

Stefan Schulz Medical University of Graz (Austria)

purl.org/steschu



# How Ontologies can Improve Semantic Interoperability in Health Care

KR4HC – ProHealth'13 Murcia, Spain, June 1<sup>st</sup>, 2013

### **Semantic Interoperability**

"... integrating resources that were developed using different **vocabularies** and different **perspectives** on the data. To achieve semantic interoperability, systems must be able to exchange data in such a way that the precise meaning of the data is readily accessible and the data itself can be translated by any system into a form that it understands."

Jeff Heflin and James Hendler (2000) Semantic Interoperability on the Web <a href="http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf">http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf</a>

"... integrating resources that were developed using different **vocabularies** and different **perspectives** ... "

Vocabularies": Terminologies / classifications / ontologies

- Provide codes that denote types of clinical entities (84114007 |heart failure|): ICD, SNOMED CT.
- "Perspectives": EHR information models
  - Provide standardized structure (section, entry, cluster, etc.) and context (past history, order, ...) for clinical data: openEHR, ISO 13606, HL7 CDA.

#### "Vocabularies": Terminologies / classifications / ontologies

#### Provide codes that denote types of clinical entities (84114007 |heart failure|): ICD, SNOMED CT.

#### Parent(s):

(Select a parent to make it the "Current Concept".) Malignant neoplasm of thorax (disorder) Neoplasm of breast (disorder)

> **Current Concept:** Malignant tumor of breast (disorder)

#### Child(ren):

(N=16) (Select a child to make it the "Current Concept".) Carcinoma of breast (disorder) Familial cancer of breast (disorder) Hormone receptor positive malignant neoplasm of breast (disorder) Local recurrence of malignant tumor of breast (disorder) Malignant lymphoma of breast (disorder) Malignant melanoma of breast (disorder) Malignant neoplasm of axillary tail of breast (disorder) Malignant neoplasm of breast lower inner quadrant (disorder) Malignant neoplasm of breast lower outer guadrant (disorder) Malignant neoplasm of breast upper inner quadrant (disorder) Malignant neoplasm of breast upper outer guadrant (disorder) Malignant neoplasm of female breast (disorder) Malignant neoplasm of male breast (disorder) Primary malignant neoplasm of breast (disorder) Sarcoma of breast (disorder)

#### Current Concept: Fully Specified Name: Malignant tumor of breast (disorder) ConceptId: 254837009 Defining Relationships: Malignant neoplasm of thorax (disorder) Is a Neoplasm of breast (disorder) Is a Group 1 Associated morphology Malignant neoplasm of primary, secondary, or uncertain origin (morphologic abnormality) Finding site Breast structure (body structure) This concept is fully defined. Qualifiers: View Qualifying Characteristics and Facts Descriptions (Synonyms): Fully Specified Name: Malignant tumor of breast (disorder) Malignant tumor of breast [379661016] Preferred: Synonym: Breast cancer [379662011] CA - Breast cancer [379663018] Synonym: Preferred: Malignant tumour of breast [379664012] Related Concepts: - All "Is a" antecedents -- All descendents/subtypes -Related concepts demo

#### "Vocabularies": Terminologies / classifications / ontologies

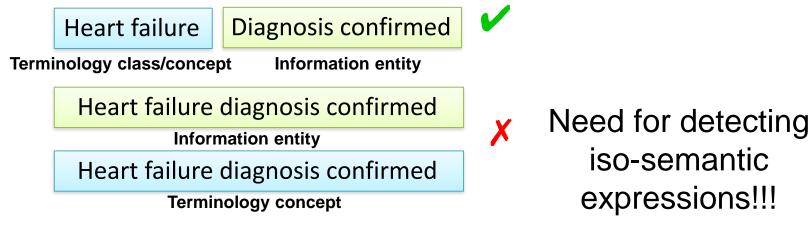
 Provide codes that denote types of clinical entities (84114007 |heart failure|): ICD, SNOMED CT.

- "Perspectives": EHR information models
  - Provide standardized structure (section, entry, cluster, etc.) and context (past history, order, ...) for clinical data: openEHR, ISO 13606, HL7 CDA.

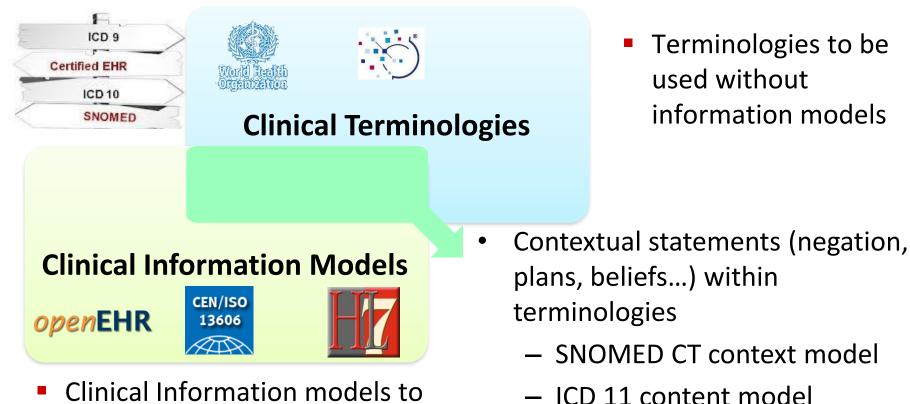
```
SECTION[at0000] matches { -- History of problem / condition
members cardinality matches {1..*; unordered} matches {
    ENTRY[at0001] matches { -- Problem / Condition
    items cardinality matches {1..*; unordered } matches {
        ELEMENT[at0002] matches { -- Diabetes Mellitus
        value matches {
            SIMPLE_TEXT[at0003] matches { -- SIMPLE_TEXT
            originalText matches {"Yes","No","Unknown"}
        }
    }
    }
}
```

#### Vocabularies": Terminologies / classifications / ontologies

- Provide codes that denote types of clinical entities (84114007 |heart failure|): ICD, SNOMED CT.
- "Perspectives": EHR information models
  - Provide standardized structure (section, entry, cluster, etc.) and context (past history, order, ...) for clinical data: openEHR, ISO 13606, HL7 CDA.



# **Overlap Terminologies / Information Models**



- Clinical Information models to be used without or with inexpressive terminologies
- Local terminology within IMs
- Postcoordination within IMs

#### **Consequence: Plurality of isosemantic encodings**

The same meaning is represented by...

"Suspected heart failure caused by ischaemic heart disease"

"Finding with explicit context" and DueTo some "Ischaemic heart disease" and FindingContext some "Suspected"

Diagnosis: Certainty: Etiology: "Heart Failure" "Suspected" "Ischaemic heart disease" ... single codes in different terminologies

... postcoordinated
 expressions in
 different terminologies

... different combinations between terminologies and information models



#### About

# FIRST YEAR DELIVERABLES

SemanticHealthNet will develop a scalable and sustainable pan-European organisational and governance process for the semantic interoperability of clinical and biomedical knowledge, to help ensure that EHR systems are optimised for patient care, public health and clinical research across healthcare systems and institutions.

Through a clinically-driven workplan, exemplified in cardiovascular medicine, *SemanticHealthNet* will capture the needs for evidence-based, patient-centred integrated care and for public health, encapsulating existing European consensus in the management of chronic heart failure and cardiovascular prevention. Experts in EHR architectures, clinical data structures, terminologies and ontology will combine, tailor and pilot their best-of-breed resources in response to the needs articulated by clinicians and public health physicians.

These exemplars will be cross-referenced with other domains and stakeholder perspectives via Clinical and Industrial Advisory Boards and interactions with other projects in Topic 5.3. The project will generalise and formalise the methods and best practices in how to combine and adapt informatics resources to support semantic interoperability, and how these can be developed and supported at scale. Health authorities, clinical professionals, ministries, vendors, purchasers, insurers are involved to ensure the project approach and results are realistically adoptable and viable, building on the SemanticHEALTH and CALLIOPE roadmaps.

A business model to justify strategic investments, including the opportunity costs for key stakeholders such as SDOs, industry, will be defined. This, and links with epSOS II and the eHealth Governance Initiative, will inform the shape of the Virtual Organisation that this Network will establish to sustain semantic interoperability developments and their adoption.

The consortium comprises more than 40 internationally recognised experts, including from USA and Canada, ensuring a global impact.

#### SEMANTICHEALTHNET

Call: FP7-ICT-2011-7 Grant agreement for: Network of Excellence (NoE) Project acronym: SemanticHealthNet Project full title: Semantic Interoperability for Health Network Grant agreement no.: 288408 Budget: 3.222.380 EURO Funding: 2.945.364 EURO Start: 01.12.2011 End: 30.11.2014

#### SUPPORTED BY



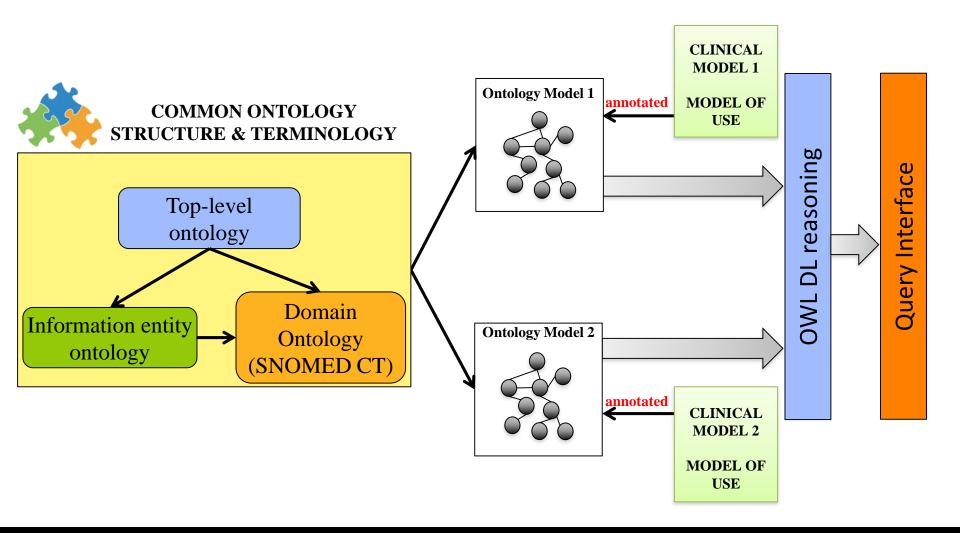
The SemanticHealthNet project is partially funded by the European Commission.

# **Challenge of SemanticHealthNet NoE**

- Create interoperability between isosemantic but heterogeneous representations of structured clinical content
- Target: optimise clinical queries and exchange of data
- Method: Formal ontologies and description logics (OWL DL)

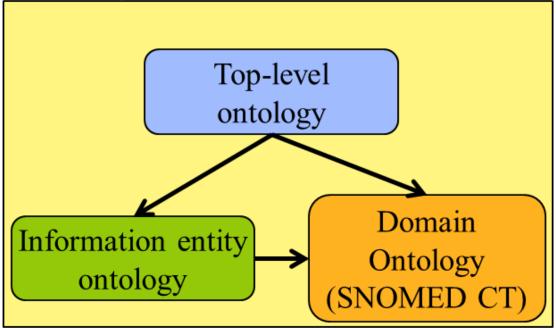
Organ Failure Diagnosis			Diagnosis	Diagnosis
Organ	Heart	•	Suspected heart failure causec by ischaemic heart disease	Heart Failure
Status	Suspected	•		Status
Caused by ischaemic heart disease	Yes No Unknown	×		Suspected  Cause Ischaemic heart disease

## Semantic interoperability by ontology annotations



### **Ontologies for SemanticHealthNet**





# **Different views on (formal) ontologies**

#### Computer science view

- Ontologies are purpose-oriented formal models of meaning (conceptualizations)
- Cognitive / linguistic view
  - Ontologies are concept systems or systems of semantic reference (no clear distinction from thesauri)
  - Also adopted by parts of the Semantic Web community
  - Not clearly distinguished from knowledge representation in general
- Philosophy view (scientific realism)
  - Ontology is the study of what there is
  - Formal Ontologies give precise mathematical formulations of the properties and relations of certain entities.

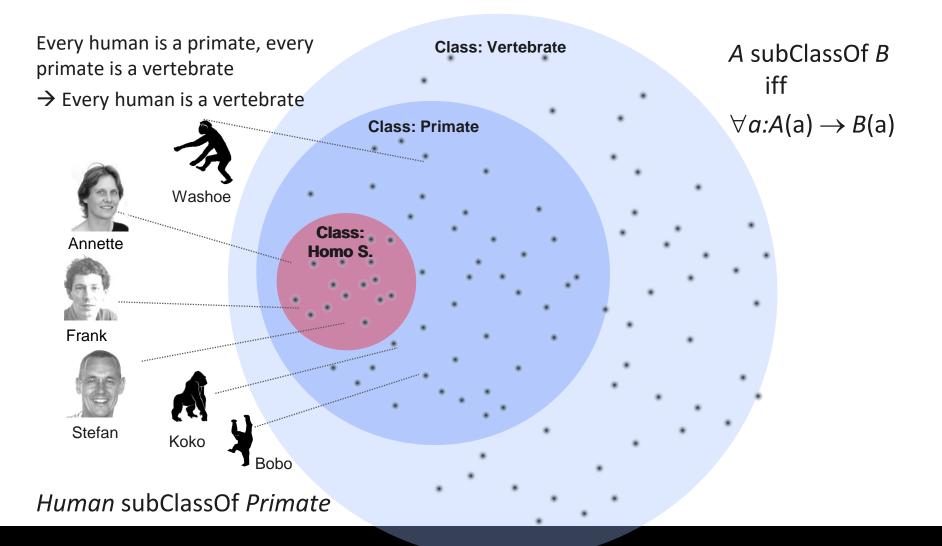
Quine O. On what there is. In: Gibson R. Quintessence - Basic Readings from the Philosophy of W. V. Quine. Cambridge: Belknap Press, Harvard University, 2004. Schulz S, Stenzhorn H, Boeker M, Smith B: Strengths and limitations of formal ontologies in the biomedical domain. RECIIS - Electronic Journal in Communication, Information and Innovation in Health, 2009; 3 (1): 31-45:

## **Basic principles we subscribe to**

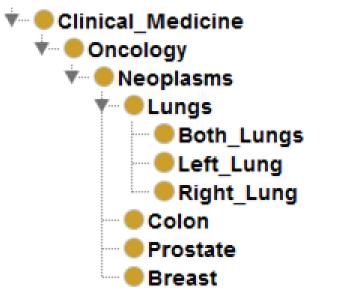
- Ontologies as formal systems (using appropriate language)
- Ontological commitment supported by
  - disjoint upper-level categories (process, object, ...)
  - closed set of basic relations
  - constraining axioms
- Clear division between classes and individuals
- Equivalence and Subsumption statements
- Aristotelian definitions (genus differentia)
- Naming conventions
- Design patterns und guidelines

→ towards "evidence-based" ontology engineering

# **Class-individual distinction not discretionary**



### Intuitive taxonomies ≠ good taxonomies

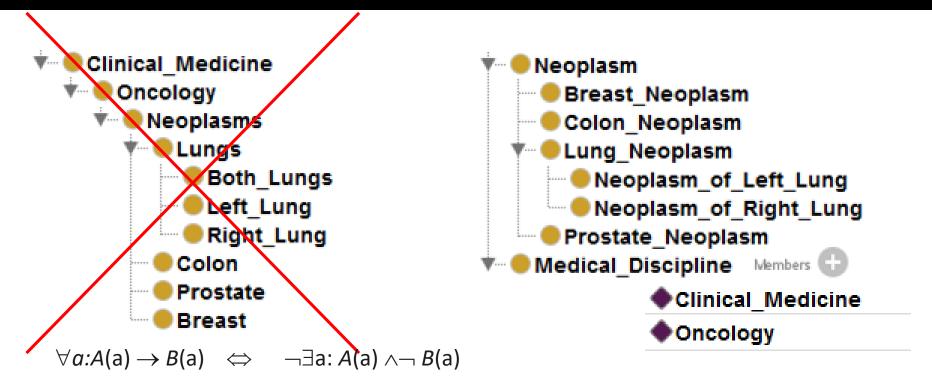


 $\forall a: A(a) \rightarrow B(a) \quad \Leftrightarrow \quad \neg \exists a: A(a) \land \neg B(a)$ 

Test :

- there is no neoplasms that is not an oncology
- there is no prostate that is not a neoplasm
- there is no oncology that is not a clinical medicine

### Intuitive taxonomies ≠ good taxonomies



#### Test :

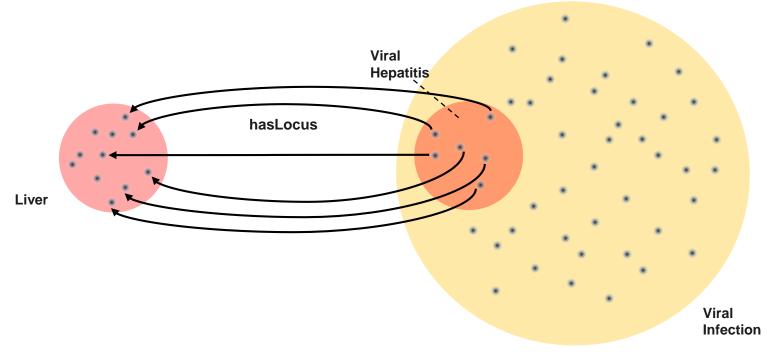
- oncology is an instance of a medical discipline
- there is no prostate neoplasm that is not a neoplasm

Labelling !

http://en.wikipedia.org/wiki/OntoClean

Schober D, Smith B, Lewis SE, Kusnierczyk W, Lomax J, Mungall C, Taylor CF, Rocca-Serra P, Sansone SA. Survey-based naming conventions for use in OBO Foundry ontology development. BMC Bioinformatics. 2009 Apr 27;10:125. doi: 10.1186/1471-2105-10-125.

### **Aristotelian Definitions do not permit exceptions**



FOL:  $\forall x: Hepatitis(x) \leftrightarrow ViralInfection(x) \land \exists Liver: C(z) \land hasLocus(x,z)$ 

*OWL-DL: ViralHepatitis* equivalentTo *ViralInfection* and **hasLocus some** *Liver* Test :

- There is no viral hepatitis that is not located in a liver
- There in no viral hepatitis that is not a viral infection

# Always investigate the ontological commitment

Lung\_Neoplasm
 Neoplasm\_of\_Left\_Lung
 Neoplasm\_in\_both\_lungs
 Neoplasm\_of\_Right\_Lung
 Neoplasm\_in\_both\_lungs

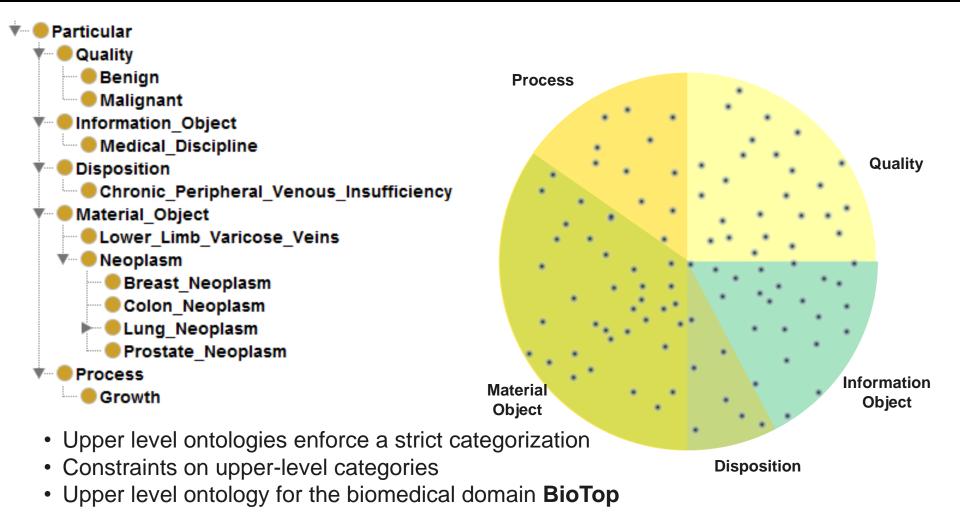
- Which are exactly the instances?
- Does the label tell us what is meant?
- Is there an implicit context?

Disease
Chronic\_Peripheral\_Venous\_Insufficiency
Lower\_Limb\_Varicose\_Veins

Test :

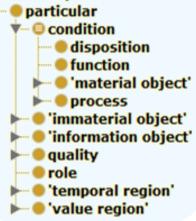
- There is no neoplasm in both lungs that is not a neoplasm in the left lung OR There is no patient with neoplasm in both lungs that is not a patient with the neoplasm in the left lung
- There is no varicose vein in the lower limb that is not a chronic peripheral venous insufficiency OR There is no patient with varicose lower limb veins that is not a patient with a chronic peripheral venous insufficiency

# Upper level ontologies partition the domain into disjoint and exhaustive categories

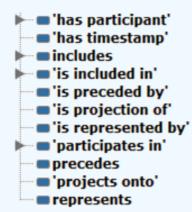


# BioTopLite provides a small set of toplevel classes, relations, and axioms

#### **Toplevel Categories**



#### **Basic relations**



- Precise formulations about generic and defining properties of basic categories of a domain
- Logical Framework (Description logics)
- OWL DL (Web Ontology Language) complete and decidable language - compromise between expressiveness and performance (EXPTIME)

#### Constraining axioms

- 'has part' only process
- 'has participant' some Thing
- 'is bearer of' only 'process quality'
- is part of only process
- projects onto' only 'temporal region'
- 'projects onto' some 'temporal region'
- 'realization of' only disposition
- includes only (process or 'process quality')

- Automated reasoning enables checking consistency, equivalence and subsumption
- Ontologies play an increasing role in new generation of biomedical

Elena Beißwanger, Stefan Schulz, Holger Stenzhorn and Udo Hahn

BioTop: An Upper Domain Ontology for the Life Sciences - A Description of its Current Structure, Contents, and Interfaces to OBO Ontologies in Applied Ontology, Volume 3, Issue 4, Pages 205-212, IOS Press, Amsterdam, December 2008

# Ontology development should be guideline-based

#### Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics

URL: http://www.purl.org/goodod/guideline

Version 1.0 December 2012

Send feedback to: martin.boeker@uniklinik-freiburg.de ludger.jansen@uni-rostock.de

Schulz S<sup>1,3</sup>, Seddig-Raufie D<sup>1</sup>, Grewe N<sup>2</sup>, Röhl J<sup>2</sup>, Schober D<sup>1</sup>, Boeker M<sup>1</sup>, Jansen L<sup>2</sup>

 <sup>1</sup>: Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg
 <sup>2</sup>: Institute of Philosophy, University of Rostock
 <sup>3</sup>: Department of Medical Informatics, University of Graz

11th December 2012

http://purl.org/googod/guideline

### Formal ontologies vs. human conceptualizations

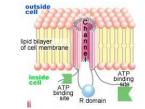
- Built around taxonomies of classes
  - ATTENTION: our intuitive way of hierarchically organize terms is not strictly taxonomic (e.g. Anatomy)
- State what is true for all individual members of a class (instances of a type)
- Requires to distinguish between classes and individuals
  - ATTENTION: human language is often misleading, e.g.
     "London is a big city" vs. "The liver is a big organ"
- Individuals commit to upper-level categories
  - ATTENTION: our thinking fuses mutually dependent entities that belong to different categories, e.g. Cancer (growth process vs. mass of malignant tissue)
- Upper level categories should be made explicit
  - Explicit upper level ontology common understanding
  - Implicit upper level ontology of each of us misunderstanding

### **Beware of creating "Nontologies"**

- OWL syntax without description logics semantics
- Formal reasoning leads to incorrect entailments
- Examples: NCI thesaurus, Radlex
- Many other ontologies contain problematic axioms that contradict the intended meaning
- Example (NCI thesaurus):

Calcium-Activated\_Chloride\_Channel-2 subClassOf Gene\_Product\_Expressed\_In\_Tissue some Lung and Gene\_Product\_Expressed\_In\_Tissue some Mammary\_Gland and Gene\_Product\_Expressed\_In\_Tissue some Trachea

Ureter\_Small\_Cell\_Carcinoma subclassOf Disease\_May\_Have\_Finding some Pain





Schulz S, Schober S, Tudose I, Stenzhorn H: The Pitfalls of Thesaurus Ontologization – the Case of the NCI Thesaurus. AMIA Annu Symp Proc, 2010: 727-731

## Large parts of knowledge are not ontological

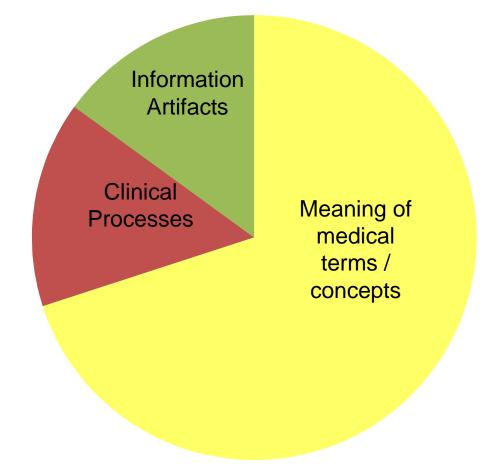
- Ontology ≠ Knowledge representation
  - "There are very few interesting items of knowledge that are truly ontological in this strict sense" (Alan Rector)
  - antinomy:  $\delta v \tau o \varsigma$  (being  $\rightarrow$  ontology) vs.  $\epsilon \pi i \sigma \tau \eta \mu \eta$  (knowledge  $\rightarrow$  epistemology)
- Ontology is not appropriate for
  - Default knowledge
    - "The hand has 5 fingers" (unless otherwise stated)
  - Probabilistic knowledge
    - Mesothelioma is a rare cancer
  - Contingent knowledge
    - Aspirin prevents myocardial infarction
    - Jaundice is a typical symptom of hepatitis
- Can ontology represent clinical information?

# Can formal ontology represent both information and clinical information?

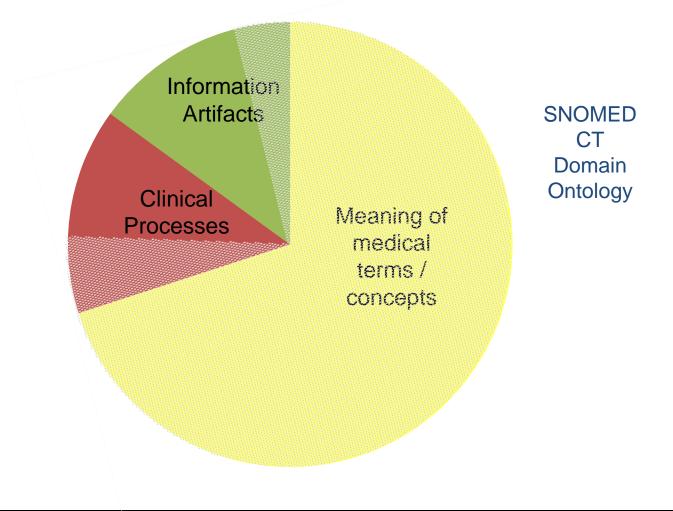
#### **Clinical Terminologies**

#### **Clinical Information Models**

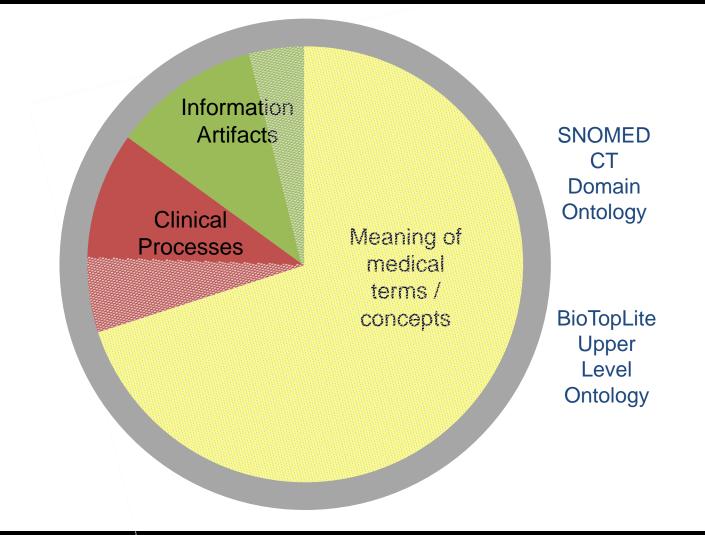
### **Ontologies used and created in SemanticHealthNet**



### **Ontologies used and created in SemanticHealthNet**



### **Ontologies used and created in SemanticHealthNet**

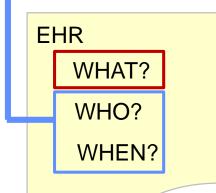


# Basic representational pattern for terminology binding

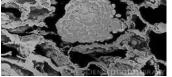
Demographics Time stamps Metadata



 Example: Diagnosis (statement about clinical condition)







# Basic representational pattern for terminology Demographics

Time stamps

Patient X

Metadata

 Example: Diagnosis (statement about clinical condition)

**OWL** annotation of an information item

InformationEntity and hasQuality InformationItemQuality and **isAboutSituation** <u>only</u> (*ClinicalSituation* and ...) Neoplasia EHR WHAT? WHO? WHEN?

#### **Example:**

#### "Suspected heart failure caused by ischaemic heart disease"

#### **Example:**

#### "Suspected heart failure caused by ischaemic heart disease"

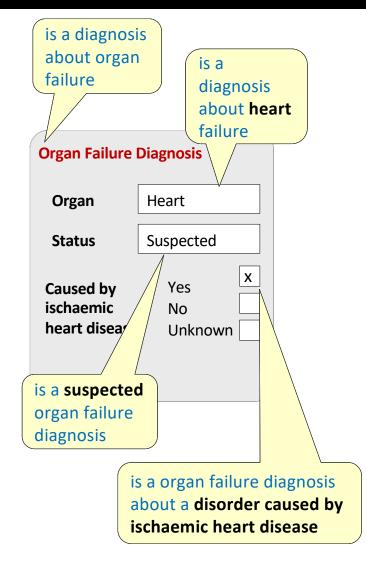
- One code or postcoordinated expression in SNOMED CT
- Reference to two kinds of disorders (ontological types / concepts)
- Semantic relation between both
- Epistemic context: represents state of knowledge about a clinical situation
- Not clear whether there is really some heart failure at all!
- Many entries in EHRs must not be interpreted as factual statements
- Blending of ontological and epistemic information in one code characteristic for many clinical terminologies

#### "Suspected heart failure caused by ischaemic heart disease"

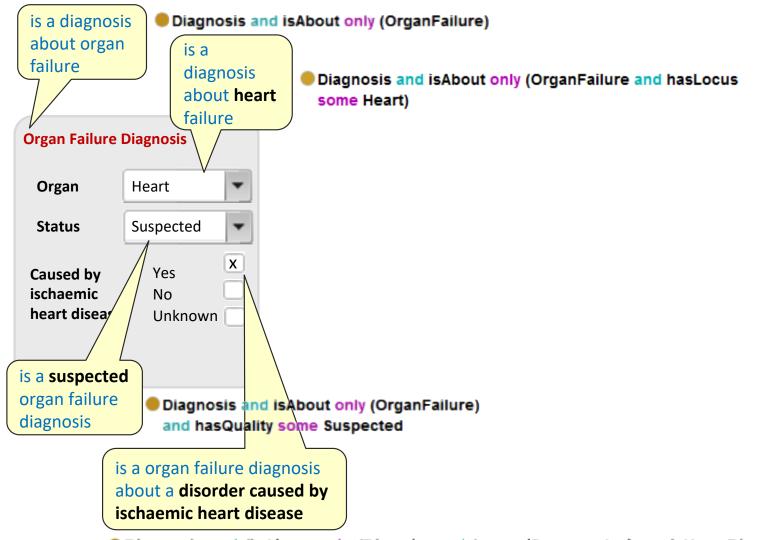
- Three heterogeneous representations of the same statement
- Three different atomic information entities

Organ Failure Diagnosis			Diagnosis	Diagnosis
Organ	Heart	•	Suspected heart failure caused by ischaemic heart disease	Heart Failure
Status	Suspected	-		Status
Caused by ischaemic heart disease	Yes No Unknown	×		Suspected  Cause Ischaemic heart disease

#### "Suspected heart failure caused by ischaemic heart disease" Annotation 1

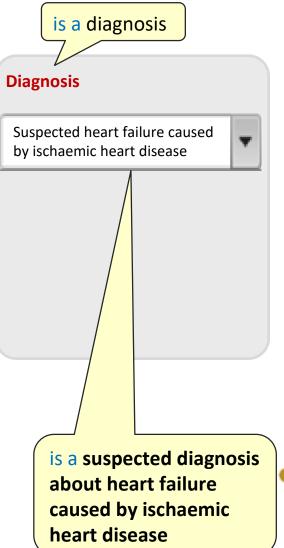


#### "Suspected heart failure caused by ischaemic heart disease" Annotation 1



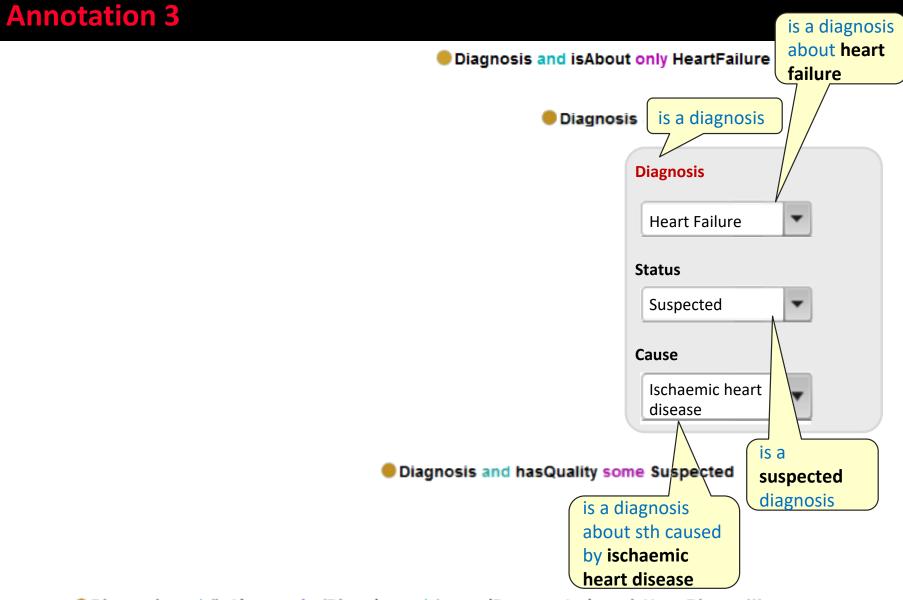
Diagnosis and (isAbout only (Disorder and (causedBy some IschaemicHeartDisease)))

## "Suspected heart failure caused by ischaemic heart disease" Annotation 2



Diagnosis and isAbout only (HeartFailure and (causedBy some IschaemicHeartDisease)) and (hasQuality some Suspected)

#### "Suspected heart failure caused by ischaemic heart disease"



Diagnosis and (isAbout only (Disorder and (causedBy some IschaemicHeartDisease)))

#### **One diagnosis instance for each model**

- Diagnosis\_2 Type Diagnosis
- Diagnosis\_2 Type Diagnosis and (hasQuality some Suspected) and (isAbout only (HeartFailure and (causedBy some lschaemicHeartDisease)))

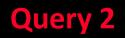
Organ Failure Diagnosis		Diagnosis		Diagnosis	
Organ	Heart 💌		Suspected heart failure caused by ischaemic heart disease		Heart Failure
Status	Suspected 💌			-	Status
Caused by ischaemic	Yes X				Suspected
heart disease					Cause
					Ischaemic heart disease

- Diagnosis\_1 Type Diagnosis and (isAbout only (Disorder and (causedBy some IschaemicHeartDisease)))
- Diagnosis\_1 Type Diagnosis and (isAbout only OrganFailure)
- Diagnosis\_1 Type Diagnosis and (hasQuality some Suspected)
- Diagnosis\_1 Type Diagnosis and (isAbout only (OrganFailure and (hasLocus some Heart)))
- Diagnosis\_1 Type Diagnosis

- Diagnosis\_3 Type Diagnosis and (isAbout only HeartFailure)
- Diagnosis\_3 Type Diagnosis and (hasQuality some Suspected)
- Diagnosis\_3 Type Diagnosis and (isAbout only
- (Disorder and (causedBy some lschaemicHeartDisease)))
- Diagnosis\_3 Type Diagnosis

#### Query 1

Query:					
Query (class expression)					
Diagnosis and isAbout only (HeartFailure and (causedBy s	some IschaemicHeartDisease)) and (hasQuality some Suspected)				
Execute Add to ontology					
Query results	r				
Equivalent classes (1)					
Diagnosis_about_suspected_heart_failure_caused_by	_ischaemic_heart_disease 🛛 🔹 💿				
Ancestor classes (11)					
Diagnosis	0				
Diagnosis_about_condition	0				
Diagnosis_about_disorder_caused_by_ischaemic_hea	rt_disease 🛛 📀				
Diagnosis_about_heart_disorder	0				
Diagnosis_about_heart_failure	0				
Diagnosis_about_heart_failure_caused_by_ischaemic	_heart_disease 🛛 😨				
Diagnosis_about_organ_failure					
Diagnosis_about_suspected_condition	0				
Diagnosis_about_suspected_organ_failure	?				
InformationArtefact	0				
Thing	2				
Instances (3)					
♦ Diagnosis_3	0				
♦ Diagnosis_1 All three info	mation instances found 🛛 👔 🕡				
♦ Diagnosis_2	2				



Query:	0801						
Query (class expression)							
Diagnosis_about_heart_failure <mark>and</mark> Diagnosis_about_suspected_condition <mark>and</mark> Diagnosis_about_disorder_caused_by_ischaemic_heart_disease							
Execute Add to ontology							
Query results							
	_						
Equivalent classes (1)	Super classes						
Diagnosis_about_suspected_heart_failure_caused_by_ischaemic_heart_disease	Ancestor classes						
Instances (3)	🗹 Equivalent classes						
♦ Diagnosis_3	Subclasses						
♦ Diagnosis_1 All three information instances found 2	Descendant classes						
♦ Diagnosis 2	🗹 Individuals						

# Open issues (I)

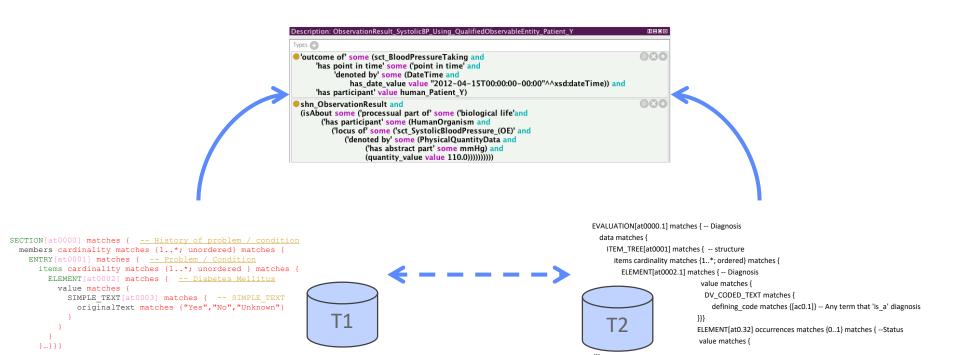
- Accept semantic resources as they are (including what is considered "bad practice" by some), or more prescriptive approach (enforce distinction between terminology and information model)
- Is OWL appropriate to provide appropriate patterns to express "second-order" statements? Alternatives?
- Does the required expressivity (OWL DL + concrete domains) render the framework intractable?
- Query languages: DL, SPARQL, combinations?
- Semantic annotation of formal clinical guidelines?

# **Open issues (II)**

- Is it realistic that IM developers will invest efforts into correctly use OWL for IM annotations?
- Education, training, modification of engineering and maintenance workflows?
- Will the ontological foundation of clinical terminologies be reliable and quality assured?
- To which extent semantic standards will be adopted at all?
- Possibility to use approach for semantic interpretation of text-mined content

# Which are the main scenarios of use?

- Semantic abstraction for querying as demonstrated: using ontology based representation and querying (generic)
- Semantically-enriched data transfer: via semantic abstraction difficult. Rule-based approach (non-generic) ?



## **Further readings**



Stefan Schulz: How Ontologies can Improve Semantic Interoperability in Health Care

## **Ontology on the Web**

- Description Logics: <u>http://dl.kr.org/</u>
- Protégé: <u>http://protege.stanford.edu/</u>
- Bioontology: <u>http://www.bioontology.ch/</u>
- Buffalo Ontology Site: <u>http://ontology.buffalo.edu/smith/</u>
- OBO Foundry: <u>http://obofoundry.org/</u>
- Bioportal: <u>http://bioportal.bioontology.org/</u>
- SNOMED CT: <u>http://www.ihtsdo.org/snomed-ct/</u> <u>http://terminology.vetmed.vt.edu/sct/menu.cfm</u>
- CO-ODE (Pizza ontology): http://www.co-ode.org/
- GoodOD Guideline: <u>http://www.iph.uni-rostock.de/GoodOD-</u> <u>Guideline.1299.0.html</u>

# CLINICAL INFORMATION PATTERNS (WHAT, HOW)

#### **PAST HISTORY OF CONDITION / SITUATION:**

shn:InformationEntity and shn:isAboutSituation only (*btl:BiologicalLife* and btl:hasProcessualPart some <u>shn:ClinicalSituation</u>)

**PRIMARY DIAGNOSIS OF CONDITION / SITUATION:** 



shn:InformationEntity and shn:isAboutSituation only <u>shn:ClinicalSituation</u> and btl:outcomeOf some sct:DiagnosticProcedure

SYMPTOM RECORD

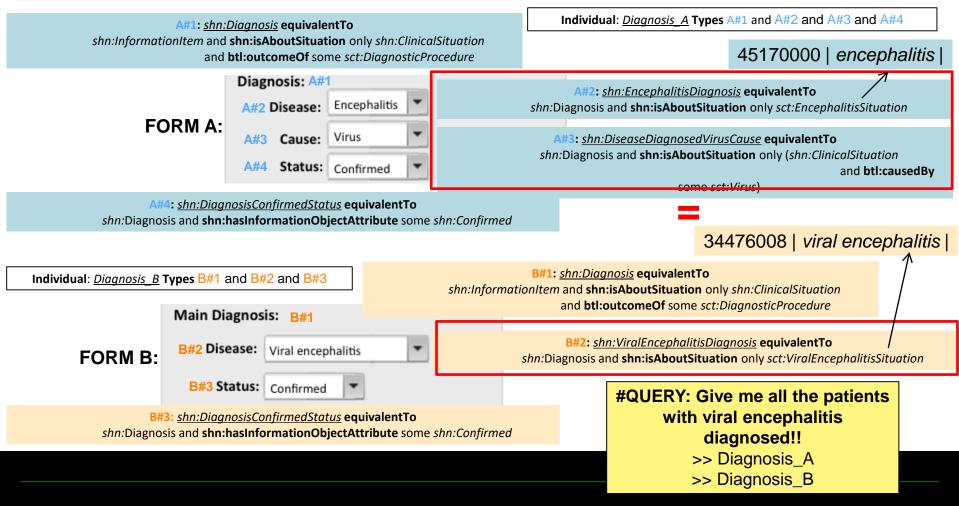
sct:SwollenAnkle

shn:InformationEntity and shn:isAboutSituation only <u>shn:ClinicalSituation</u> and btl:outcomeOf some sct:EvaluationForSignsAndSymptoms

## I. QUERY EXPRESSIVITY

**Confirmed viral encephalitis diagnosis** 

## The ability to support pre/post-coordination



## **II. SUBSUMPTION TESTING**

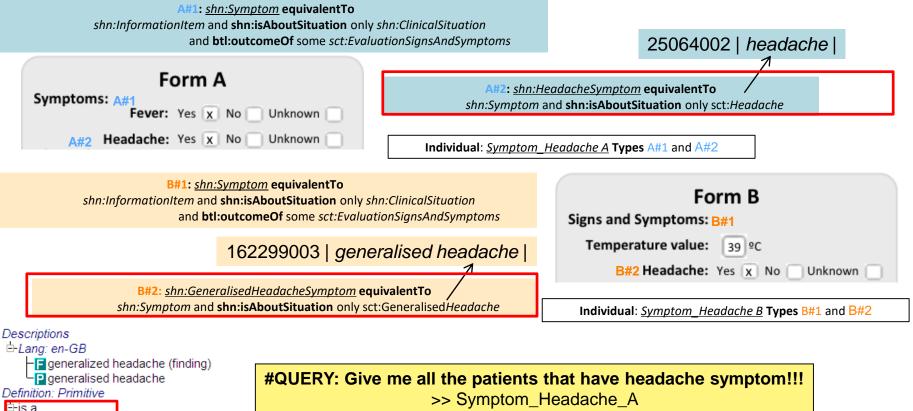
headache

head structure

Inding site

Headache vs. Generalised headache

#### >> A specific condition means also a more general one

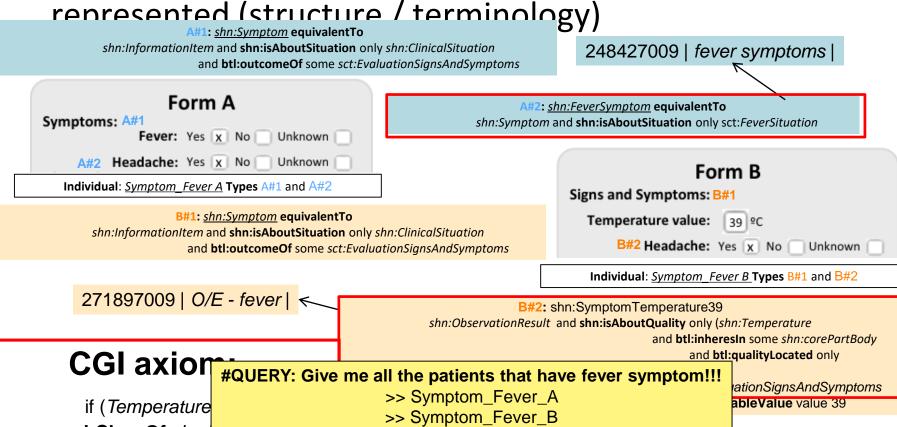


>> Symptom\_Headache\_B

## **III.CONTEXT AWARENESS**

**Fever** 

# Awareness of the **context**, independently of where it is



subClassOf shn: - eversymptom