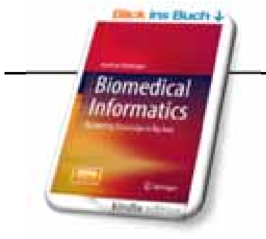


Andreas Holzinger

VO 709.049 Medical Informatics

28.10.2015 11:15-12:45

Lecture 03

Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)

a.holzinger@tugraz.at

Tutor: markus.plass@student.tugraz.at

http://hci-kdd.org/biomedical-informatics-big-data

Schedule

- 1. Intro: Computer Science meets Life Sciences, challenges, future directions
- 2. Back to the future: Fundamentals of Data, Information and Knowledge
- **3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)**
- 4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use
- 5. Semi structured and weakly structured data (structural homologies)
- 6. Multimedia Data Mining and Knowledge Discovery
- 7. Knowledge and Decision: Cognitive Science & Human-Computer Interaction
- 8. Biomedical Decision Making: Reasoning and Decision Support
- 9. Intelligent Information Visualization and Visual Analytics
- 10. Biomedical Information Systems and Medical Knowledge Management
- 11. Biomedical Data: Privacy, Safety and Security
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

Keywords of the 3th Lecture

- Biomedical Ontologies
- Classification of Diseases
- International Classification of Diseases (ICD)
- Medical Subject Headings (MeSH)
- Modeling biomedical knowledge
- Ontology Languages (OL)
- Resource Description Framework (RDF)
- Standardized Medical Data
- Systematized Nomenclature of Medicine (SNOMED)
- Unified Medical Language System (UMLS)
- Work domain model (WDM)

- … have acquired background knowledge on some issues in standardization and structurization of data;
- … have a general understanding of modeling knowledge in medicine and biomedical informatics;
- … got some basic knowledge on medical Ontologies and are aware of the limits, restrictions and shortcomings of them;
- … know the basic ideas and the history of the International Classification of Diseases (ICD);
- … have a view on the Standardized Nomenclature of Medicine Clinical Terms (SNOMED CT);
- … have some basic knowledge on Medical Subject Headings (MeSH);
- … understand the fundamentals and principles of the Unified Language System (UMLS);

Advance Organizer (1/2)

- **Abstraction** = process of mapping (biological) processes onto a series of concepts (expressed in mathematical terms);
- **Biological system** = a collection of objects ranging in size from molecules to populations of organisms, which interact in ways that display a collective function or role (= collective behaviour);
- **Coding** = any process of transforming descriptions of medical diagnoses and procedures into standardized code numbers, i.e. to track health conditions and for reimbursement; e.g. based on Diagnosis Related Groups (DRG)
- **Data model** = definition of entities, attributes and their relationships within complex sets of data;
- **DSM** = Diagnostic and Statistical Manual for Mental Disorders
- **Extensible Markup Language (XML)** = set of rules for encoding documents in machine-readable form.
- **GALEN** = Generalized Architecture for Languages, Encyclopedias and Nomenclatures in Medicine is a project aiming at the development of a reference model for medical concepts
- **ICD** = International Classification of Diseases, the archetypical coding system for patient record abstraction (est. 1900)
- **Medical Classification** = provides the terminologies of the medical domain (or at least parts of it), there are 100+ various classifications in use;
- **MeSH** = Medical Subject Headings is a classification to index the world medical literature and forms the basis for UMLS

Advance Organizer (2/2)

- **Metadata** = data that describes the data;
- **Model** = a simplified representation of a process or object, which describes its behaviour under specified conditions (e.g. conceptual model);
- **Nosography** = science of description of diseases;
- **Nosology** = science of classification of diseases;
- **Ontology** = structured description of a domain and formalizes the terminology (concepts-relations, e.g. IS-A relationship provides a taxonomic skeleton), e.g. gene ontology;
- **Ontology engineering** = subfield of knowledge engineering, which studies the methods and methodologies for building ontologies;
- **SNOMED** = Standardized Nomenclature of Medicine, est. 1975, multitaxial system with 11 axes;
- **SNOP** = Systematic Nomenclature of Pathology (on four axes: topography, morphology, etiology, function), basis for SNOMED;
- **System features** = static/dynamic; mechanistic/phenomenological; discrete/continous; deterministic/stochastic; single-scale/multi-scale
- **Terminology** = includes well-defined terms and usage;
- **UMLS** = Unified Medical Language System is a long-term project to develop resources for the support of intelligent information retrieval;

Glossary

- ACR = American College of Radiologists
- API = Application Programming Interface
- DAML = DARPA Agent Markup Language
- DICOM = Digital Imaging and Communications in Medicine
- DL = Description Logic
- ECG = Electrocardiogram
- EHR = Electronic Health Record
- FMA = Foundational Model of Anatomy
- FOL = First-order logic
- GO = Gene Ontology
- ICD = International Classification of Diseases
- IOM = Institute of Medicine
- KIF = Knowledge Interchange Format, a FOL-based language for knowledge interchange.
- LOINC = Logical Observation Identifiers Names and Codes
- MeSH = Medical Subject Headings
- MRI = Magnetic Resonance Imaging
- NCI = National Cancer Institute (US)
- NEMA = National Electrical Manufacturer Association
- OIL = Ontology Inference Layer (description logic)
- OWL = Ontology Web Language
- RDF = Resource Description Framework
- RDF Schema = A vocabulary of properties and classes added to RDF
- SCP = Standard Communications Protocol
- SNOMED CT = Systematized Nomenclature of Medicine – Clinical Terms
- SOP = Standard Operating Procedure
- UMLS = Unified Medical Language System

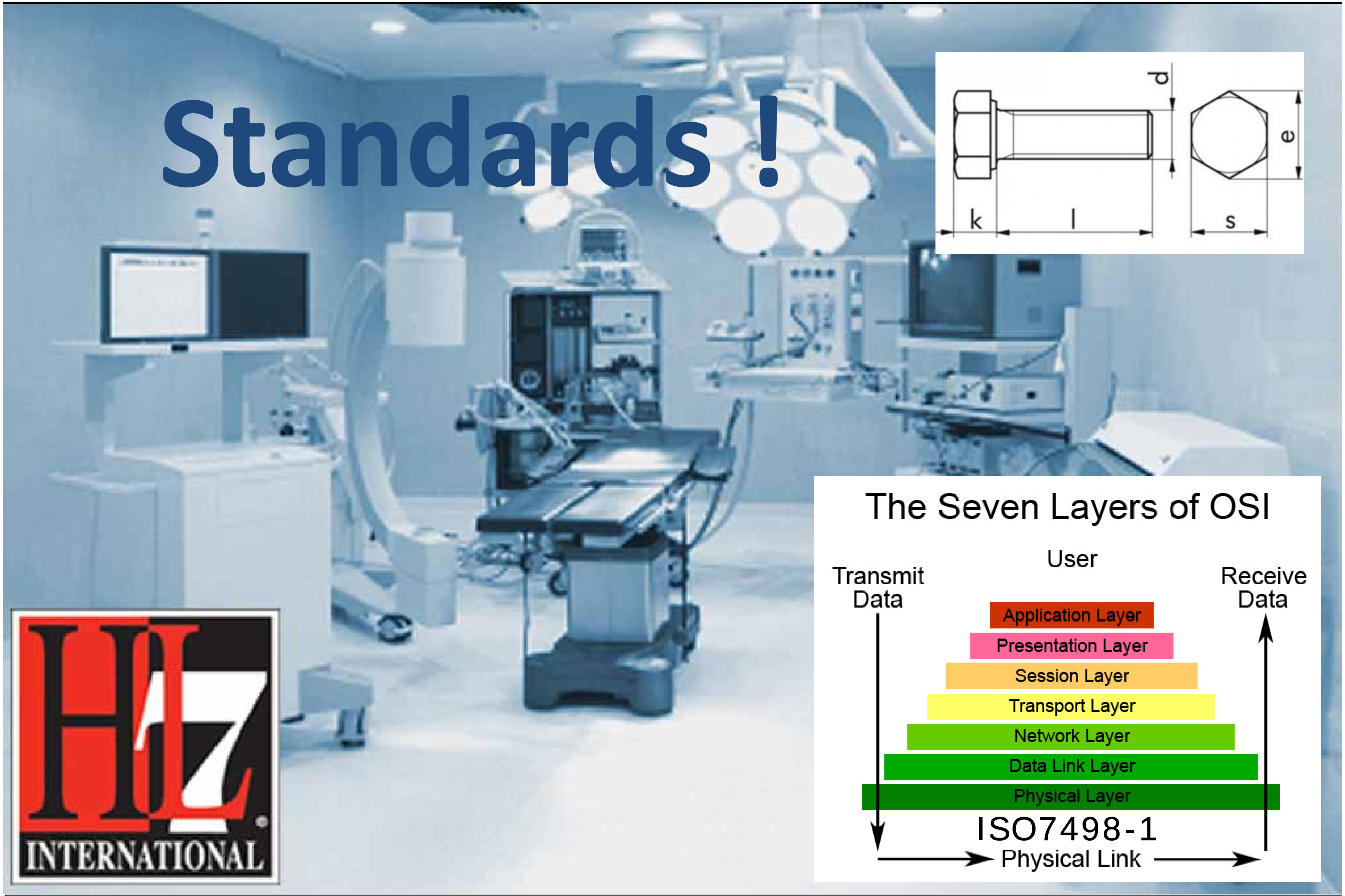
- To find a trade-off between standardization and **personalization** [1];
- The large amounts of **non-standardized data** and **unstructured information** ("free text") [2];
- **Low integration** of standardized terminologies in the daily clinical practice (Who is using e.g. SNOMED, MeSH, UMLS in daily routine?);
- **Low acceptance** of classification codes amongst practitioners;

1. Holmes, C., Mcdonald, F., Jones, M., Ozdemir, V., Graham, J. E. 2010. Standardization and Omics Science: Technical and Social Dimensions Are Inseparable and Demand Symmetrical Study. Omics-Journal of Integr. Biology, 14, (3), 327-332.
2. Holzinger, A., Schantl, J., Schroettner, M., Seifert, C. & Verspoor, K. 2014. Biomedical Text Mining: State-of-the-Art, Open Problems and Future Challenges. In: LNCS 8401. Berlin Heidelberg: Springer pp. 271-300.


FOR A FAIR SELECTION EVERYBODY HAS TO TAKE THE SAME EXAM: PLEASE CLIMB THAT TREE

Standards ?

Standards !



d
e
k
l
s
The Seven Layers of OSI
User
Transmit Data
Receive Data
Application Layer
Presentation Layer
Session Layer
Transport Layer
Network Layer
Data Link Layer
Physical Layer
ISO7498-1
Physical Link
HL7
INTERNATIONAL

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. BME-19, NO. 5, SEPTEMBER 1972 331

Standardization and Health Care

J. H. U. BROWN, SENIOR MEMBER, IEEE, AND DEWITT JAMES LOWELL

***Abstract*–In order to deliver reasonable health care to all people, it is essential that standards be established. Standards vary with the type of control and with the approach desired in determining the quality of care. This paper discusses various kinds of standards and their application in the health care field. Standards may be determined as a process or as a direct regulation. It is probable that regulation of standards by process is the most satisfactory method.**

Introduction

SOCIETY cannot exist without a yardstick by which its accomplishments or failures are measured. Such yardsticks are called *standards*. They are created by the need for regulation and control as an escape from anarchy or to motivate towards greater achievement. In the ultimate, society dictates these limits by the demands it places upon itself. Standards provide opportunities for security and augmentation of process and output by virtue of the goal and process structure that they provide.

arbiter may be the market place or agencies that rely on expertise from many sources to set acceptable standards of quality or performance. For these reasons, the final moderator may be found in a governmental authority, and its delegation into a system of regulation, law, and judicial action, so that an established code can become the focal point of resolution.

The Objectives of Standardization

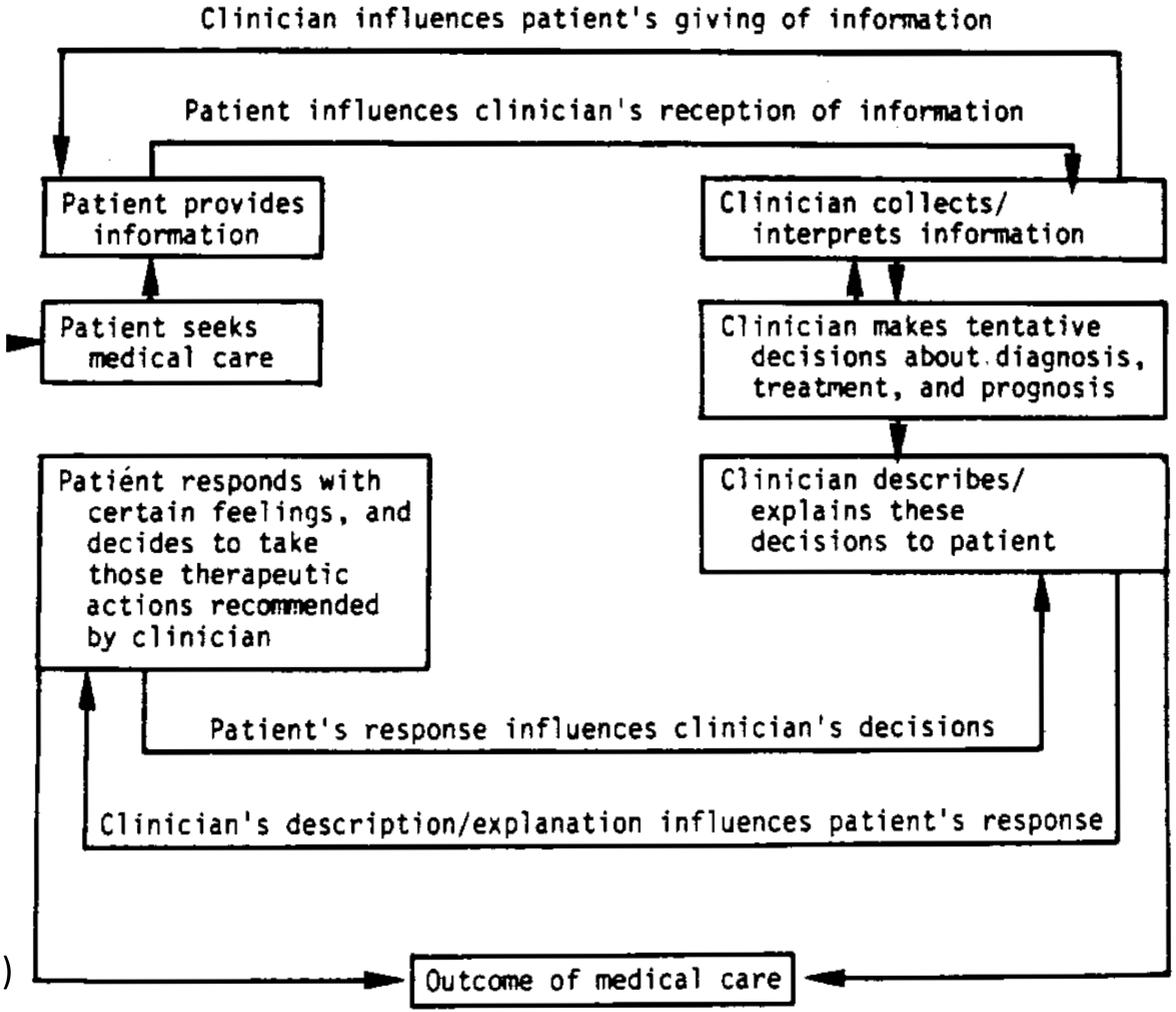
Standards have value within themselves in that they help establish quality. However, they accomplish more for society than the mere establishment of a level of quality and performance. A standard allows coordination of effort between producers so that like products can be produced. It permits the reproduction of similar units in mass quantity and permits the consumer to judge one product or service against another by performance. It establishes *freedom* of *interchange* of material and ideas, and permits the activity in one part of society

Brown, J. H. U. & Loweli, D. J. (1972) Standardization and Health Care. *IEEE Transactions on Biomedical Engineering, BME-19, 5, 331-334.*

- Medical (clinical) data are defined and detected disturbingly “soft” …
- … having an obvious degree of **variability** and **inaccuracy**.
- Taking a medical history, the performance of a physical examination, the interpretation of laboratory tests, even the definition of diseases … are surprisingly **inexact**.
- Data is defined, collected, and interpreted with a degree of variability and inaccuracy which falls far short of the standards **which engineers do expect from most data**.
- Moreover, standards might be **interpreted variably** by different medical doctors, different hospitals, different medical schools, different medical cultures, …

Komaroff, A. L. (1979) The variability and inaccuracy of medical data. *Proceedings of the IEEE, 67, 9, 1196-1207.*

Slide 3-3: The patient-clinician dialogue (from 1979)

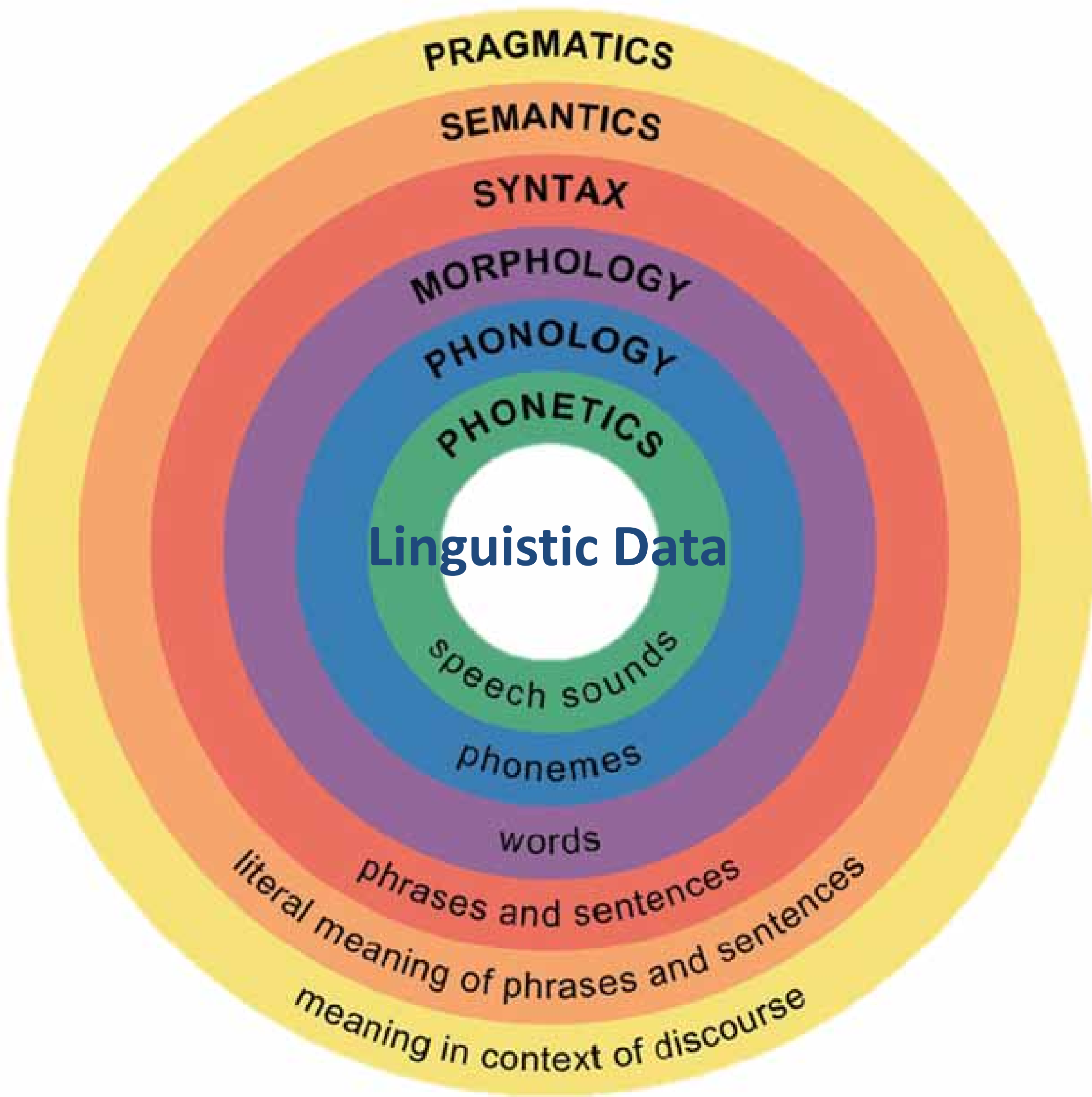


Komaroff (1979)

- … ensures that information is interpreted by all users with the same understanding;
  - supports the reusability of the data,
  - improves the efficiency of healthcare services and
  - avoids errors by reducing duplicated efforts in data entry;

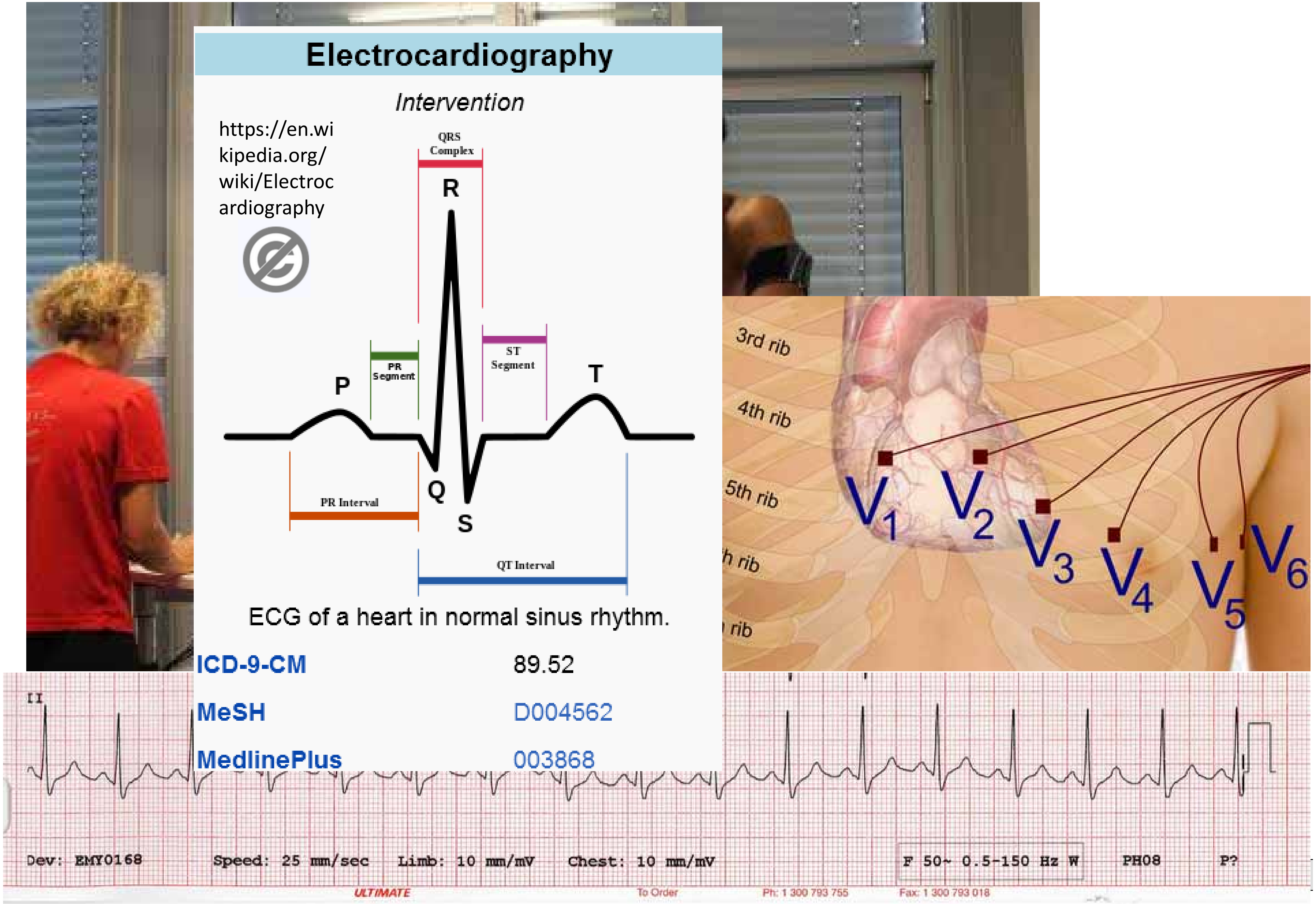
- Data standardization refers to
  - a) the data content;
  - b) the terminologies that are used to represent the data;
  - c) how data is exchanged; and
  - iv) how knowledge, e.g. clinical guidelines, protocols, decision support rules, checklists, standard operating procedures are represented in the health information system (refer to IOM).
- Elements for sharing require standardization of identification, record structure, terminology, messaging, privacy etc.
- The most used standardized data set to date is the **International Classification of Diseases (ICD),** which was first adopted in 1900 for collecting statistics (Ahmadian et al. 2011)

Slide 3-5: Complex Example: Non-Standardized Data

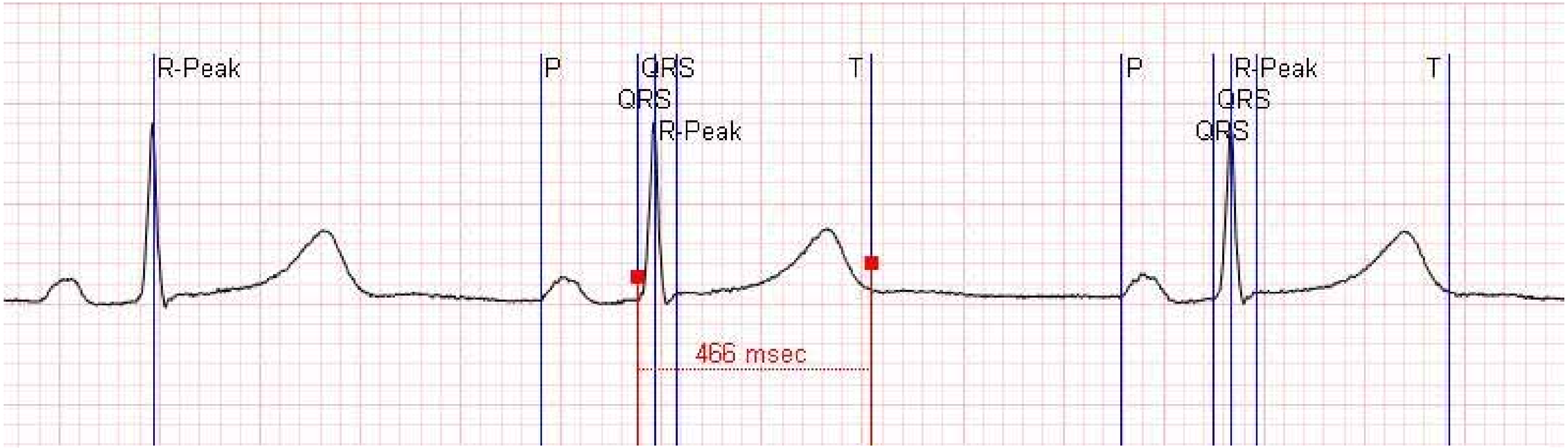


Thomas, J. J. & Cook, K. A. 2005. *Illuminating the path: The research and development agenda for visual analytics, New York, IEEE Computer Society Press.*

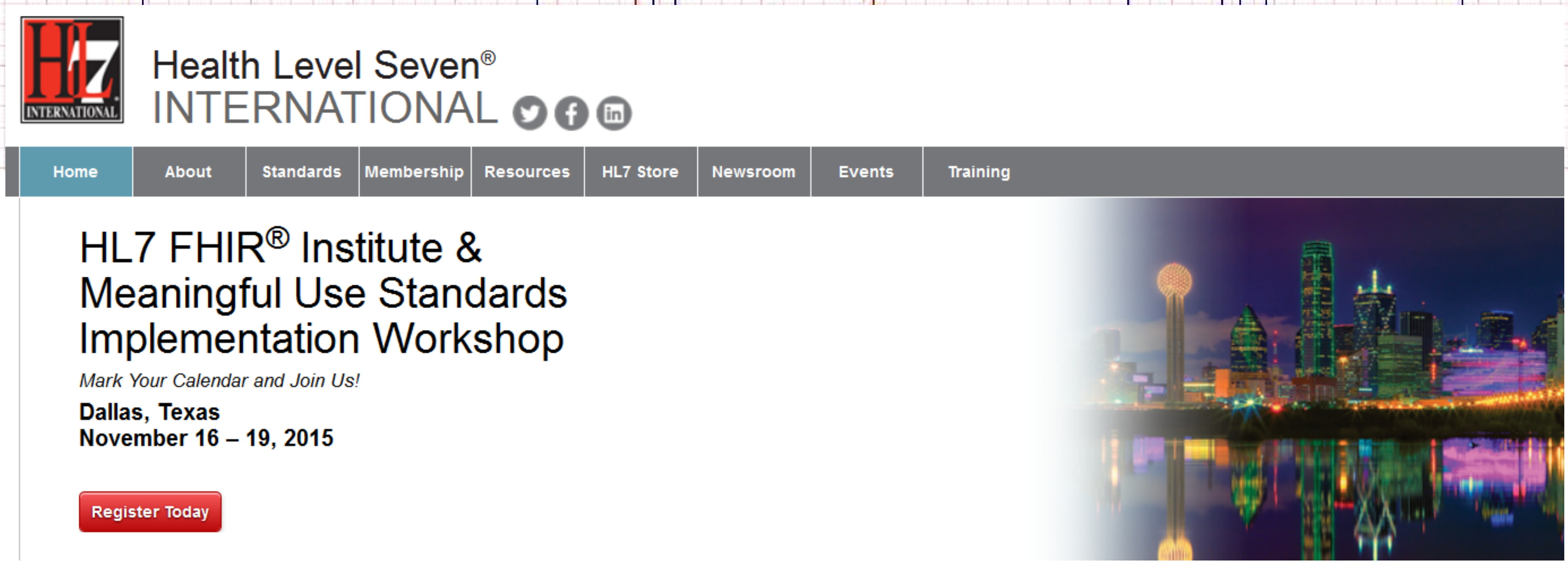
Example: ECG



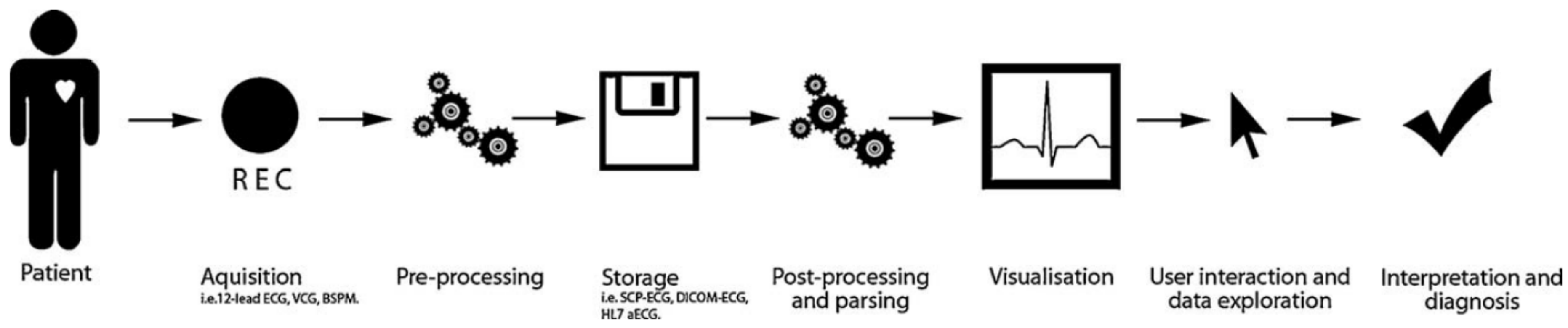
Electrocardiography
Intervention
https://en.wikipedia.org/wiki/Electrocardiography
QRS Complex
R
PR Segment
ST Segment
P
T
Q
S
PR Interval
QT Interval
ECG of a heart in normal sinus rhythm.
ICD-9-CM 89.52
MeSH D004562
MedlinePlus 003868
3rd rib
4th rib
5th rib
V1
V2
V3
V4
V5
V6
II
Dev: EMY0168
Speed: 25 mm/sec
Limb: 10 mm/mV
Chest: 10 mm/mV
F 50~ 0.5-150 Hz W
PH08
P?
ULTIMATE
To Order
Ph: 1 300 793 755
Fax: 1 300 793 018

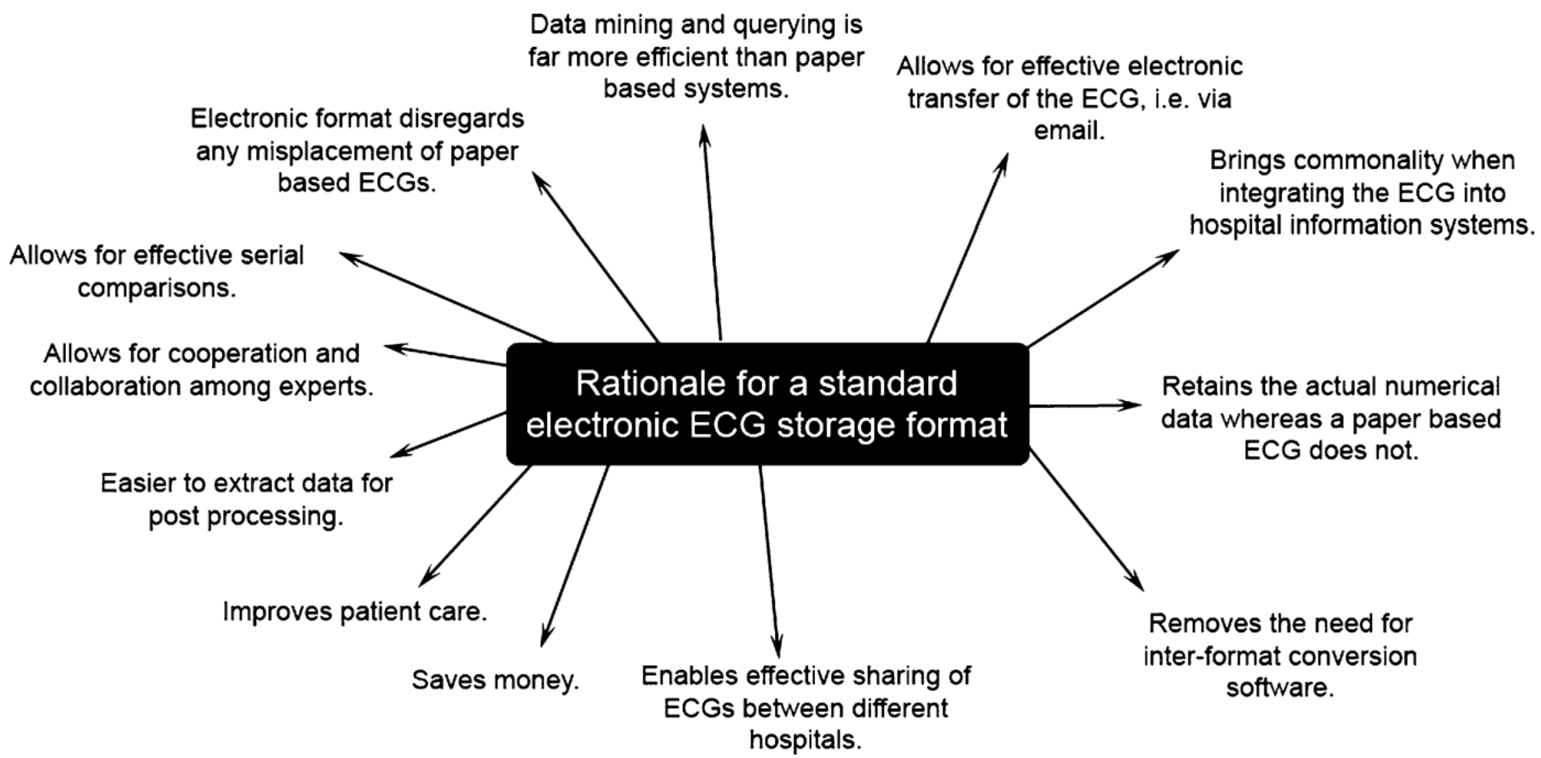


R-Peak
P
QRS
QRS
R-Peak
T
P
R-Peak
QRS
QRS
T
466 msec



HL7 INTERNATIONAL
Health Level Seven®
INTERNATIONAL
Home
About
Standards
Membership
Resources
HL7 Store
Newsroom
Events
Training
HL7 FHIR® Institute &
Meaningful Use Standards
Implementation Workshop
Mark Your Calendar and Join Us!
Dallas, Texas
November 16 – 19, 2015
Register Today





Bond, R. R., Finlay, D. D., Nugent, C. D. & Moore, G. (2011) A review of ECG storage formats. *International Journal of Medical Informatics 80, 10, 681-697.*

- There has been a large number of ECG storage formats proclaiming to promote interoperability.
- There are three predominant ECG formats:
  - SCP-ECG (1993, European Standard, Binary data)
  - DICOM-ECG (2000, European Standard, Binary data)
  - HL7 aECG (2001, ANSI Standard, XML data)
- A mass of researchers have been proposing their own ECG storage formats to be considered for implementation (= proprietary formats).
- Binary has been the predominant method for storing ECG data

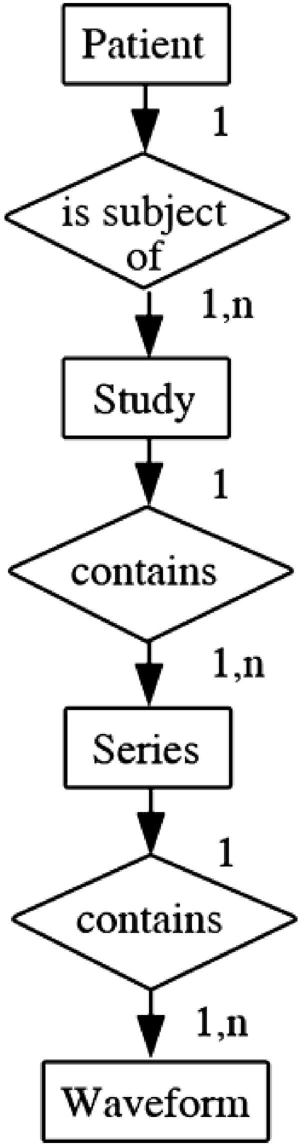
Bond, R. R., Finlay, D. D., Nugent, C. D. & Moore, G. (2011) A review of ECG storage formats. *International Journal of Medical Informatics, 80, 10, 681-697.*

- Overview on current ECG storage formats

ECG format	Year	Method of implementation	Specification	Viewers
SCP-ECG	1993	BINARY	Can be freely downloaded from the Internet [7].	Freely available SCP-ECG Viewer made by EcgSoft [8].
DICOM-WS 30	2000	BINARY	Can be freely downloaded from the Internet [5].	Freely available DICOM-ECG viewer made by Charruasoft [9].
HL7 aECG	2001	XML	The XML Schema can be used as the specification or the implementation guide by AMPS [6].	Freely available aECG viewer by AMPS [10].
ecgML	2003	XML	Can be freely downloaded from the Internet [11].	None currently exist. Under development.
MFER	2003	BINARY	Can be freely downloaded from the Internet [12].	Freely available MFER viewer [13].
Philips XML	2004	XML	The specification is packaged with the actual product.	Philips viewer. Not freely available.
XML-ECG	2007	XML	Can be freely downloaded from the Internet [14].	XML-ECG viewer [14]. Not freely available.
mECGml	2008	XML	Can be freely downloaded from the Internet [15].	mECGml mobile viewer [15]. Not freely available.
ecgAware	2008	XML	Can be freely downloaded from the Internet [16].	TeleCardio viewer [16]. Not freely available.

Bond, R. R., Finlay, D. D., Nugent, C. D. & Moore, G. (2011) A review of ECG storage formats. *International Journal of Medical Informatics, 80, 10, 681-697.*

Slide 3-10: Example of a Binary ECG file



31	20	08	20	18	20	55	49	2a	20	31	2e	32	2e	38	32	1 . . UI* 1.2.82
36	2e	30	2e	31	2e	33	34	34	37	31	2e	32	2e	34	34	6.0.1.34471.2.44
2e	36	2e	32	30	30	32	31	31	32	32	30	39	31	30	30	.6.2002112209100
30	2e	2e	31	08	20	20	20	44	41	08	20	32	30	30	32	0..1. DA. 2002
31	31	32	32	08	20	23	20	44	41	08	20	32	30	30	32	1122. # DA. 2002
31	31	32	32	08	20	2a	20	44	54	0e	20	32	30	30	32	1122. * DT. 2002
31	31	32	32	30	39	31	30	30	30	08	20	30	20	54	4d	1122091000. 0 TM
06	20	30	39	31	30	30	30	08	20	33	20	54	4d	06	20	. 091000. 3 TM.
30	39	31	30	30	30	08	20	50	20	53	48	20	20	08	20	091000. P SH .
60	20	43	53	04	20	45	43	47	20	08	20	70	20	4c	4f	` CS. ECG . p LO
08	20	55	6e	6b	6e	6f	77	6e	20	08	20	90	20	50	4e	. Unknown . □ PN
20	20	08	20	60	10	50	4e	20	20	08	20	70	10	50	4e	. `.PN . p.PN
20	20	08	20	90	10	4c	4f	06	20	45	4c	49	32	35	30	. □.LO. ELI250
10	20	10	20	50	4e	06	20	73	6d	69	74	68	20	10	20	. . PN. smith
20	20	4c	4f	08	20	53	42	4a	2d	31	32	33	20	10	20	LO. SBJ-123 .
30	20	44	41	08	20	31	39	35	33	30	35	30	38	10	20	0 DA. 19530508.
40	20	43	53	02	20	4d	20	10	20	20	10	4c	4f	20	20	@ CS. M . .LO
10	20	10	10	41	53	20	20	10	20	20	10	44	53	20	20	. ..AS . .DS
10	20	30	10	44	53	20	20	18	20	20	10	4c	4f	20	20	. 0.DS . .LO

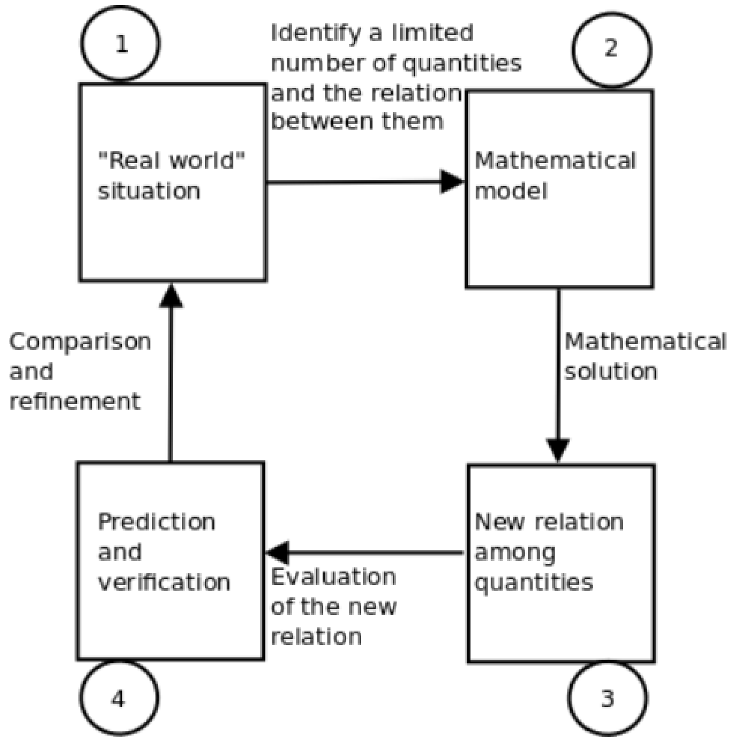
Bond et al. (2011)

Slide 3-11: Example of a XML ECG file

```
<sequenceSet>

    <component>
        <sequence>
            <code code="TIME_ABSOLUTE" codeSystem="2.16.840.1.113883.5.4"
                codeSystemName="ActCode" displayName="Aboslute Time"/>

            <value xsi:type="GLIST_TS">

                <head value="20021122091000.000"/>

                <increment value="0.002" unit="s"/>

            </value>
        </sequence>

    </component>

    <component>
```

Bond et al. (2011)

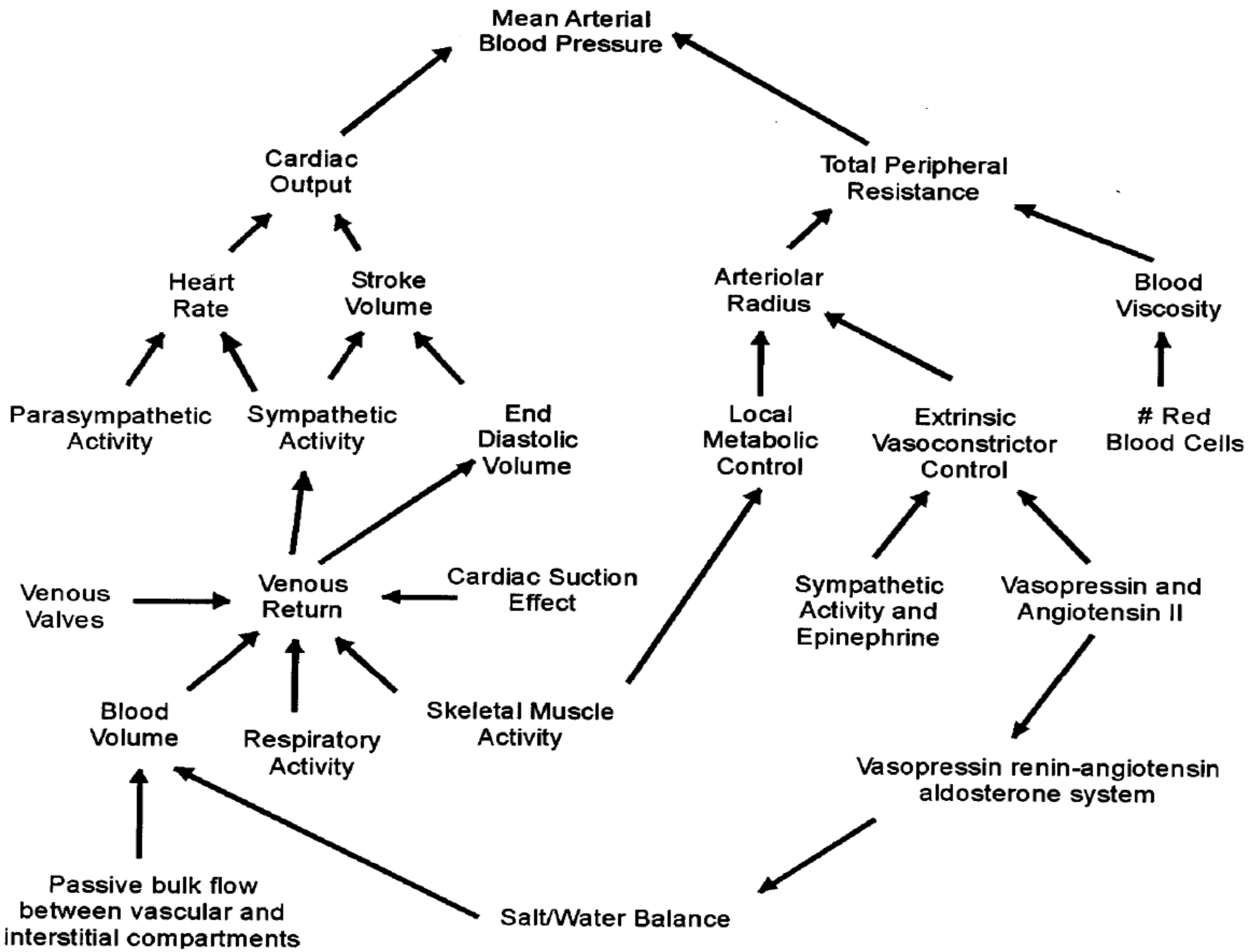
How do we represent biomedical knowledge?



Mathematical Logic	Psychology	Biology	Statistics	Economics
Aristotle				
Descartes				
Boole	James		Laplace	Bentham Pareto
Frege Peano			Bernoullii	Friedman
Goedel Post Church Turing Davis Putnam Robinson	Hebb Bruner Miller Newell, Simon	Lashley Rosenblatt Ashby Lettvin McCulloch, Pitts Heubel, Weisel	Bayes Tversky, Kahneman	Von Neumann Simon Raiffa
Logic PROLOG	SOAR **KBS, Frames**	**Connectionism**	**Causal Networks**	**Rational Agents**

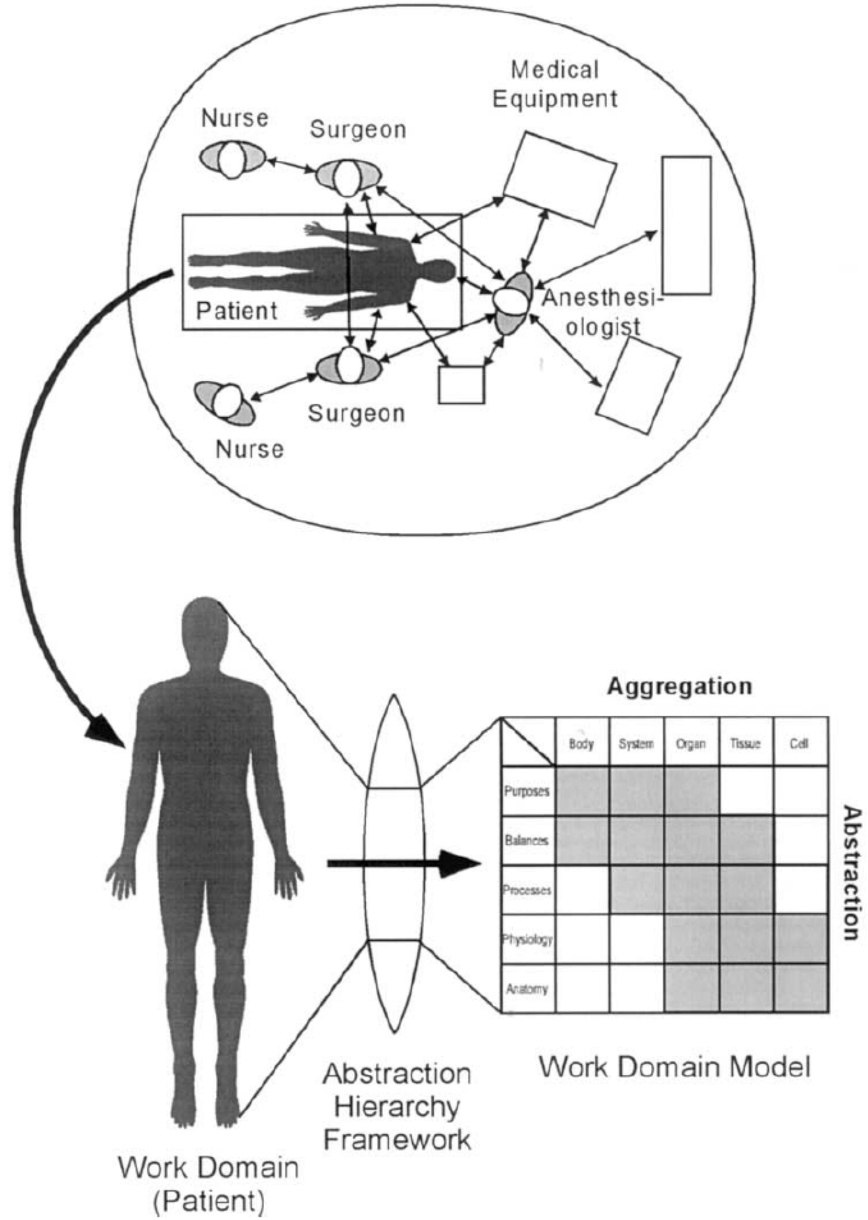
Davis, R., Shrobe, H. , Szolovits, P. 1993 What is a knowledge representation? AI Magazine, 14, 1, 17-33.

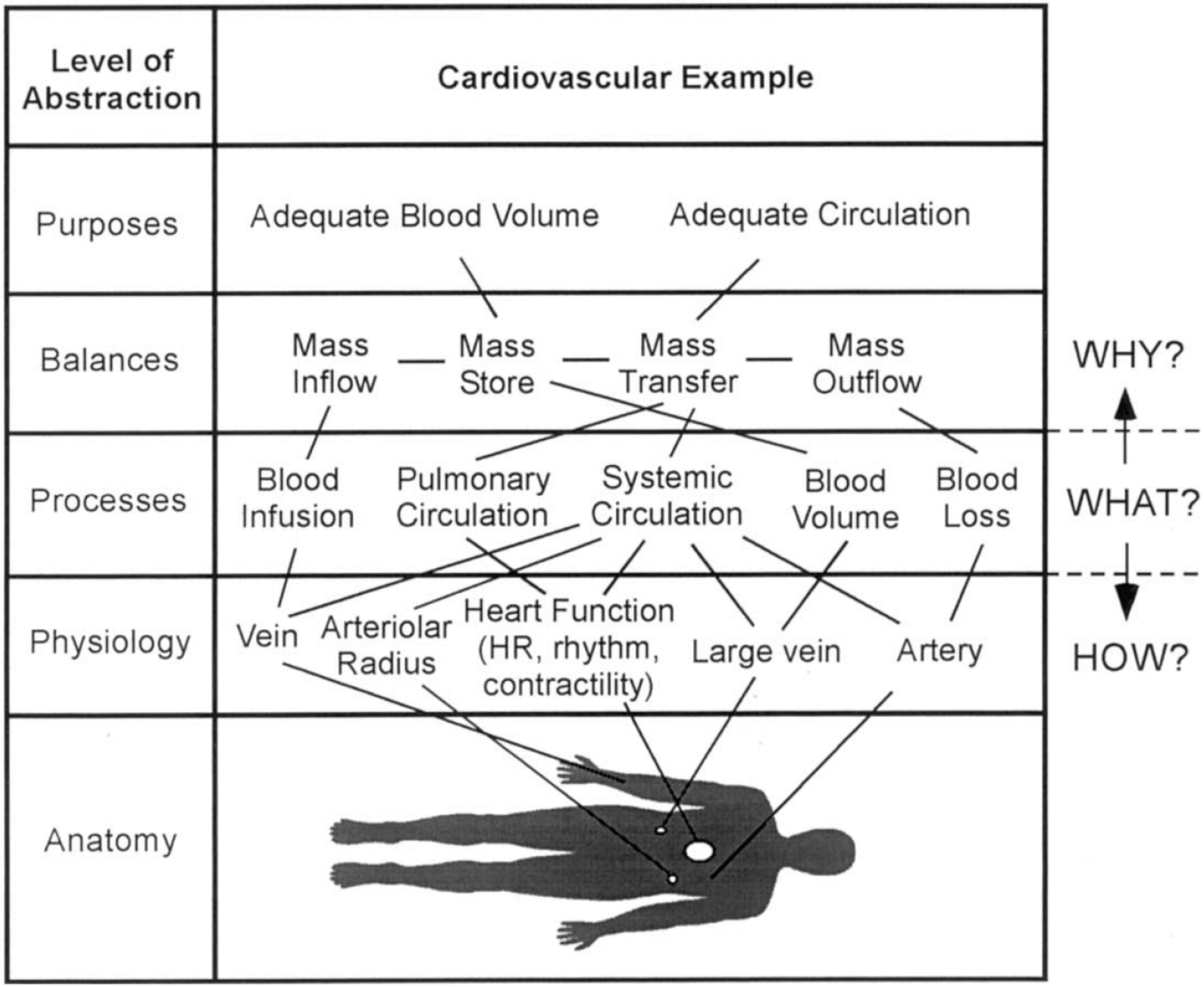
Slide 3-12 Example for Modeling of biomedical knowledge



Hajdukiewicz, J. R., Vicente, K. J., Doyle, D. J., Milgram, P. & Burns, C. M. (2001) Modeling a medical environment: an ontology for integrated medical informatics design. *International Journal of Medical Informatics, 62, 1, 79-99.*

Hajdukiewicz, J. R., Vicente, K. J., Doyle, D. J., Milgram, P. & Burns, C. M. (2001) Modeling a medical environment: an ontology for integrated medical informatics design. *International Journal of Medical Informatics, 62, 1, 79-99.*





Hajdukiewicz et al. (2001)

Slide 3-15: WDM of: (a) the human body

Level of Aggregation

a)

Level of Abstraction	Body	System	Organ	Tissue	Cell
Purposes	Homeostasis (Maintenance of Internal Environment)	Adequate Circulation, Blood Volume, Oxygenation, Ventilation	Adequate Organ Perfusion, Blood Flow	Adequate Tissue Oxygenation and Perfusion	Adequate Cellular Oxygenation and Perfusion
Balances	Balances: Mass and Energy Inflow, Storage, and Outflow *	System Balances: Mass and Energy Inflow, Storage, Outflow, and Transfer *	Organ Balances: Mass and Energy Inflow, Storage, Outflow, and Transfer *	Tissue Balances: Mass and Energy Inflow, Storage, Outflow, and Transfer *	Cellular Balances: Mass and Energy Inflow, Storage, Outflow, and Transfer *
Processes	Total Volume of Body Fluid, Temperature, Supply: O_2, Fluids, Nutrients, Sink: CO_2, Fluids, Wastes	Circulation, Oxygenation, Ventilation, Circulating Volume	Perfusion Pressure, Organ Blood Flow, Vascular Resistance	Tissue Oxygenation, Respiration, Metabolism	Cell Metabolism, Chemical Reaction, Binding, Inflow, Outflow
Physiology		System Function	Organ Function	Tissue Function	Cellular Function
Anatomy			Organ Anatomy	Tissue Anatomy	Cellular Anatomy

* Balances include: Water, Salt, Electrolytes, pH, O_2, CO_2

Hajdukiewicz et al. (2001)

Slide 3-16: WDM of: (b) the cardiovascular system

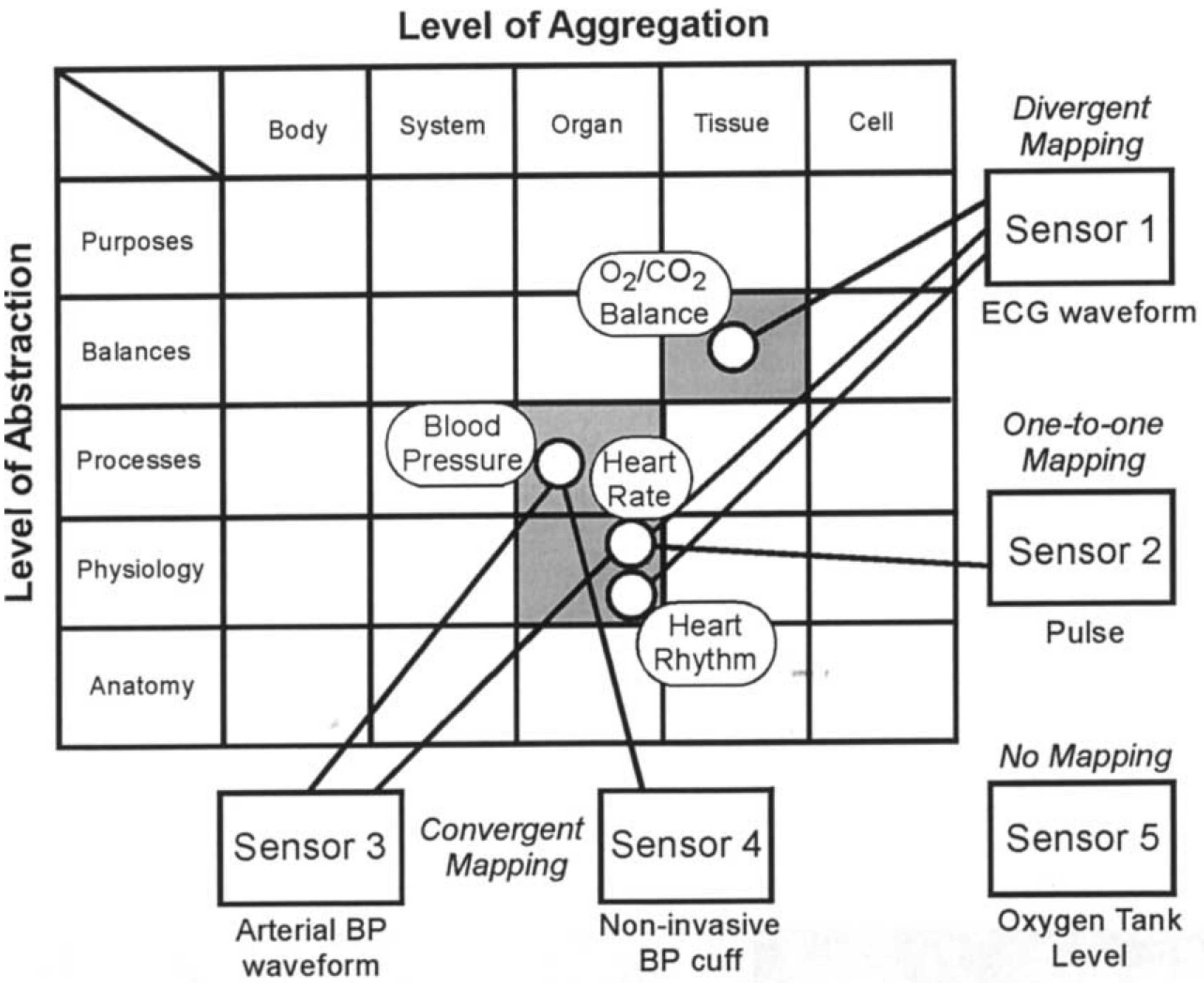
b)

Level of Abstraction

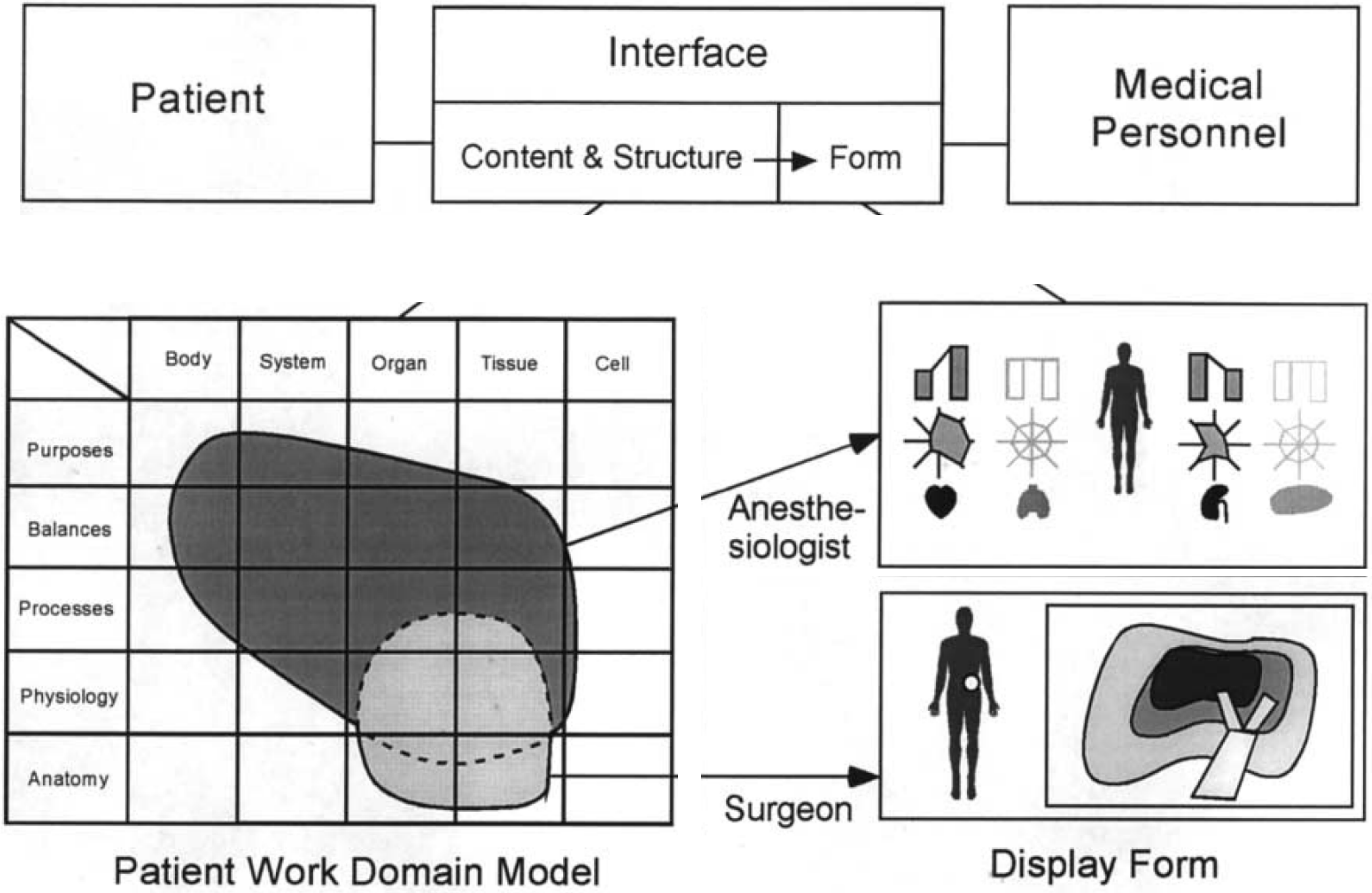
	System	Subsystem	Organ	Component
Purposes	Adequate Circulation and Blood Volume			
Balances	Cardiovascular System: Mass Inflow, Storage, and Outflow	Pulmonary and Systemic Systems: Balance Mass Flows; Mass Inflow, Storage, Outflow, and Transfer	Organ Vascular Network: Balance Mass Flows; Mass Inflow, Storage, Outflow , and Transfer	Vascular Components: Balance Mass Flows; Mass Inflow, Storage, Outflow, and Transfer
Processes	Circulation, Volume, Fluid Supply and Sink	Pulmonary and Systemic Circulation (Pressure, Flow, Resistance) and Volume, Fluid Supply and Sink	Cardiac Output, Organ Circulation (Pressure, Flow, Resistance), Fluid Supply and Sink from each Vascular Network	Circulation through Vascular Components (Pressure, Flow, Resistance), Vascular Blood Volume, Fluid Supply and Sink
Physiology	Cardiovascular System Function	Pulmonary and Systemic System Function	Cardiac Function (Heart Rate, Rhythm)	Atrial and Ventricular Function; Arterial, Arteriolar, Capillary, Venule, Venous Function
Anatomy			Cardiac Anatomy	Atrial, Ventricular, and Vascular Anatomy

Hajdukiewicz et al. (2001)

Slide 3-17: Example: Mapping OR sensors onto the WDM

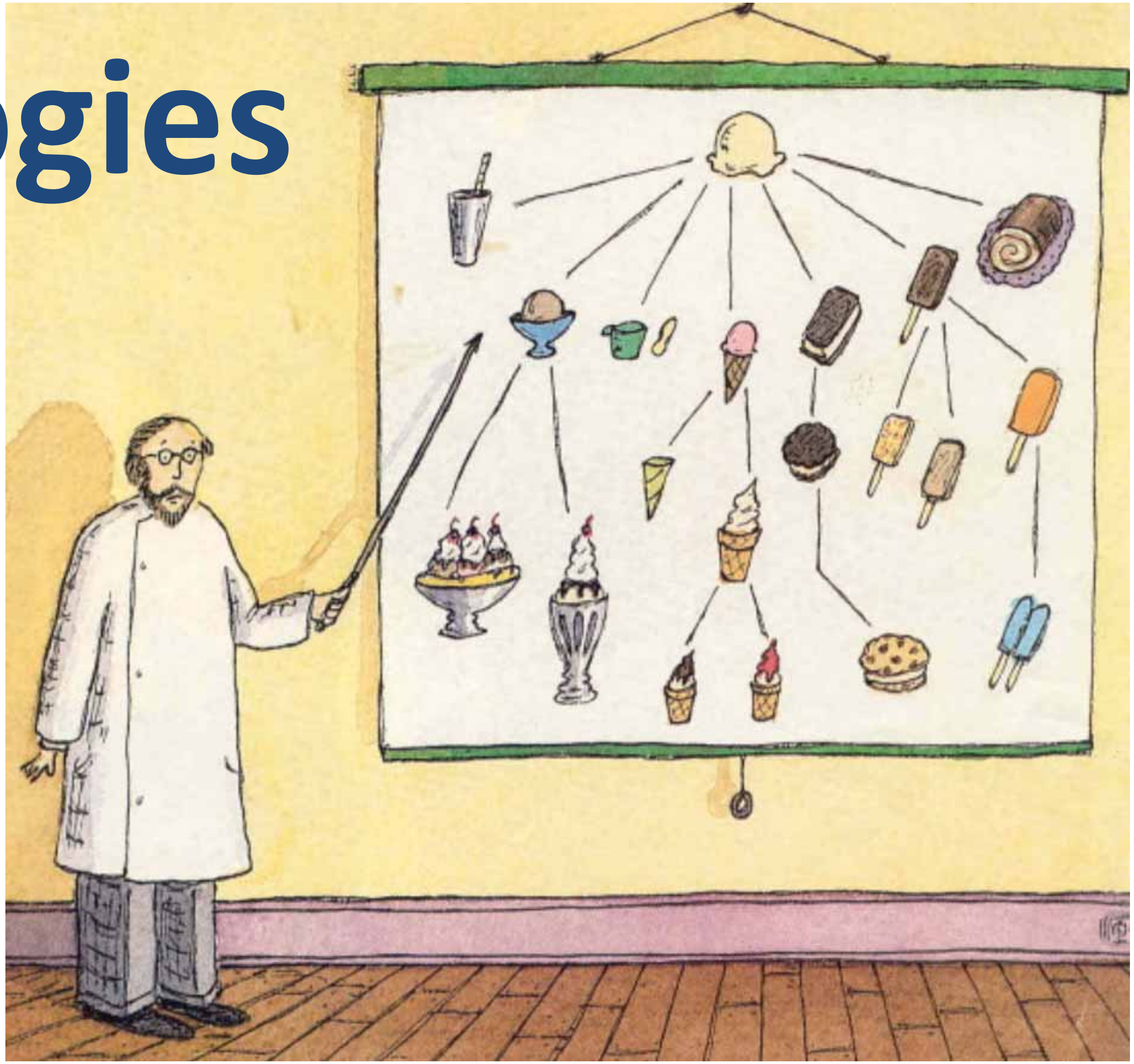


Hajdukiewicz et al. (2001)

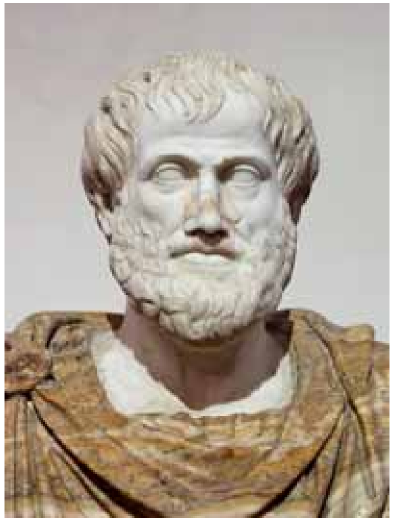


Hajdukiewicz et al. (2001)

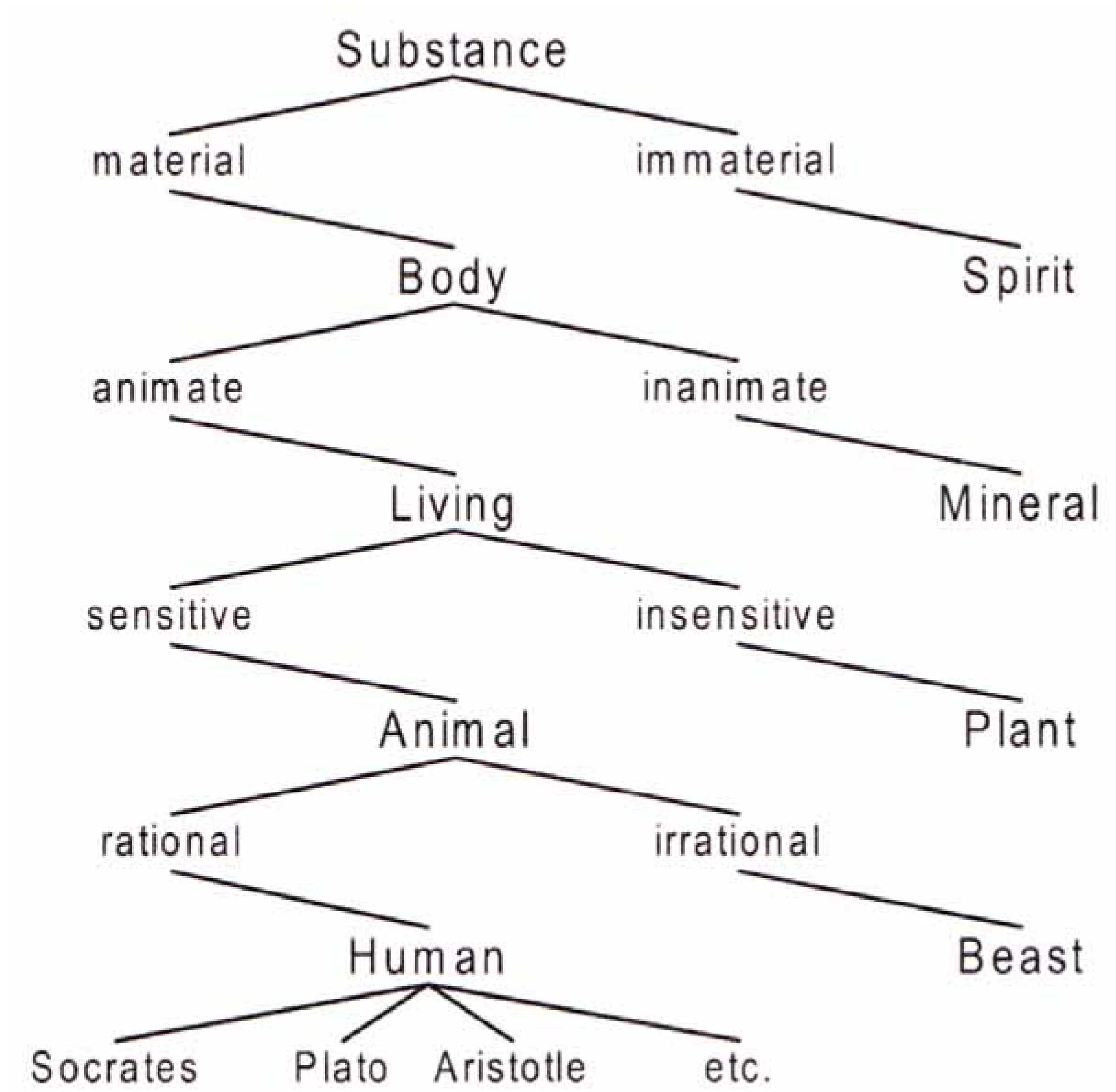
Ontologies



Slide 3-19: A simple question: What is a Jaguar?



* 384 BC † 322 BC



Simonet, M., Messai, R., Diallo, G. & Simonet, A. (2009) Ontologies in the Health Field. In: Berka, P., Rauch, J. & Zighed, D. A. (Eds.) *Data Mining and Medical Knowledge Management: Cases and Applications. New York, Medical Information Science Reference, 37-56.*

Later: Porphyry (≈ 234-305) tree

- Aristotle attempted to **classify the things in the world** - where it is employed to describe the existence of beings in the world;
- Artificial Intelligence and Knowledge Engineering deals also with **reasoning about models of the world**.
- Therefore, AI researchers adopted the term 'ontology' to describe **what can be computationally represented** of the world within a program.
- **“An ontology is a formal, explicit specification of a shared conceptualization”.**
  - A 'conceptualization' refers to an **abstract model** of some phenomenon in the world by having identified the relevant concepts of that phenomenon.
  - 'Explicit' means that the type of concepts used, and the constraints on their use are **explicitly defined.**

Studer, R., Benjamins, V. R. & Fensel, D. (1998) Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering, 25, 1-2, 161-197*.

- Ontology = a structured description of a domain in form of **concepts ↔ relations**;
- The **IS-A relation** provides a taxonomic skeleton;
- Other relations reflect the **domain semantics**;
- Formalizes the **terminology** in the domain;
- Terminology = terms definition and usage in the specific **context**;
- Knowledge base = **instance classification** and **concept classification**;
- Classification provides the **domain terminology**

…

Slide 3-23: Additionally an ontology may satisfy:

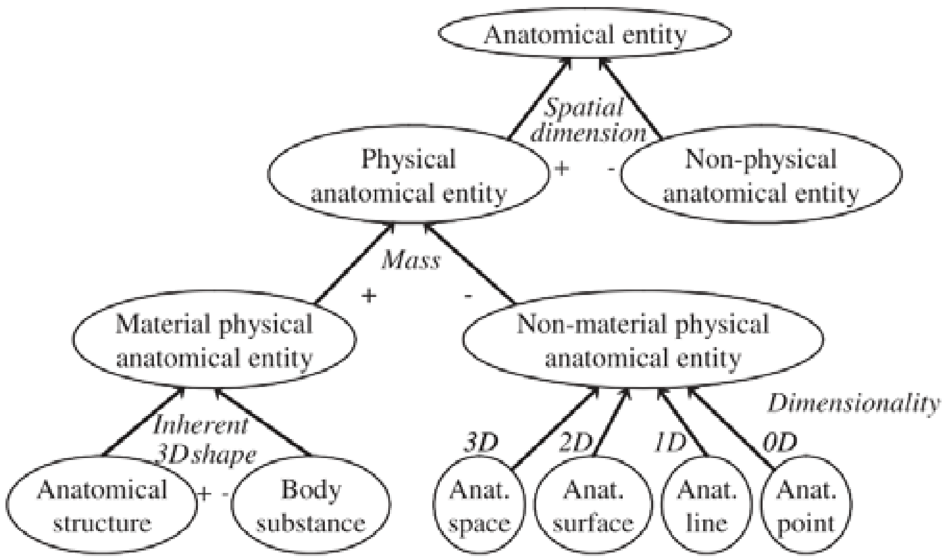
(1) In addition to the IS-A relationship, partitive (meronomic) relationships may hold between concepts, denoted by PART-OF. Every PART-OF relationship is irreflexive, asymmetric and transitive. IS-A and PART-OF are also called hierarchical relationships.

(2) In addition to hierarchical relationships, associative relationships may hold between concepts. Some associative relationships are domain-specific (e.g., the branching relationship between arteries in anatomy and rivers in geography).

(3) Relationships r and r' are inverses if, for every pair of concepts x and y, the relations $\langle x, r, y \rangle$ and $\langle y, r', x \rangle$ hold simultaneously. A symmetric relationship is its own inverse. Inverses of hierarchical relationships are called INVERSE-IS-A and HAS-PART, respectively.

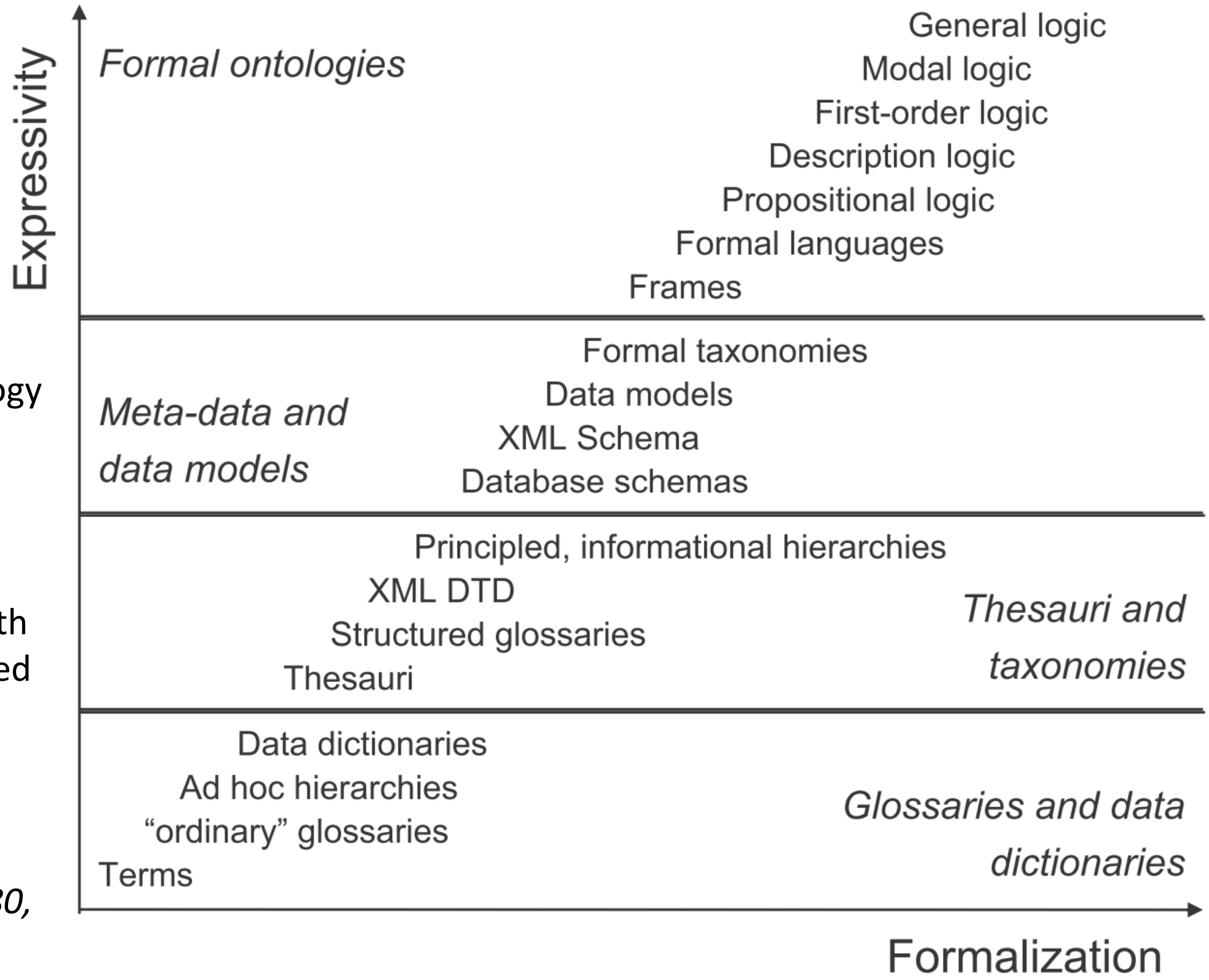
(4) Every non-taxonomic relation of x to z, $\langle x, r, z \rangle$, is either inherited ($\langle y, r, z \rangle$) or refined ($\langle y, r, z' \rangle$ where z' is more specific than z) by every child y of x. In other words, every child y of x has the same properties (z) as it parent or more specific properties (z').

Zhang, S. & Bodenreider, O. 2006. Law and order: Assessing and enforcing compliance with ontological modeling principles in the Foundational Model of Anatomy. *Computers in Biology and Medicine, 36, (7-8), 674-693.*

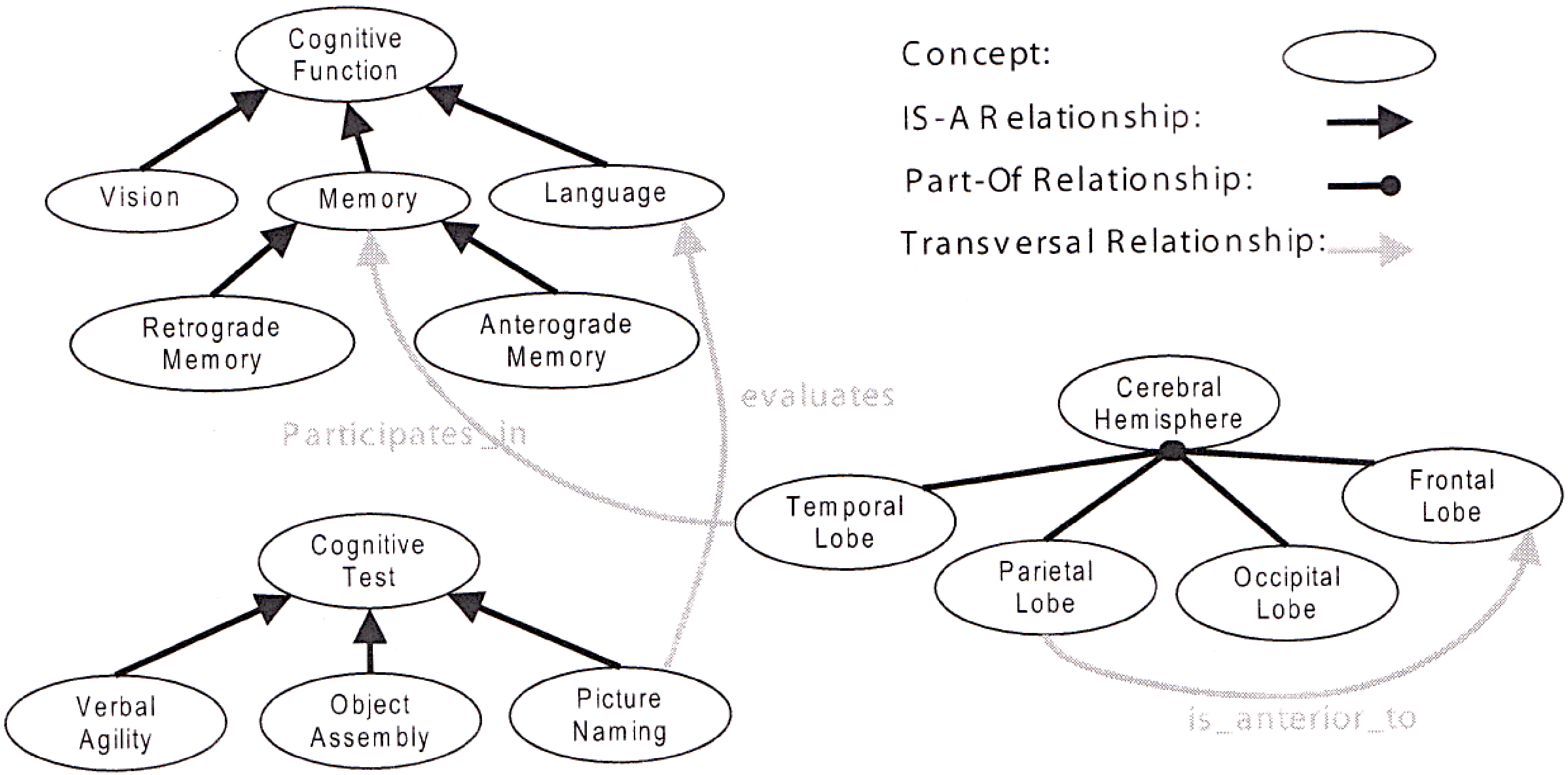


Slide 3-24: Ontologies: Taxonomy

Blobel, B. (2011) Ontology driven health information systems architectures enable pHealth for empowered patients. *International Journal of Medical Informatics, 80, 2, e17-e25.*



Slide 3-25 Example of a conceptual structure from CogSci



Simonet, M., Messai, R., Diallo, G. & Simonet, A. (2009) Ontologies in the Health Field. In: Berka, P., Rauch, J. & Zighed, D. A. (Eds.) *Data Mining and Medical Knowledge Management: Cases and Applications. New York, Medical Information Science Reference, 37-56.*

Name	Ref.	Scope	# concepts	# concept names				Subs. Hier.	Version / Notes
				Min	Max	Med	Avg		
SNOMED CT	[21]	Clinical medicine (patient records)	310,314	1	37	2	2.57	yes	July 31, 2007
LOINC	[24]	Clinical observations and laboratory tests	46,406	1	3	3	2.85	no	Version 2.21 (no "natural language" names)
FMA	[25]	Human anatomical structures	~72,000	1	?	?	~1.50	yes	(not yet in the UMLS)
Gene Ontology	[28]	Functional annotation of gene products	22,546	1	24	1	2.15	yes	Jan. 2, 2007
RxNorm	[31]	Standard names for prescription drugs	93,426	1	2	1	1.10	no	Aug. 31, 2007
NCI Thesaurus	[34]	Cancer research, clinical care, public information	58,868	1	100	2	2.68	yes	2007_05E
ICD-10	[36]	Diseases and conditions (health statistics)	12,318	1	1	1	1.00	no	1998 (tabular)
MeSH	[38]	Biomedicine (descriptors for indexing the literature)	24,767	1	208	5	7.47	no	Aug. 27, 2007
UMLS Meta.	[41]	Terminology integration in the life sciences	1,4 M	1	339	2	3.77	n/a	2007AC (English only)

Bodenreider, O. (2008) Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Methods of Information In Medicine, 47, Supplement 1, 67-79.*

- **1) Graph notations**
  - Semantic networks
  - Topic Maps (ISO/IEC 13250)
  - Unified Modeling Language (UML)
  - Resource Description Framework (RDF)
- **2) Logic based**
  - Description Logics (e.g., OIL, DAML+OIL, OWL)
  - Rules (e.g. RuleML, LP/Prolog)
  - First Order Logic (KIF – Knowledge Interchange Format)
  - Conceptual graphs
  - (Syntactically) higher order logics (e.g. LBase)
  - Non-classical logics (e.g. Flogic, Non-Mon, modalities)
- **3) Probabilistic/fuzzy**

Slide 3-28 Example for (1) Graphical Notation: RDF

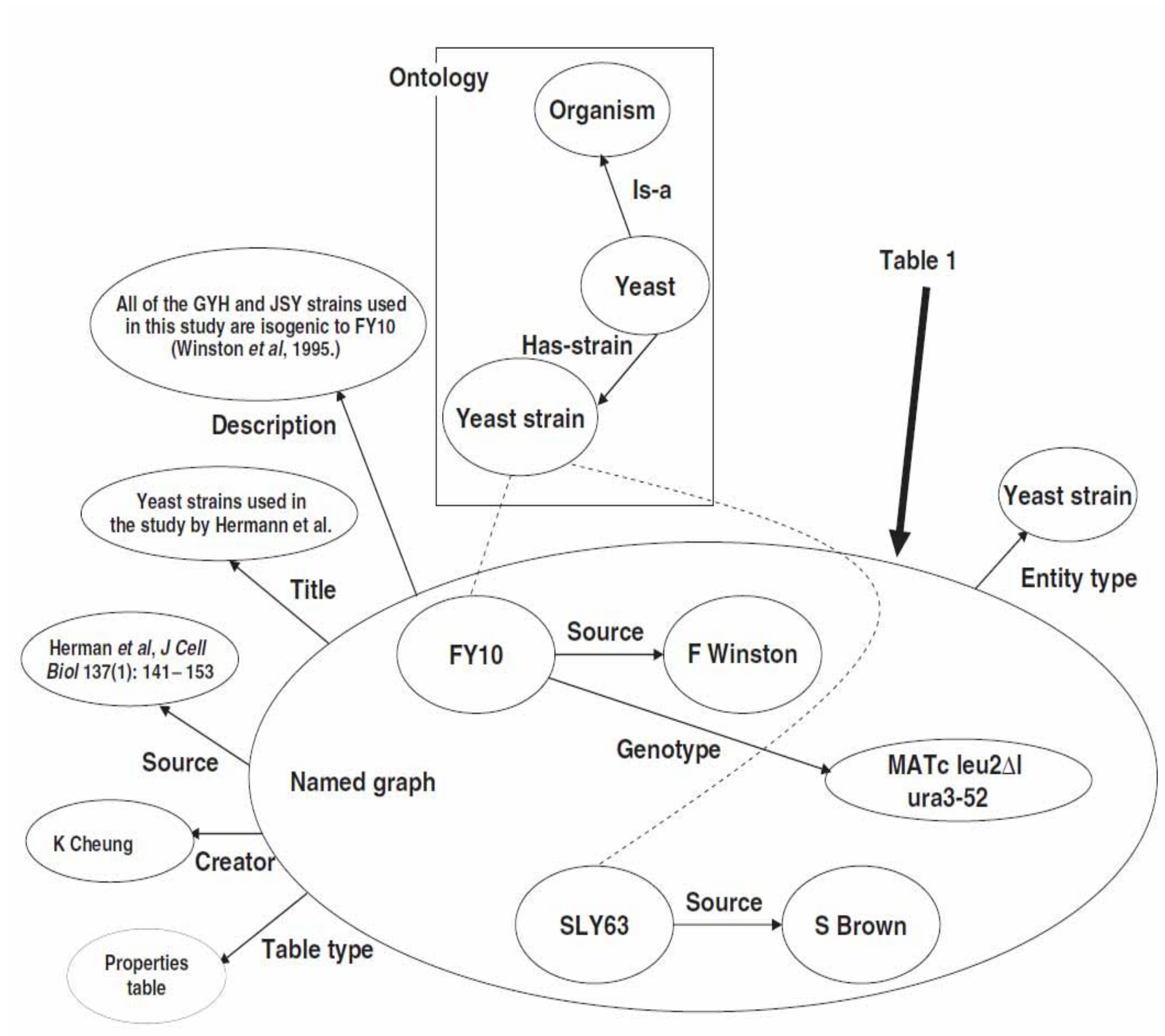


Table 1 Yeast strains used in the study by Hermann *et al* (1997)

Name	Genotype*	Source
FY10	*MATα leu2Δ1 ura3-52*	F Winston
FY22	*MATα his3Δ200 ura3-52*	F Winston
GHY1	*MATα leu2Δ1 his3Δ200 ura3-52 mdm20-1*	This study
JSY707	*MATα his3Δ200 ura3-52 tpm1D::HIS3*	This study
JSY948	*MATα leu2Δ1/leu2Δ1 ura3-52/ura3-52*	This study
JSY999	*MATα leu2Δ1 his3Δ200 ura3-52*	This study
JSY1065	*MATα leu2Δ1 his3Δ200 ura3-52 mdm20D::LEU2*	This study
JSY1084	*MATα leu2Δ1 his3Δ200 ura3-52 tpm1D::HIS3*	This study
JSY1138	*MATα leu2Δ1/leu2Δ1 his3Δ200/his3Δ 200 ura3-52/ura3-52 tpm1D::HIS3/ + mdm20D::LEU2/ +*	This study
JSY1285	*MATα leu2Δ1 his3Δ200 ura3-52 tpm2D::HIS3*	This study
JSY1340	*MATα leu2Δ1 his3Δ200 ura3-52 mdm20D::LEU2*	This study
JSY1374	*MATα leu2Δ1/leu2Δ1 his3Δ200/his3Δ200 ura3-52/ura3-52 tpm2D::HIS3/ + mdm20D::LEU2/ +*	This study
ABY1249	*MATα leu2-3,112 ura3-52 lys2-801 ade2-101 ade3 bem2-10*	A Bretscher
IGY4	*MATα leu2-3,112 his3Δ200 ura3-52 lys2-801 ade2 sac6D::LEU2*	A Adams
SLY63	*MATα leu2-3,112 ura3-52 trp1-1 his6 myo2-66*	S Brown

Cheung, K.-H., Samwald, M., Auerbach, R. K. & Gerstein, M. B. 2010. Structured digital tables on the Semantic Web: toward a structured digital literature. *Molecular Systems Biology, 6, 403.*

Slide 3-29: Example for (2) Web Ontology Language OWL

DL = Description Logic

Concept inclusion,
Speak: All C1 are C2

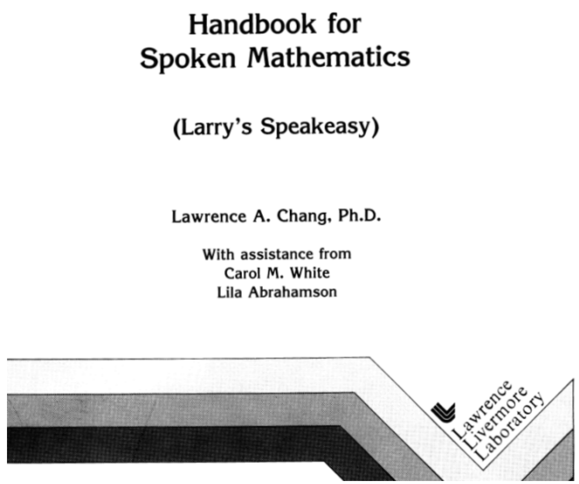
Concept equivalence
Speak: C1 is equivalent to C2

Axiom	DL syntax	Example
Sub class	$C_1 \sqsubseteq C_2$	Alga $\sqsubseteq$ Plant $\sqsubseteq$ Organism
Equivalent class	$C_1 \equiv C_2$	Cancer $\equiv$ Neoplastic Process
Disjoint with	$C_1 \sqsubseteq \neg C_2$	Vertebrate $\sqsubseteq$ $\neg$Invertebrate
Same individual	$x_1 \equiv x_2$	Blue_Shark $\equiv$ Prionace_Glauca
Different from	$x_1 \sqsubseteq \neg x_2$	Sea Horse $\sqsubseteq$ $\neg$Horse
Sub property	$P_1 \sqsubseteq P_2$	has_mother $\sqsubseteq$ has_parent
Equivalent property	$P_1 \equiv P_2$	treated_by $\equiv$ cured_by
Inverse	$P_1 \equiv P_2{}^-$	location_of $\equiv$ has_location$^-$
Transitive property	$P^+ \sqsubseteq P$	part_of$^+$ $\sqsubseteq$ part_of
Functional property	$\top \sqsubseteq \leq 1P$	$\top \sqsubseteq \leq 1$has_tributary
Inverse functional property	$\top \sqsubseteq \leq 1P^-$	$\top \sqsubseteq \leq 1$has_scientific_name$^-$

Bhatt, M., Rahayu, W., Soni, S. P. & Wouters, C. (2009) Ontology driven semantic profiling and retrieval in medical information systems. *Web Semantics: Science, Services and Agents on the World Wide Web, 7, 4, 317-331.*

Helpful: Handbook for Spoken Mathematics

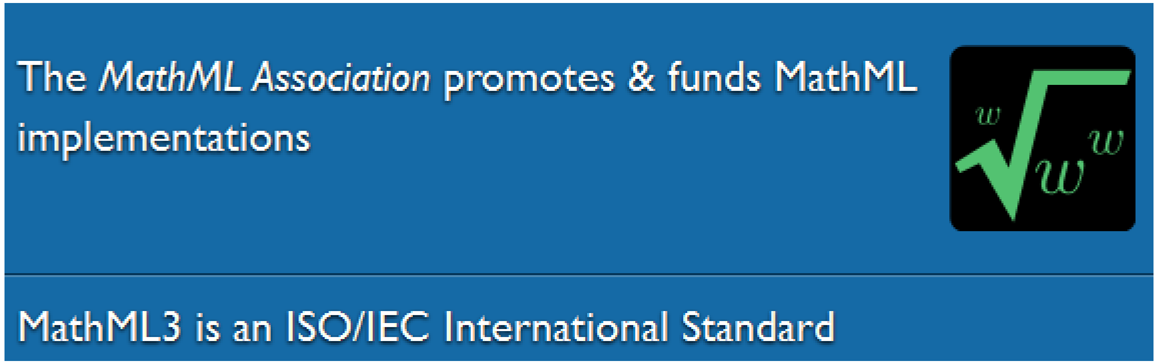
web.efzg.hr/dok/MAT/vkojic/Larrys_speakeasy.pdf



HELPFUL: https://en.wikipedia.org/wiki/List_of_mathematical_symbols

LaTeX Symbols : http://www.artofproblemsolving.com/wiki/index.php/LaTeX:Symbols

Math ML: http://www.robinlionheart.com/stds/html4/entities-mathml



Slide 3-30: OWL class constructors

Intersection/conjunction of concepts,
Speak: C1 and ... Cn

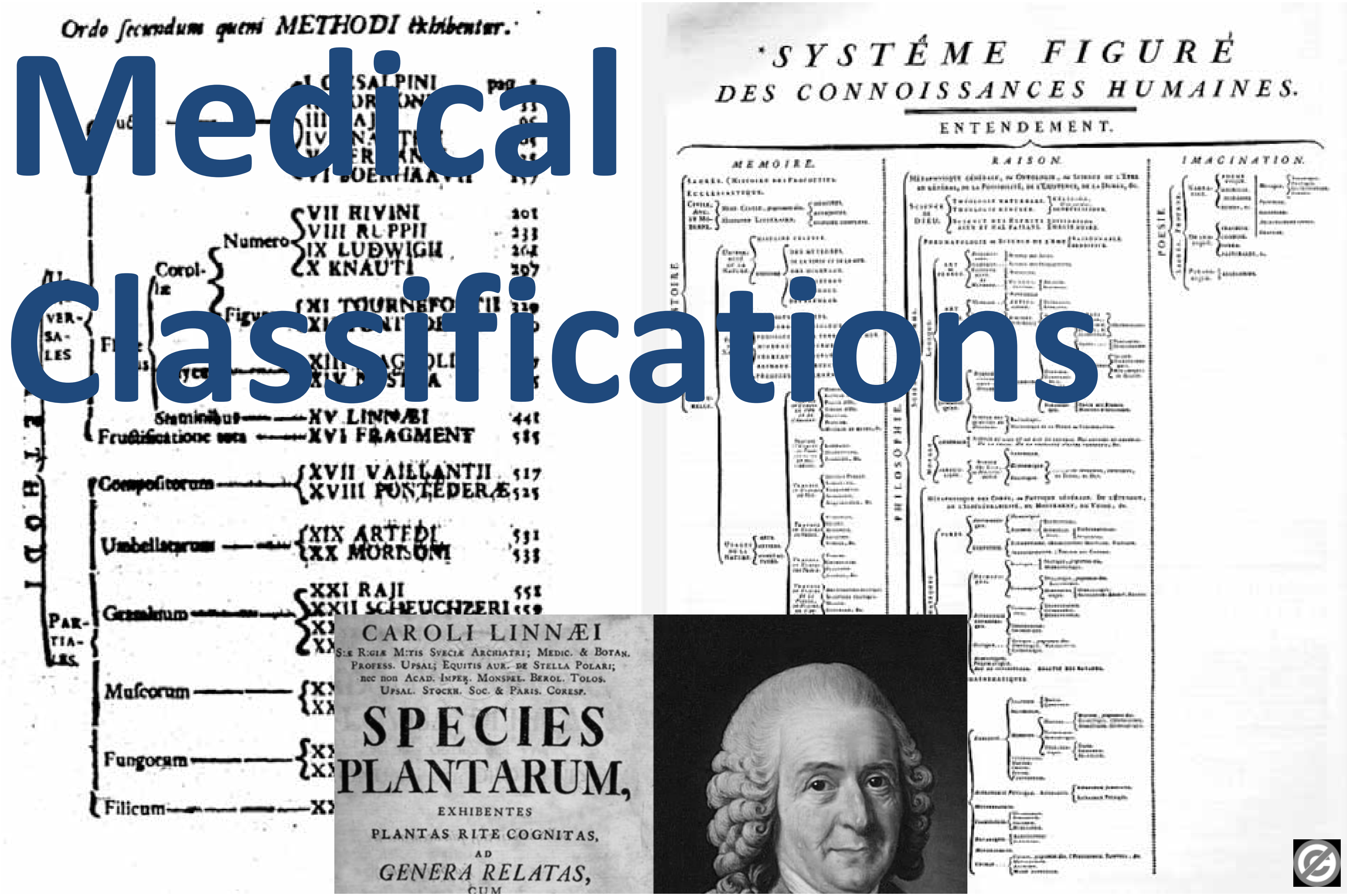
Constructor	DL syntax	Example
Intersection	$C_1 \sqcap \ldots \sqcap C_n$	Anatomical_Abnormality ⊓ Pathological_Function
Union	$C_1 \sqcup \ldots \sqcup C_n$	Body_Substance ⊔ Organic_Chemical
Complement	$\neg C$	¬Invertebrate
One of	$x_1 \sqcup \ldots \sqcup x_n$	Oestrogen ⊔ Progesterone
All values from	$\forall P.C$	∀co_occurs_with.Plant
Some values	$\exists P.C$	∃co_occurs_with.Animal
Max cardinality	$\leq nP$	≤ 1has_ingredient
Min cardinality	$\geq nP$	≥ 2has_ingredient

Universal Restriction
Speak: All P-successors are in C

Existential Restriction
Speak: An P-successor exists in C

Bhatt et al. (2009)

Medical Classifications



- Since the classification by Carl von Linne (1735) approx. 100+ various classifications in use:
  - **I**nternational **C**lassification of **D**iseases (ICD)
  - **S**ystematized **No**menclature of **Med**icine (SNOMED)
  - **Me**dical **S**ubject **H**eadings (MeSH)
  - **F**oundational **M**odel of **A**natomy (FMA)
  - **G**ene **O**ntology (GO)
  - **U**nified **M**edical **L**anguage **S**ystem (UMLS)
  - **L**ogical **O**bservation **I**dentifiers **N**ames & **C**odes (LOINC)
  - **N**ational **C**ancer **I**nstitute Thesaurus (NCI Thesaurus)

Slide 3-32: International Classification of Diseases (ICD)

World Health Organization

Health topics | Data and statistics | Media centre | Publications | Countries | Programmes and projects | Abo

Search

Classifications

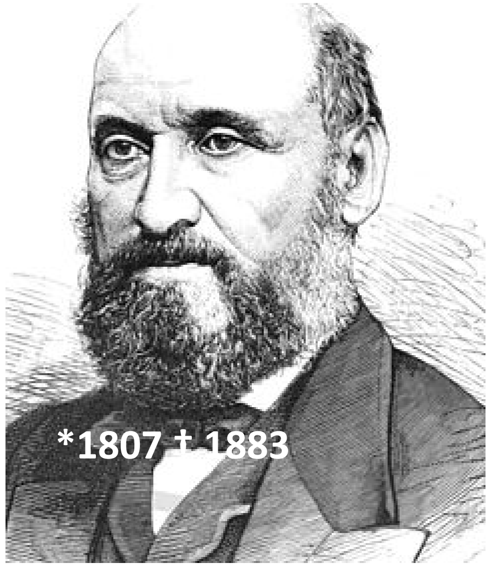
- Family of International Classifications
- Family of International Classifications network
- Classification of Diseases (ICD)
- Classification of Functioning, Disability and Health (ICF)
- Classification of Health Interventions (ICHI)
- Frequently asked questions

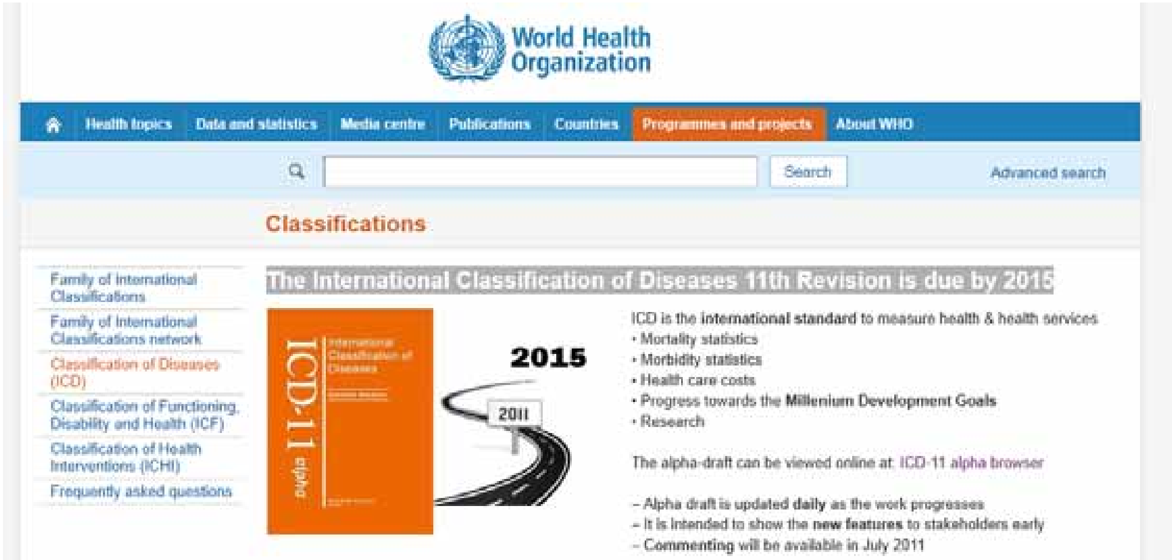
International Classification of Diseases (ICD)

ICD-10 was endorsed by the Forty-third World Health Assembly in May 1990 and came into use in WHO Member States as from 1994. The classification is the latest in a series which has its origins in the 1850s. The first edition, known as the International List of Causes of Death, was adopted by the International Statistical Institute in 1893. WHO took over the responsibility for the ICD at its creation in 1948 when the Sixth Revision, which included causes of morbidity for the first time, was published. The World Health Assembly adopted in 1967 the WHO Nomenclature Regulations that stipulate use of ICD in its most current revision for mortality and morbidity statistics by all Member States.

http://www.who.int/classifications/icd/en

- 1629 London Bills of Mortality
- 1855 **William Farr** (London, one founder of medical statistics): List of causes of death, list of diseases
- 1893 von Jacques Bertillot: List of causes of death
- 1900 International Statistical Institute (ISI) accepts Bertillot’s list
- 1938 5th Edition
- 1948 WHO
- 1965 ICD-8
- 1989 ICD-10
- 2015 ICD-11 due
- 2018 ICD-11 adopt





- 1965 SNOP, 1974 SNOMED, 1979 SNOMED II
- 1997 (Logical Observation Identifiers Names and Codes (LOINC) integrated into SNOMED
- 2000 SNOMED RT, 2002 SNOMED CT

INTERNATIONAL HEALTH TERMINOLOGY
STANDARDS DEVELOPMENT ORGANISATION

239 pages

SNOMED CT® Technical Reference Guide

January 2011 International Release
(US English)

http://www.isb.nhs.uk/documents/isb-0034/amd-26-2006/techrefguid.pdf

Slide 3-35: SNOMED Example Hypertension

A

24184005|Finding of increased blood pressure (finding) ➔
38936003|Abnormal blood pressure (finding) AND
roleGroup SOME
(363714003|Interprets (attribute) SOME
75367002|Blood pressure (observable entity))

B

12763006|Finding of decreased blood pressure (finding)➔
392570002|Blood pressure finding (finding) AND
roleGroup SOME
(363714003|Interprets (attribute) SOME
75367002|Blood pressure (observable entity))

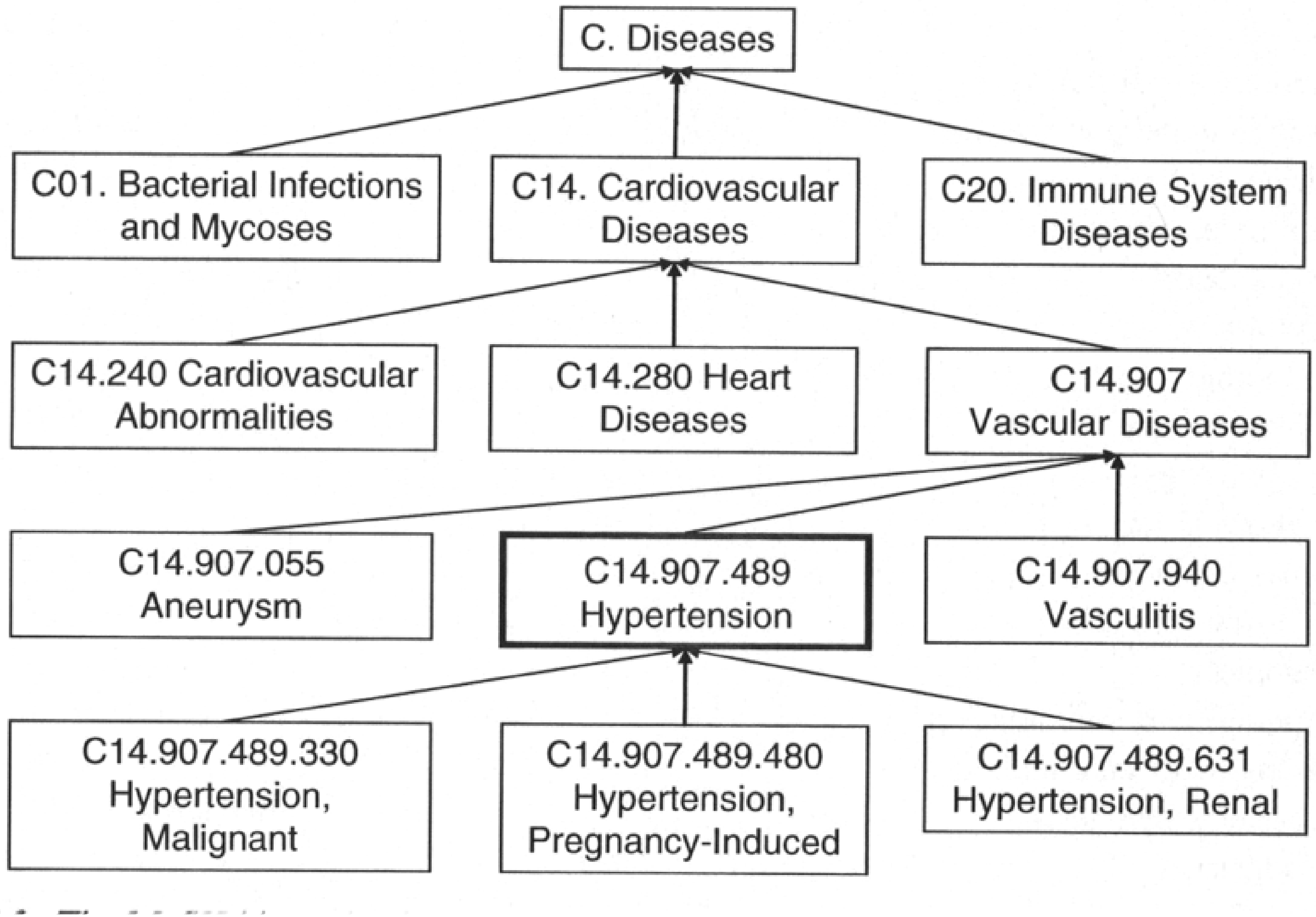
Rector, A. L. & Brandt, S. (2008) Why Do It the Hard Way? The Case for an Expressive Description Logic for SNOMED. *Journal of the American Medical Informatics Association, 15, 6, 744-751.*

- MeSH thesaurus is produced by the National Library of Medicine (NLM) since 1960.
- Used for cataloging documents and related media and as an index to search these documents in a database and is part of the metathesaurus of the Unified Medical Language System (UMLS).
- This thesaurus originates from keyword lists of the Index Medicus (today Medline);
- MeSH thesaurus is polyhierarchic, i.e. every concept can occur multiple times. It consists of the three parts:
  - 1. MeSH Tree Structures,
  - 2. MeSH Annotated Alphabetic List and
  - 3. Permuted MeSH.

Slide 3-37: The 16 trees in MeSH

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

Slide 3-38: MeSH Hierarchy: e.g. heading Hypertension 1/2



Hersh, W. (2010) *Information Retrieval: A Health and Biomedical Perspective. New York, Springer.*

Slide 3-39: MeSH Example Hypertension 2/2

National Library of Medicine - Medical Subject Headings

2011 MeSH

MeSH Descriptor Data

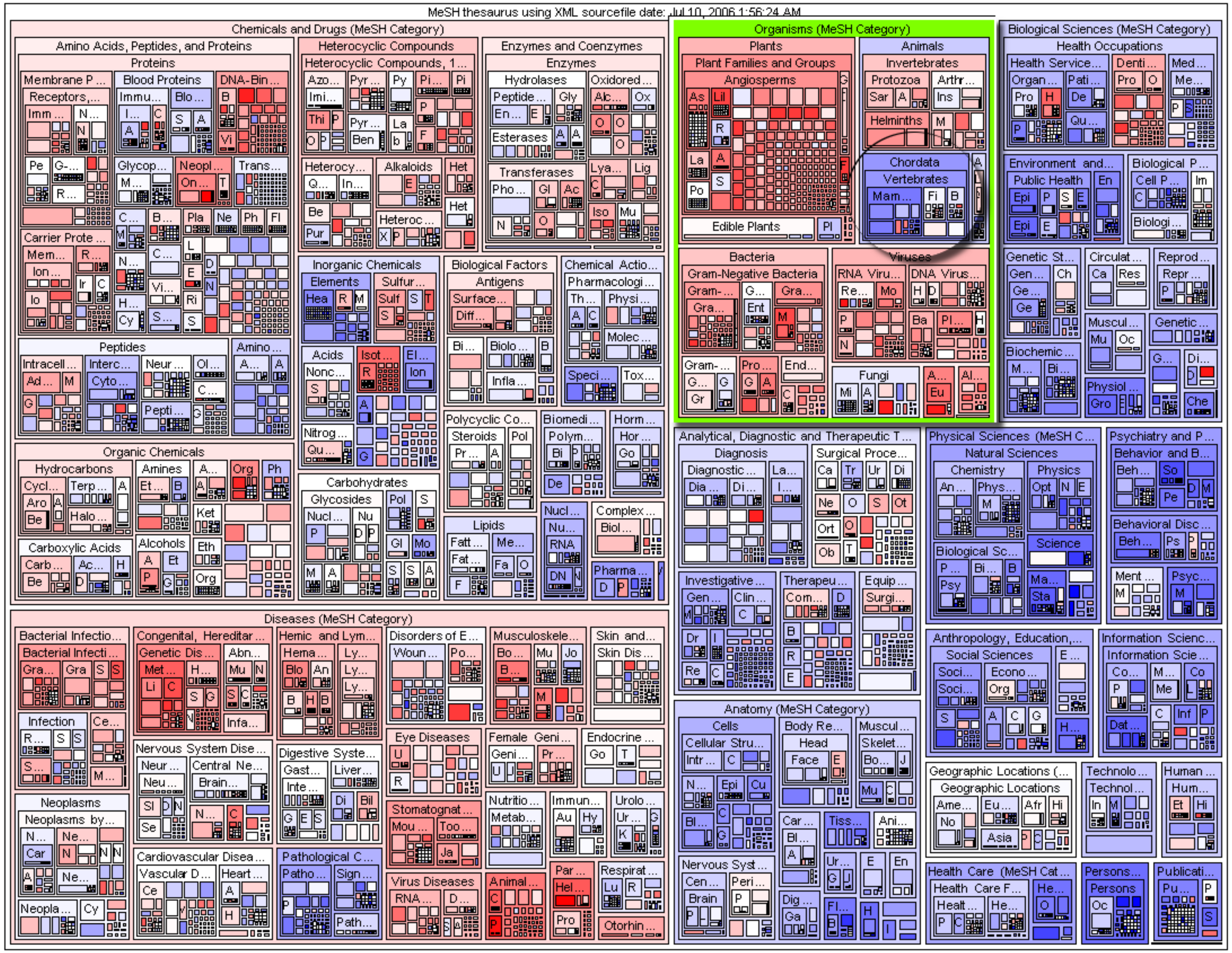
Return to Entry Page

Standard View. Go to Concept View; Go to Expanded Concept View

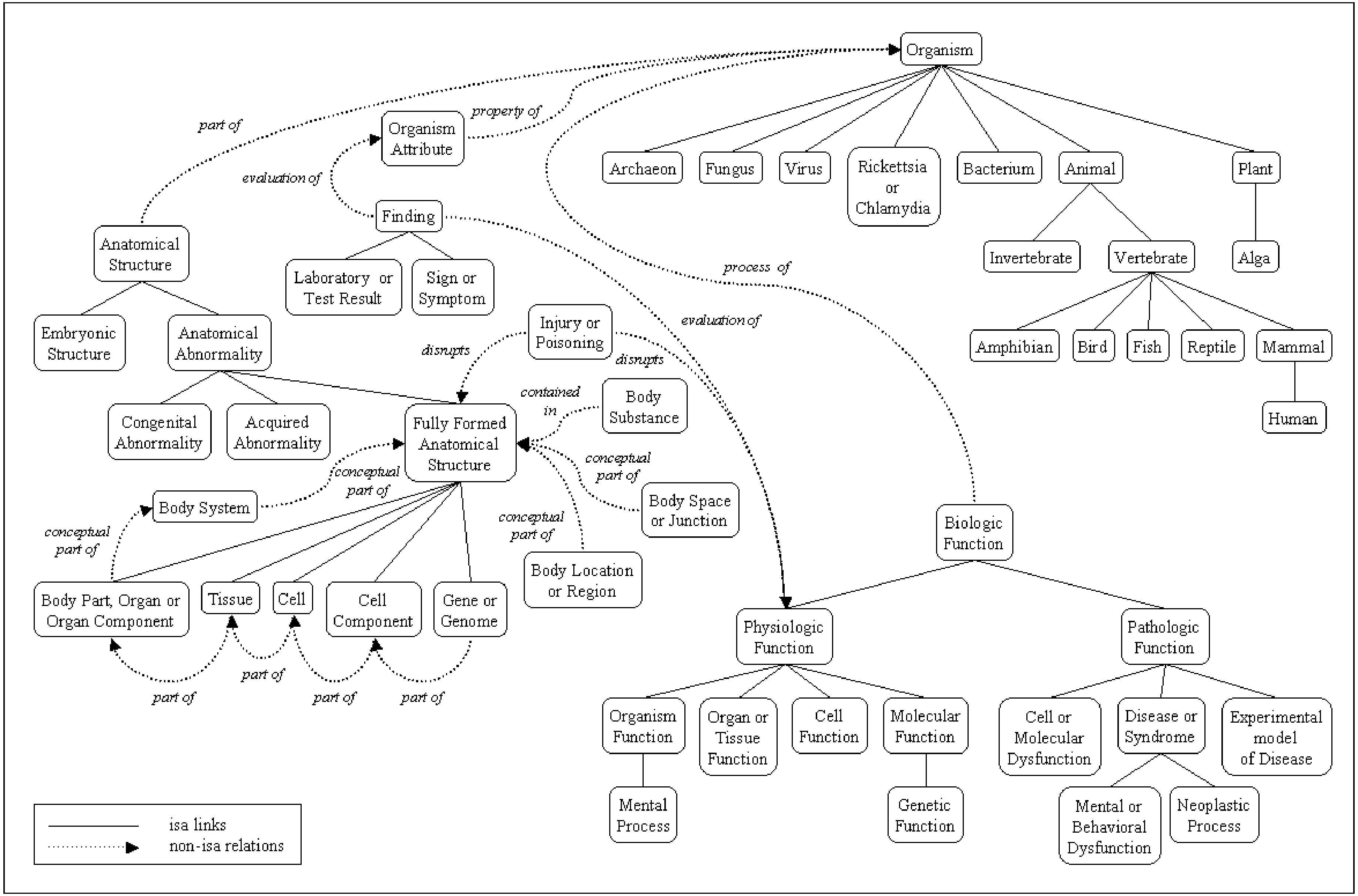
MeSH Heading	Hypertension
Tree Number	C14.907.489
Annotation	not for intracranial or intraocular pressure; relation to BLOOD PRESSURE: Manual 23.27; Goldblatt kidney is HYPERTENSION, GOLDBLATT see HYPERTENSION, RENOVASCULAR; hypertension with kidney disease is probably HYPERTENSION, RENAL, not HYPERTENSION; venous hypertension: index under VENOUS PRESSURE (IM) & do not coordinate with HYPERTENSION; PREHYPERTENSION is also available
Scope Note	Persistently high systemic arterial BLOOD PRESSURE. Based on multiple readings (BLOOD PRESSURE DETERMINATION), hypertension is currently defined as when SYSTOLIC PRESSURE is consistently greater than 140 mm Hg or when DIASTOLIC PRESSURE is consistently 90 mm Hg or more.
Entry Term	Blood Pressure, High
See Also	Antihypertensive Agents
See Also	Vascular Resistance
Allowable Qualifiers	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UR US VE VI
Date of Entry	19990101
Unique ID	D006973

http://www.nlm.nih.gov/mesh/

Slide 3-40: MeSH Interactive Tree-Map Visualization (see L 9)

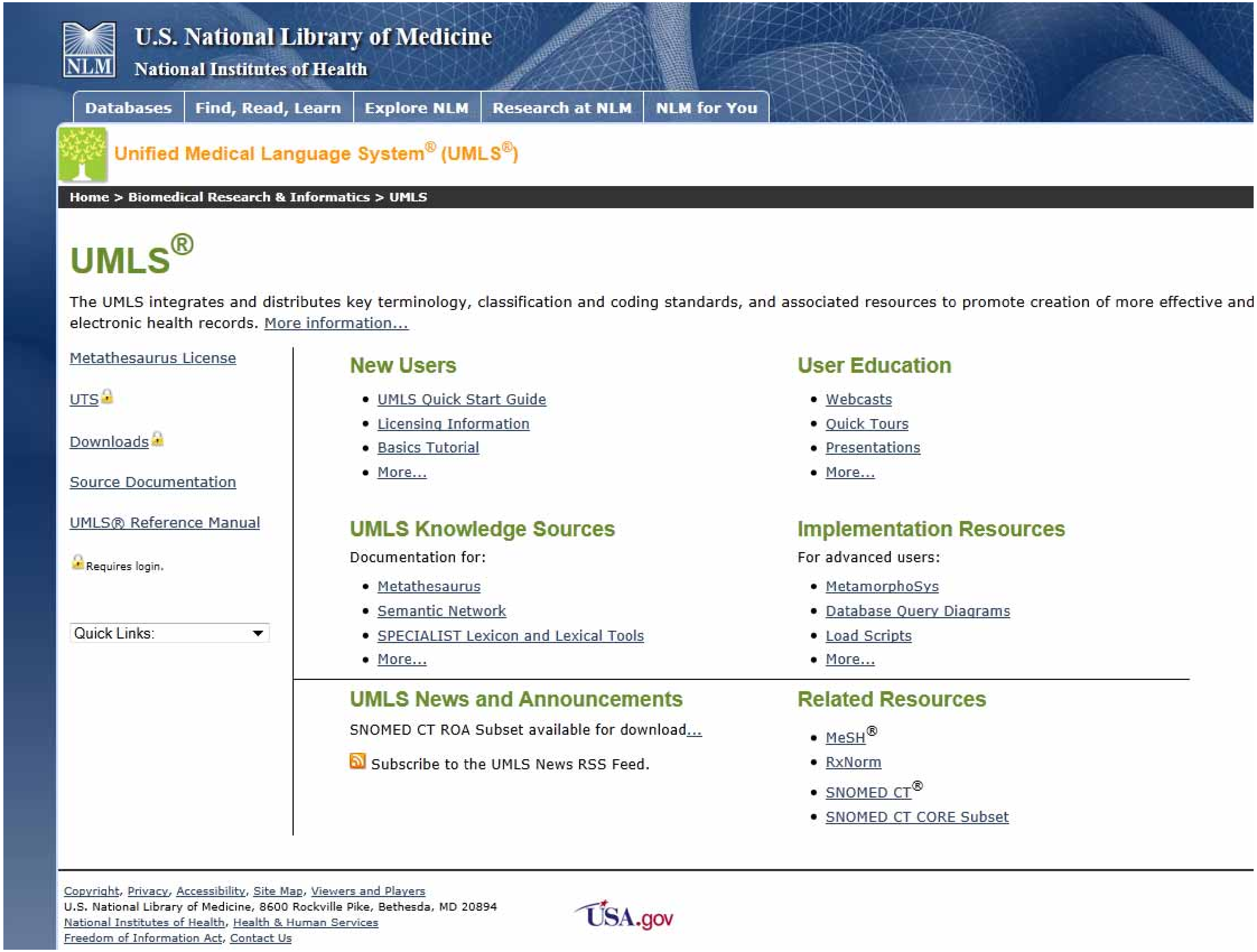


Eckert, K. (2008) A methodology for supervised automatic document annotation. *Bulletin of IEEE Technical Committee on Digital Libraries TCDL, 4, 2.*



Organism
part of
property of
Organism Attribute
evaluation of
Finding
Archaeon
Fungus
Virus
Rickettsia or Chlamydia
Bacterium
Animal
Plant
Anatomical Structure
Laboratory or Test Result
Sign or Symptom
process of
Invertebrate
Vertebrate
Alga
Embryonic Structure
Anatomical Abnormality
Injury or Poisoning
evaluation of
disrupts
disrupts
Amphibian
Bird
Fish
Reptile
Mammal
contained in
Body Substance
Human
Congenital Abnormality
Acquired Abnormality
Fully Formed Anatomical Structure
conceptual part of
conceptual part of
Body Space or Junction
Body System
conceptual part of
Biologic Function
conceptual part of
Body Location or Region
Body Part, Organ or Organ Component
Tissue
Cell
Cell Component
Gene or Genome
Physiologic Function
Pathologic Function
part of
part of
part of
part of
Organism Function
Organ or Tissue Function
Cell Function
Molecular Function
Cell or Molecular Dysfunction
Disease or Syndrome
Experimental model of Disease
Mental Process
Genetic Function
Mental or Behavioral Dysfunction
Neoplastic Process
isa links
non-isa relations

Slide 3-42: http://www.nlm.nih.gov/research/umls/



U.S. National Library of Medicine
National Institutes of Health

Databases | Find, Read, Learn | Explore NLM | Research at NLM | NLM for You

Unified Medical Language System® (UMLS®)

Home > Biomedical Research & Informatics > UMLS

UMLS®

The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and electronic health records. More information...

Metathesaurus License

UTS

Downloads

Source Documentation

UMLS® Reference Manual

Requires login.

Quick Links:

New Users

- UMLS Quick Start Guide
- Licensing Information
- Basics Tutorial
- More...

User Education

- Webcasts
- Quick Tours
- Presentations
- More...

UMLS Knowledge Sources

Documentation for:

- Metathesaurus
- Semantic Network
- SPECIALIST Lexicon and Lexical Tools
- More...

Implementation Resources

For advanced users:

- MetamorphoSys
- Database Query Diagrams
- Load Scripts
- More...

UMLS News and Announcements

SNOMED CT ROA Subset available for download...

Subscribe to the UMLS News RSS Feed.

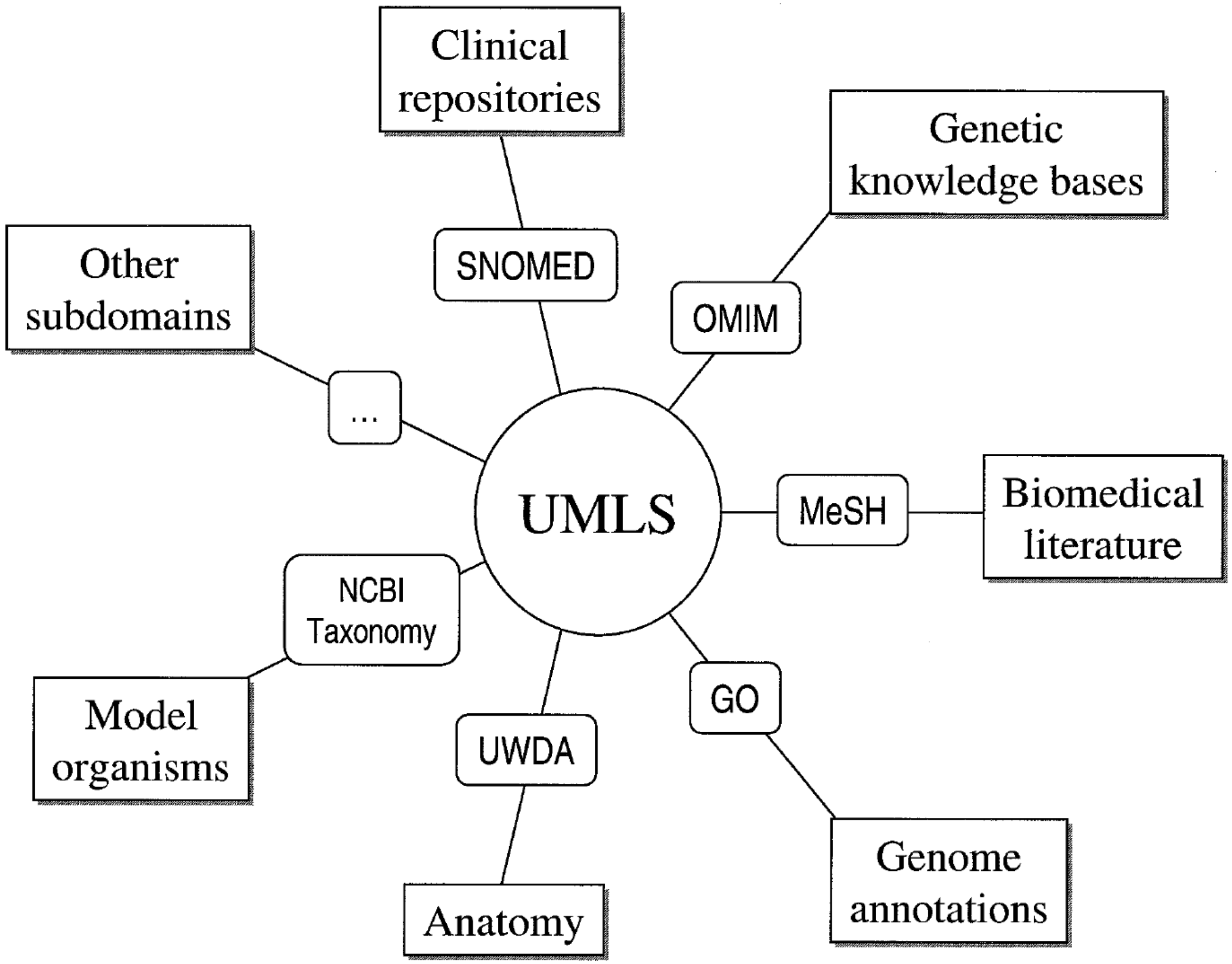
Related Resources

- MeSH®
- RxNorm
- SNOMED CT®
- SNOMED CT CORE Subset

Copyright, Privacy, Accessibility, Site Map, Viewers and Players
U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health, Health & Human Services
Freedom of Information Act, Contact Us

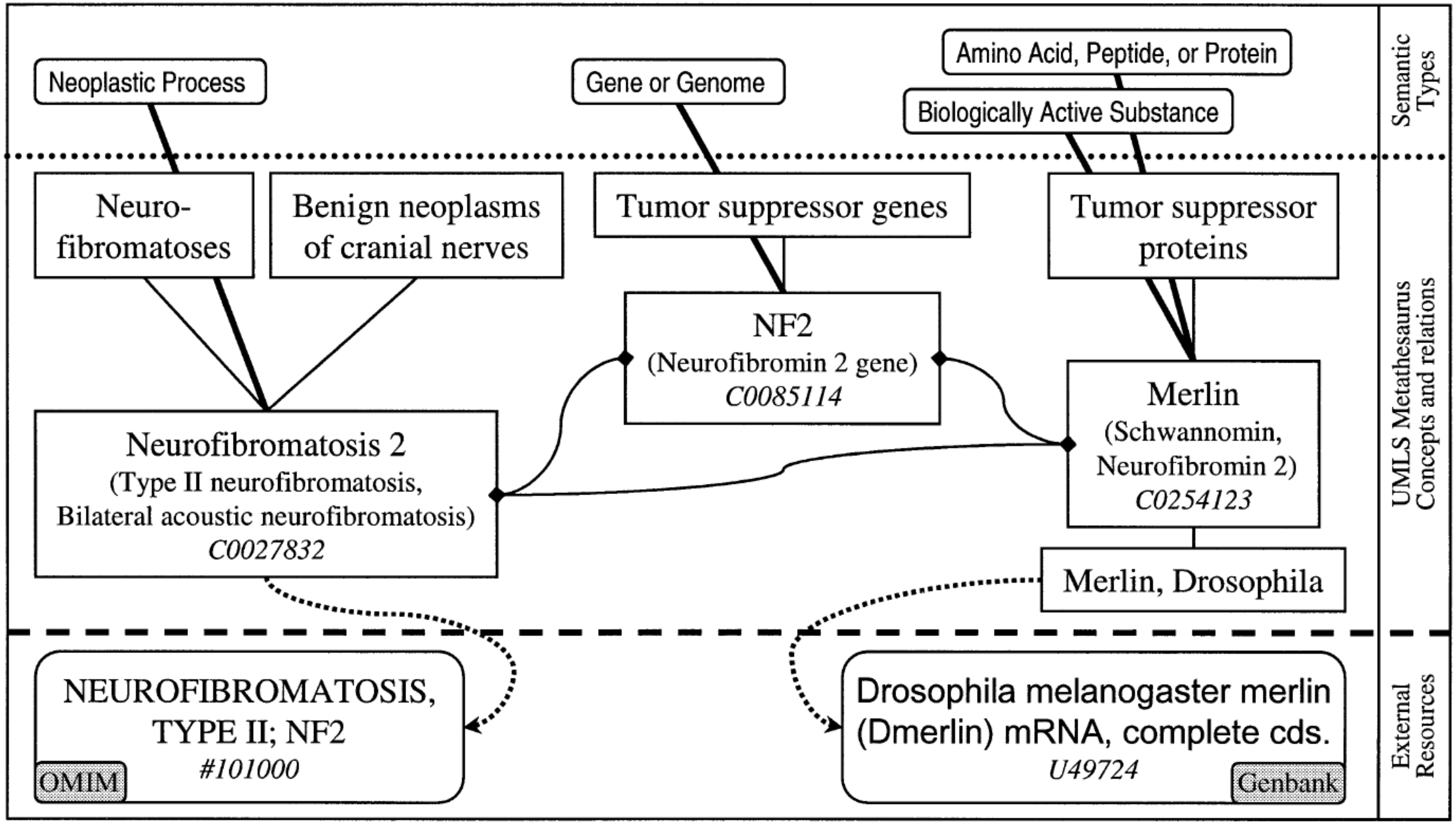
USA.gov

Slide 3-43: UMLS Metathesaurus integrates sub-domains



Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research, 32, D267-D270.*

Slide 3-44: Example of proteins and diseases in the UMLS



Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research, 32, D267-D270.*

- Data fusion – Data integration in the life sciences
- Self learning stochastic ontologies [1]
- Interactive, integrative machine learning and ontologies
- Never ending learning machines [2] for building knowledge spaces
- Integrating ontologies in daily work
- Knowledge and **context awareness**

[1] Ongenae, F., Claeys, M., Dupont, T., Kerckhove, W., Verhoeve, P., Dhaene, T. & De Turck, F. 2013. A probabilistic ontology-based platform for self-learning context-aware healthcare applications. Expert Systems with Applications, 40, (18), 7629-7646.

[2] Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E. R. & Mitchell, T. M. 2010. Toward an Architecture for Never-Ending Language Learning. Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10). Atlanta: AAAI. 1306-1313.

Thank you!

Sample Questions (1)

- What is the proportion of structured/standardized versus weakly structured/non-standardized data?
- What are the benefits of standardized data?
- Which problems are involved in dealing with medical data?
- What is still a remaining big problem in the health domain … even with standardized data?
- What constitutes data standardization?
- What is the most used standardized data set in medical informatics today?
- Which are the three predominant ECG data formats?
- What is the advantage/disadvantage between binary data and XML data?
- What is the purpose of modeling biomedical knowledge?
- Provide examples for various abstraction levels of a Work Domain Model!
- What can be done with a Work Domain Model?
- What is the origin of ontologies?
- Please provide the classic definition of an ontology!
- What does domain semantics mean?
- What constitutes the classification of an ontology?

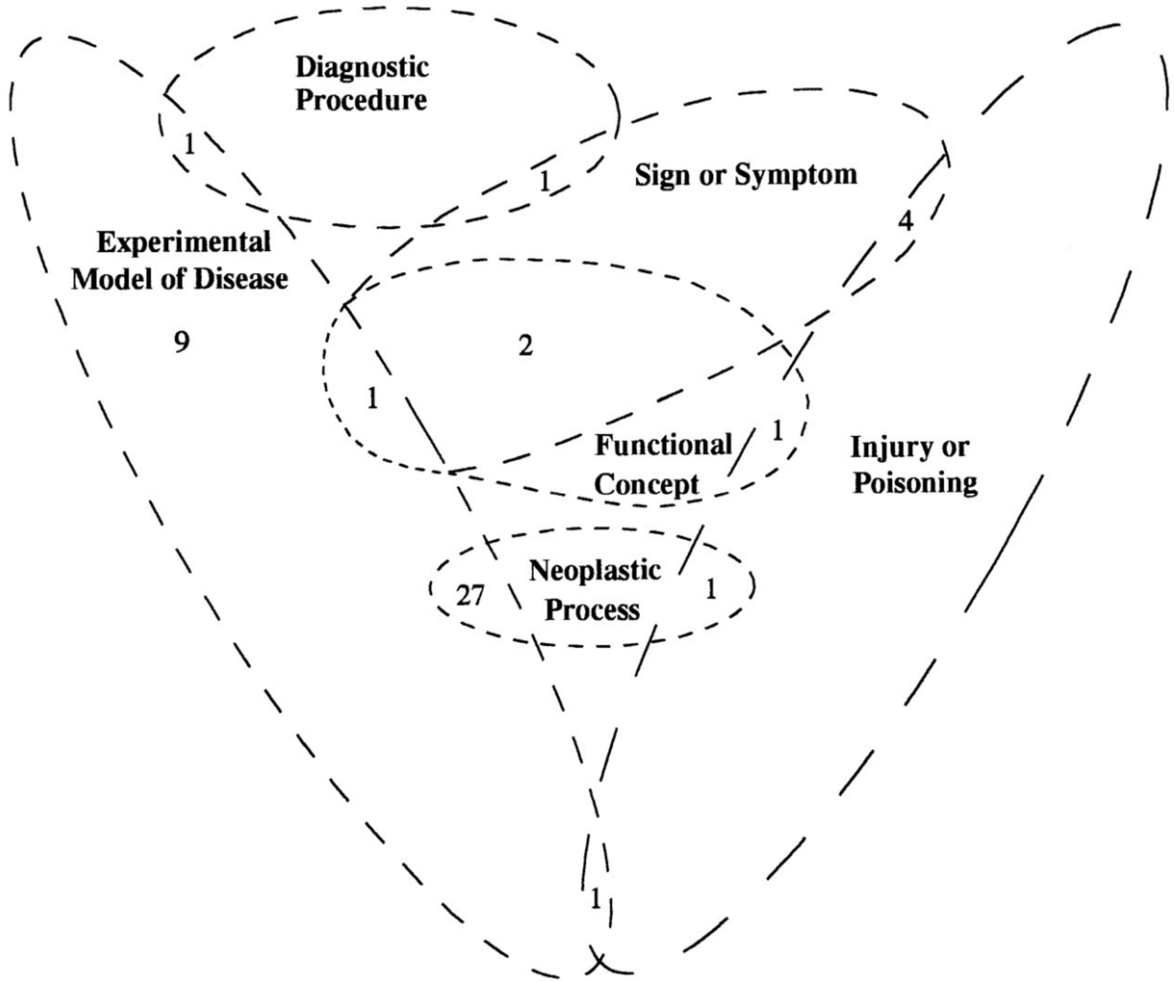
Sample Questions (2)

- Provide an overview about the most important biomedical ontologies!
- What are typical ontology languages?
- Please provide some examples of typical OWL axioms!
- What is a OWL class constructor?
- How do you start the development of an ontology?
- What are typical layers of abstraction – on the example of a Breast Cancer Imaging Ontology?
- What does “semantic enrichment” of a medical ontology mean?
- Within an ontology based architecture: what does the so called Knowledge Layer include?
- What are the roots of the ICD?
- What is the advantage of SNOMED-CT?
- What does polyhierachic thesaurus mean? Please provide an example for such a thesaurus!
- How can I expand queries with the MeSH Ontology?
- What is the major component of the UMLS?
- What is the main purpose of the Gene Ontology?

Some useful links

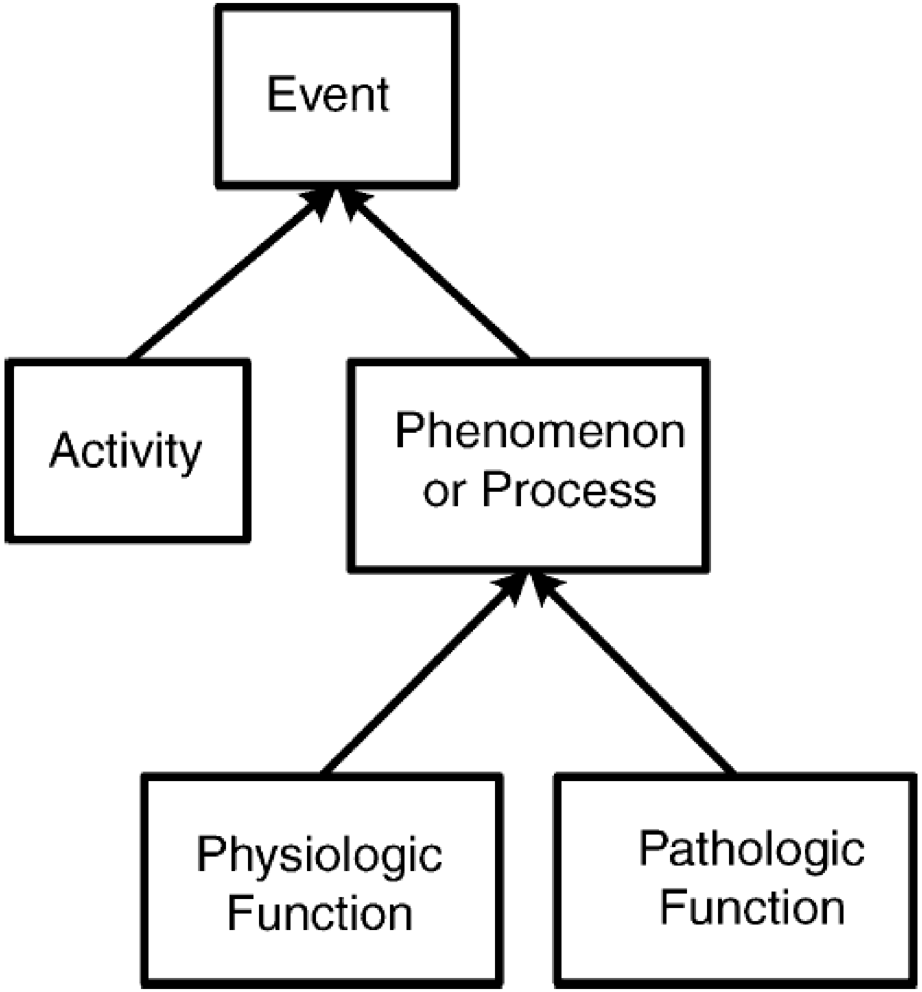
- http://wiki.hl7.org
- http://snomed.dataline.co.uk/
- https://github.com/drh-uth/MEDRank
- http://www.nlm.nih.gov/mesh/
- http://www.nlm.nih.gov/research/umls/
- http://www.geneontology.org/
- http://www.who.int/classifications/icd/en/

Backup-Slide: UMLS: Six semantic types and intersections



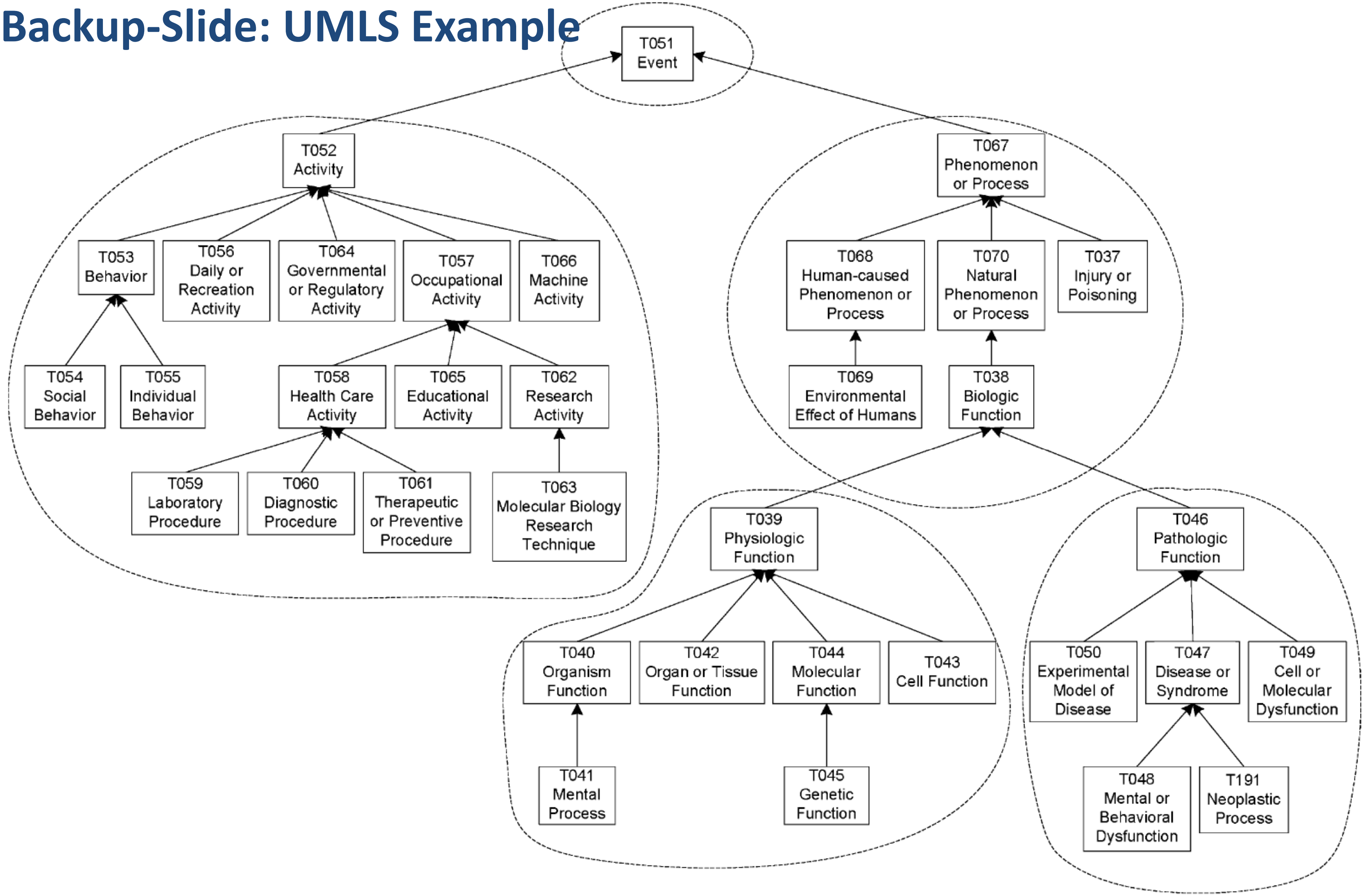
Gu, H., Perl, Y., Geller, J., Halper, M., Liu, L.-m. & Cimino, J. J. (2000) Representing the UMLS as an Object-oriented Database: Modeling Issues and Advantages. *Journal of the American Medical Informatics Association, 7, 1, 66-80.*

Backup-Slide: Metaschema hierarchy



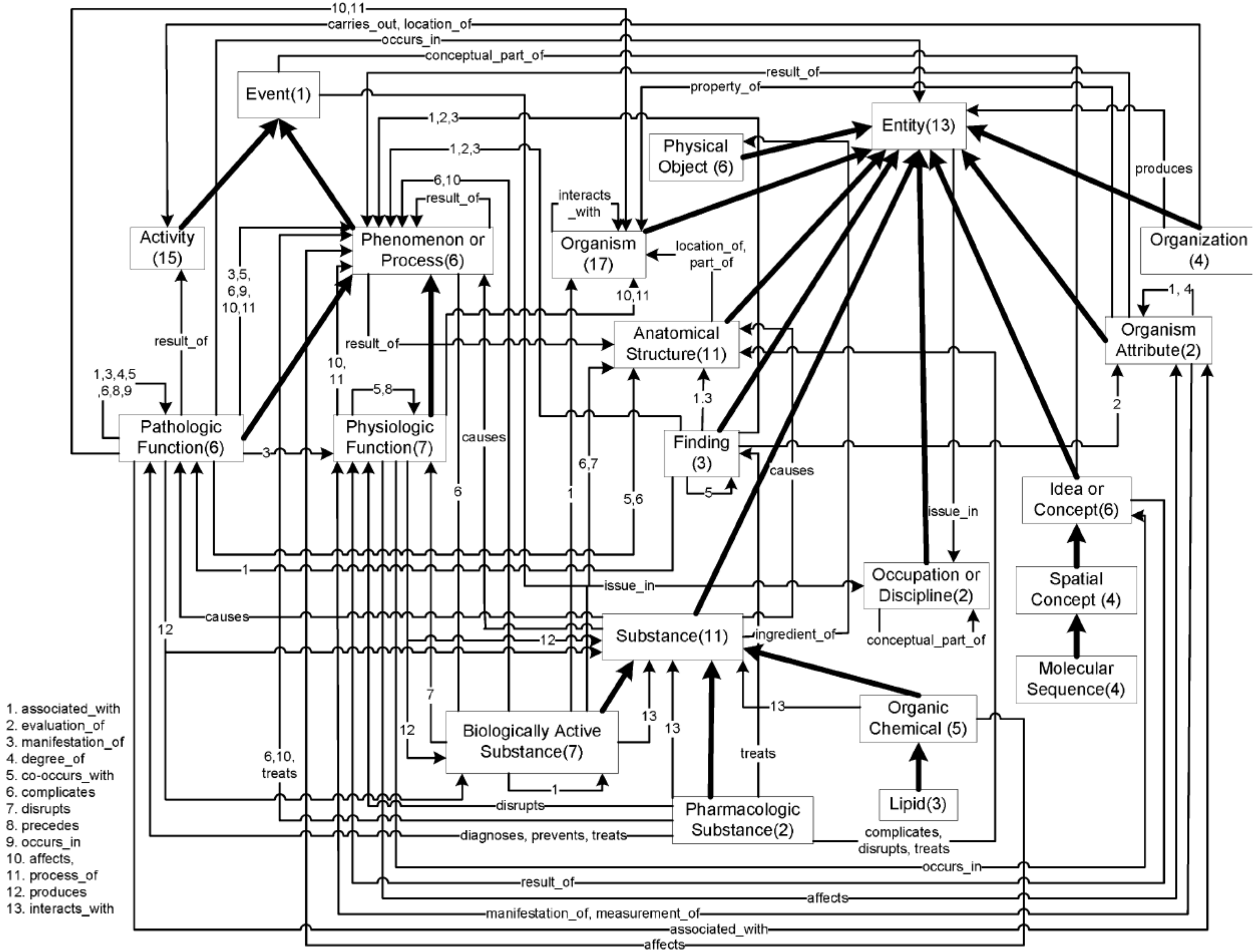
Zhang, L., Hripcsak, G., Perl, Y., Halper, M. & Geller, J. (2005) An expert study evaluating the UMLS lexical metaschema. *Artificial Intelligence in Medicine, 34, 3, 219-233.*

Backup-Slide: UMLS Example



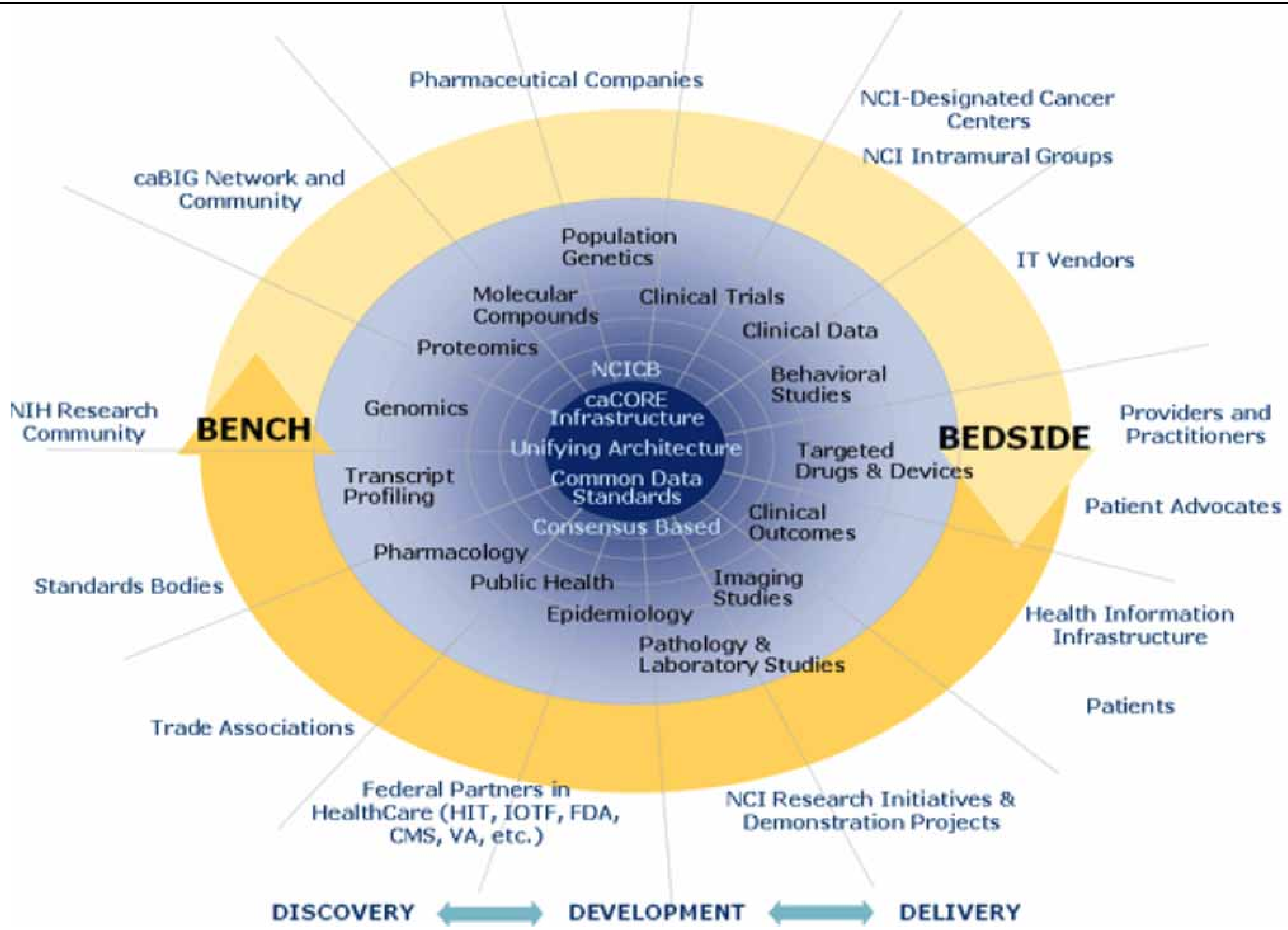
Zhang, L., Hripcsak, G., Perl, Y., Halper, M. & Geller, J. (2005) An expert study evaluating the UMLS lexical metaschema. *Artificial Intelligence in Medicine, 34, 3, 219-233.*

Backup-Slide: Lexical Metaschema

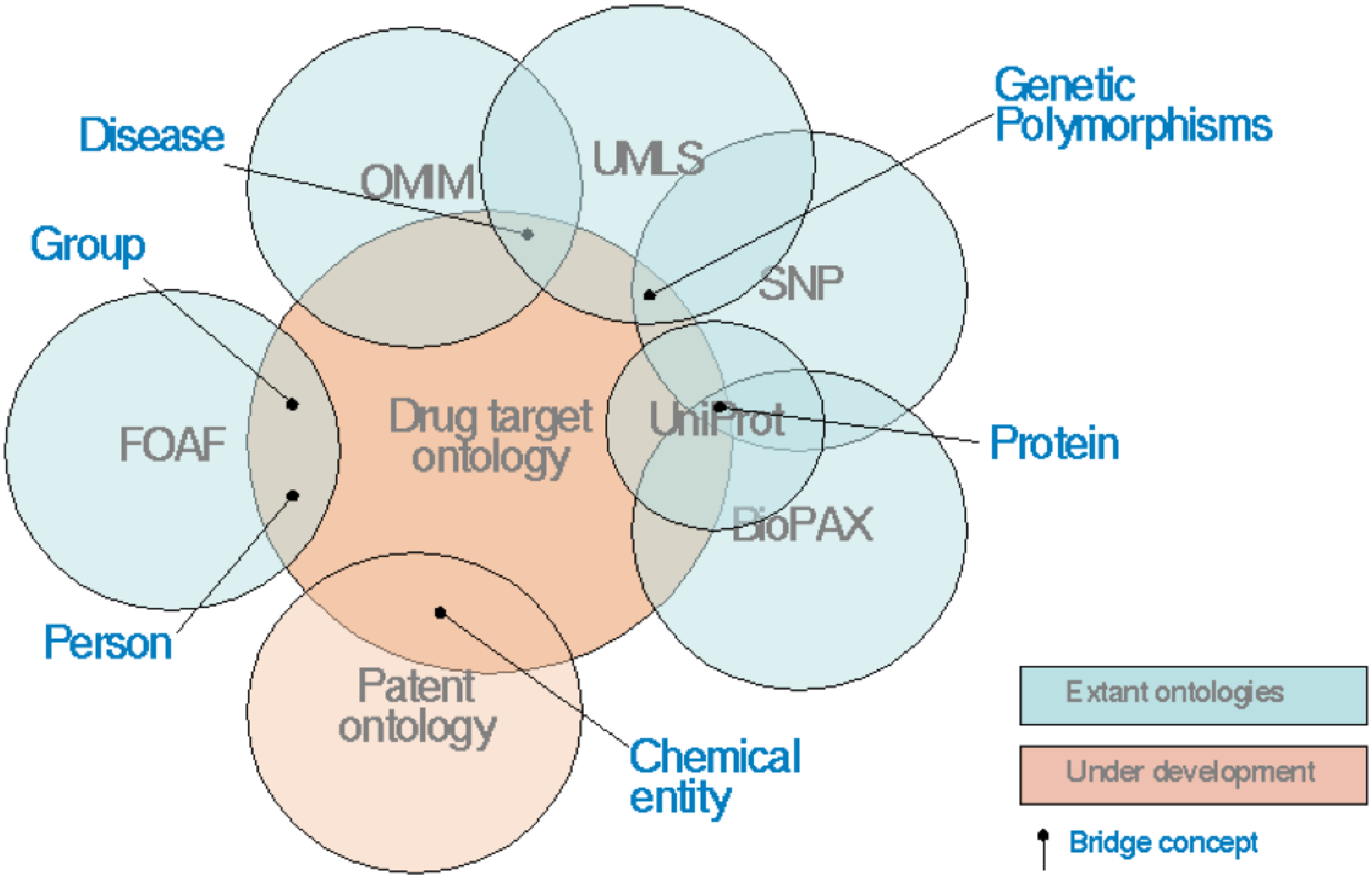


Zhang et al. (2005)

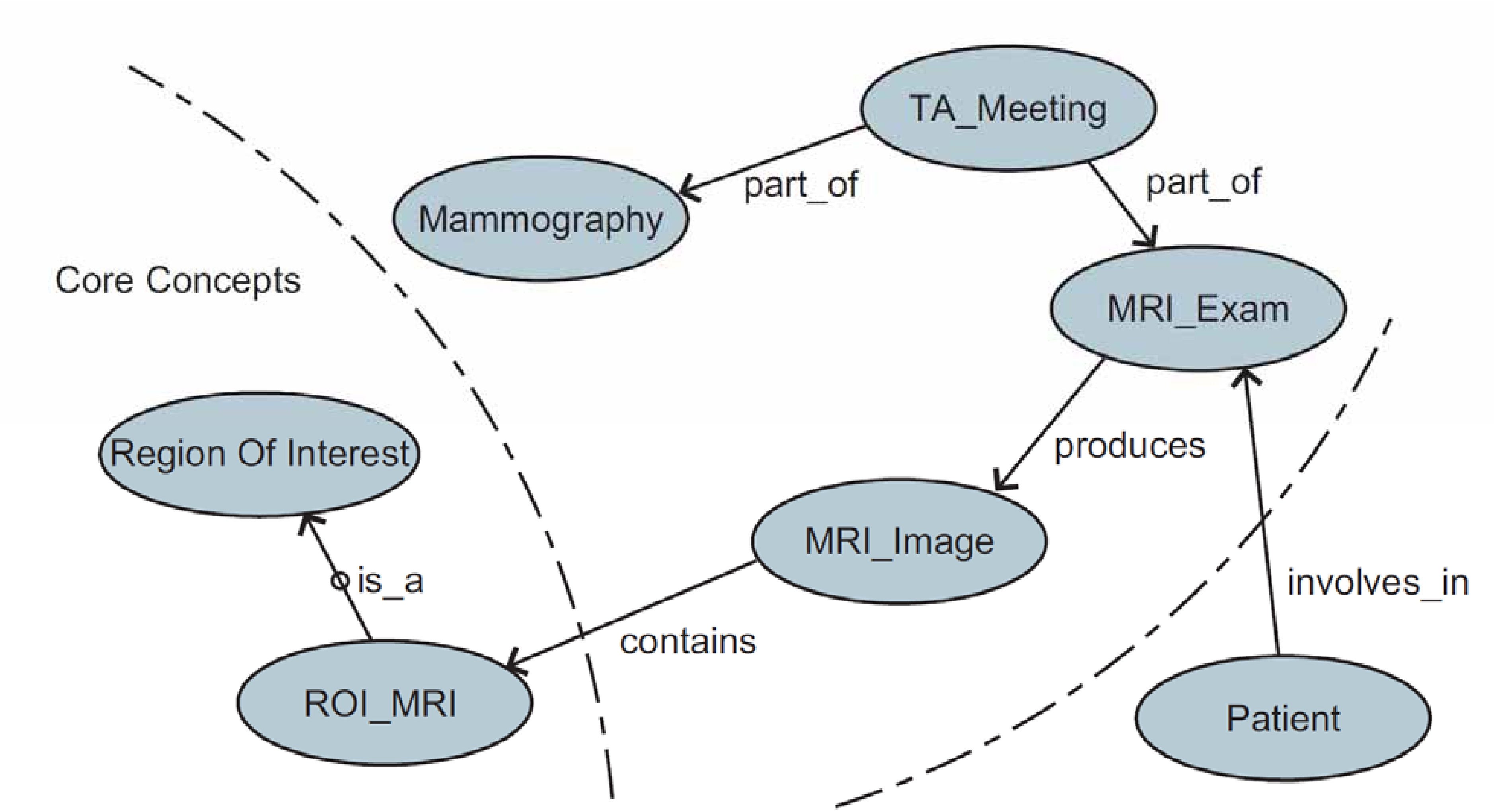
Backup Slide: National Cancer Institute Integration Effort



Backup Slide: Examples for well-known ontologies

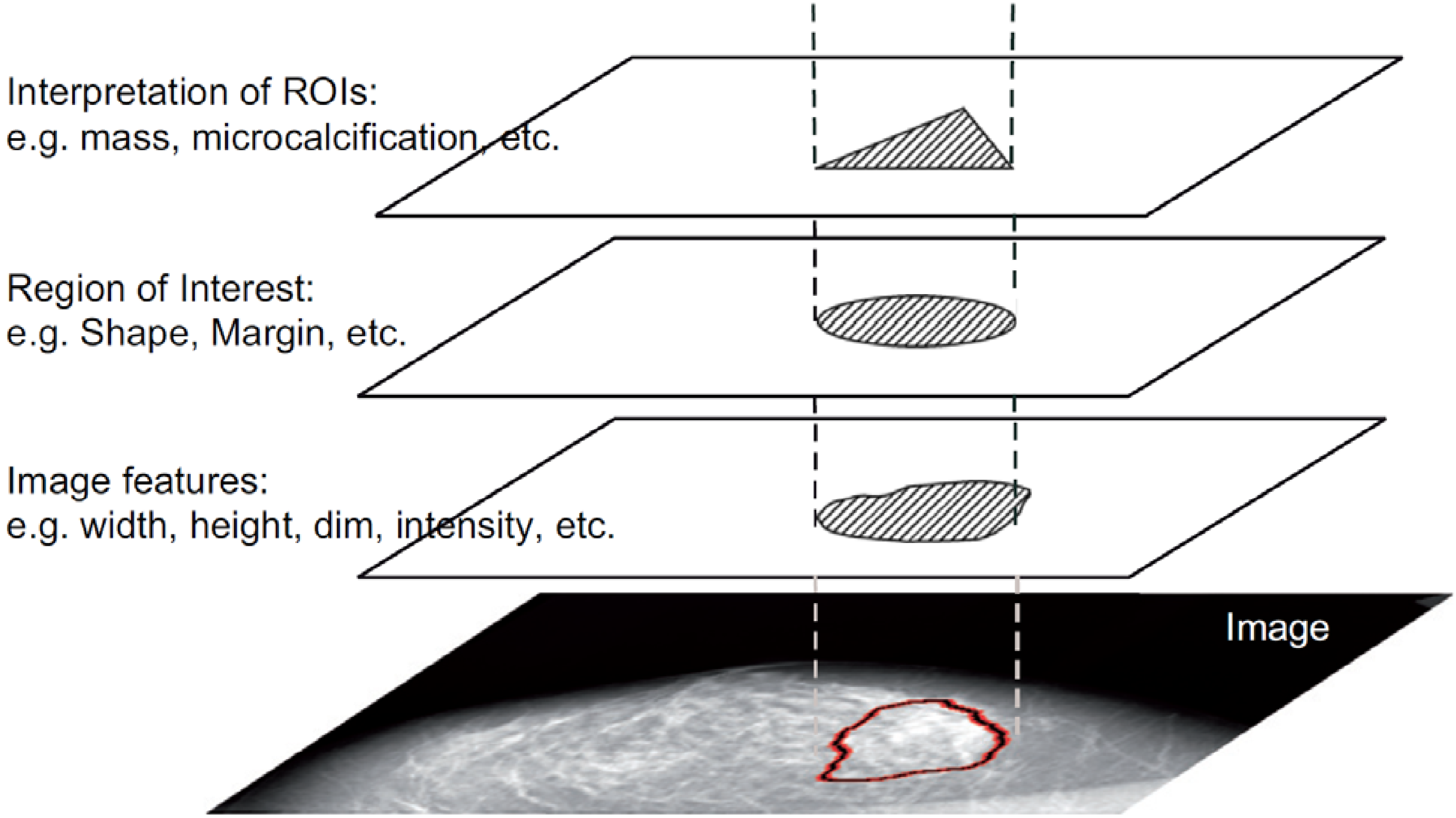


Example: Ontol. Development: Gradually enriching BCIO



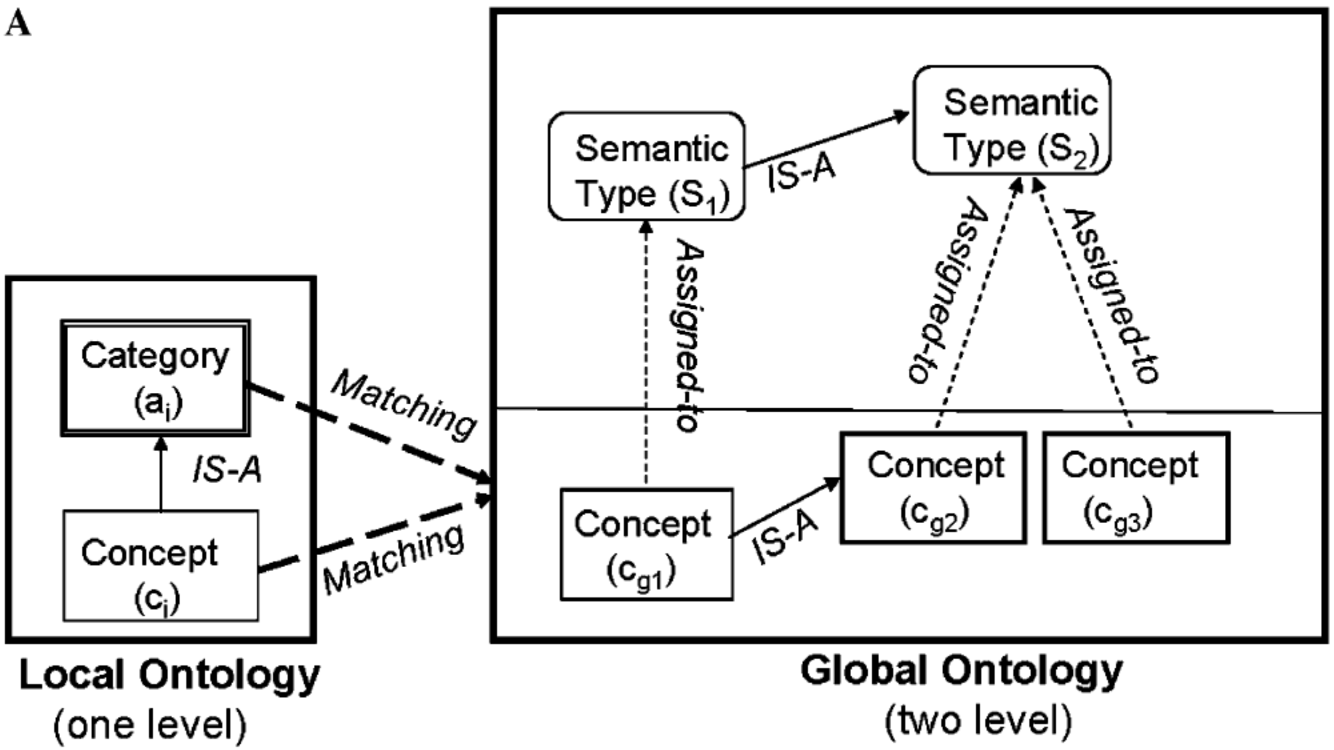
Hu, B., Dasmahapatra, S., Dupplaw, D., Lewis, P. & Shadbolt, N. (2007) Reflections on a medical ontology. *International Journal of Human-Computer Studies, 65, 7, 569-582.*

Example: Layer of abstraction



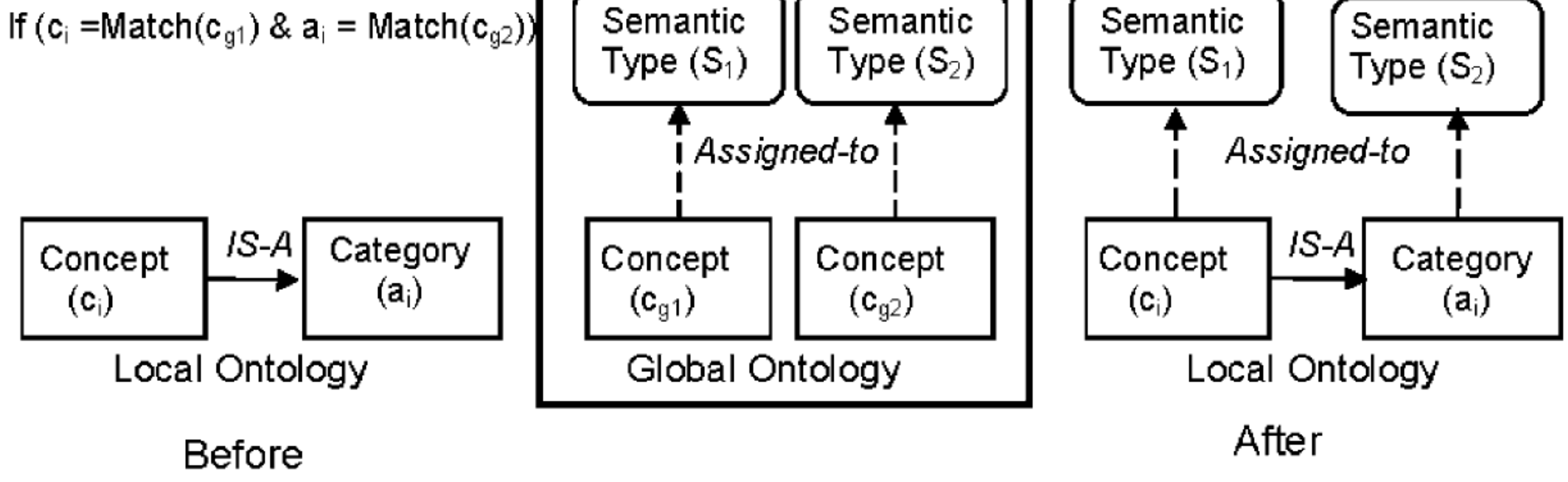
Hu et al. (2007)

Backup-Slide: Medical Ontologies Semantic Enrichment

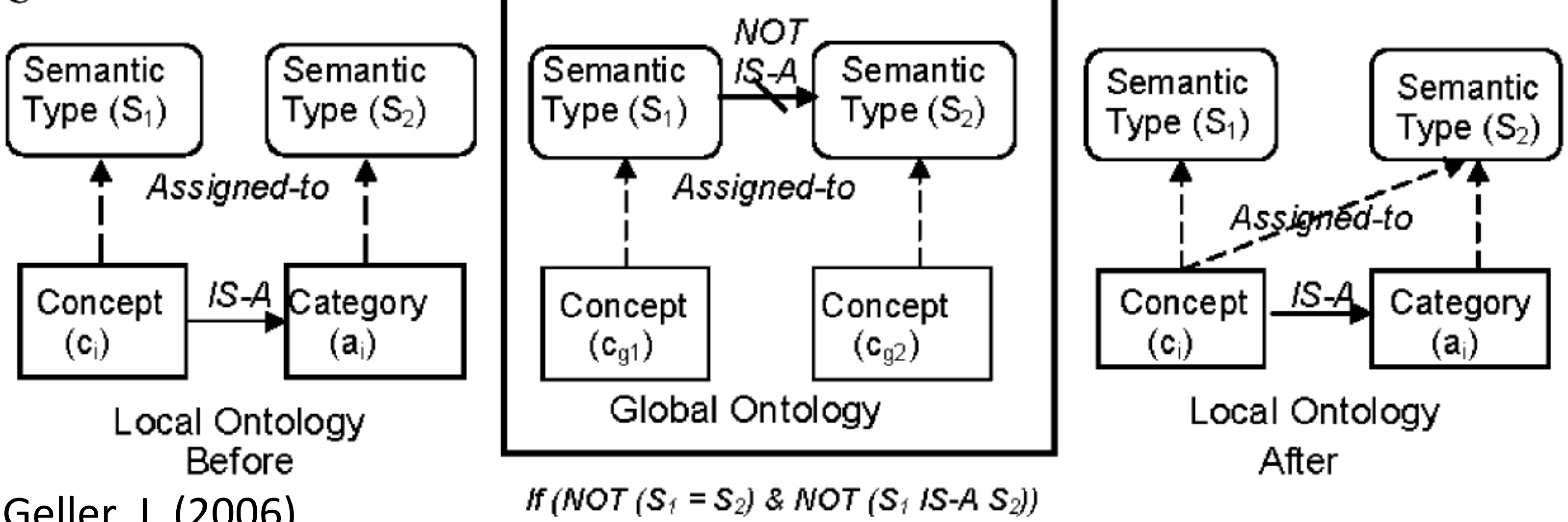


Lee, Y. & Geller, J. (2006) Semantic enrichment for medical ontologies. *Journal of Biomedical Informatics, 39,* ***2, 209-226.***

B

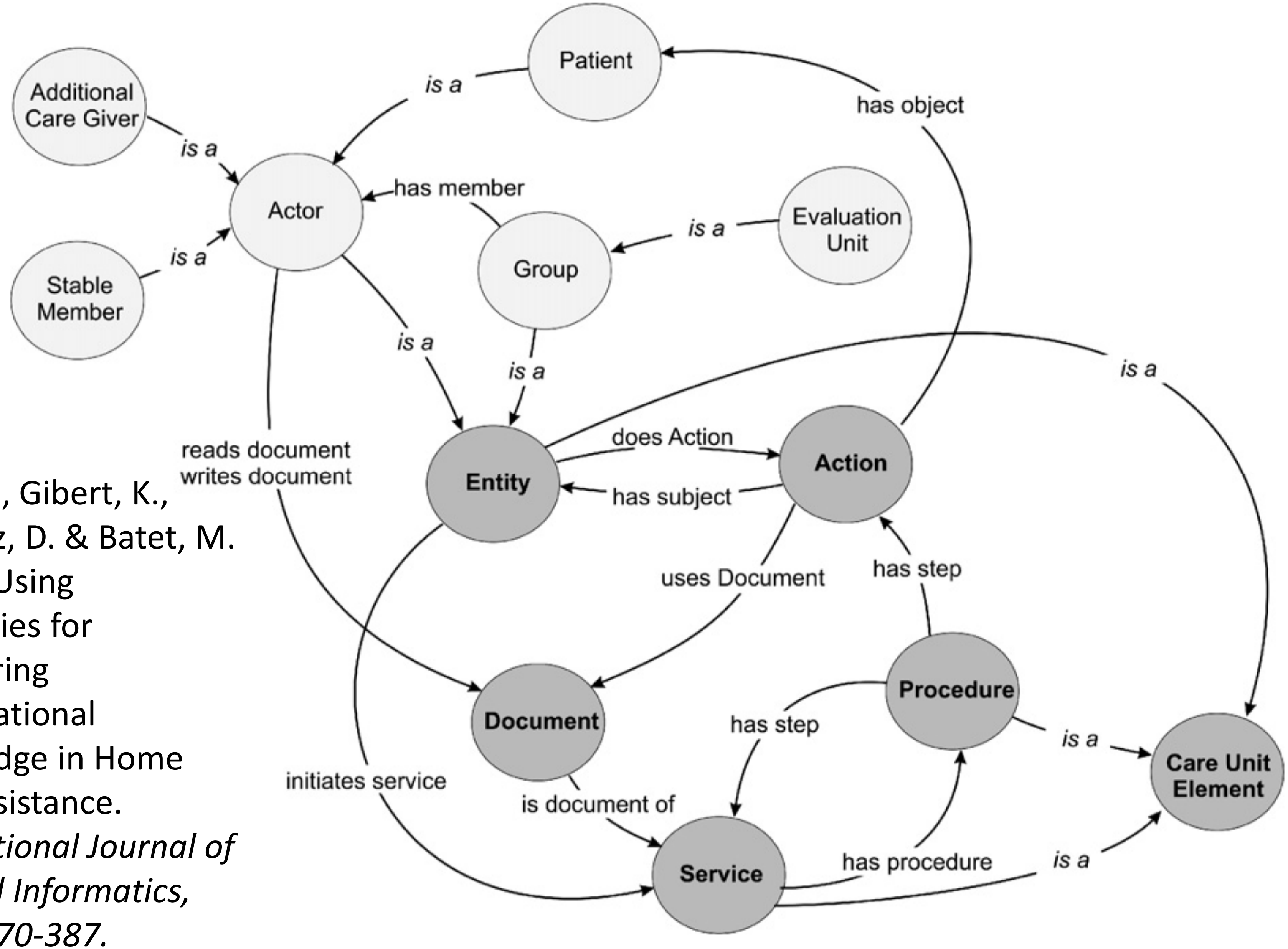


C



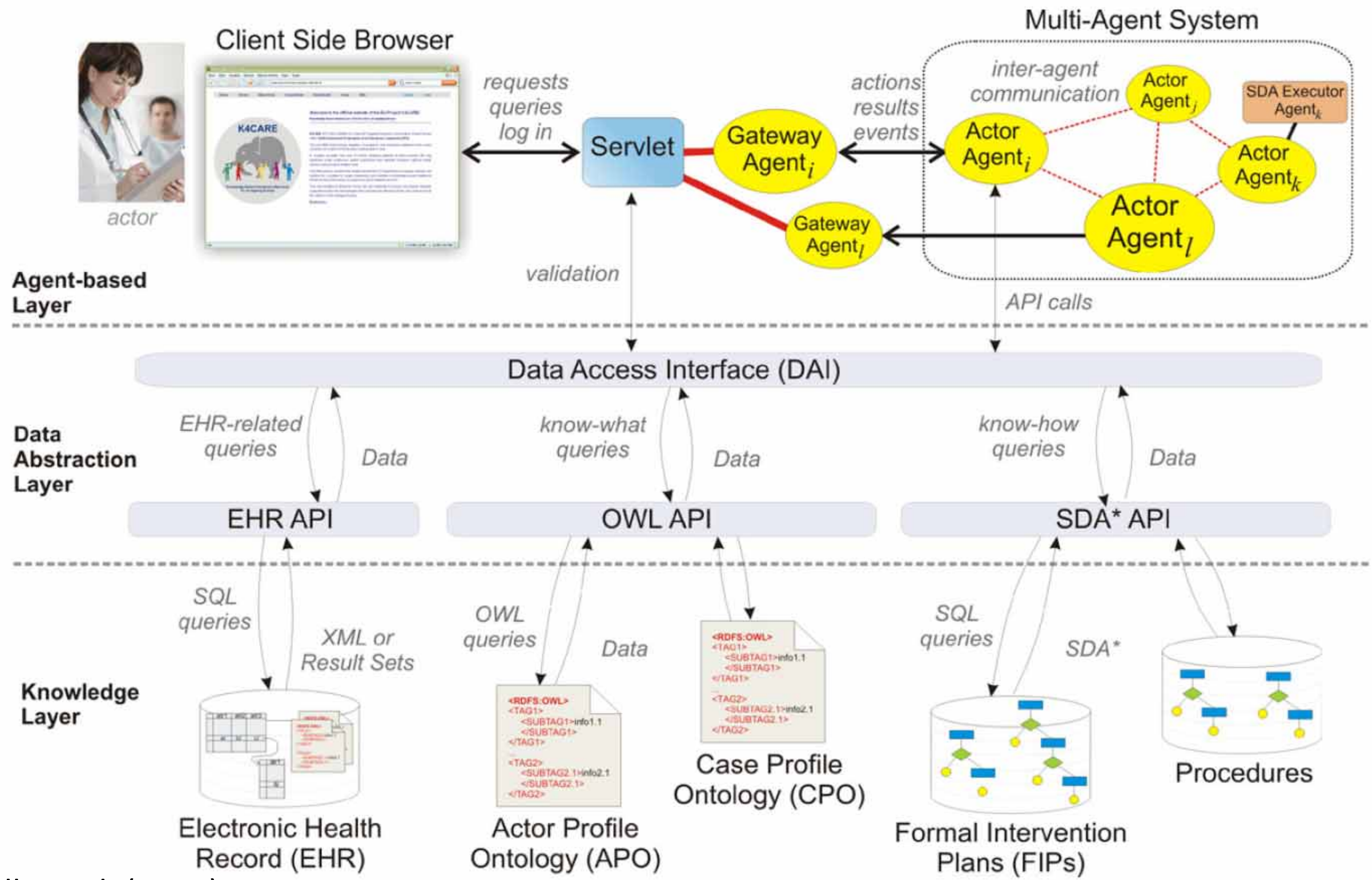
Lee, Y. & Geller, J. (2006)

Backup-Slide: General structure of Actor Profile Ontology



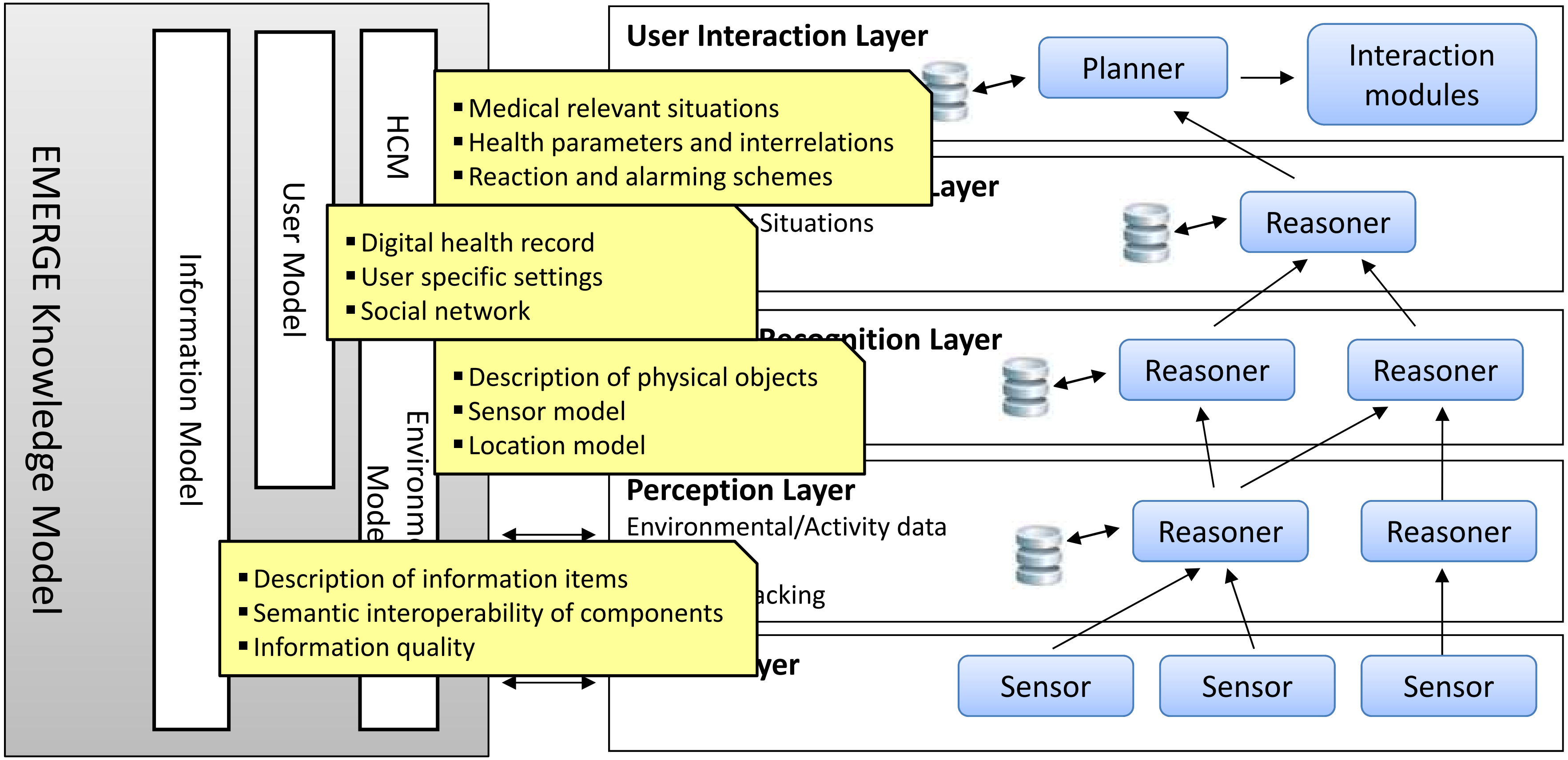
Valls, A., Gibert, K., Sánchez, D. & Batet, M. (2010) Using ontologies for structuring organizational knowledge in Home Care assistance. *International Journal of Medical Informatics, 79, 5, 370-387.*

Backup-Slide: General structure of Actor Profile Ontology



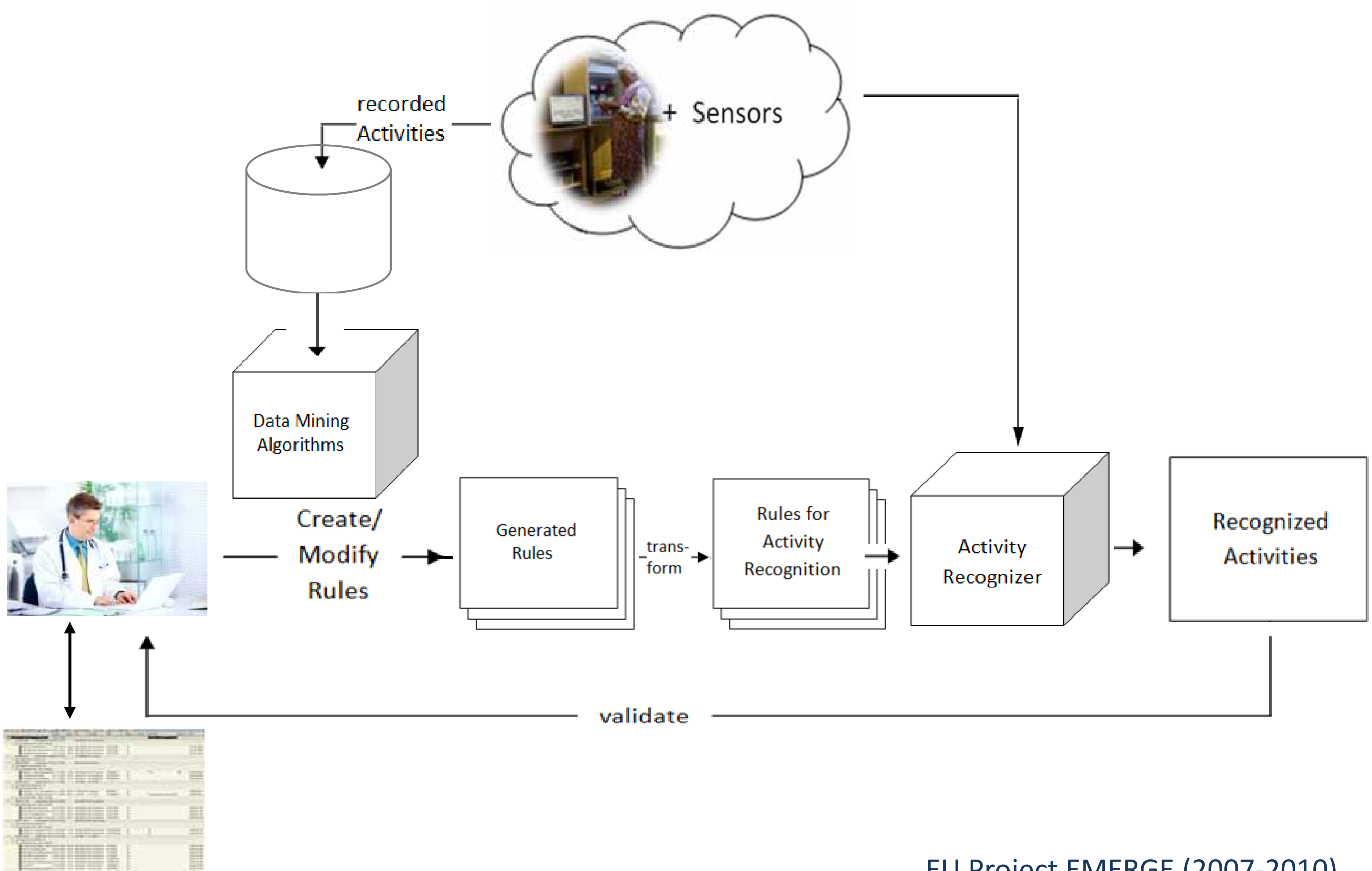
Valls et al. (2010)

Backup-Slide: Example for an OWL DL application



EU Project EMERGE (2007-2010)

Backup-Slide: Example for supervised ontology learning



EU Project EMERGE (2007-2010)

Backup-Slide: Expanding Queries with the MeSH Ontology

MeSH contains two organization files:

1) an alphabetic list with bags of synonymous and related terms, called records, and
2) a hierarchical organization of descriptors associated to the terms.

We consider that a term is a set of words (no word sequence order), that is:

$$t = \{w_1, \cdots, w_{|t|}\} \text{ where } w \text{ is a word}$$

A bag of terms is defined as:

$$b = \{t_1, \cdots, t_{|b|}\}$$

a term t exists in the query q $(t \in q)$ if:

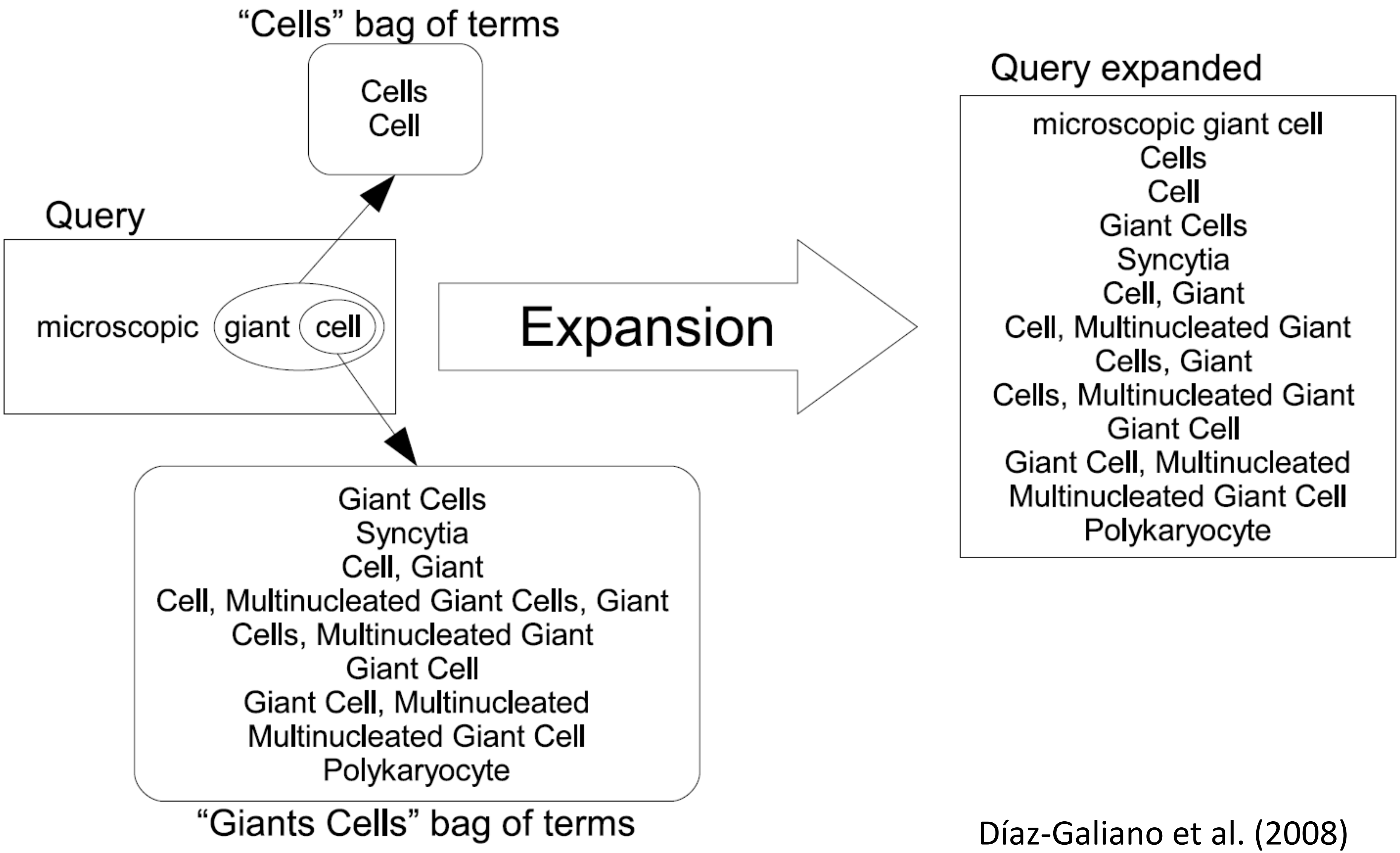
$$\forall w_i \in t, \exists w_j \in q / w_i = w_j$$

Therefore, if all the words of a term are in the query, we generate a new expanded query by adding all its bag of terms:

$$q \text{ is expanded with } b \text{ if } \exists t \in b / t \in q$$

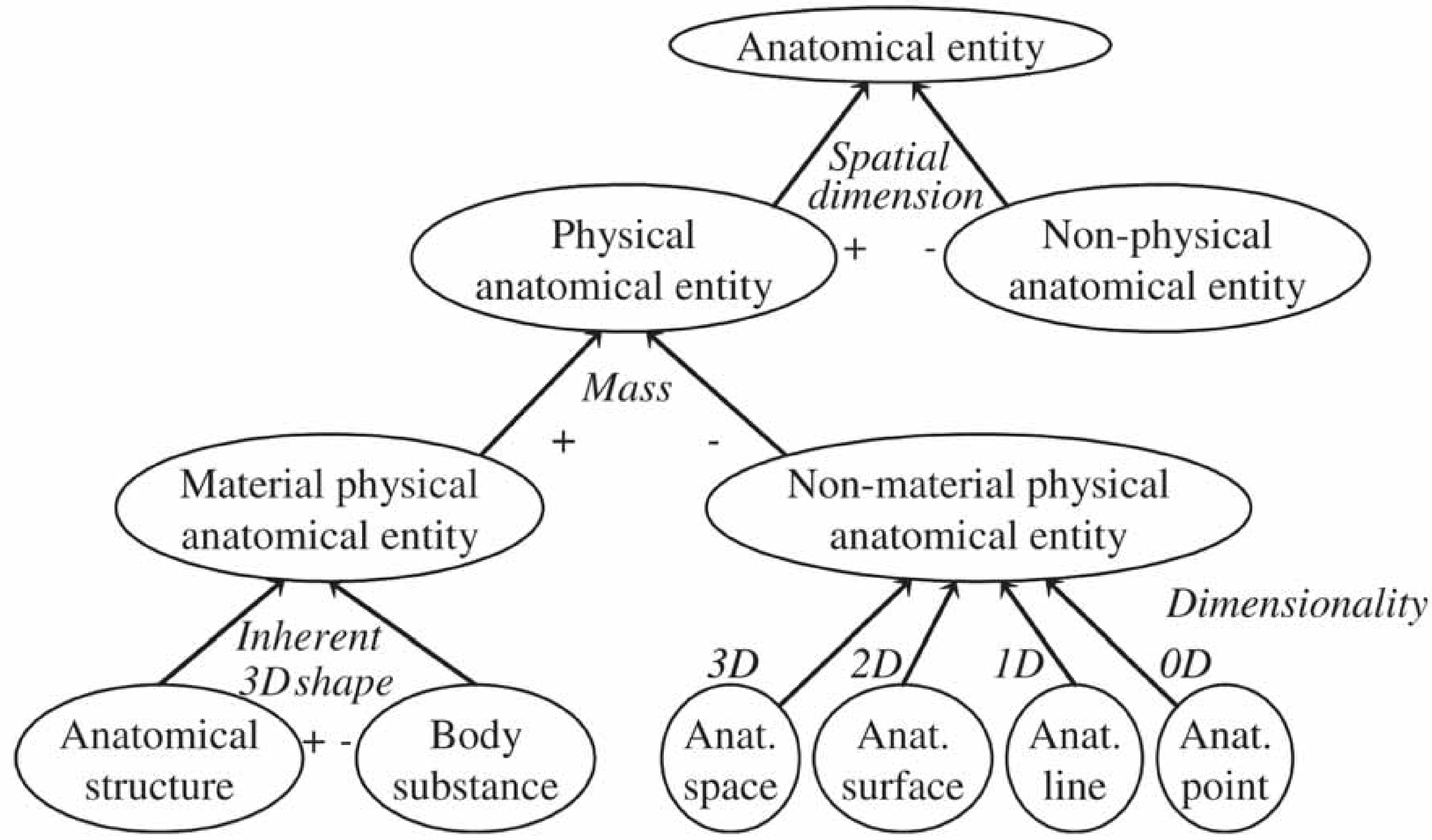
Díaz-Galiano, M. et al. (2008) Integrating MeSH Ontology to Improve Medical Information Retrieval. In: Peters, C. et al. (Eds.) *Advances in Multilingual & Multimodal Information Retrieval, Lecture Notes in Computer Science 5152. Berlin, Heidelberg, New York, Springer, 601-606.*

Backup-Slide: Expanding Queries with the MeSH Ontology



Díaz-Galiano et al. (2008)

Backup-Slide: Foundational Model of Anatomy (FMA)



Zhang, S. & Bodenreider, O. (2006) Law and order: Assessing and enforcing compliance with ontological modeling principles in the Foundational Model of Anatomy. *Computers in Biology and Medicine, 36, 7-8, 674-693.*