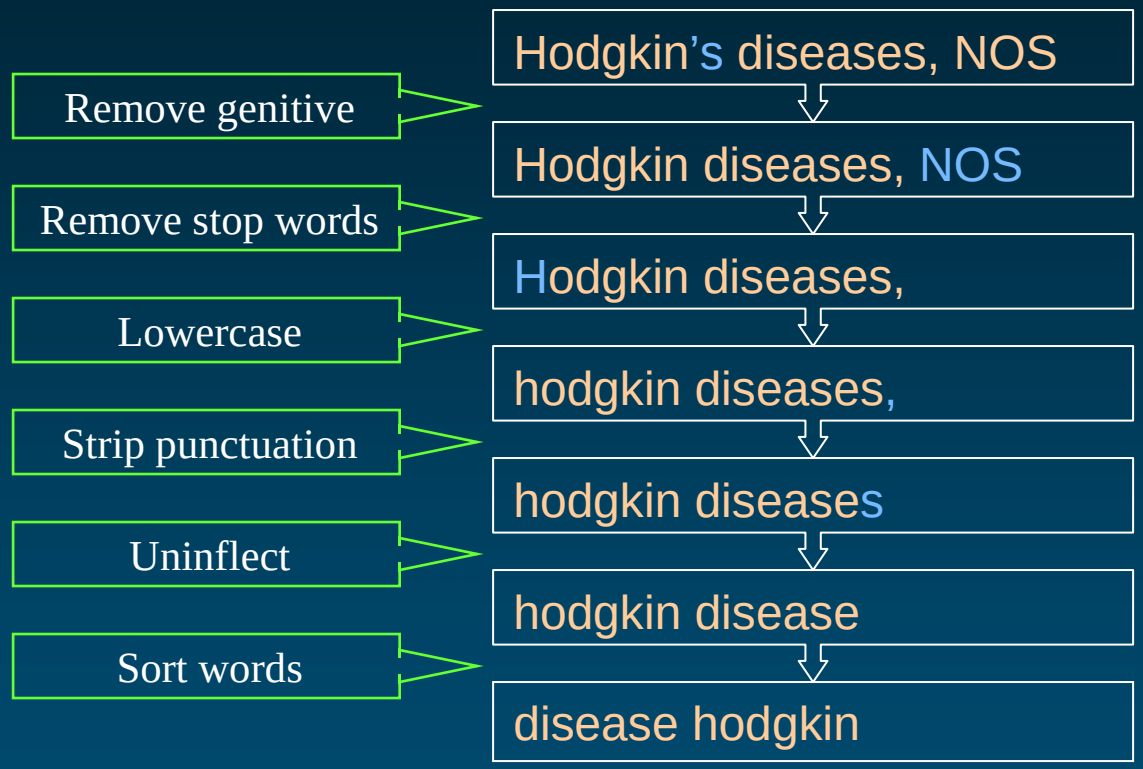
- Model of the mapping
- Model of the task (context of use)

- Other terminology services
  - Subsetting terminologies
  - Helping define value sets
  - Self-generating terminologies (from orthogonal ontologies)
  - Extending terminologies

Lexico-syntactic level

- Lots of developments in the past 15 years
- Stable for English [McCray, AMIA 1994]
  - UMLS SPECIALIST Lexicon
  - Lexical tools (e.g., lvg, spelling correction module)
- Underway for other languages
  - Spanish (NLM)
  - German (Freiburg)
  - French (UMLF)

Normalization



Normalization: Example

Hodgkin Disease
HODGKINS DISEASE
Hodgkin's Disease
Disease, Hodgkin's
Hodgkin's, disease
HODGKIN'S DISEASE
Hodgkin's disease
Hodgkins Disease
Hodgkin's disease NOS
Hodgkin's disease, NOS
Disease, Hodgkins
Diseases, Hodgkins
Hodgkins Diseases
Hodgkins disease
hodgkin's disease
Disease, Hodgkin

normalize → disease hodgkin

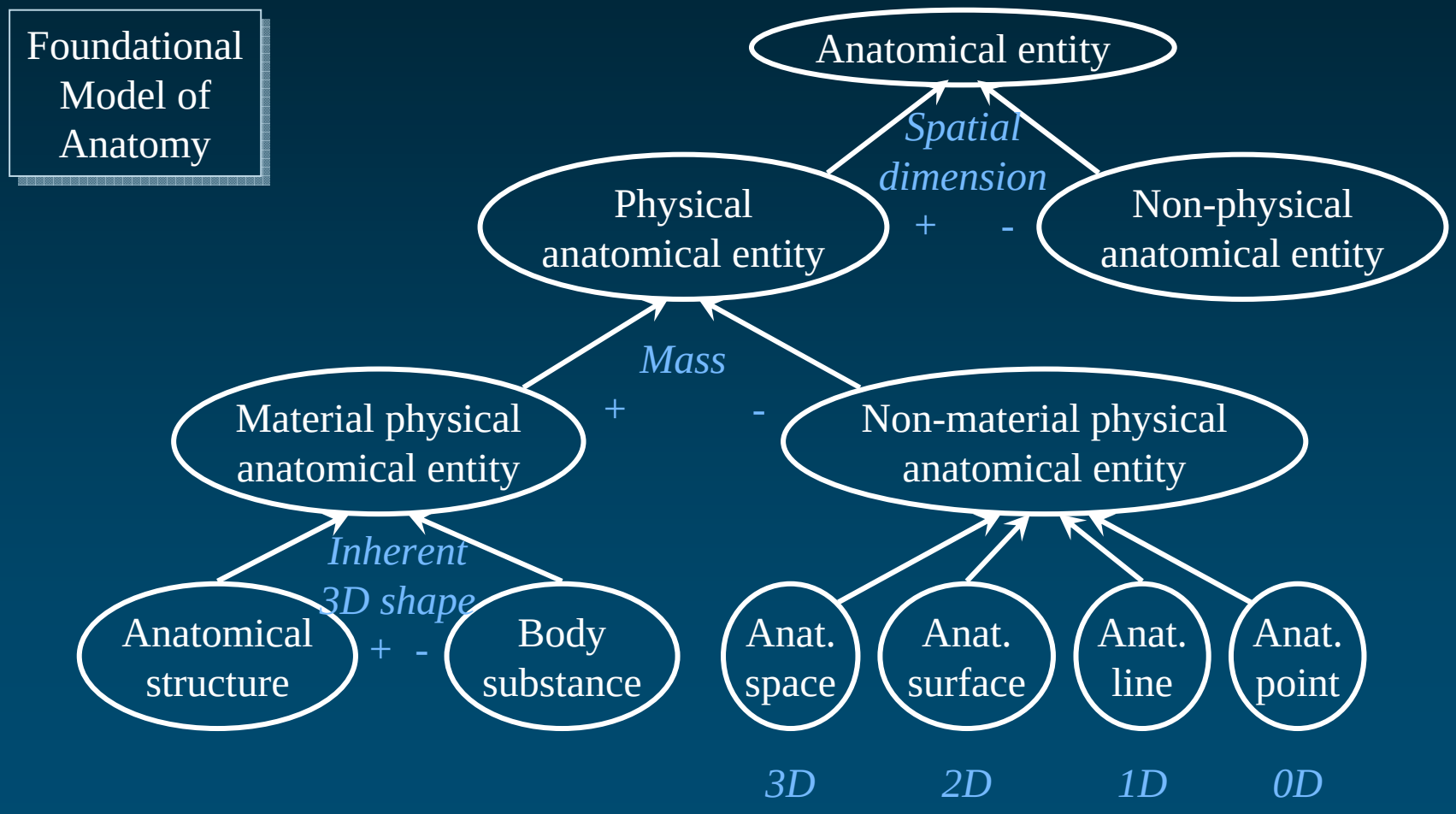
Lexical issues

- Normalization was developed essentially for clinical terms
- Known issues
  - Drug names
  - Chemicals
- New issues with biological corpora
  - Gene names

Semantic level

- Limited progress in the past 15 years
- Single most important contribution: SNOMED CT
- Main source of labeled relations in the UMLS i.e., explicit classificatory criteria
- Few other vocabularies in the UMLS contribute labeled relations in large numbers
  - NDFRT
  - RxNorm

Explicit classificatory principle



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] +
- 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] +

agent/cause

location

stage in life

Semantic issues

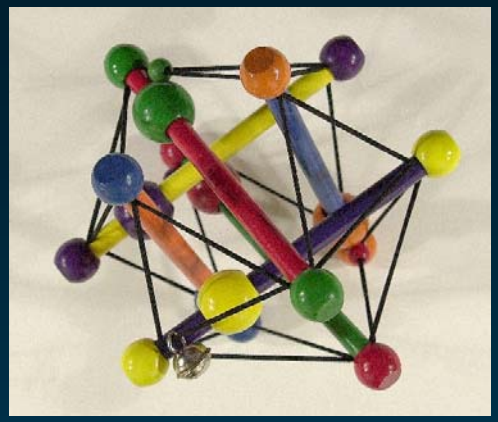
- Lack of classificatory principles explicitly stated and represented in ontologies
- Lack of trans-ontological (associative) relations represented in ontologies

- Result in
  - Inconsistent representations
    - e.g., Prevention of X / X
  - Maintenance issues
    - e.g., modification of a given term should trigger the review of dependent terms

Summary

Summary

- Terminology vs. ontology
- Terminology vs. terminology services
- Usefulness vs. elegance



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