University of Pisa, Italy
June 14, 2007

NETTAB 2007 - A Semantic Web for Bioinformatics

Bio-ontologies

The cream in the Semantic Web layer cake

Olivier Bodenreider

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

Semantic Web pastry

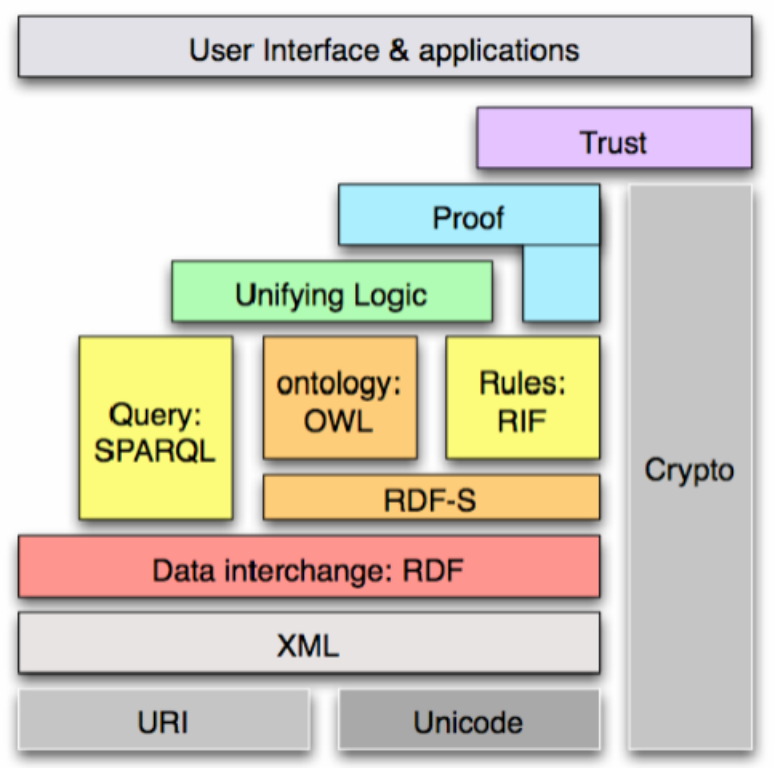
Semantic Web layer cake

http://www.axis-of-aevil.net/img/2005_09/tyrnicake1.jpg

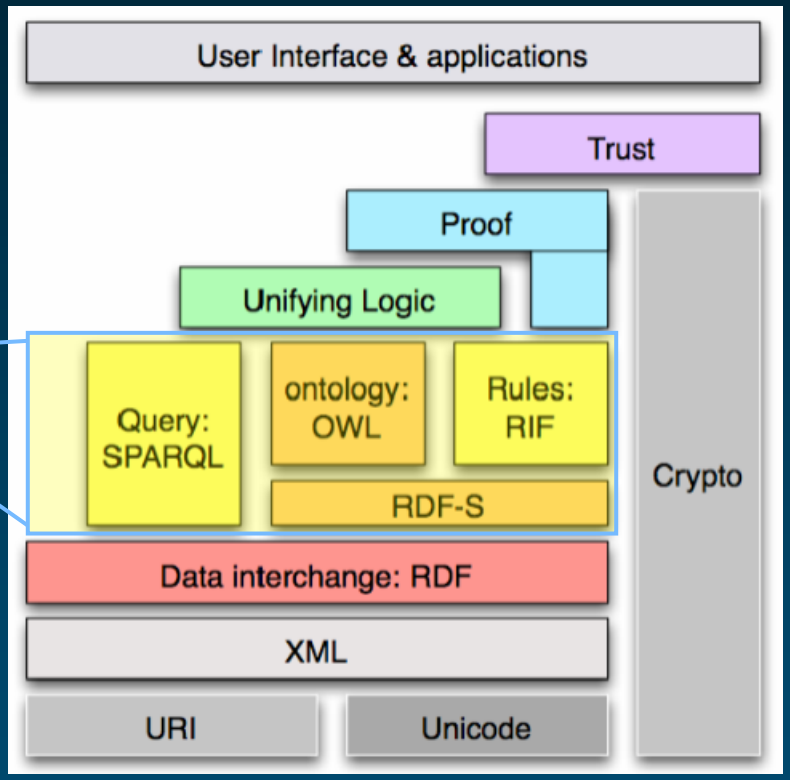
Semantic Web layer cake

http://www.cookingwithkristina.com/uploaded_images/fudgy-702607.jpg

Semantic Web layer cake



Semantic Web layer cake



Outline

- Historical perspective
- Modern bio-ontologies
- Tools and formalisms
- Institutionalization of bio-ontologies
- Bio-ontologies and Semantic Web

Briefings in Bioinformatics

BRIEFINGS IN BIOINFORMATICS. VOL 7. NO 3. 256–274 doi:10.1093/bib/bbl027

Bio-ontologies: current trends and future directions

Olivier Bodenreider and Robert Stevens

Submitted: 23rd June 2006; Received (in revised form): 10th July 2006

http://bib.oxfordjournals.org/cgi/reprint/7/3/256?ijkey=1ejwW7ipyG1ASiI&keytype=ref

Before we called them bio-ontologies

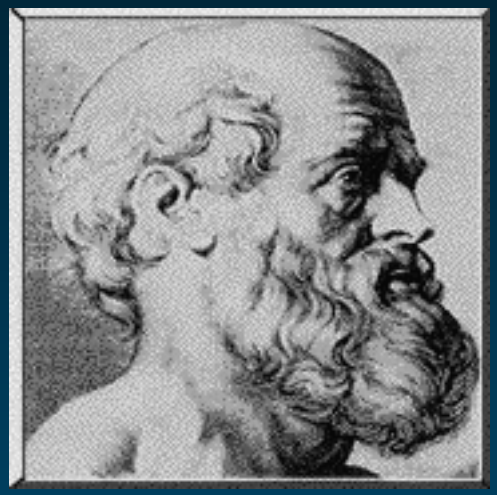
A brief history of biomedical terminologies

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

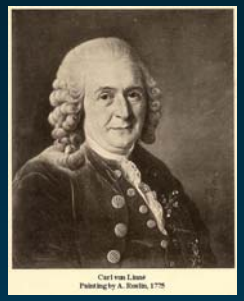
To support a theory of diseases

- 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)

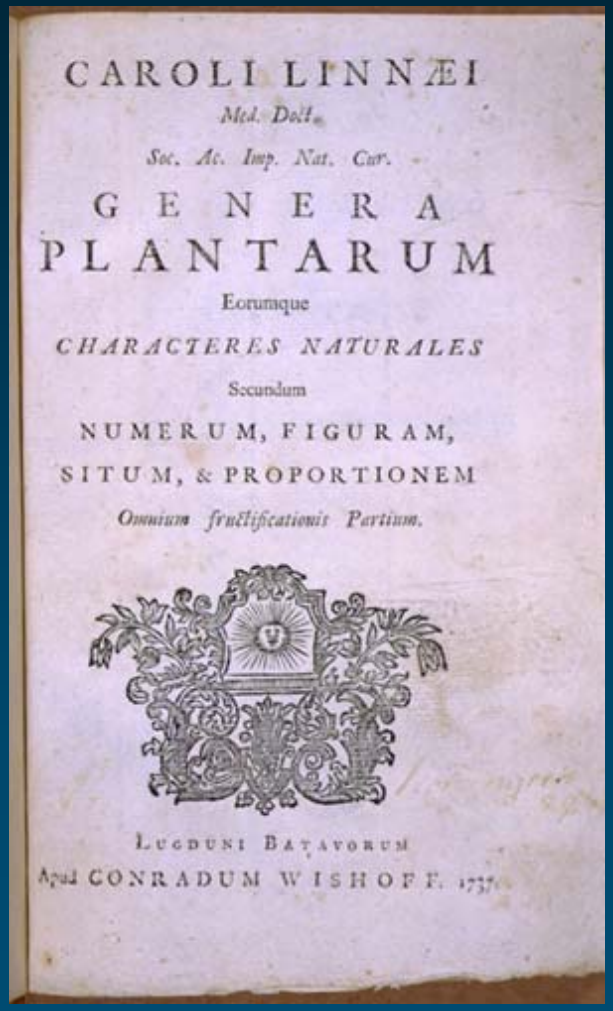
- Carolus Linnaeus (1707-1778)
  - *Genera Plantarum* (1737)
  - *Genera Morborum* (1763)



- 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

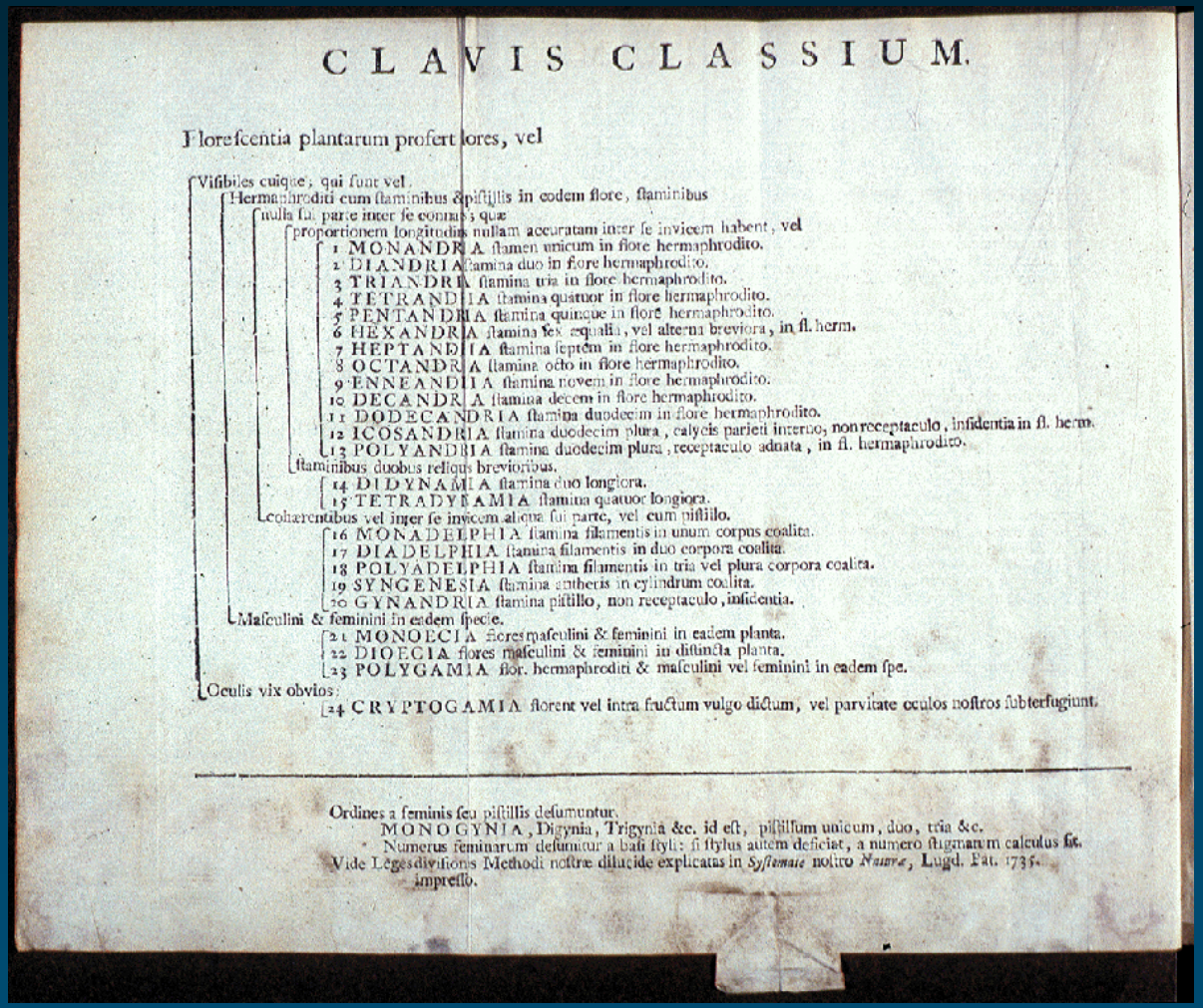
Secundum

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 reliqui brevioribus.
        - 14 DIDYNAMIA stamina duo longiora.
        - 15 TETRADYNAMIA stamina quatuor longiora.
    - coherentibus 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

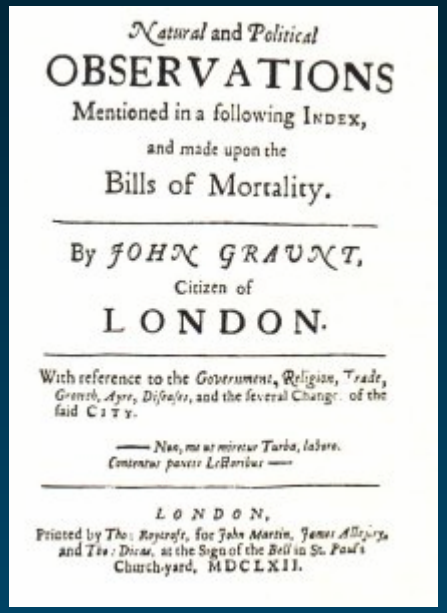
- Four categories (W. Cullen)
  - Fevers
  - Nervous disorders
  - Cachexias
  - Local diseases

"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.

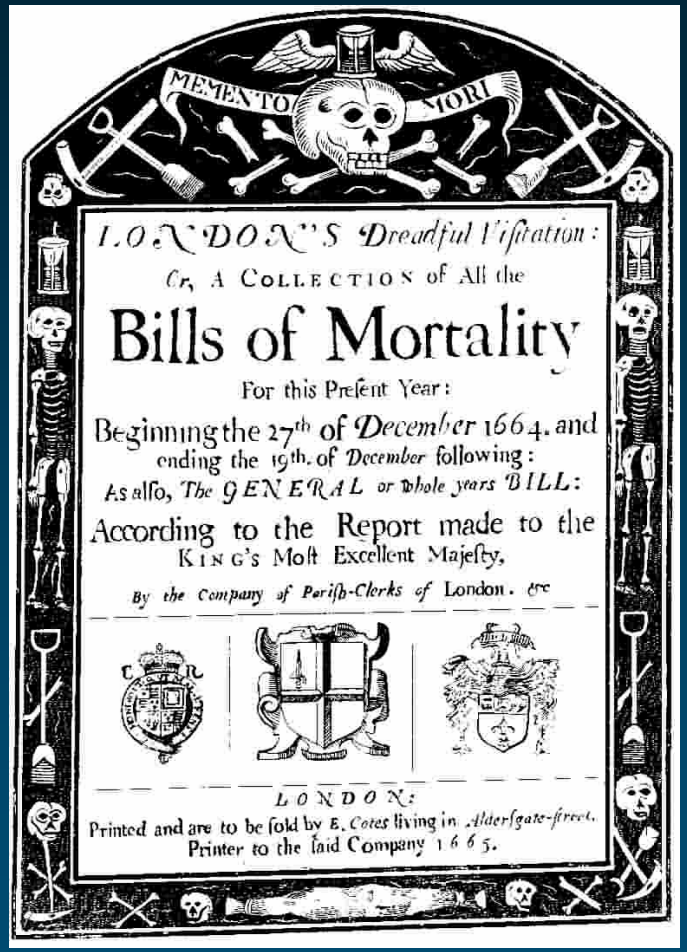
By JOHN GRAUNT,
Citizen of
LONDON.

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



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	Flux and Small Pox	655	Plague	68596
Ague and Feaver	5257	Found dead in Streets, fields, &c.	2	Planet	6
Appoplex and Suddenly	116	French Pox	86	Pleurie	15
Bedrid	12	Frighted	23	Poisoned	1
Blasted	5	Gout and Sciatica	27	Quinsie	35
Bleeding	16	Grief	46	Rickets	157
Bloody Flux, Scowring & Flux	185	Griping in the Guts	1288	Rising of the Lights	397
Burnt and Scalded	8	Hanged & made away themselves	7	Rupture	34
Calenture	3	Headmould shot & Mould fallen	14	Scurvy	17
Cancer, Gangrene and Fistula	56	Jaundies	120	Shingles and Swine pox	2
Canker, and Thrush	121	Impostume	227	Sores, Ulcers, broken and bruised Limbs	82
Childbed	625	Kild by severall accidents	46	Spleen	14
Christomes and Infants	1258	Kings Evill	86	Spotted Feaver and Purples	1929
Cold and Cough	65	Leprosie	2	Stopping of the Stomack	334
Collick and Winde	134	Lethargy	14	Stone and Strangury	98
Consumption and Tisick	4888	Livergrown	21	Suddat	1
Convulsion and Mother	2036	Measlom and Headach	12	Teeth and Worms	2614
Distracted	5	Measles	7	Vomiting	31
Dropsie and Tympany	1478	Murthered and Shot	9	Wenn	7
Drowned	53	Overlaid & Starved	45		

Christened: Males 5114, Females 4853, In all 9967

Buried: Males 4856, Females 4837, In all 9693 — Of the Plague 68595

Increased in the Burials in the 130 Parishes and at the Pest-house this year 70169

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)
  - ChEBI
  - SNOMED CT
  - LOINC

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)

Modern bio-ontologies

Biomedical ontologies (and terminologies)

- The OBO family
  - Ontologies and terminologies
  - Gene Ontology
  - Mostly biological ontologies
- UMLS
  - Ontologies and terminologies
  - MeSH, SNOMED CT
  - Mostly clinical ontologies

Open Biological Ontologies

- Extended family of the Gene Ontology (GO)

- Collaborative development
  - http://obo.sourceforge.net/
- National Center for Biomedical Ontology
  - http://bioontology.org/
- OBO Foundry
  - http://obofoundry.org/
  - Promote best practices in ontology development
  - 10 inclusion criteria

Open Biological Ontologies (OBO)

Main Criteria Ontologies Browse Project CVS Subscribe Contact

OBO Ontology Browser

Browse the tree by clicking on the category names; click on an ontology name to view more information on it.

- anatomy
- animal natural history and life history
- chemical
- development
- ethology
- evidence codes
- experimental conditions
- genomic and proteomic
- metabolomics
- OBO relationship types
- phenotype
- taxonomic classification
- vocabularies

http://obo.sourceforge.net/

Integrating subdomains

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

(Barry Smith)

OBO ontologies Examples

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

UMLS Source Vocabularies (2007AA)

- 139 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.5M names
  - 1.4M concepts
  - 16M relations
- Common presentation

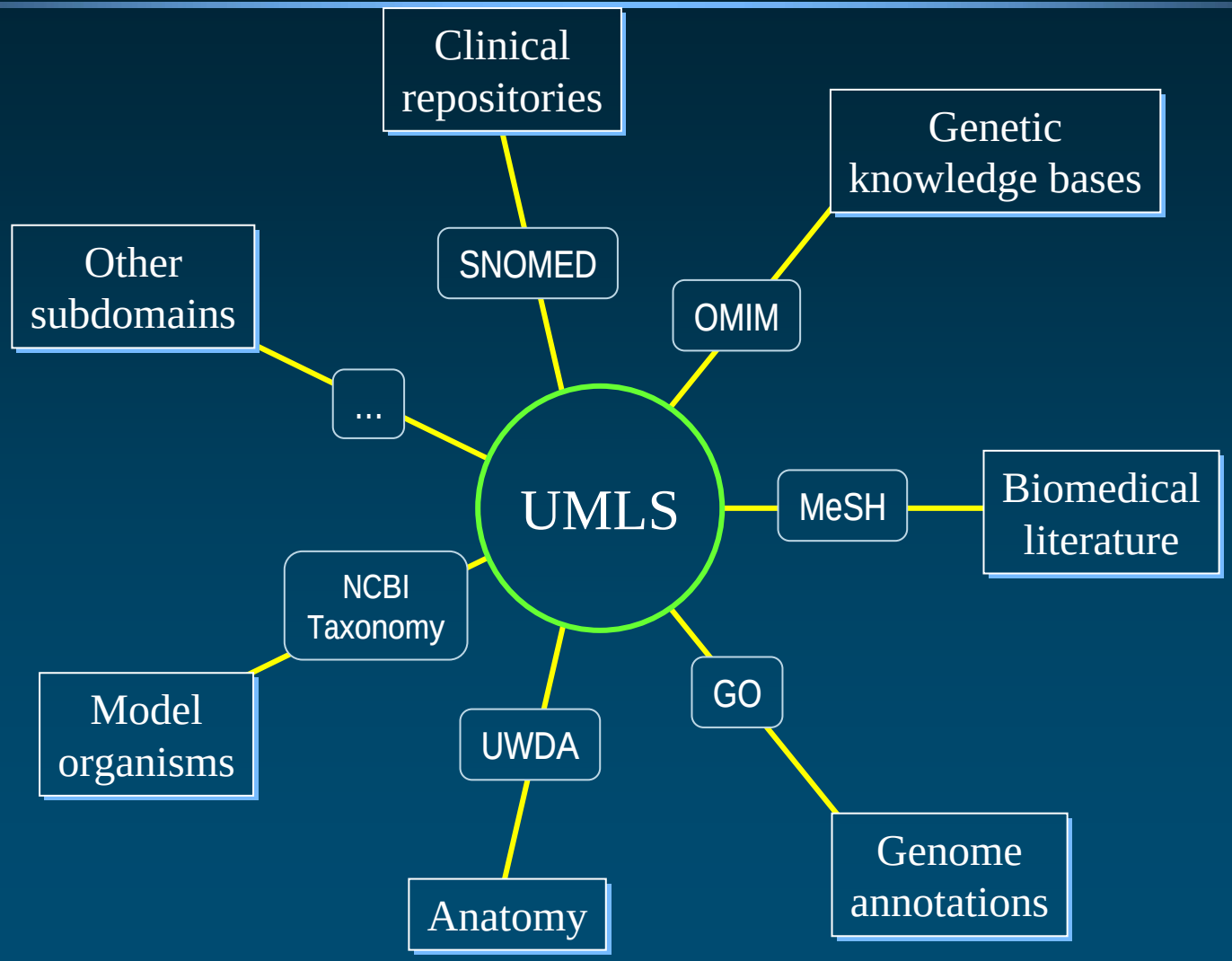
Biomedical terminologies in UMLS

- General vocabularies
  - anatomy (UWDA, Neuronames)
  - drugs (RxNorm, First DataBank, Micromedex, …)
  - medical devices (UMD, SPN)
- Several perspectives
  - clinical terms (SNOMED CT)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - data exchange terminologies (HL7, LOINC)

Biomedical terminologies in UMLS

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

Integrating subdomains



Tools and formalisms for bio-ontologies

Three examples

Three examples

- Foundational Model of Anatomy
  - Protégé-frames
- Gene Ontology
  - OBO-Edit
- NCI Thesaurus
  - OWL DL

- Conversions

Foundational Model of Anatomy (FMA)

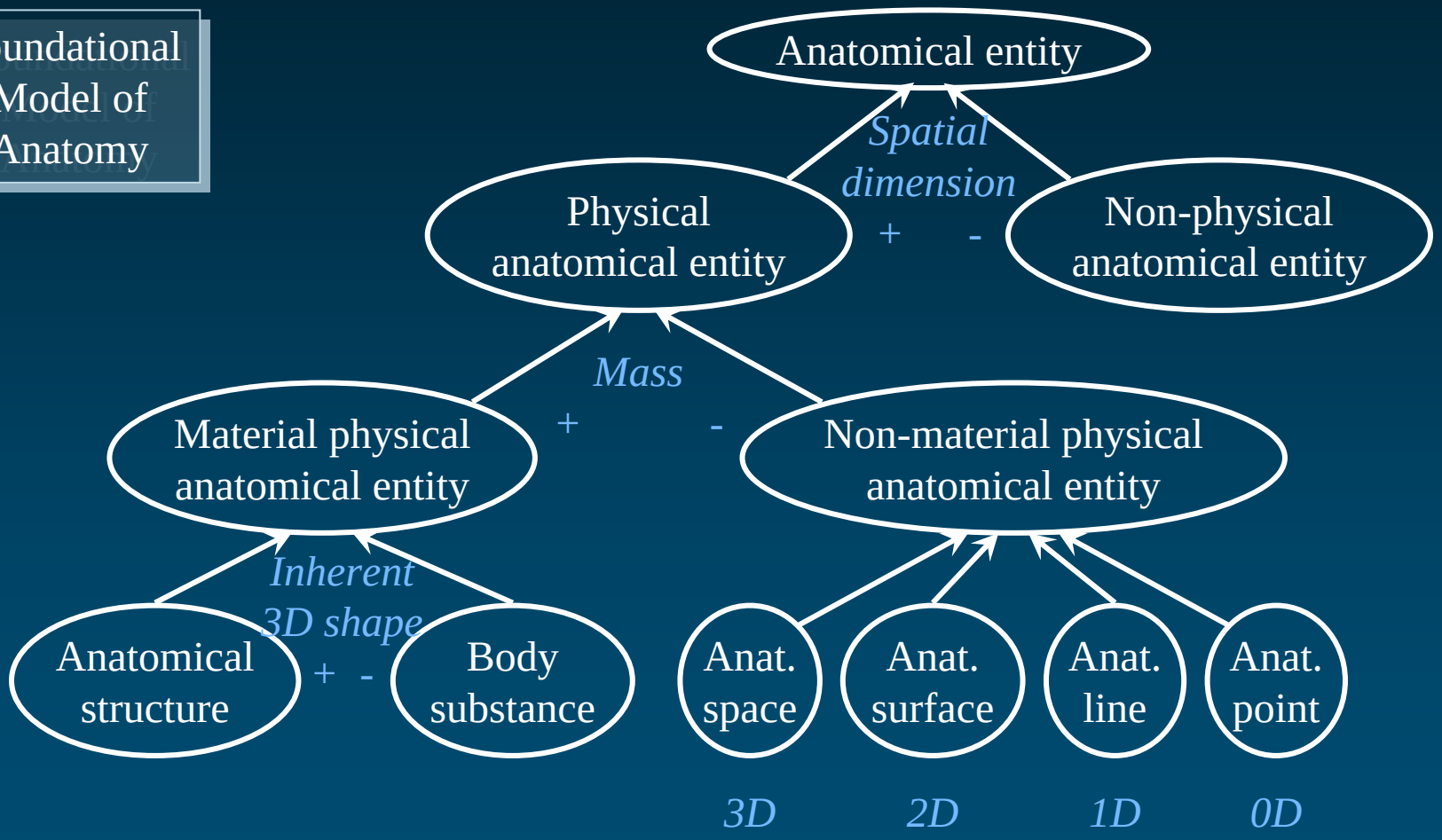
http://sig.biostr.washington.edu/projects/fm/index.html

- University of Washington
- Canonical anatomy
- 75,000 anatomical entities
- Synonyms
- Relationships
  - Isa
  - Part of (5 subtypes)
  - Topological, etc.
- Frame-based / Protégé

http://protege.stanford.edu/

Explicit classificatory principle

Foundational Model of Anatomy



FMA Conversions

- OWL DL
  - Golbreich et al., JWS 2006
- OWL Full
  - Noy and Rubin, SMI Tech Report 2007
- OBO
  - http://obofoundry.org/cgi-bin/detail.cgi?id=fma_lite

Gene Ontology

http://www.geneontology.org/

- GO Consortium
- Annotation of gene products (Molecular functions, Cellular components, Biological processes)
- 24,000 terms
- Synonyms
- Isa and part of relations
- OBO-Edit / OBO
- Also available in RDF and OWL DL

OBO-Edit
The OBO Ontology Editor

http://oboedit.org/

OBO format

- Used to represent many ontologies in the OBO family (Open Biological Ontologies)

http://www.godatabase.org/dev/doc/obo_format_spec.html

- Essentially a subset of OWL DL

```
[Term]
id: GO:0019563
name: glycerol catabolism
namespace: biological_process
def: "The chemical reactions and pathways resulting in the breakdown of glycerol …
subset: gosubset_prok
exact_synonym: "glycerol breakdown" []
exact_synonym: "glycerol degradation" []
xref_analog: MetaCyc:PWY0-381
is_a: GO:0006071 ! glycerol metabolism
is_a: GO:0046174 ! polyol catabolism
```

NCI Thesaurus

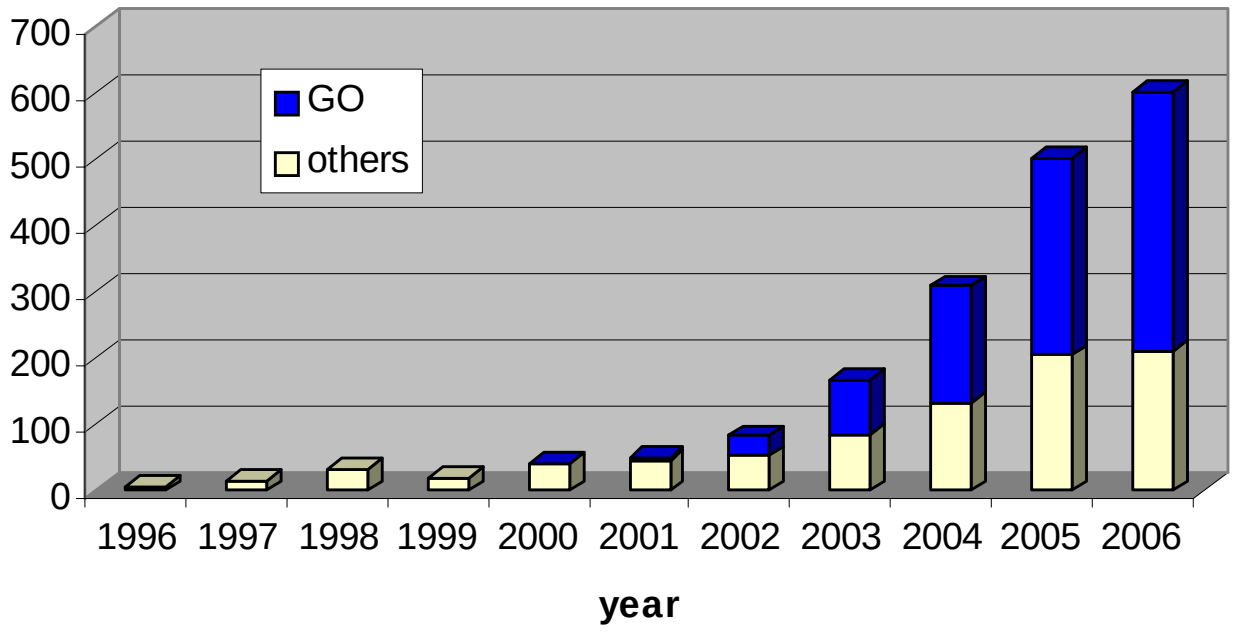
http://nciterms.nci.nih.gov/NCIBrowser/

- National Cancer Institute
- Cancer research
- 54,000 concepts
- 150,000 concept names
- Relations
  - Isa
  - Associative (87 relationship types)
- OWL DL

http://protege.stanford.edu/

Institutionalization of bio-ontologies

Bio-ontologies have become mainstram



Some institutions Bio-ontologies

- National Center for Biomedical Ontology
  - http://bioontology.org/
- OBO Foundry
  - http://obofoundry.org/
  - Promote best practices in ontology development

- Other ontology centers
  - NCOR – National Center for Ontology Research (US)
  - ECOR – European Center for Ontology Research

Some institutions Semantic Web

- W3C Health Care and Life Sciences Interest Group
  - http://www.w3.org/2001/sw/hcls/
  - BioRDF
  - BioOnt

Bio-ontologies
and Semantic Web

Use cases for a biomedical SW

- Integration
  - Data/Information
  - E.g., translational research
- Hypothesis generation
- Knowledge discovery

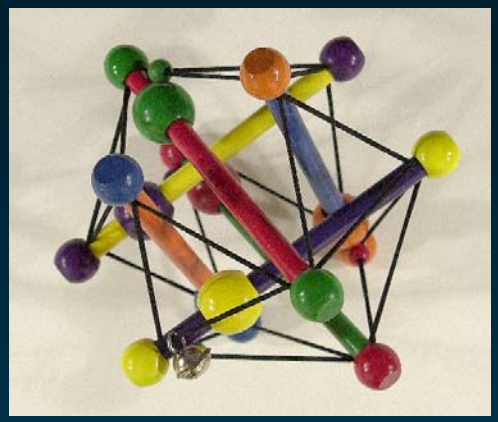
- Clinical data
  - Aggregation, sharing, exchange
  - Support for clinical decision

Some issues

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

Summary

- Biomedical information integration is a good use case for the Semantic Web
  - Semantic Web technologies
  - Ontologies
- Ontologies
  - Identification
  - Mapping
  - Reasoning



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