

Patterns, Properties and Minimising Commitment Reconstruction of the GALEN Upper Ontology in OWL

Alan Rector & Jeremy Rogers

Information Management Group / Bio Health Informatics Forum
Department of Computer Science, University of Manchester

rector@cs.man.ac.uk

co-ode-admin@cs.man.ac.uk

www.co-ode.org

protege.stanford.org

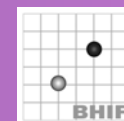


University
of Southampton



EPISTEMICS LTD
"Knowledge is our business"

1



MRC
Medical Research Council



The GALEN Upper Ontology

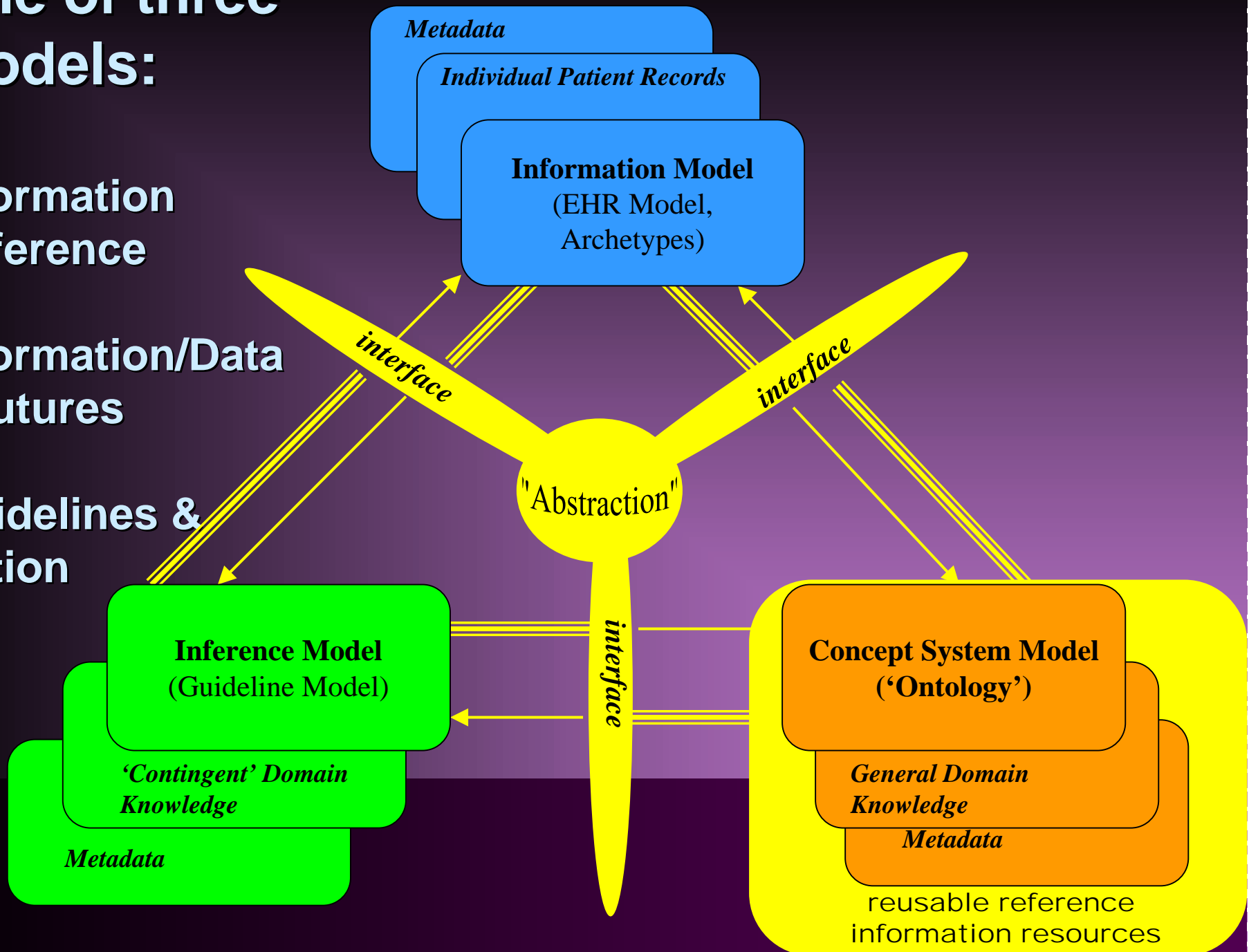
- A lightweight ‘ontology’ for practical software engineering –
 - Software informed by philosophy, but still *software* with a specific task
 - *Representing what healthcare and scientific agents at a given place and time report that they have heard, seen, thought, and done*
- Serves for a large ontology of biomedical concepts
 - Anatomy
 - Physiology
 - Diseases
 - Drugs and their uses
 - Surgical procedures
- Originally Represented in GRAIL – a non-standard DL with
 - Existential restrictions
 - Role hierarchy
 - Propagates via axioms: $R_1 \circ R_2 \rightarrow R_1$
 - Restricted General Inclusion Axioms (absorbable)
 - “Sanctioning” instead of universal restrictions

One of three models:

Information Reference

Information/Data structures

Guidelines & Action



Upper Ontology: What's it for?



The *Upper Ontology* is to enable...

Cooperation on the
Top Domain Ontology
that is to enable

Cooperation on the
Domain Content
Ontologies
that are to enable...

Cooperation on the
Information
resources

How best to reconstruct the GALEN Upper Ontology in OWL?

- With the new expressivity of OWL
- Without propagates_via axioms
- Preserving the principles of “normalisation”
 - Decomposition of primitives into disjoint trees
 - Any information should require changing in only one place
- Taking into account other work and harmonisation
 - The Digital Anatomist FMA & Harmonisation with Mouse Developmental and Adult Anatomy in SOFG
 - OntoClean
 - Barry Smith’s work on Formal Ontology
- Identifying issues that transcend formalism

Principles

- **An Implemented Ontology in OWL/DLs**
 - **Must be implemented and support a large ontology**
- **Must allow definition of top level domain ontology**
 - **The goal is to help domain experts reate their starting points and patterns**
- **Just enough**
 - *No distinction without a difference!*
 - **Properties are as important as Classes/Entities/Concepts**
 - **If an upper level category does not act as a domain or range constraint or have some other engineering effect, why represent it?**
 - **Exclude things that will be dealt with by other means or given**
 - **“Concrete domains”**
 - **Time and place**
 - **Designed to record what an observer has recorded at a given place and time**
 - **Non_physical – e.g. agency**
 - **Causation – except in sense of “aetiology”**

Principles 2

- **Minimal commitment**
 - **Don't make a choice if you don't have to**
- **Understandable**
 - **Experts can make distinctions repeatably/reliably**
- **Able to infer classification top *domain* concepts**
 - **'Twenty questions' – to neighbourhood**
- **Upper ontology primarily composed of 'open dichotomies'**
 - **Open to defer arguments such as whether Collectives of Physical things are physical**

Specific requirements

- **Anatomy, Physiology, Disease, Pathology (Procedures)**
- **Part-whole relations and the relation of diseases to anatomy**
- **Differences in granularity**
- **Differences in view between specialties – FMA & Mouse & GALEN**

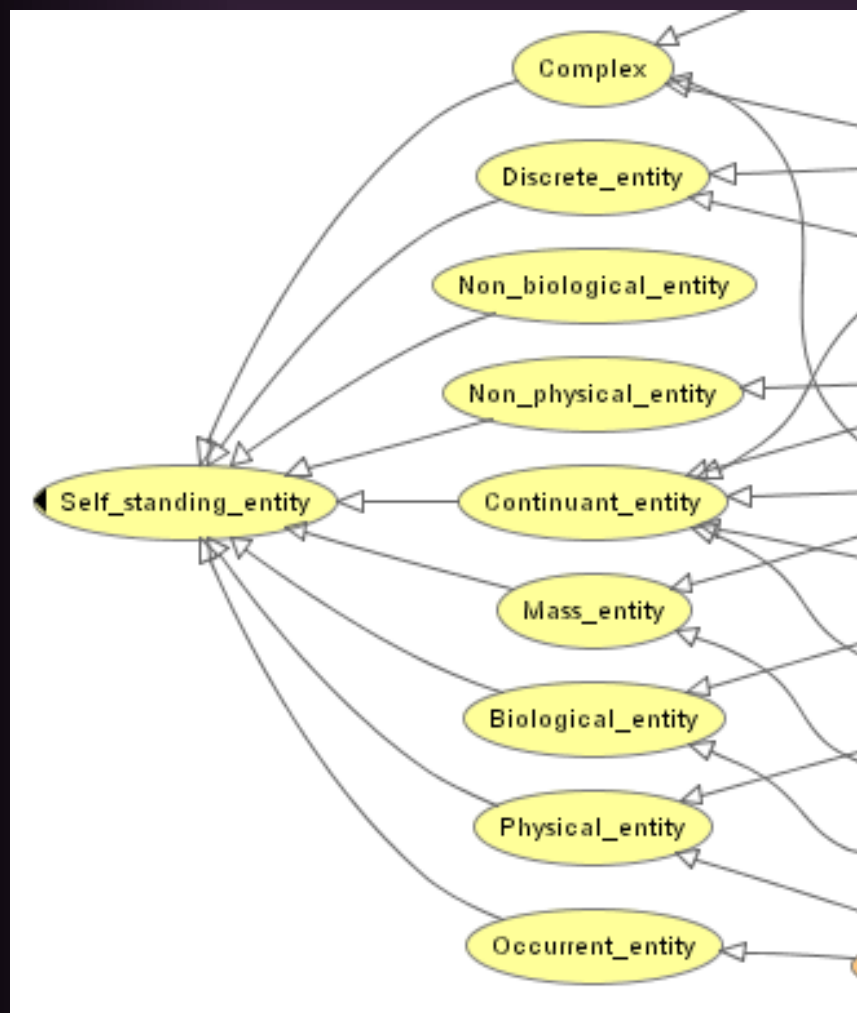
Basic distinctions

- **Self-standing vs Refining**
 - **Probably the same as ‘Independent’ / ‘sortals’**
 - **Property: is_refined_by**
 - **Self_standing_entity is_refined_by Refining_entity**
 - **Establishes the domain & range of a top property distinction**

Within Self Standing

- **Continuant vs Occurrent**
 - **Self_standing_entity participate_in Occurrent**
- **Discrete vs Mass**
 - **Discrete_entity is_constituted_of Mass_entity**
- **Physical vs Non_physical**
 - **Non_physical is_manifested_by Physical**
 - **Only physical can be material**
 - **Material defines non_material (things define holes)**
- **Biological – Non-biological**
- **Complex – all collections, relations, groups, etc.**
 - **No opposite – all arguments deferred**

Basic Distinctions



owl:Thing

Domain_entity

Refining_entity

Self_standing_entity

Biological_entity

Complex

Continuant_entity

Discrete_entity

Mass_entity

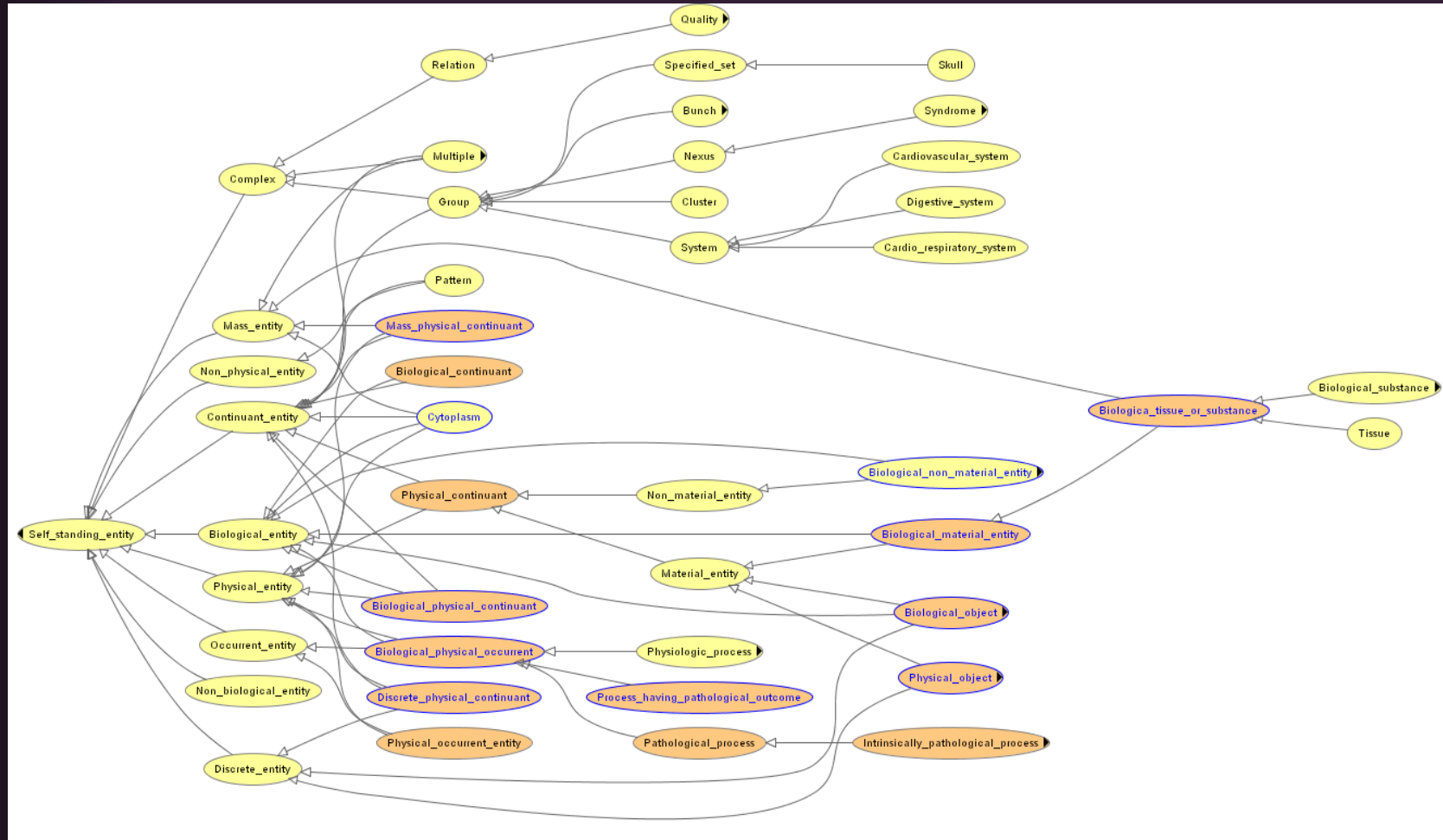
Non_biological_entity

Non_physical_entity

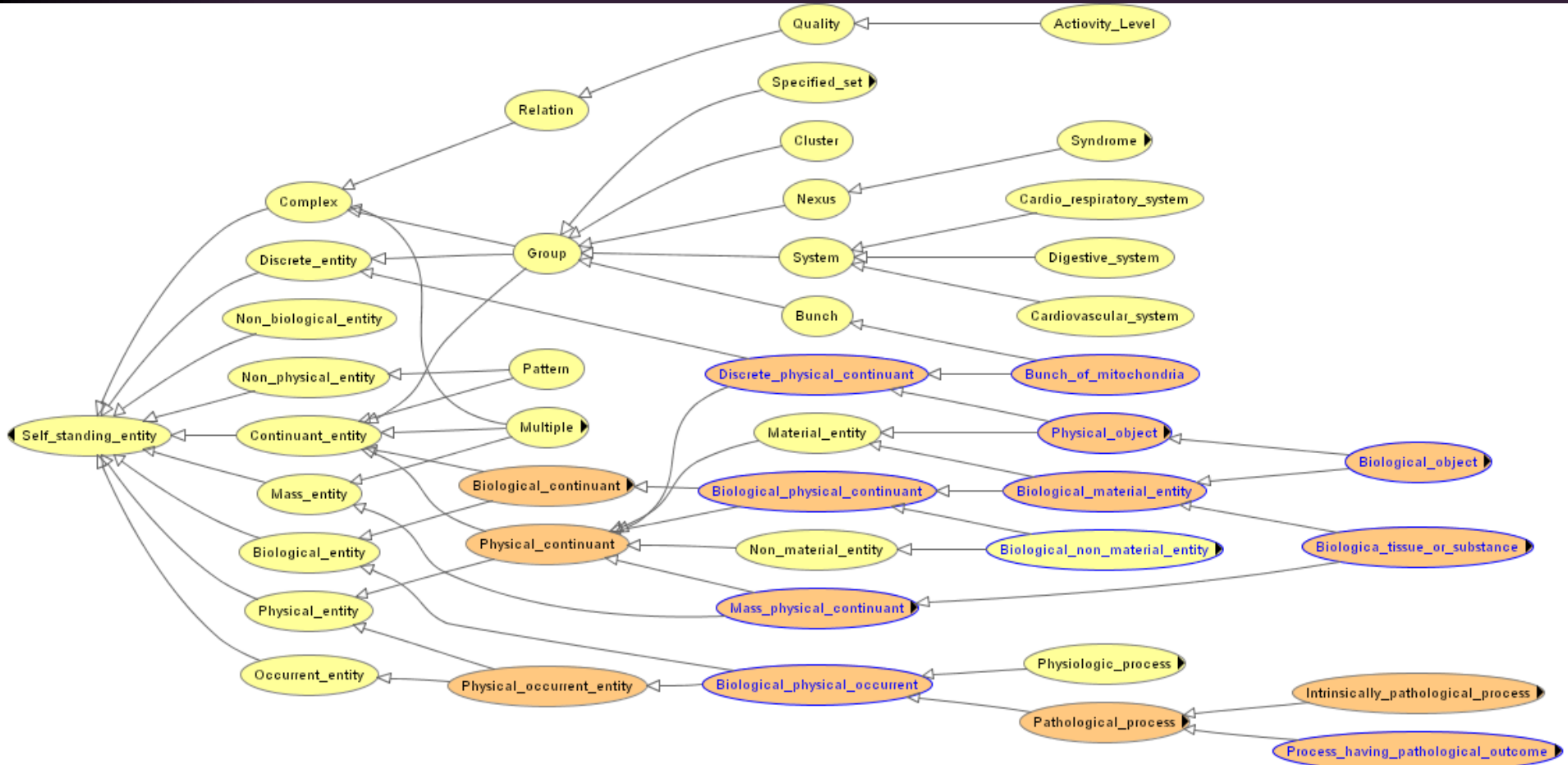
Occurrent_entity

Physical_entity

Unclassified Structure

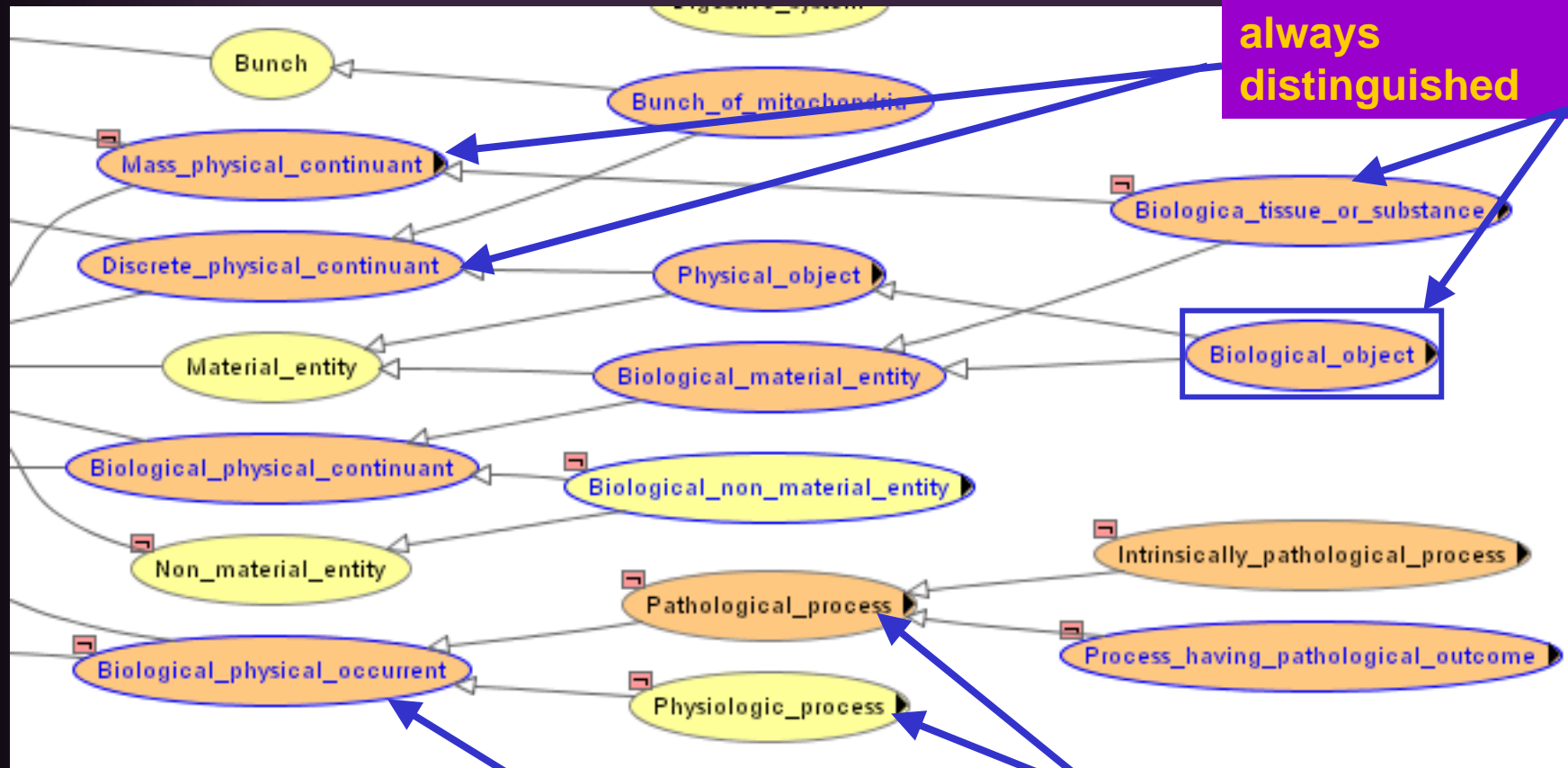


Classified Structure



Detail of Biological Object & Mass/Discrete distinction

Continuants - Mass vs Discrete always distinguished



Occurs - Mass vs Discrete distinguished case by case

“Twenty questions”

Example: What is an Organelle?

- Is it **Continuant** or **Occurrent**? *Continuant*
 - Does it happen or do things happen to it?
- Is it **physical**? *yes*
- Is it **Discrete** or **mass**? *Discrete*
 - (Can you count it?)
- If **physical & discrete**, Is it **material** or **non-material** (thing or hole)? *Material*
- Is it **Biological**? *yes*

Special questions for Biology

- Is it part of something? *yes*
 - if so, definite number or not? *yes*
 - Groups of Organelles are part of Cytoplasm`
- Is it pathological? *no*
- Therefore, it is a “Cell_part” (a subclass of Biological_object)

Before Classification

Classified simply `Biologica_entity`

- ▶ **C** Biological_entity
 - ▶ **C** Biological_continuant
 - ▶ **C** Biological_material_entity
 - ▶ **C** Biological_non_material_entity
 - ▶ **C** Biological_object
 - ▶ **C** Biological_physical_continuant
 - ▶ **C** Biological_physical_occurrence
 - ▶ **C** Cytoplasm
 - ▶ **C** Organelle
 - ▶ **C** Complex
 - ▶ **C** Continuant_entity
 - ▶ **C** Discrete_entity
 - ▶ **C** Mass_entity
 - ▶ **C** Non_biological_entity
 - ▶ **C** Non_physical_entity

RDFS:COMMENT:

Class Description

Paraphrase/descriptive syntax ▼

Class: Organelle

NECESSARILY

Self_standing_entity

Continuant_entity

Material_entity

Discrete_entity

Biological_entity

is_grain_of some (Bunch AND has_grain only

Organelle AND is_functional_part_of some Cell)

After Classification

Classified under Cell_part

- Physical_object
 - Biological_object
 - Biological_object_the
 - Body_part
 - Cell
 - Cell_part
 - Cell_membrane
 - Chromosome
 - Mitochondrion
 - Nucleus
 - Organelle
 - Ribosome
 - Lesion
 - Macro_molecule

Any of the various independent bodies in the cell - mitochondria, chloroplasts etc. There is some disagreement as to whether the nucleus should be

Class Description

Paraphrase/descriptive syntax

Class: Organelle

NECESSARILY

Self_standing_entity

Continuant_entity

Material_entity

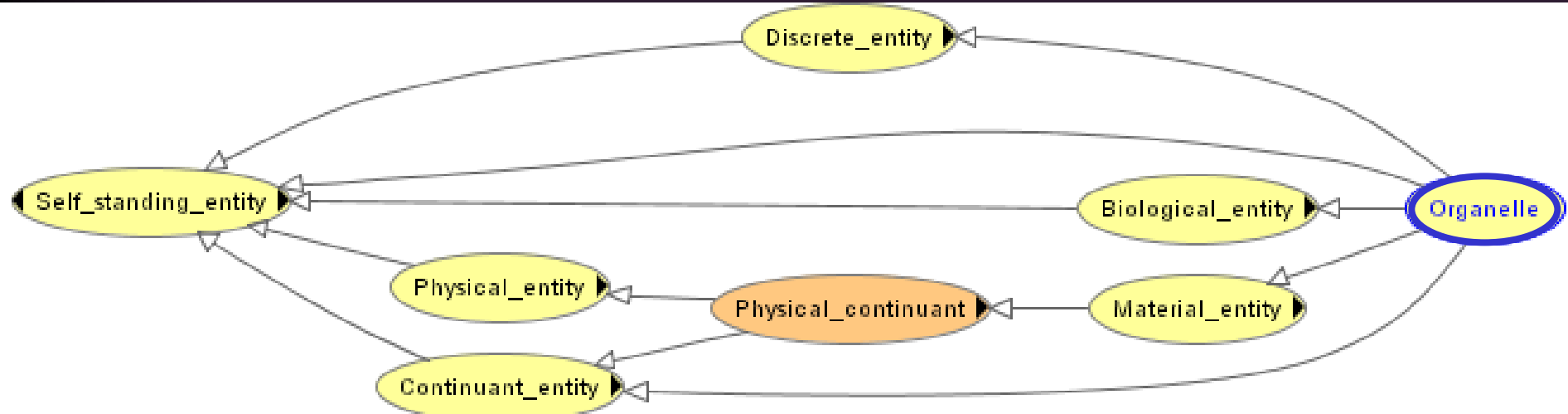
Discrete_entity

Biological_entity

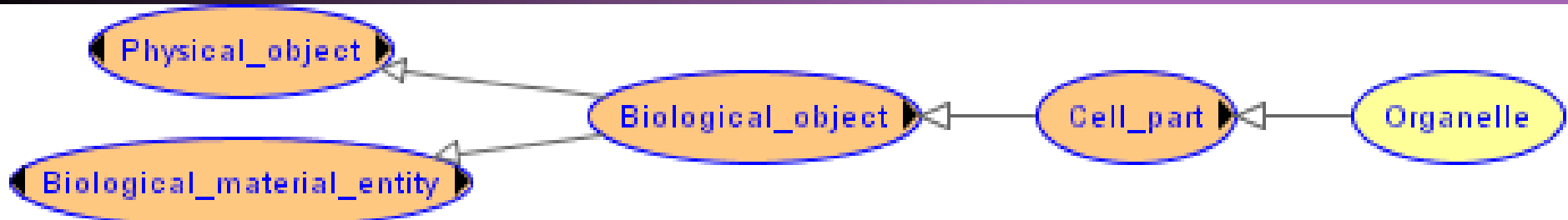
is_grain_of some (Bunch AND has_grain only

Organelle AND is_functional_part_of some Cell)

Before Classification



After Classification



GALEN & Partonomy

- **Need to distinguish**
 - **Location, ‘Locus’, Physical location**
 - **Parthood**
 - **Structural**
 - Subdivisions vs distinct parts
 - **Functional**
 - **Constituent**
 - **Membership**
 - **Containment**
- **Need to capture both FMA and Clinical intuitions**

In DL or OWL must use the Property Hierarchy

- has_location ↔ is_location_of
 - ▶ □ has_locus ↔ is_locus_of
 - ▼ □ has_physical_location ↔ is_physical_location
 - ▼ □ is_part_of ↔ has_part
 - ▼ □ is_clinically_part_of ↔ has_clinically_part_of
 - is_functional_part_of ↔ has_functional_part_of
 - ▶ □ is_structural_part_of ↔ has_structural_part_of
 - ▼ □ is_member_of ↔ has_member
 - is_grain_of ↔ has_grain
 - ▶ □ is_included_in ↔ includes
 - is_in_group_of ↔ has_group
 - is_contained_in ↔ contains
 - is_connected_to ↔ has_connection

- is_structural_part_of ↔ has_structural_part_of
 - is_defined_by ↔ defines
 - is_distinct_part_of ↔ has_distinct_part_of
 - is_suspended_in ↔ is_suspended_in
 - is_portion_of ↔ has_portion_of
 - is_constituent_of ↔ has_constituent_of
 - is_subdivision_of ↔ has_subdivision_of

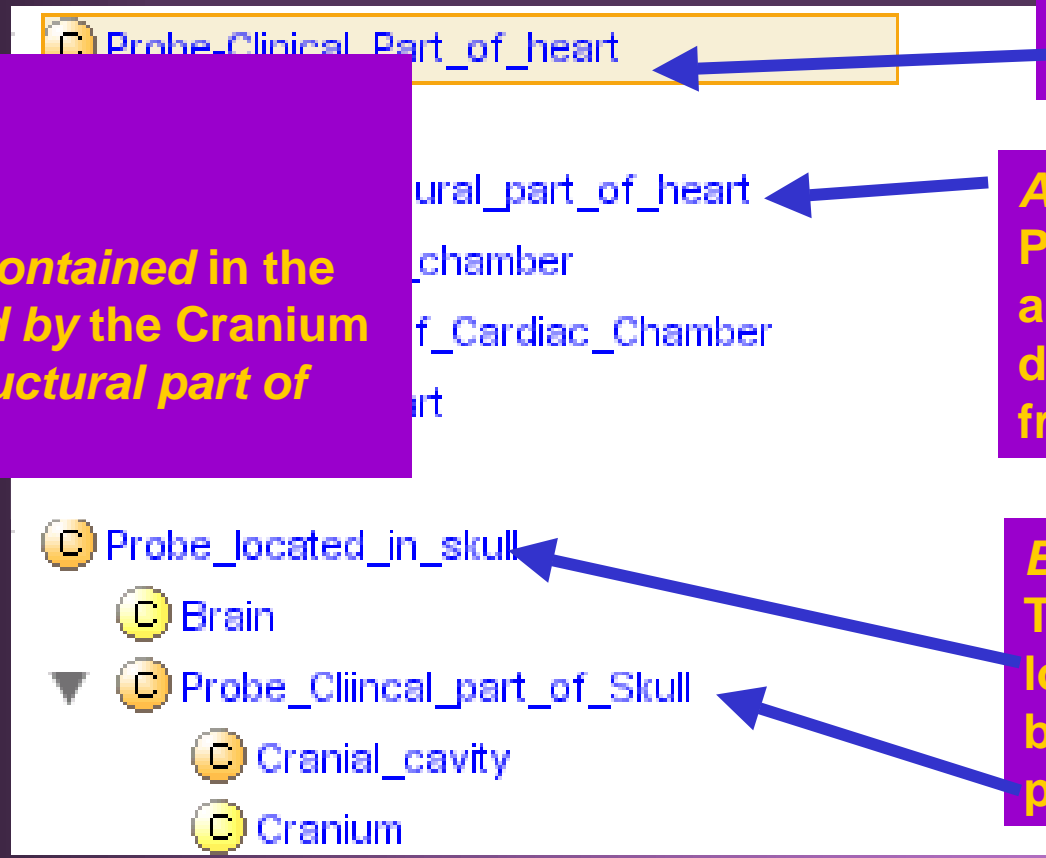
Sufficient to support multiple “views”

Formally:
 The Brain is *contained in the Cavity defined by the Cranium which is a structural part of the skull.*

Clinician's view:
 Pericardium is part of heart & Pericardiitis is a kind of Heart Disease

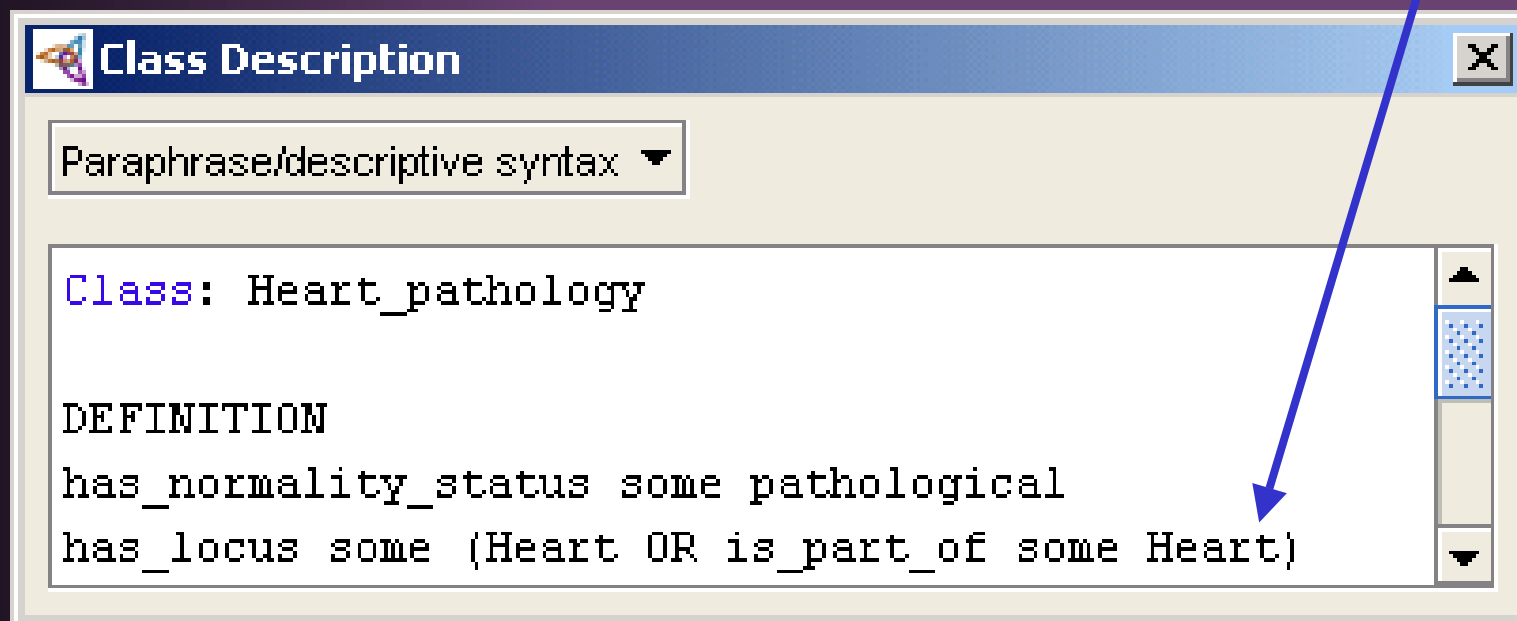
Anatomist's view:
 Pericardium is a distinct organ that develops separately from Heart

Both views:
 The Brain is located in the skull but not part of the skull



Adapted SEP Triples¹

- The disease of the part is a disease of the whole
 - Almost always, but we also want to talk about diseases of the whole only



The screenshot shows a window titled "Class Description" with a close button (X) in the top right corner. Below the title bar is a dropdown menu set to "Paraphrase/descriptive syntax". The main content area displays the following text:

```
Class: Heart_pathology

DEFINITION
has_normality_status some pathological
has_locus some (Heart OR is_part_of some Heart)
```

A blue arrow points from the text "Almost always, but we also want to talk about diseases of the whole only" in the list above to the "has_locus" line in the screenshot.

¹ Schultz & Hahn

Adapted SEP Triples Disease of the whole

- “Removal of the ‘kidney as a whole’”
 - Removal actsOn *some* Kidney
- “Removal of a ‘some part of the kidney’”
 - Removal actsOn *some*
(Kidney or is_clinical_part_of Kidney)
- “Pathology of a Kidney”
 - Pathology has_locus *some*
(Kidney or is_clinical_part_of Kidney)

Adapted SEP Triples

Complex Condition of the whole

- Cardiac_failure is a disorder of the “heart as a whole”

Class: Cardiac_failure

DEFINITION

Cardiac_pumping

has_outcome some

(Circulation_of_blood AND (

has_level_quality some

(Activity_Level AND

has_status some inadequate_status AND

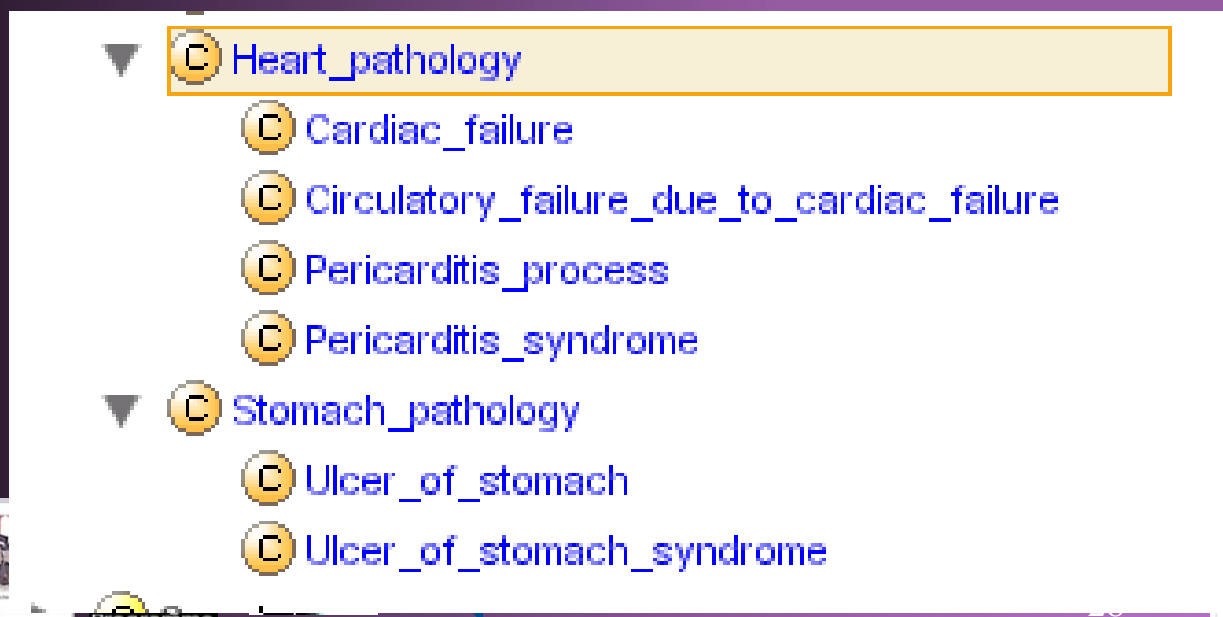
has_normality_status some pathological)))

INHERITS

has_actor Heart ← “as a whole”

What's a Disease?

- A pathological process, object, or quality?
- A “Clinical Situation”
 - A collection of a root pathology and its consequences
 - A “Syndrome”
 - If so, then for classification need policies about when to use syndromes and when not



Duals

- Many processes and their outcomes come in pairs
 - “The erosion (process) of the soil took place over many years”
 - “The erosion (physical thing) extended over many hectares”
- For software engineering reasons, we do not want to define them independently
 - Should the process or object be primary?
 - Erosion_object = Physical_continuant & is_specific_outcome_of some Erosion_occurent
 - Erosion_process = Physical_occurent & has_specific_outcome some Erosion_continuant
 - Should it be a complex?
 - Erosion = Nexus & has_occurent some Erosion_occurent
has_continuant some Erosion_continuant

Engineering Issues

- **Must be consistent within the ontology or classification fails**
 - **An occurrent cannot be a kind of continuant or visa versa**
 - **For a ‘nexus’, the site must be declared for the nexus as a whole**
 - **If for each individually, cannot express the constraint that they must be the same**
- **GALEN chose to make occurrents primary**
(with exceptions for a few complete subontologies)
- **Or is it a continuant pattern that ought to common**
 - **Would this be different from a ‘nexus’ in engineering terms?**

Granularity

Collective vs Individual

- Collectives ('multiples') of discrete entities at one level of granularity form mass entities at the next
 - e.g. multiple of grains of sand is constituent of a beach
 - multiple of red cells are a portion of blood
 - multiple of water molecules are a portion of water
 - multiple of bone cells are a portion of bone tissue
 - is a constituent of long bones
 - The concern is with the collective as a whole not its 'grains'
 - Loss or gain of grains does not affect identity of multiple
 - Not a matter of size,
 - although grains are always smaller than the multiples they make up

Current Controversies

- **Mass vs Discrete entities**
 - **Do tissues exist as distinct from the organs they constitute?**
- **Structured mass entities**
 - **Tissues, cloth, ...**
- **Scale**
 - **Fixed partitions vs case by case representation of “multiples”**

Controversies: How to argue?

- Evidence is effect on representation
 - Is there a real difference or just labelling
 - Are two solutions really isomorphic up to labelling?
 - Relative expressiveness?
 - Effect on hard cases?
 - Understandability? / Repeatability?
 - The views of domain experts
 - Whether there is a transformation from untuitive form to
 - Effect on performance?
 - Small changes can have massive effects on classification time

Summary

- **Implementation works in OWL**
 - Places top domain entities correctly using “twenty questions”
 - Captures notion of views
- **Most of parontology works**
 - Requires elaborate property hierarchy
 - Some paradigms cannot be captured – see Rector 2002
- **SEP Triples work in place of *propagates_via* axioms for coordinating parontology and diseases and procedures**
- **Granularity a matter of collectives vs individuals**
 - NOT of scale.
- **Duals still a problem**
- **Controversy over status of Tissues and other Mass Entities**

www.opengalen.org;

www.co-ode.org

www.cs.man.ac.uk/~rector/ontologies/sample-top-bio