

# Introduction to Ontology Introductions

Alan Ruttenberg  
Science Commons

# Underlying assumption

- A realist ontology is concerned with enumerating those things that exist and how they relate to each other, in particular those things they existentially depend on.
- This enables a kind of audit trail from representations to elements of the world which can be checked.
- A realist ontology is one that has the best chance of being effectively communicated, it's creation objectively moderated, and as a result being widely accepted.

# Questions to ask

- If you say there is a class, what are the particulars (syn: members, individuals, instances, entities)?
- What are the entities dependent on – without what else can't they exist?
- When do entities come into existence. When do they go out of existence?
- Is it with respect to a certain perspective (e.g. a certain granularity) that an entity can be referred to?

# The open world assumption

- Assertions we make are not (and need not be) complete
- What we don't state, or can not infer from what is stated, we have no knowledge of
- When we have all possibilities enumerated, we explicitly say so – “closure”.
- Something to watch out for – ontology may expose detail and complication – this does not imply our representations must record all of that

# Reusing terms is hard, but essential

- Some existing OBO ontologies:
  - Biological process
  - Cellular component
  - Foundational model of anatomy
  - Chemical entities of biological interest
  - Sequence ontology
  - PATO – Phenotypes/Qualities
- Each has lots of good content. If we don't reuse them we lose in at least two ways:
  - We have to redefine the terms we need
  - We have to do work to do data integration
- It's hard because we don't have adequate tools – in many cases we haven't even *conceived* of what the form of such tools will be.

# Fewer relations are better

- Most queries use relations as connectives and bind classes or instances to variables
- The more relations we have, the more possible queries we have (from a topological point of view)
- Reasoning systems have more expressivity to reason about classes than about relations
- If we can constrain the number of possibilities, we have a higher chance that queries return results
- I start with the OBO Relations ontology and add sparingly as necessary

# You: Practice thinking about upper ontology

- As I describe entities – assess whether they are
  - Continuant
    - Dependent continuant
      - Specifically Dependent
      - Generically Dependent
        - » Information content entity
    - Independent continuant
  - Occurrent
- If you don't know: ASK!

# An introduction to the Ontology of Biomedical Investigations (OBI)

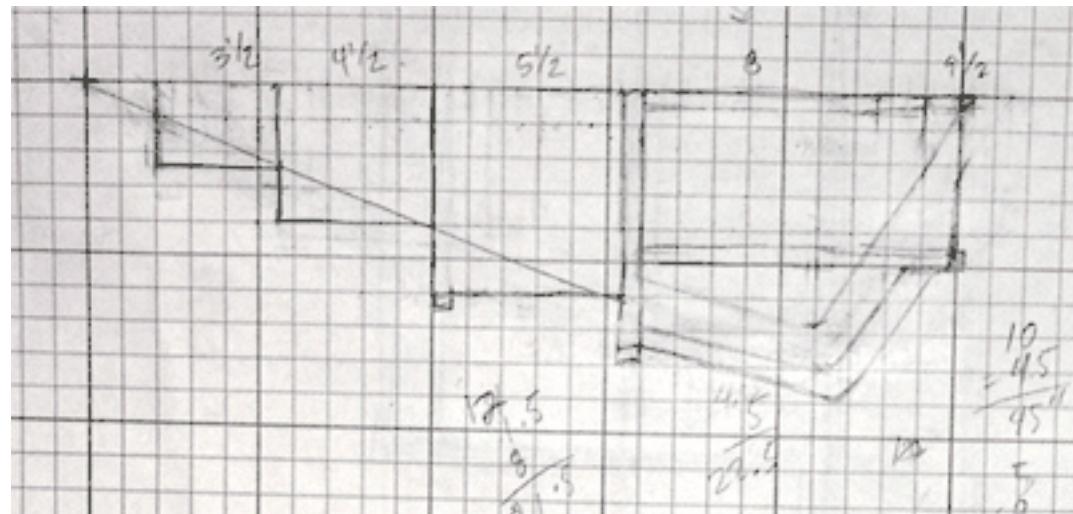
# The OBI project

- OBI is an ontology covering multiple scientific domains
- Built under the Basic Formal Ontology
- Collaboration among 19 groups – many with data repositories

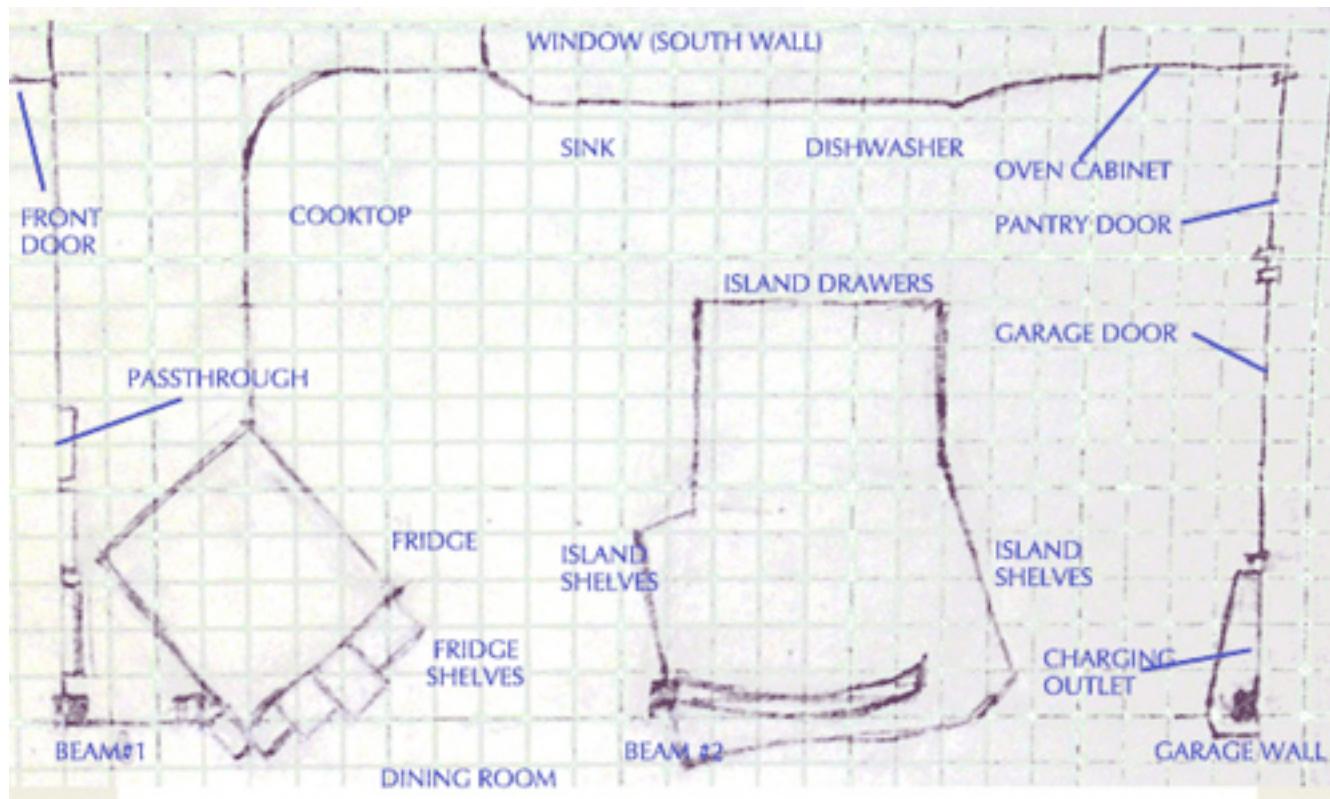
# OBI

- What's the domain?
  - materials made and produced for investigations
  - research objectives
  - experimental protocols
  - roles of people in investigations
  - processing and publication of data gathered in investigations

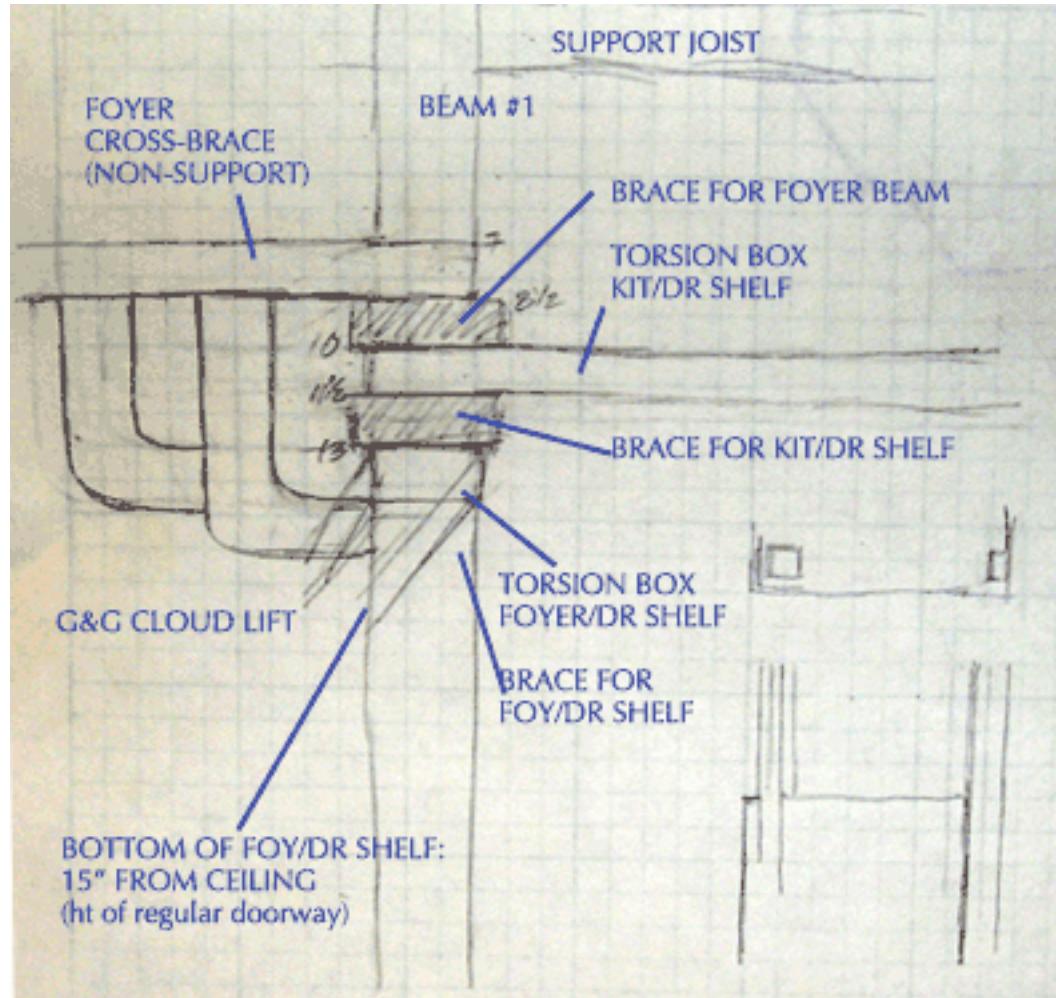
# A very rough sketch



# Add some more detail



# Even more detail



[http://www.flickr.com/photos/pricklypear\\_projects](http://www.flickr.com/photos/pricklypear_projects)

Some rights reserved

# We are somewhere in the middle of detail for OBI

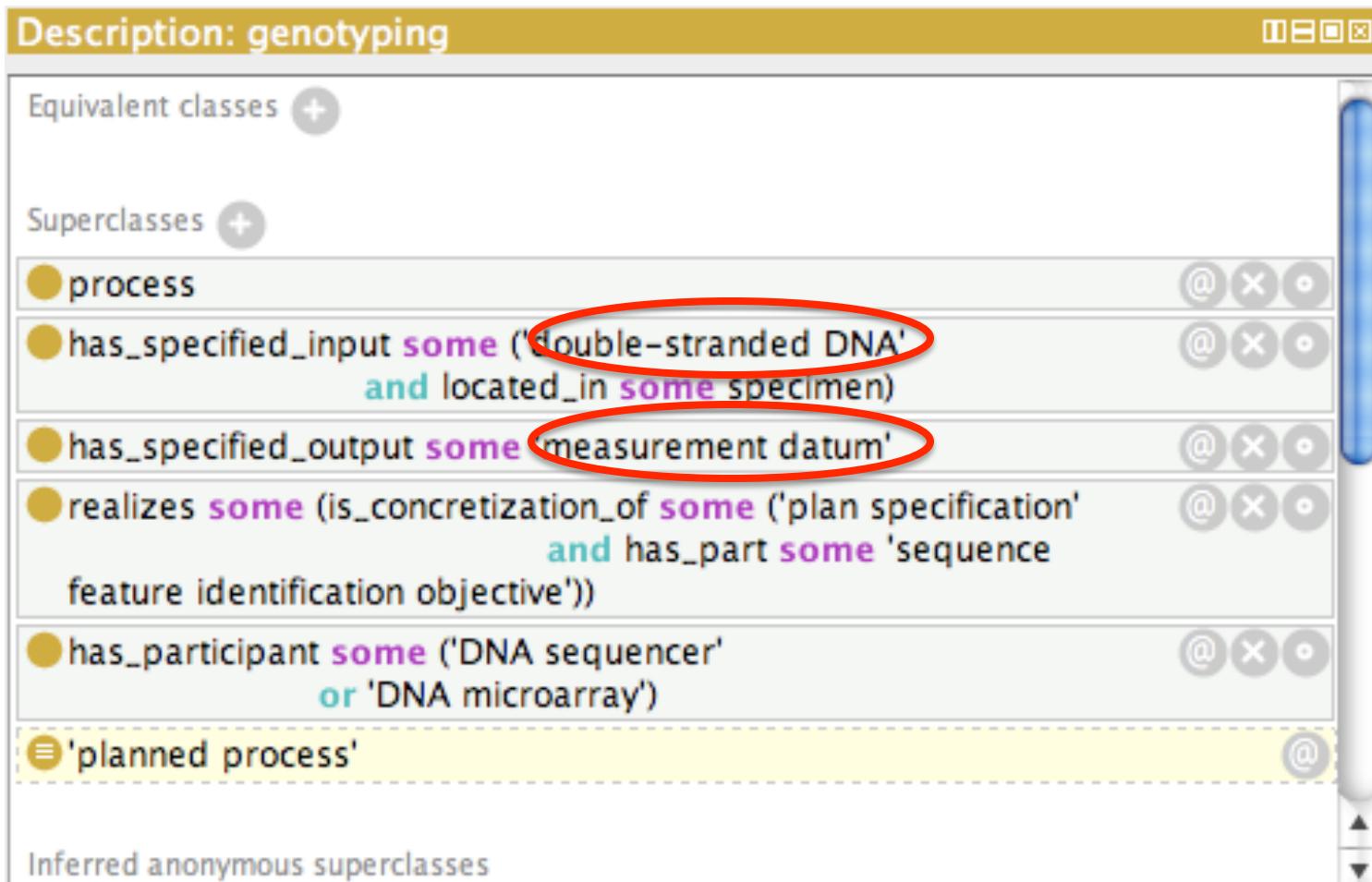
Description: genotyping

Equivalent classes +

Superclasses +

- process @ ✘ ○
- has\_specified\_input some ('double-stranded DNA'  
and located\_in some specimen) @ ✘ ○
- has\_specified\_output some measurement datum' @ ✘ ○
- realizes some (is\_concretization\_of some ('plan specification'  
and has\_part some 'sequence  
feature identification objective')) @ ✘ ○
- has\_participant some ('DNA sequencer'  
or 'DNA microarray') @ ✘ ○
- 'planned process' @

Inferred anonymous superclasses



# *Material and methods* in journal articles

## Materials

**N2a** cells were purchased from American Type Culture Collection. Insulin, IGF-1, IL-4, and dox were purchased from Sigma-Aldrich. Antibodies were purchased from Upstate Biotechnology (anti-IRS-2, anti-phospho Akt, and anti-phospho S6 kinase), BD Biosciences (Akt, S6 kinase, and IRS-1), Cell Signaling Technology (anti-phospho-mTOR and anti-mTOR), and Roche (anti-GFP). htt-exon1 (CAGCAA) constructs were obtained from A. Kazantsev (Massachusetts General Hospital, Charlestown, MA). pYFP-LC3 was obtained from T. Yoshimori (National Institute of Genetics, Mishima, Shizuoka, Japan).

<http://www.atcc.org/ATCCAdvancedCatalogSearch/ProductDetails/tabid/452/Default.aspx?ATCCNum=CCL-131&Template=cellBiology>

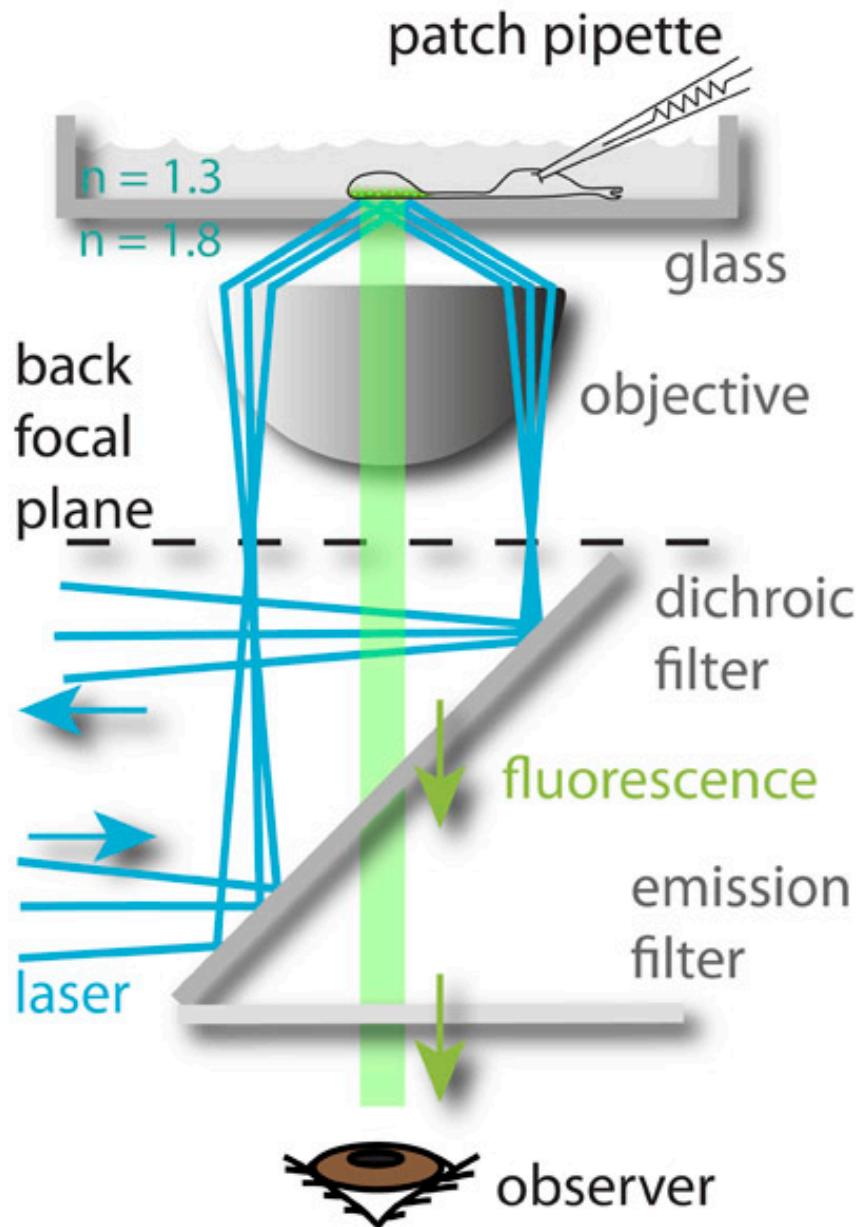
Search pubmed for N2a cells, neuro-2a cells, CCL-131...

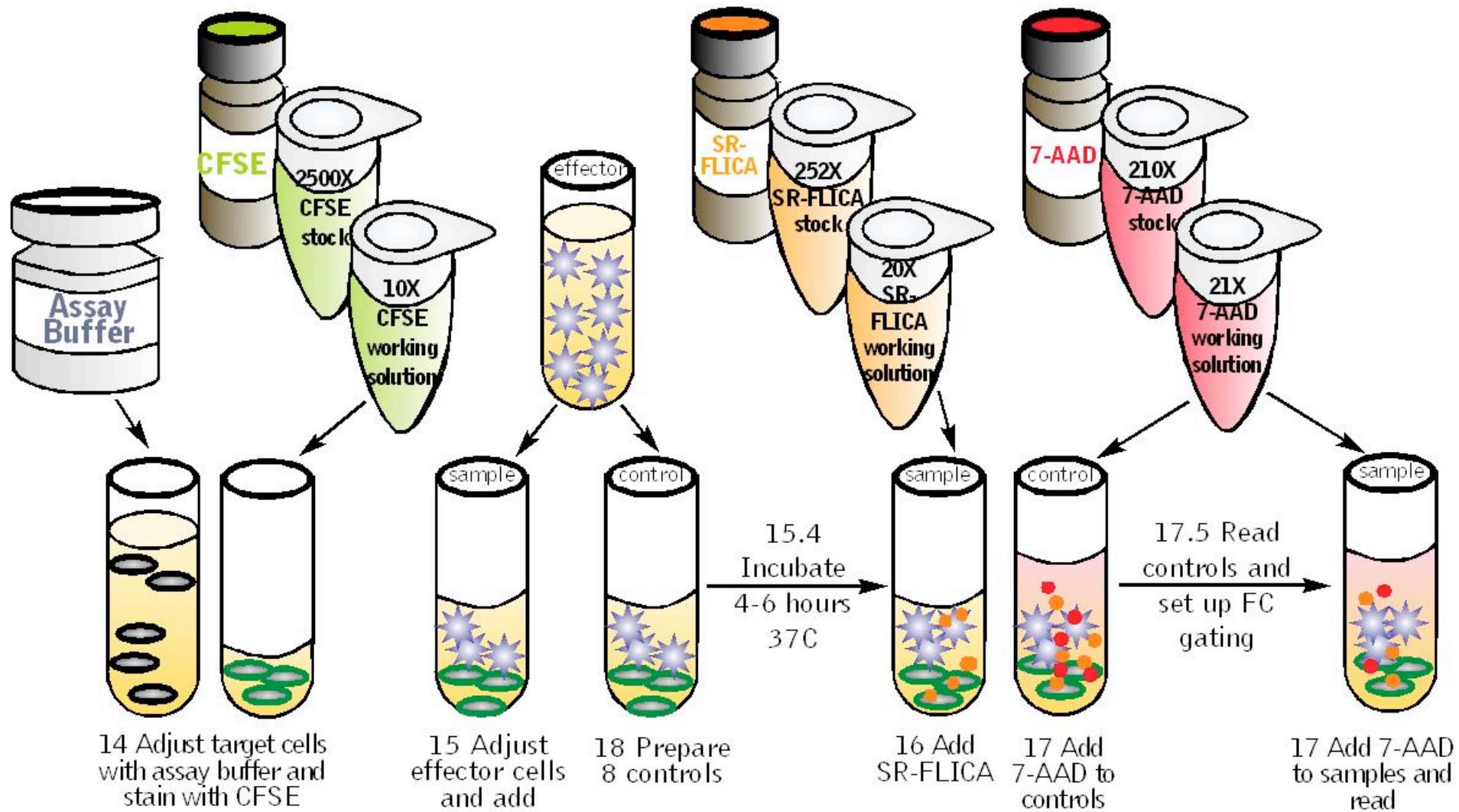
# *Material and methods* in journal Articles

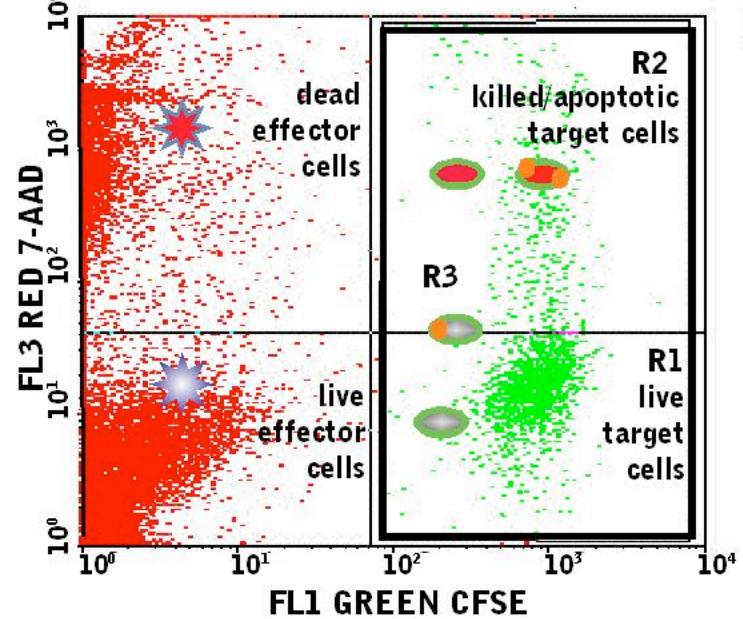
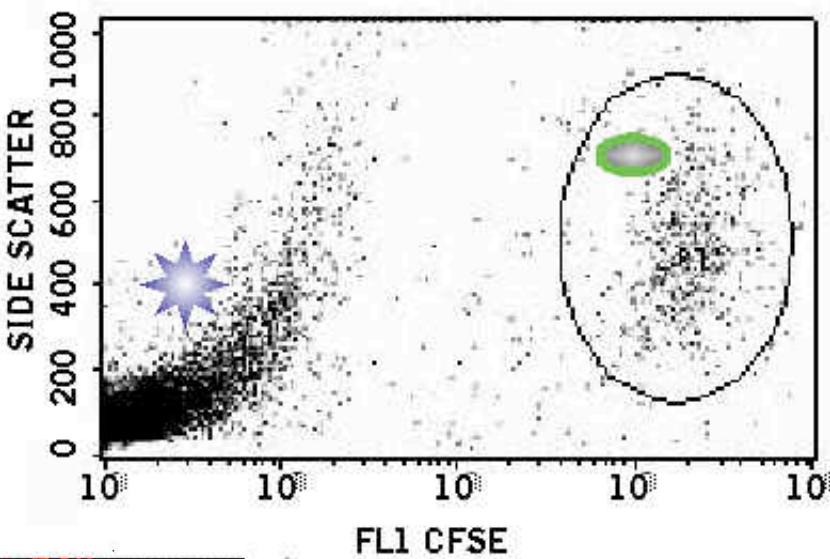
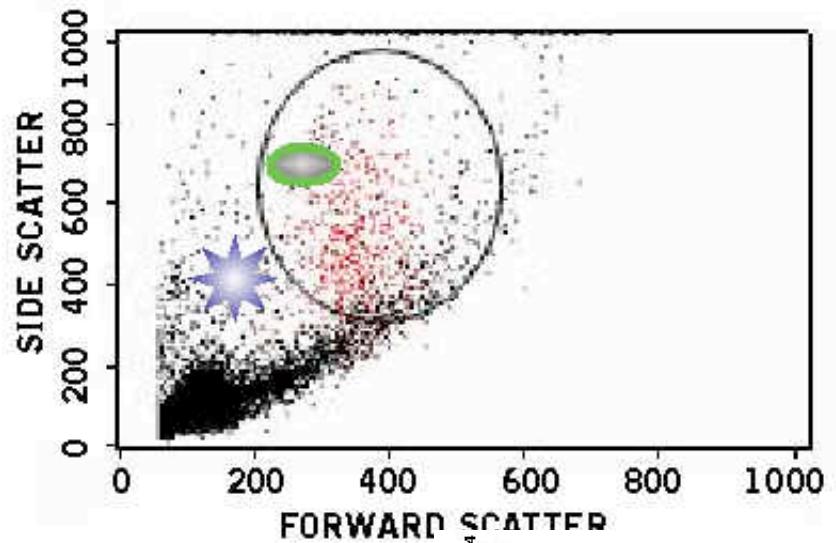
## **Creation of cell line**

N2a were selected to be tTA-positive by transfection with PCMV-tTA-IRES-neo and selection with 800 µg/ml G418. PolyQ cell lines were created by cotransfected Hela and N2a with tetO-htt (25Q, 65Q, or 103Q) exon1-mCFP and PTk-hgro (CLONTECH Laboratories, Inc.) and then selected with hygromycin using 800 and 200 µg/ml, respectively. 100 ng/ml dox was also maintained in the culture media during selection to maintain suppression of transgene expression. HeLa cells were maintained in DME with 10% FCS, whereas N2a cells were maintained in 50% DME/50% Optimem in 10% FCS.

*Autophagy-mediated clearance of huntingtin aggregates triggered by the insulin-signaling pathway*  
J. Cell Biol., Vol. 172, No. 5. (27 February 2006), pp. 719-731. by Ai Yamamoto, Laura M. Cremona, James E. Rothman







# Selected topics from OBI

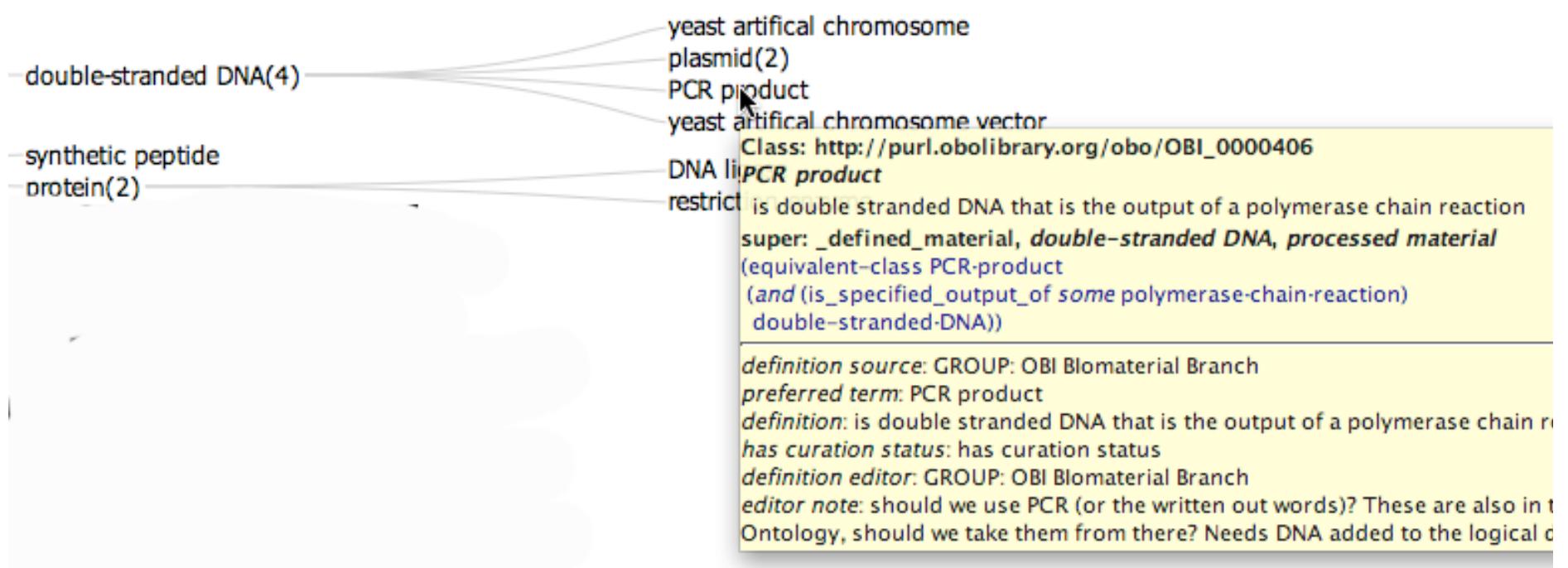
# Organisms

- Our take: A species is a universal – at a given time it has an extension which is a set of organisms.
- "An organism is material entity that is an individual living system, such as animal, plant, bacteria or virus, that is capable of replicating or reproducing, growth and maintenance in the right environment. An organism may be unicellular or made up, like humans, of many billions of cells divided into specialized tissues and organs."
- We import species terms from a subset of NCBI Taxonomy – *Viruses, Bacteria, Archaea, Eukaryota*, and skip *viroids, unclassified, and other*.
- There is an ongoing discussion of the ontology of species, with some arguing that species are individuals, with parts being the individual organisms.

# Connecting Natural to Artifactual

- Use GO for the natural term
- Relate the artifactual to the natural, e.g. via derives\_from, or as specified output of a protocol

# Example: PCR Product



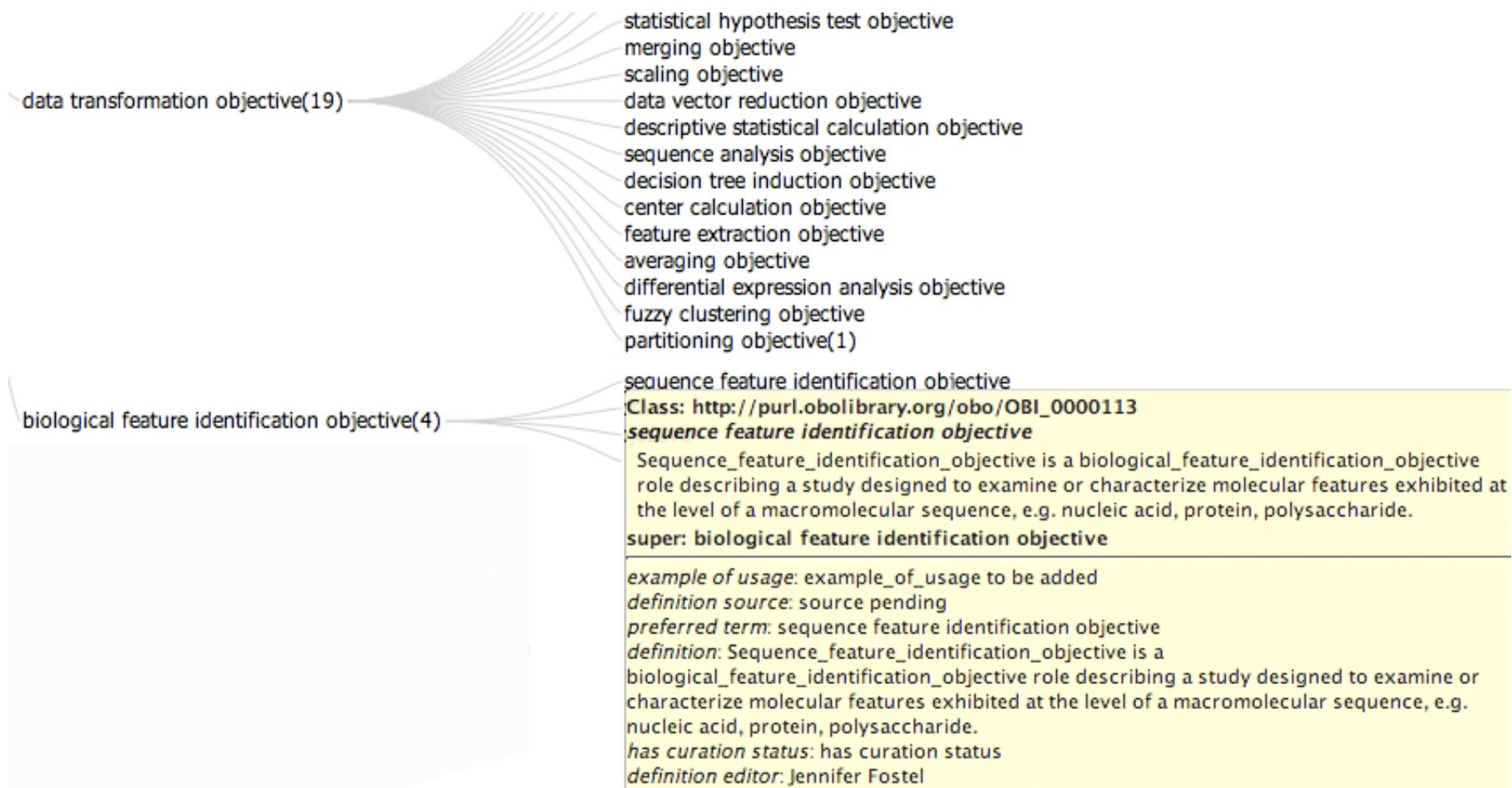
# Plan Specification

A plan specification is an information entity about a realizable entity. The specification includes parts such as objective specification, action specifications and conditional specifications. When concretized it is realized in a process in which the bearer tries to achieve the objectives, in part by taking the actions specified.

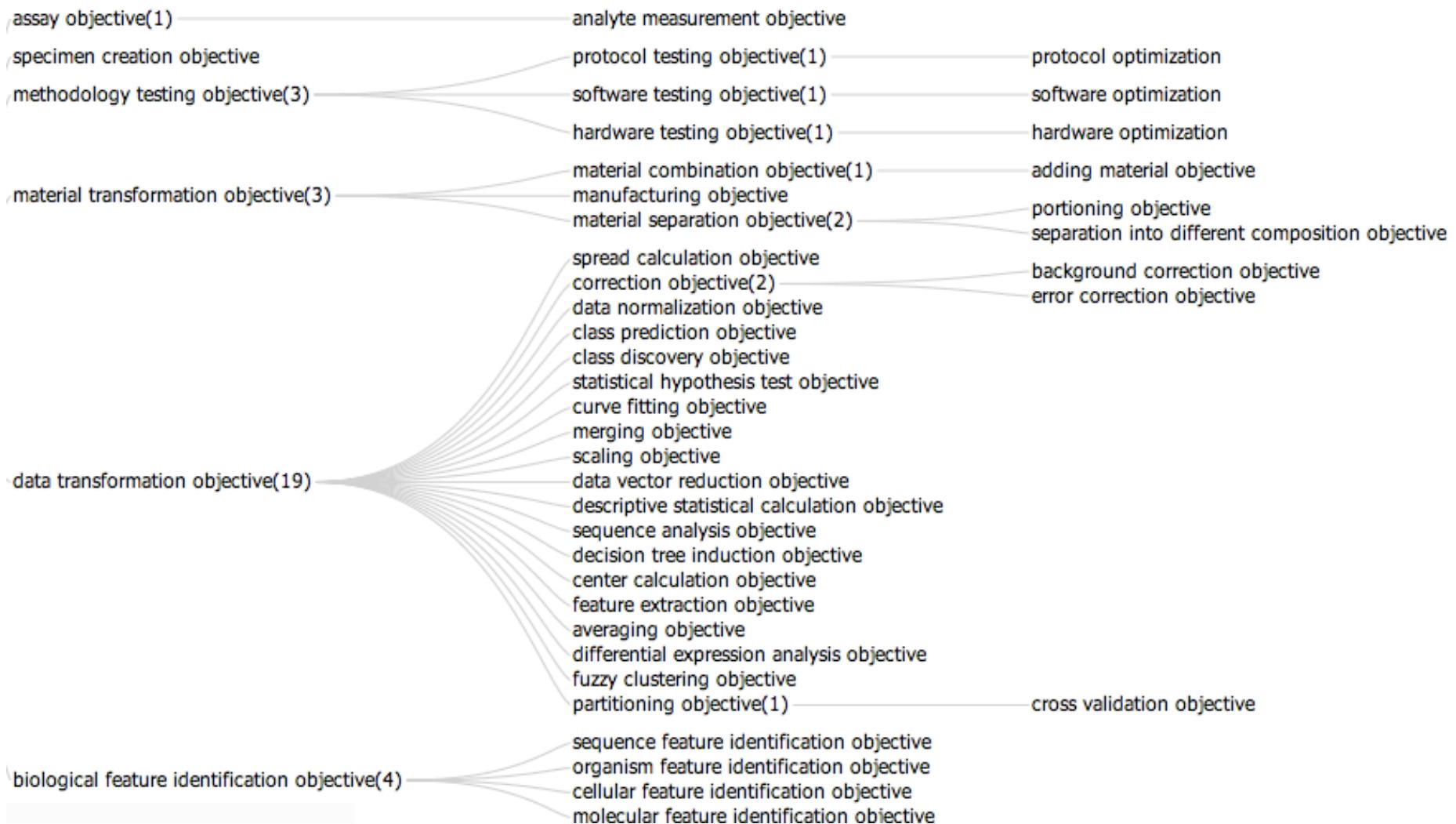
# Objective Specification

An objective specification is an information entity about a realizable entity that describes a intended process endpoint. When part of a plan specification the concretization is realized in a planned process in which the bearer tries to effect the world so that the process endpoint is achieved.

# Objective Specification



# Hierarchy below objective specification



# Achieves planned objective

This relation links a process that realizes a plan which is a concretization of a plan specification with an objective specification that is part of the plan specification. The criteria specified in the objective specification are met at the end of the process.

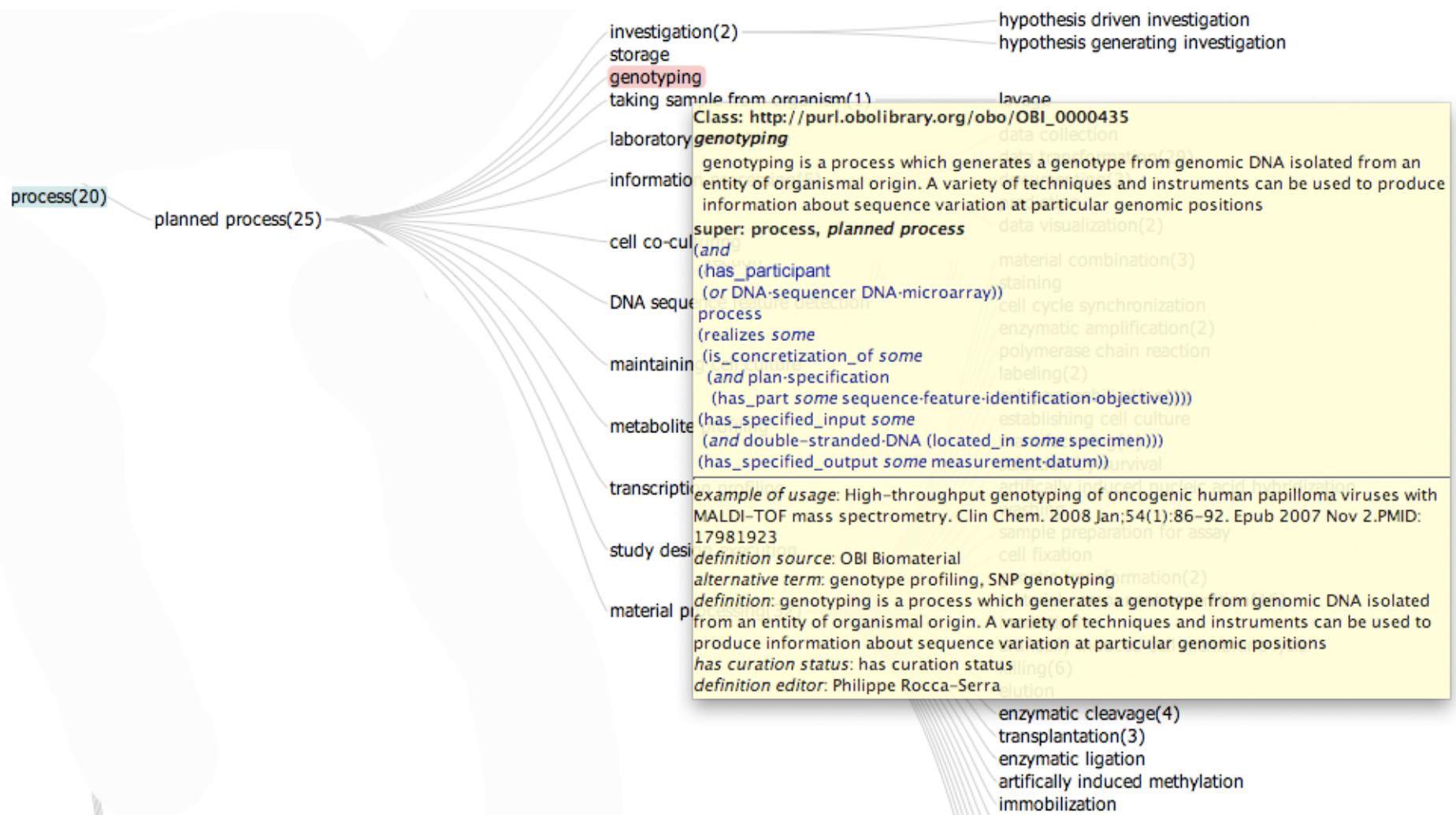
# Planned Processes

- A process that realizes a plan which is the concretization of a plan specification.
- Definition boils down to *intention*
- Previous attempts to define the class of processes that happen during investigations by their participants failed

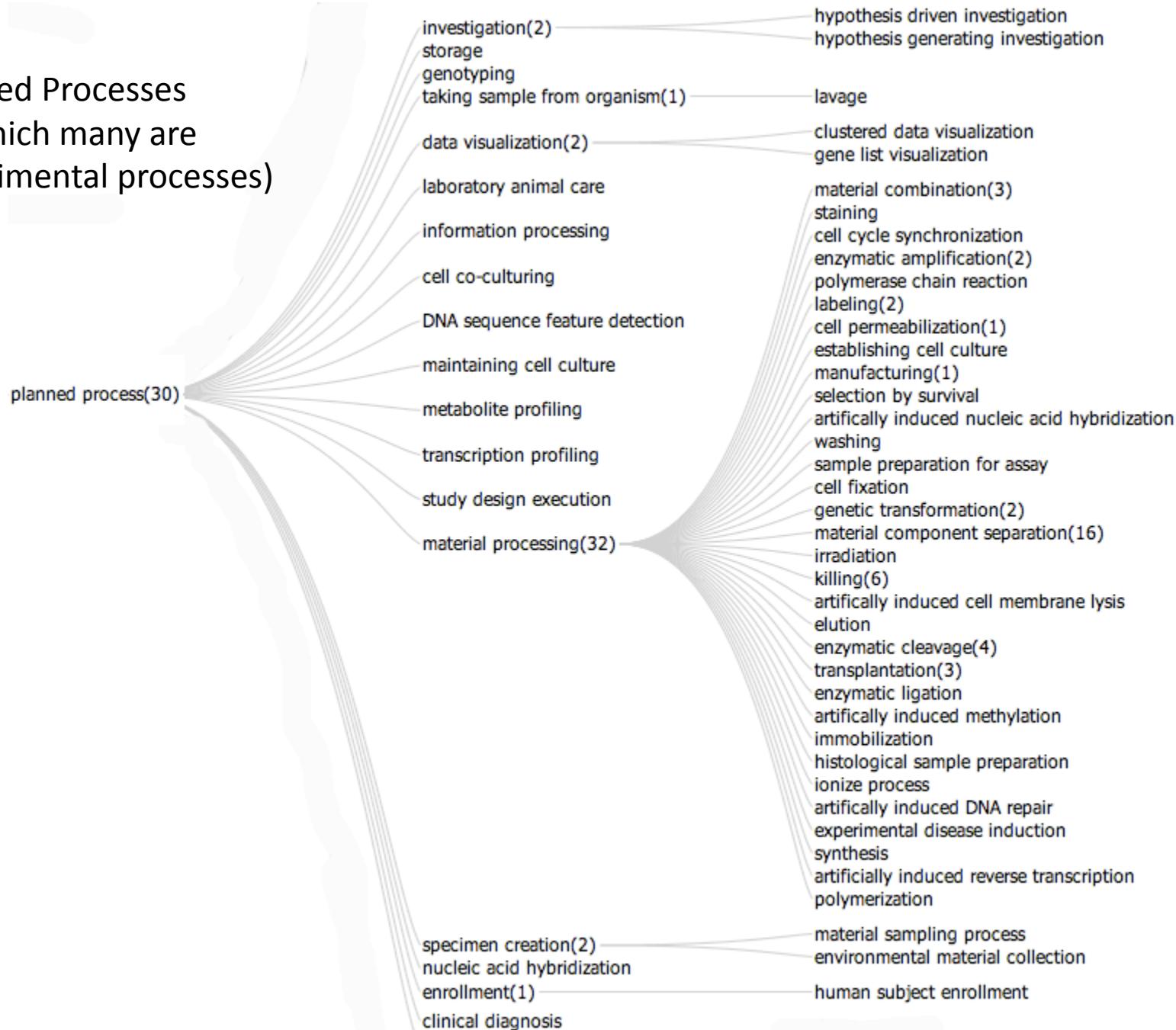
# Some upper level planned processes

<i>Specified input</i>	<i>Process type</i>	<i>Specified output</i>
Material entity and (has role some evaluant role)	<b>assay</b>	Information content entity about the evaluant
Material entity	<b>material processing</b>	Processed material
Data item or data set	<b>data transformation</b>	Data item or data set

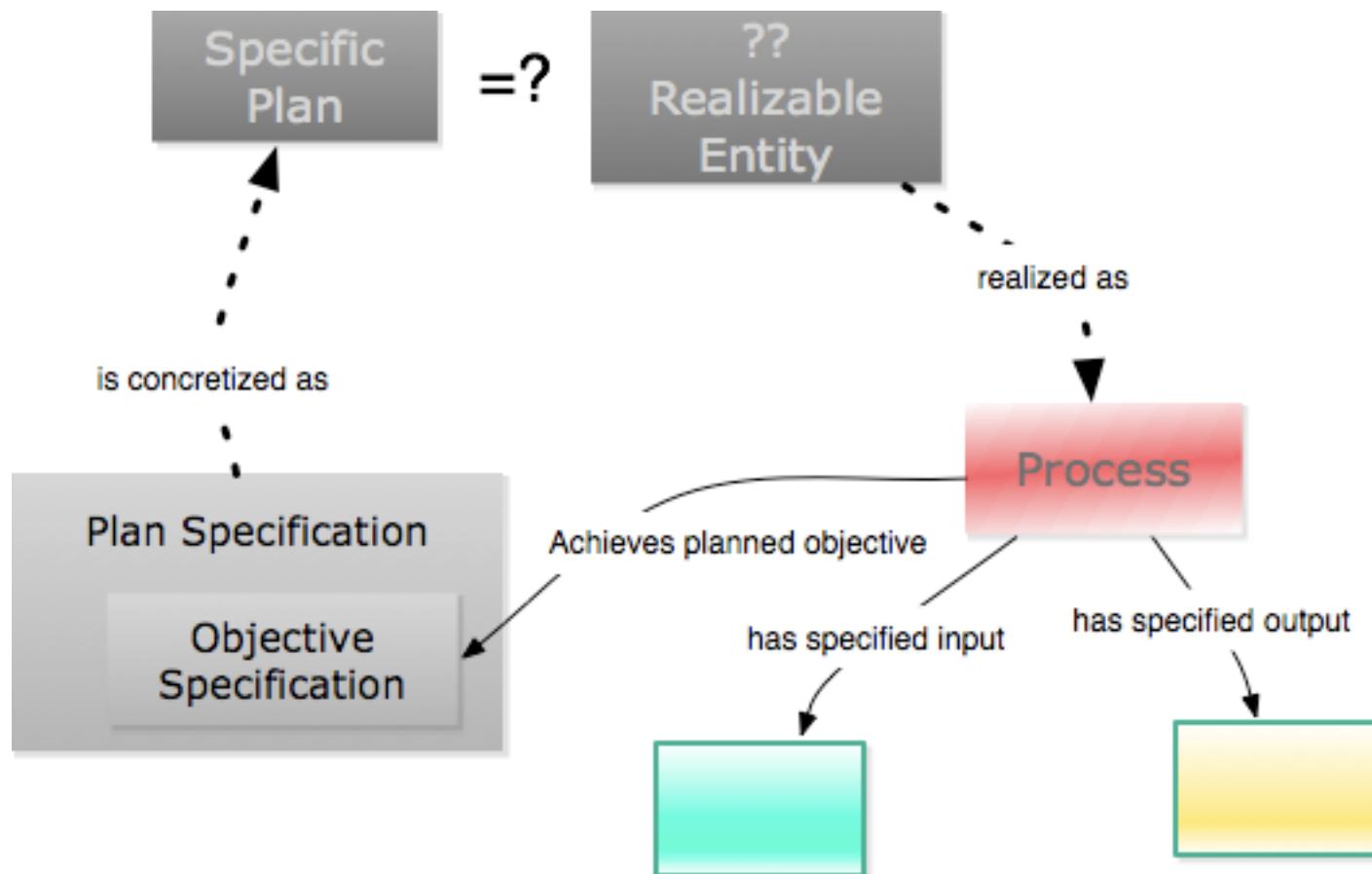
# Process: Genotyping



## Planned Processes (of which many are experimental processes)



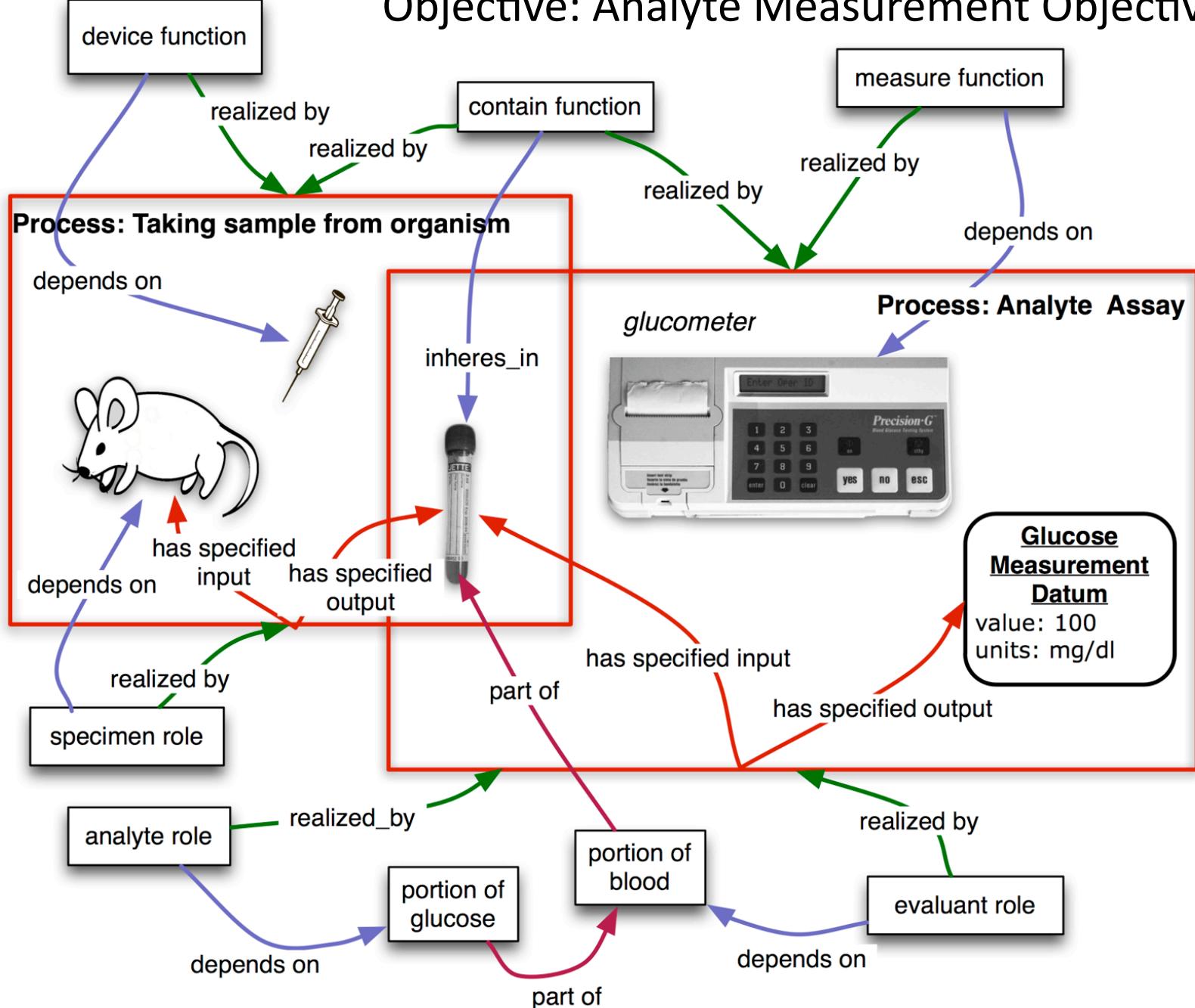
# Connecting plans, objectives, processes



# Studying diabetes. What entities and Relations?



# Objective: Analyte Measurement Objective

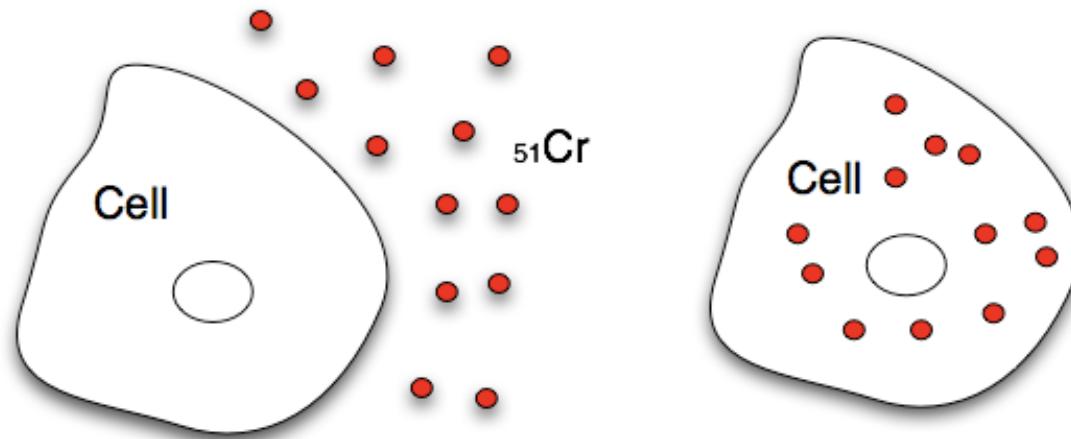


# Proxy, e.g.: Chromium Release assay

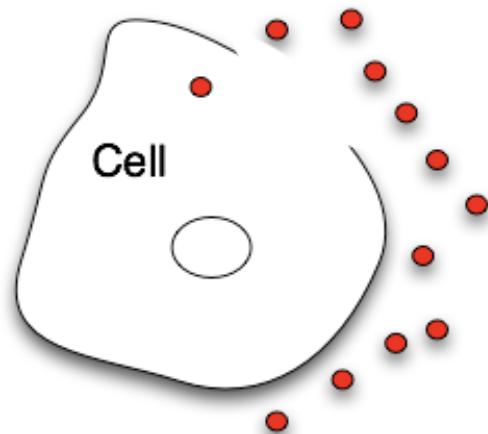
Autologous EBV-transformed B-LCL were used as target cells for the influenza virus-specific CTL assays. Equal volumes of target ( $5 \times 10^4$  cells/ml) and effector cells were added to triplicate wells of 96-well tissue culture plates, and 1:2 serial dilutions of effectors were made, producing effector-to-target (E:T) ratios of 100:1, 50:1, 25:1, and 12.5:1 (9). Spontaneous release of **51Cr** was assessed by the incubation of target cells in medium alone, and maximum release of **51Cr** was determined by the incubation of target cells in 0.1% **Triton X-100** (Sigma-Aldrich). After a 4-h incubation of the effector cells with the target cells at 37°C in 6% CO<sub>2</sub>, supernatants were collected following brief centrifugation. The supernatant was collected into filters of supernatant collection system harvesting frames (Molecular Devices Skatron, Sunnyvale, Calif.) and transferred to polystyrene tubes to be counted with the LKB 1272 Clinigamma counter (Wallac). Percent specific was determined with the following equation: **(experimental  $51\text{Cr}$  release - spontaneous  $51\text{Cr}$  release)/(maximum  $51\text{Cr}$  release - spontaneous  $51\text{Cr}$  release)  $\times 100$** .

# Chromium Release Assay

Radioactive Chromium is Absorbed into Cells

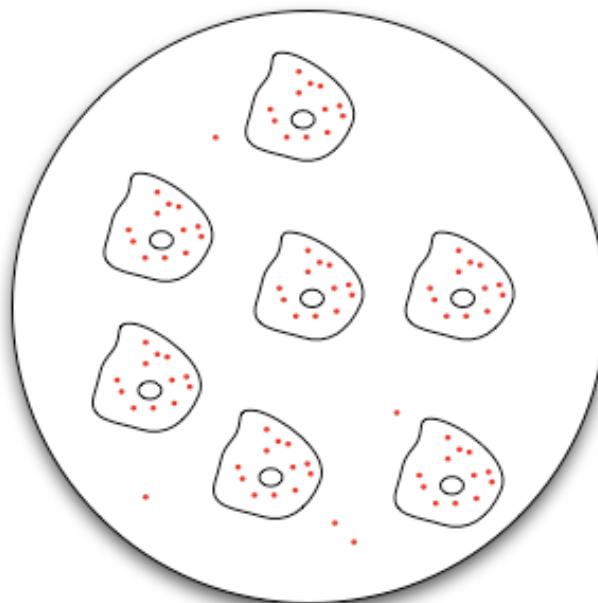


Cells die (or detergent is added to their medium) and Chromium is released

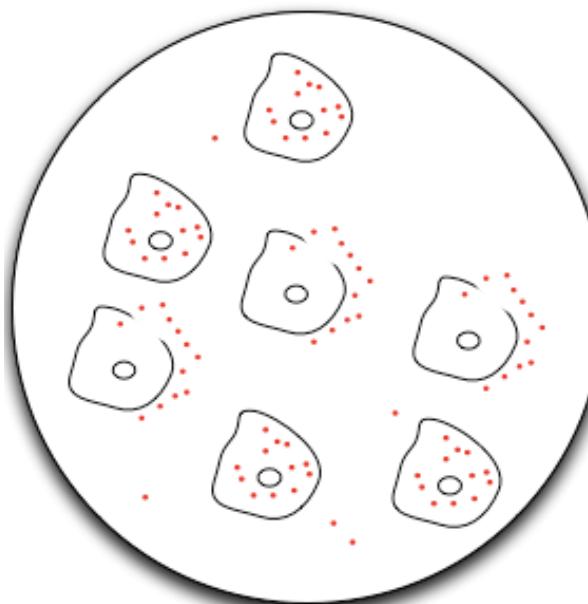


# Cells culture after various treatments

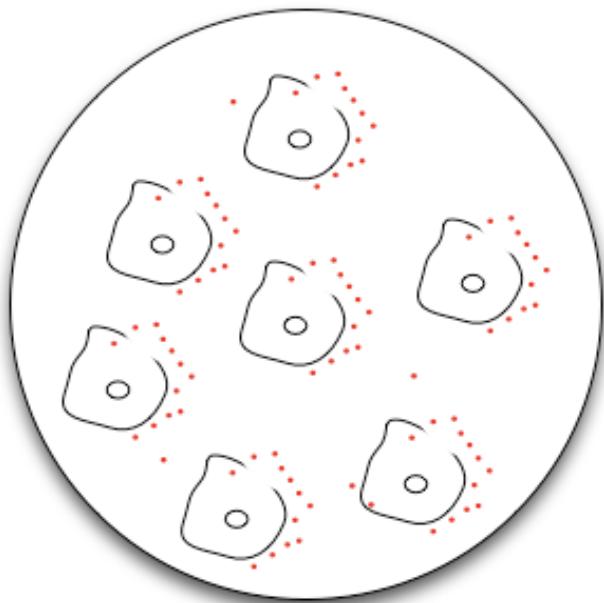
Left alone



Something toxic added

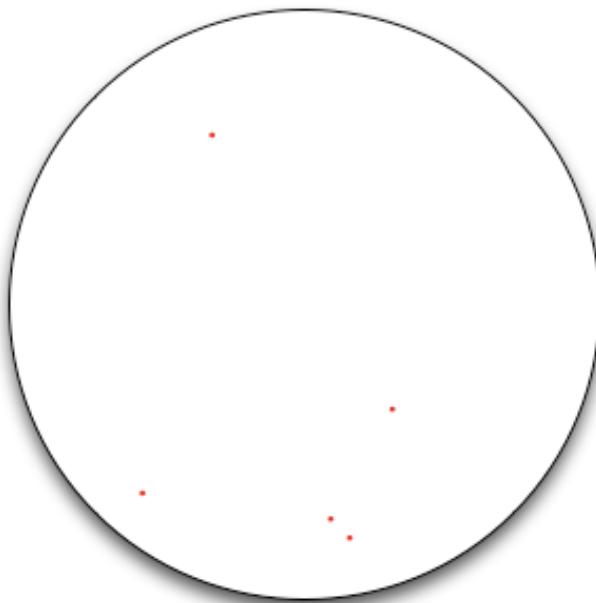


Detergent added

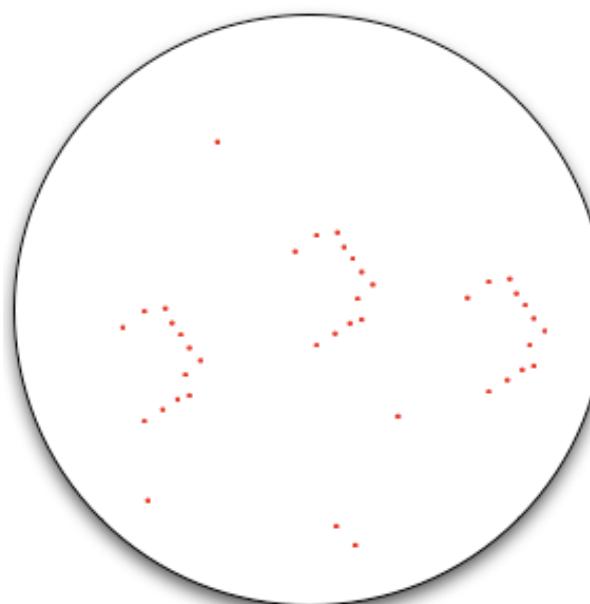


# Cells culture after various treatments

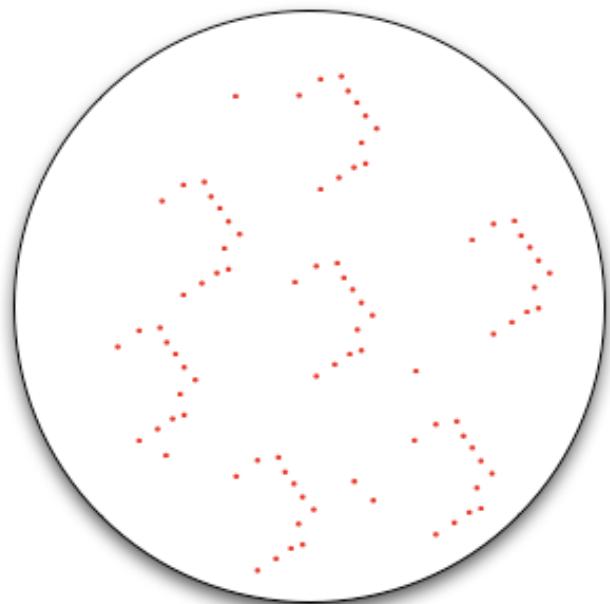
Left alone



Something toxic added



Detergent added



Dead cells: Hard to measure. Radiation: Easy to measure.  
Radiation is *Proxy for* dead cells in this experiment.

# Proxy for

- A relation between continuant instances c1 and c2 where within a protocol application, measurement of c1 is related to what would be the measurement of c2.
- *position on a gel is\_proxy\_for mass and charge of molecule in an western blot. Fluorescent intensity is\_proxy\_for amount of protein labeled with GFP.*
- *A260/A280 (of a DNA sample) is\_proxy\_for DNA-purity.*
- *NMR Sample scan is a proxy for sample quality.*

# Material Entity

material\_entity(24)

organism(4)  
Molecular entity(10)  
physical document  
cellular\_component(1)  
processed material(28)  
protein complex(1)  
anatomical entity(3)  
specimen(5)  
cell(4)  
flat\_object\_part  
soil(1)  
whole organism preparation(1)  
scattered molecular aggregate  
chemical solution(2)  
material information bearer  
enzyme(4)  
organ section  
object\_aggregate  
object  
fragment derived from protein  
organization(1)  
collection (of entities of organismal origin)(4)  
chimera  
instrument(5)

# Some devices

device(88)

- acquisition computer
- liquid chromatography mass spectrometry platform
- charge plate
- voltage amplifier(2)
- platform
- DNA sequencer
- flow cytometer analyzer(25)
- needle
- NMR console(1)
- photomultiplier tube
- syringe
- gel tank
- sonicator
- flow cell(1)
- oligonucleotide synthesizer
- microarray(2)
- mass analyzer
- hybridization station
- multispectral imaging flow cytometer(1)
- light source(2)
- ion detector
- microarray platform
- microarray wash station
- autosampler(1)
- polystyrene tube

# Qualities

- Imported from the Phenotypic Quality Ontology (PATO)
- Monadic qualities of continuants: frozen, solid...
- Relational qualities of continuants: anterior to, molecular concentration...

# Realizable entities

- Disposition: disease...
- Function: device function, heart function...
- Role: analyte role, drug role
- Person Roles: Principal investigator, IRB member
- Challenges: correct categorization disposition/ quality and role/function

# Challenges

- Devices versus instruments
- Specimens versus samples
- Intention-based versus intrinsic definitions
- Tracking origin
  - “Conferred qualities”
    - Names mean that processes have happened
  - Organism of origin
    - Think about proximity – Surgical extraction versus further processing

# Acknowledgements

The OBI Consortium

<http://purl.obolibrary.org/obo/obi>

See the OBI Poster at the ICBO Conference  
Poster Session

# Introduction to the Information Artifact Ontology

# Goals of this section

- Motivate the need for such an ontology
  - What problem should it help solve
  - What should its scope be
- Expose the current state of the IAO
  - Relations
  - Generically dependent continuants
  - Processes
  - Material entities
- Enumerate open issues and future plans

# Scope of IAO

- Aiming to be a mid-level ontology
- Information content entities
- Processes that consume or produce information content entities
- Material bearers of information
- Relations in which one of the relata are information content entities

# Examples of information content entities

- Novels
- Legal documents
- Charts
- Symbol
- Traffic directions
- Recipes
- Computer programs
- XML files
- File formats
- “Referring particulars”
- Ontologies
- Class descriptions
- Sentences
- URIs
- Simulation models
- Ideas
- Questions
- Hypotheses
- Databases
- Licenses
- Poems
- Journal article
- Rejection letters
- Things that can be true or false
- Propositions
- Advertisements
- Specifications
- Serial numbers
- Model numbers
- Formulas

# Examples of information bearers

- Books
- Photographic prints
- CDs
- Hard disks
- Bound journals
- Traffic Signs
- Billboards

# Examples of processes

- Writing
- Documenting
- Recording
- Measuring
- Communicating
- Transcribing
- Encoding
- Client-server processes

# Examples of relations

- is about
- denotes
- mentions
- is measurement of
- encodes
- is rendering of
- is topic of
- makes claim
- cites
- is translation of

# What problems should the IAO Solve

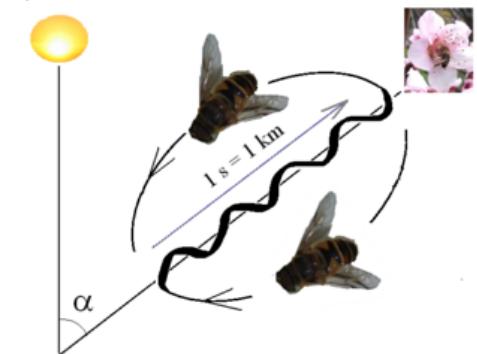
- Information and that which the information is about is often confused.
- Enable the representation of different ways that information relates to the world, sufficient for, at least, representing biomedical investigations.
- Example: Provenance. In a record about a mammalian cell, what speaks about where the cell comes from and what speaks about where the information comes from?
- Consistency with Realist ontology. In what sense does information exist? What is it dependent on. What are the particulars? When does it come into existence and when does it go out of existence?

# Stance of IAO

- 1 Information entities are created by a sentient, intentionally, or by a machine made for that purpose.
- 2 Information is defined by what it is at creation.
- 3 Information content entities are related to other things by being “about” them – they are dependents, not independent entities

# 1 Information entities are intentional

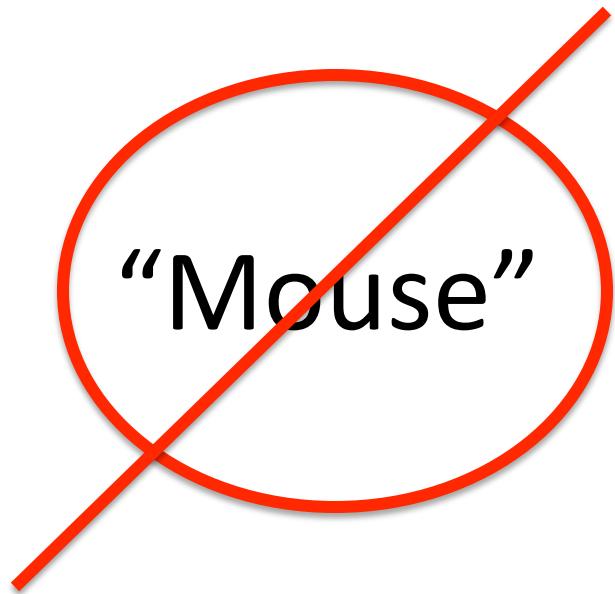
- Information entities are created by a sentient, intentionally, or by a machine made for that purpose.
- Purpose of this scope: I don't want information entities created by wind, or random patterns on beaches
- Consequences of this scope:
  - Bee waggle dances: out?
  - Ant pheromone communication out?
  - Communication is a broader class of processes than information entity transmission



## 2 Information is defined by intention at creation

- Some possible denotations of Werner's "Wednesday".
  - As the answer to the question: What was yesterday?  
Denotes yesterday.
  - As the answer to the question: What day of the week does the OWL working group meet? A scattered temporal region
  - As the answer to the question: Which day of the week were you born? A particular day in the past.
- Often the "question" isn't obvious in the message
- There are almost always multiple *interpretations*. However there is a single utterer. Better strategy to start there.

# 3 Information entities do not stand on their own

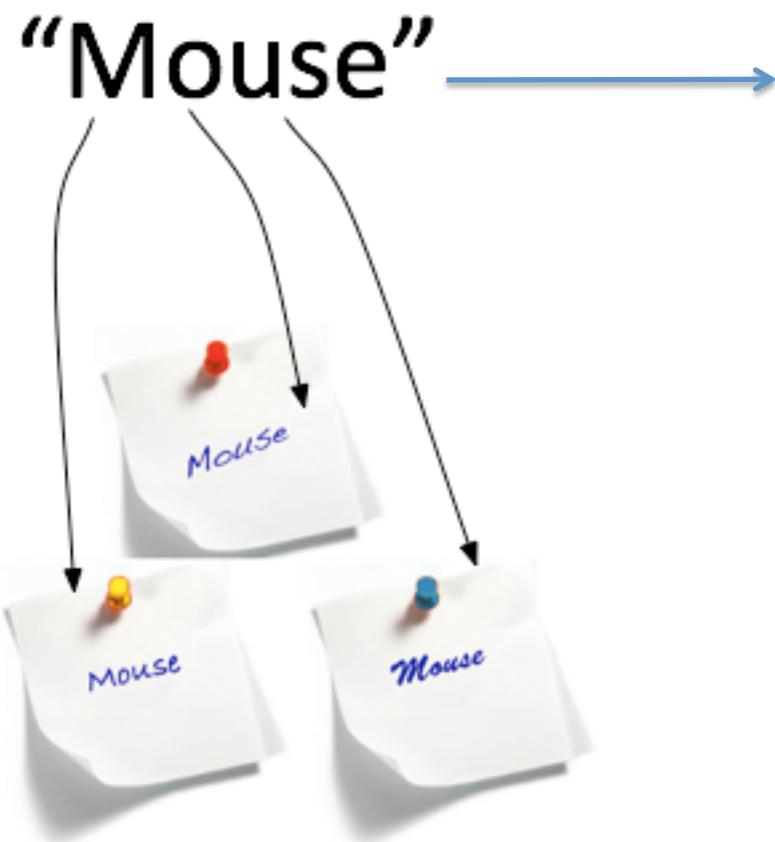


Information entities related to something (are about them)

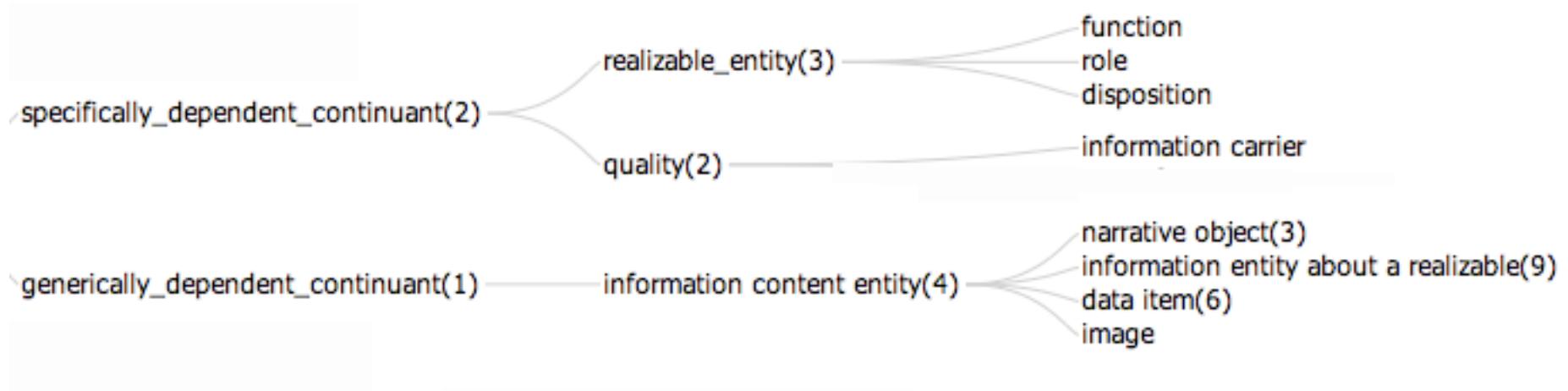
“Mouse” →



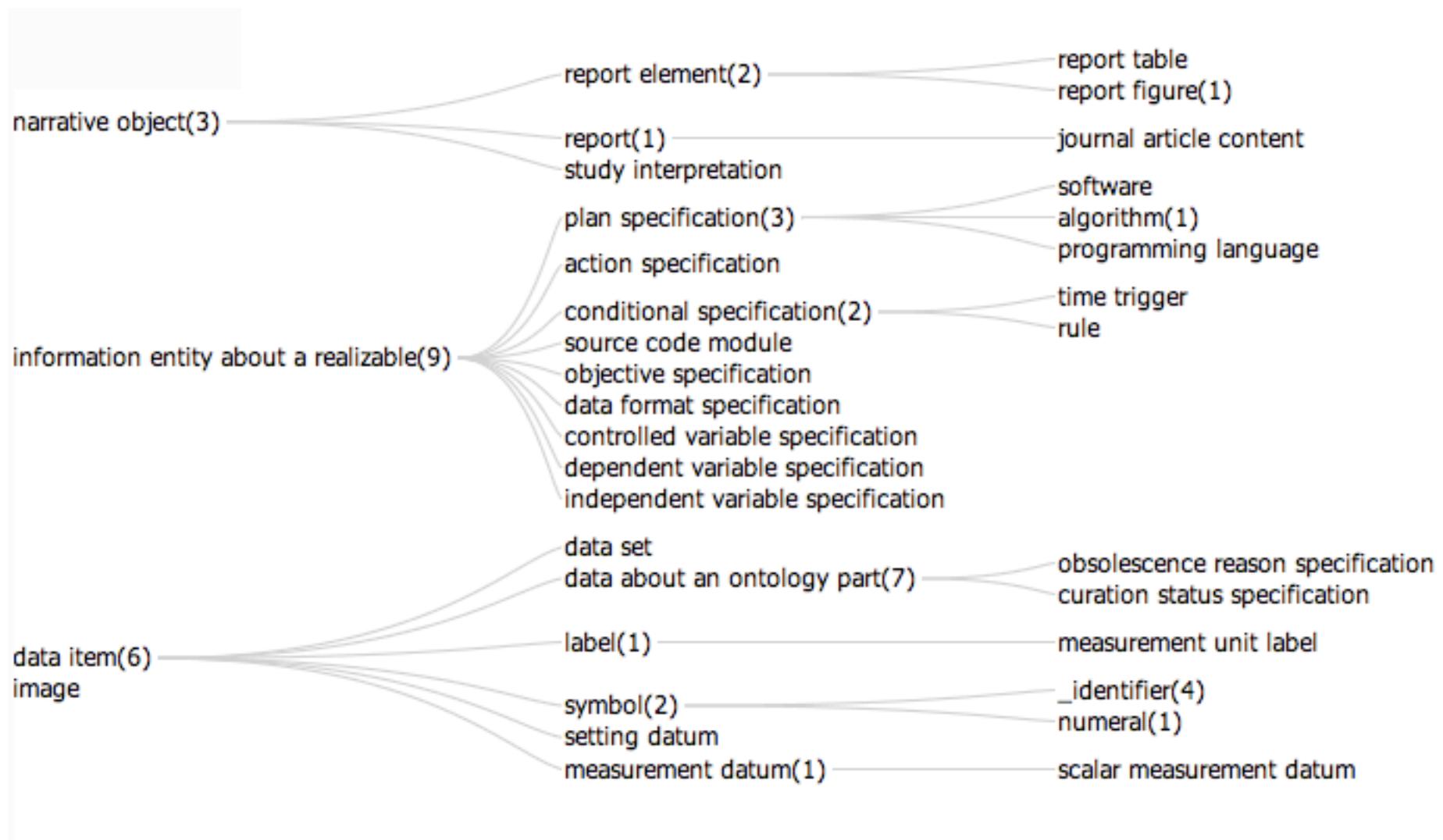
# Information entities are generically dependent



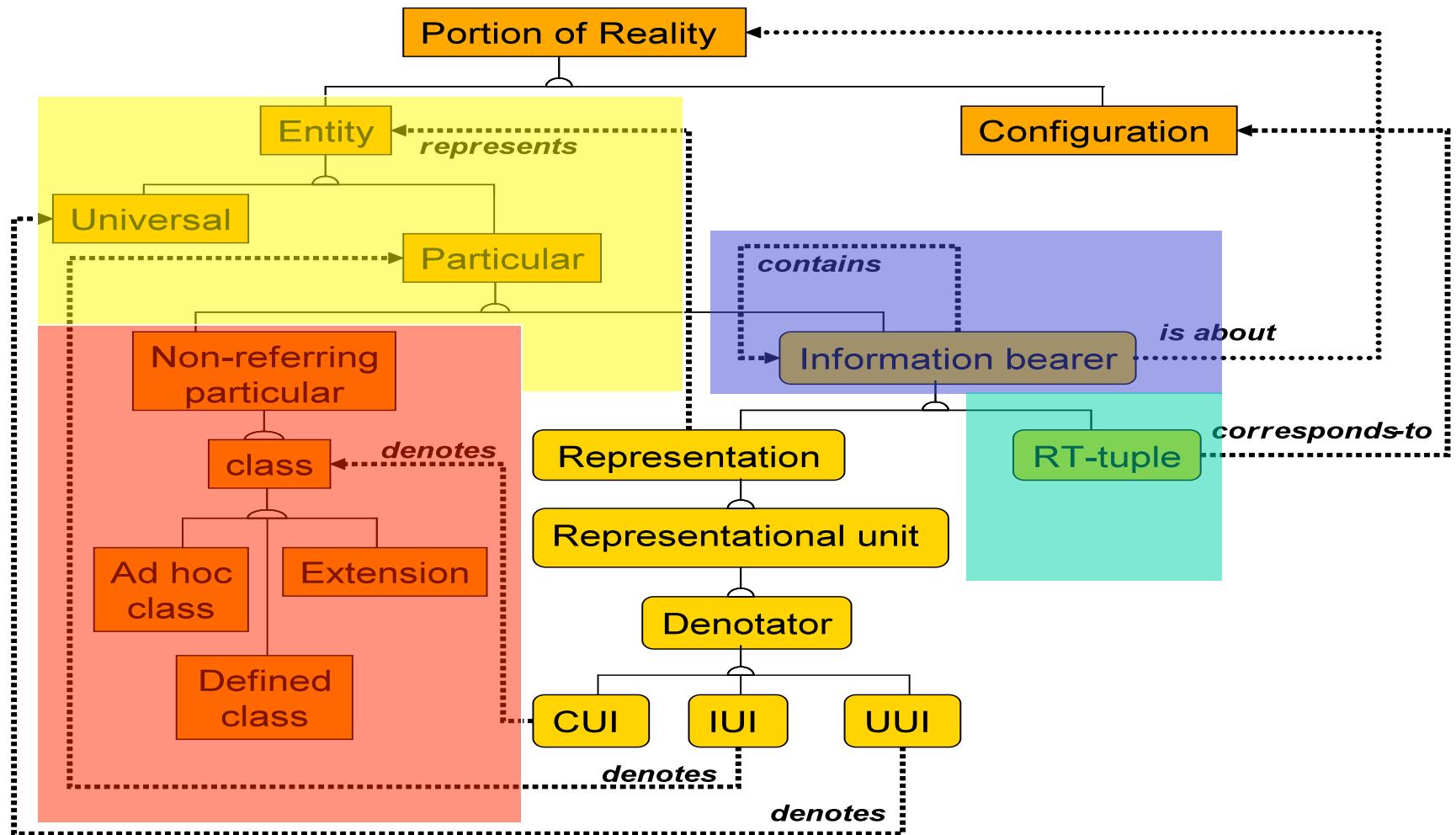
# IAO Overview (dependent continuants)



# IAO Overview, more detail



# Ceuster's 3 levels



# Selected topics in IAO

# \_identifier

- There are a bunch of short strings that are sometimes called “identifiers”.
- What is common about them?
- What is different?

# Model numbers

**easyCyte 8HT** Technology Analysis Specifications Support

**NEW guava easyCyte 8HT System at a glance...**

The new easyCyte 8HT system provides even more powerful cell analysis capabilities. Dual blue 488 nm and red 635 nm excitation lasers provide eight detection parameters, including six fluorescent colors plus forward and side scatter for size and morphology determination. High throughput analysis is made possible with a robotic sample tray that automatically handles a 96-well microplate and up to 10 sample tubes. Plus, this system is the only one capable of running our new InCyte software.

**Intuitive software** provides real-time data acquisition and analysis, letting you visualize up to eight plots simultaneously, while still accessing operation and data analysis functions – all from the same laptop screen

Microcapillary flow cell requires no sheath fluid and is user-replaceable

Six-color detection made possible by two excitation lasers (red & blue)

Small footprint saves valuable laboratory space  
Width: 20.3 in (51.5 cm)  
Depth: 23.4 in (59 cm)  
Height: 10.0 in (25.4 cm)  
(does not include laptop)

Wash vial offers a high-pressure purge to easily clear obstructions from the flow cell

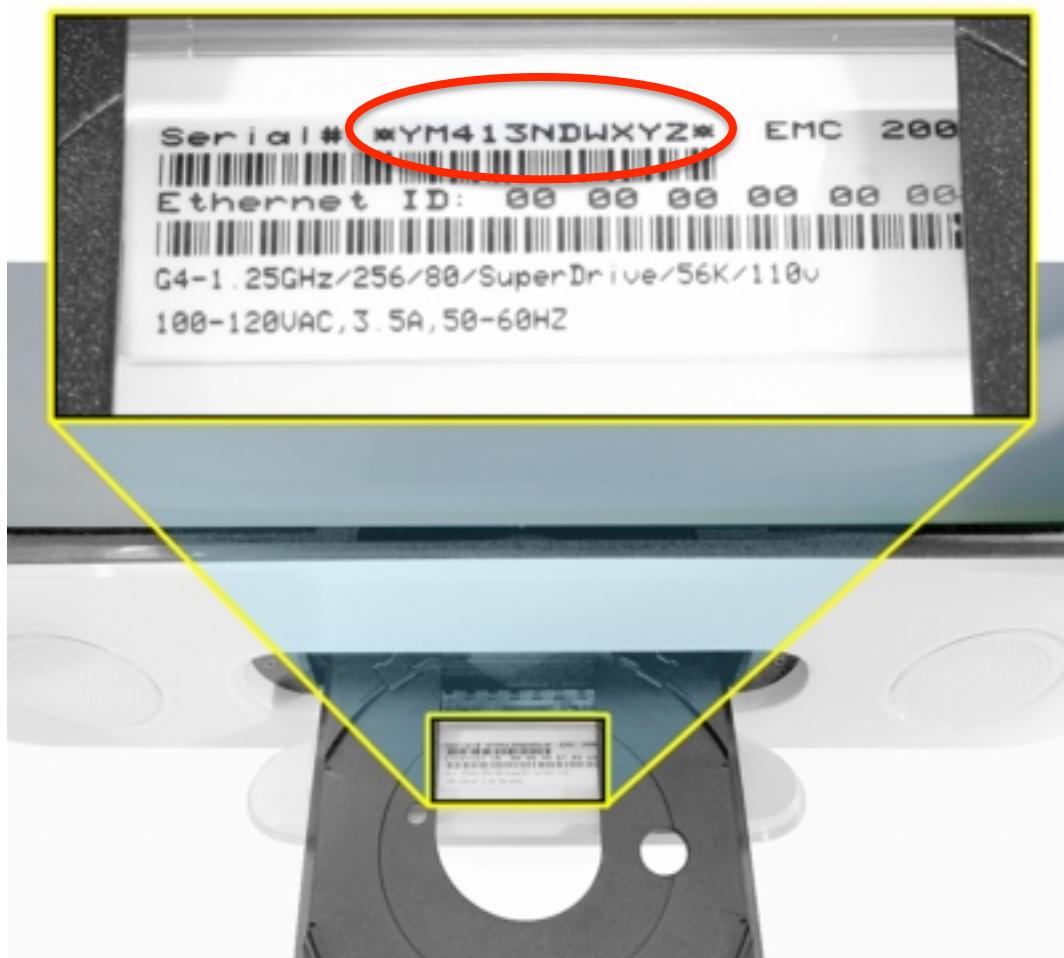
Waste vial collects less than 80 mL of waste in a typical 8-hour workday

Robotic sample tray provides walk-away automation for a 96-well microplate or up to 10 sample tubes

# Model numbers

A model number is an information content entity specifically borne by catalogs, design specifications, advertising materials, inventory systems and similar that is about manufactured objects of the same class. The model number is an alternative term for the class. The manufactured objects may or may not also bear the model number. Model numbers can be encoded in a variety of other information objects, such as bar codes, numerals, or patterns of dots.

# Serial Numbers



# Serial Number

A serial number is an information content entity which is a unique sequence of characters borne by part of manufactured product or its packaging that is assigned to each individual in some class of products, and so can serve as a way to identify an individual product within the class. Serial numbers can be encoded in a variety of other information objects, such as bar codes, numerals, or patterns of dots.

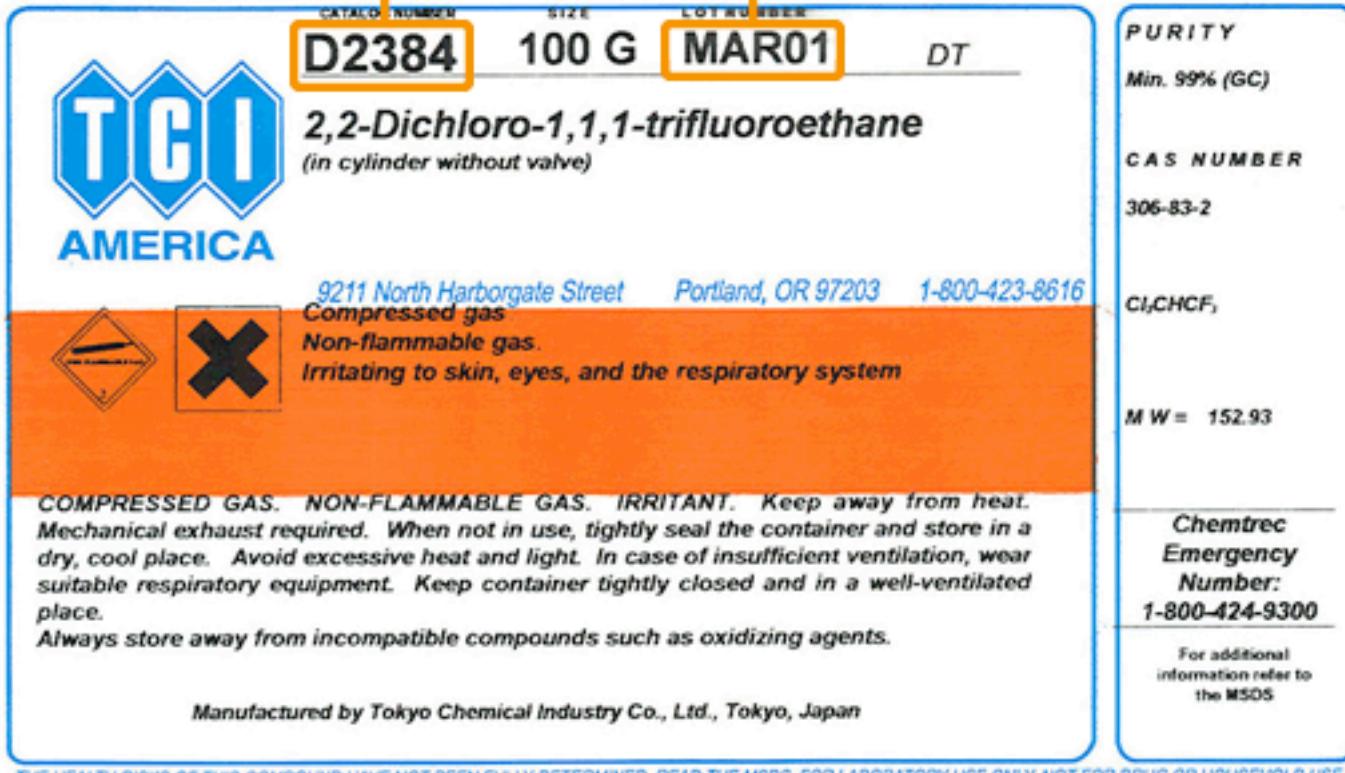
# Lot Number

A lot number is an information content entity which is an identical sequence of character borne by part of manufactured product or its packaging for each instances of a product class in a discrete batch of an item. Lot numbers are usually assigned to each separate production run of an item. Manufacturing as a lot might be due to a variety of reasons, for example, a single process during which many individuals are made from the same portion of source material. Lot numbers can be encoded in a pattern of other information objects, such as bar codes, numerals, or patterns of dots.

# Lot number

Product Number  
(5 digits)

Lot Number  
(4 or 5 digits)



# Lot numbers are important! (in a different way than other “identifiers”)



Search

## Werner Boston Baked Beans

[Email this page](#) [Print this page](#) [Change font size](#)

[FDA Home](#) > [Peanut Butter and other Peanut Containing Products Recall List Home](#)

**Information current as of 12 PM June 12, 2009**

**3916 entries in list**

[Peanut Containing Product Recall \(Salmonella Typhimurium Outbreak\): Main Page](#)

**Note:** This list includes human and pet food subject to recall in the United States since January 2009 related to peanut products distributed by Peanut Corporation of America. This list will be updated as new information is received. The information is current as of the date indicated. Once included, all human and pet food recalls will remain listed. If we learn that any information is not accurate, we will revise the list as soon as possible. When available, this database also includes photos of recalled human and pet food products that have been voluntarily submitted by recalling firms to the FDA to assist the public in identifying those products that are subject to recall.

Category: Candy

Brand Name: Werner

Product Description: Boston Baked Beans

Recalling Firm: Werner Gourmet Meat Snacks Inc.

Size: 3 to 9 ounce sizes

Lot Numbers: Mar. 20, 2008 to Feb. 9, 2010 with lot number 7152 thru. 7334

Stock Number:

Product Code: 7 09972 80026 6

Label: <http://www.fda.gov/oc/po/firmrecalls/photos/werner.jpg>

# Data item

A data item is an information content entity that is intended to be a truthful statement about something (modulo, e.g., measurement precision or other systematic errors) and is constructed/acquired by a method which reliably tends to produce (approximately) truthful statements.

# Data item

- Data sets
  - Measurement datum
  - In Werner's schema: RT-TRIPLES that correspond to a configuration
- 
- Data about an ontology (= metadata)?
  - (note current version of IAO has other children that are probably wrongly placed)

# Material information bearer

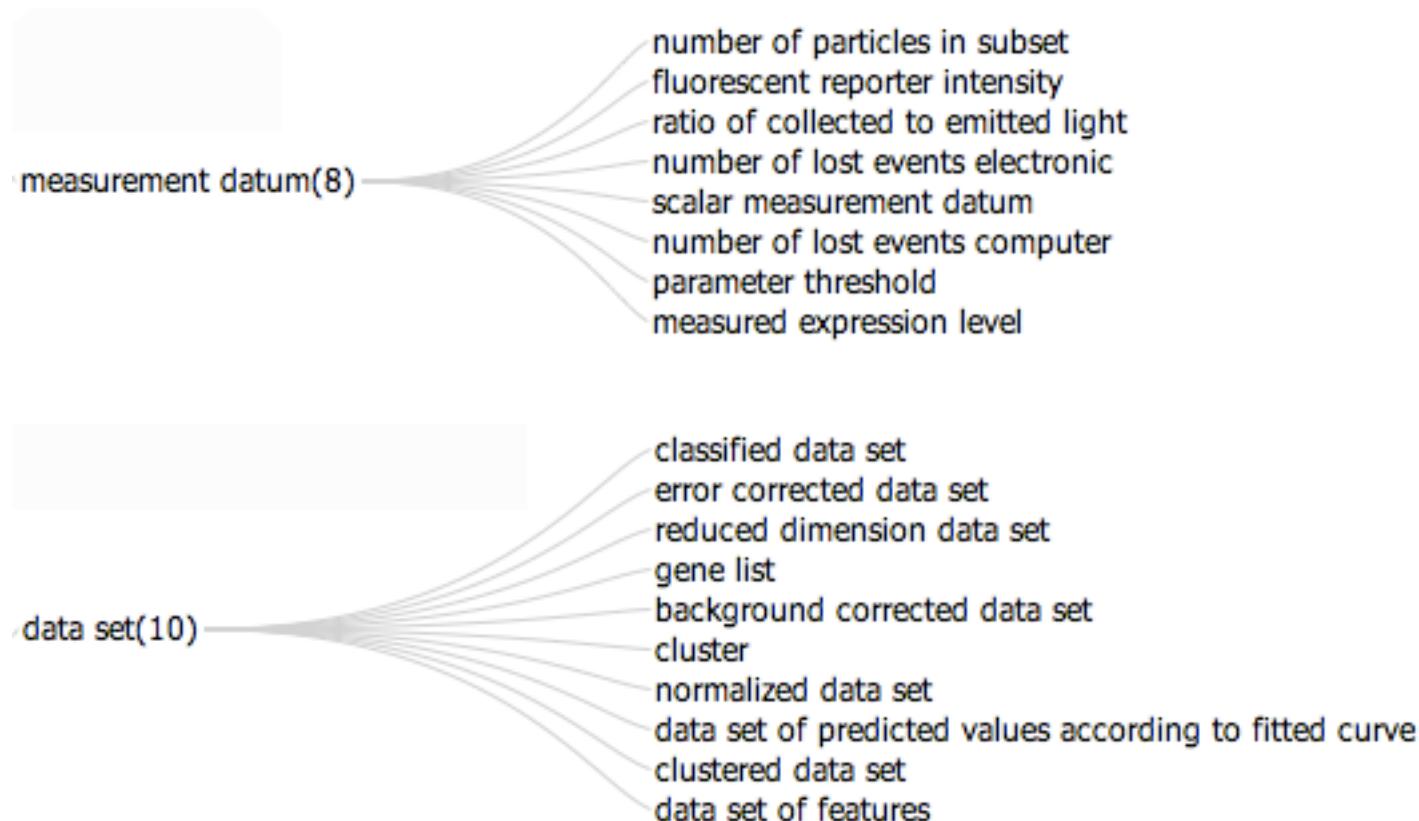
- A defined class of material entities that carry information.
- Material entity and *is\_bearer\_of* some (*is\_concretization\_of* some information content entity)
- Books, Photographic prints, CDs, Hard disks, Bound journals, Traffic Signs, Billboard

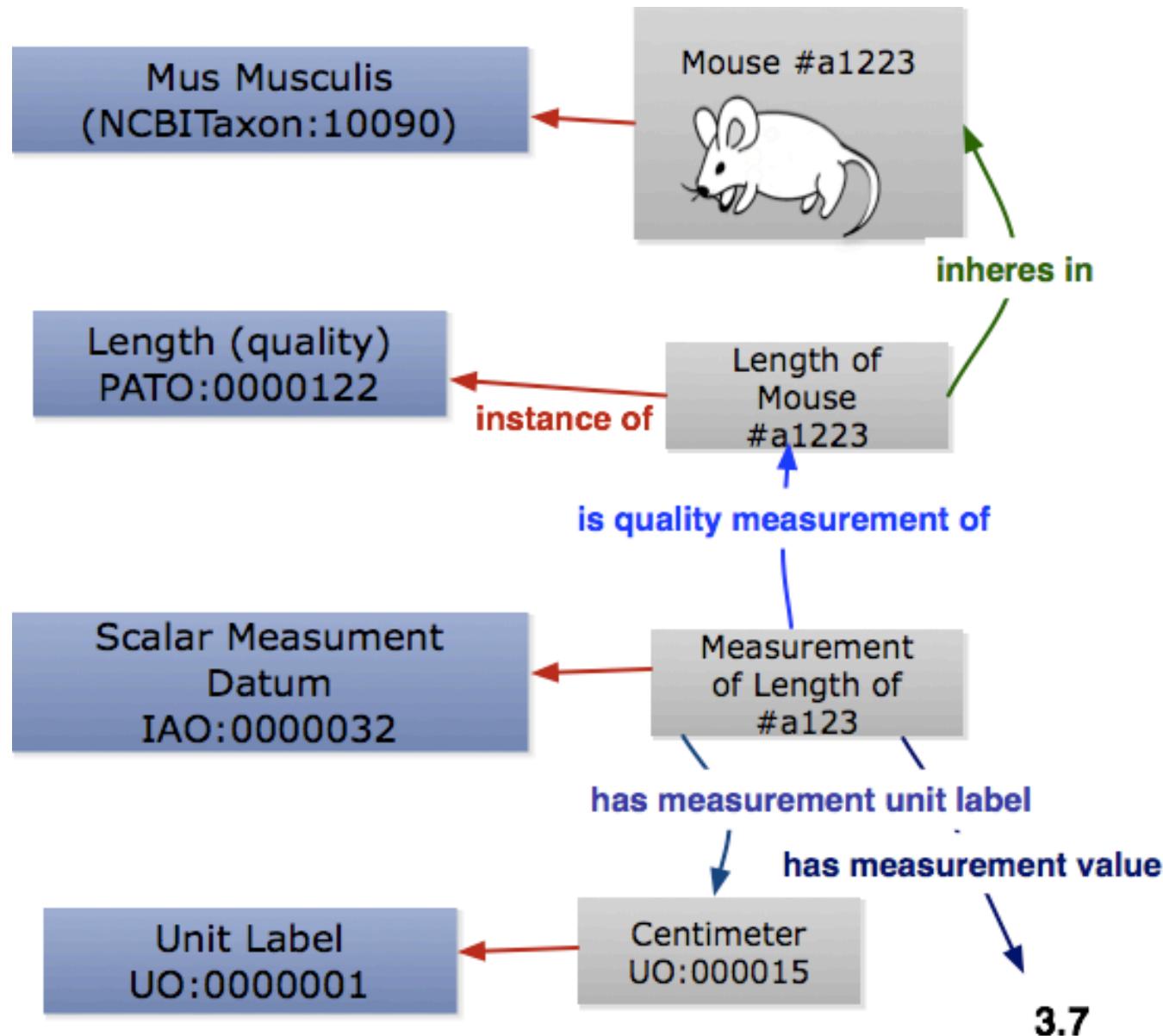


# Measurement Data

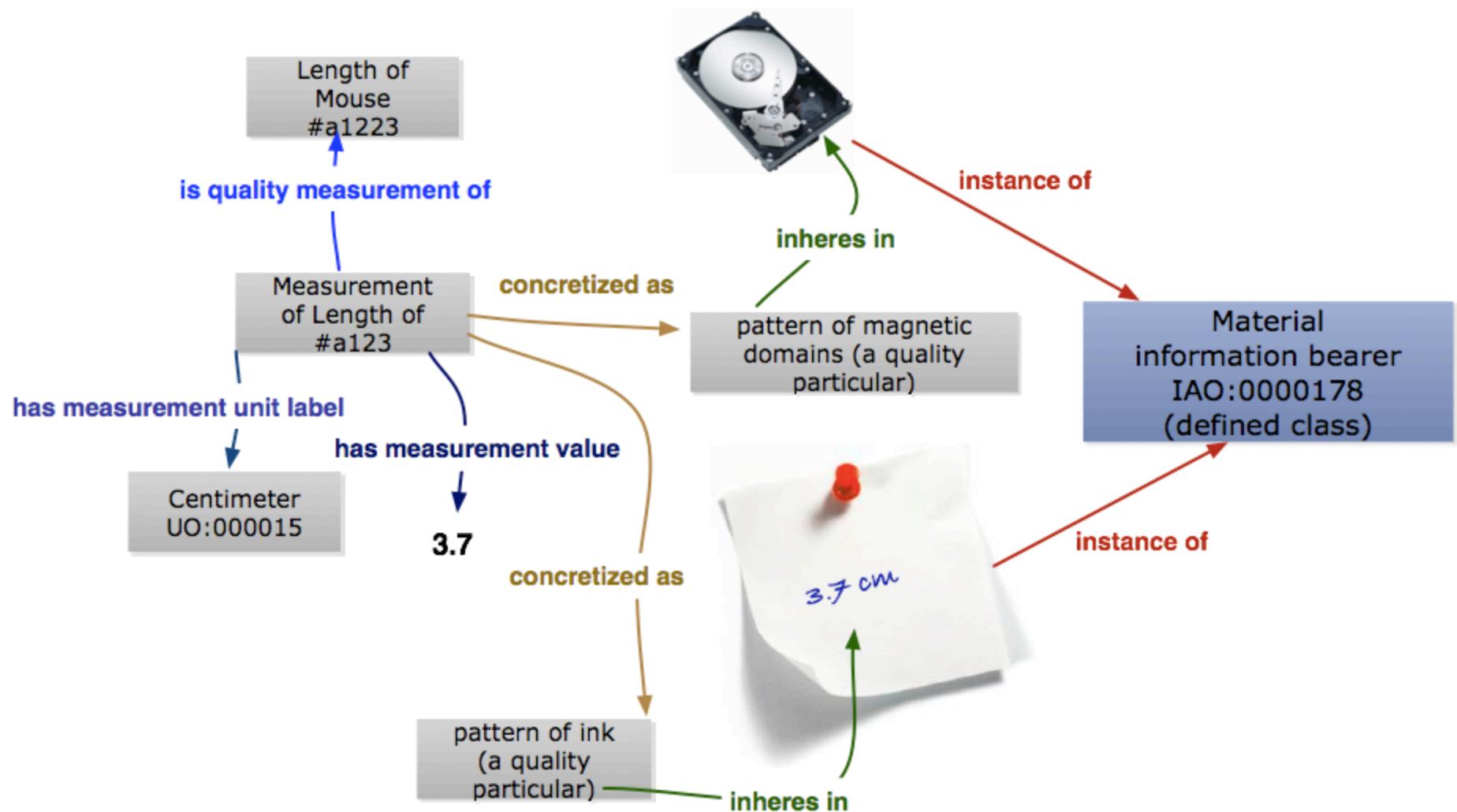
- Measurements are often wrong. Some types of measurement are *always* wrong.
- A measurement of something is not that something
- We don't want to lose the connection, however, between the the measurement and the thing measured.
- There are many types of measurements
- We start with “scalar” measurements.
  - (Oppose with “vector” measurements, such as direction in space)

# First term requests from OBI





# Information entity, concretized



# Narrative objects

- Current definition: A narrative object is an information content entity that is a set of propositions.
- Intention: Capture narratives – the main content of journal articles, books, stories.
- Issue 1: Proposition not currently defined – roughly: a logical sentence that can be true or false.
- Issue 2: do we only text? What about a story told by mime. Does music count? (no) what about an oral report. Regarding definition, saying it is a set of propositions means we lose the idea that wording matters. Maybe adjust saying a narrative object has some relationship to a set of propositions.

# Narrative object - Study interpretation

- A study interpretation is a narrative object interpreting a study result. Examples include discussion of whether a hypothesis is false, whether the study failed to address the hypothesis, and whether the study results have led to new hypotheses.

# Narrative Object - Reports

- A report is a narrative object assembled by an author for the purpose of providing information for the audience. A report is the output of a documenting process and has the objective to be consumed by a specific audience. The topic of the report is on something that has completed. A report is not a single figure. Examples of reports are journal article, patent application, grant progress report, case report (not patient record)

# Narrative Object – Report

- A report element is a narrative object in which information is presented and consumed by a human being, and is part of a report. Examples of report elements are figure (dot plot), table, text portion (may include a movie or audio clip on a web page).

# Diagnosis: “John has Leukemia”

- Suppose there is some question of whether this is true. What are the information entities and what do they denote?
- Background: Based on symptoms of fatigue, elevated white blood cell count, enlarged lymph nodes
- “John” denotes a particular person. What does “Leukemia” denote

# What leukemia is

- **Leukemia is a form of cancer** that begins in the blood-forming cells of the bone marrow—the soft, inner part of the bones.
- The cellular basis of leukemia is that white blood cells change in a way that causes them to proliferate more than they should, leading to an excess accumulation of them.
- Whether or not John has leukemia hinges on whether his white blood cells have changed in this way.

# There are a variety of possibilities for the observed symptom

- Acute lymphocytic leukemia
- Acute myelogenous leukemia (AML)
- Allergy, especially severe allergic reactions
- Chronic lymphocytic leukemia
- Chronic myelogenous leukemia
- Drugs, such as corticosteroids and epinephrine
- Hairy cell leukemia
- Measles
- Myelofibrosis
- Other bacterial infections
- Other viral infections
- Polycythemia vera
- Rheumatoid arthritis
- Smoking
- Stress, such as severe emotional or physical stress
- Tissue damage, such as from burns
- Tuberculosis
- Whooping cough

<http://www.mayoclinic.com/health/high-white-blood-cell-count/MY00161/DSECTION=causes>

# Possibilities for what “Leukemia”

- It only refers to a disease of John, if it is true.
- Other possibilities:
  - The type leukemia disease
  - Previous instances of leukemia the doctor is familiar with
  - The *idea* of leukemia happening to John
- What to represent – Two possibilities
  - Encapsulate these possibilities with a type “Diagnosis”, an information artifact, and document it as having these possibilities. Relate the entity to John and the other entities. Allow it to give rise to a *different* information artifact if it turns out to be true, in which “leukemia” *does* refer to John’s disease.
  - Don’t let doctors say things that don’t necessary refer (Werner’s strategy)
  - Accept fallibilism. Record it as referring to John’s disease and erass it if it turns out to be false.

# Study Design

## (information about a realizable)

- A study design is a plan specification comprised of protocols that are executed as part of an investigation. The objective of the study design is to produce (information about the study subject). The study design specifies the purpose of the different data produced in the execution of the study design.

# Study variables (information about a realizable)

- independent variable specification is an information entity about a realizable which is part of a study design. Independent variables are entities whose values are controlled or which are selected to determine its relationship to an observed phenomenon (the dependent variable). In such an experiment, an attempt is made to find evidence that the values of the independent variable determine the values of the dependent variable (that which is being measured). The independent variable can be changed as required, and its values do not represent a problem requiring explanation in an analysis, but are taken simply as given.

# Study variables (information about a realizable)

- Dependent variable specification is part of a study design. The dependent variable is the event studied and expected to change when the independent variable varies.

# Models

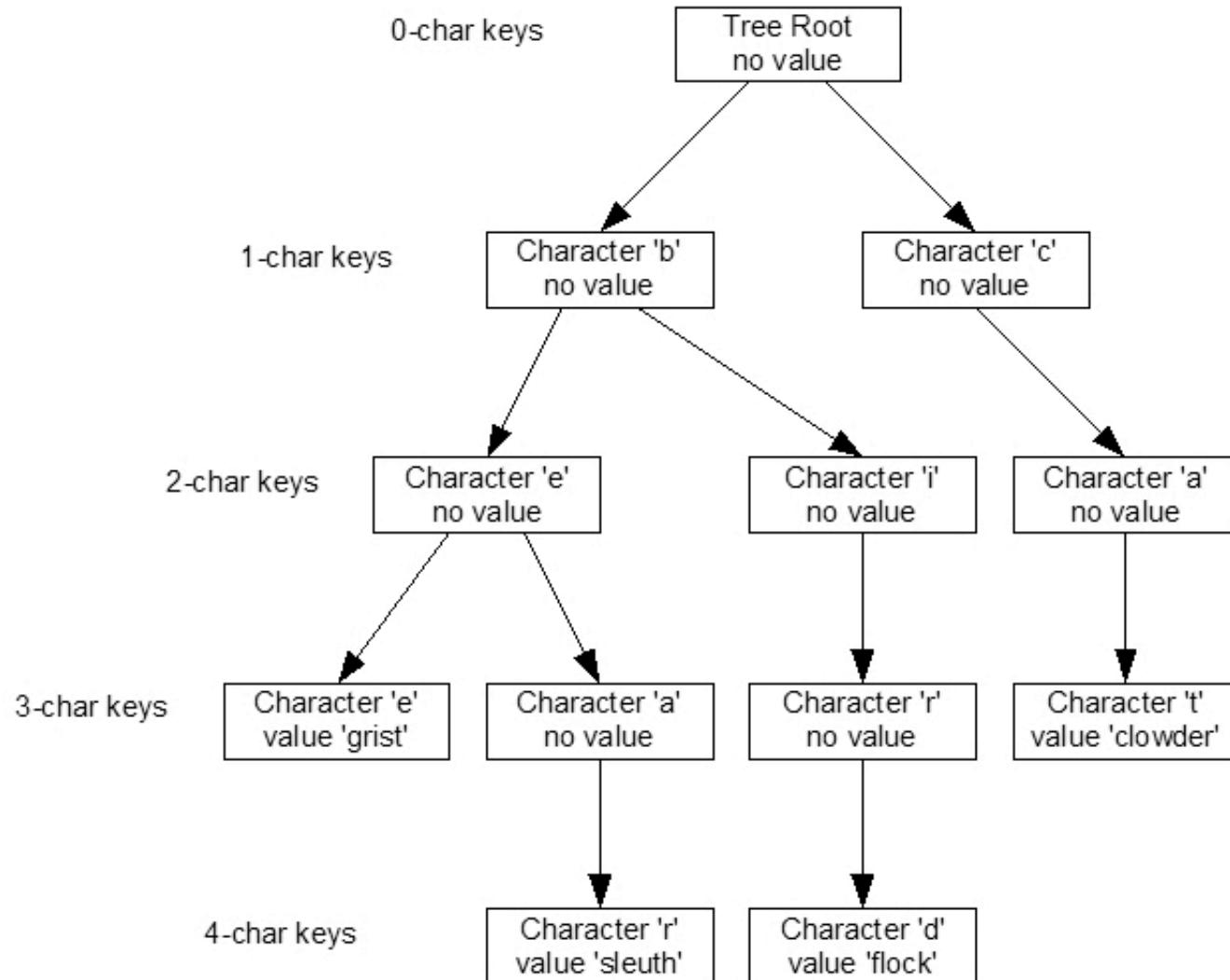
- Example: Network data structure that is a model of protein interactions



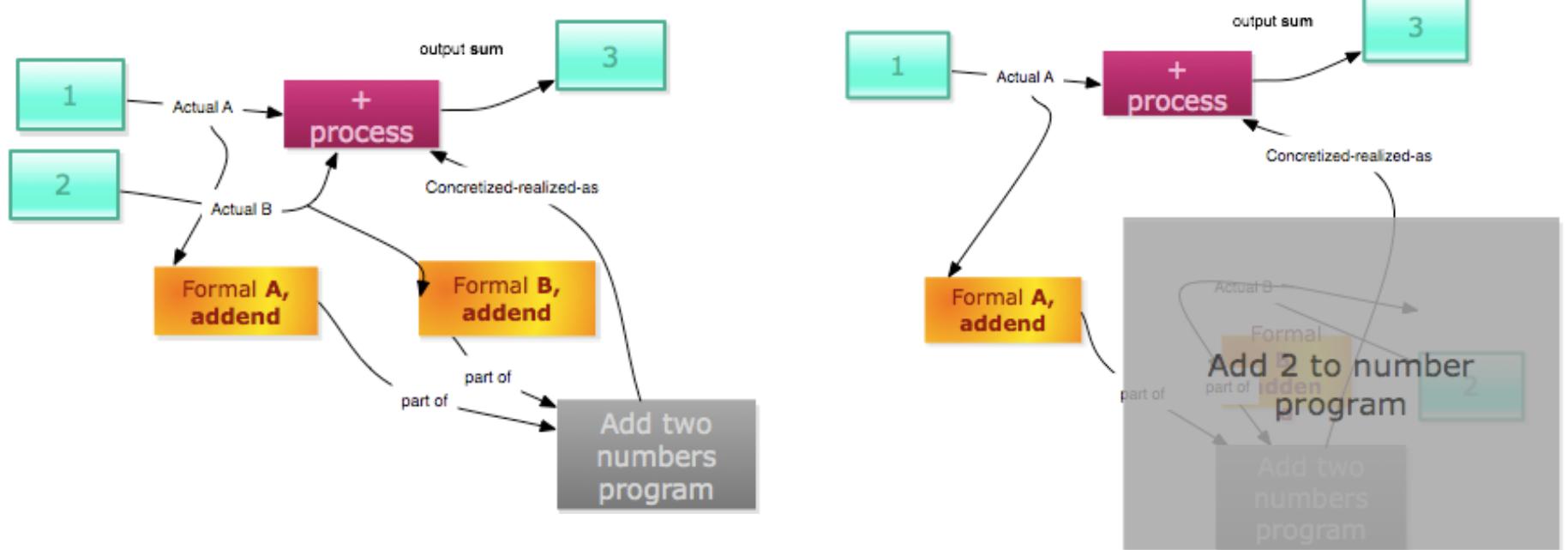
# Protein Interaction Network

- In this network model, each edge denotes a process type and each node denotes a protein type, instances of which participate in instances of this process.
- One edge connects PARP-1 (Uniprot:P09874) and BUB-3 (uniprot:O43684)
- The process denoted is protein amino acid poly-ADP-ribosylation (GO:0070212)

# Not all network data structures are models



# What are parameters?

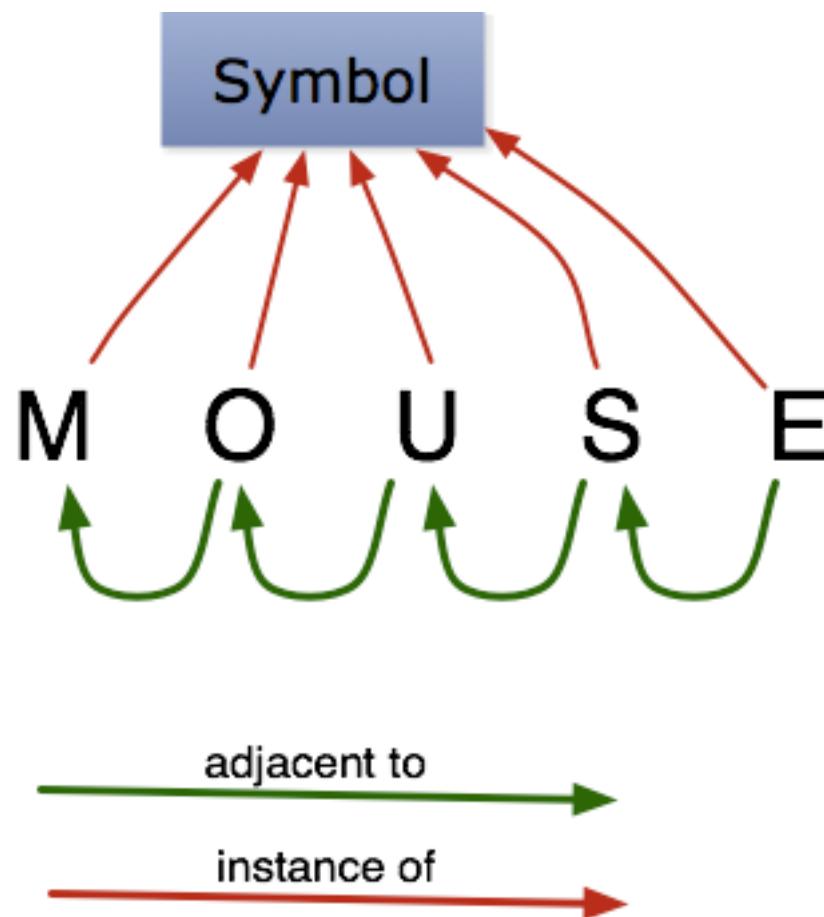


Are parameters made of the same stuff programs are?

# Encodings

- The problem: Same content information entities can be encoded in different formats.
- Suppose we have a postal address. That address can be written out in English characters, or it can be encoded in a bar code.
- A table of numbers can be encoded in several flavors of Microsoft Excel format.
- How do we describe this situation?

# Spatial encoding



# Example of alternate encodings

- Suppose we have an association (dictionary)

```
{
```

```
"bee"->"grist",
```

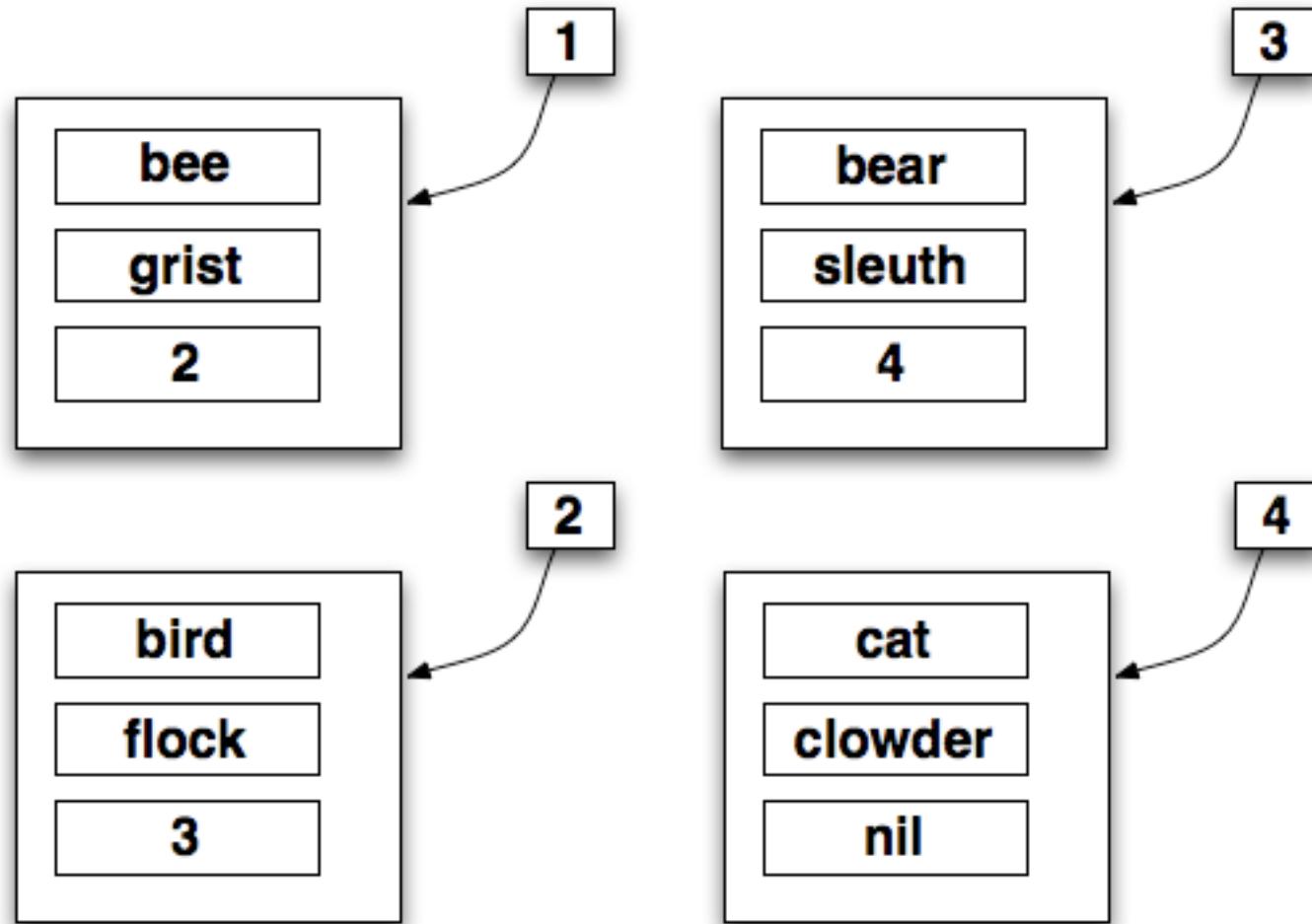
```
"bear"->"sleuth",
```

```
"bird"->"flock",
```

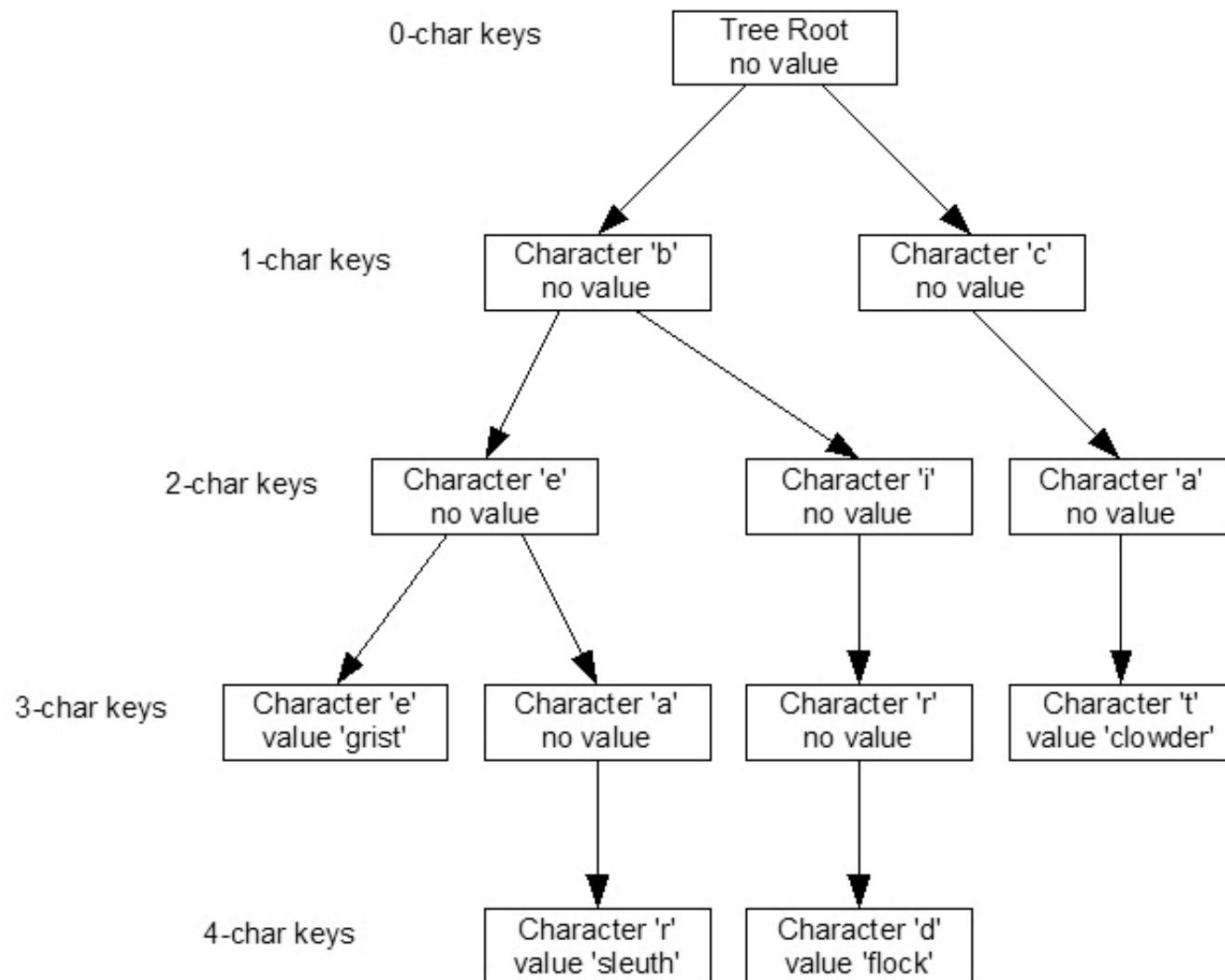
```
"cat"->"clowder"
```

```
}
```

# Encode as a linked list



# Encode as trie

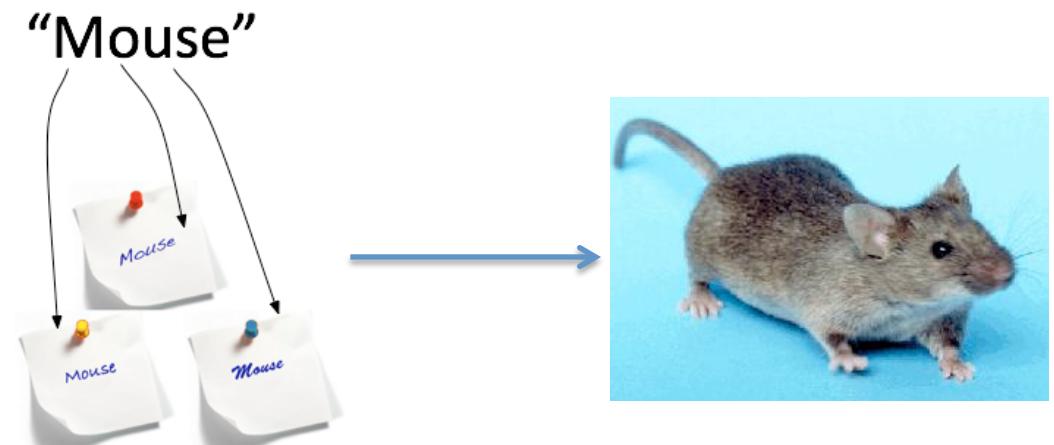


# What do we see about encodings

- The encoding is an information content entity
- Each encoding has a core with the same information content entities as components
- Each has components whose denotation is internal – to other components of the encoded entity or components of the constituent entities
- Each encoding requires a different algorithm for accessing a given element
- The denotation of the original information entities remains unchanged.

# Questions about encoding

- Are encodings distinct entities
- If they are, is there an additional dependence that needs to be added to our picture of any information content entity?
- Are concretizations the same as encodings



# Suggested answers for encoding questions

- Q: Are encodings distinct entities?
- A: Yes. Otherwise we can't explain what is the same when we have two encodings.
- Q: Are concretizations the same as encodings
- No: Encodings, in the sense described, are generically dependent. They can be concretized as qualities of different materials (at least in the patterns of charge in DRAM on different computers)
- Q: If they are, is there an additional dependence that needs to be added to our picture of any information content entity?
- A: Not sure:
  - *Against*: Even though there might be a multiplicity of concretizations of an information entity, each concretization might be of a different encoding.
  - *For*: Each concretization appears to be of an encoding
  - *Perhaps*: A generic dependence on *some* encoding (I'm not smart enough to know how to say this)

# Bibliographic information

- There are a number of ontologies and schemas for bibliographic information
  - Bibo
  - Cito
  - SWAN
  - MARC
  - Medline...
- Should IAO venture there?

# Against IAO representing bibliographic information

- There's too many schemas already

# In favor of IAO representing bibliographic information

- IAO represents the connection between information entities and the things they represent. Most (all?) existing schema don't
- IAO integrates into the OBO Foundry framework, which distributes the work of representing the non referring particulars.
- IAO might be able to provide a framework for integrating different schemas (i.e. encodings)

# Current status of IAO OWL

- In flux – mainly serving OBI's needs
- Focusing currently on how to organize, rather than on breadth
- Relations sketched, but not finished
- Lots of questions – some answers
- Still some confusion about principal axis of classification – intention? what about?
- An open project:  
<http://code.google.com/p/information-artifact-ontology/>

# Acknowledgments

The IAO is a descendant of the DENRIE branch of the OBI Project, edited by Chris Stoeckert and other OBI team members.

Currently IAO developers are:

- Alan Ruttenberg,
- Barry Smith and
- Jonathan Rees
- Melanie Courtot
- Bjoern Peters
- Werner Ceusters
- James Malone
- Paolo Ciccarresse
- Holger Stenzhorn
- Philippe Rocca Serra
- Jenifer Fostel
- Larisa Soltadova