Background

- Working in ontology since 1975, with bio-ontologists and clinical ontologists since 2002
- Working with Gene Ontology since 2004
- Co-PI of the Protein Ontology (NIH/NIGMS)
- Coordinating Editor of the OBO (Open Biomedical Ontologies) Foundry

NCBO

- Dissemination and Ontology Best Practices of the National Center for Biomedical Ontology (PI Mark Musen, Stanford)
- http://bioontology.org

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Example ontologies

Basic Formal Ontology (BFO)
Common Anatomy Reference Ontology (CARO)
Environment Ontology (EnvO)
Foundational Model of Anatomy (FMA)
Infectious Disease Ontology (IDO)
Ontology for Biomedical Investigations (OBI)
Ontology for Clinical Investigations (OCI)
Phenotypic Quality Ontology (PATO)
Relation Ontology (RO)

Ontologies and terminologies examined

SNOMED
Unified Medical Language System
National Cancer Institute Thesaurus
HL7 Reference Information Model
International Classification of Functioning, Disability and Health

Collaborations

Cleveland Clinic Semantic Database for Cardiovascular Surgery Ontology
Duke University Laboratory of Computational Immunology
German Federal Ministry of Heath
European Union Emergency Patient Summary Initiative
University of Pittsburgh Medical Center
University of Texas Southwestern Medical Center

Collaborations (Brain and Behavior)

UB Task Force for Ontology-Based IT Support for Large-Scale Field Studies in Psychiatry

Jacobs Neurological Institute, University at Buffalo

Ontology Task Force (San Diego) of the Biomedical Informatics Research Network (BIRN)

Neurocommons/Science Commons (MIT)

Agenda • Day 1

- **Introduction: What is an ontology and what is it useful for?**
- Basic Formal Ontology: An upper-level ontology to support scientific research
- Open Biomedical Ontologies (OBO) and the Web Ontology Language (OWL)
- The OBO Relation Ontology

Multiple kinds of data in multiple kinds of silos

Lab / pathology data
Electronic Health Record data
Clinical trial data
Patient histories
Medical imaging
Microarray data
Protein chip data
Flow cytometry
Mass spec
Genotype / SNP data

How to *find* your data?

How to reason with data when you find it?

How to understand the significance of the data *you* collected 3 years earlier?

How to integrate with other people's data?

Part of the solution must involve consensus-based, standardized terminologies and coding schemes

Ontologies facilitate retrieval of data

by allowing grouping of annotations

brain	**20**
hindbrain	**15**
rhombomere	**10**

Query brain without ontology **20**
Query brain with ontology **45**

Making data (re-)usable through standards

- Standards provide
  - common structure and terminology
  - single data source for review (less redundant data)
- Standards allow
  - use of common tools and techniques
  - common training
  - single validation of data

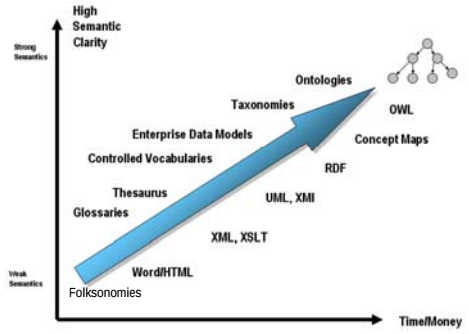
Unifying goal: integration

- within and across domains
- across different species
- across levels of granularity (organ, organism, cell, molecule)
- across different perspectives (physical, biological, clinical)

Problems with standards

- Standards involve considerable costs of re-tooling, maintenance, training, ...
- They pose risks to flexibility
- May break legacy solutions which work locally
- Not all standards are of equal quality
- Bad standards create lasting problems
- 'Ontology' = good standards in terminology

Leo Obrst: The Ontology Spectrum



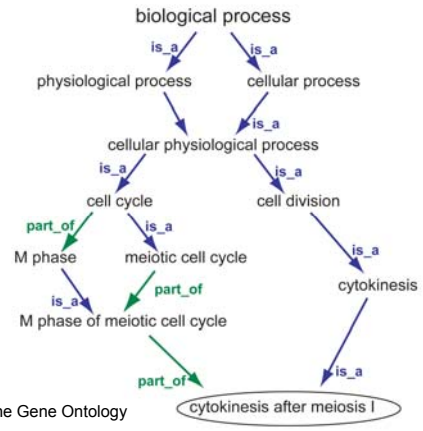
http://www.mkbergman.com/?m=20070516 (slightly modified)

The wisdom of clouds (folksonomies ...)

07 2007 2008 51207 51507 51607 51707 academy accs andy blog centennial ceremony charlottesville coffee collabratory commencement compression concert conference datlat digital diskutility dtlt dv dvd encoders engl375mm eve6 fa07 fa08 faculty facultyacademy ffmpegx foucault fredericksburg fredericksburgnormalandindustrialinstitute freshman fsem100j globalization grad graduation gravatar greenlaw groom header historic historical history homecoming img jim jimschair js kenmore learning marker markers mary mashup mashups microsoft nmc nmc2007 ontology patrick popfly presentation reverend reverendjimtshirt rush sa07 sa2007 semantic seminar student studentacademy symposium teaching technologies transfer tshirt umw umwblogs umwead university video virginia virginiahistoricalmarkers visualizing wall

Ontologies are, at least, controlled structured vocabularies

providing definitions and reasoning including support for automatic validation of ontology structure



from the Gene Ontology

NIH Mandates for Sharing of Research Data

Investigators submitting an NIH application seeking $500,000 or more in any single year are expected to include a *plan for data sharing*

(http://grants.nih.gov/grants/policy/data_sharing)

Program Announcement Number: PAR-07-425

Title: Data Ontologies for Biomedical Research (R01)

NIH Blueprint for Neuroscience Research, (http://neuroscienceblueprint.nih.gov/)
National Cancer Institute (NCI), (http://www.cancer.gov)
National Center for Research Resources (NCRR), (http://www.ncrr.nih.gov/)
National Eye Institute (NEI), (http://www.nei.nih.gov/)
National Heart Lung and Blood Institute (NHLBI), (http://http.nhlbi.nih.gov)
National Human Genome Research Institute (NHGRI), (http://www.genome.gov)
National Institute on Alcohol Abuse and Alcoholism (NIAAA), (http://www.niaaa.nih.gov/)
National Institute of Biomedical Imaging and Bioengineering (NIBIB), (http://www.nibib.nih.gov/)
National Institute of Child Health and Human Development (NICHD), (http://www.nichd.nih.gov/)
National Institute on Drug Abuse (NIDA), (http://www.nida.nih.gov/)
National Institute of Environmental Health Sciences (NIEHS), (http://www.niehs.nih.gov/)
National Institute of General Medical Sciences (NIGMS), (http://www.nigms.nih.gov/)
National Institute of Mental Health (NIMH), (http://www.nimh.nih.gov/)
National Institute of Neurological Disorders and Stroke (NINDS), (http://www.ninds.nih.gov/)
National Institute of Nursing Research (NINR), (http://www.ninr.nih.gov)

PAR-07-425 Purpose

Optimal use of informatics tools and data resources depends upon explicit understandings of concepts related to the data upon which they compute. This is typically accomplished by a tool or resource adopting a formal controlled vocabulary and ontology.

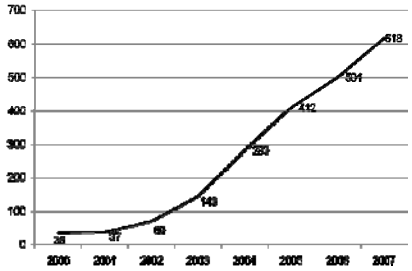
Currently, there is no convenient way to map the knowledge that is contained in one data set to that in another data set, primarily because of differences in language and structure

... in some areas there are emerging standards.

Examples include:

- the Unified Medical Language System (**UMLS**),
- the **Gene Ontology**,
- the **caBIG** project,
- Open Biomedical Ontologies (**OBO**)

Uses of 'ontology' in PubMed abstracts



Types of ontologies

	Upper-level integrating ontologies	Domain ontologies
Ontologies in support of science		
Administrative ontologies		

Types of ontologies

	Upper-level integrating ontologies	Domain ontologies
Ontologies in support of science	*BFO (Basic Formal Ontology)* *DOLCE, SUMO*	*GO* *FMA* *SNOMED*
Administrative ontologies (e-commerce, etc.)	*FOAF top level: person, topic, document, primary topic ...*	*Amazon.com ontology* *Library of Congress Catalog*

Scientific ontologies vs. administrative ontologies

BFO, GO, FMA …

vs.

Library of Congress Catalog, Yahoo ontology, FirstGov Life Events Taxonomy, …

Part of our goal is realized if we can create controlled terminologies

In science we can and must go further than this

Why build scientific ontologies?

There are many ways to create terminologies

Multiple terminologies will not solve our data silo problems

We need to constrain terminologies so that they converge

Evidence-based terminology development

Q: What is to serve as constraint?

A1: Authority?

A2: First in field (Founder effect)?

A3: Best candidate terminology?

A4: Reality, as revealed, incrementally, by experimentally-based science

The standard methodology

- Pragmatics is everything
- It is easier to write useful software if one works with a simplified model
- ("…we can't know what reality is like in any case; we only have our concepts…")
- This looks like a useful model to me
- (One week goes by:) This other thing looks like a useful model to him
- Data in Pittsburgh does not interoperate with data in Vancouver
- Science is siloed

The methodology of ontological realism

- Find out what the world is like by doing science, talking to other scientists and working continuously with them to ensure that you don't go wrong
- Build representations adequate to this world, not to some simplified model in your laptop
- Ontology is ineluctably a multi-disciplinary enterprise – need to work hard to overcome the resultant terminological confusions

Our first job is in to create a common understanding of terms such as:

- universal, type, kind, class
- instance
- model
- representation
- data

Entity =def

anything which exists, including things and processes, functions and qualities, beliefs and actions, documents and software

Scientific ontologies have special features

Every term must be such that the developers of the ontology believe it to refer to some entity on the basis of the best current scientific evidence

(Important role of instances that we can observe in the laboratory)

Administrative ontologies

- Entities may be brought into existence by the ontology itself. (Convention ...)
- Highly task-dependent – reusability and compatibility not (always) important
- Can be secret
- Are comparable to software artifacts

For scientific ontologies

openness, reusability and compatibility with neighboring scientific ontologies are crucial

- Scientific ontologies must evolve gracefully
- Scientific ontologies must be evidence-based
- Scientific ontologies are comparable to scientific theories

The central distinction

***universal* vs. instance**

(*catalog* vs. inventory)

(*science text* vs. diary)

(*human being* vs. Arnold Schwarzenegger)

Science texts are representations of universals in reality

= representations of what is *general* in reality

Clinical guidelines are representations of universals in reality

diseases, therapies, diagnostic procedures (measurements) are generals, with particular instances in particular patients

Ontologies are representations of universals in reality

aka kinds, types, categories, species, genera, ...

A	515287	**DC3300 Dust Collector Fan**
B	521683	**Gilmer Belt**
C	521682	**Motor Drive Belt**

universals

Catalog vs. inventory

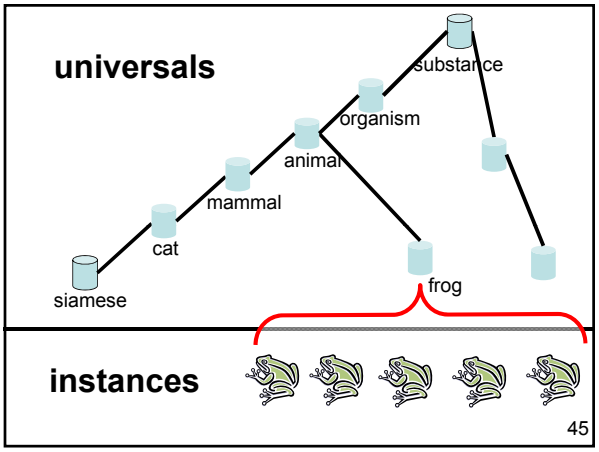
For scientific ontologies

it is *generalizations* (universals) that are important

For databases it is (normally) instances that are important

= particulars in reality:

- patient #0000000001
- headache #000000004
- MRI image #23300014, etc.



In a scientific ontology

every node in the ontology should represent both universals *and* the corresponding instances in reality

every term should reflect instances – it is instances which form the objects of our experiments

to do this is hard work …

Each term in an ontology represents exactly one universal

For this reason ontology terms should be **singular nouns**

headache

human being

drug administration

An ontology is a representation of universals

We learn about universals in reality from looking at the results of scientific experiments as expressed in the form of scientific theories – which describe, not what is particular in reality, but what is general

A photographic image is a representation of an instance

Three Levels to Keep Straight

- Level 1: the entities in reality, both instances and universals
- Level 2: cognitive representations of this reality on the part of scientists ...
- Level 3: publicly accessible concretizations of these cognitive representations in textual and graphical artifacts

Ontology development

starts with: Level 2 = the cognitive representations of practitioners or researchers in the relevant domain

results in: Level 3 representational artifacts (comparable to maps, science texts, dictionaries)

Domain =def.

a portion of reality that forms the subject-matter of a single science or technology or mode of study;

proteomics
HIV
demographics
...

Representation =def.

an image, idea, map, picture, name or description ... of some entity or entities

two kinds of representation:

analogue (photographs)
digital/composite/syntactically structured

Representational units =def

terms, icons, alphanumeric identifiers ... which refer, or are intended to refer, to entities

and which are minimal ('atoms')

Composite representation =def

a representation

(1) built out of representational units

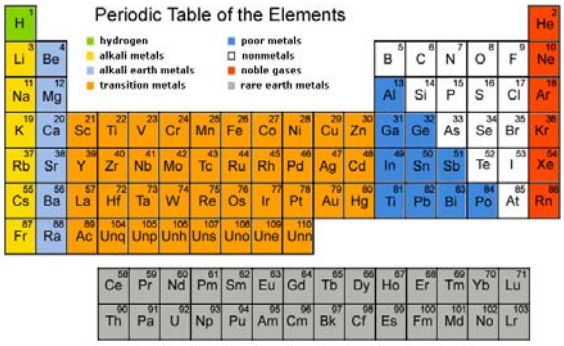
which

(2) form a structure that mirrors, or is intended to mirror, the entities in some domain

Analogue representations

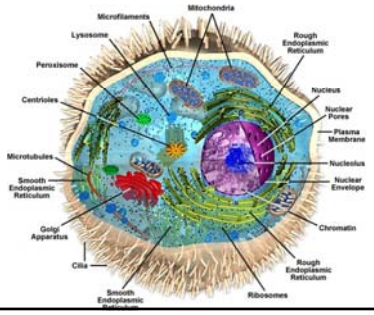
The Periodic Table

H 1	Periodic Table															He 2
Li 3	Be 4											B 5	C 6	N 7	O 8	F 9	Ne 10
Na 11	Mg 12											Al 13	Si 14	P 15	S 16	Cl 17	Ar 18
K 19	Ca 20	Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36
Rb 37	Sr 38	Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54
Cs 55	Ba 56	La 57	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
Fr 87	Ra 88	Ac 89	Rf 104	Ha 105	?? 106												
Lanthanide Series		Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70	Lu 71		
Actinide Series		Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96	Bk 97	Cf 98	Es 99	Fm 100	Md 101	No 102	Lr 103		



We can't take photographs of universals

But we can create cartoons and diagrams

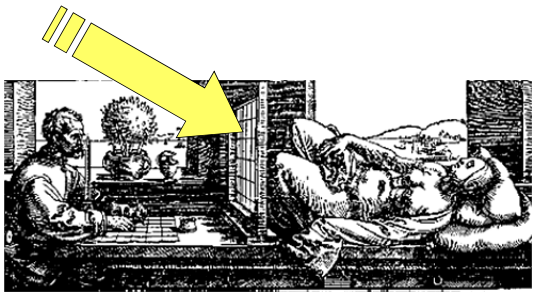


Cognitive representations

Representational artifacts

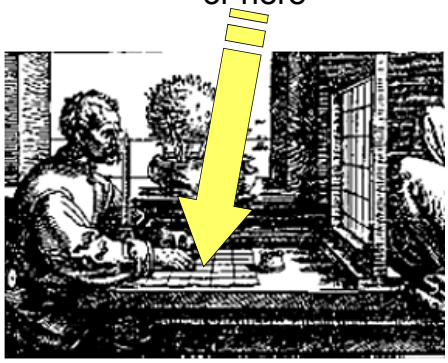
Reality

Ontologies are here



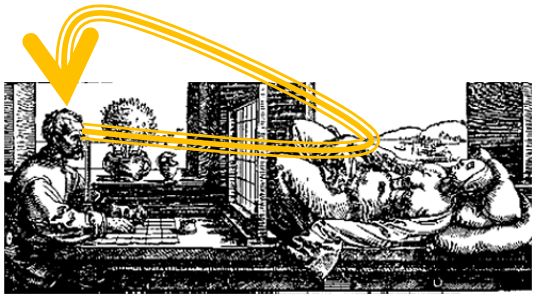
61

or here



62

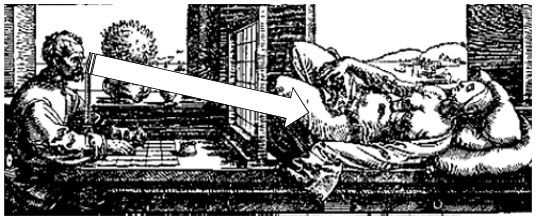
Ontologies do *not* represent concepts in people's heads



63

Like the scientific theories from which they derive, they represent universals in reality

e.g. *leg*



64

Compare the typical relations used in medical ontologies

part_of
connected_to
adjacent_to
causes
treats ...

65

"leg" is not the name of a concept

concepts do not stand in
- part_of
- connected_to
- adjacent_to
- causes
- treats ...

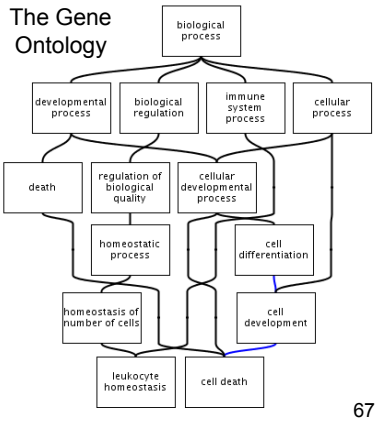
relations to each other

66

The Gene Ontology

is_a —

part_of —



How do we know which general terms designate universals?

Roughly: terms used in a plurality of sciences to designate entities about which we have a plurality of different kinds of testable propositions / laws

(compare: cell, electron, membrane ...)

Class =def.

a maximal collection of particulars referred to by a general term

the class *A* =def. the collection of all particular *A*'s

where '*A*' is a general term (e.g. 'brother of Elvis fan', 'cell')

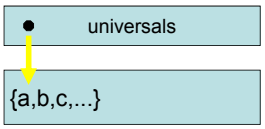
Classes are on the same level as the instances which they contain

Extension =def

the collection of all particular *A*'s, where '*A*' is the name of a universal

universals vs. their extensions

The extension of the *universal A* is the class of *A's* instances



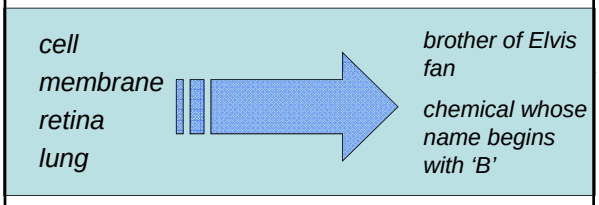
collections of particulars

Problem

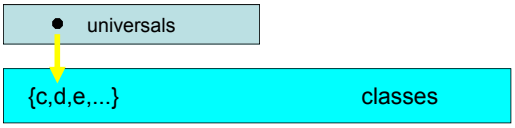
The same general term can be used to refer both to universals and to collections of particulars.

HIV is an infectious retrovirus
HIV is spreading very rapidly through Asia

a spectrum of cases



Not all classes correspond to universals



Administrative ontologies often go beyond universals

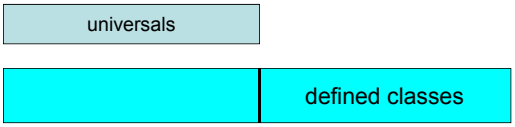
Fall on stairs or ladders in water transport injuring occupant of small boat, unpowered

Railway accident involving collision with rolling stock and injuring pedal cyclist

Non-traffic accident involving motor-driven snow vehicle injuring pedestrian

ICD (WHO International Classification of Diseases)

universals vs. classes



Defined class =def

a class defined by a general term which does not designate a universal

person called 'Chris'

person with diabetes in Maryland on 4 June 1952

OWL (Ontology Web Language) is a good representation of defined classes

sibling of Finnish spy

member of Abba aged > 50 years

property-owning farm employee

such set-theoretic combinations are at the heart of many administrative ontologies

(Scientific) Ontology =def.

a representational artifact whose representational units (which may be drawn from a natural or from some formalized language) are intended to represent

1. universals in reality
2. those relations between these universals which obtain universally (= for all instances)

*lung **is_a** anatomical structure*

*lobe of lung **part_of** lung*

Agenda • Day 1

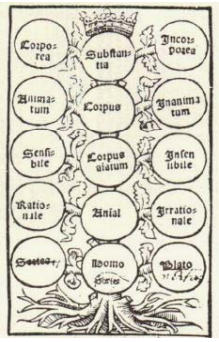
- Introduction: What is an ontology and what is it useful for?
- **Basic Formal Ontology: An upper-level ontology to support scientific research**
- Open Biomedical Ontologies (OBO) and the Web Ontology Language (OWL)
- The OBO Relation Ontology

Ontology

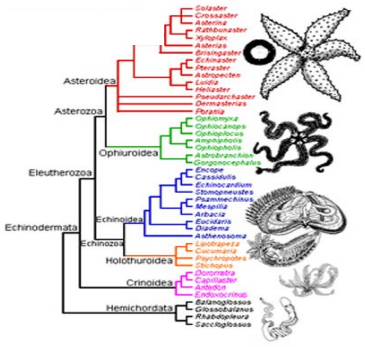
the science of the kinds and structures of objects, properties, events, processes and relations in every domain of reality

World's first ontology

(from Porphyry's *Commentary* on Aristotle's *Categories*)



Linnaean Ontology



Contemporary top-level ontologies

DOLCE = Domain Ontology for Linguistic and Cognitive Engineering

SUMO = Suggested Upper Merged Ontology

BFO = Basic Formal Ontology

Each of these ontologies

is not just a system of categories

but a *formal theory*

with definitions, axioms, theorems

designed to provide the resources for reference ontologies built to represent the entities in specific domains

of sufficient richness that terminological incompatibilities can be resolved intelligently rather than by brute force

BFO is a very small ontology to support integration of scientific research data

SUMO contains many portions which are more properly conceived of as domain ontologies (airports, bacteria, ...)

DOLCE is tilted towards objects of general thought and communication (fiction, mythology, ...)

Basic Formal Ontology

- a true upper level ontology
- no interference with domain ontologies
- no interference with issues of cognition
- no negative entities
- no putative fictions
- a small subset of DOLCE but with more adequate treatment of instances, universals, relations and qualities

http://www.ifomis.org/bfo/

Groups and Organizations using BFO

AstraZeneca - Clinical Information Science
BioPAX-OBO
BIRN Ontology Task Force (BIRN OTF)
Computer Task Group Inc.
Duke University Laboratory of Computational Immunology
Dumontier Lab
INRIA Lorraine Research Unit
Kobe University Graduate School of Medicine
Language and Computing
National Center for Multi-Source Information Fusion
Ontology Works
Science Commons - Neurocommons
University of Texas Southwestern Medical Center

Ontologies using BFO

BioTop: A Biomedical Top-Domain Ontology
Common Anatomy Reference Ontology (CARO)
Foundational Model of Anatomy (FMA)
Gene Ontology (GO)
Infectious Disease Ontology
Ontology for Biomedical Investigations (OBI
Ontology for Clinical Investigations (OCI)
Phenotypic Quality Ontology (PaTO)
Protein Ontology (PRO)
RNA Ontology (RnaO)
Senselab Ontology
Sequence Ontology (SO)
Subcellular Anatomy Ontology (SAO)
Vaccine Ontology (VO)

Realist Perspectivalism: The philosophical basis of BFO

There is a multiplicity of ontological perspectives on reality, all equally veridical i.e. transparent to reality

Ontologies are windows on reality

The Time Problem

The tumor developed in John's lung over 25 years

The Problem

____ developed in _____ over 25 years

process

state

The Problem

The tumor developed in the lung over 25 years

substances

things

objects

continuants

The Problem

The tumor developed in the lung over 25 years

what is it that participates in this process of tumor development?

parthood here not determinate

The Problem

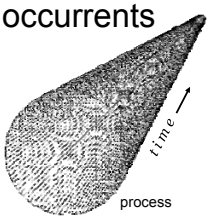
The tumor developed in the lung over 25 years

substances

process

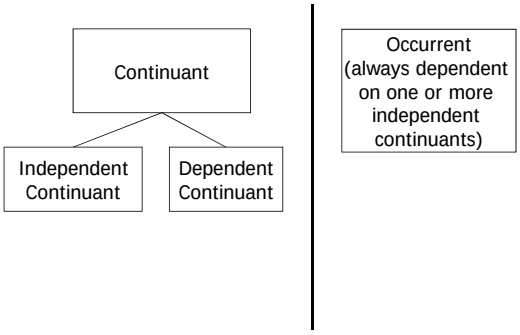
gluing these two types of entities together yields ontological monsters

Continuants vs occurrents



In preparing an inventory of reality we keep track of these two different kinds of entities in two different ways

BFO: the very top



An alternative approach: Fourdimensionalism

- only processes (occurrents) exist
- time is just another dimension, analogous to the three spatial dimensions
- substances are analyzed away as worms/fibers within the four-dimensional plenum
- fourdimensionalism brings benefits especially for computational purposes

There are no substances

Bill Clinton does not exist

Rather: there exists within the four-dimensional plenum a continuous succession of processes which are similar in a Billclintonizing way

Fourdimensionalism (“everything is flow”) is right in everything it says

But it is incomplete

Realist Perspectivalism

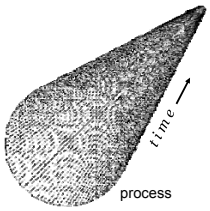
There is a multiplicity of ontological perspectives on reality, all equally veridical = transparent to reality

Fourdimensionalism is one veridical perspective among others

Cf. particle vs. wave ontologies used in quantum mechanics

Snapshot ontology

Video ontology



Continuants and Occurrents

Two Orthogonal, Complementary Perspectives

stocks and **flows**
commodities and **services**
product and **process**

anatomy and **physiology**

Continuant entities
- have continuous existence in time
- preserve their identity through change
- exist *in toto* if they exist at all

Occurrent entities
- have temporal parts
- unfold themselves phase by phase
- exist only in their phases/stages

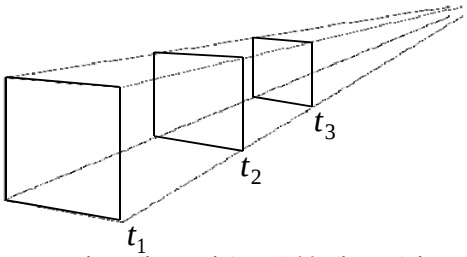
You are a substance

Your *life* is a process

***You* are 3-dimensional**

Your *life* is 4-dimensional

Many SNAPshot Ontologies



here time exists outside the ontology, as an *index* or *time-stamp*

mereology works without restriction (parthood is everywhere determinate) in every SNAP ontology

Note that, while, the *views* are instantaneous, the *objects viewed* endure

Three kinds of continuant entities

1. Substances (Independent)
2. SNAP Dependent Entities
3. Spatial regions, contexts, niches, environments, settings

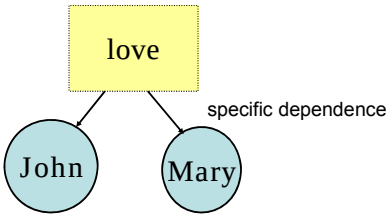
Dependent continuants:

one-place:

your temperature, color, height
my knowledge of French
the whiteness of this cheese

relational dependent continuants

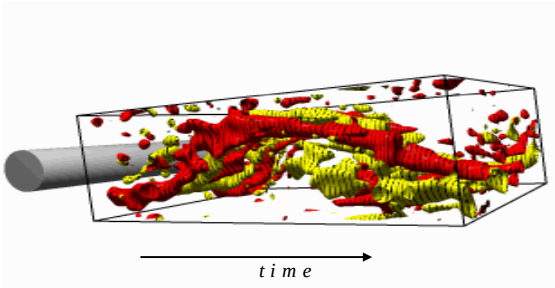
stand in relations of one-sided dependence to a plurality of substances simultaneously



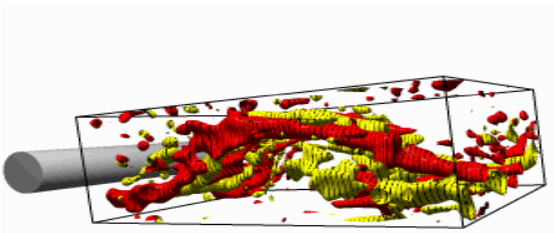
Dependent continuants

Functions, qualities, roles …
dispositions, plans, shapes, diseases ...

The world of processes

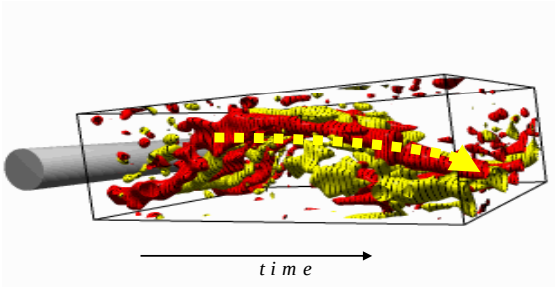


Occurrents



here time exists as part of the domain of the ontology

mereology works without restriction everywhere here and boundaries are mostly fiat



Processes, too, are dependent on substances

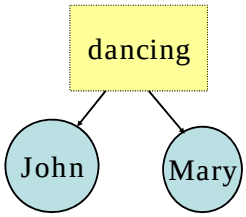
One-place vs. relational processes

One-place processes:
- your getting warmer
- your getting hungrier

Examples of relational processes

kissings, thumpings, conversings, dancings, …

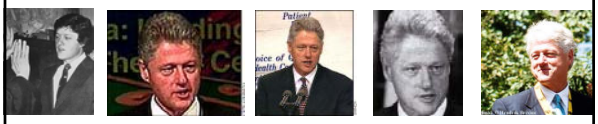
join their carriers together into *collectives* of greater or lesser duration



Two kinds of occurrent entities

1. Processes (including events, beginnings, endings = process-boundaries)
2. Spatio-temporal regions

How do you know whether an entity is a continuant or an occurrent?



problem cases

forest fire
the Olympic flame
epidemic
hurricane
traffic jam
ocean wave

forest fire

a process

a pack of monkeys jumping from tree to tree and eating up the trees as they go

(anthrax spores are little monkeys)

The Epidemic (Continuant)

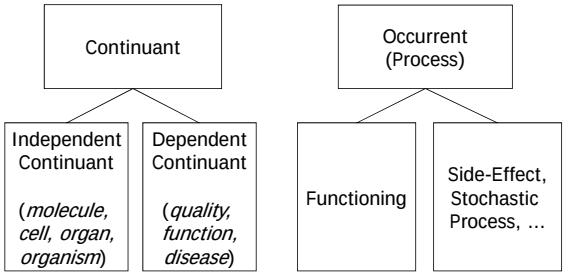
The Spread of an Epidemic (Occurrent)

Three dichotomies

- instance vs. universal
- continuant vs. occurrent
- dependent vs. independent

- universals exist in reality *through their instances*

BFO



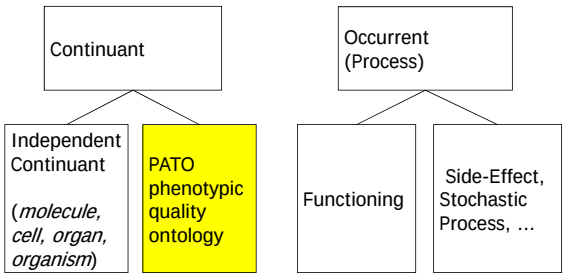
BFO

all terms included in the ontology are intended to designate universals in reality

in conformity with the basic principle of science-based ontology

science-based ontologies are windows on reality

Phenotype Ontology



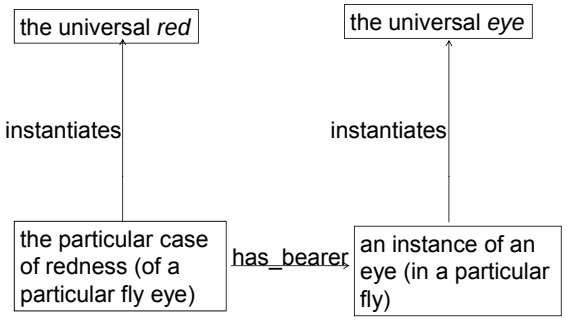
An example of a quality

- The particular redness of the left eye of a single individual fly
  - An *instance* of a quality universal
- The color 'red'
  - A quality universal
- Note: the eye does not instantiate 'red'
- PATO represents **quality universals:** color, temperature, texture, shape …

Qualities are *dependent* entities

- Qualities require (depend on) *bearers,* which are independent continuants

  Example:
  - A shape requires a physical object as its bearer
  - If the physical object ceases to exist (e.g. it decomposes), then the shape ceases to exist



What a quality is NOT

- Qualities are not measurements
  - Instances of qualities exist independently of their measurements
  - Qualities can have zero or more measurements
- These are not the names of qualities:
  - percentage
  - process
  - abnormal
  - high
- Open problem: how relate qualities such as length to measurement values?

Agenda • Day 1

- Introduction: What is an ontology and what is it useful for?
- Basic Formal Ontology: An upper-level ontology to support scientific research
- **Open Biomedical Ontologies (OBO) and the Web Ontology Language (OWL)**
- The OBO Relation Ontology

Gene Ontology

constructed in 1998 by researchers studying the genome of three model organisms: Drosophila melanogaster (fruit fly), Mus musculus (mouse), and Saccharomyces cerevisiae (brewers' or bakers' yeast)

developed its own flat-file (GO-)format

OBO (Open Biomedical Ontologies)

created 2001 in Ashburner and Lewis

a shared portal for (so far) 58 ontologies
http://obo.sourceforge.net
with a common OBO flatfile format

Animal natural history and life history	ADW	protege source
Arabidopsis development	TAIR	arabidopsis development.obo
Arabidopsis gross anatomy	TAIR	po anatomy.obo
Biological imaging methods	FBbi	image.obo
Biological process	GO	gene ontology.obo
BRENDA tissue / enzyme source	BTO	BrendaTissue.obo
C. elegans development	WBls	worm development.obo
C. elegans gross anatomy	[none]	[none]
Cell type	CL	cell.obo
Cellular component	GO	gene ontology.obo
Cereal plant development	GRO	cereals development.obo
Cereal plant gross anatomy	GRO	po anatomy.obo
Cereal plant trait	TO	plant trait.obo
Chemical entities of biological interest	CHEBI	chebi.obo
Dictyostelium discoideum anatomy	DDANAT	dictyostelium anatomy.obo
Drosophila development	FBdv	fly development.obo
Drosophila gross anatomy	FBbt	fly anatomy.obo
Event (INOH pathway ontology)	IEV	event.obo
Evidence codes	ECO	evidence code.obo
eVOC (Expressed Sequence Annotation for Humans)	EV	evoc.obo.tar (v2.7)

OBO builds on the principles successfully implemented by the GO

- ontologies should be
- open
- orthogonal
- instantiated in a well-specified syntax
- designed to share a common space of identifiers

In 2004 reform efforts initiated, e.g. linking GO formally to other ontologies and data sources

GO term: **osteoblast differentiation**
Synonym: **osteoblast cell differentiation**
GO id: **GO:0001649**
Definition: **Processes whereby a relatively unspecialized cell acquires the specialized features of an osteoblast, the mesodermal cell that gives rise to bone.**

id: CL:0000062
name: osteoblast
def: "A bone-forming cell which secretes an extracellular matrix. Hydroxyapatite crystals are then deposited into the matrix to form bone."
is_a: CL:0000055
relationship: develops_from CL:0000008
relationship: develops_from CL:0000375

Osteoblast differentiation: Processes whereby an osteoprogenitor cell or a cranial neural crest cell acquires the specialized features of an osteoblast, a bone-forming cell which secretes extracellular matrix.

GO

+

Cell type

=

New Definition

The methodology of cross-products

compound terms in ontologies to be defined as cross-products of simpler terms:

E.g *elevated blood glucose* is a cross-product of PATO: *increased concentration* with FMA: *blood* and CheBI: *glucose*.

= factoring out of ontologies into discipline-specific modules (orthogonality)

The methodology of cross-products

enforcing use of RO in linking terms drawn from Foundry ontologies serves

to reduce arbitrariness and ambiguity which marks existing approaches to post-composition

makes the results of post-composition algorithmically processable in virtue of the logical definitions provided by the RO

General Principle for Definitions

Definitions should be of the following form

an A =def. a B which Cs

where B is the is_a parent of A and C is some differentia

General Principle for Definitions

Definitions specify necessary and sufficient conditions. Thus if we are defining what it is to be an A, then the definition might read, for example:

x is an A =def. x is a B which has features F1, F2, F3.

This definition would be correct if and only if every B which has features F1, F2, and F3 is an A, and everything which is an A has features F1, F2, and F3.

Accessing Ontologies

Ontology Lookup Service

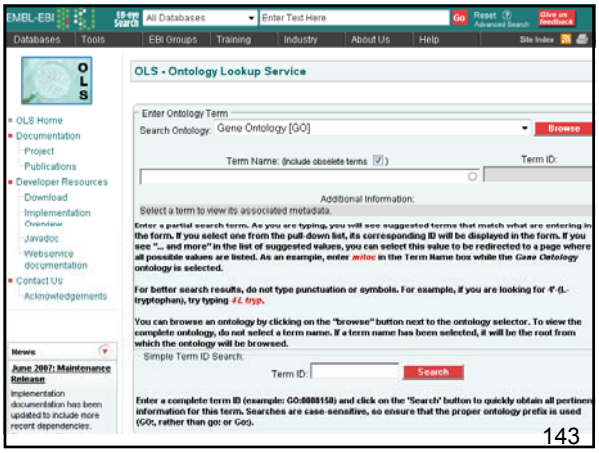
www.ebi.ac.uk/ontology-lookup/

QuickGO: http://www.ebi.ac.uk/ego/

OBO: http://obo.sourceforge.org

NCBO Bioportal

http://www.bioontology.org/**bioportal**.html



Mark Musen: In biology, many ontology developers are almost hobbyists

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

NCBO will offer

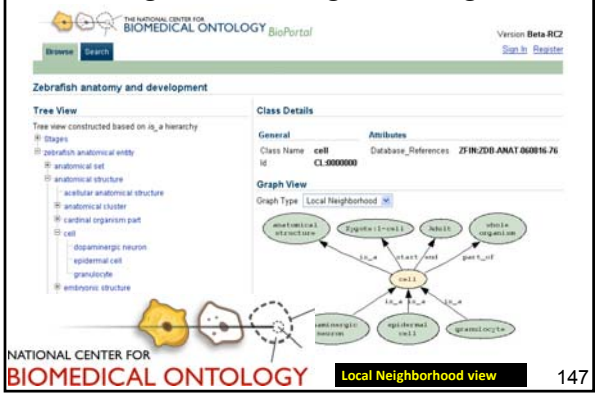
- Technology for uploading, browsing, and using biomedical ontologies
- Methods to make the online "publication" of ontologies more like that of journal articles
- Tools to enable the biomedical community to put ontologies to work on a daily basis

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

http://bioportal.bioontology.org

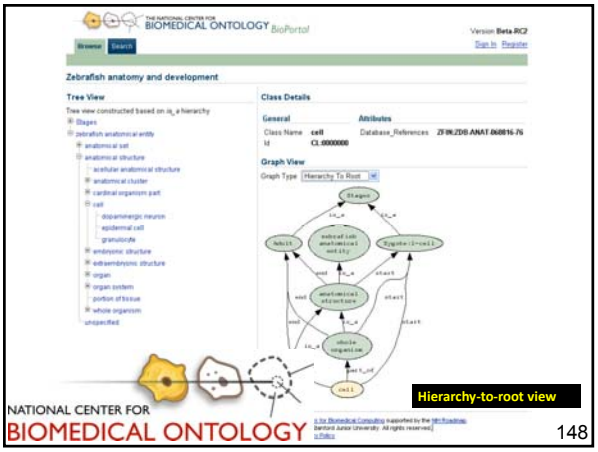
NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Browsing/Visualizing Ontologies



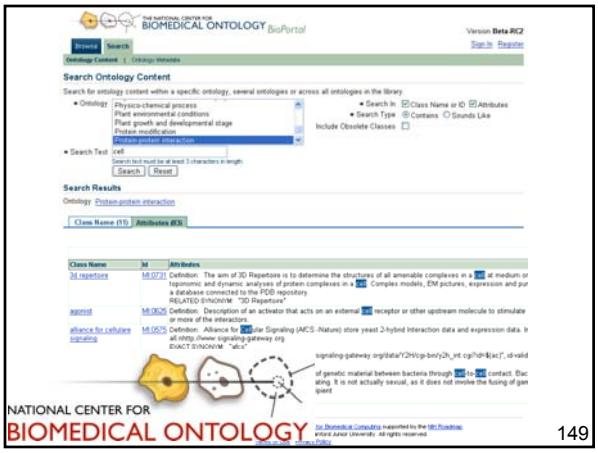
NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Local Neighborhood view



NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Hierarchy-to-root view



NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

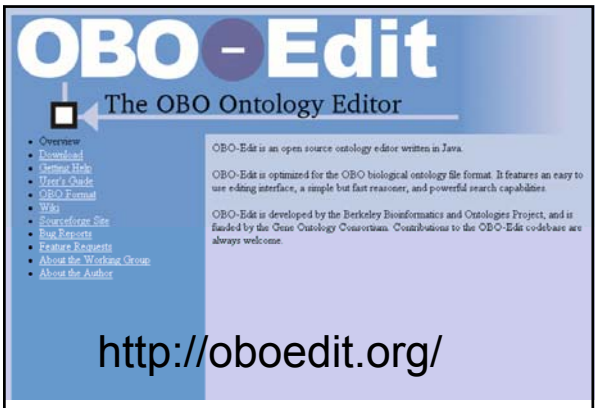
Goals for BioPortal

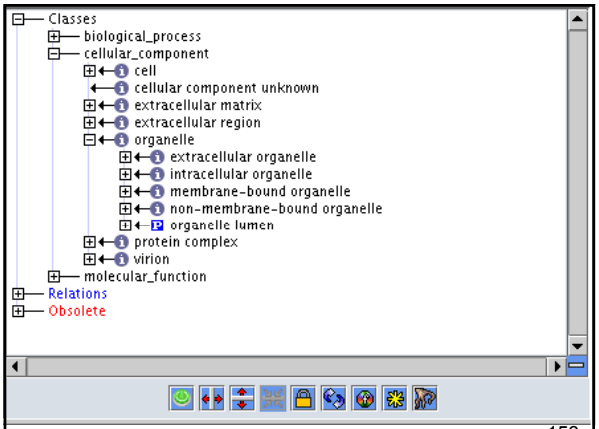
- Web accessible repository of ontologies for the biomedical community
  - Archived locally
  - Anywhere in cyberspace
- Support for ontology
  - Peer review
  - Annotation (marginalia)
  - Versioning
  - Alignment
  - Search

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Building Ontologies: The Software

OBO-Edit and Protégé-OWL

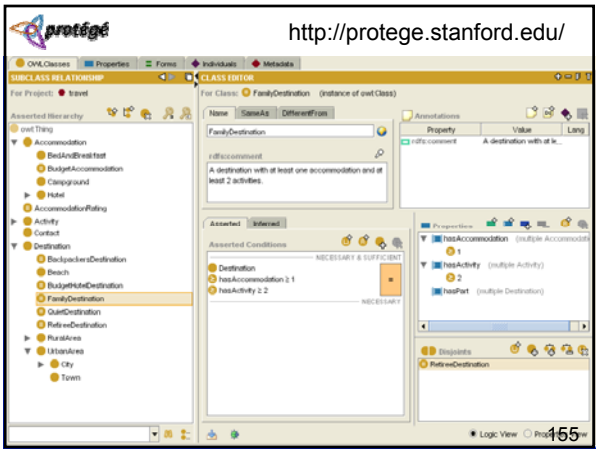


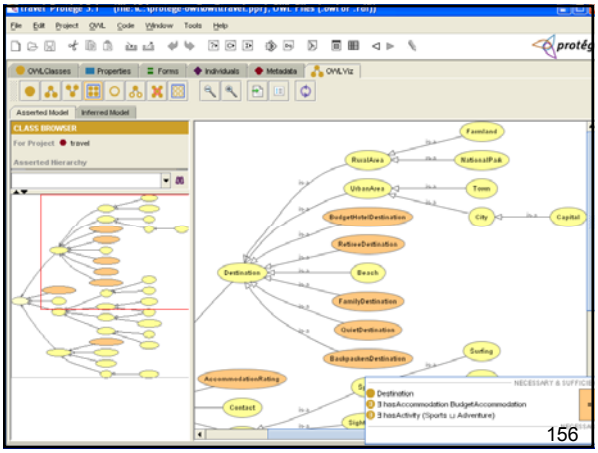



```
[Term]
id: j.0:OBI_200083
name: survival analysis
def: "__file:///data/cluster/cjm/obo/website/utils/obo-
all/obi/obi.owl#__Description1" []
is_a: j.0:OBI_200082  ! longitudinal data analysis
property_value: j.0:curation_status "graph_position_temporary"
xsd:string

[Term]
id: j.0:OBI_200084
name: time series analysis
def: "__file:///data/cluster/cjm/obo/website/utils/obo-
all/obi/obi.owl#__Description1" []
is_a: j.0:OBI_200067  ! _unclassified
property_value: j.0:curation_status "graph_position_temporary"
xsd:string

[Term]
id: j.0:OBI_200085
name: mass spectrometry analysis
def: "__file:///data/cluster/cjm/obo/website/utils/obo-
all/obi/obi.owl#__Description1" []
is_a: j.0:OBI_200067  ! _unclassified
property_value: j.0:curation_status "graph_position_temporary"
xsd:string
```



Protégé OWL tutorial

access by opening Protege OWL and looking under "Help"

Example of OWL Syntax

Class(pp:cow partial pp:vegetarian)
Class(pp:mad+cow complete
  intersectionOf(pp:cow restriction(pp:eats
    someValuesFrom(intersectionOf(pp:brain
      restriction(pp:part_of someValuesFrom
      pp:sheep))))))

(There can be no mad cows.)

OWL Manchester Syntax

OWL	DL Symbol	Manchester OWL Syntax Keyword	Example
someValuesFrom	∃	some	hasChild some Man
allValuesFrom	∀	only	hasSibling only Woman
hasValue	∋	value	hasCountryOfOrigin value England
minCardinality	≥	min	hasChild min 3
cardinality	=	exactly	hasChild exactly 3
maxCardinality	≤	max	hasChild max 3

http://www.co-ode.org/resources/reference/manchester_syntax/

The OBO-OWL controversy

2006: Creation of OBO → OWL DL converters make OBO Foundry annotated data immediately accessible to Semantic Web data integration projects

OBO now the principal entry point for creation of web-accessible biomedical data

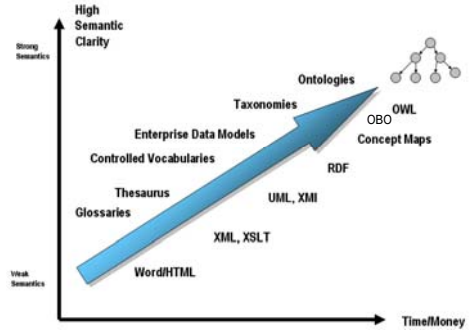
OBO / OWL pros and cons

OBO simple, more easily readable, used by many biologists, restricted expressivity

OWL is a recognized standard, more tools available, less restricted expressivity (a fragment of FOL*)

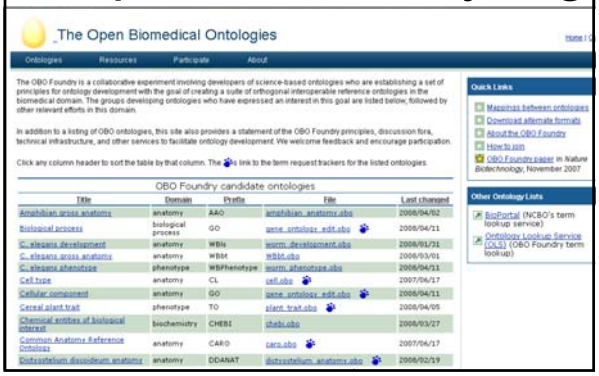
*FOL = first order logic, highly expressive, non-decidable (good framework for formulating complex relations before you start OBO or OWL coding)

Leo Obrst: The Ontology Spectrum



http://www.mkbergman.com/?m=20070516 (slightly modified)

http://obofoundry.org



How to build an ontology

work with scientists with data annotation needs to create an initial top-level classification

find ~50 most commonly used terms corresponding to universals in reality

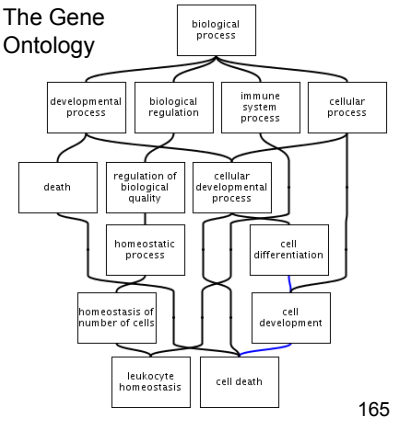
arrange these terms into an informal *is_a* hierarchy according to the universality principle

A is_a B → every instance of *A* is an instance of *B*

draw on the main BFO divisions, filling in missing terms to give a complete hierarchy

recruit domain scientists with data annotation needs to help populate the lower levels of the hierarchy and foster data integration

The Gene Ontology



Principle of Low Hanging Fruit

Include even absolutely trivial assertions (assertions you know to be universally true)

cellular development process is_a cellular process

cell death is_a death

pneumococcal bacterium is_a bacterium

Computers need to be led by the hand

Principle of Singular Nouns

Terms in ontologies represent universals

One universal per term

Avoid

death is_a biological processes

Principle

Avoid confusing between words and things

Avoid confusing between concepts in our minds and entities in reality

Recommendation: avoid the word ‘concept’ entirely

Ontologies are about universals (entities) in reality

Univocity

Terms should have the same meanings on every occasion of use.

(= They should refer to the same universals)

Basic ontological relations such as *is_a* and *part_of* should be used in the same way by all ontologies

Universality

Ontologies are made of relational assertions

They should include only those which hold universally

pneumococcal bacterium causes pneumonia

The all-some principle

Universality

Often, order will matter:

We can assert

adult transformation_of child

but not

child transforms_into adult

Universality

viral pneumonia caused by virus

but not

virus causes virus pneumonia

Universality

results analysis later_than protocol-design

but not

protocol-design earlier_than results analysis

Positivity

Complements of universals are not themselves universals.

Terms such as

non-mammal

non-membrane

other metalworker in New Zealand

do not designate universals in reality

Ontology of universals ≠ logic of terms

There are no conjunctive and disjunctive universals. Avoid terms like:

anatomic structure, system, or substance

musculoskeletal and connective tissue disorder

175

Objectivity

Which universals exist in reality is not a function of our knowledge.

Terms such as

unknown

unclassified

unlocalized

arthropathies not otherwise specified

do not designate universals in reality.

176

Keep epistemology separate from ontology

If you want to say that

We do not know where *A*'s are located

do not invent a new universal of

A's with unknown locations

(A well-constructed ontology should grow linearly; we should not need to *delete* universals or relations because of *increases* in knowledge)

177

Keep sentences separate from terms

If you want to say

I surmise that this is a case of pneumonia

do not invent a new universal of *surmised pneumonias*

Confusion of 'findings' in medical terminologies

178

Why do we need rules/standards for good ontology?

Ontologies must be intelligible both to humans (for annotation and curation) and to machines (for reasoning and error-checking): the lack of rules for classification leads to human error and blocks automatic reasoning and error-checking

Intuitive rules facilitate training of curators and annotators

Common rules allow alignment with other ontologies

179

Example: Vaccine Ontology (VO)

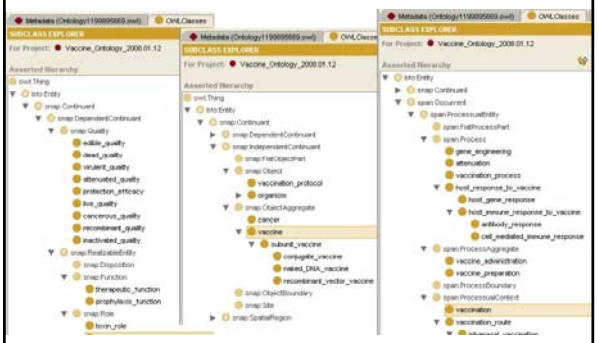
- Follow OBO Foundry principles
- Utilize the Basic Formal Ontology (BFO)
- Include two forms:
  - A disease-neutral core VO: ~60 terms in current draft
  - Disease-specific extensions: use *E. coli* and *Brucella* as models
- Collaboration: University of Michigan, University at Buffalo, and Duke University
- URL: http://www.violinet.org/wiki/index.php/Vaccine_Ontology

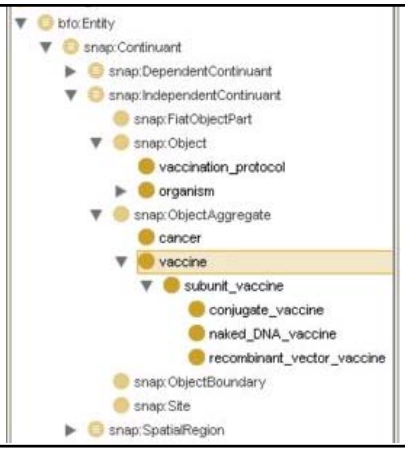
180

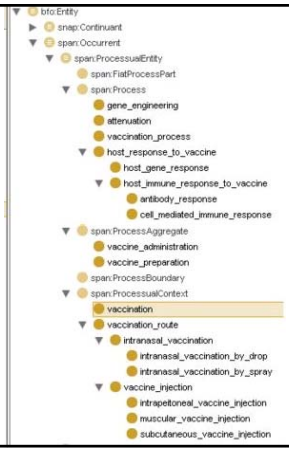
Vaccine Ontology

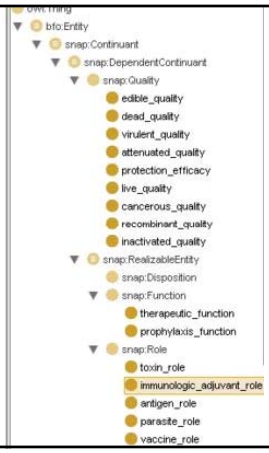
- VO will be used for natural language processing (NLP)
  - Retrieve VO-related vaccine data from literature
  - Study host-vaccine interaction pathways
  - VO scope can also be expanded by NLP and statistical analysis
- Available vaccine data in the VIOLIN vaccine database:
  http://www.violinet.org/

Vaccine Ontology Overview



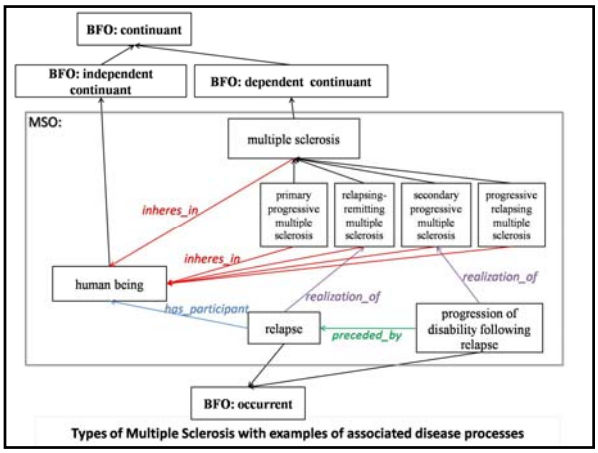






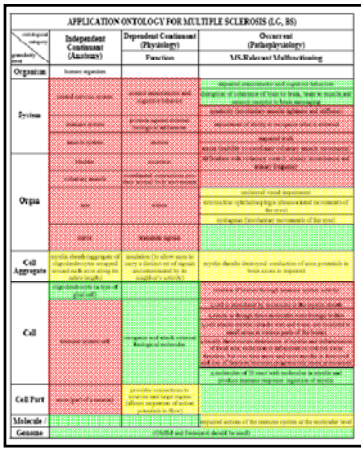
quality
function
role

Another example: ontology of multiple sclerosis



Types of Multiple Sclerosis with examples of associated disease processes

ontological category / granularity	Anatomical Structure (BFO: Independent Continuant)	Examples of Associated Functions (BFO: Dependent Continuant)	Examples of Associated Malfunctioning (BFO: Occurrent)
Organism	human organism		
Organ system	musculoskeletal system	to produce bodily motions	muscle spasm
	nervous system	to control and regulate sensorimotor activities and cognitive behavior	spasticity of skeletal muscle
			feeling prickly sensations
Organ system subdivision	immune system	to protect against external biological and non-biological agents and influences, and against cancer cells	autoimmune-induced process of inflammation
Organ	neuraxis	to provide central control of sensorimotor activities and to control cognitive behaviors	process of accumulating inflammation
	spinal nerve	to conduct electrical impulses from the spinal cord to target organs	sensation of pain in the absence of appropriate stimulus
	skeletal muscle	to contract in coordination with other skeletal muscles to produce normal body movements	muscle spasm
	extra-ocular muscle	to produce and control eye movements	eyes moving without voluntary control
			seeing double images
	optic nerve	to transmit light and visual images from the retina to the brain	axon-glial injury
Organ part	cerebellum	to coordinate movement	slowing of rapid repeating movements
	spinal cord	to convey sensory and motor signals and messages between brain, body proper, and limbs	symptoms of transverse myelitis
Tissue	neural tissue of neuraxis	to control and regulate sensorimotor activities and cognitive behavior	demyelination of affected site
Cell	T cell	to secrete lymphokines and cytokines and to assist in both antibody- and cell-mediated immune responses	T cell stimulating neutrophils and macrophages to promote inflammation in area of lesion
			T cell reacting to myelin molecules as non-self antigens



Draft MS Ontology

to apprehend what is unknown requires a *complete* demarcation of the relevant space of alternatives

Agenda • Day 1

- Introduction: What is an ontology and what is it useful for?
- Basic Formal Ontology: An upper-level ontology to support scientific research
- Open Biomedical Ontologies (OBO) and the Web Ontology Language (OWL)
- **The OBO Relation Ontology**

Realist Perspectivalism

There is a multiplicity of ontological perspectives on reality, all equally veridical i.e. transparent to reality

Two Cardinal Perspectives

1. Occurrents vs. Continuants

2. Granularity
(Micro vs. Meso vs. Macro)

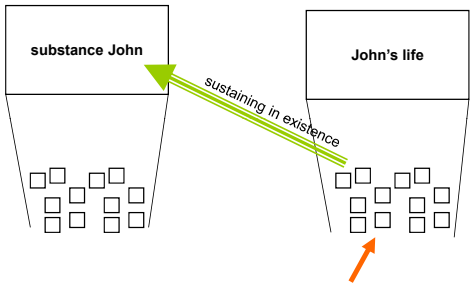
Instance-level relations

Mary's part is an instance-level part of Mary

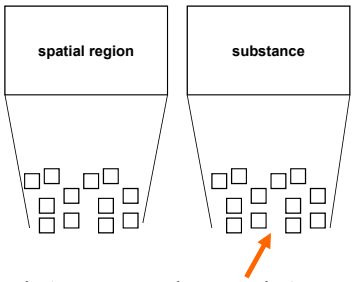
Instance-Level Part-Whole

Basic relation which holds exclusively between entities of the same top-level ontological category

Relations crossing the continuant/occurrent border are never part-relations

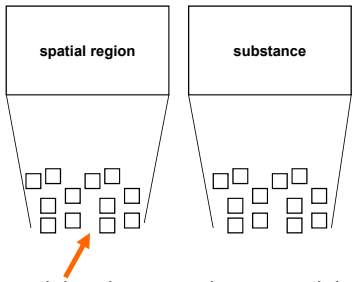


Granularity



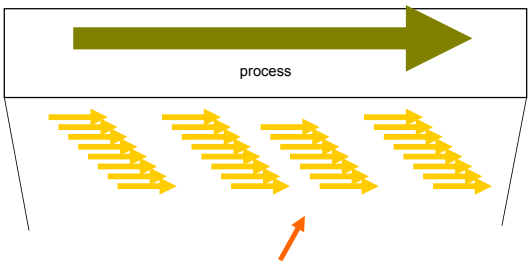
parts of substances are always substances

Granularity



parts of spatial regions are always spatial regions

Granularity



parts of processes are always processes

Intra-granular and cross-granular parthood

across continuants

Kevin's arm is part of Kevin
Kevin's molecule is part of Kevin

across occurrents

Kevin's leg-movement is part of Kevin's running
Kevin's cytometabolism is part of Kevin's running

How link continuants and occurrents together on the instance level?

via other formal relations, for example dependnce

DEPENDENCE

- one entity needs another entity to serve as its bearer
- quality depends on object
  John's suntan depends on John
- process depends on object
  John's sleeping depends on John

Objects participate in processes

PARTICIPATION

(a species of dependence)

Participation

A substance participates in a process

A runner participates in a race

A voter participates in an election

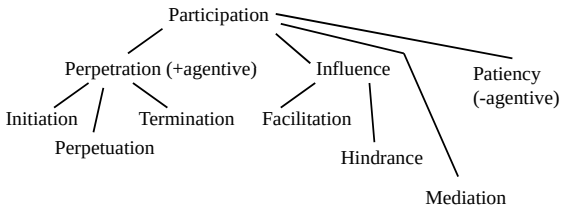
Axes of variation of participation

activity/passivity

direct/mediated

benefactor/malefactor
(conducive to existence) **[MEDICINE]**

Subtypes of participation



Initiation

A substance initiates a process:

The referee starts the race

The attorney initiates the process of appeal

Perpetuation

A substance sustains a process:

The singer sings the song

The charged filament perpetuates the emission of light

Termination

A substance terminates a process:

The operator terminates the projection of the film

The judge terminates the imprisonment of the pardoned convict

REALIZATION

A very general relation between a dependent continuant entity and a process

The *power* to legislate is realized through the passing of a law

The *role* of antibiotics in treating infections is via the killing of bacteria

Realization

the **execution** of a plan, algorithm
the **expression** of a function
the **exercise** of a role
the **realization** of a disposition

Material examples

performance of a symphony
projection of a film
expression of an emotion
utterance of a sentence
application of a therapy
course of a disease
increase of temperature

Realizable dependent entities

plan
function
role
disposition
algorithm

} **continuants**

Their realizations

execution
expression
exercise
realization
application
course

} **occurrents**

Continuant → Occurrent

Participation
Independent Continuant → Process

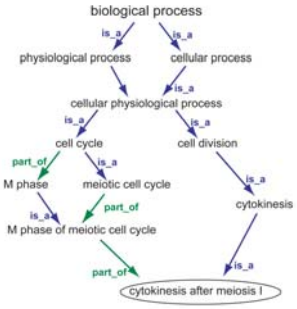
Realization
Dependent Continuant → Process

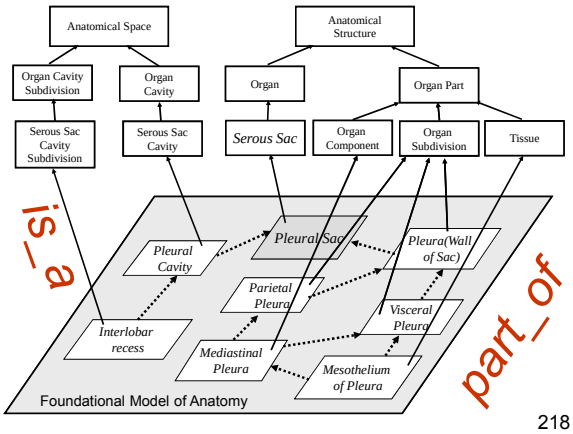
Still on the instance level

Instance-Level Relations

part_of, depends_on, realizes, ...

But ontologies are representations not of instances but of universals





OBO Relation Ontology 1.0

Foundational	***is_a*** ***part_of***
Spatial	***located_in*** ***contained_in*** ***adjacent_to***
Temporal	***transformation_of*** ***derives_from*** ***preceded_by***
Participation	***has_participant*** ***has_agent***

"Relations in Biomedical Ontologies", *Genome Biology*, April 2005

Kinds of relations

<universal, universal>: *is_a*, *part_of*, ...

<instance, universal>: this explosion
instance_of the universal *explosion*

<instance, instance>: Mary's heart
part_of Mary

Key idea

To define ontological relations like

part_of, develops_from

we need to take account not only of universals but also of their *instances* at specific *times*

(→ link to Electronic Health Record)

Key idea

To define ontological relations like

part_of, develops_from

we need to take account of both universals and their *instances* and *time*

(→ link to Electronic Health Record)

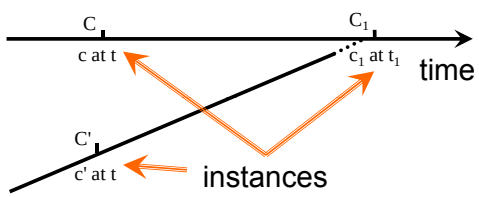
part_of
for occurrent universals is atemporal

A part_of B =def.
given any particular ***a,***
if ***a*** is an instance of *A*,
then there is some instance ***b*** of *B*
such that
a is an **instance-level part_of *b***

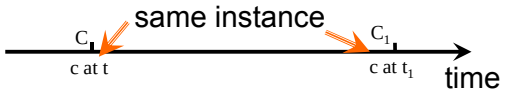
part_of
for continuant universals is time-indexed

A part_of B =def.
given any particular ***a*** and any time ***t,***
if ***a*** is an instance of *A* at ***t***,
then there is some instance ***b*** of *B*
such that
a is an **instance-level part_of *b*** at ***t***

derives_from
(ovum, sperm → zygote ...)



transformation_of

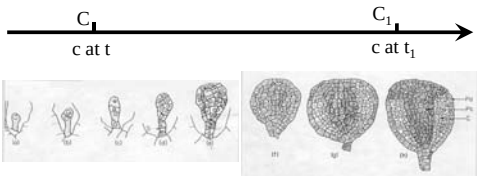


pre-RNA → mature RNA
child → adult

transformation_of

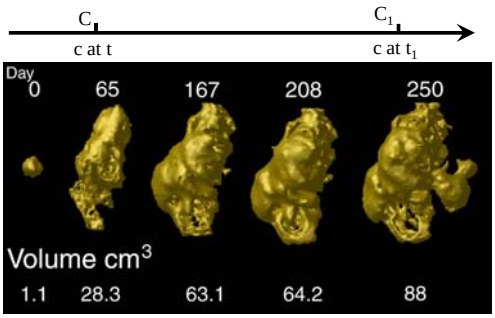
C_2 *transformation_of* C_1 =def. any instance of C_2 was at some earlier time an instance of C_1

- fetus *transformation_of* embryo
- larva *transformation_of* pupa
- adult *transformation_of* child



embryological development

tumor development



229

is_a (for occurrents)

A ***is_a*** *B* =def

For all *x*, if *x* **instance_of** *A* then *x* **instance_of** *B*

cell division ***is_a*** *biological process*

230

is_a (for continuants)

A ***is_a*** *B* =def

For all *x, t* if *x* **instance_of** *A* **at** *t* then *x* **instance_of** *B* **at** *t*

abnormal cell is_a cell
adult human is_a human
but not: *adult is_a child*

231

These definitions should support cross-ontology reasoning

Whichever *A* you choose, the instance of *B* of which it is a part will be included in some *C*, which will include as part also the *A* with which you began

The same principle applies to the other relations in the OBO-RO:

located_at, *transformation_of*, *derived_from, adjacent_to*, etc.

232

A part_of B, *B part_of C* ...

The **all-some** structure of the definitions in the OBO-RO allows

cascading of inferences

(i) within ontologies

(ii) between ontologies

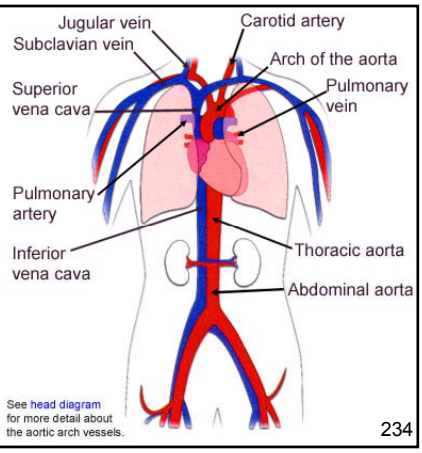
(iii) between ontologies and EHR repositories of instance-data

233

Continuity

Attachment

Adjacency



234

Modes of Connection

Modes of connection:

- *attached_to* (muscle to bone)
- *synapsed_with* (nerve to nerve, nerve to muscle)
- *continuous_with* (= share a fiat boundary)

a continuous_with *b* = *a* and *b* are continuant instances which share a fiat boundary

This relation is always symmetric at the instance level:

if *x* **continuous_with** *y* , then *y* **continuous_with** *x*

continuous_with (relation between universals)

A continuous_with B =Def.

for every instance *x* of *A* at *t*
there is some instance *y* of *B at t* such that
x **continuous_with** *y at t*

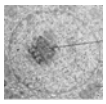
continuous_with as a relation between universals is not always symmetric

Consider *lymph node* and *lymphatic vessel:*

Each lymph node is continuous with some lymphatic vessel, but there are lymphatic vessels (e.g. lymphs and lymphatic trunks) which are not continuous with any lymph nodes

instance level

this nucleus is adjacent to this cytoplasm
implies:
this cytoplasm is adjacent to this nucleus



universal level

nucleus adjacent_to cytoplasm
Not: *cytoplasm adjacent_to nucleus*

Adjacent_to as a relation between universals is not always symmetric

Consider

seminal vesicle adjacent_to urinary bladder

Not: *urinary bladder adjacent_to seminal vesicle*

Applications

Expectations of symmetry e.g. for interactions may hold only at the instance level

if *A* interacts with *B*, it does not follow that *B* interacts with *A*

if *A* is expressed simultaneously with *B*, it does not follow that *B* is expressed simultaneously with *A*

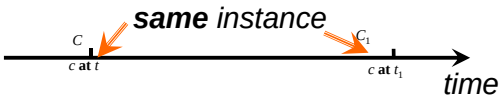
transformation_of

A transformation_of B =Def.

Every instance of *A* was at some earlier time an instance of *B*

– adult transformation_of child

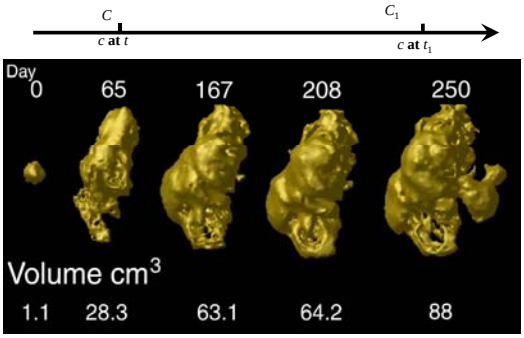
transformation_of



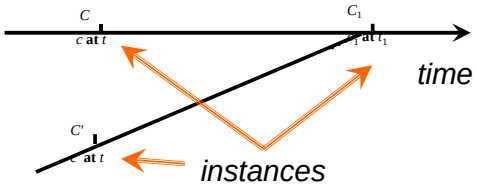
pre-RNA → mature RNA

child → adult

tumor development

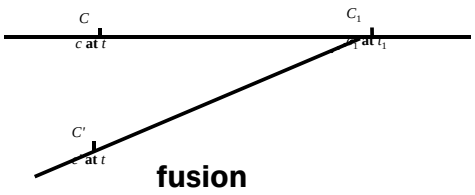


derives_from

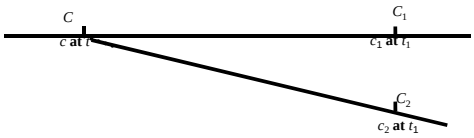


zygote derives_from *ovum*
sperm

two continuants fuse to form a new continuant

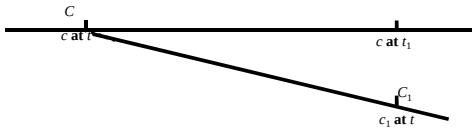


one initial continuant is replaced by two successor continuants



fission

one continuant detaches itself from an initial continuant, which itself continues to exist



budding

Primitive Relations

- c **instance_of** C **at** t - a primitive relation between a continuant instance and a universal which it instantiates at a specific time
- p **instance_of** P - a primitive relation between a process instance and a universal which it instantiates holding independently of time
- c **part_of** c1 **at** t - a primitive relation between two continuant instances and a time at which the one is part of the other
- p **part_of** p1, r **part_of** r1 - a primitive relation of parthood, holding independently of time, either between process instances (one a subprocess of the other), or between spatial regions (one a subregion of the other)

Primitive Relations

- c **located_in** r **at** t - a primitive relation between a continuant instance, a spatial region which it occupies, and a time
- r **adjacent_to** r1 - a primitive relation of proximity between two continuants
- t **earlier** t1 - a primitive relation between two times
- c **derives_from** c1 - a primitive relation involving two distinct material continuants c and c1
- p **has_participant** c **at** t - a primitive relation between a process, a continuant, and a time

Defined Instance-Level Relations

p **occurring_at** t =def. for some c, p **has_participant** c **at** t.

p **preceded_by** p1 =def. for all t, t1, if p **occurring_at** t and p1 **occurring_at** t1, then t1 **earlier** t

Defined Instance-Level Relations

t **first_instant** p =def. p **occurring_at** t, and for all t1, if t1 **earlier** t, then not p **occurring_at** t1

t **last_instant** p =def. p **occurring_at** t and for all t1, if t **earlier** t1, then not p **occurring_at** t1 p

Overlaps on the level of instances

x **overlaps** y **at t** =def. there is some z such that z is **part_of** x **at t** and z is **part_of** y **at t**

Overlaps on the level of universals

X *overlaps* Y =def. for every t and every x, if x **instance_of** X at t, then there is some instance y of Y at t such that (x **overlaps** y at t)

Note that it can be the case that X *overlaps* Y as thus defined, even though Y does not *overlap* X.

Thus uterine tracts *overlaps* urinogenital sysem

but not uriongenital system *overlaps* uterine tract (because of male urinogenital systems)

Proposed new relations on the level of universals

about – between an information object and an object to which it refers

inheres_in

depends_on

output_of

has_input

has_function

has_quality

realization_of

from http://www.bioontology.org/wiki/index.php/RO:Main_Page

New relations

A *depends_on* B =def. every instance of A is such that it cannot exist unless some instance of B exists

apoptosis depends_on cell

death depends_on organism

…

New Gene Ontology 'Regulates' Relations

def: "A relation between a process and a process. A regulates B if the unfolding of A affects the frequency, rate or extent of B. A is called the regulating process, B the regulated process“

A regulates B =def. A is a process type and B is a process type and every instance of A is such that its unfolding affects the frequency, rate or extent of some instance of B.

Positive and Negative Regulation

positively_regulates def: "A regulation relation in which the unfolding of the regulating process *increases* the frequency, rate or extent of the regulated process"

negatively_regulates def: "A regulation relation in which the unfolding of the regulating process *decreases* the frequency, rate or extent of the regulated process"

The Granularity Gulf

most existing data-sources are of fixed, single granularity

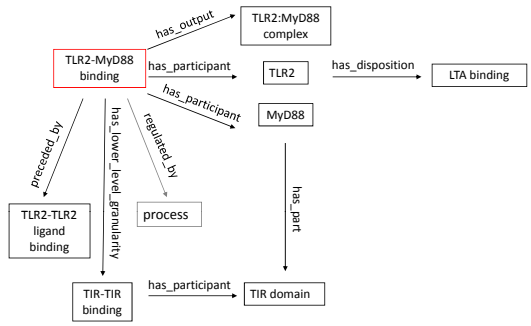
many (all?) clinical phenomena *cross* granularities

Advantages of the methodology of enforcing commonly accepted coherent definitions

promote quality assurance (better coding)

guarantee automatic reasoning across ontologies and across data at different granularities

yields connection to times and instances in EHR



TLR-2 signalling pathway

with thanks to Lindsay Cowell

AN INTRODUCTION TO BIOMEDICAL ONTOLOGY 2
Sunday April 13, 2008

Barry Smith

University at Buffalo
http://ontology.buffalo.edu/smith

Agenda · Day 2

An ontological introduction to biomedicine: Defining organism, function and disease

The Gene Ontology (GO), the Foundational Model of Anatomy (FMA) and the Infectious Disease Ontology (IDO)

The OBO Foundry: A suite of biomedical ontologies to support reasoning and data integration

Applications of ontology outside biomedicine

Agenda

An ontological introduction to biomedicine: Defining organism, function and disease

The Gene Ontology (GO), the Foundational Model of Anatomy (FMA) and the Infectious Disease Ontology (IDO)

The OBO Foundry: A suite of biomedical ontologies to support reasoning and data integration

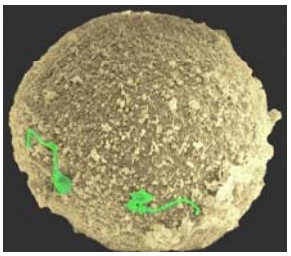
Applications of ontology outside biomedicine

Defining ‘organism’

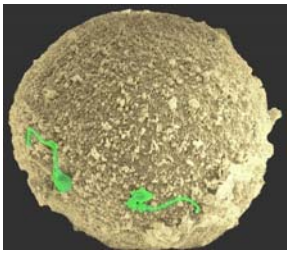
Organism =def. an independent continuant, made of matter, which …

To fill in the gap, we consider the question: When does an organism begin to exist?

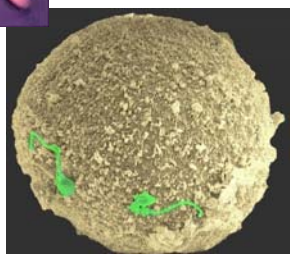
First there are two:

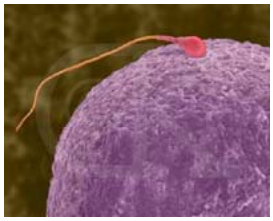


first there are two:



first there are two:

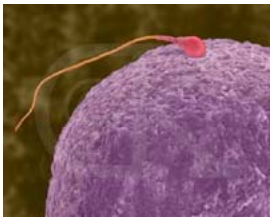




... and then there is one

This is an organism

This is not (yet) an organism



So where is the threshold?

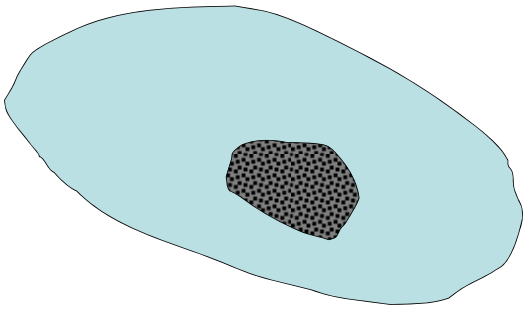
a. zygote (single cell) (day 0)
b: multi-cell (days 0-3)
c. morula (day 3)
d. early blastocyst (day 4)
e. implantation (days 6-13)
f. gastrulation (days 14-16)
g. neurulation (from day 16)
h. formation of the brain stem (days 40-43)
i. end of first trimester (day 98)
j. viability (around day 130)
k. sentience (around day 140)
l. quickening (around day 150)
m. birth (day 266)
n. the development of self-consciousness

Methodology for answering this question

Set forth criteria which an entity must satisfy to be an organism

And establish at which point in human development these criteria are first satisfied by an entity which can be transtemporally identical with the adult human being

Is the zygote already an organism?

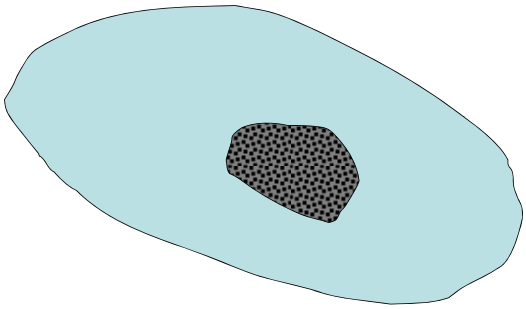


283

and is it the *same* organism as this?

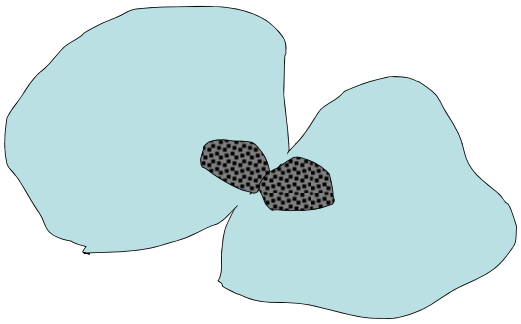
284

the problem is that this, almost immediately,



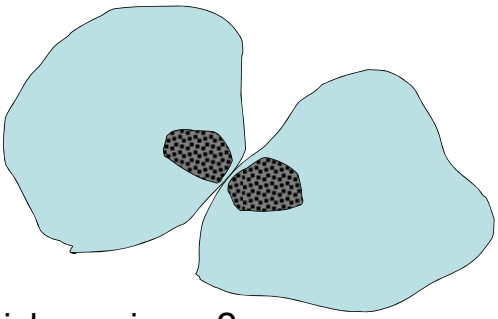
285

becomes this…



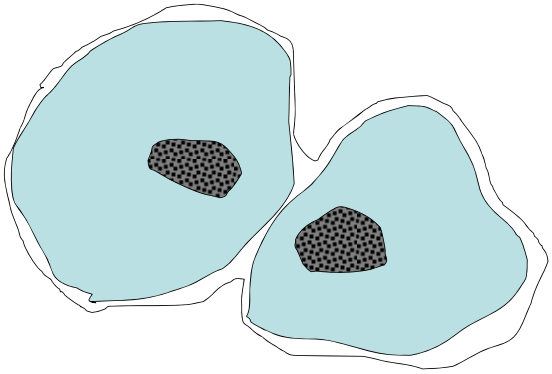
286

…and then cleavage



which one is me?

287



2 cells plus *zona pellucida*

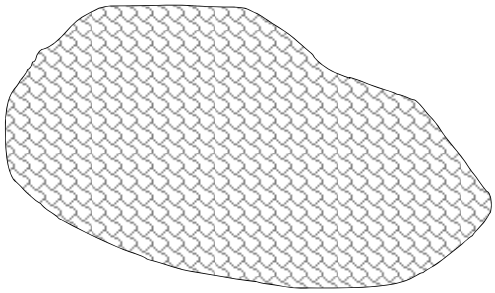
288

is 1 of the cells at the 2-cell stage *me?*

these two cells of this new organism are cytoplasmically differentiated

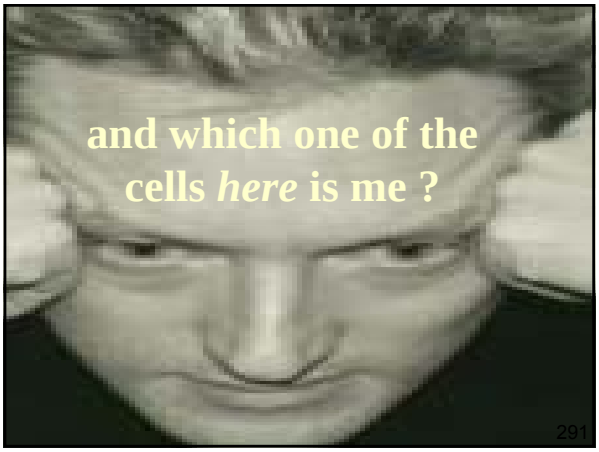
289

… but now, more cleavages, create a cell mass



which one of ***these*** cells is me?

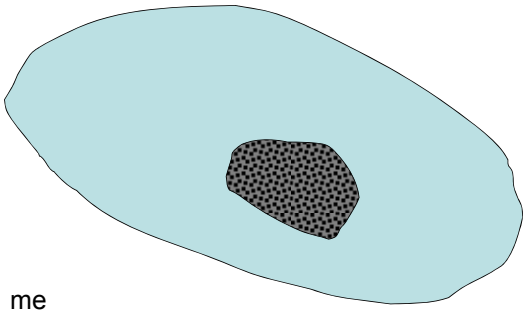
290



291

292

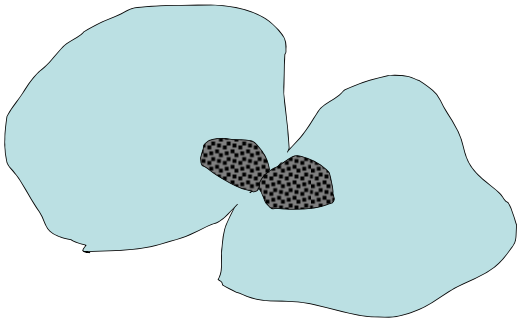
An alternative story



me

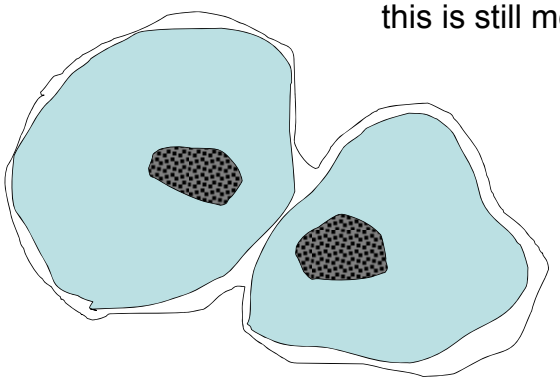
293

still me (all of it)

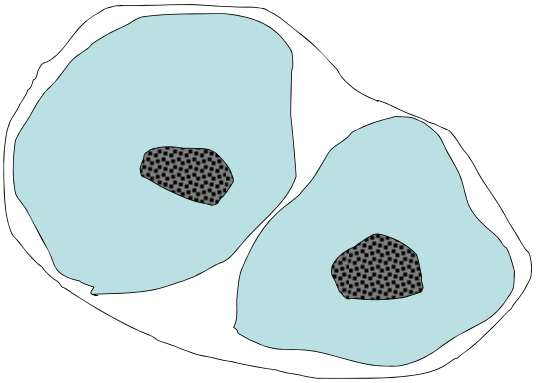


294

this is still me



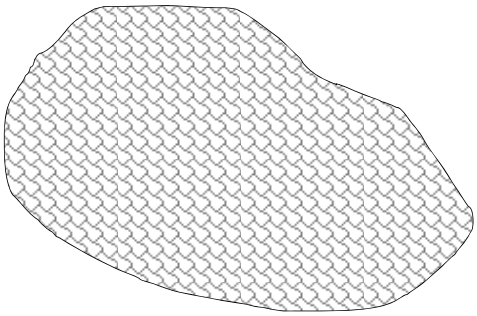
295



2 cells plus *zona pellucida*

296

This is still me:
I was once a whole blastula (60 cells)



297

Methodology for determining which if these two accounts of organism formation is correct

What are the criteria which an entity must satisfy to be an organism?

298

First criterion

An organism must be an independent continuant.

More specifically it must be what Aristotle referred to under the term ‘substance’

(= a maximally self-connected independent continuant)

299

Conditions on Substance

1. Each substance is an entity which persists through time and remains numerically one and the same
2. Each substance is a bearer of change. (John is now warm, now cold)
3. Each substance is extended in space (The spatial parts of John are, for example, his arms and legs, his cells and molecules.)
4. Each substance possesses its own complete, connected external boundary
5. Each substance is connected in the sense that its parts are not separated from each other by spatial gaps. (Substances are thereby distinguished from heaps or aggregates of substances) (Exceptions: blood cells, immune system parts)
6. Each substance is an independent entity (Contrast: smiles, blushes)

300

Second criterion

An organism must be a relatively isolated causal system

301

Conditions on Relatively Isolated Systems

7. The external boundary of the entity is established via a physical covering (for example a membrane)

8. The events transpiring inside this covering divide between those with characteristic magnitudes (of temperature, etc.) inside a spectrum of allowed values and those outside

9. The covering serves as shield to protect the entity from damaging causal influences

10. The entity contains its own mechanisms for maintaining sequences of events falling within the spectrum of allowed values (mechanisms of self-repair)

302

These two criteria are to a degree independent

A block of ice is a substance, but it is not a relatively isolated causal system.

An orbiting space-ship, with its sophisticated mechanisms for self-repair, is both a substance and a causally isolated system.

Siamese twins may be one substance, but two causally isolated systems.

An amoeba is both a substance and a causally isolated system yet still *divisible*

303

Being a relatively isolated causal system is realized to different degrees by different entities.

Being a substance is realized always to the same degree: either wholly or not at all.

All substantial change is (practically) instantaneous.

304

Substantial change

two drops of water flow together and become one

an ameoba splits and becomes two

305

'Substance' has to do with existence and structure. 'Causal system' has to do with function and functioning.

Being a relatively isolated causal system is often realized through modules organized hierarchically (nesting).

Thus functions, too, are often organized modularly.

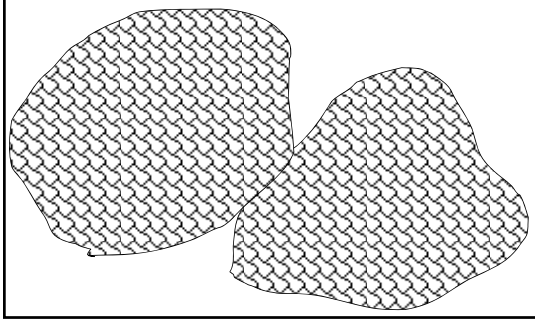
306

Was I ever a blastula? (a whole blastula?)

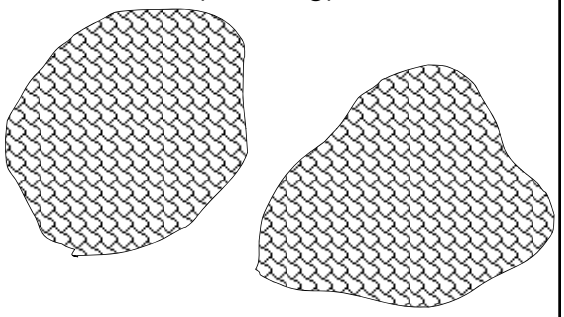
The blastula is **a single substance:** its cells together form a connected whole with a common physical boundary

But it lacks its own internal mechanisms in virtue of which its several parts would in case of disturbance work together as a whole to restore stability

If I was ever a blastula then I am such that it was once possible that *this* happened to me



blastulae are subject to division (twinning)



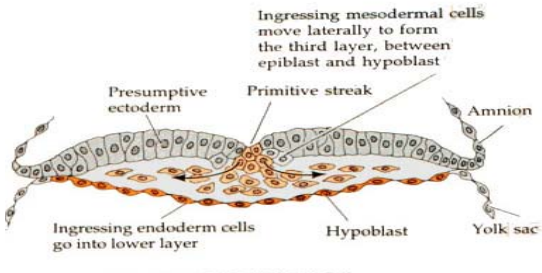
Gastrulation (Day 16)

Hypothesis: Gastrulation transforms the blastula from a putative cluster of cells into a single heterogeneous entity—a whole multicellular individual living being which has a body axis and bilateral symmetry and its own mechanisms to protect itself and to restore stability in face of disturbance.

Lewis Wolpert

"It is not birth, marriage or death, but gastrulation, which is truly the most important event in your life."

Gastrulation



Gastrulation is analogous to the transformation of a mass of copper threads into a single integrated circuit

Neurulation (begins day 16)

transforms the gastrula by establishing the beginning of the central nervous system.

= a second nd massive migration of cells and topological folding and connecting and subsequent cell specialization yielding the neural tube

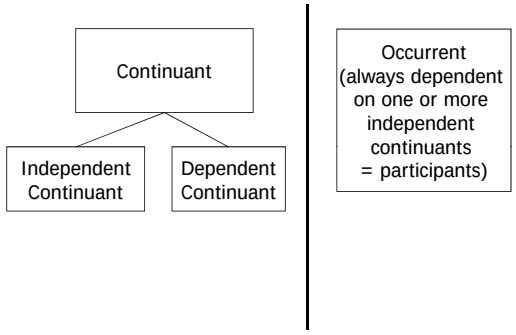
a. zygote (single cell) (day 0)
b: multi-cell (days 0-3)
c. morula (day 3)
d. early blastocyst (day 4)
e. implantation (days 6-13)
f. gastrulation (days 14-16)
g. neurulation (from day 16)
h. formation of the brain stem (days 40-43)
i. end of first trimester (day 98)
j. viability (around day 130)
k. sentience (around day 140)
l. quickening (around day 150)
m. birth (day 266)
n. the development of self-consciousness (some time after birth)

Further reading

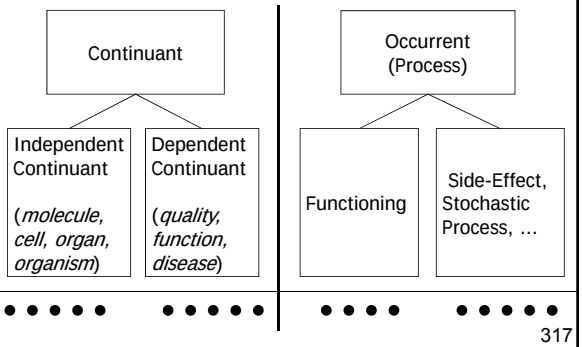
Barry Smith and Berit Brogaard, "Sixteen Days", *The Journal of Medicine and Philosophy*, 28 (2003), 45–78.

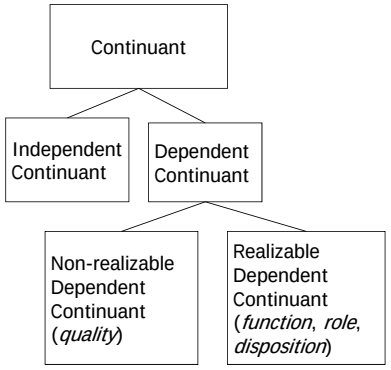
http://ontology.buffalo.edu/smith/articles/embryontology.htm

What is a function?



BFO





the function of a screwdriver
the function of a heart

roughly: functions are beneficial dispositions hard-wired into an entity

(a) by its maker

(b) by evolution

What is a disposition?

An object has a disposition to *M* when *C* =def. it is physically structured in such a way that it *M*s when *C*.

e.g. An object has a disposition to shatter when dropped

A disposition is a realizable dependent continuant

The process of shattering is the *realization* of the disposition we call 'fragility'

The parts of the organism have functions

They are designed to ensure that the events transpiring inside the organism remain within the spectrum of allowed values and to respond when they move outside this spectrum of allowed values

What is a biological function?

First proposal: an entity x has a biological function if and only if x is part of an organism and has a disposition to act reliably in such a way as to contribute to the organism's survival

the function *is* this disposition

e.g. your heart is disposed to pump blood

Problem of aging and death

are there parts of the organism involved in bringing about or responding gracefully to aging processes?

is *this* their function?

Problem of reproductive organs

some organisms are such that the exercise of their reproductive organs brings death

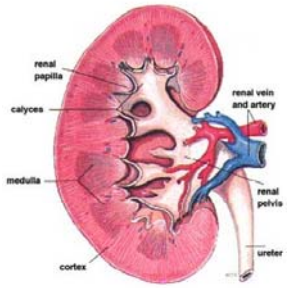
Perhaps: an entity has a biological function if and only if it is part of an organism and has a disposition to act reliably in such a way as to contribute to the **group's** survival?

seems too remote – think of my left upper molar

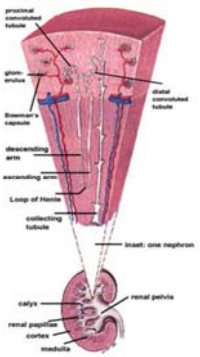
Functions are organized in modular hierarchies

The function of each functional part is: *to contribute to the functioning of the next larger whole*

We need to understand 'function' in relation to the immediate environing whole of the part in question. From this perspective the group seems structurally too far away



The function of the kidney is to purify blood

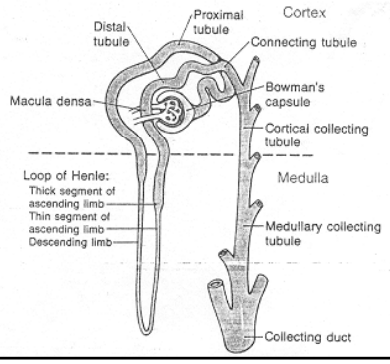


The *nephron* is the cardinal functional unit of the kidney

Functions

- to regulate the concentration of water and soluble substances like sodium salts in the blood
- to eliminate wastes from the body
- to regulate blood volume and pressure
- to control levels of electrolytes and metabolites
- to regulate blood pH

functional segments within the nephron
15 different cell types



… an entity has a biological function if and only if it is part of an organism and has a disposition to act *reliably* in such a way as to …

Function is what gives rise to *normal* activity

Normality ≠ statistical normality

That sperm exercise their function (to penetrate an ovum) is *rare*

That human adults have 32 teeth is *rare*

Functions and Malfunctionings

This is a screwdriver
This is a good screwdriver
This is a broken screwdriver

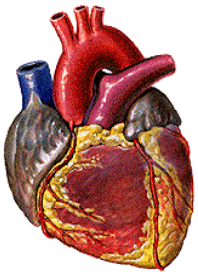
This is a heart
This is a healthy heart
This is an unhealthy heart

Functions are associated with certain characteristic *process shapes*

Screwdriver: rotates and simultaneously moves forward simultaneously transferring torque from hand and arm to screw

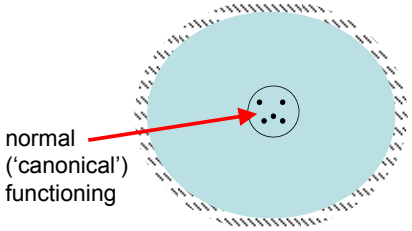
Heart: performs a contracting movement inwards and an expanding movement outwards

Functions and Prototypes

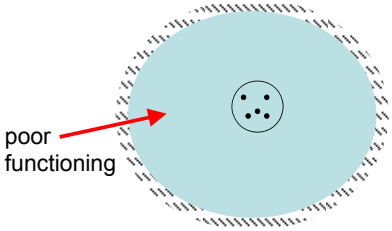


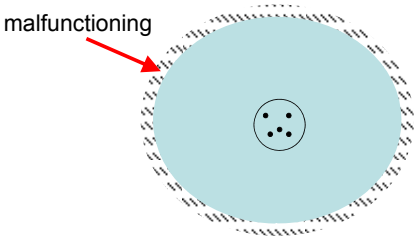
In its functioning, a heart creates a four-dimensional process shape.
Good hearts create other process shapes than sick hearts do.

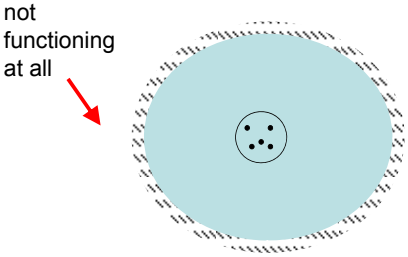
Prototypes



Map of process shapes







Not functioning at all

leads to **death,** *modulo*
internal factors:
- plasticity
- redundancy (2 kidneys)
- criticality of the system involved

external factors:
- prosthesis (dialysis machines, oxygen tent)
- special environments
- assistance from other organisms

What is health?

Boorse: the state of an organism is theoretically healthy, i.e., free from disease, in so far as its mode of functioning conforms to the natural design of that kind of organism

What clinical medicine is for

to eliminate malfunctioning by fixing broken body parts

(or to prevent the appearance of malfunctioning by intervening, e.g. at the molecular level, before the breaks develop)

What, then, is function?

The Gene Ontology

represents only what is normal in the realm of (molecular) functioning

= what pertains to normal ('wild type') organisms (in all species)

The Gene Ontology is a canonical ontology

The GO is a canonical representation

"The Gene Ontology is a computational representation of the ways in which gene products normally function in the biological realm"

Nucl. Acids Res. 2006: 34.

The Foundational Model of Anatomy
a representation of canonical anatomy

a representation of universals, and relations between universals, deduced from the qualitative observations of the *normal* human body, the structure generated by the coordinated expression of the organism's own structural genes

Model organisms

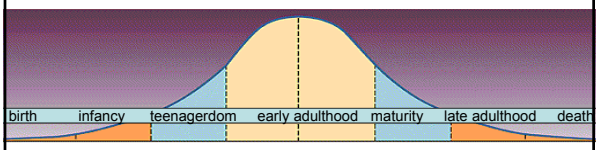
you can buy a mouse with the prototypical mouse *Bauplan* according to a precise genetical specification

A solution to the problem of defining function

For each type of organism there is not only a canonical *Bauplan*, but also a canonical life plan (canonical life *Gestalt*)

= the physiological counterpart of canonical anatomy

the canonical human life (plan)



For all animals the canonical life plan includes:

- canonical embryological development
- canonical growth
- canonical reproduction
- canonical aging
- canonical death

For humans

first, mewling and puking
then creeping like snail unwillingly to school
then sighing like furnace with woeful ballad made to his mistress' eyebrow
then a soldier full of strange oaths
then justice in fair round belly
then the lean and slipper'd pantaloon
then second childishness and mere oblivion, sans teeth, sans eyes, sans taste, sans everything.

As You Like It, II.vii.139-166

Family	Work	Money
Adoption	Employment	Bankruptcy
Aging	Injury	Budgeting
Birth	Job Seeking	Charitable Contributions
Child care	Re-employment	College
Death	Small Business	Credit
Disability	Self-employment	Disasters
Divorce	Telecommuting	Home Improvement
Domestic Violence	Unemployment	Home Purchase
Driving	Volunteering	Home Selling
Elder Care	Workplace Violence	Insurance
Empty Nesting		Investing
Health		IRS Audit
Illness		Lawsuits
Kids		Mortgage
Marriage		Property
Parenting		Renting
Retirement		Saving
Schooling		Taxes
Teenagers		Trusts
Travelling		Wills

FirstGov Life Events Taxonomy

What does every human canonical life involve?

9 months of development

...

cycles of waking, sleeping; eating and not eating; drinking and not drinking

...

death

Iberall and McCulloch 20 action modes:

Action Modes	% of time
Sleeps	30
Eats	5
Drinks	1
Voids	1
Sexes	3
Works	25
Rests (no motor activity, indifferent internal sensory flux)	3
Talks	5
Attends (indifferent motor activity, involved sensory activity)	4
Motor practices (runs, walks, plays, etc.)	4
Angers	1
Escapes (negligible motor and sensory input)	1
"Anxioius-es"	2
"Euphorics"	2
Laughs	1
Aggresses	1
Fears, fights, flights	1
Interpersonally attends (body, verbal or sensory contact)	8
Envies	1
Greeds	1
Total:	100% +/- 20% of time involvement

Water balance (from hour to hour)

Water balance (in the long run)

What does "function" mean?

Initial version:

an entity has a biological function if and only if it is part of an organism and has a disposition to act reliably in such a way as to contribute to the organism's survival

Improved version

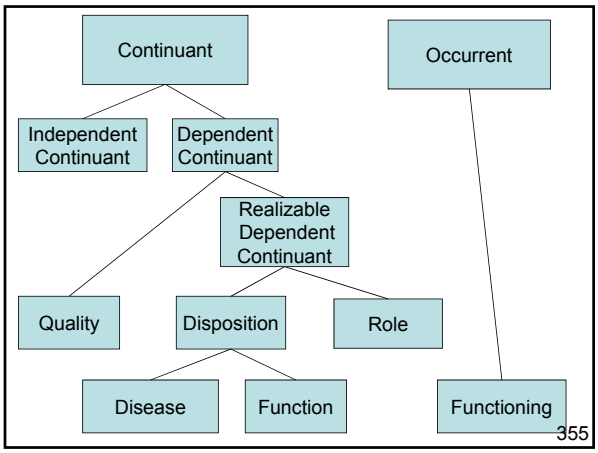
an entity has a biological function if and only if it is part of an organism and has

a disposition to act reliably in such a way as to contribute to the organism's realization of the canonical life plan for an organism of that type

What is disease?

functions are, roughly, good dispositions relevant to the realization of the canonical life plan for an organism of the relevant type

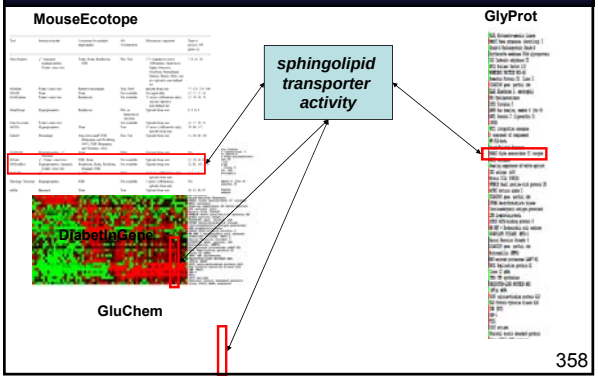
diseases are (even more roughly) counterpart bad dispositions



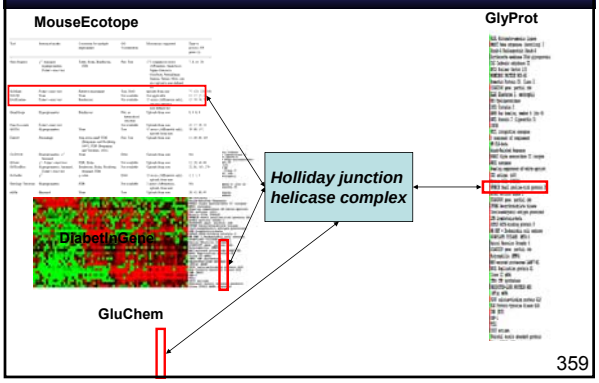
Agenda · Day 2

- An ontological introduction to biomedicine: Defining organism, function and disease
- **The Gene Ontology (GO), the Foundational Model of Anatomy (FMA) and the Infectious Disease Ontology (IDO)**
- The OBO Foundry: A suite of biomedical ontologies to support reasoning and data integration
- Applications of ontology outside biomedicine

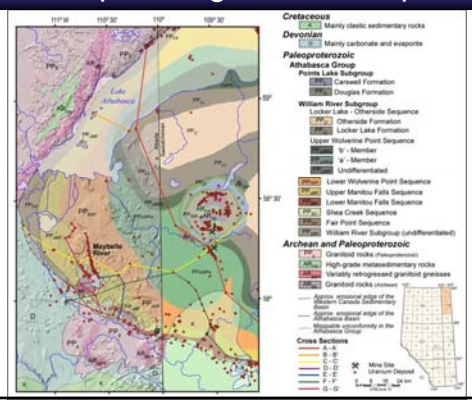
The Idea of Common Controlled Vocabularies



ontologies are legends for data



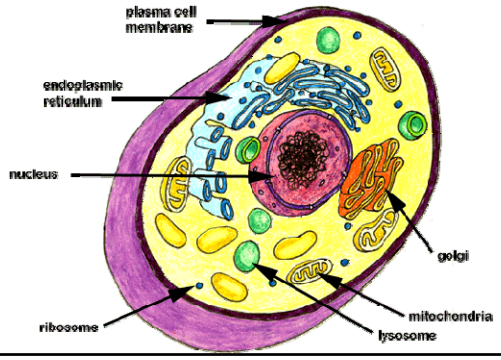
compare: legends for maps



common legends allow (cross-border) integration

361

compare: legends for diagrams



362

legends

help human beings use and understand complex representations of reality

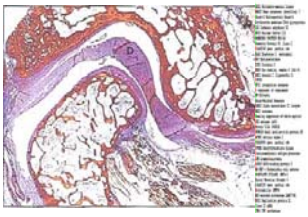
help human beings create useful complex representations of reality

help computers process complex representations of reality

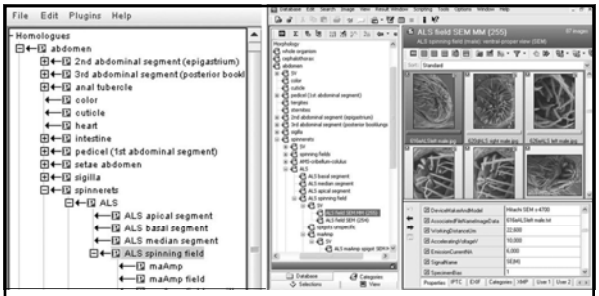
help glue data together

363

Annotations using common ontologies can yield integration of image data



364



Ramirez et al.
Linking of Digital Images to Phylogenetic Data Matrices Using a Morphological Ontology
Syst. Biol. 56(2):283–294, 2007

365

The Gene Ontology

a *structured representation of* attributes of gene products, which can be used by researchers in many different disciplines who are focused on one and the same biological reality

366

The GO works

by providing a common set of terms for describing different types of data

- across species (human, mouse, yeast, ...)
- across granularities (molecule, cell, organ, organism, population)
- across technologies (Microarray, CT, MRI, ..

and so provide for enhanced access to and reasoning with data

The methodology of annotations

Model organism databases employ scientific curators who use the experimental observations reported in the biomedical literature to associate GO terms with entries in gene product and other molecular biology databases

Example of use of the GO

A study of 11 breast and 11 colorectal cancers found 13,023 genes

The GO tells you what is standard functioning for these genes

By tracking deviations from this standard, in part through use of GO, 189 genes were identified as being mutated at significant frequencies and thus as providing targets for diagnostic and therapeutic intervention.

Sjöblöm T, *et al. Science*. 2006 ;314:268-74.

Uses of GO to throw light on

genes involved in occupational bronchitis in humans (PMID 17459161)

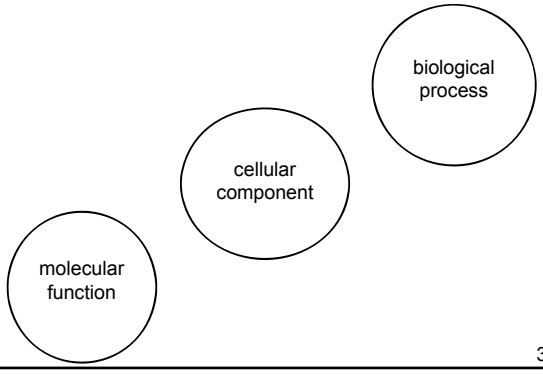
immune system involvement in abdominal aortic aneurisms in humans (PMID 17634102)

prevention of ischemic damage to the retina in rats (PMID 17653046)

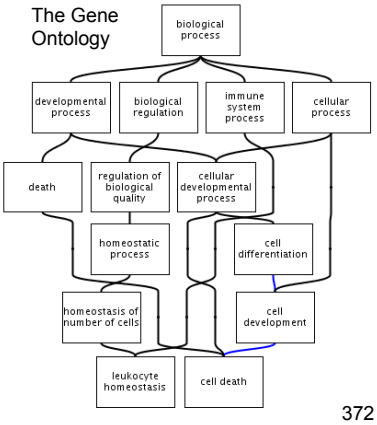
how the white spot syndrome virus affects cell function in shrimp (PMID 17506900)

...

GO's three ontologies

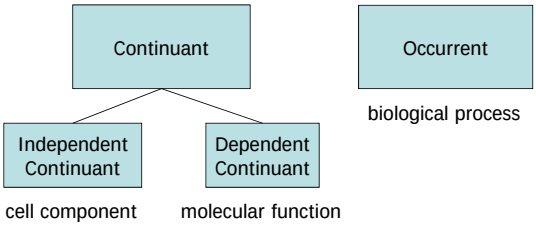


The Gene Ontology



no connections between the three separate ontologies

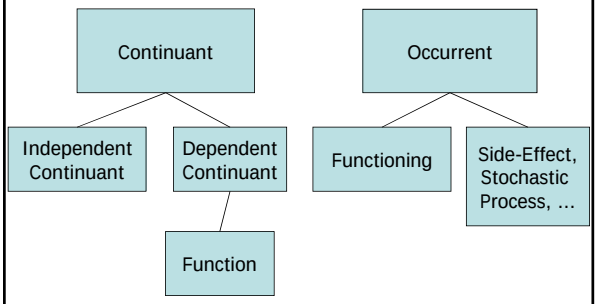
research on dependence relations



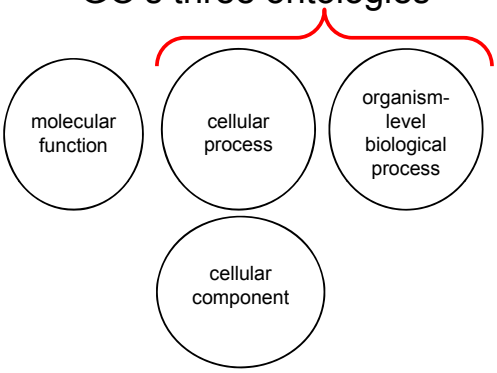
Kumar A., Smith B, Borgelt C. Dependence relationships between Gene Ontology terms based on TIGR gene product annotations. *CompuTerm* 2004, 31-38.

Bada M, Hunter L. Enrichment of OBO Ontologies. *J Biomed Inform*. 2006 Jul 26

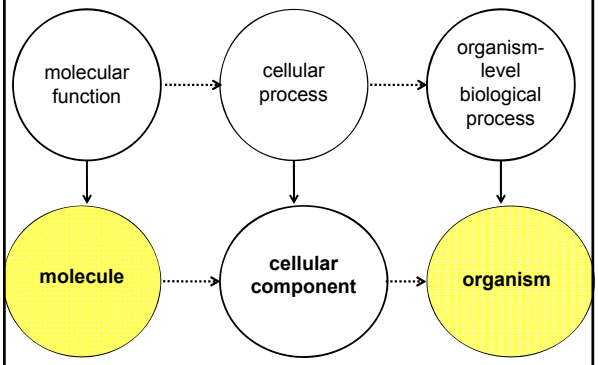
Top-Level Ontology



GO's three ontologies

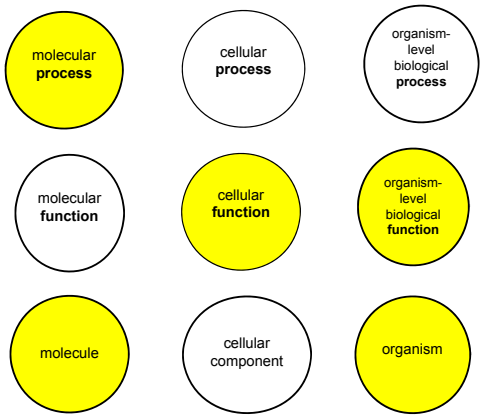


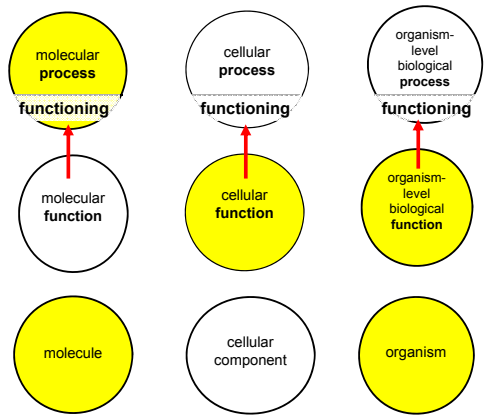
Normalization of Granular Levels



need to separate function from process

not all processes are realizations of functions





Glossary

Instance: A particular entity in spatio-temporal reality.

Type: A general kind instantiated by an open-ended totality of instances which share certain qualities and propensities in common of the sort that can be documented in scientific literature

Glossary

Gene product instance: A molecule that is generated by the expression of a DNA sequence and which plays some significant role in the biology of the organism.

Gene product type: A type of gene product instance.

Glossary

Biological process instance (aka "occurrence"): A change or complex of changes on the level of granularity of the cell or organism, mediated by one or more gene products.

Biological process type: A type of biological process instance.

Glossary

Cellular component instance: A part of a cell, including cellular structures, macromolecular complexes and spatial locations identified in relation to the cell

Cellular component type: A type of cellular component.

Glossary

Molecular function instance: The propensity of a gene product instance to perform actions, such as catalysis or binding, on the molecular level of granularity.

Molecular function type: A type of molecular function instance.

Glossary

Molecular function execution instance (aka "functioning"): A process instance on the molecular level of granularity that is the result of the action of a gene product instance.

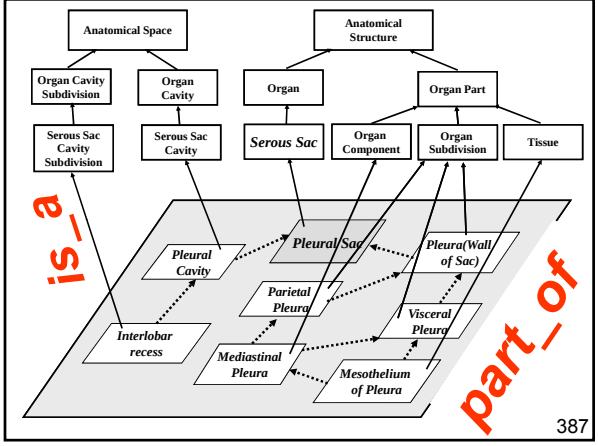
Molecular function execution type: A type of molecular function execution instance (aka "a type of functioning")

Warning re GO's use of the word 'activity'

The Foundational Model of Anatomy (FMA)

Department of Biological Structure, University of Washington, Seattle

Professor
Cornelius Rosse
Biological Structure



The FMA

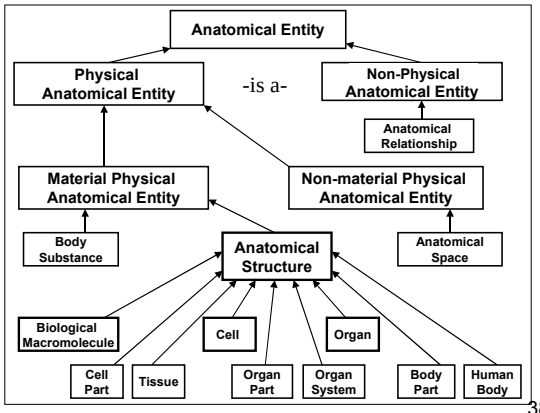
is organized in a graph-theoretical structure involving two principal sorts of links or edges:

is-a (= *is a subtype of*)

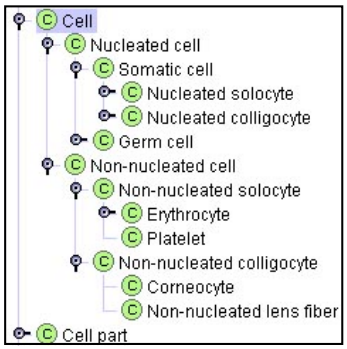
(*pleural sac* ***is-a*** *serous sac*)

part-of

(*cervical vertebra* ***part-of*** *vertebral column*)



at every level of granularity



anatomical structure (cell, lung, nerve, tooth)

result from the coordinated expression of structural genes

have their own 3-D shape

portion of body substance

inherits its shape from container

urine

menstrual flood

blood ...

anatomical space

cavities, conduits

anatomical attribute

mass

weight

temperature

your temperature

its value now

anatomical relationship

located_in

contained_in

adjacent_to

connected_to

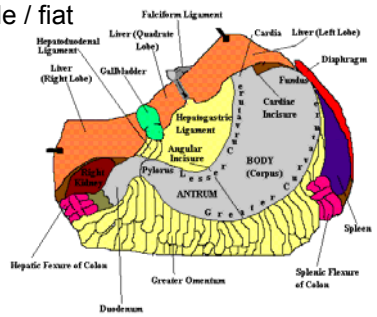
surrounds

lateral_to (West_of)

anterior_to

boundary

bona fide / fiat



Generalizing beyond the FMA

Model organism research seeks results valuable for the understanding of human disease.

This requires the ability to make reliable cross-species comparisons, and for this anatomy is crucial.

But different MOD communities have developed their anatomy ontologies in uncoordinated fashion.

Multiple axes of classification

Functional: cardiovascular system, nervous system
Spatial: head, trunk, limb
Developmental: endoderm, germ ring, lens placode
Structural: tissue, organ, cell
Stage: developmental staging series

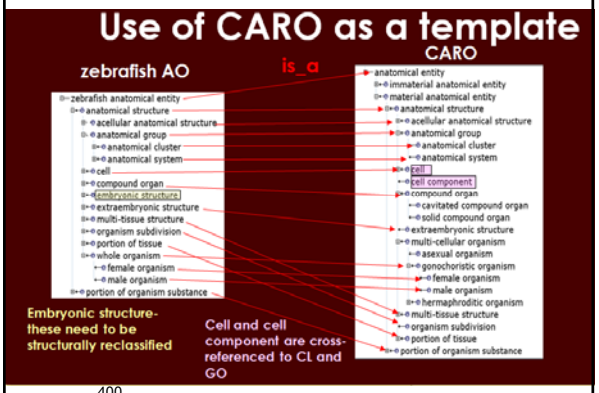
CARO – Common Anatomy Reference Ontology

for the first time provides guidelines for model organism researchers who wish to achieve comparability of annotations

for the first time provides guidelines for those new to ontology work

See Haendel et al., "CARO: The Common Anatomy Reference Ontology", in: Burger (ed.), *Anatomy Ontologies for Bioinformatics:* Springer, in press.

Use of CARO as a template



CARO-conformant ontologies already in development:

Fish Multi-Species Anatomy Ontology (NSF funding received)
Ixodidae and Argasidae (Tick) Anatomy Ontology
Mosquito Anatomy Ontology (MAO)
Spider Anatomy Ontology
Xenopus Anatomy Ontology (XAO)

undergoing reform: Drosophila and Zebrafish Anatomy Ontologies

The Infectious Disease Ontology

We have data

TBDB: Tuberculosis Database, including Microarray data
VFDB: Virulence Factor DB
TropNetEurop Dengue Case Data
ISD: Influenza Sequence Database at LANL
PathPort: Pathogen Portal Project
...

We need to annotate these data

to allow retrieval and integration of
- sequence and protein data for pathogens
- case report data for patients
- clinical trial data for drugs, vaccines
- epidemiological data for surveillance, prevention
- ...

Goal: to make data deriving from different sources comparable and computable

IDO needs to work with

Disease Ontology (DO) + SNOMED CT
Gene Ontology Immunology Branch
Phenotypic Quality Ontology (PATO)
Protein Ontology (PRO)
Sequence Ontology (SO)
...

We need common controlled vocabularies to describe these data in ways that will assure comparability and cumulation

What content is needed to adequately cover the infectious disease domain?
- Host-related terms (e.g. carrier, susceptibility)
- Pathogen-related terms (e.g. virulence)
- Vector-related terms (e.g. reservoir,
- Terms for the biology of disease pathogenesis (e.g. evasion of host defense)
- Population-level terms (e.g. epidemic, endemic, pandemic,)

IDO Processes

- process
  - colonization
  - damaging host cells and tissues in subo
  - dormancy period
  - emergence
  - epidemiological spread of disease
  - host recovery period
  - immunization
  - infection
  - infection treatment process
  - infectious disease course
  - infectious disease progression
  - initiation of infection
  - intrinsic incubation period
  - outbreak
  - pathogen adherence to host cell
  - pathogen evasion of host immune resp
  - pathogen life cycle
  - pathogen outbreak
  - pathogen spread through host
  - pathogenesis
  - penetration of epithelial barrier
  - penetration of host cell
  - prevention of infection
  - transmission

IDO Qualities

- quality of host
  - case of infectious disease
  - co-infected
  - immunity
  - immunocompromised
  - immunosuppressed
  - infected
  - refractoriness
  - resistance
  - susceptibility
  - swollen
  - temperature
- quality of host population
  - herd immunity
  - infectious disease
  - prevalence
- quality of infectious disease
  - endemic
  - incidence
- quality of pathogen
  - toxicity
- quality of pathogen
  - contagious
  - dormancy
  - drug resistance
  - endogenous
  - exogenous
  - invasive
  - latency
  - pathogenicity
  - tropism
  - virulence
  - zoonotic

IDO Roles

- role
  - colonizer
  - factor
    - adhesion factor
    - colonization factor
    - virulence factor
      - invasin
      - toxin
        - endotoxin
        - exotoxin
  - fusion protein
  - host
    - carrier
  - mobile genetic element
  - mode of transmission
  - pathogen
  - pathogenicity island
  - reservoir
  - route of entry
  - sign
  - symptom
  - treatment
  - vector

what is a role?

a realizable independent continuant that is not the consequence of the nature of the independent continuant entity which bears the role (contrast: disposition)

the role is optional (someone else assigns it, the entity acquires it by moving it into a specific context)

IDO provides a common template

IDO works like CARO.

It contains terms (like 'pathogen', 'vector', 'host') which apply to organisms of all species involved in infectious disease and its transmission

Disease- and organism-specific ontologies built as refinements of the IDO core

Disease-specific IDO test projects

MITRE, Mount Sinai, UTSouthwestern **– Influenza**
- **Stuart Sealfon, Joanne Luciano,**

IMBB/VectorBase – Vector borne diseases (*A. gambiae, A. aegypti, I. scapularis, C. pipiens, P. humanus*)
- Kristos Louis

Colorado State University – Dengue Fever
- Saul Lozano-Fuentes

Duke – Tuberculosis
- Carol Dukes-Hamilton

Cleveland Clinic – Infective Endocarditis
- Sivaram Arabandi

University of Michigan – Brucilosis
- Yongqun He

Agenda · Day 2

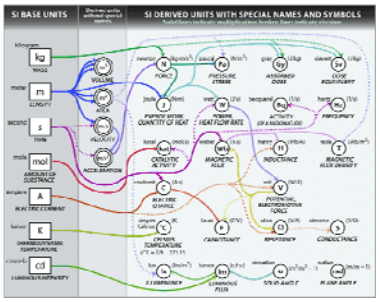
- An ontological introduction to biomedicine: Defining organism, function and disease
- The Gene Ontology (GO), the Foundational Model of Anatomy (FMA) and the Infectious Disease Ontology (IDO)
- **The OBO Foundry: A suite of biomedical ontologies to support reasoning and data integration**
- Applications of ontology outside biomedicine

In the olden days

people measured lengths using inches, ulnas, perches, king's feet, Swiss feet, leagues of Paris, etc., etc.

on June 22 1799 everything changed

we now have the International System of Units



The SI is a Controlled Vocabulary

Each SI unit is represented by a symbol, not an abbreviation. The use of unit symbols is regulated by precise rules.

The symbols are designed to be the same in every language.

Use of the SI system makes scientific results comparable

The SI is an Ontology

Quantities are universals

one each for each 'quantitative' dimension of reality

(= dimension which can be apportioned into homogeneous units, and thus associated with quantitative measures)

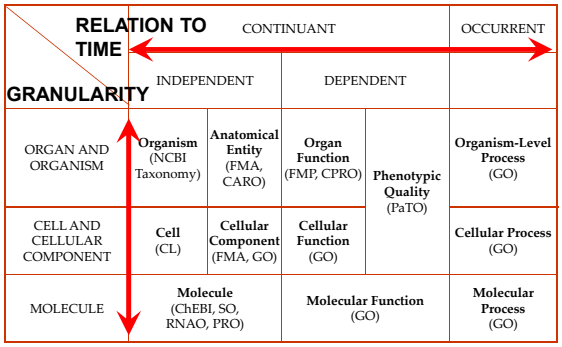
Goal of OBO Foundry

to provide a suite of controlled structured vocabularies for the callibrated annotation of data to support integration and algorithmic reasoning across the entire domain of biomedicine

current list of Foundry ontologies:
http://obofoundry.org

see also Coordinated Evolution of Ontologies to Support Biomedical Data Integration, Nature Biotechnology 25 (2007)

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO)	**Biological Process** (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	**Cellular Component** (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RNAO, PRO)		**Molecular Function** (GO)		Molecular Process (GO)



rationale of OBO Foundry coverage (homesteading principle)

421

OBO FOUNDRY CRITERIA

The ontology is open and available to be used by all.

The ontology is in a common formal language.

The developers of the ontology agree in advance to collaborate with developers of other OBO Foundry ontology where domains overlap.

422

OBO FOUNDRY CRITERIA

UPDATE: The developers of each ontology commit to its maintenance in light of scientific advance, and to soliciting community feedback for its improvement.

423

OBO FOUNDRY CRITERIA

IDENTIFIERS: The ontology possesses a unique identifier space within OBO.

VERSIONING: The ontology provider has procedures for identifying distinct successive versions.

The ontology includes textual definitions for all terms.

424

OBO FOUNDRY CRITERIA

CLEARLY BOUNDED: The ontology has a clearly specified and clearly delineated content.

DOCUMENTATION: The ontology is well-documented.

USERS: The ontology has a plurality of independent users.

425

OBO FOUNDRY CRITERIA

ORTHOGONALITY: They commit to working with other Foundry members to ensure that, for any particular domain, there is community convergence on a single controlled vocabulary.

426

OBO FOUNDRY CRITERIA

COMMON ARCHITECTURE: The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology

How to submit **ontologies** to the Foundry

First step is to join one or more mailing lists (http://obofoundry.org)

1.to become familiar with the Foundry's collaborative methodology

2.to identify members with overlapping expertise

3.submit new ontology resources for informal consideration by existing members

How to submit ***single terms*** to Foundry ontologies

Submit to ontology trackers/editor(s)

Orthogonality brings division of labor; so almost all development decisions can be made by the authors of single ontologies.

In cases of overlap, editors of involved ontologies will negotiate

In cases where these negotations bring no satisfactory outcomes, OBO Foundry editors adjudicate

All decisions are revisable

PROPOSED NEW CRITERIA

- OBO Foundry Ontologies should be organized in such a way as to reflect the top-level categories of dependent and independent / continuant and occurrent
- INSTANTIABILITY: Every term in an ontology should correspond to instances in reality
- Use singular nouns

PROPOSED NEW CRITERIA

- Use terms which form part of ordinary (including technical) English; do not use phrases like EV-EXP-IGI
- Use Aristotelian definitions (An A =def. a B which Cs)
- Employ cross-products and compositionality in building terms and definitions

THESE CRITERIA

provide **guidelines (traffic laws)** to new groups of ontology developers in ways which can ensure coordination of effort and provide for cumulation of benefits of lessons learned

The OBO Foundry map provides a navigational guide for those who need to find ontology resources

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RNAO, PRO)		Molecular Function (GO)		Molecular Process (GO)

building out from this original map

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)			
MOLECULE	(ChEBI, SO, RNAO, PRO)		Molecular Function (GO)			Molecular Process (GO)

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		Cellular Pathology ????	
MOLECULE	(ChEBI, SO, RNAO, PRO)		Molecular Function (GO)			Molecular Process (GO)

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function ???? (GO???)		Cellular Pathology ????	
MOLECULE	(ChEBI, SO, RNAO, PRO)		Molecular Function (GO)			Molecular Process (GO)

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)			
MOLECULE	(ChEBI, SO, RNAO, PRO)		Molecular Function (GO)			Molecular Process (GO)

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placeholder)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)			
MOLECULE	2- and 3-D Structure (RNAO)	(PRO)	Molecular Function (GO)			Molecular Process (GO)
	Small Molecule (ChEBI)	1-D Sequence (SO)				

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placehold er)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)			
MOLECULE	2- and 3-D Structure (RNAO)	(PRO)	Molecular Function (GO)			Molecular Process (GO) ?????
	Small Molecule (ChEBI)	1-D Sequence (SO)				Reactome

ORGAN AND ORGANISM	Organism (NCBI Taxonomy / placeholder)	Anatomical Entity (FMA, CARO)	Organ Function (placehold er)	Phenotypic Quality (PATO)	Disease (DO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)			
MOLECULE	2- and 3-D Structure (RNAO)	(PRO)	Molecular Function (GO)	Phenotypic Quality of Molecule ????		Molecular Process (GO) ?????
	Small Molecule (ChEBI)	1-D Sequence (SO)				Reactome

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO)	**Biological Process** (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	**Cellular Component** (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RNAO, PRO)		**Molecular Function** (GO)		Molecular Process (GO)

Building out from the original GO

clinical data includes

- clinical records
- clinical trial data
- demographic data
- National Hospital Discharge Survey
- National Ambulatory Medical Care Surveys
- MEDPAR
- Medicare's national claims data base

Community / Population Ontology

- family, clan
- ethnicity
- religion
- diet
- social networking
- education (literacy ...)
- healthcare (economics ...)
- household forms
- demography
- public health
- ...

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO)	**Biological Process** (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	**Cellular Component** (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)		**Molecular Function** (GO)		Molecular Process (GO)

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	**Family, Community, Deme, Population**		Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO)	**Biological Process** (GO)
	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)			
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)		**Molecular Function** (GO)		Molecular Process (GO)

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
COMPLEX OF ORGANISMS	**Family, Community, Deme, Population**		Organ Function (FMP, CPRO)	**Population Phenotype**	**Population Process**
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)		Phenotypic Quality (PaTO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)		**Molecular Function** (GO)		Molecular Process (GO)

The Environment Ontology

OBO Foundry
Genomic Standards Consortium
National Environment Research Council (UK)
USDA, Gramene, J. Craig Venter Institute, …

http://environmentontology.org/

RELATION TO TIME / GRANULARITY	CONTINUANT				OCCURRENT
	INDEPENDENT			DEPENDENT	
COMPLEX OF ORGANISMS	**Family, Community, Deme, Population**		ENVIRONMENT	Organ Function (FMP, CPRO)	**Population Phenotype**	**Population Process**
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	(FMA, CARO)			Phenotypic Quality (PaTO)	Biological Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cell Com-ponent (FMA, GO)		Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)			**Molecular Function** (GO)		Molecular Process (GO)

RELATION TO TIME / GRANULARITY	CONTINUANT			
	INDEPENDENT			
COMPLEX OF ORGANISMS	**Family, Community, Deme, Population**		ENVIRONMENT	← **Environment of population**
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	(FMA, CARO)		← **Environment of single organism**
CELL AND CELLULAR COMPONENT	Cell (CL)	Cell Com-ponent (FMA, GO)		← **Environment of cell**
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)			← **Molecular environment**

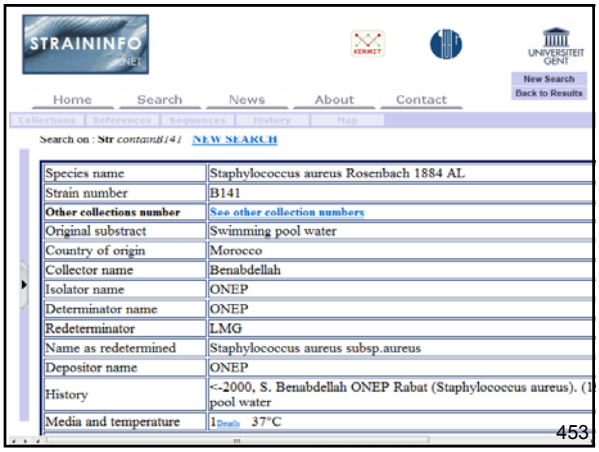
RELATION TO TIME / GRANULARITY	CONTINUANT			
	INDEPENDENT			
COMPLEX OF ORGANISMS	Family, Community, Deme, Population		ENVIRO	← **Environment of population**
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	(FMA, CARO)		← **Environment of single organism***

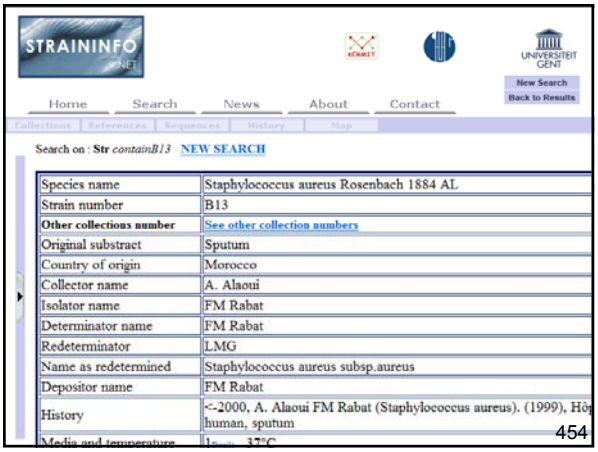
* The sum total of the conditions and elements that make up the surroundings and influence the development and actions of an individual.

RELATION TO TIME / GRANULARITY		CONTINUANT
		INDEPENDENT
COMPLEX OF ORGANISMS	ENVIRONMENT	biome / biotope, territory, habitat, neighborhood, ...
ORGAN AND ORGANISM		work environment, home environment; host/symbiont environment; ...
CELL AND CELLULAR COMPONENT		extracellular matrix; chemokine gradient; ...
MOLECULE		hydrophobic surface; virus localized to cellular substructure; active site on protein; pharmacophore ...

Applications of EnvO in biology

- Support the annotation of meta-data related to:
  - Data about biological samples produced from various technologies
    - Metagenomics, Metabolomics, Proteomics, Transcriptomics, Genomics...
  - Data produced from remote sensing equipment
  - Images
    - Web 2.0, tagging
  - Physical holdings
    - Museum artifacts, (preserved) biological samples / organisms
  - ...anything that has an ***environment***





to enhance coordination of research

Obesity, Physical Activity & Built Space in New York City
Environmental and Family Influences on Adolescent Overweight
Walking on Campus: Correlates & Web Tools
Measuring Physical Activity Affordances in Preschool Outdoor Environments
Environmental Predictors of Elderly Obesity
Neighborhood Food Environment, Diet & Health: Quasi-Experimental Study

How EnvO currently works for information retrieval

Retrieve all experiments on organisms obtained from:
- deep-sea thermal vents
- arctic ice cores
- rainforest canopy
- alpine melt zone

Retrieve all data on organisms sampled from:
- hot and dry environments
- cold and wet environments
- a height above 5,000 meters

Retrieve all the omic data from soil organisms subject to:
- moderate heavy metal contamination

Environment = totality of circumstances external to a living organism or group of organisms

- pH
- evapotranspiration
- turbidity
- available light
- predominant vegetation
- predatory pressure
- nutrient limitation …

extending EnvO to the clinical domain

- dietary patterns (Food Ontology: FAO, USDA) ... allergies
- neighborhood patterns
  - built environment, living conditions
  - climate
  - social networking
  - crime, transport
  - education, religion, work
  - health, hygiene
- disease patterns
  - bio-environment (bacteriological, ...)
  - patterns of disease transmission (links to IDO)

a new type of patient data

a patient's *environmental history*

use EnvO and the community ontology to mine relations between disease phenotypes and environmental patterns and patterns of community behavior

e.g. for cows

Another way the OBO Foundry is being used

The Senselab/NeuronDB* comprehends three types of neuronal properties:

voltage gated conductances

neurotransmitter receptors

neurotransmitter substances

Many questions immediately arise: what *are* receptors? Proteins? Protein complexes? The Foundry framework provides an opportunity to evaluate such choices.

* http://senselab.med.yale.edu/

Senselab/NeuronDB

The GO Molecular Function (MF) ontology already has classes such as *receptor activity* (GO_0004872) plus subclasses describing receptor activities already referred to in NeuronDB.

This provides a roadmap for further development. Review the 130 receptor classes to see if they exist in MF, where not, create subclasses and submit to GO for future inclusion. We can then e.g. take advantage of GO Annotations to find the proteins that correspond to these receptor classes in different species.

Agenda · Day 2

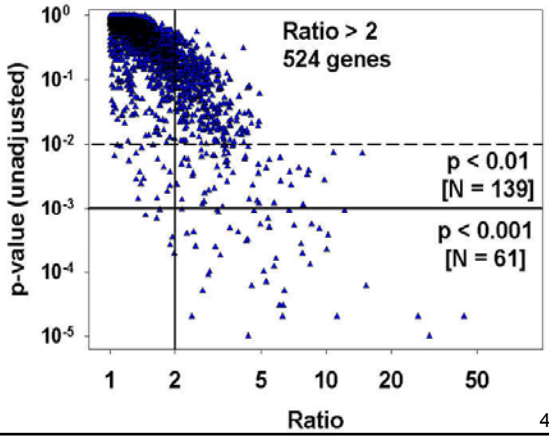
- An ontological introduction to biomedicine: Defining organism, function and disease
- The Gene Ontology (GO), the Foundational Model of Anatomy (FMA) and the Infectious Disease Ontology (IDO)
- The OBO Foundry: A suite of biomedical ontologies to support reasoning and data integration
- **Applications of ontology outside biomedicine**

Towards an ontology of science

To make experimental data computationally accessible we need ontologies to describe the data

(1) from the point of view of their relation to *biological reality*

(2) from the point of view of the *evidence* that supports them



The problem of data provenance

High throughput experimentation data is meaningless unless the users of the data have detailed information concerning how it was obtained

- which protocoll
- which staining
- which equipment
- which settings
- which statistical tools ...

We need to annotate data

in terms of how the data was obtained and processed

A new kind of ontology is required, an ontology of experimental design, evidence, statistics, data transformations applied ...

Three proposals

EXPO: The Experiment Ontology

The MGED Ontology

OBI: The Ontology for Biomedical Investigations

EXPO

The Ontology of Experiments

L. Soldatova, R. King
Department of Computer Science
The University of Wales, Aberystwyth

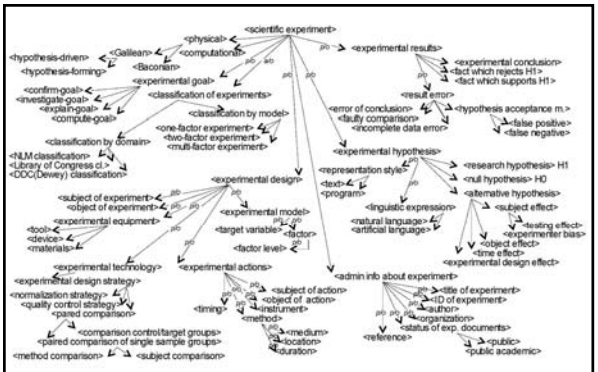
EXPO Formalisation of Science

The goal of science is to increase our knowledge of the natural world through the performance of experiments.

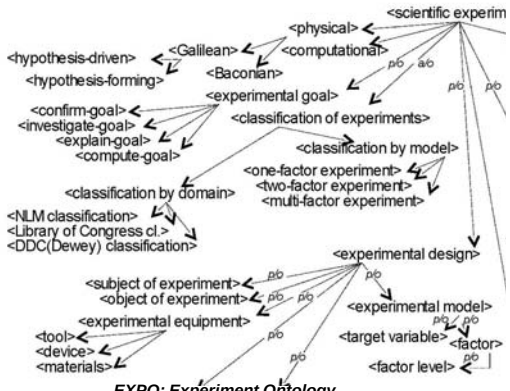
This knowledge should, ideally, be expressed in a formal logical language.

Formal languages promote semantic clarity, which in turn supports the free exchange of scientific knowledge and simplifies scientific reasoning.

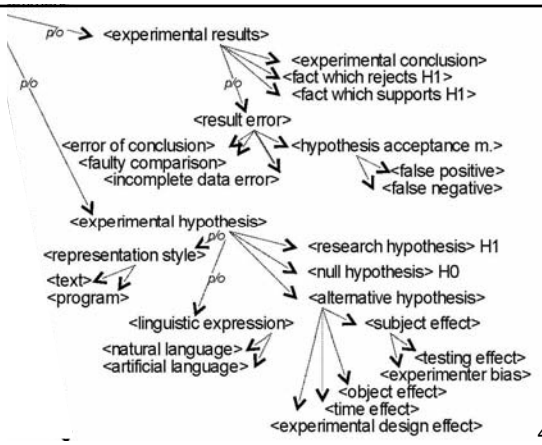
We need a formal language to describe experiments

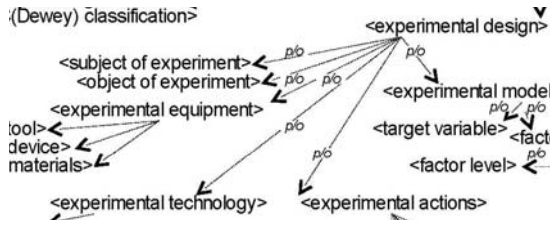


EXPO: Experiment Ontology



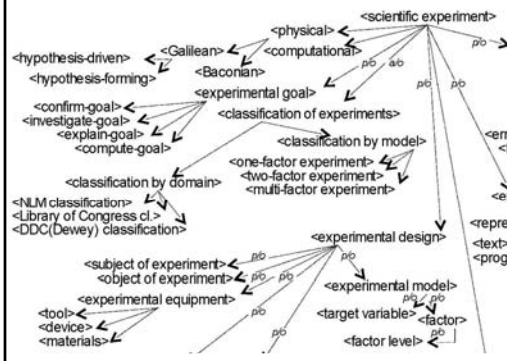
EXPO: Experiment Ontology





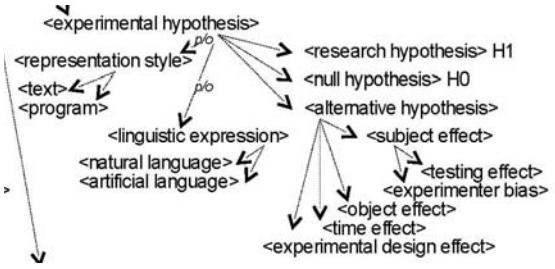
experimental actions part_of experimental design
subject of experiment part_of experimental design

representational style part_of experimental hypothesis

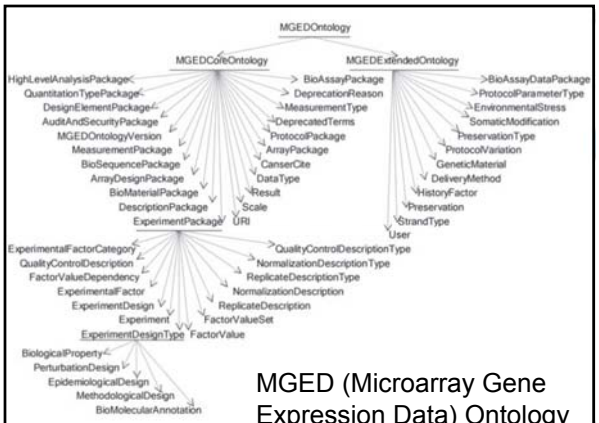


equipment part_of experimental design
(confuses object with specification)

Role of Philosophy of Science



EXPO: Experiment Ontology



MGED (Microarray Gene Expression Data) Ontology

MGED Ontology

Individual =def. name of the individual organism from which the biomaterial was derived

Experiment =def. The complete set of bioassays and their descriptions performed as an experiment for a common purpose. ... An experiment will be often equivalent to a publication.

MGED Ontology

Chromosome =Def A biological sequence that can be placed on an array

Chromosome =Def An abstraction used for annotation

OBI

The Ontology for Biomedical Investigations

To provide a resource for the unambiguous description of the components of biomedical investigations such as the design, protocols and instrumentation, material, data and universals of analysis and statistical tools applied to the data

OBI Collaborating Communities

Crop sciences Generation Challenge Programme (GCP),
Environmental genomics MGED RSBI Group,
www.mged.org/Workgroups/rsbi
Genomic Standards Consortium (GSC),
www.genomics.ceh.ac.uk/genomecatalogue
HUPO Proteomics Standards Initiative (PSI), psidev.sourceforge.net
Immunology Database and Analysis Portal, www.immport.org
Immune Epitope Database and Analysis Resource (IEDB),
http://www.immuneepitope.org/home.do
International Society for Analytical Cytology, http://www.isac-net.org/
Metabolomics Standards Initiative (MSI),
Neurogenetics, Biomedical Informatics Research Network (BIRN),
Nutrigenomics MGED RSBI Group, www.mged.org/Workgroups/rsbi
Polymorphism
Toxicogenomics MGED RSBI Group, www.mged.org/Workgroups/rsbi
Transcriptomics MGED Ontology Group

Background of OBI

http://obi.sf.net

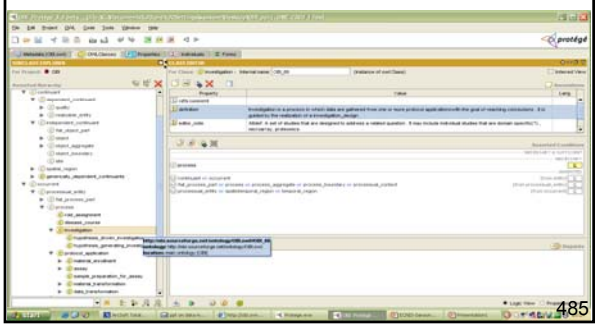
Omics standardization effort initiatives (Genomic Standards Consortium, MGED, PSI, MSI)
Semantic web
BIRN Biomedical Informatics Research Network
European Bioinformatics Institute
National Cancer Institute
Vendors and manufacturers (ontologically organized catalogs)

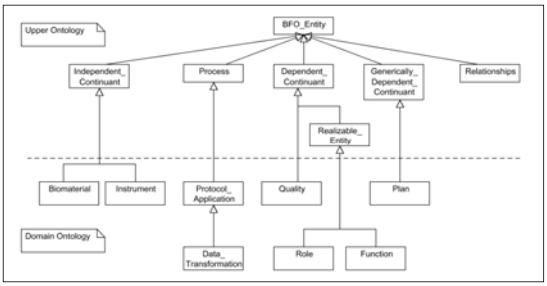
- Plurality of (prospective) uses
  - Driving data entry and annotation
    - Indexing of experimental data, minimal information lists, x-db queries
  - Text-mining
    - Benchmarking, enrichment, annotation
  - Encoding facts from literature
- Long term
  - Algorithmic science

OBI – Tools and Documentation

http://obi.sf.net

- Open source, standards compliant, version management
  - Ontology Web Language (OWL) using Protégé editor
  - OBI.owl files are available from the OBI SVN Repository





List of OBI Branches

- Biomaterial (lead editor: Susanna Sansone)
- Data Transformation (Tina Boussard)
- Digitial Entity and Non-realizable Information Entity (Chris Stoeckert)
- Function (Bill Bug)
- Instrument (Daniel Schober)
- Plan (investigation design - protocol) (Philippe Rocca-Serra)
- Protocol Application (Bjoern Peters)
- Qualities (Chris Mungall)
- Role (Jennifer Fostel)
- Relations (Liju Fan)

Ontology of Biomedical Investigation

Function Branch Report

with thanks to Bill Bug, BIRN OTF, UC San Diego

OBI Functions

- the function of a ***birth canal*** to enable ***transport***
- the function of the ***heart*** in the ***body*** to ***pump blood***
- the function of ***reproduction*** in the ***transmission*** of ***genetic material***
- the digestive function of the ***stomach*** to ***nutriate*** the ***body***
- the function of a ***hammer*** to ***drive*** in ***nails***
- the function of a ***computer program*** to ***compute mathematical equations***
- the function of an ***automobile*** to provide ***transportation***
- the function of a ***judge*** in a ***court of law***

OBI: Function

the function of a ***heart*** to ***pump blood***

the function of a ***high pressure liquid chromatagraphic (HPLC) system*** to ***separate molecules*** based on their ***solubility properties***

the function of the ***Tail Flick Analgesia test*** to ***measure pain sensitivity*** in ***mice*** and ***rats*** as they ***respond*** to the ***application of heat*** to a ***small area of their tails***.

the function of an ***antibody-coated Enzyme-linked Immunosorbant Assay (ELISA) multi-well plate*** to ***identify the presence*** of a ***specific molecule*** based on its ***matching epitopes binding*** to the ***immobilized antibodies*** coating the plate wells;

the function of the ***Cy5 coupled-ligand*** to ***separate cells*** in a ***Fluorescence-Activated Cell Sorter (FACS)***

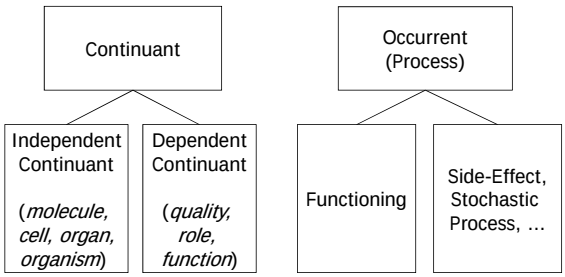
the function of ***semi-permeable dialysis tubing*** to ***separate solutes*** by selectively ***restricting diffusion*** by solute ***size*** and ***generating osmotic pressure***.

the function of an ***electromagnetic lens*** in an ***electron microscope*** to ***direct the trajectory*** of the incident ***electron beam*** to systematically ***raster*** across a ***specimen*** to ***construct*** a composite ***image***.

Institutional Entities

Research teams
Funding agencies
Regulatory bodies
IRBs
Vendors
Manufacturers
...

What is an organization?



Towards an Ontology of Information Entities

Information Entities in Science

protocol
database
ontology
gene list
publication
result
...

Information Entities in Scientific Experimentation

serial number
batch number
grant number
person number
name
(building) address
email address
URL
...

What is a credit card number?

1. not a mathematical object (Plato)
2. not a contingent object with physical properties, taking part in causal relations
3. but a historical object, with a very special provenance, relations analogous to those of ownership, existing only within a nexus of institutions of certain types

W3C Schemaweb Ontologies (http://www.schemaweb.info/)

MusicBrainz Metadata Vocabulary

Musical Baton Vocabulary

Beer Ontology

Kissology

Ebiquity Publication Ontology

http://ebiquity.umbc.edu/ontology/publication.owl

```
- <owl:ObjectProperty rdf:ID="author">
  <rdfs:label>Resource Author</rdfs:label>
  <rdfs:domain rdf:resource="#Resource" />
  <rdfs:range
    rdf:resource="http://ebiquity.umbc.edu/ontology/person.owl#Person
    " />
```

author_of is a relation between a resource and a person

Ebiquity Publication Ontology

http://ebiquity.umbc.edu/ontology/publication.owl

```
- <owl:DatauniversalProperty rdf:ID="chapter">
  <rdfs:label>Publication Chapter</rdfs:label>
  <rdfs:domain rdf:resource="#Publication" />
  <rdfs:range
    rdf:resource="http://www.w3.org/2001/XMLSchema#string" />
```

chapter_of is a relation between a publication and a string

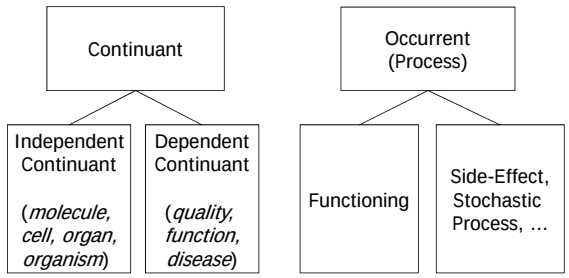
What is a string?

XML Schema Part 2: Datatypes

http://www.w3.org/TR/xmlschema-2/#string

[Definition:] The **string** datatype represents character strings in XML. The ·value space· of **string** is the set of finite-length sequences of characters (as defined in [XML 1.0 (Second Edition)]) that ·match· the Char production from [XML 1.0 (Second Edition)]. A character is an atomic unit of communication; it is not further specified except to note that every character has a corresponding Universal Character Set code point, which is an integer.

What is a protocol?



Is a protocol a string?

Nature Protocols

vs.

The protocol McDoe has been following in project #334 since March

universals and instances

universal: human being
Instance: Leo Tolstoy

universal: novel
Instance: *War and Peace*

universal: book
Instance: this copy of *War and Peace*

Rule for universals: their names are pluralizable
There are two laptops, two rabbits, …
There cannot be two Leo Tolstoys

Specific vs. generic dependence

The pdf file which was just copied from your laptop to my laptop

The novel *War and Peace*

The UniProt database

The Gene Ontology

What is a database?

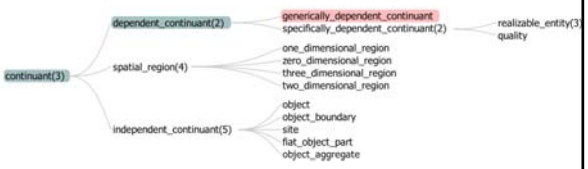
Is UniProt a universal or an instance?

If UniProt were a universal, and the copy of UniProt on my laptop were an instance, then

1. universals would include massively arbitrary kluges (is *War and Peace* a universal?)
2. there would be many UniProts and many *War(s) and Peaces.*

Hence UniProt is an instance.

Generically dependent continuant

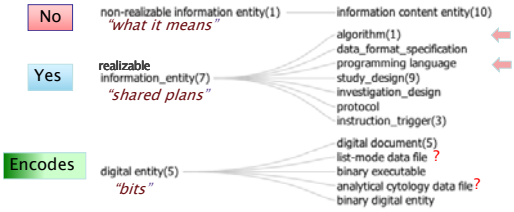


A continuant that is dependent on one or other independent continuant bearers. Every instance of a generically dependent continuant A requires at least one instance of an independent continuant B to bear it but *which* instance(s) of B do this can change from time to time.

The bearers: People, organizations, paper, computer hardware, magnetic and optical media

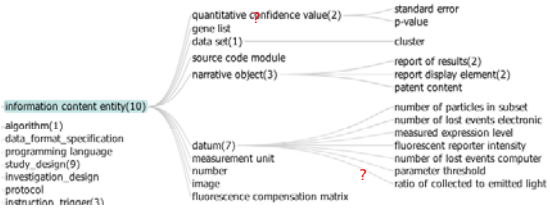
with thanks to Alan Ruttenberg, Neurocommons

Realizables

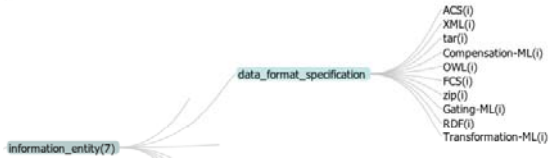


Plans, by their nature, suggest something will be happening. Non-realizable information entities do not. Digital entities encode either.

Information Content Entities



The case for instances in OBI?

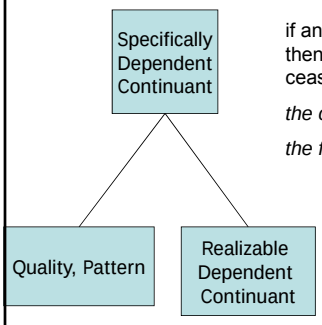


Are "file format" annotations best represented as
- instances of specifications?
- universals of digital objects?

Information objects

pdf file
poem
symphony
algorithm
symbol
sequence
molecular structure

Specifically Dependent Continuants

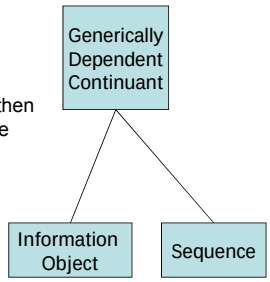


if any bearer ceases to exist, then the quality or function ceases to exist

the color of my skin

the function of my heart

Generically Dependent Continuants



if one bearer ceases to exist, then the entity can survive, because there are other bearers

the pdf file on my laptop

the DNA (sequence) in this chromosome

Generically dependent continuants

are *realized* through being concretized in specifically dependent continuants

(the plan in your head, the protocol being realized by your research team)

Generically dependent continuants are distinct from types / universals

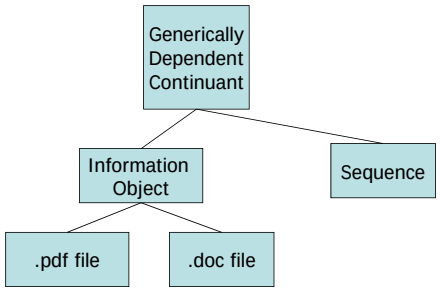
they have a different kind of provenance

- **'a'** as universal (type)
- **'a'** as letter of the Roman alphabet
- aspirin as product of Bayer GmbH
- aspirin as molecular structure

Если должник перестал выплачивать ссуду, и кредитор требует чтобы вы, как поручитель или совместный заемщик, взяли на себя выплаты ссуды или угрожает подать на вас в суд, обратитесь за юридической консультацией.

Внекоторыхограниченныхситуациях,поручитель может оспорить иск, даже если он подписал контракт. Вы должны незамедлительно обратиться за советом, если вы считаете, что работник учреждения, которое выдало ссуду, брокер или продавец, который организовал финансы, обманул или ввел вас в заблуждение. Вы также должны незамедлительно обратиться за юридической консультацией, если на момент подписания контракта:

Generically Dependent Continuants



instances

Generically dependent continuants

are concretized in specifically dependent continuants

Beethoven's 9th Symphony is concretized in the pattern of ink marks which make up this score in my hand

Generically dependent continuants

do not require specific media (paper, silicon, neuron …)

Towards an Ontology of Information

Basic rule of evidence-based ontology: all terms in an ontology must have instances in reality

Ontologies must be anchored to reality

How anchor information (propositions, logical content)?

First: through human acts of using language

The Ontology of Speech Acts

requesting, questioning, answering, ordering, imparting information, promising, commanding, baptising

Social acts which "are performed in the very act of speaking"

Some social acts can be purely internal

envy

forgiveness

waiving a claim

Some social acts depend on uptake

they must be not only directed towards other people

but also registered by their addressees

523

Some social acts depend on external circumstances

For example *commands, marryings, baptisings*

depend on

relations of authority

524

Some social acts give rise to successor entities

Promising gives rise to claims and obligations (e.g. to debts)

Marrying gives rise to marital bond

Promoting gives rise to new role on the part of the promotee

525

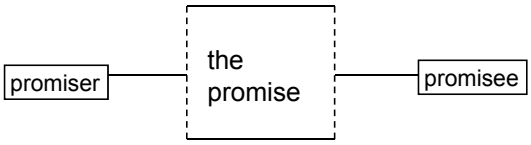
Some social acts give rise to tendencies

Promising, commands, requests gives rise to tendencies to realization of their content

Tendencies can be blocked …

526

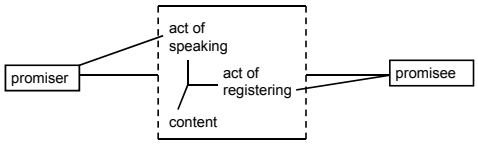
The Structure of the Promise



relations of one-sided dependence

527

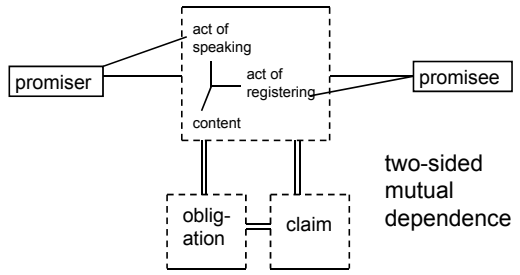
The Structure of the Promise



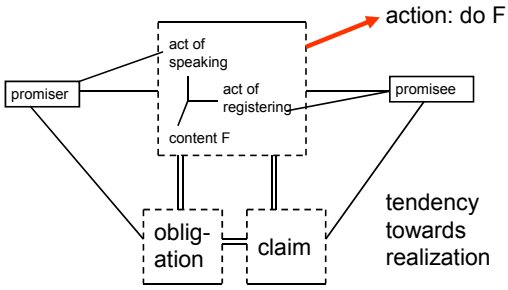
three-sided mutual dependence

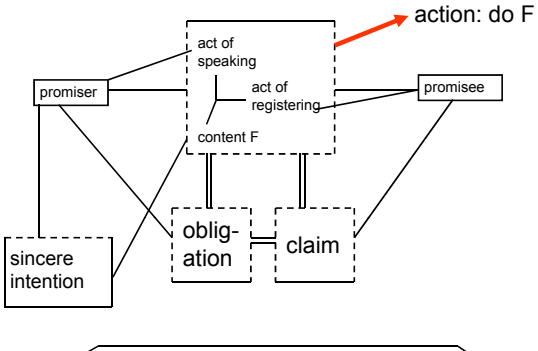
528

The Structure of the Promise



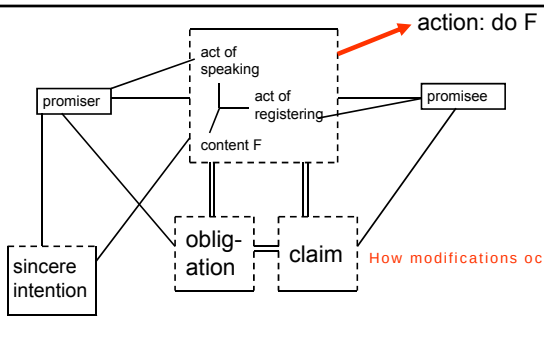
The Structure of the Promise

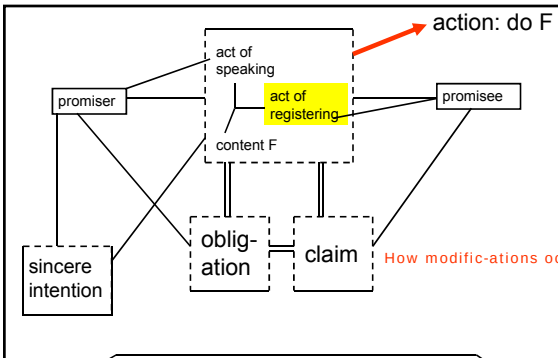


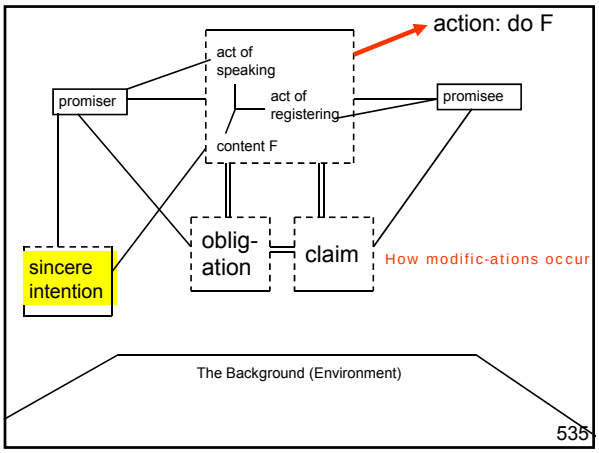


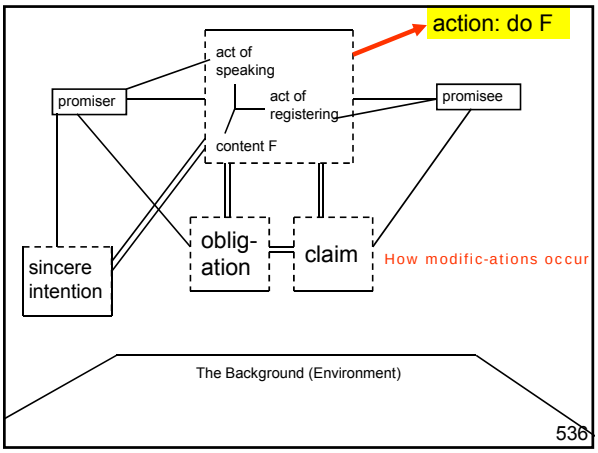
Modifications of Social Acts

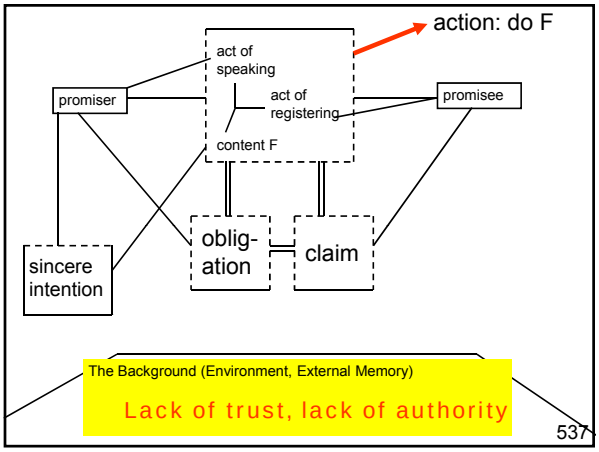
- Sham promises
- Lies as sham assertions (cf. a forged signature); rhetorical questions
- Social acts performed in someone else's name (representation, delegation)
- Social acts with multiple addresses
- Conditional social acts











The Ontology of Claims and Obligations (Endurants)

Debts
Offices, roles
Licenses
Prohibitions
Rights
Laws

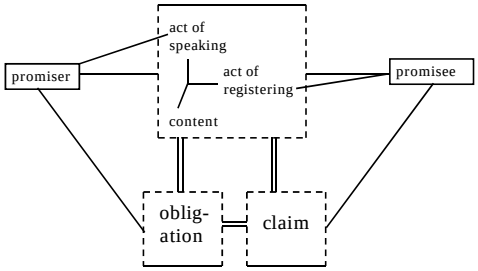
538

Three sorts of objects

1. Necessary Objects (intelligible; timeless) – e.g. the number 7 (Plato)
2. Contingent Objects (knowable only through observation; historical; causal) – e.g. Bill Clinton (positivists)
3. Objects of the third kind (intelligible, but have a starting point in time) – e.g. claims, obligations …

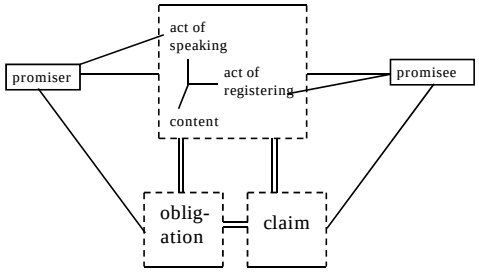
539

Material Ontology of Social Interaction

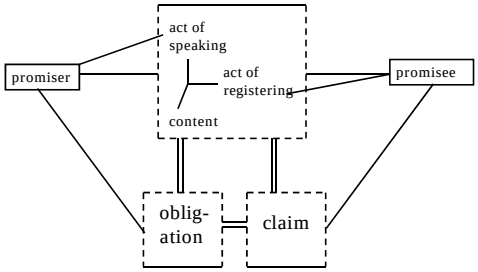


540

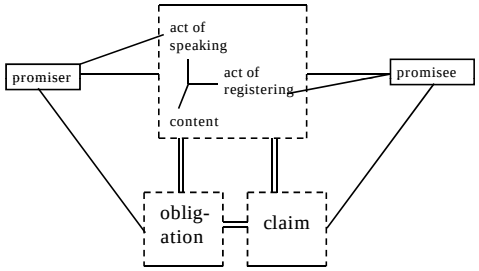
A Window on Reality



Universals



Instances



Biomedical Ethics Ontology

Continuants
- Subject
  - Animal
  - Human
- Sample
  - Tissue
    - Human tissue
    - Animal tissue
- Institutional Review Board
  - IRB member
  - IRB Chair
- Document
  - Study Design
  - Human Subject Study Application
  - Consent form

Occurrents
- Study
- Review
  - Full review
  - Continuing review
  - Expedited Review
- Event
  - Adverse event
    - Related
    - Non-related
- Ethical Duty
- Ethical Lapse
- Risk
  - Minimal risk
  - Non-minimal risk