Stefan Schulz
Medical Informatics
Research Group
University Medical Center
Freiburg, Germany

Biomedical Ontologies
What are they (for)?
Understanding / Semantic Interoperability

Enables understanding between human and computational agents

Common language: Ontologies and Terminology Systems
Ontologies and Terminology Systems

- aka *Knowledge Organization Systems*: Systems that support semantic interoperability by communicating and processing information:
  - In a structured form
  - Well-defined
  - Unambiguous
  - Processable by machines
  - Understandable by humans

- Life Sciences: major focus for the development of ontologies and terminological systems
Literature on Biomedical Terminologies and Ontologies
Purpose of this Talk

- What are Ontologies
- What are they for?
Introduction - Current Systems
Terminological Clarification
What do Formal Ontologies Represent?
Terminologies vs. Formal Ontologies
Practice of Good Ontology
Outlook
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
A cruise through the archipelago of systems for biomedical knowledge organization
MeSH
Medical Subject Headings
MeSH Tree Structures - 2006

1. Anatomy [A]
2. Organisms [B]
   - Animals [B01] +
   - Algae [B02] +
   - Bacteria [B03] +
   - Viruses [B04] +
   - Fungi [B05] +
   - Plants [B06] +
   - Archaea [B07] +
   - Mesomyctozoa [B08] +
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. Physical Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology and Food and Beverages [J]
11. Humanities [K]
12. Information Science [L]
13. Persons [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographic Locations [Z]
Hierarchical principle: broader term / narrower term (not a taxonomy)
Bacteria [B03]
Gram-Positive Bacteria [B03.510]
Gram-Positive Cocci [B03.510.400]
Staphylococcaceae [B03.510.400.790]

► Staphylococcus [B03.510.400.790.750]
  Staphylococcus aureus [B03.510.400.790.750.100]
  Staphylococcus epidermidis [B03.510.400.790.750.343]
  Staphylococcus haemolyticus [B03.510.400.790.750.400]
  Staphylococcus hominis [B03.510.400.790.750.425]

Return to Entry Page
National Library of Medicine - Medical Subject Headings

2006 MeSH

MeSH Descriptor Data

Return to Entry Page

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Staphylococcus aureus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>B03.510.400.790.750.100</td>
</tr>
<tr>
<td>Annotation</td>
<td>Infection = STAPHYLOCOCCAL INFECTIONS &amp; do not bother to coord with S. aureus unless particularly discussed (index IM); DF: STAPH AUREUS</td>
</tr>
<tr>
<td>Scope Note</td>
<td>Potentially pathogenic bacteria found in nasal membranes, skin, hair follicles, and perineum of warm-blooded animals. They may cause a wide range of infections and intoxications.</td>
</tr>
<tr>
<td>Allowable Qualifiers</td>
<td>CH CL CY DE EN GD GE IM IP ME PH PY RE UL VI</td>
</tr>
<tr>
<td>Entry Version</td>
<td>STAPH AUREUS</td>
</tr>
<tr>
<td>Previous Indexing</td>
<td>Staphylococcus (1966-1974)</td>
</tr>
<tr>
<td>Online Note</td>
<td>use STAPHYLOCOCCUS AUREUS to search MICROCOCCUS PYOGENES 1975-91, use STAPHYLOCOCCUS 1966-74</td>
</tr>
<tr>
<td>History Note</td>
<td>76, was MICROCOCCUS PYOGENES see under STAPHYLOCOCCUS 1963-75, MICROCOCCUS PYOGENES was see STAPHYLOCOCCUS AUREUS 1976-91</td>
</tr>
<tr>
<td>Unique ID</td>
<td>D013211</td>
</tr>
</tbody>
</table>

MeSH Tree Structures

- Bacteria [B03]
  - Gram-Positive Bacteria [B03.510]
    - Gram-Positive Cocci [B03.510.400]
      - Staphylococcoccae [B03.510.400.790]
        - Staphylococcus [B03.510.400.790.750]
          - Staphylococcus aureus [B03.510.400.790.750.100]
            - Staphylococcus epidermidis [B03.510.400.790.750.343]
            - Staphylococcus haemolyticus [B03.510.400.790.750.400]
            - Staphylococcus hominis [B03.510.400.790.750.425]
GO
Gene Ontology
Part of (partonomy)

Is a (taxonomy)
ICD
International Classification of Diseases
Internationale Statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
10. Revision
Version 2006

German Modification

Vierstellige Ausführliche Systematik

Kapitelübersicht

<table>
<thead>
<tr>
<th>Kapitel</th>
<th>Gliederung</th>
<th>Titel</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A00-B99</td>
<td>Bestimmte infektiöse und parasitäre Krankheiten</td>
</tr>
<tr>
<td>II</td>
<td>C00-D48</td>
<td>Neubildungen</td>
</tr>
<tr>
<td>III</td>
<td>D50-D99</td>
<td>Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems</td>
</tr>
<tr>
<td>IV</td>
<td>E00-E99</td>
<td>Endokrine, Ernährungs- und Stoffwechselkrankheiten</td>
</tr>
<tr>
<td>V</td>
<td>F00-F93</td>
<td>Psychische und Verhaltensstörungen</td>
</tr>
<tr>
<td>VI</td>
<td>G00-G99</td>
<td>Krankheiten des Nervensystems</td>
</tr>
<tr>
<td>VII</td>
<td>H00-H59</td>
<td>Krankheiten des Auges und der Augenanhängegebilde</td>
</tr>
<tr>
<td>VIII</td>
<td>H60-H85</td>
<td>Krankheiten des Ohres und des Warzenfortsatzes</td>
</tr>
<tr>
<td>IX</td>
<td>I00-I99</td>
<td>Krankheiten des Kreislaufsystems</td>
</tr>
<tr>
<td>X</td>
<td>J00-J99</td>
<td>Krankheiten des Atmungssystems</td>
</tr>
<tr>
<td>XI</td>
<td>K00-K93</td>
<td>Krankheiten des Verdauungssystems</td>
</tr>
<tr>
<td>XII</td>
<td>L00-L99</td>
<td>Krankheiten der Haut und der Unterhaut</td>
</tr>
<tr>
<td>XIII</td>
<td>M00-M99</td>
<td>Krankheiten des Muskel-Skelett-Systems und des Bindegewebes</td>
</tr>
<tr>
<td>XIV</td>
<td>N00-N99</td>
<td>Krankheiten des Urogenitalsystems</td>
</tr>
<tr>
<td>XV</td>
<td>O00-O99</td>
<td>Schwangerschaft, Geburt und Wochenbett</td>
</tr>
<tr>
<td>XVI</td>
<td>P00-P95</td>
<td>Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben</td>
</tr>
<tr>
<td>XVII</td>
<td>Q00-Q99</td>
<td>Angeborene Fehlbildungen, Deformitäten und Chromosomenanomalien</td>
</tr>
<tr>
<td>XVIII</td>
<td>R00-R99</td>
<td>Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind</td>
</tr>
<tr>
<td>XIX</td>
<td>S00-T98</td>
<td>Verletzungen, Vergiftungen und bestimmte andere Folgen äußerer Ursachen</td>
</tr>
<tr>
<td>XX</td>
<td>V01-V98</td>
<td>Äußere Ursachen von Morbidität und Mortalität</td>
</tr>
<tr>
<td>XXI</td>
<td>Z00-Z99</td>
<td>Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen</td>
</tr>
<tr>
<td>XXII</td>
<td>U00-U99</td>
<td>Schlüsselnummern für besondere Zwecke</td>
</tr>
</tbody>
</table>

Kapitel I:

Bestimmte infektiöse und parasitäre Krankheiten (A00-B99)
Kapitel II:

Neubildungen (C00-D48)

C00-C07 Bosartige Neubildungen
   C00-C07 Bosartige Neubildungen an genau bezeichneten Lokalisationen, als primär festgestellt oder vermutet, ausgenommen lymphatisches, blutbildendes und verwandtes Gewebe
   C00-C14 Lippe, Mundhöhle und Pharynx
   C15-C26 Verdauungssorgane
   C30-C39 Atmungsorgane und sonstige intrathorakale Organe
   C40-C41 Knochen und Gelenkkapsel
   C43-C44 Haut
   C45-C49 Mesotheliales Gewebe und Weichteilgewebe
   C50 Brustdrüse (Mamma)
   C51-C59 Weibliche Genitalorgane
   C60-C63 Mänliche Genitalorgane
   C64-C68 Hamorgane
   C69-C72 Auge, Gehirn und sonstige Teile des Zentralnervensystems
   C73-C75 Schilddrüse und sonstige endokrine Drüsen

C76-C80 Bosartige Neubildungen unge nau bezeichnet, sekundärer und nicht näher bezeichneter Lokalisationen

C81-C96 Bosartige Neubildungen des lymphatischen, blutbildenden und verwandten Gewebes, als primär festgestellt oder vermutet

C97 Bosartige Neubildungen als Primartumoren an mehreren Lokalisationen

D00-D08 In-situ-Neubildungen

D10-D35 Gutartige Neubildungen

D37-D48 Neubildungen unsicherer oder unbekannter Verhaltens [siehe Hinweis am Anfang der Krankheitsgruppe D37-D48]

Kapitel III:

Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems (D50-D90)

D50-D52 Alimentare Anämien

D55-D59 hämatologische Anämien

D56-D58 Aplastische und sonstige Anämien

D55-D59 Koagulopathien, Purpura und sonstige hamorrhagische Diathesen

D70-D77 Sonstige Krankheiten des Blutes und der blutbildenden Organe

D80-D90 Bestimmte Störungen mit Beteiligung des Immunsystems

Kapitel IV:
Class / subclass Relation (is_a)

Inkl.: Mittelohr
Exkl.: Mesothelium (C45.0)

C30.0 Bösartige Neubildung der Nasenhöhle und des Mittelohres
Conchae nasales
Naseninnernasum
Nasenknorpel
Nasenseptum
Vestibulum nasi

Exkl.: Bulbus olfactorius (C72.2)
Haut der Nase (C43.0, C44.3)
Hinterrand des Nasenseptums und der Choanen (C11.3)
Nase o.n.A. (C75.0)
Nasenbein (C41.02)

C30.1 Mittelohr
Cellulae mastoideaer
Innenohr
Tuba auditiva [Eustachio]

Exkl.: Gehörgang (äußerer) (C43.2, C44.2)
Haut des (äußeren) Ohres (C43.2, C44.2)
Knöcherner Gehörgang (Meatus) (C41.01)
Ohrknorpel (C49.0)

C31.0 Bösartige Neubildung der Nasennebenhöhlen
Sinus maxillaris [Kieferhöhle]
Antrum maxillare [Highmore-Höhle]

C31.1 Sinus ethmoidalis [Siebbeinzellen]

C31.2 Sinus frontalis [Stirnhöhle]

C31.3 Sinus sphenoidalis [Keilbeinhöhle]

C31.8 Nasennebenhöhlen, mehrere Teilbereiche übergreifend
[Siehe Hinweis 5 am Anfang dieses Kapitels]

C31.9 Nasennebenhöhlen, nicht näher bezeichnet

C32.0 Bösartige Neubildung des Larynx
Glottis
Lig. vocale [echtes Stimmband] o.n.A.
Ventriculus laryngis

C32.1 Supraglottis
Anepiglottische Falte, laryngeale Seite
Epiglottis (suprahyoideal Anteil) o.n.A.
Hintere (laryngeale) Fläche der Epiglottis
Flecht vestibularis
SNOMED
Clinical Terms
acute myocardial infarction of anterior wall

Details of 'acute myocardial infarction of anterior wall'

ConceptStatus Current

Descriptions
- acute myocardial infarction of anterior wall (disorder)
- acute myocardial infarction of anterior wall
- acute anterior myocardial infarction

Fully defined by...
- acute myocardial infarction
- acute

Course
- Associated morphology
  - acute infarct
- Finding site
- structure of anterior myocardium

Qualifiers
- Onset
  - sudden onset
  - gradual onset
- Severity
  - severities
- Episodicity
  - Episodicities

Legacy codes
- SNOMED: D3-15110
- CTV3ID: Xa0YL

Hierarchy for 'myocardial infarction'
- injury of anatomical site
- structural disorder of heart
- myocardial necrosis
- myocardial disease
- myocardial infarction
  - acute myocardial infarction
  - acute anteroseptal myocardial infarction
  - acute anteroseptal myocardial infarction
  - acute non-Q-wave infarction - anteroseptal
  - acute infarction of papillary muscle
  - acute myocardial infarction NOS
    - acute myocardial infarction of anterior wall
      - acute myocardial infarction of anterolateral wall
      - acute myocardial infarction of inferolateral wall
      - acute myocardial infarction of inferoposterior wall
      - acute myocardial infarction of lateral wall
      - acute myocardial infarction of inferior wall
SNOMED CT Facts (I)

- SNOMED CT is a terminology consisting of terms used in health & health care,
- attached to concept codes with multiple terms per code
- structured according to logic-based representation of meanings
- increasingly guided by ontological principles
- Current size:
  - 283,000 Concepts
  - 732,000 Terms
  - 923,000 Concept – Concept Relations
Since 2007: Maintained by IHTSDO (International Health Terminology standards development organization)

Members: Australia, Canada, Denmark, Lithuania, The Netherlands, New Zealand, Sweden, UK, USA.

Annual budget ~ 5 M€
Different Purposes – Heterogeneous Approaches
Different Purposes – Heterogeneous Approaches

- **MeSH** [Medical Subject Headings]:
  Hierarchy (broader / narrower) of *descriptors*, used for indexing biomedical publications for literature retrieval support

- **GO** [Gene Ontology]:
  Hierarchy (is_a / part_of) of controlled terms for describing gene and gene product properties

- **ICD** [International Classification of Diseases]:
  Strict Hierarchy of non-overlapping classes for classifying statistically relevant health conditions

- **SNOMED CT** [Systematized Nomenclature of Medicine – Clinical Terms]:
  Hierarchical system of concepts with (partially) logic-based concept definitions
<table>
<thead>
<tr>
<th>Other Biomedical Knowledge Organization Systems: Medicine</th>
<th>Source: UMLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI/RHEUM</td>
<td></td>
</tr>
<tr>
<td>Alcohol and Other Drug Thesaurus</td>
<td></td>
</tr>
<tr>
<td>Alternative Billing Concepts</td>
<td></td>
</tr>
<tr>
<td>Beth Israel Vocabulary</td>
<td></td>
</tr>
<tr>
<td>Canonical Clinical Problem Statement System</td>
<td></td>
</tr>
<tr>
<td>Clinical Classifications Software</td>
<td></td>
</tr>
<tr>
<td>Clinical Terms Version 3 (CTV3) (Read Codes)</td>
<td></td>
</tr>
<tr>
<td>Common Terminology Criteria for Adverse Events</td>
<td></td>
</tr>
<tr>
<td>COSTAR</td>
<td></td>
</tr>
<tr>
<td>COSTART</td>
<td></td>
</tr>
<tr>
<td>CRISP Thesaurus</td>
<td></td>
</tr>
<tr>
<td>Current Dental Terminology 2005 (CDT-5)</td>
<td></td>
</tr>
<tr>
<td>Current Procedural Terminology</td>
<td></td>
</tr>
<tr>
<td>Diseases Database</td>
<td></td>
</tr>
<tr>
<td>DSM-III-R</td>
<td></td>
</tr>
<tr>
<td>DSM-IV</td>
<td></td>
</tr>
<tr>
<td>DXplain</td>
<td></td>
</tr>
<tr>
<td>Gene Ontology</td>
<td></td>
</tr>
<tr>
<td>HCPCS Version of Current Dental Terminology 2005 (CDT-5)</td>
<td></td>
</tr>
<tr>
<td>Healthcare Common Procedure Coding System</td>
<td></td>
</tr>
<tr>
<td>HL7 Vocabulary Version 2.5</td>
<td></td>
</tr>
<tr>
<td>HL7 Vocabulary Version 3.0</td>
<td></td>
</tr>
<tr>
<td>Home Health Care Classification</td>
<td></td>
</tr>
<tr>
<td>HUGO Gene Nomenclature</td>
<td></td>
</tr>
<tr>
<td>ICD10</td>
<td></td>
</tr>
<tr>
<td>ICD-9-CM</td>
<td></td>
</tr>
<tr>
<td>ICPC</td>
<td></td>
</tr>
<tr>
<td>ICPC2 - ICD10 Thesaurus</td>
<td></td>
</tr>
<tr>
<td>ICPC2-ICD10 Thesaurus</td>
<td></td>
</tr>
<tr>
<td>International Classification of Primary Care</td>
<td></td>
</tr>
<tr>
<td>International Classification of Primary Care 2nd Edition</td>
<td></td>
</tr>
<tr>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
<td></td>
</tr>
<tr>
<td>JAMAS Japanese Medical Thesaurus (JJMT)</td>
<td></td>
</tr>
<tr>
<td>Library of Congress Subject Headings</td>
<td></td>
</tr>
<tr>
<td>LOINC 2.15</td>
<td></td>
</tr>
<tr>
<td>Master Drug Data Base</td>
<td></td>
</tr>
<tr>
<td>McMaster University Epidemiology Terms</td>
<td></td>
</tr>
<tr>
<td>Medical Dictionary for Regulatory Activities Terminology (MedDRA)</td>
<td></td>
</tr>
<tr>
<td>Medical Entities Dictionary</td>
<td></td>
</tr>
<tr>
<td>Medical Subject Headings</td>
<td></td>
</tr>
<tr>
<td>MEDLINE (1996-2000)</td>
<td></td>
</tr>
<tr>
<td>MEDLINE (2001-2006)</td>
<td></td>
</tr>
<tr>
<td>MedlinePlus Health</td>
<td></td>
</tr>
<tr>
<td>Topics_2004_08_14</td>
<td></td>
</tr>
<tr>
<td>Micromedex DRUGDEX</td>
<td></td>
</tr>
<tr>
<td>Multum MediSource Lexicon</td>
<td></td>
</tr>
<tr>
<td>NANDA nursing diagnoses: definitions &amp; classification</td>
<td></td>
</tr>
<tr>
<td>National Drug Data File Plus Source Vocabulary</td>
<td></td>
</tr>
<tr>
<td>National Drug File - Reference Terminology</td>
<td></td>
</tr>
<tr>
<td>National Library of Medicine Medline Data</td>
<td></td>
</tr>
<tr>
<td>NCBI Taxonomy</td>
<td></td>
</tr>
<tr>
<td>NCI SEER ICD Neoplasm Code Mappings</td>
<td></td>
</tr>
<tr>
<td>NCI Thesaurus</td>
<td></td>
</tr>
<tr>
<td>Neuronames Brain Hierarchy</td>
<td></td>
</tr>
<tr>
<td>Nursing Interventions Classification</td>
<td></td>
</tr>
<tr>
<td>Nursing Outcomes Classification</td>
<td></td>
</tr>
<tr>
<td>Omaha System</td>
<td></td>
</tr>
<tr>
<td>Online Congenital Multiple Anomaly/Mental Retardation Syndromes</td>
<td></td>
</tr>
<tr>
<td>Online Mendelian Inheritance in Man</td>
<td></td>
</tr>
<tr>
<td>Patient Care Data Set</td>
<td></td>
</tr>
<tr>
<td>Perioperative Nursing Data Set</td>
<td></td>
</tr>
<tr>
<td>Pharmacy Practice Activity Classification</td>
<td></td>
</tr>
<tr>
<td>Physician Data Query</td>
<td></td>
</tr>
<tr>
<td>Physicians’ Current Procedural Terminology</td>
<td></td>
</tr>
<tr>
<td>Quick Medical Reference (QMR)</td>
<td></td>
</tr>
<tr>
<td>Read thesaurus</td>
<td></td>
</tr>
<tr>
<td>Read thesaurus Americanized Synthesized Terms</td>
<td></td>
</tr>
<tr>
<td>RXNORM Project</td>
<td></td>
</tr>
<tr>
<td>SNOMED-2</td>
<td></td>
</tr>
<tr>
<td>SNOMED Clinical Terms</td>
<td></td>
</tr>
<tr>
<td>SNOMED International</td>
<td></td>
</tr>
<tr>
<td>Standard Product Nomenclature</td>
<td></td>
</tr>
<tr>
<td>Thesaurus of Psychological Index Terms</td>
<td></td>
</tr>
<tr>
<td>The Universal Medical Device Nomenclature System (UMDNS)</td>
<td></td>
</tr>
<tr>
<td>UltraSTAR</td>
<td></td>
</tr>
<tr>
<td>UMLS Metathesaurus</td>
<td></td>
</tr>
<tr>
<td>University of Washington Digital Anatomist</td>
<td></td>
</tr>
<tr>
<td>USP Model Guidelines</td>
<td></td>
</tr>
<tr>
<td>Veterans Health Administration National Drug File</td>
<td></td>
</tr>
<tr>
<td>WHO Adverse Reaction Terminology</td>
<td></td>
</tr>
<tr>
<td>WHOART</td>
<td></td>
</tr>
</tbody>
</table>
Other Biomedical Knowledge Organization Systems: Biology (OBO)

- Anatomy
  - Cell type
  - Gross anatomy
    - Animal gross anatomy
      - C. elegans gross anatomy
      - Drosophila gross anatomy
    - eVoc (Expressed Sequence Annotation for Humans)
      - Fish anatomy
        - Medaka fish anatomy and development
        - Zebrafish anatomy and development
    - Human developmental anatomy
      - Human developmental anatomy, timed
      - Human developmental anatomy, abstract
      - Mosquito gross anatomy
    - Mouse anatomy
      - Mouse adult anatomy
      - Mouse anatomy and development
  - Microbial anatomy
    - Fungal anatomy
    - Dictyostelium anatomy
  - Plant anatomy
    - Arabidopsis anatomy
    - Cereal anatomy
    - Maize anatomy
    - Plant anatomy
- Organ
- BRENDA tissue / enzyme source

- Animal natural history and life history
- Chemical
  - Chemical entities of biological interest
  - Physico-chemical methods and properties
  - Physico-chemical process
  - Systems biology
- Development
  - Animal development
    - C. elegans development
    - Drosophila development
    - eVoc (Expressed Sequence Annotation for Humans)
    - Medaka fish anatomy and development
    - Zebrafish anatomy and development
  - Mouse anatomy and development
    - Plasmodium life cycle
    - Zebrafish anatomy and development
- Plant development
  - Plant growth and developmental stage
  - Arabidopsis development
  - Cereal plant development
- Ethology
  - Habronattus courtship
  - Loggerhead nesting
- Evidence codes
- Experimental conditions
  - Biological imaging methods
  - Microarray experimental conditions
  - Nuclear magnetic resonance experiment conditions
  - Ontology for biomedical investigations
  - Physico-chemical methods and properties
  - Sample processing and separation techniques
- Genomic and proteomic
  - Gene product
    - Biological process
    - Cellular component
    - Event
    - Gene product name
    - Molecular function
    - Molecule role
    - Multiple alignment
    - Pathway
    - Protein
      - Protein covalent bond
      - Protein domain
      - Protein modification
      - Protein-protein interaction
    - Proteomics: data and process provenance
      - Sequence types and features
- Metabolomics
  - Nuclear magnetic resonance experiment conditions
- OBO relationship types
- Phenotype
  - Cereal plant trait
  - eVoc (Expressed Sequence Annotation for Humans)
  - Human diseases
  - Mammalian phenotype
  - Medaka fish anatomy and development
  - Mouse pathology
  - Plant environmental conditions
  - PATO
- Taxonomic classification
  - Fly taxonomy
  - NCBI organismal classification
  - SwissProt organismal classification
- Vocabularies
  - FlyBase controlled vocabulary
  - MESH
  - NCI Thesaurus
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
Unresolved Terminological Confusion…

- Knowledge Organization Systems: artifacts for ordering domain entities, relating word meanings or providing semantic reference:
  - Vocabularies
  - Terminologies
  - Thesauri
  - Concept Systems
  - Classifications
  - (Formal) Ontologies
Different scientific traditions: Biology, Medicine, Philosophy, Logic, Linguistics, Library and Information Science, Computer Science, Cognitive Science, International Terminology norms

Different philosophical schools of thinking: Platonism, Aristotelian Realism, Conceptualism, Relativism, Idealism, Postmodernism, Constructivism, Nominalism, Tropism,…
**Components of Knowledge Organization Systems**

**Dictionaries of Natural language Terms**
- Benign neoplasm of heart
- Benign tumor of heart
- Benign tumour of heart
- Benign cardiac neoplasm
- Gutartiger Herzumor
- Gutartige Neubildung am Herzen
- Gutartige Neubildung: Herz
- Gutartige Neoplasie des Herzens
- Tumeur bénigne cardiaque
- Tumeur bénigne du cœur
- Neoplasia cardiaca benigna
- Neoplasia benigna do coração
- Neoplasia benigna del corazón
- Tumor benigno do coração

**Hierarchically ordered Nodes and Links**

**Formal or informal Definitions**

**domain or region of DNA [GENIA]:** A substructure of DNA molecule which is supposed to have a particular function, such as a gene, e.g., c-jun gene, promoter region, Sp1 site, CA repeat. This class also includes a base sequence that has a particular function.

**Peptides [MeSH]:** Members of the class of compounds composed of AMINO ACIDS joined together by peptide bonds between adjacent amino acids into linear, branched or cyclical structures. OLIGOPEPTIDES are composed of approximately 2-12 amino acids. Polypeptides are composed of approximately 13 or more amino acids. PROTEINS are linear polypeptides that are normally synthesized on RIBOSOMES.

```
19429009|chronic ulcer of skin|
116680003|is a|=64572001|disease|
{116676008|associated morphology|=405719001|chronic ulcer|
363698007|finding site|=39937001|skin structure|}
```
What do the nodes in Formal Ontologies / Terminological Systems stand for?

- names
- concepts
- classes
- terms
- categories
- types
- entities
- descriptors
- universals
- sets
- synsets
- sorts
- properties
- descriptors
Ontology: Gradient or crisp boundary?
Ontology: Gradient or crisp boundary?
Organizing the world

Terminology

Set of terms representing the system of concepts of a particular subject field. (ISO 1087)

Formal Ontology

Ontology is the study of what there is. Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities. (Stanford Encyclopedia of Philosophy)
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
Terminologies start with human language

**Terminology**

Set of terms representing the system of concepts of a particular subject field. (ISO 1087)

**Ontology**

Ontology is the study of what there is. Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities. (Stanford Encyclopedia of Philosophy)
Entities of Language (Terms)

"benign neoplasm of heart"
"gutartige Neubildung des Herzmuskels"
"neoplasia cardiaca benigna"

Shared / Meanings / Entities of Thought (Concepts)
Example: UMLS (mrconso table)

| C0153957 | ENG | P | L0180790 | PF | S1084242 | Y | A1141630 | || | MTH | PN | U001287 | benign neoplasm of heart | 0 | N ||
| C0153957 | ENG | P | L0180790 | VC | S0245316 | N | A0270815 | || | ICD9CM | PT | 212.7 | Benign neoplasm of heart | 0 | N ||
| C0153957 | ENG | P | L0180790 | VC | S0245316 | N | A0270817 | || | RCD | SY | B727. | Benign neoplasm of heart | 3 | N ||
| C0153957 | ENG | P | L0180790 | VO | S1446737 | Y | A1406658 | || | SNMI | PT | D3-F0100 | Benign neoplasm of heart, NOS | 3 | N ||
| C0153957 | ENG | S | L0524277 | PF | S0599118 | N | A0654589 | || | RCDAE | PT | B727. | Benign tumor of heart | 3 | N ||
| C0153957 | ENG | S | L0524277 | VO | S0599510 | N | A0654975 | || | RCD | PT | B727. | Benign tumour of heart | 3 | N ||
| C0153957 | ENG | S | L0018787 | PF | S0047194 | Y | A0066366 | || | ICD10 | PS | D15.1 | Heart | 3 | Y ||
| C0153957 | ENG | S | L0018787 | VO | S0900815 | Y | A0957792 | || | MTH | MM | U003158 | Heart <3> | 0 | Y ||
| C0153957 | ENG | S | L1371329 | PF | S1624801 | N | A1583056 | || | 10004245 | MDR | LT | 10004245 | Benign cardiac neoplasm | 3 | N ||
| C0153957 | GER | P | L1258174 | PF | S1500120 | Y | A1450314 | || | DMDICD10 | PT | D15.1 | Gutartige Neubildung: Herz | 1 | N ||
| C0153957 | SPA | P | L2354284 | PF | S2790139 | N | A2809706 | || | MDRSPA | LT | 10004245 | Neoplasia cardiaca benigna | 3 | N ||

Example: UMLS (mrrel table)
Example: UMLS

Semantic relations

INFORMAL

Shared / Meanings / Entities of Thought
Formal Ontology represents the world

Set of terms representing the system of concepts of a particular subject field. (ISO 1087)

Ontology is the study of what there is (Quine). Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities. (Stanford Encyclopedia of Philosophy)
Organizing Entities

Entity Types

The type “benign neoplasm of heart”

Entities of the World

My benign neoplasm of heart
Organizing Entities

Entity Types
- Universals, classes, (Concepts)
  - abstract
  - concrete

Entities of the World
- Particulars, instances
  - The type “benign neoplasm of heart”
  - The benign neoplasm of my heart
Organizing Entities

Entities of Language
Terms, names

The string „benign neoplasm of heart“

Entity Types
Universals, classes, (Concepts)

abstract

Entity Types
concrete

Instances of

Entities of the World
Particulars, instances

The type “benign neoplasm of heart”

The benign neoplasm of my heart
Organizing Entities

represents

Gutartigen Herztumors

(the complication of my) benign heart tumor
Organizing Entities

represents

(the) benign heart tumor (is congenital)

Komplikation meines Gutartigen Herztumors
Entities of Language

...are stored in dictionaries and represented by terminologies
Database systems / information models store references to...

Entities of the World
Entity Types

... are organized in formal ontologies
Hierarchical framework for entity types

- **Taxonomy:** relates types and subtypes:
  - *Tumor of Heart is_a Tumor* equivalent to:
  - All instances of *Tumor of Heart* are instances of *Tumor* (without exceptions)

- **Relations:**
  - *instance_of* relates instances with types, all others relate instances (e.g. *part_of*) or are derived from them (e.g. *is_a*)

- **Definitions:** describe what is always true for all instances of a type
  - *Tumor of Heart has_location Heart* :
    All instances of *Tumor of Heart* are located in some *Heart*
Type / Subtype Hierarchy

- Benign Tumor of Heart
- Malignant Tumor of Heart
- Benign Tumor
- Tumor of Heart

Is_a relationships:
- Benign Tumor of Heart is_a Benign Tumor
- Tumor of Heart is_a Benign Tumor
- Tumor of Heart is_a Malignant Tumor
A classification view on Formal Ontologies

World
Hierarchies, Types, Classes, Individuals

World
Hierarchies, Types, Classes, Individuals

Type 1

World
Hierarchies, Types, Classes, Individuals

Formal Ontology

Type 1

Subtype 1.1

Subtype 1.2

Subtype 1.3

World
Hierarchies, Types, Classes, Individuals

Formal Ontology

Inflammatory Disease
Hierarchies, Types, Classes, Individuals

Formal Ontology

- Gastritis
- Hepatitis
- Pacreatitis

Is_a relationship:
- Gastritis Is_a Inflammatory Disease
- Hepatitis Is_a Inflammatory Disease
- Pacreatitis Is_a Inflammatory Disease
Hierarchies, Types, Classes, Individuals

Formal Ontology

- Gastritis
- Hepatitis
- Pacreatitis

Is_a relationships:
- Gastritis -> Inflammatory Disease
- Hepatitis -> Inflammatory Disease
- Pacreatitis -> Inflammatory Disease
Hierarchies, Types, Classes, Individuals

Formal Ontology

- Gastritis
- Inflammatory Disease
- Hepatitis
- Pacreatitis

Is_a relationships:
- Gastritis is_a Inflammatory Disease
- Hepatitis is_a Inflammatory Disease
- Pacreatitis is_a Inflammatory Disease
Informal Ontology

- **Liver** has **Inflammatory Disease** as a location.
- **Hepatitis** is an **Is_a** relationship to **Inflammatory Disease**.
- **Liver** has **Location** relationship to **Hepatitis**.
Formal Ontology

- **Inflammatory Disease**
- **Hepatitis**
- **Liver**

Relationships:
- Hepatitis is an inflammatory disease.
- Hepatitis has a location in the liver.
Relations and Definitions

Formal Ontology

Population of Virus

Is_a

Viral Hepatitis

Is_a

caused by

Liver

has Location

Hepatitis

Inflammatory Disease

Is_a
Languages for formal ontologies

- **Natural Language**
  
  “Every hepatitis is an inflammatory disease that is located in some liver”
  “Every inflammatory disease that is located in some liver is an hepatitis”

- **Logic**
  
  \[
  \forall x: \text{instanceOf}(x, \text{Hepatitis}) \iff \text{instanceOf}(x, \text{Inflammation}) \land \\
  \exists y: \text{instanceOf}(y, \text{Liver}) \land \text{hasLocation}(x, y)
  \]

Logic is computable: it supports machine inferences but…

it only scales up if it has a very limited expressivity
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
Terminologies vs. Formal Ontologies

Terminologies

- **Describe**: Meaning of human language units
- “Concepts”: aggregate (quasi)-synonymous terms
- **Relations**: informal, elastic Associations between Concepts
- **Description pattern**: Concept$_1$ Relation Concept$_2$

Formal Ontologies

- **Describe**: entities of reality as they generically are – independent of human language
- “Types”: represent the generic properties of world entities
- **Relations**: rigid, exactly defined, quantified relationships between particulars
- **Description pattern**: for all instance of Type$_1$ : there is some…
Example Hepatitis - Liver

**Terminologies**

- **Concept Hepatitis:**
  \{Hepatitis (D), Leberentzündung (D), hepatitis (E), hépatite (F)\}

- **Concept Liver:**
  \{Leber (D), liver (E), foie (F)\}

- **Relations:**
  - Hepatitis – hasLocation – Liver
  - Hepatitis – isA - Inflammation

**Formal Ontologies**

- **Type: Hepatitis:**

- **Description:**
  "Every hepatitis is an inflammatory disease that is located in some liver"
  "Every inflammatory disease that is located in some liver is an hepatitis"
Example Hand - Thumb

Terminologies

- Concept **Hand**: 
  \{Hand (D), hand (E), main (F)\}

- Concept **Thumb**: 
  \{Daumen (D), thumb (E), pouce (F)\}

- Relations:
  - **Hand** – hasPart – **Thumb**
  - **Thumb** – partOf – **Hand**

Formal Ontologies

- Type: **Thumb**:

- Description:
  "Every thumb is part of some hand"
  "Every hand has some thumb as part"
Example Aspirin - Headache

Terminologies

- **Concept Aspirin**: {Aspirin (D,E), Acetylsalicylsäure (D), ASS (D), acetylsalicylic acid (E), Acide acétylsalicylique(F)}
- **Concept Headache**: {Kopfschmerz (D), headache (E), céphalée(F)}
- **Relation**: Aspirin – treats – Headache

Formal Ontologies

- **Type**: Aspirin:
- **Description**:
  
  "For every portion of aspirin there is some disposition for treating headache"
Strengths of Formal Ontologies

- Exact, logic-based descriptions of entity types that are instantiated by real-world objects, processes, states
- Representation of stable, context-independent accounts of reality
- Use of formal reasoning methods using tools and approaches from the AI / Semantic Web community
### Deficit of Nomenclatures / Terminologies

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5-46210</td>
<td>Acute appendicitis, NOS</td>
</tr>
<tr>
<td>D5-46100</td>
<td>Appendicitis, NOS</td>
</tr>
<tr>
<td>G-A231</td>
<td>Acute</td>
</tr>
<tr>
<td>M-41000</td>
<td>Acute inflammation, NOS</td>
</tr>
<tr>
<td>G-C006</td>
<td>In</td>
</tr>
<tr>
<td>T-59200</td>
<td>Appendix, NOS</td>
</tr>
<tr>
<td>G-A231</td>
<td>Acute</td>
</tr>
<tr>
<td>M-40000</td>
<td>Inflammation</td>
</tr>
<tr>
<td>G-C006</td>
<td>In</td>
</tr>
<tr>
<td>T-59200</td>
<td>Appendix, NOS</td>
</tr>
</tbody>
</table>
Formal-ontological descriptions: Advantages

- Different description of the same thing can be automatically mapped to a canonic description by a logic-based reasoning device
- Meaning of defined classes can be unambiguously expressed
Only suitable to represent shared, uncontroversial meaning of a domain vocabulary

Supports universal statements about instances of a type:

- All $X$s are $Y$s
- For all $X$s there is some $Y$

Properties of types are properties of all entities that instantiate these types (strict inheritance)
Formal Ontologies: Limitations (II)

- Representation of context dependent knowledge
  - "Allergic Rhinitis is a common disorder (in Europe)"
- Representation of probabilistic knowledge
  - "95% of people infected with viral hepatitis recover"
  - "Smoking is a cardiovascular risk factor"
- Default / canonic knowledge
  - "Adult humans have 32 teeth"
- Dispositions:
  - "Oxazepam is indicated for anxiety disorders"
  - "Aspirin affects the gastric mucosa"

Ontology ⊂ Knowledge Representation
Continuum of knowledge

- Universally accepted assertions
- Consolidated but context-dependent facts
- Hypotheses, beliefs, statistical associations

Domain Knowledge
Formal Ontology !

Domain Knowledge

- Universally accepted assertions
- Consolidated but context-dependent facts
- Hypotheses, beliefs, statistical associations
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
Learning good ontology practice from bad ontologies...
Don’t mix up universals (Concepts, Classes) with individuals (Instances)

- subclass-of (Motor Neuron, Neuron) (FMA, OpenGALEN)
- Is_a (Motor Neuron, Neuron)
- instance-of (Motor Neuron, Neuron) (FlyBase)

But:

- instance-of (my Hand, Hand)
- instance-of (this amount of insulin, Insulin)
- instance-of (Germany, Country)
- not: instance of (Heart, Organ)
- not: instance of (Insulin, Protein)
Don’t use superclasses to express roles

- `Is_a (Fish, Animal)`
- `Is_a (Fish, Food) ??`

- `Is_a (Acetylsalicylic Acid, Salicylate)`
- `Is_a (Acetylsalicylic Acid, Analgetic Drug) ??`

Be aware of the “rigidity” of entity types
Partition the ontology by principled upper level categories

Example: DOLCE’s Upper Ontology

**Endurant (Continuant)**

*Physical*
- Amount of matter
- Physical object
- Feature

*Non-Physical*
- Mental object
- Social object

... 

**Perdurant (Occurrent)**

*Static*
- State
- Process

*Dynamic*
- Achievement
- Accomplishment

**Quality**

*Physical Qualities*
- Spatial location
- ... 

*Temporal Qualities*
- Temporal location
- ... 

*Abstract Qualities*
- ... 

**Abstract**

*Quality region*
- Time region
- Space region
- Color region

Source: S. Borgo ISTC-CNR
Limit to a parsimonious set of semantically precise Basic Relations

First version of the OBO Relation Ontology

Foundational relations

is_a
part_of

Spatial relations (connecting one entity to another in terms of relations between the spatial regions they occupy)

located_in
contained_in
adjacent_to

Temporal relations (connecting entities existing at different times)

transformation_of
derives_from
preceded_by

Participation relations (connecting processes to their bearers)

has_participant
has_agent
Avoid idiosyncratic categorization

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body structure</td>
<td>Acquired body structure, Anatomical organizational pattern</td>
</tr>
<tr>
<td>Clinical finding</td>
<td>Administrative statuses, Adverse incident outcome categories</td>
</tr>
<tr>
<td>Environment or geographical</td>
<td>Environment, Geogr. and/or political region of the world</td>
</tr>
<tr>
<td>Event</td>
<td>Abuse, Accidental event, Bioterrorism related event</td>
</tr>
<tr>
<td>Linkage concept</td>
<td>Attribute, Link assertion</td>
</tr>
<tr>
<td>Observable entity</td>
<td>Age AND/OR growth period, Body product observable</td>
</tr>
<tr>
<td>Clin. history / examination</td>
<td>Device observable, Drug therapy observable, Feature of Entity</td>
</tr>
<tr>
<td>Organism</td>
<td>Animal, Chromista, Infectious agent</td>
</tr>
<tr>
<td>Pharmaceutical / biologic product</td>
<td>Alcohol products, Alopecia preparation, Alternative medicines</td>
</tr>
<tr>
<td>Physical force</td>
<td>Altitude, Electricity</td>
</tr>
<tr>
<td>Physical object</td>
<td>Device, Domestic, office and garden artefact</td>
</tr>
<tr>
<td>Procedure</td>
<td>Administrative procedure, Community health procedure</td>
</tr>
<tr>
<td>Qualifier value</td>
<td>Action, Additional dosage instructions</td>
</tr>
<tr>
<td>Record artifact</td>
<td>Record organizer, Record type</td>
</tr>
<tr>
<td>Situation with explicit context</td>
<td>A/N risk factors, Critical incident factors</td>
</tr>
<tr>
<td>Social context</td>
<td>Community, Family, Group</td>
</tr>
<tr>
<td>Special concept</td>
<td>Namespace concept, Navigational concept, Non-current concept</td>
</tr>
<tr>
<td>Specimen</td>
<td>Biopsy sample, Body substance sample, Cardiovascular sample</td>
</tr>
<tr>
<td>Staging and scales</td>
<td>Assessment scales, Endometriosis classification of American Fertility Society</td>
</tr>
<tr>
<td>Substance</td>
<td>Allergen class, Biological substance</td>
</tr>
</tbody>
</table>
Jorge Luis Borges

"On those remote pages it is written that animals are divided into:

a. those that belong to the Emperor  
b. embalmed ones  
c. those that are trained  
d. suckling pigs  
e. mermaids  
f. fabulous ones  
g. stray dogs  
h. those that are included in this classification  
i. those that tremble as if they were mad  
j. innumerable ones  
k. those drawn with a very fine camel's hair brush  
l. others  
m. those that have just broken a flower vase  
n. those that resemble flies from a distance"
Be aware of ambiguities

- “Institution” may refer to
  1. (abstract) institutional rules
  2. (concrete) things instituted
  3. act of instituting sth.

- “Tumor”
  1. evolution of a tumor as a disease process
  2. having a tumor as a pathological state
  3. tumor as a physical object

- “Gene”
  1. a (physical) sequence of nucleotides on a DNA chain
  2. a collection of (1)
  3. A piece of information conveyed by (1)
Don’t mix up ontology with epistemology

- *Is_a* (*Infection of unknown origin, Infection*)
- *Is_a* (*Newly diagnosed diabetes, Diabetes*)
- *Is_a* (*Family history of diabetes, Diabetes*)
Don’t mix up Ontology IDs with Terms

- Glycerin Kinase
- Glycerokinase
- GK
- Glycerinkinase
„how it is expressed in human language“

Terminologies

Formal Ontologies

Information Models

„what sth. knows about “
Don’t underestimate Ontology Maintenance

- Formal Ontologies must always be maintained
  - consistent (free of logic contradiction): prerequisite for machine reasoning
  - adequate (correctly describe the domain) prerequisite to prevent erroneous deductions
- Maintenance load is much higher than with terminologies.
- Ontology maintenance is mainly task of domain experts. IT staff has supportive function
- Typical design and maintenance errors
Structure of this talk

- Introduction - Current Systems
- Terminological Clarification
- What do Formal Ontologies Represent?
- Terminologies vs. Formal Ontologies
- Practice of Good Ontology
- Outlook
Outlook

- Ontology often used a buzzword for nontologies but…
- Formal ontological principles increasingly govern the construction of Life Science Knowledge Organization Systems
- Users / domain expert must be heavily involved into ontology engineering and maintenance
- Insufficient evidence:
  - Which use cases require formal ontologies
  - In which cases informal terminology systems are sufficient?
  - Which cases require both?
  - Can existing terminologies be ontologized?
  - Can terminologies and ontologies co-exist?
- The outcome of the existing legacy systems’ move toward principled ontologies is still open
SNOMED CT

- One huge system
- Impressive domain coverage
- Considerable investments, important stakeholders
- Increasing number of (clinical) users
- Other use cases mainly unexplored (basic research, clinical trials)
- Increasing mappings to existing terminologies
- Legacy: hybrid of terminology, ontology, information model
- Overall structure idiosyncratic, disorganized
- Some major architectural weaknesses
- Unreflected use of logic, unintended entailments
- Major redesign necessary: formal foundations, editing guidelines, quality control procedures
- Risk: uncontrollable proliferation, loss of expressiveness,
- Chances: Positive input by user groups
Open Biomedical Ontologies (OBO)

- Many focused ontologies
- Increasing number of annotated sources
- Convergence to standardized syntax and semantics
- Increasingly using formal ontology principles
- Public
Thank you!

Contact:
stschulz@uni-freiburg.de

Stefan Schulz
Medical Informatics
Research Group
University
Medical Center
Freiburg, Germany