**Stefan Schulz**
**Medical University of Graz (Austria)**

**purl.org/steschu**
**steschu@gmail.com**

# Biomedical Ontologies

*Fourth Interdisciplinary School on Applied Ontology (ISAO 2018)*
*10-15 September 2018, Cape Town, South Africa*

# Goals of the lectures

- Data management in biomedical research and health care
- Overview of the entities of interest this area
- Practice “ontological thinking”
- Catch up with previous knowledge on ontology and logic
- Discuss specific ontological challenges in this domain
- Distinguish ontologies from other semantic artefacts

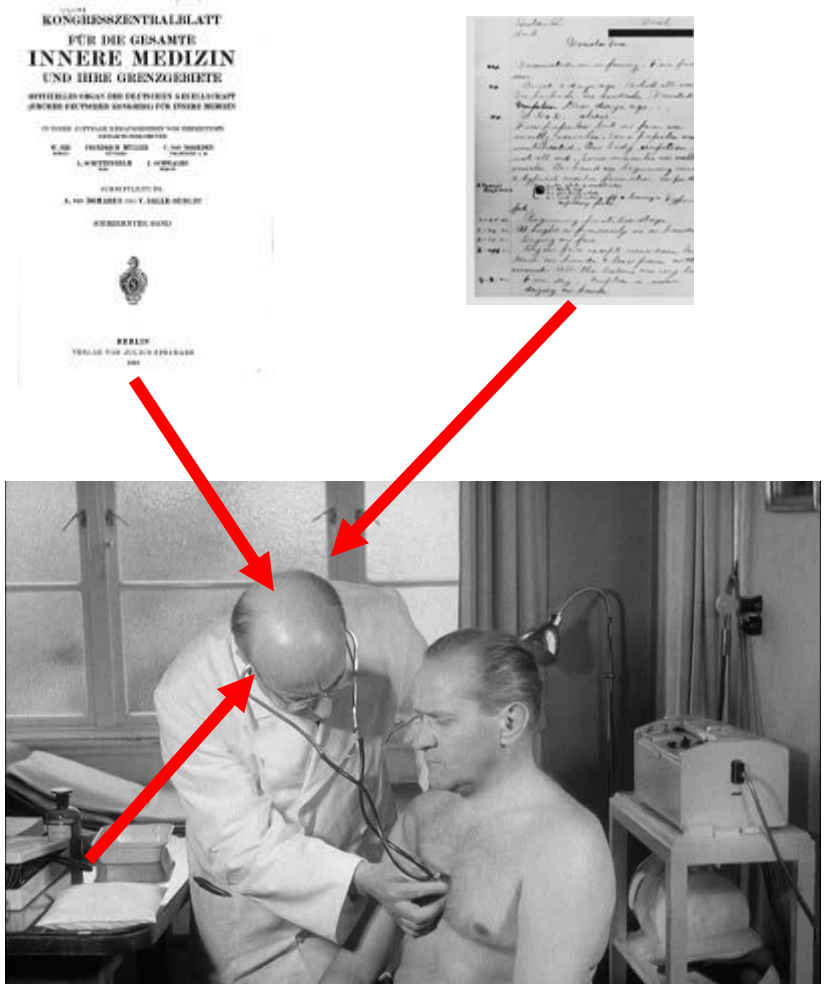
# The scope: biomedical research & health care

- Health
  - Crucial resource for well being
  - More than absence of disease
- Health care / medicine:
  - one of the world's largest and fastest-growing industries
  - > 10 percent GDP of most developed countries
- Beyond care:
  - health involves all aspects of life, e.g. diet, exercise, occupational safety
- Beyond humans: veterinary medicine

# The scope: biomedical research & health care

- Biology
  - Science that studies life and living organisms
  - Genes → Molecules → Cells → Organisms → Populations →Ecosystems
- Biomedical Science:
  - Application of biology and other natural science for diagnosis, prevention and treatment of diseases
  - Important application: pharmaceutical industry
    - total pharmaceutical revenues worldwide > 1 Trillion $
    - Cost of bringing new drug to market: > 1 Billion $

# Data in health care



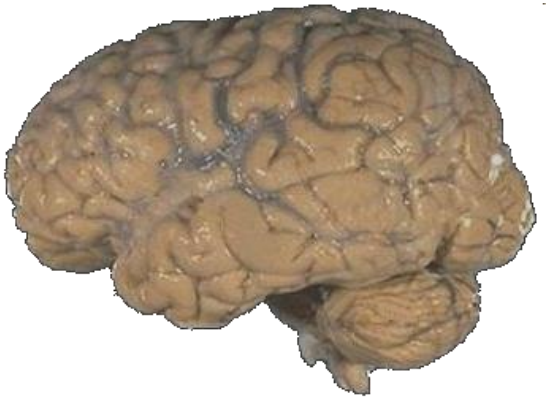
**What has changed since then?**

# Technology (r)evolution

**5 MB IBM Hard Drive, 1956** *** 100,000** **512 GB Memory Stick 2018**

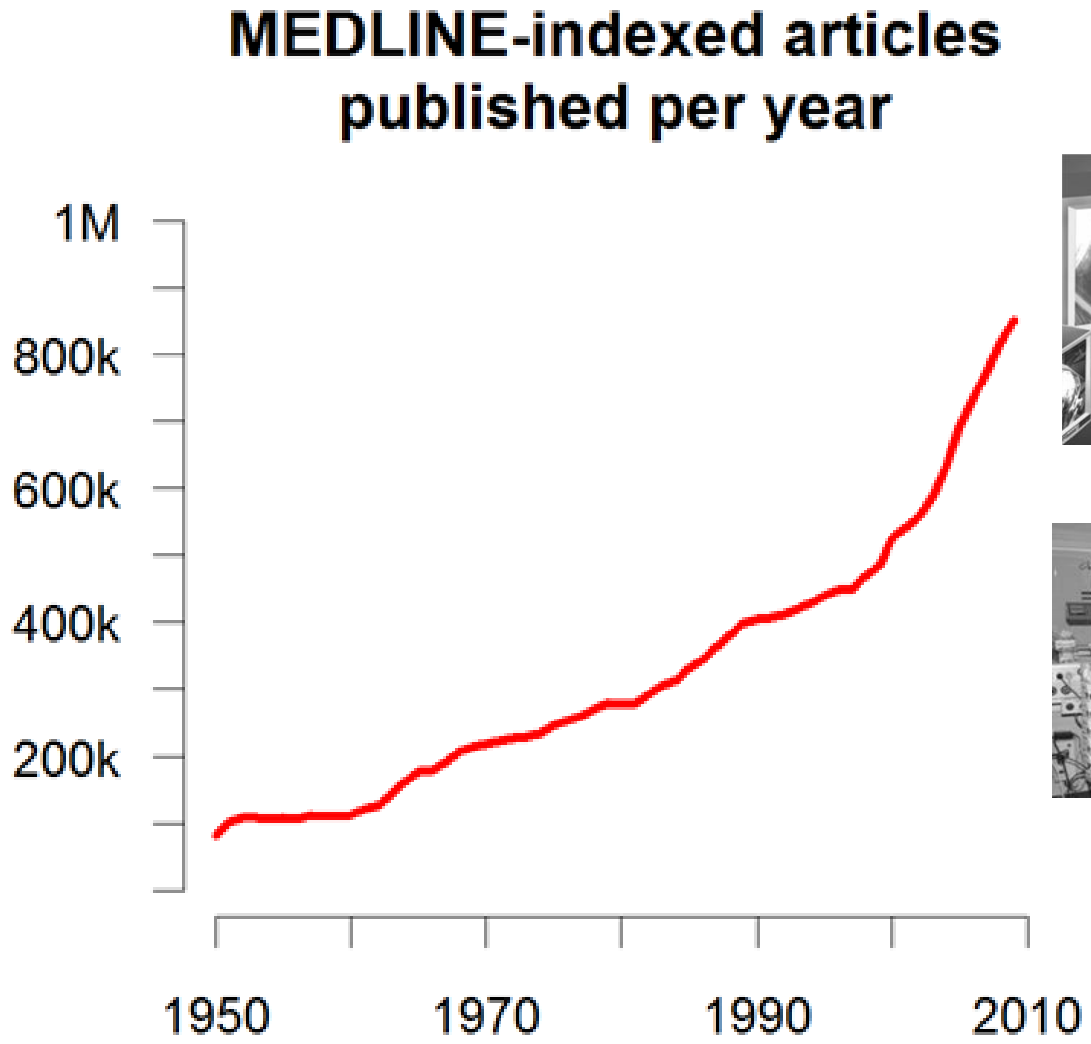
# Human evolution

**Human brain 1956**



**Human brain 2018**

# Knowledge explosion



# Data in health care

- Electronic health records
  - Substitute of traditional paper chart
  - Serve different purposes
  - Documenting the patient‘s history and progress
    - Legal requirements
    - Communication between physicians, nurses
  - Coding for billing / reimbursement
  - Special documentation
    - Clinical trials
    - Patient registries
    - Quality control

# Most clinical data is free text

St. p. TE eines exulc. sek.knot.SSM li US dors. 5/11 Level IV 2,4 mm Tumordurchm. Sentinnel LK ing. li. tumorfr.

01 CHEST, PORTABLE 1 VIEW
02 INDICATION:
03 Shortness of breath
04 COMPARISON: July 16 10 recent prior
05 FINDINGS:
06 Left central line, tip at mid-SVC.
07 Cardiac and mediastinal contours as before
08 No pneumothorax.
09 Lungs: Interval increase in right lung base
10 pulmonary opacity with air bronchograms,
11 increasing pneumonitis / atelectasis.

Paciente cardiopatia isquemico, com CRM prévia, interna para realizar ACTP + stent em ACD, via ponte de safena. Procedimento realizado com sucesso e sem intercorrências.

Planning Nieuwe afspraak binnen 6 maanden met vroegere voorafgaande adipositascontroles . De patïente moet ook PTH , folaten en cobalamine laten controleren bij labo - onderzoeken , ze doet die zelf aangezien ze verpleegster is in de provindie provincie Skåne . Moet de inname van calciumtabletten naar 3 per dag verhogen ( momenteel slechts een per dag ). Binnen 3 maanden nieuwe controle van 25- OH - vitamine D3- controle , inclusief PTH en vloeistofhuishouding . Code diagnose / behandeling Hoofddiagnose : Z090, halfjaarlijkse controle na gastric bypass wegens obesitas

**Which are the advantages / disadvantages of free text in clinical documentation?**

# Structured clinical data

## Lab results

| Lab | Result | Normal value |
|---|---|---|
| AM ACTH | <1.1 pg/mL | 7.2-63.3 pg/mL |
| AM cortisol | <0.8 μg/dL | 4-20 μg/dL |
| 30-min post-cosyntropin | 9.5 μg/dL | |
| 1-hour post-cosyntropin | 11.2 μg/dL | |
| LH | 0.3 mU/mL | 1-8 mU/mL |
| Testosterone | <20 ng/dL | 300-1,200 ng/dL |
| Prolactin | 1 ng/mL | 0-17 ng/mL |
| SHBG | 40.2 nmol/L | 19.3-76.4 nmol/L |
| Free $T_4$ | 0.61 ng/dL | 0.8-1.8 ng/dL |
| Total $T_3$ | 59 ng/dL | 75-165 ng/dL |
| TSH | 0.196 μU/mL | 0.4-5 μU/mL |
| IGF-1 | 69 ng/mL | 56-201 ng/mL |
| ANCA | <1:16 | |

Abbreviations: ACTH = adrenocorticotropic hormone; ANCA = anti-neutrophil cytoplasmic antibody; IGF-1 = insulin-like growth factor 1; LH = luteinizing hormone; SHBG = sex hormone–binding globulin; $T_3$ = triiodothyronine; $T_4$ = thyroxine; TSH = thyroid-stimulating hormone.

## Drug prescriptions

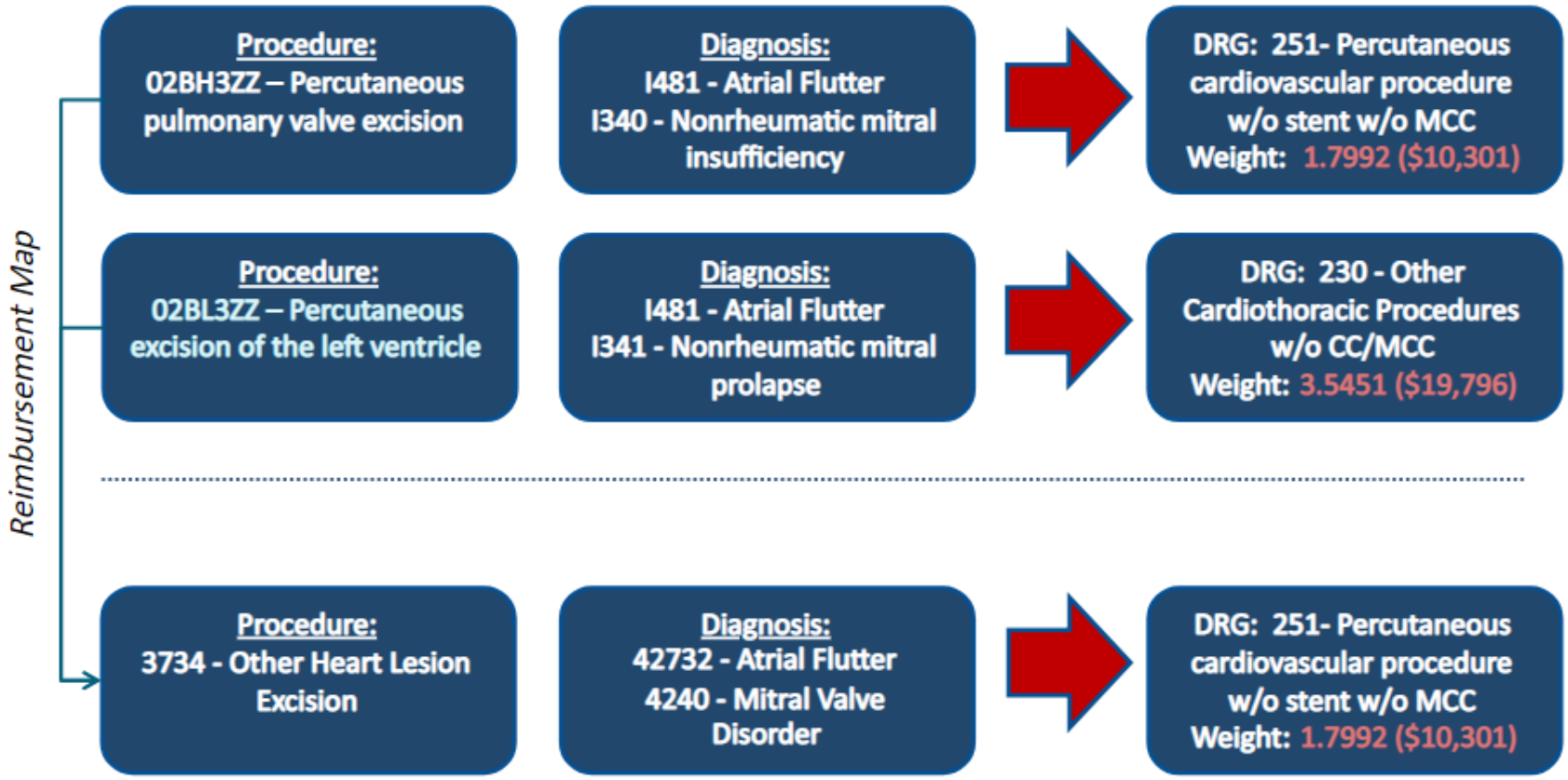
| Drug | Strength | SIG (?) |
|---|---|---|
| Aggrenox | 200-25 mg CPMP 12 | Take 1 twice daily |
| AMERGE 2.5 MG TABLET | | As instructed for Migraine Headacl |
| Aricept | 10 mg TABLET | Take 1 daily |
| Aspirin | 325 mg TABLET | Take 1 daily |
| Baclofen | | |
| BACLOFEN 10 MG TABLET | | one PO bid |
| Carbamazepine | 200 mg TABLET | Take 1 twice daily |
| Darvocet-N 100 100-650 mg TABLET | | Take 1 every 6-8 hours as needec |
| Dilantin Extended | 100 mg CAPSULE | 3-PO dedtime |
| Divalproex | 250 mg TABLET DR | Take 1 bedtime daily for 1-wk ther |
| Gabapentin | 400 mg CAPSULE | 1 capsule(s) by mouth twice a day |
| Galantamine | 4 mg TABLET | Take 1 twice daily |
| Hydromorphone | 2 mg TABLET | Take 1 twice daily as needed |
| Ibuprofen | 600 mg TABLET | Take 1 two to three times daily with |

http://www.lrec-conf.org/proceedings/lrec2016/pdf/1222_Paper.pdf

http://www.neurologyemrsoftware.com

# Abstracted, coded data (I)

## Data for billing / reimbursement

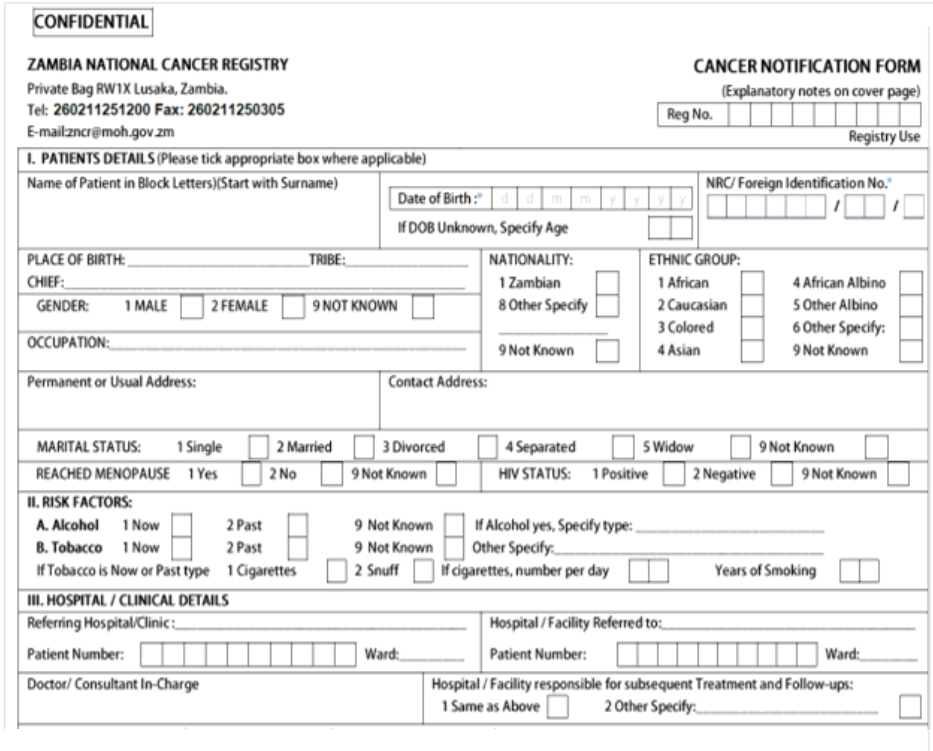



U.S. Centers for Medicare & Medicaid services https://slideplayer.com/slide/2686045/

**Motivation to produce these data? Sources of bias ?**

# Abstracted, coded data (II)

## Data for epidemiology: example cancer registry



CONFIDENTIAL

**ZAMBIA NATIONAL CANCER REGISTRY**
Private Bag RW1X Lusaka, Zambia.
Tel: 260211251200 Fax: 260211250305
E-mail:zncr@moh.gov.zm

**CANCER NOTIFICATION FORM**
(Explanatory notes on cover page)
Reg No. _____ Registry Use

**I. PATIENTS DETAILS** (Please tick appropriate box where applicable)

Name of Patient in Block Letters)(Start with Surname) | Date of Birth :* d d m m y y y y | If DOB Unknown, Specify Age | NRC/ Foreign Identification No.* / /

PLACE OF BIRTH: _____ TRIBE: _____
CHIEF: _____
GENDER: 1 MALE 2 FEMALE 9 NOT KNOWN
OCCUPATION: _____

NATIONALITY: 1 Zambian; 8 Other Specify; 9 Not Known

ETHNIC GROUP: 1 African; 2 Caucasian; 3 Colored; 4 Asian; 4 African Albino; 5 Other Albino; 6 Other Specify:; 9 Not Known

Permanent or Usual Address: | Contact Address:

MARITAL STATUS: 1 Single 2 Married 3 Divorced 4 Separated 5 Widow 9 Not Known
REACHED MENOPAUSE 1 Yes 2 No 9 Not Known | HIV STATUS: 1 Positive 2 Negative 9 Not Known

**II. RISK FACTORS:**
A. Alcohol 1 Now 2 Past 9 Not Known — If Alcohol yes, Specify type: _____
B. Tobacco 1 Now 2 Past 9 Not Known — Other Specify: _____
If Tobacco is Now or Past type 1 Cigarettes 2 Snuff — If cigarettes, number per day _____ Years of Smoking _____

**III. HOSPITAL / CLINICAL DETAILS**

Referring Hospital/Clinic: _____ Patient Number: _____ Ward: _____
Hospital / Facility Referred to: _____ Patient Number: _____ Ward: _____

Doctor/ Consultant In-Charge
Hospital / Facility responsible for subsequent Treatment and Follow-ups: 1 Same as Above 2 Other Specify: _____

Doctor/ Consultant In-Charge
Hospital / Facility responsible for subsequent Treatment and Follow-ups: 1 Same as Above 2 Other Specify: _____

**IV. TUMOUR** (Please specify primary organ or site of Cancer and exact location if possible)

Date of Diagnosis * d d m m y y y y
Primary Site: _____
Histological / Clinical Diagnosis: _____

Basis of Diagnosis (Check one or more as applicable)
1 Death Certificate Only
2 Clinical Only
3 Clinical Investigation
4 Specific Tumor Marker
5 Cytology (Lab No._____)
6 Metastasis (Lab No._____)
7 Primary Tumor (Lab No._____)
9 Not Known

Extent of Disease:
0 In Situ
1 Localised
2 Local Extension Only
3 Local Extension + Regional Nodes
4 Regional Nodes
5 Distant Metastasis
8 Not Applicable
9 Not Known

PRESENT STATUS:
1 ALIVE
2 DEAD
Date of Last Contact: d d m m y y y y
Date of Death: d d m m y y y y

If Dead Cause of Death
1 Cancer
2 Others Specify :
9 Not Known

**V. CLINICAL STAGING & TREATMENT**

Clinical Staging (cTNM)
T _____
N _____
M _____

Staging Group
0 In Situ; 1 1A; 2 1B; 3 2A; 4 2B; 5 3A; 6 3B; 7 4A; 8 4B; 9 Not Known

Treatment (Check one or more as applicable)
1 Other Treatment
2 Surgery
3 Radiotherapy
4 Chemotherapy
5 Hormones
6 Anti-Retrovirals
7 Palliative Care
9 Not Known

**VI. SOURCE OF INFORMATION**

| | | | |
|---|---|---|---|
| 1st Notification | Province | Facility Name | Date Seen: d d m m y y y y |
| 2nd Notification | Province | Facility Name | Date Seen: d d m m y y y y |
| 3rd Notification | Province | Facility Name | Date Seen: d d m m y y y y |

Other Remarks: _____ Notification Date: * d d m m y y y y

* Mandatory fields

**What's the interest of physicians to fill such forms on paper or on screen?**

http://afcrn.org/images/M_images/attachments/135/Zambia%20data%20collection%20form%20copy.jpg

# Quality problems with clinical data

- Textual data relatively accurate and complete, tailured to human readers but difficult do analyze: NLP (natural language processing systems) have to deal with multiple sublanguages and poor editing
- Structured data often not linked to international semantic standards (controlled vocabularies, ontologies)
- Limited motivation to generate good quality data:
  - Wherever users are not beneficiaries of data
  - Wherever users have to record data redundantly
- Known biases:
  - Collecting data for billing / reimbursement
  - Collecting data for quality management

# The holy grail of medical informatics…

**St. p. TE eines exulc. sek.knot.SSM li US dors. 5/11 Level IV 2,41 mm Tumordurchm. Sentinnel LK ing. li. tumorfr.**

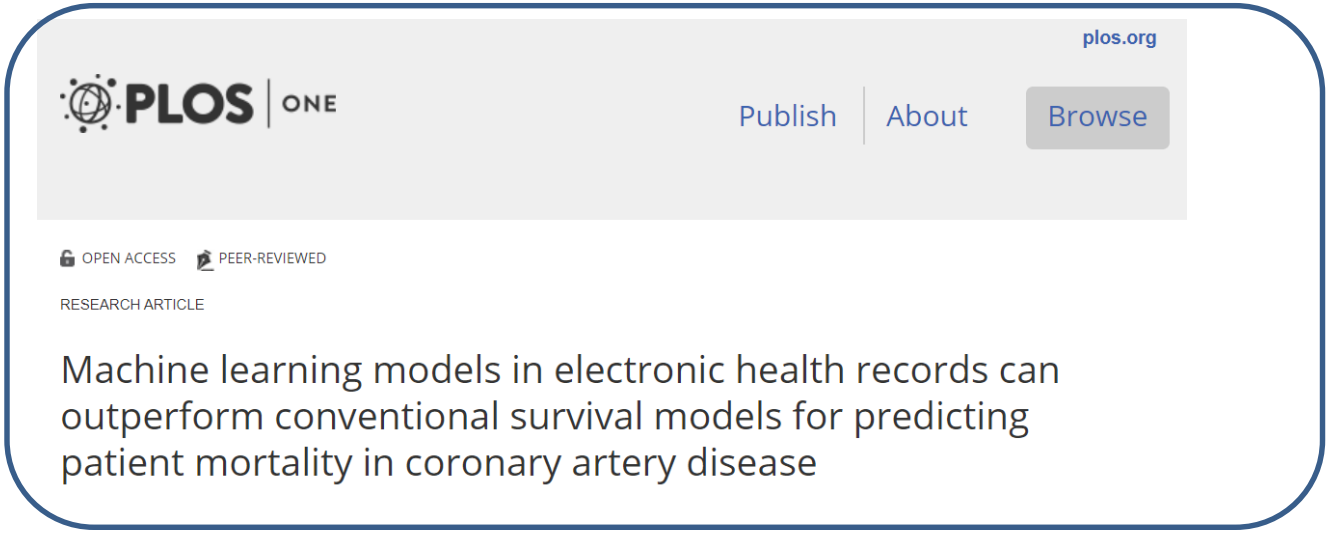
| Code (SNOMED CT, LOINC) | Value | Context |
|---|---|---|
| **254730000** \|Superficial spreading malignant melanoma of skin | | 392521001 \|History of |
| **301889008** \|Excision of malignant skin tumor | | 392521001 \|History of |
| **47224004** \|Skin of posterior surface of lower leg<br>**7771000** \|Left | | |
| **81827009** \|Diameter<br>**258673006** \|millimeter | 2.41 | |
| **258403002** \|Lymph node level IV | | |
| **94339008** \|Secondary malignant neoplasm of inguinal lymph nodes | | 15240007 \|Current<br>2667000 \|Absent |

# Primary and secondary use scenarios

- Primary use: documenting, communicating, collecting specific data for defined data analysis use cases
- Secondary use: Repurposing of clinical routine data, e.g. for
  - Building cohorts for clinical trials
  - Retrospective data analysis
  - Medical education
  - Prediction of future events

**Where do you think ontologies come into play ?**

# Primary and secondary use scenarios



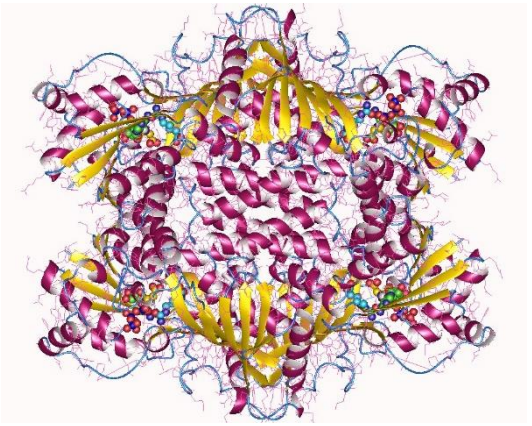
- What does this have in common?
- Is there a need for ontologies?

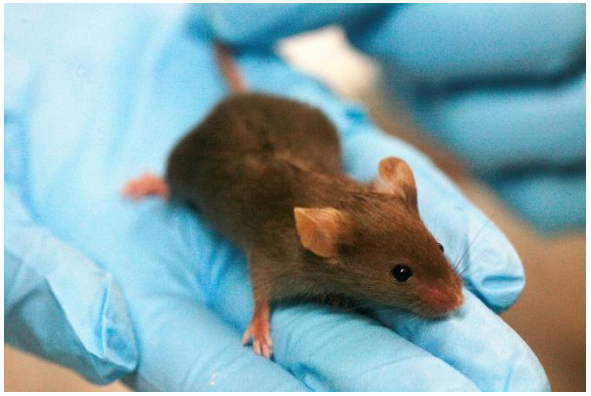
# Privacy of clinical data

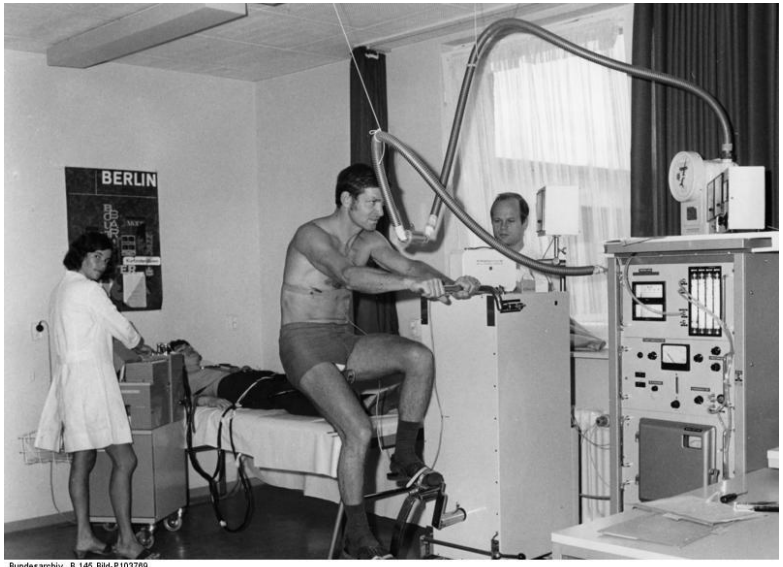
- Hippocratic oath:
  *"And whatsoever I shall see or hear in the course of my profession (…) I will never divulge, holding such things to be holy secrets"*
- Declaration of Helsinki
  *"It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects"*
- Health profession council of South Africa:
  *"Health care practitioners hold information about patients that is private and sensitive. The National Health Act (Act No. 61 of 2003) provides that this information must not be given to others, unless the patient consents or the health care practitioner can justify the disclosure. Practitioners are responsible for ensuring that clerks, receptionists and other staff respect confidentiality in their performance of their duties. "*

https://history.nih.gov/research/downloads/hippocratic.pdf
http://www.who.int/bulletin/archives/79%284%29373.pdf
http://www.hpcsa.co.za/downloads/conduct_ethics/rules/confidentiality_protecting_providing_info.pdf

# Data in biomedical sciences








Bundesarchiv, B 145 Bild-P103769
Foto: o.Ang. | 1950/1955 ca.

**What do you think is different compared to clinical data?**

# Data in biomedical sciences

- Experiments require precise documentation
  - Clinical trials use own data acquisition standards and tools
  - Lab experiments increasingly publish not only papers but also datasets
  - Primary source of scientific data peer-reviewed publications

# *Survivin*-targeting miR-542-3p overcomes HER3 signaling-induced chemoresistance and enhances the antitumor activity of paclitaxel against HER2-overexpressing breast cancer

Hui Lyu [a, b, 1], Shuiliang Wang [c, 1], Jingcao Huang [d], Bolun Wang [a, e], Zhimin He [b], Bolin Liu [a, b, *]

[a] *Department of Pathology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA*
[b] *Cancer Research Institute and Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China*
[c] *Fujian Key Laboratory of Transplant Biology, Fuzhou General Hospital, Fuzhou, Fujian, China*
[d] *Department of Hematology, Hematologic Research Laboratory, West China Hospital, Sichuan University, Chengdu, Sichuan, China*
[e] *The Second Xiangya Hospital, Central South University, Changsha, Hunan, China*

ARTICLE INFO

*Article history:*
Received 4 September 2017
Received in revised form
19 January 2018
Accepted 22 January 2018

*Keywords:*
HER3
Survivin
miRNA
Paclitaxel resistance
Breast cancer

ABSTRACT

Elevated expression of HER3, which interacts with HER2 in breast cancer cells, confers chemoresistance via phosphoinositide 3-kinase (PI-3K)/Akt-dependent upregulation of Survivin. However, the underlying mechanism is not clear. Ectopic expression or specific knockdown of HER3 in HER2-overexpressing breast cancer cells did not alter *Survivin* mRNA levels and Survivin protein stability, supporting the notion that HER3 signaling may regulate specific miRNAs that target *Survivin* to alter its protein translation. Here we showed that overexpression and specific knockdown of HER3 reduced and enhanced expression of two *Survivin*-targeting miRNAs, miR-203 and miR-542-3p, in breast cancer cells, respectively. While the specific inhibitor of either miR-203 or miR-542-3p attenuated an anti-HER3 antibody-induced downregulation of Survivin, inhibition of miR-542-3p exhibited a better efficacy than miR-203 inhibition did. Consistently, miR-542-3p mimic was much more effective than miR-203 mimic not only in inhibition of Survivin, but also in enhancement of paclitaxel-induced apoptosis in HER2-overexpressing breast cancer cells. Moreover, the combination of miR-542-3p mimic and paclitaxel, as compared with either agent alone, significantly inhibited *in vivo* tumor growth of HER2-overexpressing breast cancer cells. Collectively, our data indicated that the HER3/PI-3K/Akt signaling upregulates Survivin via suppression of miR-203 and miR-542-3p. Because miR-542-3p has three binding sites on the 3′-UTR of *Survivin* mRNA, its mimic was able to effectively downregulate Survivin *in vitro* and *in vivo*. Thus, miR-542-3p-replacement therapy is an excellent approach to overcome HER3-mediated paclitaxel resistance and significantly enhances the antitumor activity of paclitaxel against HER2-overexpressing breast cancer.

© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

# Data in biomedical sciences

- Experiments require precise documentation
  - Clinical trials use own data acquisition standards and tools
  - Lab experiments increasingly publish not only papers but also datasets
  - Primary source of scientific data peer-reviewed publications
  - On-line available
    - > 25 million abstracts via Pubmed / MEDLINE
    - Millions of full texts

# Biomedical databases

- Typical questions
  - Which genes / proteins in which organism are related to which biological processes
  - Which structure and functions do they have?
  - In which biochemical pathways are they related to which molecules?
  - Which genetic defects are related to which diseases?
  - Structured extracts of publications go into research databases, e.g. Uniprot, Ensembl, Reactome
    - By the authors
    - By database curators
    - By NLP-based algorithms

# Uniprot: example record

## P04626 (ERBB2_HUMAN)

Basket

BLAST | Align | Format | Add to basket | History

Feedback | Help video | Other tutorials and videos

Protein **Receptor tyrosine-protein kinase erbB-2**

Gene **ERBB2**

Organism *Homo sapiens (Human)*

Status Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level

## Function

Protein tyrosine kinase that is part of several cell surface receptor complexes, but that apparently needs a coreceptor for ligand binding. Essential component of a neuregulin-receptor complex, although neuregulins do not interact with it alone. GP30 is a potential ligand for this receptor. Regulates outgrowth and stabilization of peripheral microtubules (MTs). Upon ERBB2 activation, the MEMO1-RHOA-DIAPH1 signaling pathway elicits the phosphorylation and thus the inhibition of GSK3B at cell membrane. This prevents the phosphorylation of APC and CLASP2, allowing its association with the cell membrane. In turn, membrane-bound APC allows the localization of MACF1 to the cell membrane, which is required for microtubule capture and stabilization. Curated

In the nucleus is involved in transcriptional regulation. Associates with the 5'-TCAAATTC-3' sequence in the PTGS2/COX-2 promoter and activates its transcription. Implicated in transcriptional activation of CDKN1A; the function involves STAT3 and SRC. Involved in the transcription of rRNA genes by RNA Pol I and enhances protein synthesis and cell growth. 3 Publications

### Catalytic activity

ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate. PROSITE-ProRule annotation | 1 Publication

# Uniprot: annotations with Gene Ontology

**GO - Molecular function**

- ATP binding — Source: UniProtKB-KW
- ErbB-3 class receptor binding — Source: ProtInc
- identical protein binding — Source: IntAct
- phosphatidylinositol-4,5-bisphosphate 3-kinase activity — Source: Reactome
- protein C-terminus binding — Source: UniProtKB
- protein dimerization activity — Source: UniProtKB
- protein heterodimerization activity — Source: UniProtKB
- protein phosphatase binding — Source: UniProtKB
- protein tyrosine kinase activity — Source: BHF-UCL
- Ras guanyl-nucleotide exchange factor activity — Source: Reactome
- RNA polymerase I core binding — Source: UniProtKB
- transmembrane receptor protein tyrosine kinase activity — Source: BHF-UCL
- transmembrane signaling receptor activity — Source: BHF-UCL

View the complete GO annotation on QuickGO ...

**GO - Biological process**

- cell proliferation — Source: ProtInc
- cell surface receptor signaling pathway — Source: MGI
- cellular response to epidermal growth factor stimulus — Source: UniProtKB
- cellular response to growth factor stimulus — Source: UniProtKB
- enzyme linked receptor protein signaling pathway — Source: ProtInc
- ERBB2 signaling pathway — Source: Reactome
- heart development — Source: Ensembl
- MAPK cascade — Source: Reactome
- motor neuron axon guidance — Source: Ensembl
- myelination — Source: Ensembl

Explore biological databases and identify where ontologies are used

Uniprot (proteins): https://www.uniprot.org

Reactome (pathways) https://reactome.org

# Exercise (I)

- Use the following upper-level categories: Material entity, immaterial physical entity, quality, role, realizable (disposition, function), process, information entity, temporal region
- Try to relate biomedical terms to these categories
- Decide whether they denote subclasses or instances (individuals)
- Discuss additional aspects like granularity and cardinality
- Are there conflicting categorizations?

# Exercise (II)

- Sample terms:
  “cranial cavity”, “aspirin”, “road traffic accident”, “liver function”, “headache”, “social security number”, “mouse embryo”, “blood”, “carbon atom“ , “red”, “persecutory delusion”, “Groote Schuur Hospital”, “nurse”, “American College of Rheumatology recommendations for the treatment of early rheumatoid arthritis”, “death”, “acute”, “tooth extraction”, “species homo sapiens”, “39.9°C”, “Ibuprofen 300 mg Capsule”, “admission diagnosis”, “tonsillectomy”, “World Health Organisation”, “malaria”, “gunshot injury”, “DNA”, “phenotype”, “Gene”, “colon cancer”, “life”, “insulin”, “hospital”, “white blood cell”, “body mass (in kg)”, “risk of breast cancer”, “patient”

# Biomedical entities walkthrough

- Ontological analysis:
  - Inventory of middle level classes ?
  - Categorization: upper level classes ?
  - Properties: what do they have in common?
  - Relations: how can they be related?

# Material entities and immaterial spaces (I)

- By increasing cardinality:
  - Atoms, ions, small molecules, e.g. Calcium, Glucose
  - Macromolecules, e.g. proteins, nucleic acids (RNA, DNA)
  - Parts of macromolecules, e.g. gene sequences, protein sequences
  - Molecule complexes, e.g. chromatin, chromosomes
  - Cells, cell components and intracellular spaces, e.g. white blood cell, mitochondrion, cell nucleus, cell membrane, intracellular space.
  - Anatomical entities: tissues, organ parts, organs, organ systems
  - Organisms, unicellular (e.g. bacteria), multicellular
  - Populations, cohorts

# Material entities and immaterial spaces (II)

- Non-biological material entities of biomedical interest:
  - Synthesised molecules (drugs)
  - Lab devices
  - Medical devices, implants
  - Medical equipment, vehicles, buildings etc.
- Non-material physical entities
  - Geographical region
  - Habitat

# Material entities and immaterial spaces (III)

- Other aspects
  - Homomericity: part is of the same type: amount of water, amount of brain tissue etc.
  - Single objects vs. collections of same object, e.g. aspirin molecule, vs. amounts of aspirin molecules - but distinct from aspirin tablet!
  - Monomers vs. polymers: example carbohydrates, nucleic acids, proteins

**Which relations are typical for this kind of entities ?**

# Processual entities (I)

- At level below organisms
  - At molecular level: modification, transport, signal transmission, regulation of activities, e.g. gene regulation, control of transcription
  - At cellular level: mitosis, meiosis, cell death, propagation of impulses through nerves,…
  - At tissue level: immune processes
  - At level of organs and organ systems: motion, circulation, neuromuscular processes, digestion, respiration, wound healing , …

# Processual entities (II)

- With human agents on biological objects:
  - laboratory processes, omics analyses
  - Therapeutic interventions, diagnostic interventions, observing, interpreting, documenting, diagnosing, prescribing drugs, therapies
  - Health system processes: admission, discharge, billing, reimbursement, training, certification ,…
  - Lifestyle, physical exercise

**Which relations are typical for processual entities ?**

# Realisables

- Realisables exist even if not realised.
  - Ability to interact on a molecular level
  - Ability to perform cell division
  - Ability to kill pathogens
  - Ability to explode
  - Disposition of a bone to break
  - Reproductive function
  - Function of pumping blood
  - Walking function
  - Risk of breast cancer
  - Ability to lactate

**Which relations are typical for realisables? How are they related to material objects, how to processes?**

# Roles

- The role of a solvent
- The role of a substrate in a chemical reaction
- The role of a patient / of a health professional
- Employer / employee
- Parent, child, sibling,…
- (Social) gender, ethnicity
- The role of a predator / prey
- Catalyst, enzyme
- Roles in processes: active participant / passive participant / input / output
- Food as a role of a certain amount of biological matter

**Which relations are typical for roles?**

# Qualities

- Physical qualities: weight, mass, electric charge, temperature
- Qualities of processes, e.g. evolution of a disease process
- Species quality, e.g. being a human, a fish, a mushroom
- Canonicity, i.e. normal / abnormal, pathologic
- Shapes

**Which relations are typical for qualities entities ? How are they distinguished from realisables?**

# Information content entities

- Epistemology vs. ontology
- Image, e.g. X-ray
- Plans
- Thoughts, beliefs, opinions, cultural / individuals
- Results of speech acts
- Documents, i.e. results of documentation acts
- Results of observations, measurements
- Medical diagnosis, prognosis

**Which relations are typical for information content entities ?**

# Social entities

- Associations, corporations, institutions, families
- E.g. hospital, school, lab, insurance company,

**Which relations are typical for social entities ?**

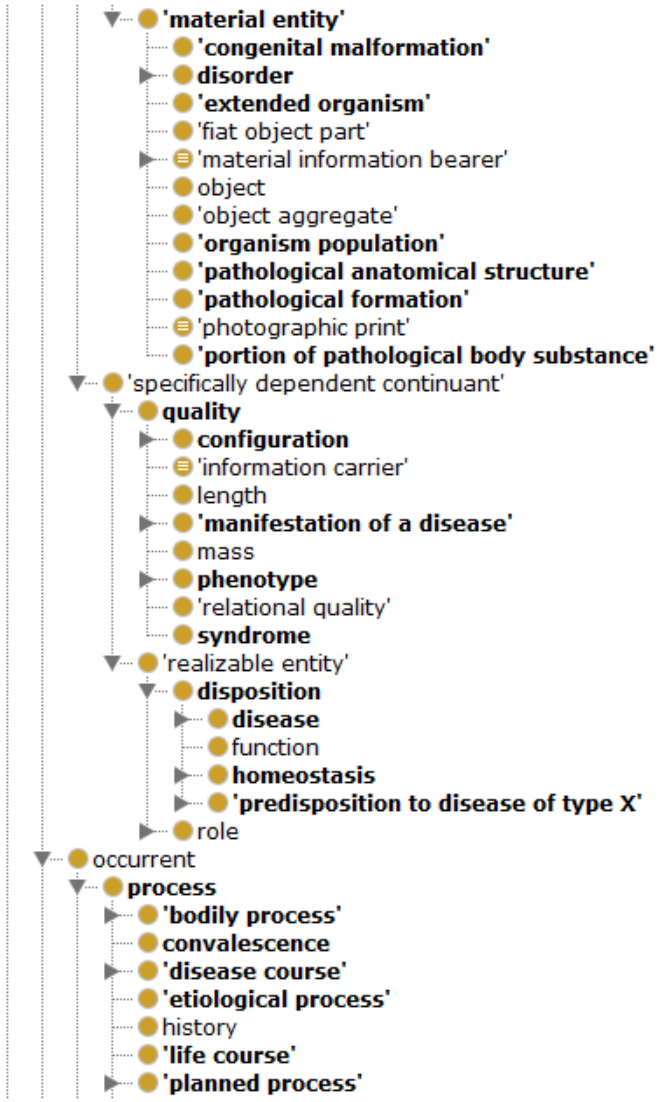
# Entity types with multiple or debatable assignment to upper-level classes (I)

- Diseases, disorders: What do a pneumonia, a club foot, a femur fracture, a seizure, an ulcer, a colon cancer have in common?
- Related entities,
  - E.g. genetic disposition -> manifestation
  - e.g. cause / mechanism of an injury -> morphology -> process
- Experiences, e.g. symptoms (individual perception of body dysfunction)?
- Delusions?

# Entity types with multiple or debatable assignment to upper-level classes (II)

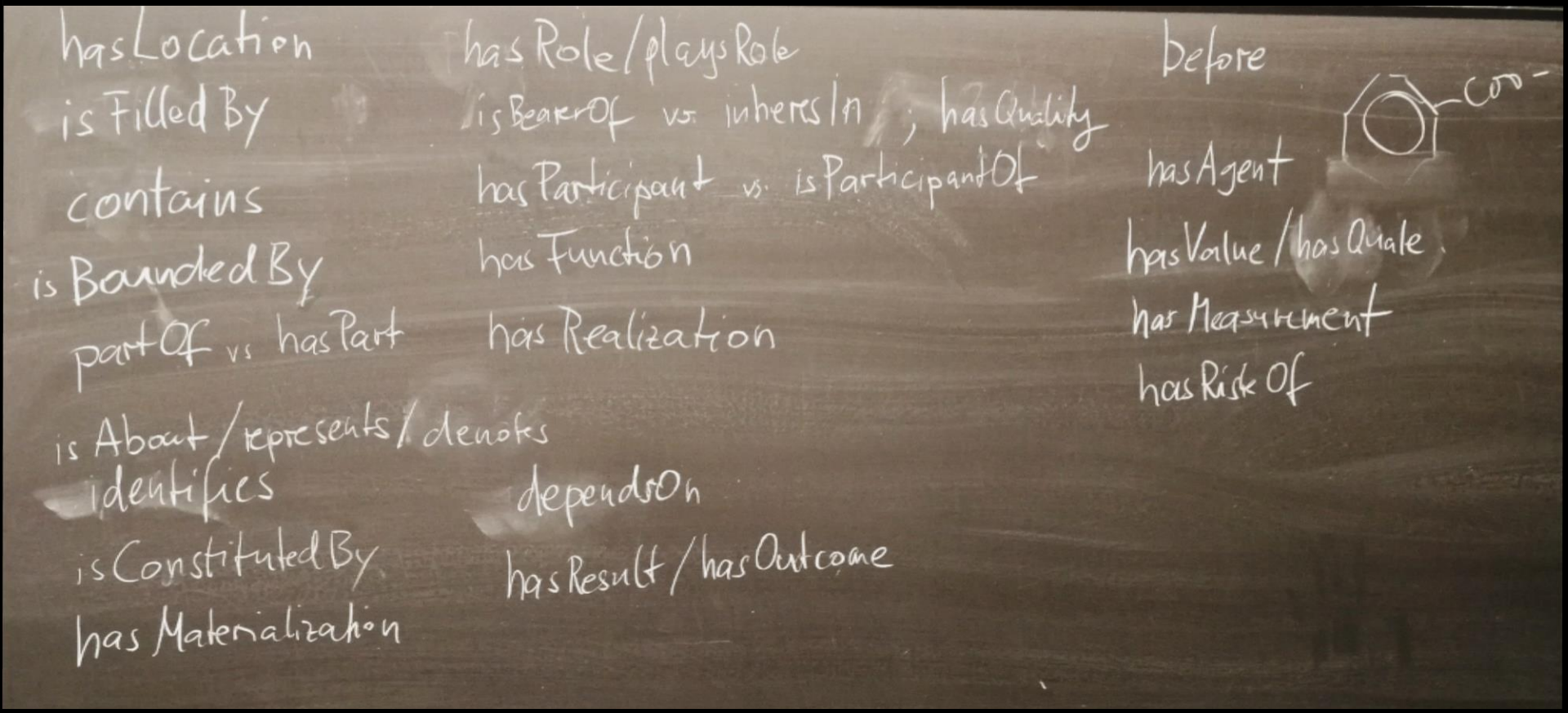
- What is the difference between the normal and the pathological?
  E.g. alopecia, vitiligo, lifestyle preferences, uncommon behaviour, ageing?
- Is this ontologically significant?
- Socioeconomic conditions
- Environment
- System
- Juridical “person”

# Example OGMS




- Ontology for general medical science
- https://bioportal.bioontology.org/ontologies/OGMS

# Ontological relations




As collected when discussing upper-level category assignment and exploring related entities

Roughly comparable with BioTop (next slide)

# BioTop ontology

- Domain-level foundational ontology for biology and medicine (BTL2 = BioTopLite v2)
- OWL-DL
- Strongly axiomatised
- Mapped to BFO and RO
- https://github.com/BioTopOntology/biotop
- Talk in JOWO 2018

Schulz, S., Boeker, M., & Martinez-Costa, C. (2017). The BioTop family of upper level ontological resources for biomedicine. *Stud Health Technol Inform*, *235*, 441-45.

# BTL2 Class Taxonomy

- 'particular at some time'
  - condition
  - disposition
  - function
  - 'immaterial object'
    - 'immaterial three dimensional physical entity'
    - 'one dimensional physical entity'
    - 'two dimensional physical entity'
    - 'zero dimensional physical entity'
  - 'information object'
    - plan
  - 'material object'
    - 'collective material entity'
      - 'amount of pure substance'
      - 'plurality of organisms'
    - compound
      - atom
      - 'compound of collective material entities'
      - 'mono molecular entity'
      - 'poly molecular composite entity'
        - 'structured biological entity'
          - cell
          - 'cellular component'
          - organism
          - 'organism part'
          - Universe
    - 'subatomic particle'
    - wave
  - process
    - action
      - 'plan execution'
    - 'collective process'
    - 'instantaneous process'
    - life
    - situation
  - quality
    - 'object quality'
      - 'physical length'
      - 'physical mass'
      - 'physical volume'
      - 'taxon quality'
    - 'process quality'
      - 'physical force'
  - role
  - 'temporal region'
    - 'point in time'
    - 'time interval'
  - 'value region'
    - 'canonicity value region'
      - 'canonical value region'
      - 'noncanonical value region'
    - 'taxon value region'

# BTL2 Relations

- 'owl:top object property'
  - 'at some time'
  - causes
    - 'has realization'
    - 'is agent in'
  - 'has condition'
  - 'has participant'
    - 'has agent'
    - 'has outcome'
    - 'has patient'
    - 'is life of'
  - includes
    - 'has part'
      - 'has boundary'
      - 'has component part'
      - 'has granular part'
    - 'is bearer of'
  - 'is caused by'
    - 'has agent'
    - 'is realization of'
  - 'is condition of'
  - 'is included in'
    - 'inheres in'
    - 'is part of'
      - 'is boundary of'
      - 'is component part of'
      - 'is granular part of'
  - 'is participant in'
    - 'has life'
    - 'is agent in'
    - 'is outcome of'
    - 'is patient in'
  - 'is preceded by'
  - 'is projection of'
  - 'is referred to at time'
  - 'is represented by'
  - precedes
  - 'projects onto'
  - represents

# BTL Axioms (examples)

**Description: MaterialObject**

Equivalent To

SubClass Of

- atSomeTime only MaterialObject
- hasBoundary only TwoDimensionalPhysicalEntity
- hasGranularPart only MaterialObject
- hasLife only Life
- hasLife some Life
- hasPart only (ImmaterialObject or MaterialObject)
- hasPart some MaterialObject
- hasPart some SubAtomicParticle
- isBearerOf only (Disposition or Function or InformationObject or ObjectQuality or Role)
- isBearerOf some PhysicalMass
- isBearerOf some PhysicalVolume
- isParticipantIn only Process
- Particular
- projectsOnto only ImmaterialThreeDimensionalPhysicalEntity
- projectsOnto some ImmaterialThreeDimensionalPhysicalEntity

**Description: InformationObject**

Equivalent To

SubClass Of

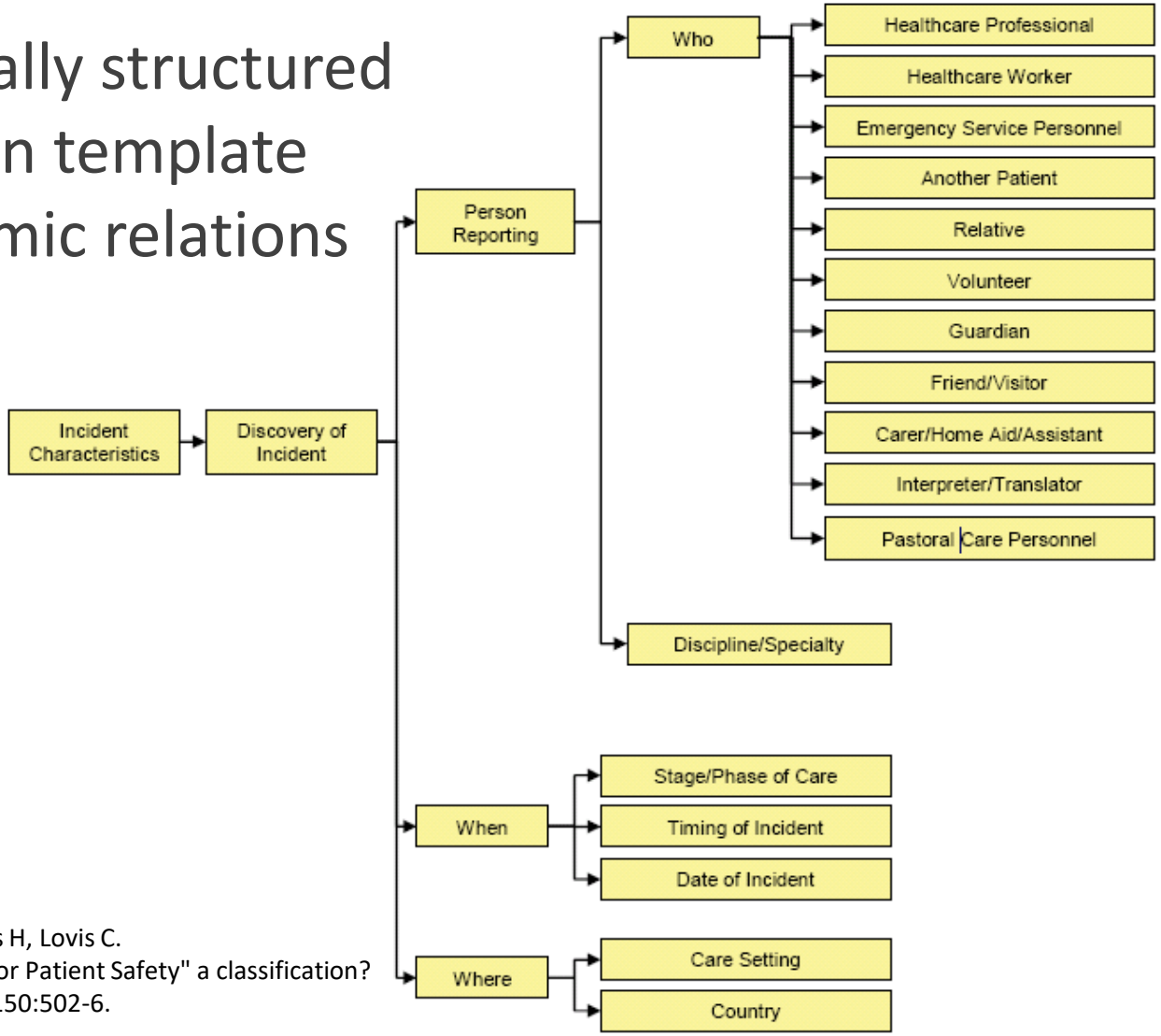
- atSomeTime only InformationObject
- Disposition or Function or InformationObject or ObjectQuality or Role
- hasLife only Life
- hasLife some Life
- hasPart only InformationObject
- hasRealization only Process
- inheresIn some MaterialObject
- isIncludedIn only (InformationObject or MaterialObject)
- isParticipantIn only Process
- isPartOf only InformationObject
- not (isBearerOf some PhysicalMass)
- not (isBearerOf some PhysicalVolume)
- Particular

# Hierarchical knowledge organization systems in biology and medicine

- ICD – International Classification of Diseases
- MeSH – Medical Subject Headings
- SNOMED CT
- OBO Foundry Ontologies
  - Gene Ontology
  - Foundational Model of Anatomy (FMA)
  - ChEBI – Chemical Entities of Biological Interest
- Meta – terminologies / Catalogies
  - UMLS – Unified Medical Language System
  - Bioportal
- Clinical Information Models

# Not all hierarchies are ontological

- Hierarchically structured information template no taxonomic relations




Schulz S, Karlsson D, Daniel C, Cools H, Lovis C.
Is the "International Classification for Patient Safety" a classification?
Stud Health Technol Inform. 2009;150:502-6.

# ICD – International Classification of Diseases

- A statistical classification of diseases, issued by WHO
  Most recent release: **ICD-11 for Mortality and Morbidity Statistics** (2018)
  Main building principles:
  - Single, mostly taxonomic hierarchies
  - Non-overlapping classes
- Rules to assure this principle:
  - E.g., Diabetes mellitus excludes Diabetes mellitus in pregnancy, which is in a different branch of the hierarchy
- "Residuals" like "other", "unspecified"
- https://icd.who.int/browse11/l-m/en

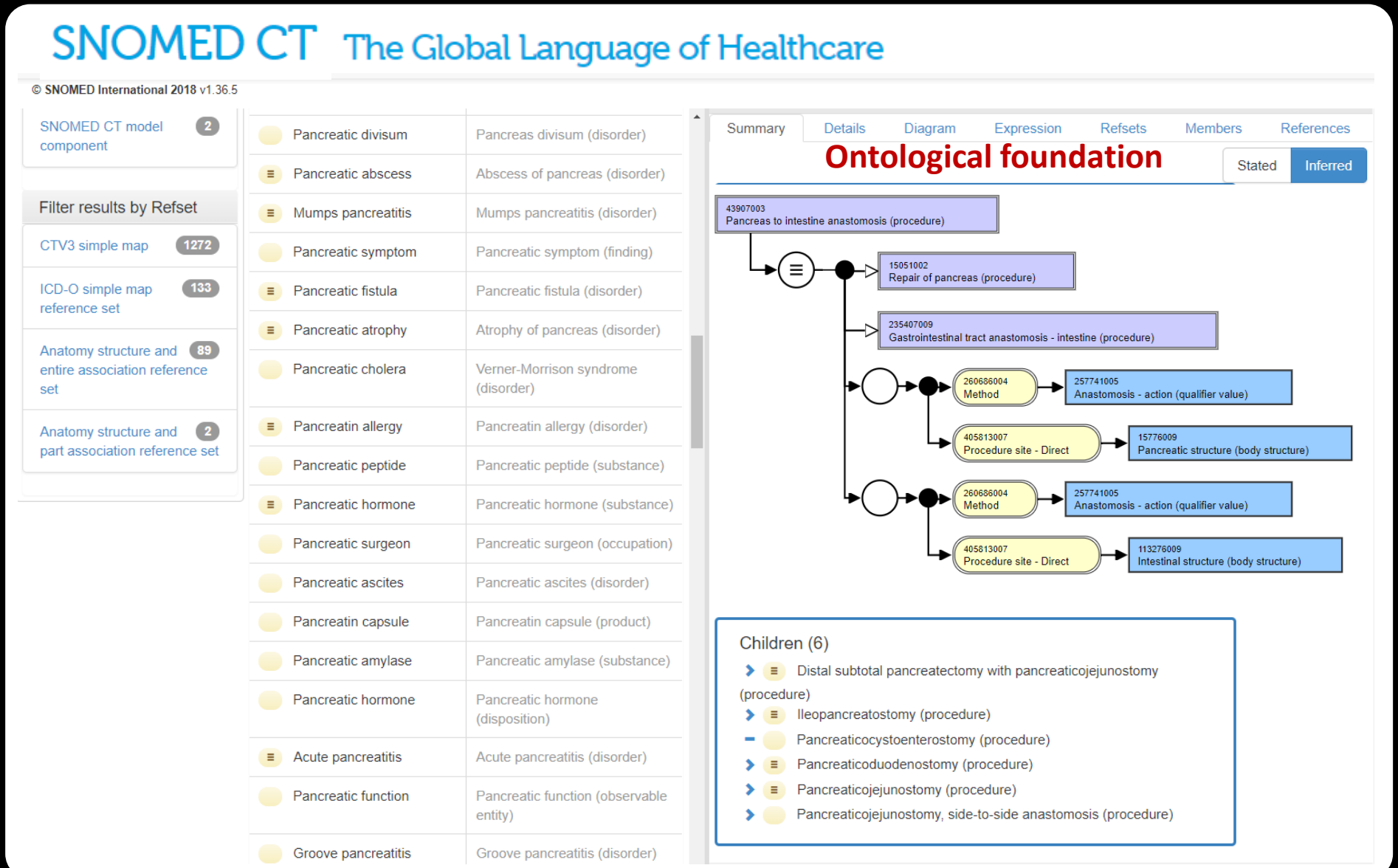
# MeSH – Medical Subject Headings

- Thesaurus for Literature Indexing in Retrieval, issued by the U.S. National Library of Medicine
- All MEDLINE literature records are manually annotated with MeSH concepts
- Multi-hierarchical (overlap of tree-like hierarchies), spans all areas of medicine and biology
- E.g. a paper indexed by “aspirin” and “stomach ulcer” would be found in a query with “antipyretics” and “gastrointestinal diseases”
- https://www.ncbi.nlm.nih.gov/mesh/

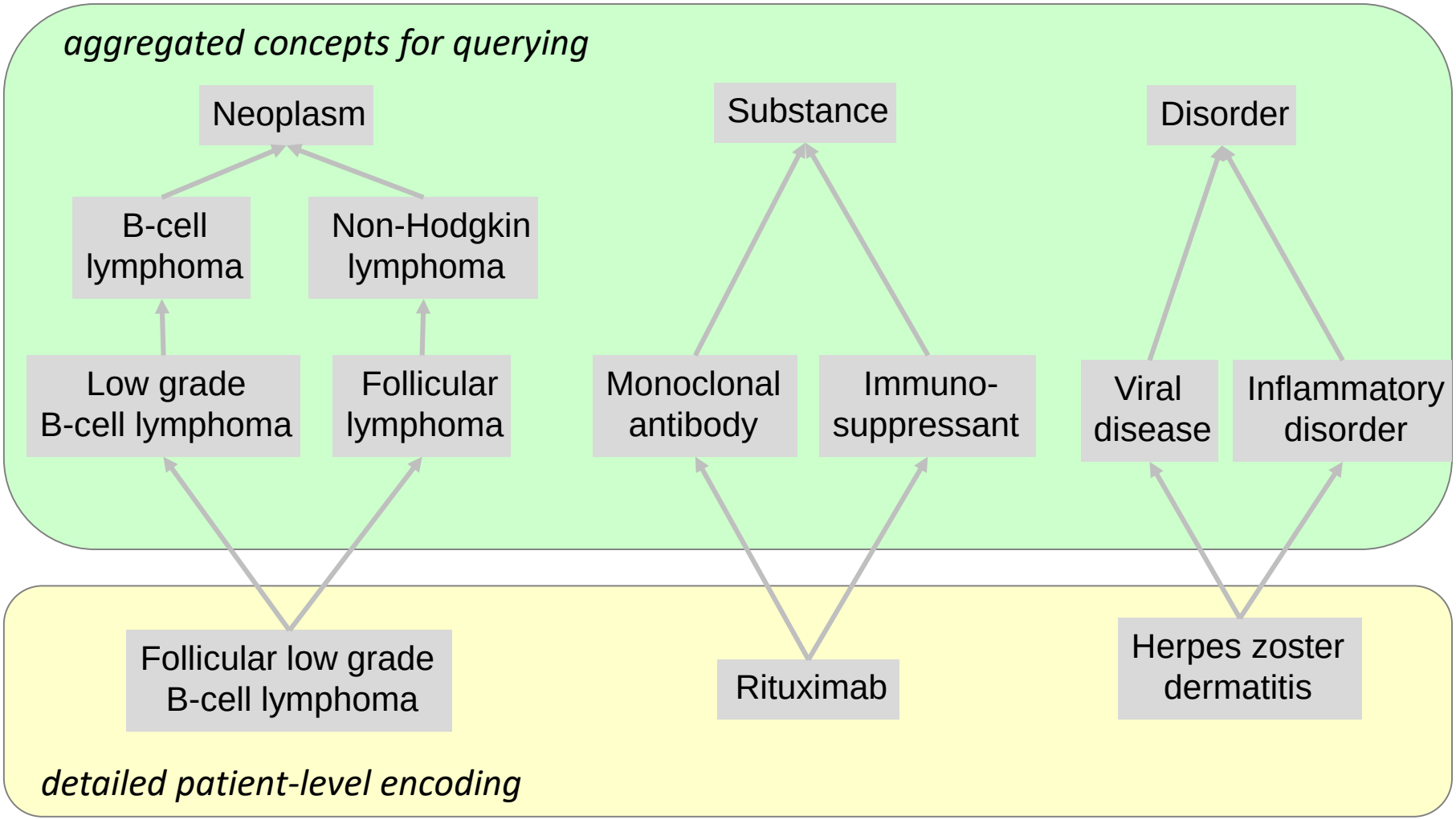
# SNOMED CT

- Ontology-based terminology for representing content of the electronic health record
- Run by an international standards organisation, requires licence for clinical use
- Distributed in a tabular form, can be transformed into OWL – EL
- Has its own OWL-like compositional syntax
- Some semantic issues unresolved
- http://browser.ihtsdotools.org/

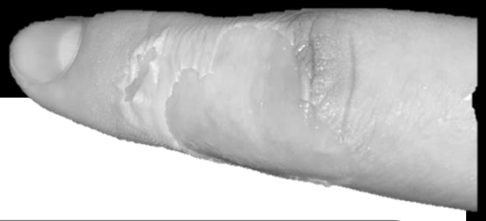
# SNOMED CT: reference terminology



# SNOMED CT – Structural benefits (I): Polyhierachies



# SNOMED CT – Structural benefits (II): Co-ordination



**Pre-coordination**

*"Verbrennung 2. Grades eines einzelnen Fingers"*

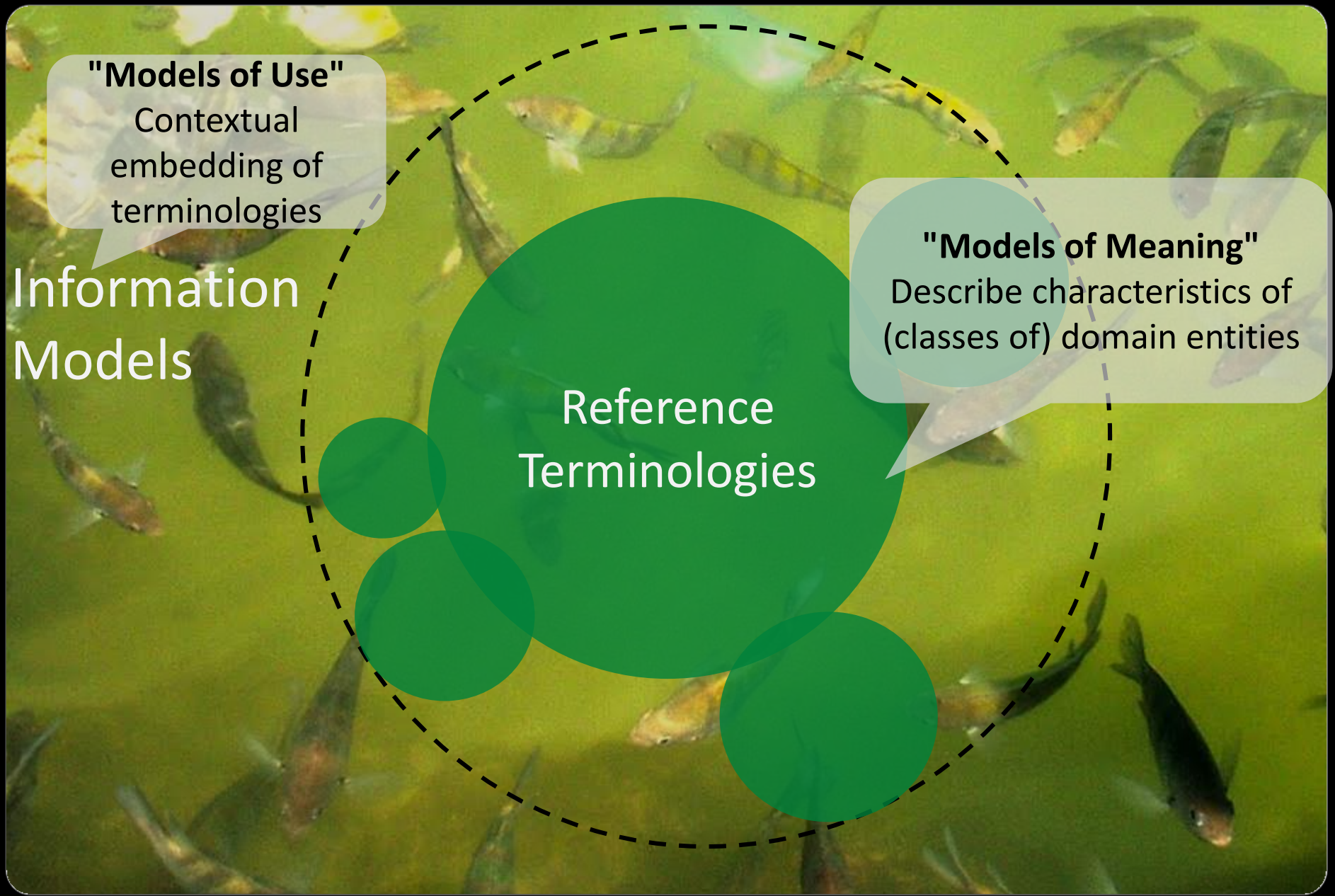
```
211908006 |Deep partial thickness burn of a single finger (disorder)|
                                    ≡
<<< 29673001 |Second degree burn of single finger, not thumb (disorder)| :
{ 116676008 |Associated morphology| = 262588000 |Deep partial thickness burn
(morphologic abnormality)|,363698007 |Finding site| = 56213003 |Skin of
finger (body structure)| }
```

**Post-coordination**

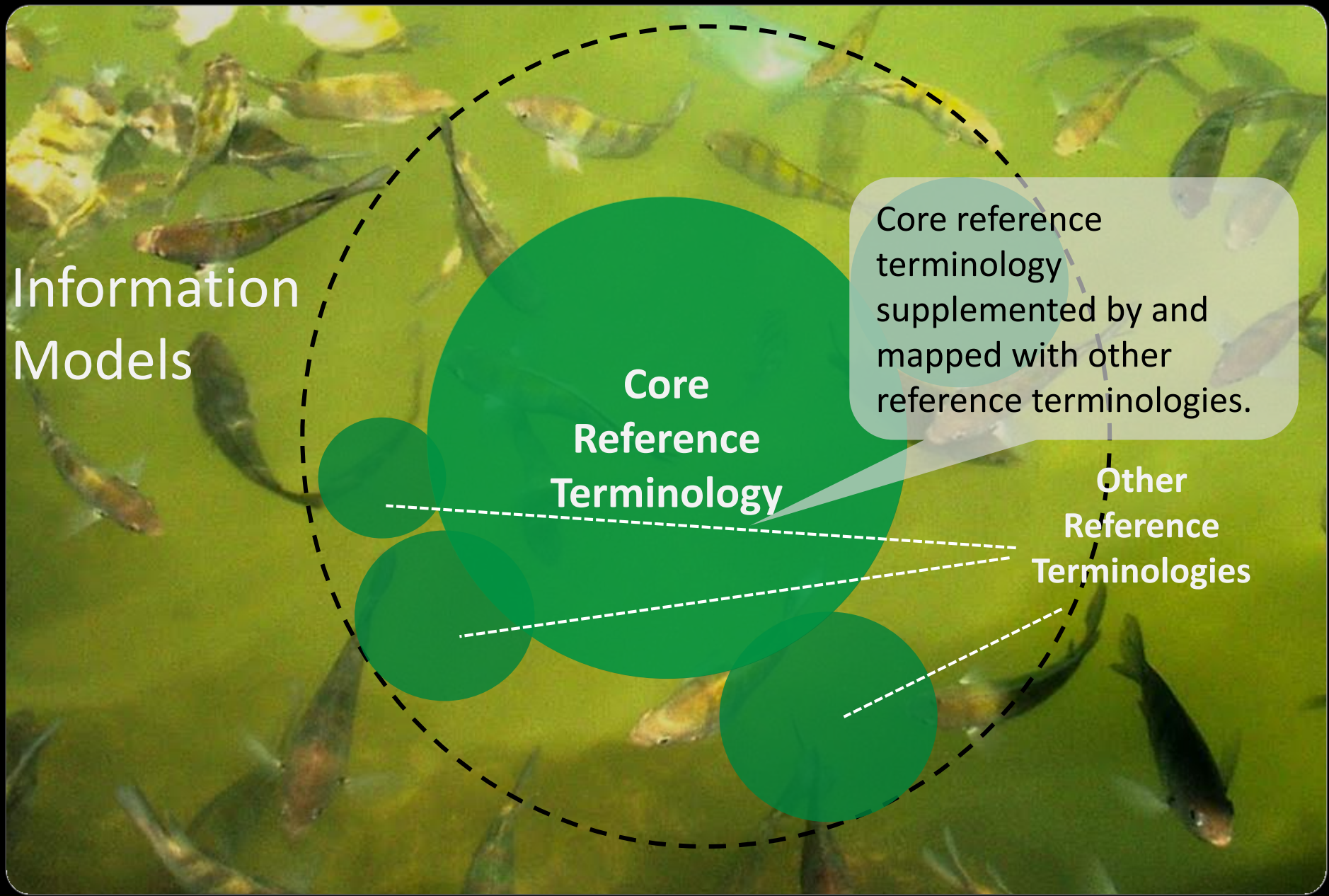
*"Verbrennung 2. Grades der Rückseite des rechten Zeigefingers"*

```
<<< 29673001 |Second degree burn of single finger, not thumb (disorder)| :
{ 116676008 |Associated morphology| = 262588000 |Deep partial thickness burn
(morphologic abnormality)|,363698007 |Finding site| = 37314006 | Skin
structure of dorsal surface of index finger (body structure) |, 272741003
|Laterality| = 24028007 |Right (qualifier value)| }
```

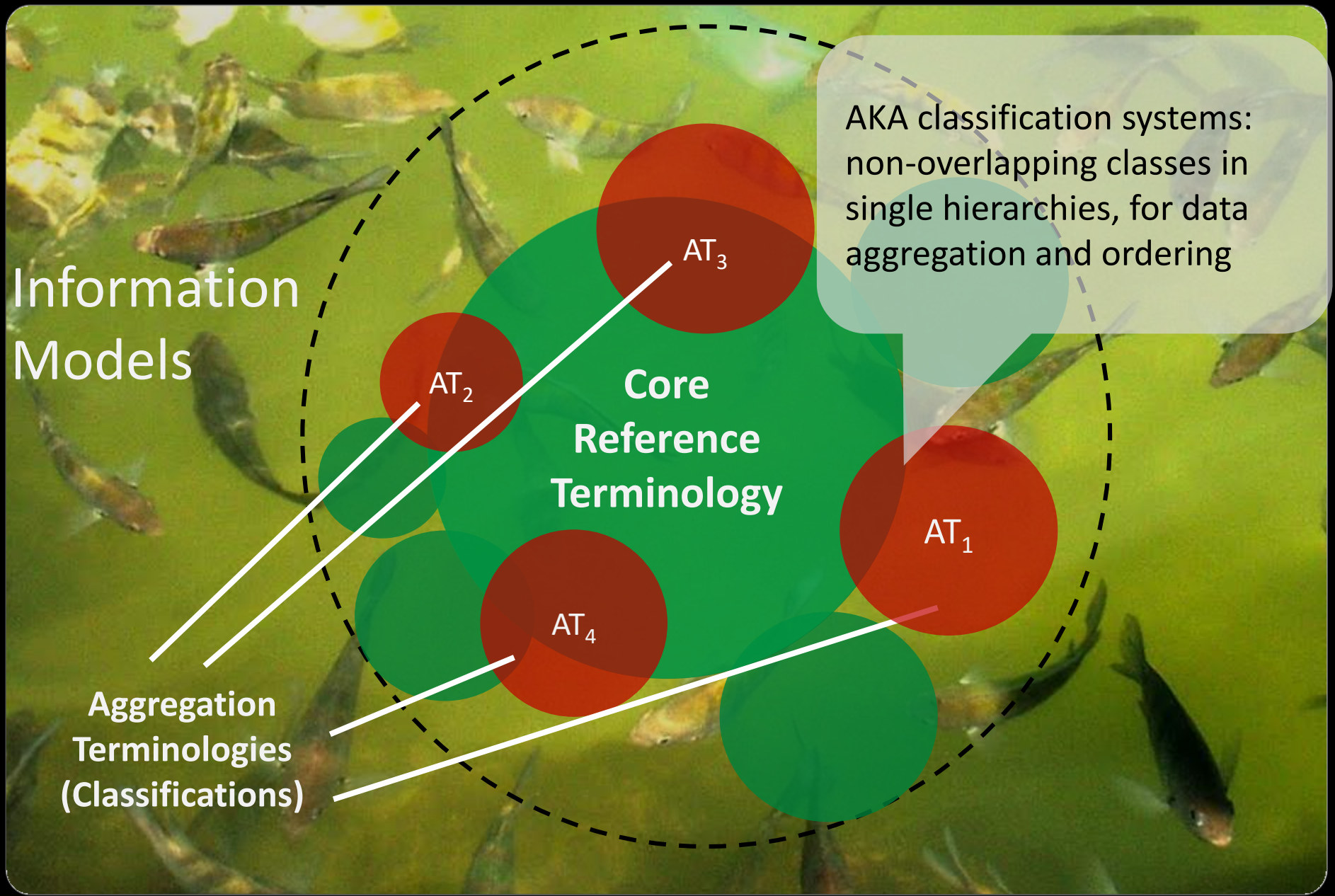
# Interoperability ecosystem



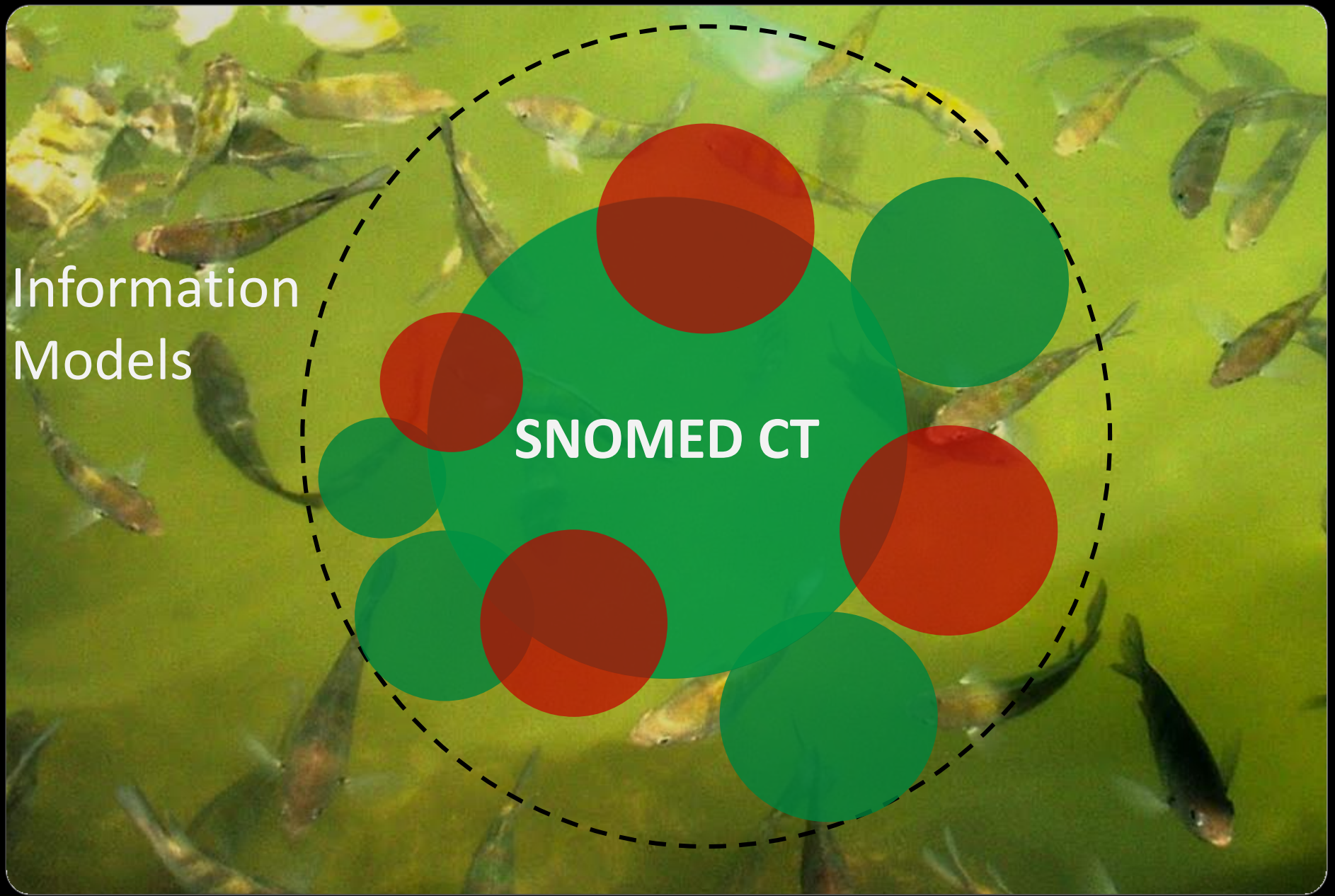
# Interoperability ecosystem



# Interoperability ecosystem

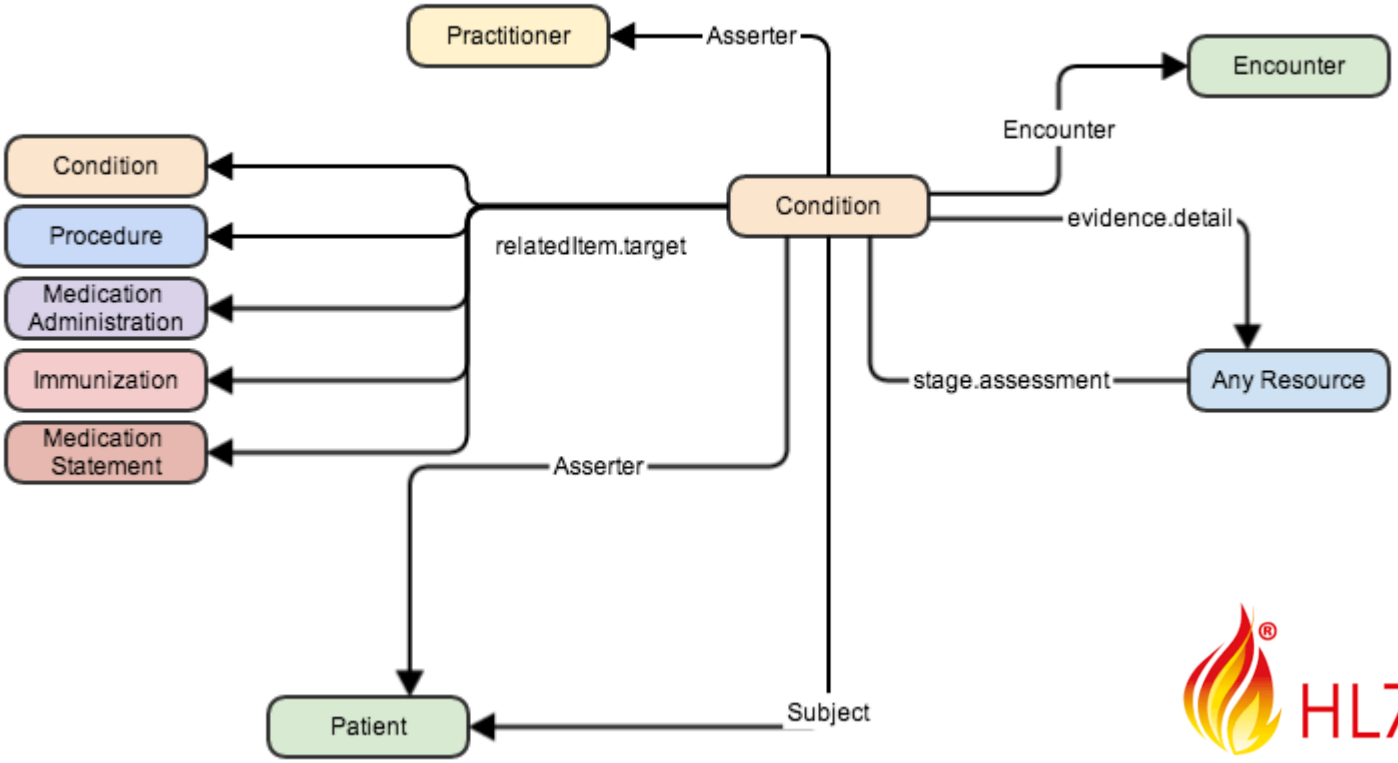


# Interoperability ecosystem



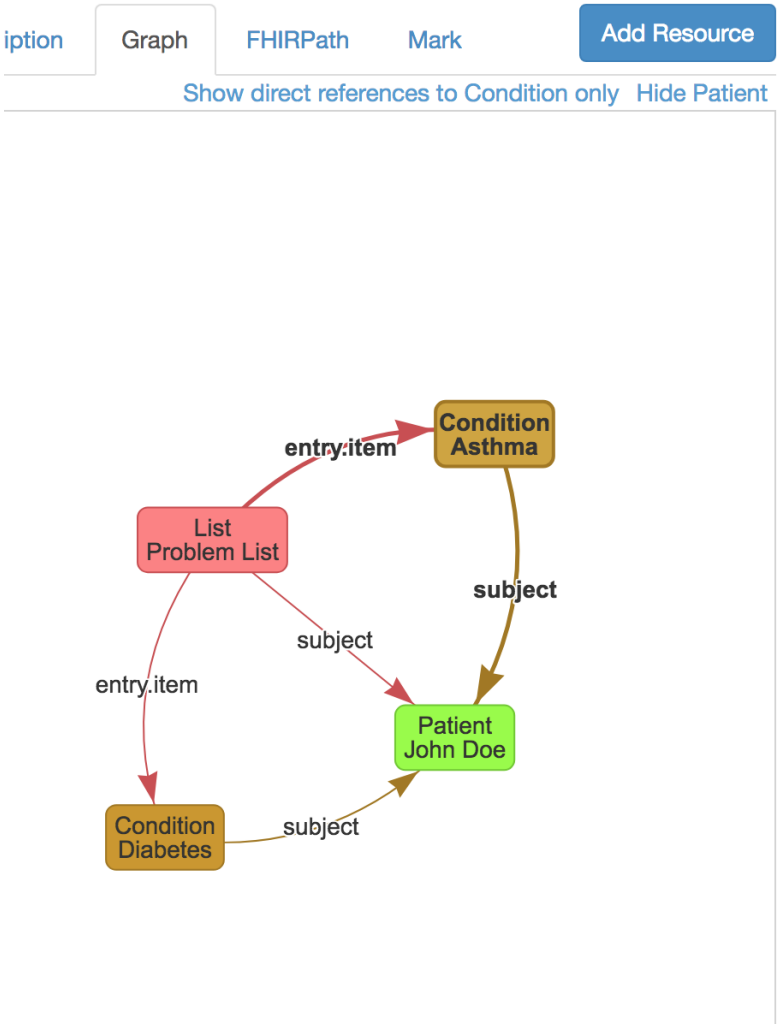
# Information models

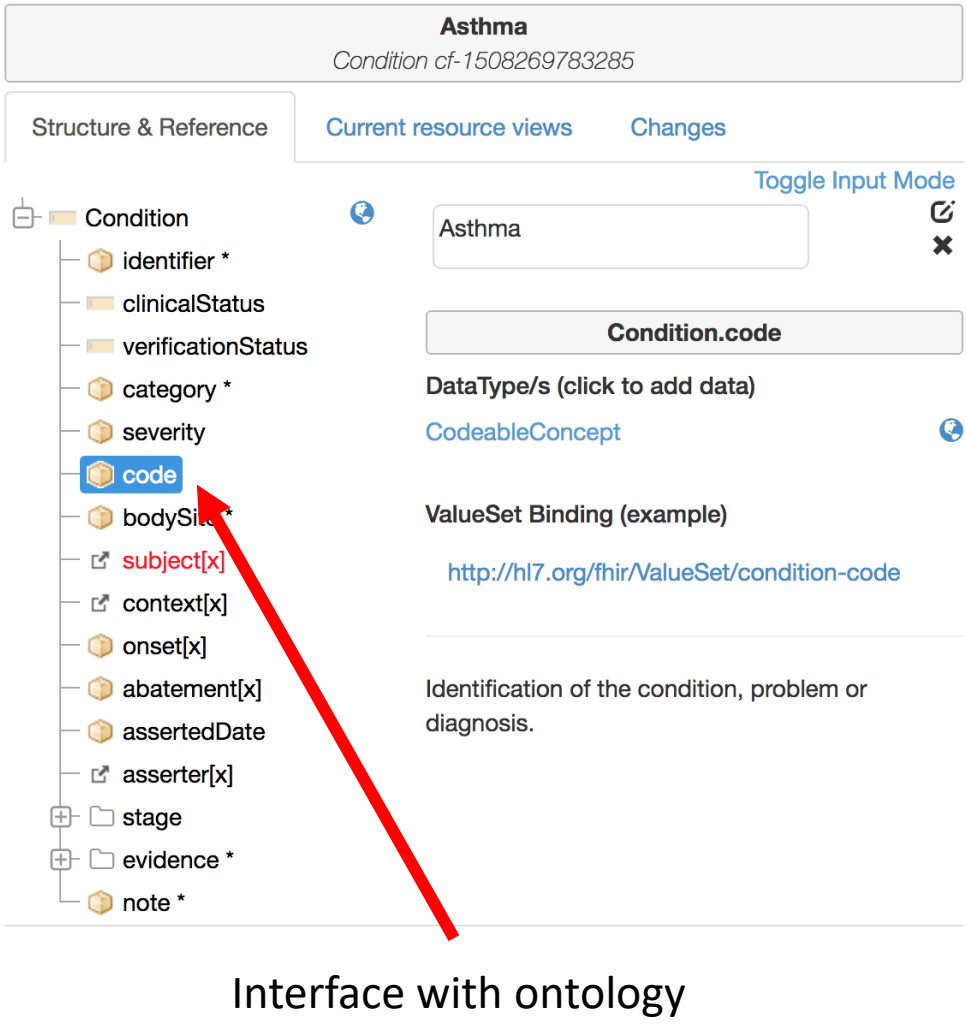
- “models of use” vs. “models of meaning”
- Recording templates for health care




HL7® FHIR®

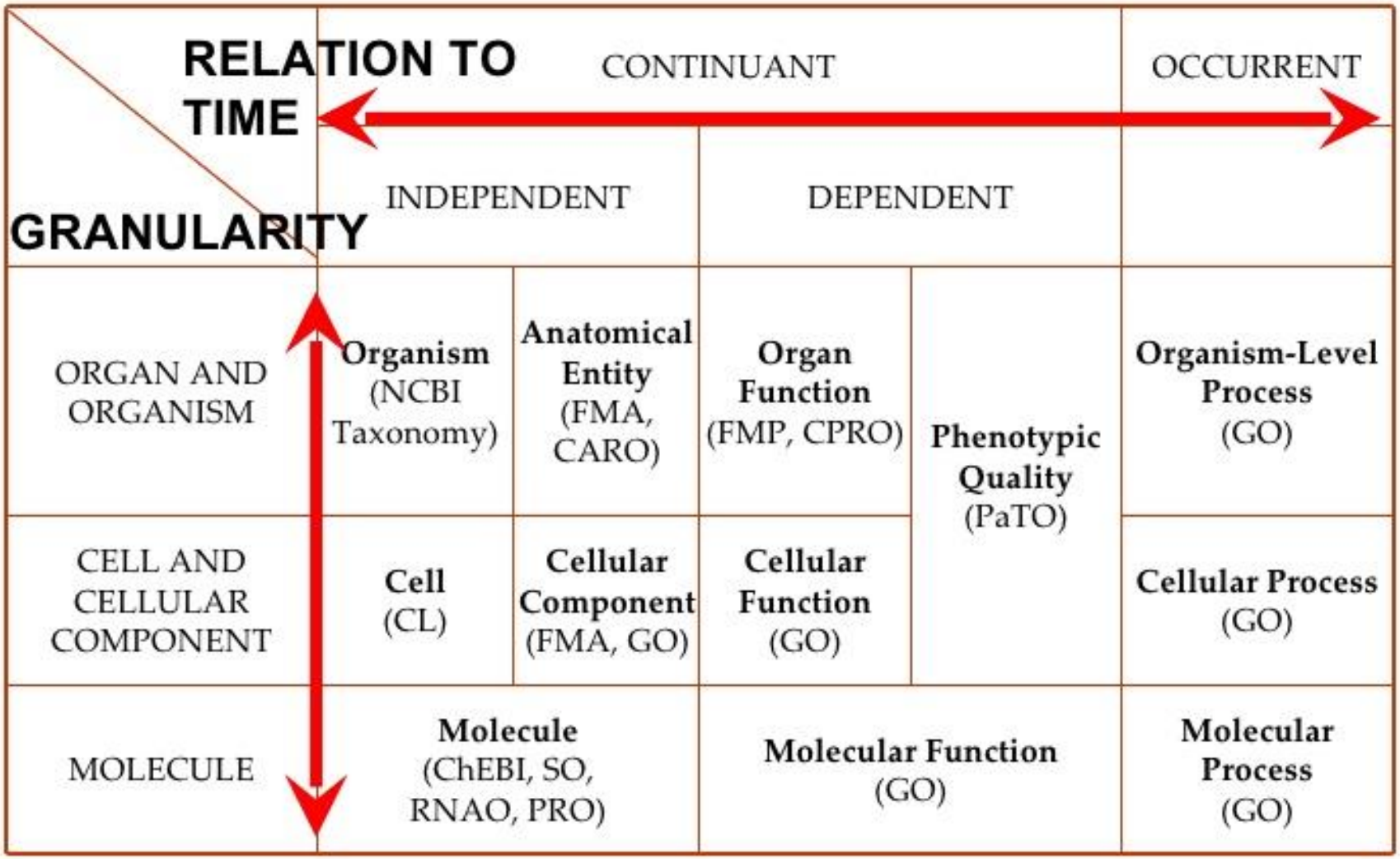
# Example: “concept” in information models




Interface with ontology

# Open biomedical ontologies



# Ontology Repositories

- UMLS – Unified Medical Language System https://uts.nlm.nih.gov/home.html
- Bioportal https://bioportal.bioontology.org/