INTRODUCTIONS

Barry Smith: NCOR, OBO Foundry
Allan Ruttenberg: Science Commons, OBO Foundry
Nigam Shah: how to use relations for reasoning over biol data
Vinaj: SRI, knowledge building based on textbooks
David Osumi-Sutherland: FlyBase anatomy development, modeling developmental biology
Larry Hunter: building systems that represent large scale biol knowledge, current represent
Peter Midford: Phenoscape, evolutionary and spatial relationships, behavior ontologies
Melissa Haendel: anatomy and phenotype ontologies, w/ Berkeley on db querying across species
Mike Bada: bio NLP tasks and ontologies
Darren Natale: Protein Info Resource, protein ontology
Karin: comp linguist, capturing relational info in text
Chris Mungall: phenotype/genotype representation
Wasila: Phenoscape, anatomy curator
Mary Dolan: MGI
Fabian Neuhaus: NIST
Suzi Lewis: GO/NCBO/OBO Foundry
David Hill: ontology development and annotation

BACKGROUND

RO v.1.0 contained binary relations only

Relations are of 3 kinds: type-type (core RO relationships); type-instance; instance-instance

Ontologies are representations of types (the general) and the goal of the RO is to create a limited repertoire of relations linking types to build these ontologies.

For example the GO contains types only, no instances. Only something that holds for all A-s will be an assertion that holds of the type A.

'All' ontologies is not ALL ontologies, there is an understanding of what the 'scope' of all is.

ACTION ITEM: scope of 'all' needs to be addressed taking into account that some ontologies (GO, FMA) are canonical, while others are not (for example, PRO includes mutant proteins)

Larry: what role does time play in 'all'?

Response: for process type, it plays no role, because the processes are the same regardless of time. However for continuants it does make a difference, For example, a cell membrane, isolated and experimented on, is used to draw conclusions about its role in the intact cell.

The need (and goal) is for a small suite of ontologies that describe the exceptions that can be used with the reference ontologies that represent the typical/normal.

Why keep RO small?

Some people are going to use the relations in RO in sloppy ways. The smaller the number of options the less the possibility of misuse. More *per se* is not bad, but if the relations are ill defined that could be bad and would poison the well.

For example, "currency has_unit $" is WRONG because there are other units of currency.

Likewise, wf we have "adjacent_to" then we don't need "adjacent_to_neuron/cytoplasm/organ". There is inherently more expressivity in the class realm than in the relation realm, thus we shall pick the former vs. the latter. But there is a

counter example of a useful releation from the Cell Ontology meeting: is_expressed _on _surface_of (aka EOSO) (CD4 receptor and cell type).

Cornelius: surface is a two dimensional entity. A gene product "*expressed _on _surface_of*" of a cell is a 3D entity, which is embedded in the cell membrane [is part of it] such that one of its surfaces is part of the external surface of the membrane. I find "expressed" to be a rather loose term. However, since it is used so universally, I don't object to its use as long as the definition of *is_expressed _on _surface_of* captures the state of affairs in reality.

The solution may be:

a. Compositional relations: could be broken down into several ontologies: RO + GO cc + cell ontology
b. Mix up UI with ontologies - show the mashed up relation even if it doesn't exist in the RO. That is: what is visible to the biologist/user is something interpretable/intuitive) and what goes into RO is whatever is necessary to capture as the formal, minimal relations)

Cornelius: *is_expressed _on _surface_of* is a good example; it belongs in an application ontology developed for particular user groups, but I question whether its current name belongs in a reference ontology. I agree that the relation representing the presence of a particular gene product within the cell membrane does.

In other words, don't just focus on one vs. the other. Further, any proposed lists of compound relational expressions should be validated by provision of definitions in terms of RO relations.

Spectrum of existing relations:

One end is the English side, what appears in papers and the other side are formal definitions

GOAL is to find a middle ground: Larry: these should be driving the development of the minimal, formal ones. If 'expressed on surface of' can be broken down into existing core ones, then good - relational cross-products. In addition to the formal definitions, non-rigorous, intuitive definitions are alos good.

Cornelius: I believe a reference ontology should contain definitions in which the differentiae take account of realty. Such Aristotelian definitions should be formulated in English and transcribed into formal definitions. When sections of such reference ontologies are adopted for incorporation in particular application ontologies, the definitions may be modified to meet the needs of, and be consonant with the views the intended users have of the ontology's domain. I assume, of course, that the RO consortium embraces the distinction between reference and application ontologies.

OPEN QUESTION: expressed_in implies located_in, what relationship does expressed_in have to existing relations in RO?

Rather than creating EOSO, could also classify receptors as cell surface receptors. Describe the same biology in different ways. Instead of capturing the location in terms of the relation, capture it in the GO/CO.

Cornelius: This may not be the whole answer. For example, in this particular case, do all cell surface receptors exist only within the cell membrane or do they also exist within the interior of the cell before they are inserted into the membrane? If they do, what should they be called prior to their insertion?

Is 'expressed' a colloquialism [euphemism] for some fundamental relation? One assertion: A is present in, i.e., *part of* or *contained in* B; second assertion: A is *expression product of* [or *product of* or some such appropriate term from the kinds of relations that designate this notion] C.

Second problem: I have always found the term *located_in* very nonspecific, though widely used.

In biological contexts it invariably implies a more specific meaning. Therefore the FMA does not employ it. *located_in* can mean *contains* [cistern of endoplasmic reticulum *contains* cytosol – or what ever the substance is; stomach *contains* gastric juice, bolus of food], *part_of* [cytoskeleton part_of cytoplasm; median nerve *part_of arm*, forearm, hand; hepatocyte *part_of* liver. It would be valid to use *located_in* in all these examples, because it is a more general location relation. However, it does not do justice to the richness of relations in biological organisms. By comparison it would not be appropriate to replace *located_in* by any of the more specific "location relations" in the following examples: toothpick *located_in* stomach; Plasmodium *located_in* erythrocyte, bullet *located_in* arm, metastasis of breast carcinoma *located_in* right lung. A foreign body, a parasite, a metastasis or embolus should not be designated either as part of or contained in the structure in which it is located; Located is the appropriate relation.

The interesting feature of the general and specific relations is the different ways they apply to relata of different sorts. The specific location relations [*part of, contains* and their inverses] work well for type-type and also instance-instance, but the general *located_in* works best for instance-instance, and not for type-type, illustrated by the bullet and metastasis examples, cases in which *part-of* would be inappropriate; in the FMA's convention [and RO's], so would be *contained-in* [more below]. I am aware that processes [normal and pathological] are often asserted to have a location [digestion *located_in* stomach; cystitis *located_in* urinary bladder], but I have never been quite at ease with such assertions. I like much better the relations introduced by RO for participants, agents etc.

There is one other example where location is the appropriate relation: coordinates, which are of two types. 1. Anatomical coordinates, such as anterior, posterior [esophagus *has-location-posterior-to* trachea; clearly an attributed or tertiary relation, provides qualitative information, pertinent to type-type and instance-instance, and is widely used. 2. Geometric coordinates, which pertain only to instances and not to types.

We have thought a lot in anatomy about distinctions between part and containment relations as well as about distinctions between different parts. Some of these thoughts are in one of our early publications before we benefited from our interactions with Barry.

http://sigpubs.biostr.washington.edu/archive/00000140/

It gives the background to several of the comments I am making here. Of course, many things have changed in the FMA since that paper.

all-some relations

All the instances of A stand in R to some B. For example: NOT - transforms_into, BUT YES - transformation_of. Or specifically, not every child transforms_into adult (single instance, apparently not on population scale), but every adult is a transformation of a child.

We (RO) don't want to become Tim Berners-Lee. No some-some relations.

Suzi: Are OBO_REL ids going to remain as the strings or will they transition to the numericals?

ACTION ITEM: move to numerical ids

Larry: Doesn't want type-type only for RO. Too constrained.

Barry: type-type require instance-type definitions

ACTION ITEM: Define instance level relations as well. Clarity is important. Relations need to be clarified whether they be used for type_type/type_instance/instance_instance.

Well-defined relationships are the ideal; clear to us, easy to explain to others

Chris: Should we split out part_of for instances from part_of for types? How would we do this?

a. Drop type-level relations (DON"T LIKE THIS)
b. Separate ids for instance and class level relations, different namespaces (THIS IS THE CHOICE, details not important right now)
c. Clarify current 'punning' semantics, same id, means different thing in different contexts (however, sometimes it's hard to tell what the context is)
d. not all type level relations can be defined in terms of instance level relations, need to work on both fronts

Long term proposal: Provide CommonLogic definition, port this to obo and owl as necessary - Fabian has some basic work done, Alan can help in translation but doesn't want to be the blocker. Have this in place for the next release.

Are ternary relationships possible? Yes.

Cyc solution (SRI): Rule macro predicates: Relation 'all' exists. Has_part A, B.

BFO (Basic Formal Ontology) is an upper ontology of entities, and the RO is an upper ontology of relations. The ro_bfo_bridge has domain and range constraints

e.g. id OBO_REL:has participant; the domain is span: Occurrent, and the Range is snap:Continuant

This is good in that they are separate and independent, but it is very awkward to maintain, BFO is primarily realized in owl, RO is in obo. The solution is to use a common language

ACTION ITEM: RO should have a commitment to an upper ontology and it shall be BFO.

RO tracker is located at http://obofoundry.org/ro. Requests can be requirements-oriented. E.g. I need a relation for this purpose, please help me formalize it.

One solution to has_part: add the qualifiers into it: all_has_part, for example

DAVID OS: REPRESENTING DEVELOPMENTAL ANATOMY

Current limitations: RO can't specify timing of transitions, can't use existing part_of relations to apply to 'sometimes'. What is needed is a system for capturing stage specific part relations.

What is developmental time?

In a developmental process, many events happen in an invariant order, these events can be used as markers of developmental progress or time (standard stage series). E.g. gastrulation is a process, which can be arbitrarily divided into stages based on marker events

Stage = BFO, fiat part of process. In other words marker events: birth, death, transformation of an anatomical structure/ beginning, end, a key point in a developmental process

Timing relationships: before, simultaneous which can both be derived from before_or_simultaneous_with

(Werner joins discussion at 11:05 am local time, during David's example of Drosophila oogenesis)

Question is: How to deal with relative processes? That is, X happens during Y only during process Z?

Ternary relations? A before B during C? - but these are not supported in OWL, so have to use binary relations.

Solution: for each type for which timing is to be captured, record a reference occurrent (see

slides for rest of detail). E.g. gastrulation reference_occurrent embryogenesis. There is still an issue with individual somite geneses - generalize to somitogenesis? (recurrent process) - individual timing relationships

Action Item: Add new timing relationships: begins_at_end_of, begins_during, ends_during, begins_before, ends_after, simultaneous_with, happens_during

In the future we hope to use ternary relationships. The solution for right now (that will work with the existing tools) is to use binary that can be mapped to ternary ones in the future. But David (and others) wants to reason using these relationships NOW. Argument for ternary relationship: can have multiple reference processes. Larry's example: oocyte maturation, follicle maturation with regards to oogenesis and with regards to life of the organism. (From Larry) Here is my example of why a single reference occurrence doesn't work (and we need a ternary relation in order to capture the reference occurrent for a particular temporal relation):

Oocyte fate determination <before> oocyte maturation [reference occurrent oogenesis]

oocyte maturation <before> follicle maturation [reference occurrent lifespan]

First, note that oocyte maturation needs to have both reference occurrents in order to be able to assert these two things. These are different in that there are a bunch of oogenesis occurents (so we need the reference occurrent), but they all end before the first follicle maturation (so we need the other reference occurrent). Incidentally, we should be able to infer Oocyte fate determination <before> follicle maturation on something like the basis that oogenesis <part-of> lifespan.

ACTION ITEM: fix phrasing to timing relationships (first_exists_during vs. begin_to_exist) to be consistent.

Fabian demonstrated an example of a reasoner that uses David/Fabian's new temporal relations. Reasoning at instance level breaks, is too slow, but reasoning at type level works.

Define: what is the same and what is new? E.g. Neuroblast vs. neuron

Alan: does NOT like succeeds, doesn't like the generalization, either have specifics or general, not both.

LUNCH BREAK occurred as a way to interrupt the escalating discussion/argument.

MELISSA: DERIVES_FROM VS. TRANSFORMATION_OF

Gave specific examples of develops_from and why this needs to be transitive: heart development. But the problem is that this is not the definition of develops_from. Do we care about identity? If not, get rid of transformation_of and keep develops_from.

develops_from

--derives_from

--transformation_of

Darren example: A -> A-PO4, is this transformation or derivation?

Define identity function first, then can decide what is transformation and what is not. Views on what identity is, is the hard part - complicated, context-dependent. For example, consider:

A -> B -> C -> D -> E

Where B derives from A, C transformation of B, C derives from D, and E buds from D

What is the relation of A to E? Is it develops from, needed from transitive closure, good for querying

Alan's issue: if there's any lack of clarity of when to use derives_from vs. transformation_of, then the system breaks down

SOLVE THE IDENTITY PROBLEM!!!!! (and solve it now)

Cell division and budding are pretty clear: division: A -> B + C whereas budding: A -> A + B

What are the examples that are clearly transformation? Rule is to stay at same level of granularity, anatomical parts down to the cellular level (NOT molecular level) - id1 => id1

caterpillar -> pupa -> butterfly

neural plate transformation of neural tube

neuron -> mature neuron

mature leaf -> senescent leaf

What are the examples that are clearly derives from? id1 => id2

[mandible derives from cranial neural crest] - a bit controversial because of the 'scaffolding' involved in bone formation

U2 neuron derives from GMC71B

leaf primordium derives from shoot apical meristem

neural crest derives from neural tube

Cornelius: Wouldn't that be: neural crest derives from neural plate? Isn't there an identifiable neural crest before the neural tube is formed? – just for my information.

Alan: Is identity reality or perspective?

Cornelius: Wouldn't boundary play an important defining role in identity and transformation? If B has the same complete external boundary as A, identity is maintained and the process A>B was a transformation. If the complete external boundaries of A and B are distinct, identity is not maintained and the process is derivation. If this proposition is valid, then identity would be reality. Does my reasoning make sense? Does this view of transformation tally with its RO definition?

If develops_from is used in cases where we don't know what to say ------ leads to a lazy way of usage. Biologists: we KNOW it's develops from, we don't know whether it's derives from or transformation of. A process needs to have a beginning and the end, The thing that develops does not have to be there at the beginning of the process but it needs to be there at some point during the process

ACTION ITEM: make a biological definition of develops_from, test at all levels of granularity, provide cases of NOT develops_from

Is transformation maturation? Is that the key? Not the single criterion. Definitely not at the RNA level. Maybe just at the organismal level.

Can one have maturation independent of some process? No.

Cornelius: Senile man *transformation_of* adult man. Surely, this would not be maturation, nor would any involution process. Round ligament of liver *transformation_of* left umbilical vein; boundary is retained, type has changed from vein to ligament and maturation does not come into play – as I understand maturation. [Some explanation: only a prenatal mammal has an umbilical vein, soon after birth the vein is transformed into a ligament; happens elsewhere as well]

COFFEE BREAK

SPATIAL RELATIONS, COMING FROM RCC8

disconnected
externally connected - adjacent?
equal
partially overlapping
tangential proper part - contained in and adjacent to
tangential proper part inverse
non-tangential proper part
non-tangential proper part inverse

Don't want to make projections of 3D onto 2D.

What do you mean by that?

If we want to use RCC keep the dimensions the same at the instance level. We should look at geospatial ontologies/descriptions for car (?) part manufacturing?

What is the difference between R and R' where: "town R county" and "island R' lake"? Or for a biological example: urine R bladder (contains? located?)

ACTION ITEM: have v. small meeting with spatial ontology (CAD, connectivity, etc.) experts + some biologists to adapt existing spatial relations to RO, who to invite? Tony Cohn. Thomas Bittner. Karen Pittman, Cycorp. Xenia Fiorentini, NIST. GOAL of this meeting: get spatial relations into RO

Cornelius: I am not very familiar with RCC, so my comments may be misdirected. I agree that the representation of spatial properties and relations of biological entities should be consistent with RCC, but I think a spatial theory of biological organisms [I should say: anatomy] needs to be an extension and specification of RCC. I believe one chief reason is that biological organisms are unique in virtue of their ability to elaborate their own parts and the spatial arrangement of these parts. The only other entity I can think of that shares this propensity is the universe. All other material entities [by that I mean the ones that have mass] are assembled from other discrete entities. I believe that RCC is limited to represent objects of the latter kind or other inanimate objects of the natural world, e.g., those of geography, and may not be sufficiently expressive to do justice to anatomy. The other reason is that limiting spatial relations in anatomy to RCC may lead to erroneous reasoning. The urine and bladder example cited in these minutes is a useful one [explored, by the way, in several of our publications].

Urine in the bladder and the bladder itself overlap in the sense that the spatial region occupied by urine falls within the spatial region of the bladder. [As far as I know, RCC does not consider boundaries and I find this one of the difficulties/disadvantages.] The same relation holds for the bladder and the pelvis. Because located_in is a transitive relation, one would conclude that there is urine in the pelvic cavity. This would, in fact be pathological state indicating that the bladder has been ruptured and would call for surgical intervention. Thus distinctions are critical between part and containment relations and between the types of continuants that qualify as contents and those that qualify as part. To arrive at the correct conclusion about relations between bladder, pelvic cavity and urine, one must in addition distinguish between part and containment relations. This, in fact was one of the objectives of the paper I cited above

I realize that some of these relations [and also nodes] may seem far-fetched, esoteric or capricious in the FMA without any explanation. They were not arrived at lightly and their justification can be found in some of the papers we published quite some time ago. I provide references to them, although I know some of you may be familiar with them

http://sigpubs.biostr.washington.edu/archive/00000155/

http://sigpubs.biostr.washington.edu/archive/00000179/

http://sigpubs.biostr.washington.edu/archive/00000222/

BTW, Tom Bittner and Maureen Donnelly would be very helpful with anatomical spatial representations.

The papers cited touch on issues I've been commenting on and also on some RCC-derived terms such as externally connected, overlap, partially overlapping etc., which also receive treatment in the FMA but under different names.

1. The entities externally connected in RCC are not connected at all, in the sense that each entity has a complete intact boundary, whereas a connection would require that their boundaries be continuous with one another over a limited area of their adjoining surfaces [such as two daughter cells an instant before their complete separation; after which they will be touching and not connected at all. The proper name of the corresponding anatomical relation is adjacency, which honors the discrete intact boundaries of the adjacent objects and specifies the relation by the additional differentia of no other entity [of the same dimension] being interposed between the adjacent entities. The FMA was the first ontology to define and employ this relation.

2. Overlap. I can take two sheets of paper and lay them on top of one another [partially or completely] and the first term I can think of to describe their relation to one another is overlap. Yet each has its own independent discrete boundary and each occupies a distinct spatial region. This is quite distinct from the relation that prevails between the bladder and urine within it, or of a cell and its nucleus. Also here, I find invoking boundaries very helpful.

3. Partially overlapping. I believe that this term is intended to suggest that a part of A, but not the whole of A, is enclosed within the spatial region occupied by B. An example cited in these minutes is the partial overlap between a bone and a [synovial] joint. It seemed to us [actually before we leaned about RCC and have been educated and indoctrinated by Barry] that the bone shared one of its parts with the joint. The fiat boundary between the unshared and shared parts of the bone is marked by the plane coinciding with the line of attachment of the joint capsule to the bone. So, we introduced the '*share part*' and '*unshared part*' distinction as a tertiary relation, which is implemented in Protégé as a slot of the part relation. [Best description is again in

http://sigpubs.biostr.washington.edu/archive/00000140/].

In addition to joints, shared parts are a characteristic feature of certain types of anatomical structures; for example, tree organs [vascular, lymphatic, biliary and tracheobronchial as well as neural trees, all have parts that are shared by the tree organ itself and the organ in which their shared parts arborize. The intrahepatic biliary ducts are part of the biliary tree and also of the liver; the branches of the femoral artery within the quadriceps muscle are part of the femoral arterial tree and also of the quadriceps, and so on. Similarly, a serous membrane such as the pleura has parts that are part of the lung as well as of the pleural sac, whereas other parts are unshared. This is true of all subtypes of 'serous sac'. [I know I am giving an anatomy lecture, but you are tinkering with anatomy and may be some of you have not thought quite as much about it as we have.] All these relations would be designated by RCC as partial overlap, which will be a very obscure term, I submit, to all health care professionals, not only to anatomists.

I am saying all this because I see that in some areas the RO consortium might be attempting to reinvent things that exist in the FMA, which is understandable since this is the prevailing practice in most ontology development, but may be there is a better way. This said, we know that the FMA is far from perfect and we would, as always, welcome any suggestion that leads to its improvement.Wasila: Spatial use cases:

teeth R bone: What is this? attached_to? connected_to? When would you use one over the other?

Cornelius: A tooth is connected to the bone of a jaw by a dento-alveolar joint [synonyms: gomphosis, peg-and-socket joint] ; def: Fibrous joint in which the periodontal ligament [synonyms: periodontium, periodontal membrane] connects a tooth to the alveolar bone of the maxilla or mandible. Example: Dento-alveolar joint of incisor tooth [FMA].

Fibrous joint: def: Nonsynovial joint in which the articulating bones or cartilages are connected by ligaments. Examples: sagittal suture, inferior tibiofibular syndesmosis, gomphosis [FMA].

bona fide boundary vs. fiat boundary Does it matter? When does it matter? It is sufficient to share a boundary? Does it have to be a bona fide boundary? Connected_to is currently a fiat boundary as in the FMA: x is connected to y if X and y share a fiat (somewhat arbitrary) boundary (doesn't reflect any physical discontinuity)

Cornelius: The dento-alveolar joint is a good use case for examining this definition [not from the FMA as cited here]. The definition holds for two relations and does not distinguish between them: *continuous-with* and *attached-to*. Both tooth [peg] and alveolus [socket] have bona fide boundaries: formed by the surface of cementum [outer layer of dentine] and of the bone of the jaw. The two surfaces are held in close proximity by a thin film of "glue', namely the periodontium or alveolar ligament, consisting of a portion of a connective tissue type. This 'glue' contains organized arrays of fine fibers, one end of which is embedded in the cementum and the other in alveolar bone. Their direction is oblique, such as to maximally resist dislodgement of the tooth [they are ripped when the tooth is pulled, some of the periodontium sticking with the tooth, the rest with alveolar bone.

It is the fibers that attach the tooth to the jaw [fiber of periodontium *attached_to* cementum, alveolar bone]; at the site where they perforate the cementum and bone, the bona fide boundary of cementum, bone and fiber is violated and we can mark it by a plane as fiat boundary [which I do not find arbitrary]. Considered at this fine level of granularity, portions of the bona fide boundaries are retained between the perforations; hence in virtue of the FMA definition, this is an attachment and not a continuity. The *continuous-with* relation is defined in the FMA as "a *connected_to* relation which holds between each anatomical entity of type A and some anatomical entity of type B such that there is no bona fide boundary between their contiguous constitutional parts". Thus the lobes of the lung are not continuous with one other, but rather are attached to one another at the root of the lung and retain bona fide surfaces elsewhere. Therefore, *connected-to* is the ancestor type of *attached-to* and *continuous-with* relations in the anatomy taxonomy of the FMA.

These relations are retrievable from the FMA but a summary of them and anatomical boundaries appears with their definitions in the most recent comprehensive account of the FMA

http://sigpubs.biostr.washington.edu/archive/00000204/

At the coarsest level of granularity one might assert that tooth *attached-to* jaw, or closer to reality

periodontal ligament *attached-to* tooth, jaw. But at these levels of granularity [resolution] one must concede that both tooth and jaw retain their complete bona fide boundaries, and none of the definitions we used so far would be valid.

Vinay: connectedness: A is connected to B means that B will move if A moves. It is grounded in physical actions/movement

Cornelius: Unfortunately that will not work in anatomy: at the coarsest level, the femur is connected to the tibia [via various things a'la periodontium] and the humerus to the ulna. Yet this connection is so designed as to allow free movement of the tibia and ulna on a stable femur and humerus in flexion and extension of the leg and arm. Note that joints connect, [some terminologies categorize them as junctions] and they have two contrasting functions: permit movement and prevent it [stabilize] between articulating pieces.

Attached_to is a connected relationship where there is a large disparity between the size of the two objects. E.g. muscle attached_to bone, tooth attached_to bone, arm attached_to body, NOT body attached_to arm,

Cornelius: None of these examples illustrate "large disparity between the sizes of the two objects". The triceps is attached to the humerus, they are of similar size, and the same is true for

all the others. The FMA specifies a requisite relation between constituent parts of the attached structures; but I am becoming overbearing with my detailed explanations, so I will not say any more.

In the case of ball and socket joint they are connected via fluid. The thought experiment is whether there are more kinds of 'connected' for things, force exerted, force connectedness? fiat connectedness?

Cornelius: Do you really think that when you are doing your push-ups, it is the film of fluid between the head of the humerus and the shallow, saucer-like socket of the glenoid that prevents your humerus popping out and sticking out toward the back of your shoulder? You may not be aware of it, but there is a huge literature and massive amounts of experimental data that document the factors for joint stability. Do you think that when you crack your knuckles, which is the sound of the fluid film breaking between the phalanx and the metacarpal head, makes your finder disconnected from the hand? I don't think there is a need for fiat connectedness.

Fabian: what kind of reasoning do you want to get out of this/support with these relationship types? Both approaches have their merits. What is the purpose? What would be useful?

Cornelius: In biomechanics, orthopedics, physiotherapy, design of orthopedic prostheses, and such like, there would indeed be a great need for an ontology of connectedness and the nature of the various connections.

He suggests, just have a single relationship called 'connected' which covers both.

Clearly, I am of a different opinion. If you are a cell biologist and are concerned chiefly with gene expression [people I have the greatest respect for and envy] then a single relationship called 'connected' suffices, I am sure.

Barry's motion: add attachment, connected, synaptic connection, get definitions from FMA which need improvement, improve them, submit to biologists for review with examples, find out why FMA used them that way. It looks like discussions on the RO list led to use of these proposed types (with same names) in the opposite direction. Is there really confusion or not? Not.

I know it sounds smug, but we feel there is no confusion about these relations if the FMA. We are aware, however, that the definitions of many [probably even all] need cleaning up and some rigor needs to be added to them by formalizing them. We feel the ones in the FMA are relevant to anatomy, but, depending on the scope of RO, that does not mean that they would all be relevant to RO.

Peter Midford: overlaps relationship

Overlaps means two things share parts, for example: synovial joint

A *part of* Bone 1

A *part_of* Synovial Joint 1

Therefore Bone 1 overlaps Synovial Joint 1

Cornelius: I think I said enough about overlap earlier, when this topic was first mentioned. Just to comment on the current example: The FMA categorizes joint as an anatomical cluster. Some members of the cluster such as the synovial capsule, some intra-articular ligaments and menisci are part only of the joint [*unshared parts*], but the parts of articulating skeletal pieces of necessity are also members of the cluster. The FMA regards these epiphyses [articular ends of bones] and their articular cartilages as *shared parts* of the bone and the joint.

Logically, everything overlaps with its parts. Are these symmetric on the instance level but not

on the type level? This seems the case for the joint example.

I do not see that: in the canonical sense, every instance of distal epiphysis of femur is a shared part of some femur and some knee joint. And every knee joint has shared part the distal epiphysis of some femur [but see below].

However uterine tract overlaps urogenital system but NOT urogenital system overlaps uterine tracts

Cornelius: This is interesting and illustrates the penetration or tenacity of dogma perpetuated by generations of anatomists. Urogenital system is a 'concept' in Barry's true sense of the word. There is a urinary system, which in both female and male consists of the kidneys, ureters, bladder and urethra. There is also a genital system, which in the female consists of ovaries, the uterine tubes, the uterus and the vagina. Urine never travels along any part of this system, and eggs never travel along any part of the urinary system. In the male the genital system consists of the testes, defent ducts, seminal vesicles, prostate and urethra. Urine travels only through the urethra and never through any other part of the genital system and sperm travel only through the urethra and never through any other part of urinary system. The urethra is derived from the phallic part of the urogenital sinus of the embryo, which becomes laid open to the amniotic cavity [or the embryo's exterior] when the urogenital membrane breaks down. In the female the opened up lumen of the sinus becomes the vestibule of the vagina [the open space between the labia minora], whereas in the male the phallic part of the sinus is rolled up again to form the penile urethra. In the vestibule of the vagina urine and discharged eggs may mix; but semen and urine never mix. Let me ask you, does the common passage for sperm and urine justify inventing the overlapping of the genital urinary systems into one entity? I do not think an ONTOLOGY should admit such a 'concept'!

Moreover, the classical systems are supposed to be defined in accord with their distinct functions. But many different anatomical entities have the disposition for the same kind of function and many different functions are inherent in a variety of instances of anatomical structure. It is true that only parts of the gastrointestinal system have the potential for digestion, but is that not the same potential as a lysosome or a vacuole in a phagocyte has for breaking down certain substances? The kidney secretes angiotensinogen, erythropoietin and renin among other things beside excreting urine. Should it overlap with the endocrine system? Should a macrophage which secretes cytokines?

I know that it is highly unlikely that my arguments will have any influence on such long perpetuated beliefs, but Barry advocates that a good ontology should be constructed on the basis of evidence rather than beliefs. So I vote for not having overlap between organ systems. As anatomical structures, some organ systems may have some shared parts.

ACTION ITEM: Add symmetric_overlaps/mutally_overlaps, modify the proposed definition.

If the femur overlaps knee joint then the reverse is knee joint overlaps femur. Are both true? Is it important to capture that this is a symmetric overlap?

Is there any part of the knee joint that is also part of the femur? Is the lower end of the femur part of the knee joint? Above we concluded yes. But think about another one: The tracheobronchial tree is one organ with the trachea and bronchi being its parts. Each lung is an organ distinct from the tracheobronchial tree. Some pats of the tracheobronchial tree are enclosed within each lung [the intrapulmonary parts of the bronchi]. The intrapulmonary bronchi are parts of the both the tree and the lung. Is any part of the lung also part of the tracheobronchial tree? This situation is the same as branches of all trees [vascular, neural and duct trees – e.g., biliary tract] which arborize within organs. Is any part of the liver part of the biliary tree or any part of the triceps part of the radial nerve?

Wasila notes that in the spatial ontology top level contains both dependent and independent continuants: anatomical axis, anatomical axis direction, anatomical compartment, anatomical

compartment boundary, anatomical gradient, anatomical region - included margins, anatomical section, anatomical surface

True.

What is a character description? Characters from a phylogenetic perspective, cover the variation in a single bone, how is it shaped differently across different species. Sample character: shape of posterior margin of lateral ethmoid has one phenotype in one species and another phenotype in another species. Of another example: the Character is the presence or absence of bone X. Or the position of bone X with regard to bone Y, where bone X may be posterior to Bone Y or Bone X may be anterior to Bone Y.

"posterior margin of lateral ethmoid" must refer to a part of the bone [3D entity] if it expresses a phenotype, whereas all the spatial ontology entities of the FMA cited above are 2D or 1D entities, and cannot have a phenotype.

ACTION ITEM: make GO spatial-related process definitions consistent with spatial ontology

ACTION ITEM: add more axis-related terms to spatial list (adaxial/abaxial, apical/basical, caudal-cephalic)

ACTION ITEM: take anything relational out of Spatial and move to RO, check out whether 'vicinity of' needs to stay

ACTION ITEM: remove spatial stuff from FMA

Cornelius: We would really like to know what is meant by *spatial stuff* in the FMA. And what is wrong with it? If it is wrong, we would like to amend it. Could you also give some examples of *spatial-related process* in GO? It is not easy to find the answer.

ACTION ITEM: decide what needs to stay in PATO and what needs to go into RO - punt to tomorrow

ACTION ITEM: Define distal and proximal, ternary relationships?

ACTION ITEM: Add medial and lateral?

Cornelius: There are other anatomical spatial descriptors: superior and inferior, anterior and posterior, rostral caudal, dorsal and ventral [you list most in the summary]. Some of these terms have different meanings in quadrupeds and bipedal, erect organisms. Some are used interchangeably within and between species and that creates a lot of confusion about deciding between equivalents. For example, is the anterior lobe of the prostate of the mouse equivalent to anterior or superior lobe in the human? One is particularly prone to benign and the other to malignant neoplasia in the human and using mouse as a model organism mandates to identify the correspondences.

CHRIS: IS_AGENT

Is there really agency ? NO because of the absence of intent. But still need something more specific than has_participant, e.g. heart process has_participant heart

Alan : NO NEED FOR AGENT! (see his very complicated slide about how one can get around this using roles and owl). The only role related relation Allan thinks is necessary: realization_of

Fabian: participates in?

Is it one or is it a bunch? Punt to tomorrow.

DAVID: REGULATES

Fabian: can processes that regulate a process be a part of that process? Yes. Good.

Ignoring regulations of qualities for now? Yes. There are not targets in the GO that already exist.

Alan: so many things affect transcription, how do you distinguish some that regulate and some that don't

By restricting to occurrents, we avoid problems that could arise from things like 'temperature' and 'nucleotide concentration' affecting the 'rate, frequency, extent.' I.e. RNA pol II does not regulate transcription, it carries out transcription.

What about other parts of the transcription machinery? If they're always there, they do transcription. If they're sometimes there, they are regulating.

If it's static canonically, it does not regulate.

Cornelius: I have found regulates always quite ambiguous and have not seen a satisfactory definition that would distinguish between the different usages.

Alan has problems with the current definition: one that includes all the things based on your intuition. cite the right entities. We should have some procedure so that I can take two processes that are happening in the vicinity of the other process that let's me tell whether one is regulatory and the other is not. Right now, it's not workable. Define rate. Who are the participants? What is the rate of a process? Duration makes sense. What is frequency of a process? His sense is that the current definition is completely lexically driven. Define in terms of a continuant which feeds into the process (did I catch this right?). In as many processes as possible.

There are participants in the BMP signaling pathway that directly do the regulation of transcription. There are proximate processes that affect transcription.

Suzi: it's the other way around. There is big effort in GO to use regular syntax when naming things. These things got named this because of what they did.

The rate of a process is how many times does this get initiated?

Larry: Can Alan improve 'regulates' any further?

Qualities of occurrents? Can they exist?

causation - regulation - How far back do you go? What is the scope? Basic metabolic processes regulated EVERYTHING. Where is the beginning and end? What is direct and indirect?

possible solution: use comparatives, why this instead of that? The comparative approach helps us do science. All of experimentation is comparative.

Why isn't the production of BMP part of this process? Probably yes, but this figure was taken out of a paper.

Is regulates transitive? If A regulates B and B regulates C, does A regulate C?

ACTION ITEM: improve on the current 'regulates' definition referring to continuants which interact and fulfill their functions, then see if this definition works and can be applied to the use case/s that we have

When known, the most proximal step to a process, should be the regulator of that process. In the example, binding of SMAD4 complex regulates transcription, all the previous stuff regulates transcription by inference.

Cornelius: Doesn't that mean simply that process A precedes process B, i.e., is dependent on it. Doesn't regulation imply modifying an ongoing process, such as its rate?

Do we need 'regulates' AND 'indirectly_regulates'? Not really.

Is regulation a mix of instructive and permissive signals?

Cornelius: May be you should consider processes other than transcription that are regulated, such as flow, rate of cell production, muscle contraction, muscle tone; you might gain a more comprehensive insight into regulation.

Hh signaling regulates limb development? Melissa/David: yes David OS: BS! -lots of argument, David H changes his mind, Hh signalling part of limb development. In summary, instructive signals would be part of the process

BMP signaling is required for dorsal/ventral patterning, it is instructive, it needs to be a part of.

Darren: If all we know that BMP binding the receptor results in change in tsc, then create BMP signaling pathway regulates transcription. As we know more, add more information to the ontology, change the annotation to something more specific.

Larry: why take away the more general and only keep the detail?

How many senses of 'regulates' are being lumped together in the relationship 'regulates'?

END DAY ONE - GOING TO DINNER at 6:20 pm after beginning at 8:50 am

Day 2: starting at 9:13 am with tasks from last night

(1) Melissa/David: Biological definition of develops_from (at the cellular level) and the continuity of cell lineage.

Darren: c develops from d iff there is a continuity of lineage extending from d to c, where:

1. the existence of d preceds the existence of C
2. some portion of the material of d is retained in the material of c
3. one can trace the event or set of events linking d to c

What is 'material'?

In the FMA it means that an entity has mass. Physical means that it has spatial dimension. A cell is material so are respiratory gases but the cavity of a pulmonary alveolus or that of a phagosome is immaterial.

Alan: still need and should have developmental specific definition, build the specific definitions (developmental, protein) first, then build the general one later

cell1 parent_of cell 2 (single division process occurs)

cell1 in_lineage_with cell2, if cell1 = cell2 or cell2 is some sequence of parent relationships to the cell2

(develops_from within a single organism) - gets around Larry's lung developed from his grandmother's ovary

anatomical structure 1 develops from anatomical structure 2 iff some cell in anatomical structure 2 in lineage with some cell in anatomical structure 1.

Is there a problem with immaterial entities? Eg. blastocoel. Not really, we don't care about them.

How about someone else caring?

What about secretions, things that are built out of secreted structures? How do we extend to those? Use 'secretes'?)

The FMA distinguishes 'portion of secreted substance' and 'portion of excreted substance' as two of the several subtypes of portion of body substance, but their instantiation is not complete.

This basic definition appears to apply on multiple levels: cellular, biochemical, anatomical structures and organisms. We will try this out.

Suzi: some events that result in something new, some events that result in 'rearrangement of material'

Alan: we want at least the cell biological level defined, good logical definition in the terms that biologists use and understand, build from the cellular level up.

(2) ontologist's justification of why the two children terms are important ;

Biologists don't care explicitly, use it implicitly. Ignore the two children for now.

Melissa: homology relation? 'descends_from'

homologous structures vs. analogous structures. For example, mouse limb homologous to human limb, but bird wing is analogous to bat wing

character = quality (shape of bone margin)

character state = variability in quality

Peter:

Review of tree building based on three characters, looking at the various character states independently reveals different things about the individual ancestral states. Comparing two trees derived from two different sets of characters reveals different phylogenetic relationships.

Homology evidence codes: (in use by Phenoscape using Phenote)

inferred from morphological similarity
inferred from positional similarity
inferred from developmental similarity
inferred from compositional similarity
inferred from gene expression similarity
inferred from phylogeny

Definition: (see RO meeting wiki but remove taxon reference and replace by 'population', remove the Instance 3-ary relation - Melissa editing now)

directly_descends_from = (definition, get rid of partially a copy, get rid of determination is instance_level

descends_from = ? in_lineage_with from the earlier discussion on develops_from

This looks good to Barry, fix italicization issue with definition

common ancestor forelimb

```
      /              \
     /                \       descends_from
    /                  \
   /                    \
  /                      \
bat wing          human arm
```

bat wing <-homologous_to -> human arm

ACTION ITEM: Two relations: 'descends_from' (not the final string), homologous_to both with definitions - generalize definition, work on last bit of definition relating to genetic material and genetic programs

What do you use to describe the relationship between a parent anatomical structure and a child anatomical structure

Darren: current definition is restricted to anatomical structures, huge problem for proteins/genes

Chris: structure of definition is fine, need to generalize homology to apply to proteins/genes, anatomical structure could broadly refer to character states

Melissa: need homologous_to NOW for anatomical structures

Note: Just because two anatomical structures are homologous, it does not mean that its parts are homologous. Structures that are homologous could use completely different programs.

David OS: transitive closure is important only between two 'generations'

1001
0001
0101
0110
0100

David OS: evolved_from relationship. The homologous_to relationship does not involve time, can hold between simultaneously existing entities: human arm and bat wing. In contrast, evolves_from does involve time: human arm and 'ancestral forelimb'

Is evolves_from built from directly_descends_from? Yes. Figure out strings to make them parallel.

BREAKTIME

Regulates discussion: not reopened to avoid non-constructive and never ending discussion. David, Tanya and Chris will come up with a more rigorous definition of 'regulates' and present that to the group for review.

Larry: epistemic relations or how do we know what we know?

supports
is_supported_by (inverse)
contradicts
is_contradicted_by (inverse)

What is scientific evidence? (evidence that supports a hypothesis - 'citeable'?)

--peer review

--capable of being communicated

--secret information can't be scientific evidence

Making physical objects scientific evidence is a problem???

--"bug" with pins in it? scientific specimen: is this evidence?

--how about clones?

information entity - something that is communicated

GO gene_association file (GAF) use case

3 epistemic columns:

*DB:reference: <assertion> is_supported_by <DB:reference>

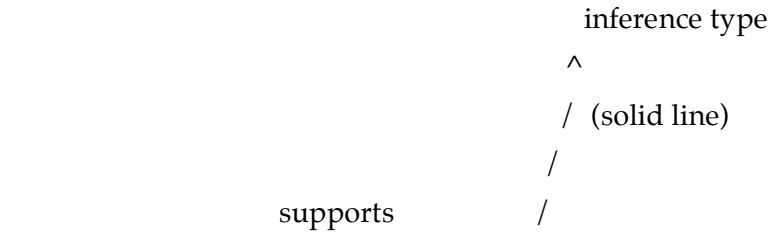
*NOT qualifier - flips evidence, not assertion: <assertion> is_contradicted_by <DB:reference>

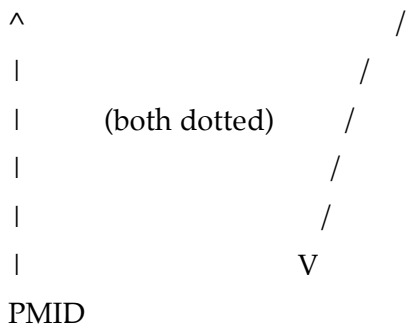
*evidence code

How to handle one piece of evidence that supports two (or more) assertions in different ways.

I.e. sentence 1 provides info for IDA and IC, how do we capture these multiple roles?

1. sub-relations: subclass binary supports, e.g. supports_by_direct_assay
2. ternary relations: Make supports a ternary relation: supports <assertion> <evidence> <evidence type>
3. PMID = ECO type
4. represent the inference





asserted sentences have a speaker

asserted sentences = things we believe are true

(When all fails, speak louder and/or faster.)

partial constraints for supports?

Barry: add to ROIL

Tanya: has(function)

ACTION ITEM: table: map relationship_types to RO relations if known, if not put?, also add range of things that go into term2, send to group (obo-discuss, ro). For protein modification, use psi-mod

(potential target for)

Mike: Biochemical relations

Built GO to CheBI cross products

Functions defined in terms of processes by which they can be realized:

kinase function = catalytic function THAT is_realized_in GO:kinase activity

Definitions are wrong!

THEY WANT: Catalytic processes and Catalytic activities defined in terms of processes they are catalyzing:

GO:kinase activity = GO:molecular_function THAT results_in_catalysis_of GO:phosphorylation

This (Mike thinks) necessitates an ontology of types of (bio)chemical reactions. GO lays out potential to do something vs. the doing something. These both exist. We will not instantiate BOTH function AND functioning. It is simpler to make links between functions and processes.

Process(function) = realization of function, and Function (process) = realized_as process

Alan: automatic generation of one term when the other is made

DavidH: how about creating links between MF and BP and the relationship describes that link?

Discussion about Mike's definitions and how they could be improved.

David H: We will create a relationship describing that a gene product has the potential to execute a molecular function (has disposition to realize?). This way the MF ontology describes the realization (execution) of a molecular function. These executions will then be allowed to be part_of BPs. All MFs are part_of some type of BP.

ACTION ITEM: Larry will create and use too many relations (TMRs) Mike has made now as 'macros' or shortcuts while the RO people write the underlying framework that will support them. If some of them don't break down properly, these will be fixed. Relations will be sent to Alan together with Mike for training. Code named: Mike And Chris Relations Ontology, or ROC, ROIL, MAC-RO, MAC-ROP

Chris: cross products, GO biological process and cell ontology

There are a bunch of different proposed relations: see ro_proposed: aka Chris' macros (how many? 12, development, anatomy and cell, for chemicals about 10)

ACTION ITEM: find an example of Chris' macro where there are two relations that identify two participants to verify claim - source and destination for a transport event

ACTION ITEM: again, find a way to translate the macros into a formal definition

Pattern: x results in A x2; A is term from another ontology

Alternate: for some y A(y); participant (x2,y) ^

(tried to transcribe the above from Fabian's writing on the board but missed some parts)

Necessary is easy, necessary and sufficient is hard

SUMMARY OF ACTION ITEMS

1. *ACTION ITEM: scope of 'all' needs to be addressed taking into account that some ontologies (GO, FMA) are canonical, while others are not (for example, PRO includes mutant proteins)*
2. *ACTION ITEM: move to numerical ids*

3. *ACTION ITEM: Define instance level relations as well. Clarity is important. Relations need to be clarified whether they be used for type_type/type_instance/instance_instance.*
4. *ACTION ITEM: RO should have a commitment to an upper ontology and it shall be BFO.*
5. *Action Item: Add new timing relationships: begins_at_end_of, begins_during, ends_during, begins_before, ends_after, simultaneous_with, happens_during*
6. *[action item] stages and temporal relations: are there cases where we do not need the reference process?*
7. *ACTION ITEM: fix phrasing to timing relationships (first_exists_during vs. begin_to_exist) to be consistent.*
8. *ACTION ITEM: make a biological definition of develops_from, test at all levels of granularity, provide cases of NOT develops_from*
9. *ACTION ITEM: have v. small meeting with spatial ontology (CAD, connectivity, etc.) experts + some biologists to adapt existing spatial relations to RO, who to invite? Tony Cohn. Thomas Bittner. Karen Pittman, Cycorp. Xenia Fiorentini, NIST. GOAL of this meeting: get spatial relations into RO*
10. *[action item] spatial relations: add to ro attachment, connected, synaptic (use fma definitions, work on them, circulate to biologists to vet)*
11. *ACTION ITEM: Add symmetric_overlaps/mutally_overlaps, modify the proposed definition.*
12. *ACTION ITEM: make GO spatial-related process definitions consistent with spatial ontology. Fix, harmonize, extract from appropriate places -- anything relational from spatial.obo to ro*
13. *ACTION ITEM: add more axis-related terms to spatial list (adaxial/abaxial, apical/basical, caudal-cephalic). Make precise what is meant by medial, lateral -- dorsal, ventral -- proximal, distal -- anterior, posterior*
14. *ACTION ITEM: take anything relational out of Spatial and move to RO, check out whether 'vicinity of' needs to stay*
15. *ACTION ITEM: remove spatial stuff from FMA; see my earlier questions.*
16. *ACTION ITEM: decide what needs to stay in PATO and what needs to go into RO - punt to tomorrow*
17. *ACTION ITEM: Define distal and proximal, ternary relationships?*
18. *ACTION ITEM: Add medial and lateral?*
19. *ACTION ITEM: improve on the current 'regulates' definition referring to continuants which interact and fulfill their functions, then see if this definition works and can be applied to the use case/s that we have*
20. *ACTION ITEM: Two relations: 'descends_from' (not the final string), homologous_to both with definitions - generalize definition, work on last bit of definition relating to genetic material and genetic programs. new relations descends_from (already used so need a different string!), homologous_to can be inferred.*
21. *[action item] evidence: add supports and contradicts relationships, small working group for domain and range*
22. *ACTION ITEM: table: map relationship_types to RO relations if known, if not put? Also add range of things that go into term2, send to group (obo-discuss, ro). For protein modification, use psi-mod (potential target for)*
23. *[action item] tair: make table of tair relations, column for ro term if it exists, send to ro group make request for needed terms*

24. *ACTION ITEM: Larry will create and use too many relations (TMRs) Mike has made now as 'macros' or shortcuts while the RO people write the underlying framework that will support them. If some of them don't break down properly, these will be fixed. Relations will be sent to Alan together with Mike for training. Code named: Mike And Chris Relations Ontology, or ROC, ROIL, MAC-RO, MAC-ROP*
25. *ACTION ITEM: find an example of Chris' macro where there are two relations that identify two participants to verify claim - source and destination for a transport event*
26. *ACTION ITEM: again, find a way to translate the macros into a formal definition*