Foundations of the Semantic Web: Ontology Engineering

Building Ontologies 5

Ontology Patterns
Upper Ontologies

Alan Rector & colleagues

Special acknowledgement to Jeremy Rogers & Chris Wroe

An Old Problem

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

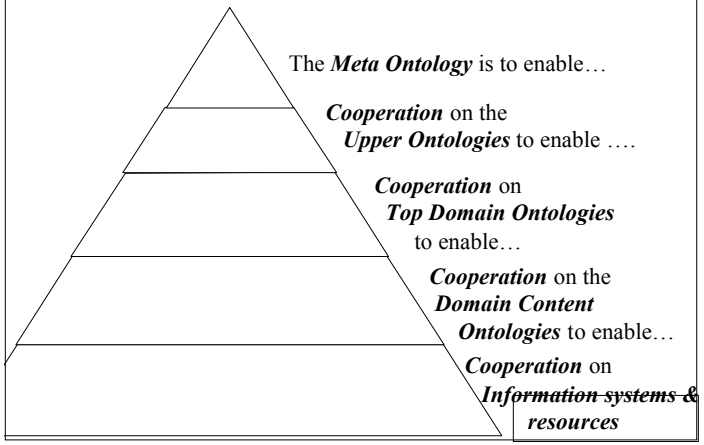
a. those that belong to the Emperor
b. embalmed ones
c. those that are trained
d. suckling pigs
e. mermaids
f. fabulous ones
g. stray dogs
h. those that are included in this classification
i. those that tremble as if they were mad
j. innumerable ones
k. those drawn with a very fine camel's hair brush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

From *The Celestial Emporium of Benevolent Knowledge*, Borges

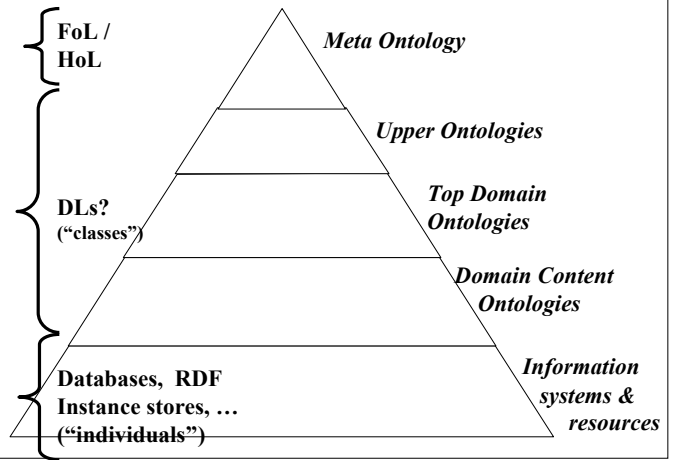
We know it is wrong – but why?

- **Do we really mean wrong?**
- **Many upper ontologies**
  - **Some very abstract, some less so**
  - **Dolce/OntoClean my favourite current compromise besides**
    - **See Guarino and Welty: http://www.loa-cnr.it/DOLCE.html**
      - **doc paper is a readable summary if you can get past the vocabulary**
      - **Also Guarino's home page**
  - **Others**
    - **SUO (Standard Upper Ontology)**
    - **John Sowa's work – see Google**
    - **OpenCyc**
    - **OpenGALEN**
- ***There is no one way!***
  - ***No matter how much some people want to make it a matter of dogma***

Ontology Layers: What's it for?



Where do DLs fit in?



How best to construct an Upper Ontology in OWL?

- With the new expressivity of OWL
- Using the principles of "normalisation"
  - Decomposition of primitives into disjoint trees
  - Any information should require changing in only one place
- Focus on the relations
  - Upper ontology entities should constrain relations
    - otherwise they are a distinction without a difference
- Taking into account other work and harmonisation
  - Eg. for anatomy, 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 an 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
  - the Digital Anatomist's Foundational Model of Anatomy (FMA)
  - Mouse embryo and adult Anatomy
  - GALEN anatomy
  - 'Usual clinical usage'

Upper Ontologies are different

- Domain ontologies are built from trees
  - Disjoint single hierarchies
  - Distinguished by "Rigid Predicates"
- Upper ontologies are built from dichotomies
  - "Dichotomy" – a distinction between two categories
- The goal
  - Be able to ask a few questions and position anything approximately in the right place in the ontology.

The Properties Hierarchy Properties are as important as Classes

- Basic meaning analogous to classes
  - *p*
    p_sub
    p_sub_sub
  - Anything linked by *p_sub_sub* is linked by *p_sub*; Anything linked by *p_sub* is linked by *p*
  - For all xy . x *p_sub_sub* y → x *p_sub* y → x *p* y
  - *p_sub_sub* SOME C→*p_sub* SOME C → *p* SOME C
- A powerful means of inference used in, amongst other things:
  - Part-whole relations
  - "Participations" in processes
  - "Views"
    - allowing different applications to see different aspects of a property
  - Lots of work arounds
    - Transitive property with a non-transitive subproperty

This time begin from the top

- The very top
  - Domain_entity
    - Always good practice to provide your own top
    - You may want to create 'probes' or do other nasty work arounds.
      - The real ontology is under Domain Entity

Basic distinctions

- **Self-standing vs Refining**
  - **Self standing**
    - **Person, computer, idea…**
  - **Refining**
    - **big, serious, efficient, …**
  - **Self_standing_entity *is_refined_by* Refining_entity**
    - **Establishes the domain & range of a top property distinction**
  - **Question: Does it make sense on its own?**
    - **If so, self_standing.**

More Traditional Distinction Independent vs Dependent

- **Independent**
  - **Entities which you can talk about on their own**
    - **Birds, ideas, courses, light…**
- **Dependent**
  - **Entities that only make sense in the context of other things**
    - **Colours, qualities, groups, collections, relations, …**
- **Alternatively - Pierce gave up and Sowa agreed to follow…**
  - **"Firsts", "Seconds" and "Thirds"**
    - **Look up John Sowa on Google**

Within Self Standing

- **Continuant vs Occurrent**
  - **Self_standing_entity *participates_in* Occurrent_entity**
- **Physical vs Non_physical**
  - **Non_physical *is_manifested_by* Physical**
  - **Only physical an be material**
    - **Material defines non_material (things define holes)**
- **Discrete vs Mass**
  - **Discrete_entity *is_constituted_of* Mass_entity**
- **Complex – all collections, relations, groups, etc.**
  - **No opposite – all arguments deferred**
  - **Complex *has_member* Self_standing_entity**
- **(Biological – Non-biological)**
  - **Artifacts, Natural_non_biological**
    - **Exclusive? Think about it**

Continuant vs Occurrent

- ***"Processes happen to things"***
- **Continuants *participate_in* Occurrents**
  - **Occurrents can also participate in other Occurrents**
    - **But only occurrents can be participated in**
      - **One justification for the difference - Occurrent is domain for *has_parfticipant***
- ***Continuants* ("endurants")**
  - **Things that retain their form over time**
    - **People, books, desks, water, ideas, universities, …**
- ***Occurrents ("perdurants")***
  - **Things that occur during time**
    - **Living, writing a book, sitting at a desk, the flow of water, thinking, building the university, ...**
- **Question: Do things happen to it? then *Continuant***
  Does it happen or occur? then *Occurrent*.

Processes act on things

- **One form of participation is acting on**
  - **Linguists call it "agency" but that label gets muddled up with legal agency and responsibility**
  - **Occurrent *acts_on* Self_standing_entity**

Processes have outcomes

- **One form of acting-on something is having it as an outcome outcome**
- **Represented in the property hierarchy**
  - ***has_participant***
    acts_on
    has_outcome
- **Occurrent *has_outcome* Self_standing_entity**
  - **Outcomes can be either Continuants or Occurrents**
    - **But only Occurrents have outcomes**
      - **Check the Domain and Range of *has_participant***

Physical vs non-Physical

- **Physical entities *manifest* non-physical patterns**
  Physical entities *embody* non-physical agents
- **Physical entities have energy or mass and occupy space or time**
  - **bodies, electricity, water, buildings, burning, cavities, planes and lines formed by the intersection of physical things…**
- **Nonphysical things**
  - **Describe "Patterns"**
    - **Forms, styles, 'oeuvres', …**
  - **Describe "psycho-social phenomena"**
    - **Organisations, agents, institutions, ideas**
- **Question: Does it have mass or energy? Does it occupy space at some time? Then it is (probably) physical.**

Material vs Non-material Physical things

- **Within Physical_entities**
  - **The problem of holes**
    - **Material things define non-material things**
      - **The room defines the interior of the room**
      - **The glass defines the space in the glass**
      - **The donut defines the hole in the donut**
      - **The intersection of the walls defines the corner (a line)**

Discrete vs Mass

- **Things *are made of* Stuff**
  Discrete_entities are constituted of Mass_entities
  - **The statue vs the clay of which the statue is made**
  - **The liver vs the tissue that makes up the liver**
  - **The table top vs the wood that constitutes the table top**
- **Discrete things can be counted**
  Mass things can only be measured
  - **Guarino calls them "Amount of matter"**
    - **An instance of a mass stuff is an amount of that stuff**
- **Questions: Can I count it? then it is probably *discrete***

  If I make a plural, is it odd or something different?
  e.g. "waters", "papers", "thinkings", or do plurals
  mean different kinds e.g. "paints", "tissues"?
  do I say pieces/drops/lumps of it?
  then it is probably *mass*

Discrete vs Mass
Cognitivist vs Realist

- **Cognitivist**
  - **Two entities can occupy the same space and time**
    - **The clay is different from the statue**
      - **If I replace some of the clay, it is still the same statue**
      - **The properties of the clay are different from the properties of the statue**
      - **There is different *information* to be conveyed about the clay than there is to be conveyed about the statue**
- **Realist**
  - **In any one time-space extent, there can be exactly one physical entity**
    - **Different lumps of stuff are parts of it at different times**

Things have parts

- **A common pattern**
  - **Define the thing and a class for parts of the thing**
    - **Organ & Organ_part**
    - **Building & Building_part**
    - **Course & Course_part**
    - **Book & book part**
    - **…**
  - **Distinctions are usually derived from domain considerations rather than ontology**
    - **E.g. "organ" has a special meaning for (some) anatomists**

Complexes vs (Monads)

- **Complexes**
  - ***Aggregations***
    - **NOT mathematical sets**
      - **Entities where we are interested in the collective properties rather than the individual properties**
    - **No standard classification but ours is**
      - ***Group* – e.g. Flocks of geese, schools of fish, crowds…**
        Discrete collections of discrete things
      - ***Collective* – e.g. metal – atoms, tissue-cells,**
        Mass collections of discrete things
  - **Relations**
    - **Reified relations that bring two or more things together with specific roles or aspects**
      - **E.g. 'marriage', 'partnership', …**

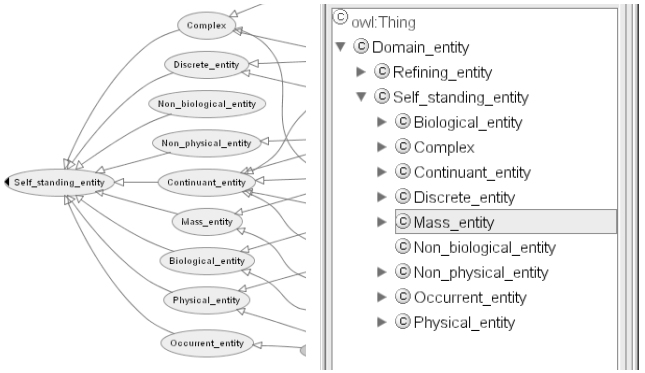
Granularity
Collective vs Individual

- **Collectives of discrete entities at one level form mass entities at the next**
  - **e.g. Collective of grains of sand is constituent of a beach**
    Collective of red cells are a portion of blood
    Collective of water molecules are a portion of water
    Collective 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**

Complexes vs (Monads)

- **Dangerous to say that anything is not a complex**
  - **Some things are definitely complexes**
    - **But almost anything can be viewed as a complex of some sort**

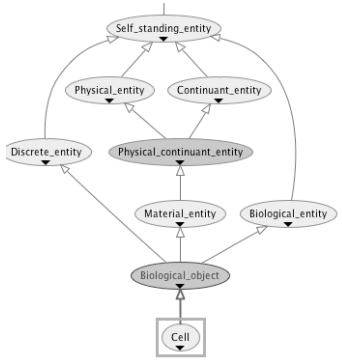
Basic Distinctions



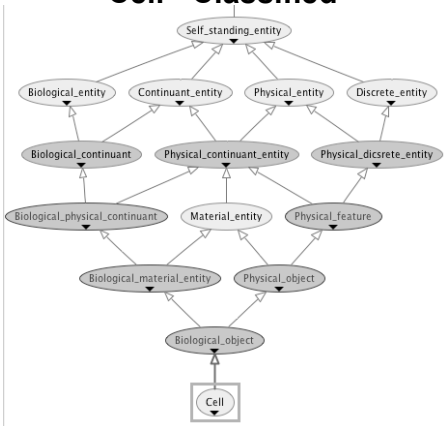
Unclassified Structure

Classified Structure
Looking from the top not always helpful

A better way to explore an ontology - Pick something and look at it from bottom:
A Cell - Unclassified



And its classification
Cell - Classified



Nonphysical entities
A real problem for for Librarians, Organisations & the law

- **What is "Hamlet"? What is "Lord of the Rings"?**
  - **The script for hamlet in the library?**
  - **The original folio?**
  - **A performance?**
  - **Can I own "Hamlet"?**
    Can I own "Lord of the Rings"?
    "A DVD of 'Lord of the Rings'"
    "The script to 'Lord of the Rings'"
    "A copy of the book 'Lord of the Rings'"
    "The first edition of the 'Lord of the Rings'"
    "A copy of the first edition of the 'Lord of the Rings'"

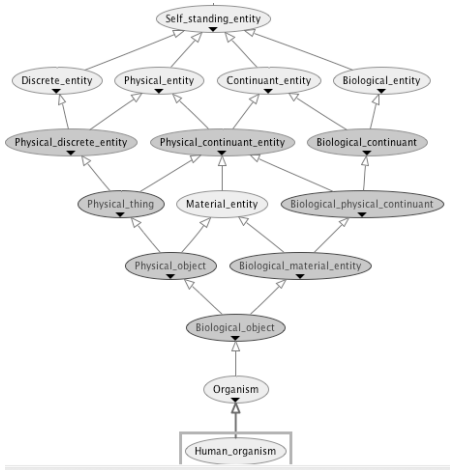
Agents and Actors

- Occurrents have actors
  - The actor in the *Process_of_erosion* is the *River*
  - The actor in the *Breaching_of_the_levy* is *Hurricane_Rita*
- Some actors are special and take responsibility and have legal status
  - We call them 'agents'
    - At least Person and Organisation
    - Possibly God, other animals, …
      - A good argument for the 'informationalist view'
        - » If there is information about it, then it is worth representing.
    - In this ontology - *Potential_agent* are things that can be agents.
      - 'Agent' is something which is an agent for some Occurrent.

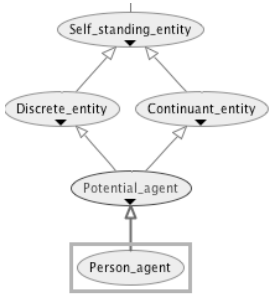
Agents – a problem for lawyers

- Is the agent Alan a different entity from Alan's Body?
- Who owns my body? Before death? After death?
  - In England, I do before death; my next of kin, after death, unless I am an executed felon,
  - but other jurisdictions have different laws
  - Hard to avoid dualism in legal ontologies!
    - When my body dies, 'I' cease to exist, but my body still exists, it is just dead
    - For the informationalist it
- Can animals be agents
  - In biology? In Law?

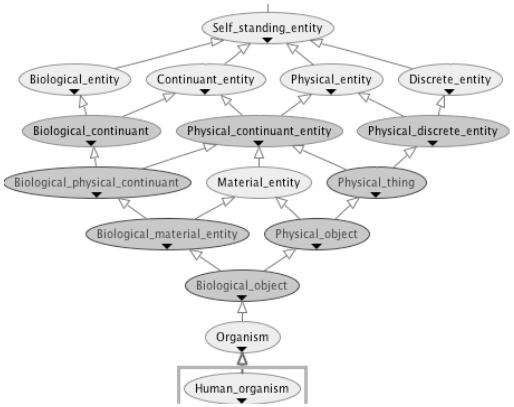
Human Organism (classified)



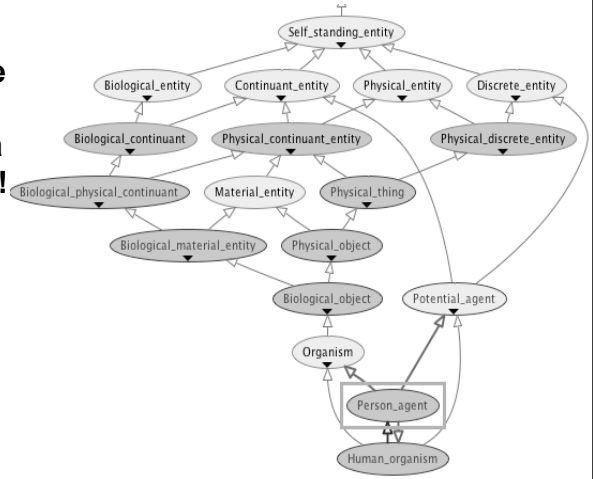
Person Agent (Classified)



Human_organism (classified)



If we equate them, what a tangle!



Acts

- **Occurrents whose actors are Agents are *Acts_***
  Use a special subproperty - *has_agent* - to avoid ambiguity
  - ***Acts* are the top level concepts in many management ontologies**
  - **A thorough study of responsibility, agency, and authorisation is a course for the business and law departments**

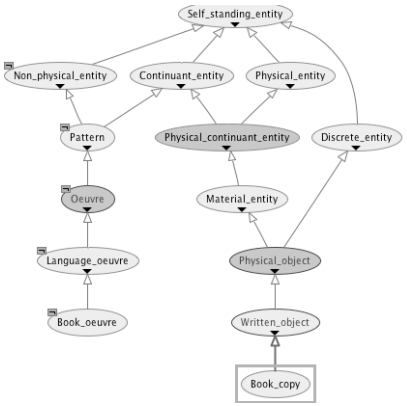
Artifacts

- **"Artifacts" are the physical outcome of "Acts"**
  - **Can have arguments about whether they must be things or can also be processes such as performances**
- **Some would divide the world into**
  - **Artifacts - the made world**
  - **Biological - the evolved world (& things derived from it)**
  - **Non-biological - the accumulated world**
    - **Awkward case: Coal**

"Oeuvres"

- The non-physical patterns of intellectual work
  - Patterns that are the outcome of Acts by Agents
  - Hamlet is the outcome of an act of playwriting by Shakespeare
  - This copy of Hamlet is the result of an act of printing by Oxford University Press
    - And is a manifestation of the Oeuvre "Hamlet"
  - This performance of Hamlet is the result of an Act of Performance by the Royal Exchange Theatre Company
    - And is also a manifestation of the Oeuvre "Hamlet"

Book_oeuvre & Book_copy before classification



After classification

OntoClean / Upper & Domain Ontologies

- OntoClean is a meta-ontology for checking whether an ontology is sensible
  - DOLCE is an upper ontology built on the principles of OntoClean
- Rigid vs Non-rigid properties
  - Rigid predicates never change through the life of the entity
    - Many rigid predicates difficult to pin down
      - What makes a horse a horse? A dog a dog?
  - OntoClean principles as applied to OWL
    - Distinctions between primitive classes should be based on rigid predicates (Aristotle's differentia)
      - When normalised, the trees should be based on rigid predicates
        - » All non-rigid predicates should be represented by restrictions
        - » All classification based on non-rigid properties should be performed by the classifier

Now we can say why this looks odd

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

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

"Twenty questions" Example: What is an Organelle?

(The small organs inside cells – mitochondria, chlorplasts, etc)

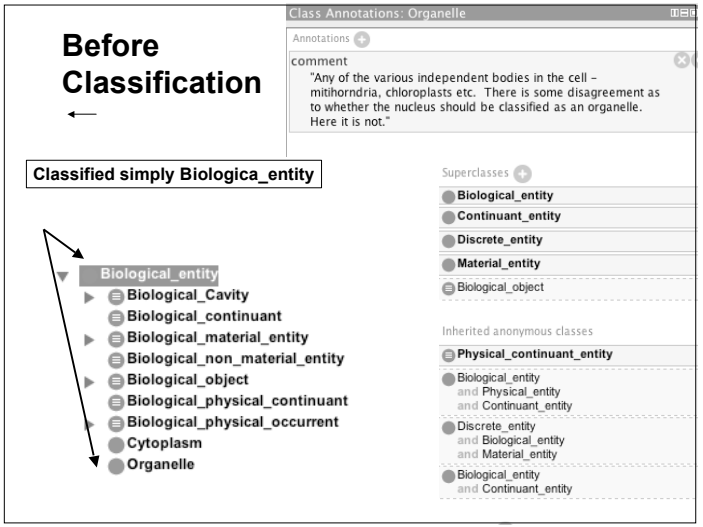
- Is it Continuant or Occurrent? *Continuant*
  - Does it happen or do things happen to it?
- Is it physical? *yes*
- Is it Discrete or mass? *Discrete*
  - (Can you count it?)
- If physical & discrete, Is it material or non-material (thing or hole)? *Material*
- Is it Biological? *yes*

Further questions

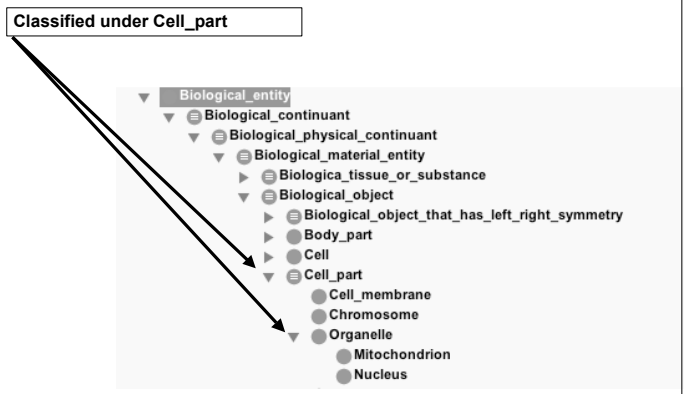
- Is it part of something? *yes*
  - if so, definite number or not? *no*
    - Collectives of Organels are part of Cytoplasm`
- Therefore, it is a "Cell_part" (a subclass of Biological_object)

Before Classification

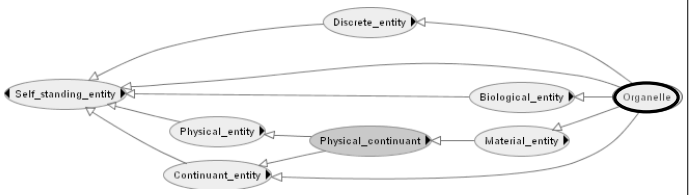
←



After Classification



Before Classification



After Classification

"Twenty Questions"
Cytoplasm (the substance that fills the cells)

- **Is it Continuant or Occurrent?** ***Continuant***
- **Is it physical? yes. Is it material?** ***yes, yes***
- **Is it discrete or mass?** ***mass***
- **Is it biological?** ***yes***

- **Then it must be a "Tissue_or_substance"**

What is Digestion

- **Is it Continuant or occurrent? -** ***occurrent***
- **Is it physical? -** ***yes***
- **Is it discrete or mass?** ***? defer***
- **Is it biological?** ***yes***
- **If so is it pathological** ***no***

- **Then it must be a "Biological_physical_occurrent"**
  - **Name chosen deliberately to defer mass/discrete choice**

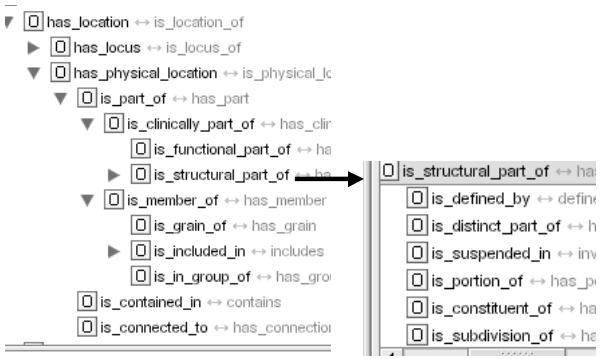
The Properties Hierarchy

- **Basic meaning analogous to classes**
  - ***p***
    p_sub
    p_sub_sub
    - **Anything linked by *p_sub_sub* is linked by *p_sub*;**
      Anything linked by *p_sub* is linked by *p*
    - **For all xy . x *p_sub_sub* y → x *p_sub* y →** x *p* y
    - ***p_sub_sub* SOME C→*p_sub* SOME C → *p* SOME C**
- **A powerful means of inference used in, amongst other things:**
  - **Part-whole relations**
  - **"Participations" in processes**
  - **"Views"**
    - **allowing different applications to see different aspects of a property**
  - **Lots of work arounds**
    - **Transitive property with a non-transitive subproperty**

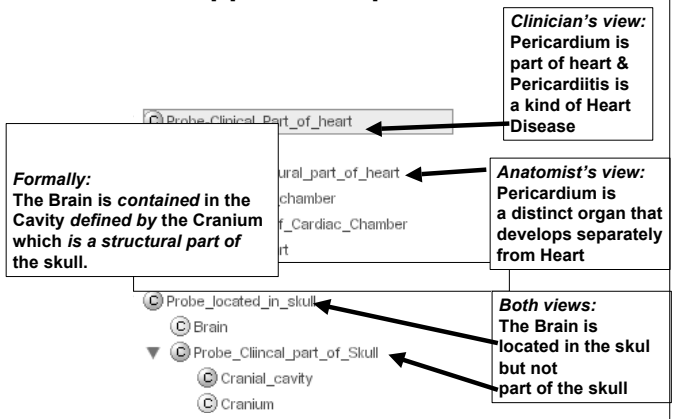
Views and the Property Hierarchy

- **The property hierarchy is as important as the class hierarchy**
  - **E.g. For different flavours of part of, containment, etc.**
  - **For direct variants of transitive relations**
  - **For many other inferences**
    - **Can sometimes get around the lack of variables**

Consider the part-of hierarchy



Sufficient to support multiple "views"



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 "collectives"
- Anything to do with agents

Controversies: How to argue?

- Evidence is effect on representation
  The test is 'faithful communication'
  - 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

OntoClean & Dolce One Upper Ontology

- Owl version Provided in the lab – see also URL
  - http://www.loa-cnr.it/DOLCE.html
  - Vocabulary:
    - "Predicate" – "Class"
      - i.e. a Class is equivalent to a one-place predicate
        - » the Class 'C' is equivalent to the predicate C(x)
    - Sortal – "Self-standing entity"
      - To a good first approximation
    - "Amount of matter" - "Mass_entity"
  - OntoClean is a meta ontology & methodology for ontology building
    - An ontology about the properties of concepts
    - used to constrain
    - DOLCE is an upper ontology that conforms to Ontoclean

Alternative Upper Ontology: BFO

Basic Formal Ontology
- http://www.ifomis.uni-saarland.de/bfo/

– Vocabulary
  - "Universal" - "Class"
    - (with some arguments)
  - "Relation" - "Property"
  - "Fiat" vs "Bona fide" boundaries & entities
    - High level division between "natural" and human imposed
– Commitments
  - All universals should have a (direct) instance
  - "Granularity"
    - Each ontology expected to be at a specific level of granulity
  - Uses "realist" view that rejects the notion of "Amount of Substance" and "constitutes" relation.
    - "The statue *is* the amount of clay" rather than "The statue is constiuted of the amount of clay"
  - Used by OboFoundry -
    - http://obofoundry.org/wiki/index.php/Main_Page
    - Extensions and ontologies surround the Gene Ontology