preprint version of paper published in *Drug Discovery Today: BIOSILICO*, 2004:2(6);246-252

On Controlled Vocabularies in Bioinformatics:
A Case Study in the Gene Ontology

Barry Smith[1,2] and Anand Kumar[1]

[1]Institute for Formal Ontology and Medical Information Science,
Saarland University, Saarbrücken, Germany.
Tel. + 49 341 97 16170, Fax. +49 341 97 16179
[2]Department of Philosophy, University at Buffalo

phismith@buffalo.edu, anand.kumar@ifomis.uni-saarland.de

Abstract

The automatic integration of information resources in the life sciences is one of the most challenging goals facing biomedical informatics today. Controlled vocabularies have played an important role in realizing this goal, by making it possible to draw together information from heterogeneous sources secure in the knowledge that the same terms will also represent the same entities on all occasions of use. One of the most impressive achievements in this regard is the Gene Ontology (GO), which is rapidly acquiring the status of a de facto standard in the field of gene and gene product annotations and whose methodology has been much intimated in attempts to develop controlled vocabularies for shared use in different domains of biology. As the GO Consortium has recognized, however, its controlled vocabulary is as currently constituted marked by a number of problematic features which are characteristic of much recent work in bioinformatics and which are destined to raise increasingly serious obstacles to the automatic integration of biomedical information in the future. Here we survey some of these problematic features, focusing especially on issues of compositionality and syntactic regimentation.

Keywords:
Ontology, Gene Ontology, Controlled Vocabulary, Terminology, Gene Product Annotation, Compositionality, Information Extraction, Syntax

GO’s Three Ontologies

The Gene Ontology (GO) [1] is an important tool for the representation and processing of gene- and gene-product-related information across all species. It provides a ‘controlled vocabulary,’ designed to support the work of researchers in biomedicine by enabling them to report the results by using a common terminology in annotating genes and gene products.

When a gene is identified, three important types of questions need to be addressed:

- Where is it located in the cell?
- What functions does it have on the molecular level?
- To what biological processes do these functions contribute?

GO’s controlled vocabulary is correspondingly built out of three terminologies consisting of cellular component, molecular function, and biological process terms, respectively. As of March 15, 2004 GO comprehends 1395 component terms, 7291 function terms, and 8479 process terms. These form three separate graphs, whose primary purpose is to allow researchers annotating genes and gene products to locate where the features and attributes they are addressing in their work might lie (their position in logical space) in relation to other, more familiar features and attributes and thus either to pick out corresponding terms already existing within GO’s controlled vocabulary or to localize corresponding gaps in the existing hierarchies and so recommend new terms which need to be included.

GO’s **Cellular Component Ontology** consists of terms such as *flagellum*, *chromosome*, *ferritin*, and *virion*, terms which (with a few exceptions – above all *cell* itself, and *extracellular matrix* and *extracellular space*) relate to entities properly included within a single cell. All cellular components are, like the cell itself, continuant entities (entities which *endure* – which means that they are such as to preserve their identity over time even while undergoing changes of various sorts). [2] This ontology is the counterpart in the GO environment of what is otherwise called *anatomy* (though GO contains also a fragmentary ontology of anatomical structures at levels of granularity higher than that of the cell in its treatment of terms such as *fat body development*, *gonad development*, *thyroid gland development*, and so forth, in its biological process ontology). The purpose of the cell component ontology is intended to allow biologists to register the physical structure with which a gene or gene product is associated.

GO's **Molecular Function (Activity) Ontology** consists of terms such as *ice nucleation activity*, *binding*, and *protein stabilization activity*. The GO definition of *molecular function* is: "activities, such as catalytic or binding activities, at the molecular level." This ontology is thus intended to consist of processes, which is to say occurrent entities which do not *endure* but rather *occur*. Where the level of granularity of the entities captured by GO's cellular component ontology is that of the *cell*, the molecular function ontology comprehends functions of both intracellular and extracellular molecules. Such functions are also, somewhat confusingly, referred to as 'activities'.

GO's **Biological Process Ontology** consists of terms such as *glycolysis* or *death* or *adult walking behavior*, terms referring to entities at both the cellular and the whole organ or organism levels of granularity. A biological process is defined in GO as: "A phenomenon marked by changes that lead to a particular result, mediated by one or more gene products." Molecular function and biological process terms are thus clearly closely interrelated: both refer to occurrent entities, which means: entities which unfold themselves in time.

What, now, is the relation between biological processes and molecular functions in the GO framework? Certainly there is such a relation on the side of the corresponding entities in reality. Thus the biological process of *anti-apoptosis*, for example, clearly stands in some relation to the molecular function labeled *apoptosis inhibitor activity*. GO's curators attempt to clarify this relationship as follows: "A biological process is accomplished via one or more ordered assemblies of molecular functions." This would suggest that molecular functions are *constituents* of biological processes, so that they would stand to such processes in a *part-of* relation. The problem is, however, that GO's authors insist at the same time that the relation *part-of* obtains only between entities within a single ontology. Thus while they can capture the relatively unproblematically parthood relations which obtain between biological processes and their biological process parts they have no means of capturing the relations between biological processes and the molecular functions which underlie them. Thus receptor binding (a molecular function) and signal transduction (a biological process) are not related in the Gene Ontology, and neither are transcription factor activity and development.

This is not as yet a fair ground for criticizing GO. Its principle according to which the three GO ontologies should be kept free of crosslinks between them is a design choice which has borne considerable practical fruit. The problem turns rather on one unanticipated consequence of this design choice, which consists in the fact that – as is illustrated by the still not satisfactorily resolves instability in GO's handling of its function terms – this choice serves as an obstacle to the understanding of such terms of the part of GO's curators.

This problem can be to some degree alleviated externally by appealing to the fact that the terms in question do in any case become unified indirectly, where single gene products are simultaneously annotated via terms from different GO ontologies. Thus of the 84,833 annotations within the Gene Ontology Annotation from TIGR (GOAT) Database, [3] more than half were simultaneously annotated to terms within two of GO's ontologies, and more than 10% were annotated to terms from all three ontologies. We are currently analyzing these cases as a basis for extending GO by establishing corresponding cross-ontology links between the corresponding terms. [4]

GO as a 'Controlled Vocabulary'

GO's considerable success is testimony to the wisdom of a number of other crucial choices made by the GO Consortium in the early stages of its development. Above all, the adoption of a relatively simple graph-theoretic architecture (see Figure 1) involving just two kinds of edges (labeled *is_a* and *part_of*) meant that work on populating GO could proceed very quickly. Such work does not require the completion of complex protocols, but can be carried out intuitively by the expert biologist, who is subject to few formal constraints when incorporating new terms and definitions.

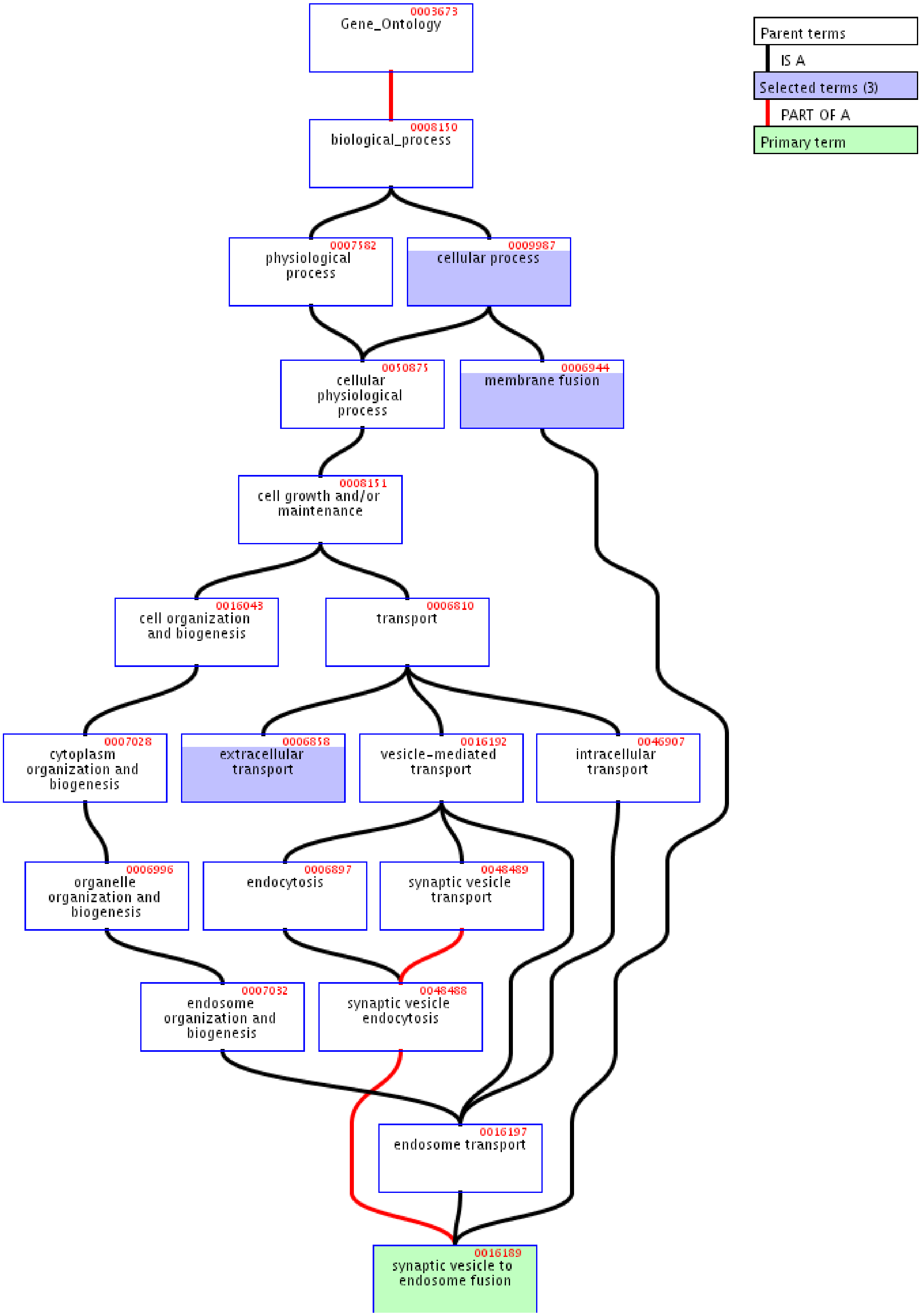

Figure 1. Relations between sample GO terms as rendered graphically by the QuickGO browser (http://www.ebi.ac.uk/ego/) © European Bioinformatics Institute

In a series of recent papers we have attempted to show, however, that there are also certain unintended negative consequences of these choices also. More precisely, we have argued that the authors of the Gene Ontology have ignored certain benefits which can accrue through the application of formal and syntactic rigor in the formulation of terms and definitions. The upshot is that there are aspects of GO's current architecture that are predestined to cause ever more serious problems as GO increases in size. For on the one hand, as the GO Consortium itself accepts, it will in the future 'be increasingly difficult to maintain the semantic consistency we desire without software tools that perform consistency checks and controlled updates.' [1] Yet on the other hand much of the information that GO contains is, under current policies, not capable of being accessed or manipulated by software tools.

For this, adherence to basic principles of logic is required, and such principles are thus destined to play a vital role in GO and similar bio-ontologies in the future as the obstacles to manual inspection and curation become ever more significant. If formal tools are to be employed for purposes of curation, however, then this means that the information content of GO must be accessible to such tools. This means in turn that the language of GO must approximate ever more to the condition of a compositional language, that is, to a language wherein the meaning of each compound expression is a function of the meanings of its constituent parts. The GO consortium has acknowledged the significance of this fact and under the auspices of the Open Biological Ontologies umbrella organization it is currently embarking on a program of reform which is in important respects in conformity with the proposals advanced in our earlier papers. In [5] we focused especially on inadequacies in GO's specification of the relations between its *function* and *process* ontologies, and on associated problems with GO's recent adoption of the suffix 'activity' to its function terms. In [6] we generalized this critique by pointing to certain inadequacies in GO's treatment of the relations between entities at different levels of granularity. In [7; see also 8, 9] we pointed to a series of difficulties and unclarities in the two foundational relations *is-a* and *part-of*, which constitute the edges of the three graph-theoretic hierarchies from out of which GO is constituted. In [10] we added discussion of formal inadequacies in GO's definitions, attempting to show how adherence to formal organizing principles drawn from philosophical ontology, principles which represent best practices in classification and definition, can lead to benefits in eliminating certain characteristic types of error by which GO has hitherto been affected.

Here we turn to those aspects of GO's architecture which have to do with its status as a 'controlled vocabulary.' We can summarize our argument as follows:

1. with the development of modern formal disciplines (formal logic, and the computational disciplines which have arisen in its wake) we have learned a great deal about the criteria which must be satisfied if a language is to be structured in such a way that the information content expressed by its means can be extracted via automatic procedures that can support logical reasoning tools;

2. GO's controlled vocabulary has been developed in large part without concern for these criteria – this applies both to GO's terms, and also to the definitions associated therewith; accordingly GO has been used primarily in support of statistics-based methodologies oriented around string searches and pattern recognition, and much of its information content thereby remains unaccessed;

3. aspects of GO's current design, above all the expressive paucity which flows from the absence of relations between the terms of its three constituent ontologies, have led its curators to bend the rules of term-formation in order, in effect, to simulate such relations by constructing artificial terms within which corresponding relational expressions are embedded;

4. such artificial terms, however, correspond to no biological natural kinds – they are, precisely, artifacts of the Gene Ontology itself; and because they are constructed by stretching the rules for term-formation they create difficulties for human biologists when curating and applying GO in ways which often go hand in hand with characteristic types of coding errors.

GO terms[*]

The problem of expressive paucity created by GO's limited repertoire of relations between its terms is to some extent counteracted through the policy of constructing special terms which

[*] We use the term 'term', in what follows, to designate single nodes of the Gene Ontology and also GO synonyms. The entities to which GO terms refer we call 'classes,' and the individual objects, processes and functions in reality by which such classes are instantiated we call 'instances.'

simulate representations of the missing relations within the very terms themselves. This is achieved by means of special operators such as *with*, *within*, *without*, *in*, *site of*, *acting on*, or *resulting in*, in terms such as:

> electron transporter, transferring electrons **within** the noncyclic electron transport pathway of photosynthesis activity
>
> oxidoreductase activity, **acting on** diphenols and related substances as donors, oxygen as acceptor
>
> oxidoreductase activity, **acting on** paired donors, **with** oxidation of a pair of donors **resulting in** the reduction of molecular oxygen to two molecules of water

Some of these operators – for example *involved* and *involving* – were initially sometimes used to simulate the presence of *part_of* and similar relations crossing boundaries between distinct ontologies. Others – for example *during* – are used in order to simulate the presence of the machinery for representing temporal relations. Yet others – for example *within*, *site of* – are used to simulate spatial relations and to compensate for the fact that GO has no means of expressing the relation *is_located_at* – in spite of the importance precisely of cell locations to its general mission.

Such construction of special terms on the part of GO's authors and curators has thus far been uncontrolled. The result is that the operators in question are used in inconsistent ways. This in turn means that the information they express remains opaque to software tools.

Consider GO's uses of *involved in* as for example in:

1. hydrolase activity, acting on acid anhydrides, **involved in** cellular and subcellular movement *is_a* hydrolase activity, acting on acid anhydrides
2. asymmetric protein localization **involved in** cell fate commitment *is_a* cell fate commitment
3. cell-cell signaling **involved in** cell fate commitment *is_a* cell fate commitment
4. protein secretion **involved in** cell fate commitment *synonym of* protein secretion

Assertion 1. is correct to the degree that there are indeed two subtypes of *hydrolase activity, acting on anhydrides*: those which are and those which are not involved in cellular and subcellular movement. The term at issue, however:

> hydrolase activity, acting on acid anhydrides, involved in cellular and subcellular movement hydrolase activity,

which was taken over by GO from the Enzyme Commission, has been declared obsolete. This is because it is a function term which contains reference also to biological processes (*cellular and subcellular movement*), precisely contravening the principle which disallows links between GO's three constituent ontologies.

The relations at issue in 2. and 3. are erroneously classified as *is_a* relations, since inspection reveals that we have to deal here rather with relations of *part_of*. Thus the instances of asymmetric protein localization which are involved in instances of cell fate commitment in fact form parts of the corresponding instances of cell fate commitment. (The problem with 2. and 3. can be seen by pointing out that the assertions in question have the same form as: breathing involved in running *is_a* running.)

4. equates the class of instances of protein secretion which are involved in instances of cell fate commitment with the class of instances of protein secretion. This, again, is an example of erroneous coding, since there are also instances of protein secretion which are not involved in cell fate commitment, which flows in part from GO's idiosyncratic understanding of 'synonym' (http://www.geneontology.org/GO.synonyms.html).

Similar problems arise, too, in connection with the expression '*site of*', another example of an operator that is used by GO in its efforts to compensate for the expressive paucity of its repertoire of relations via the construction of artificial terms. Use of '*site of*' effectively converts the relation *is_located_at* into an *is_a* relation between specially constructed component terms. But this, too, proves to be a source of errors – reinforcing our general point that to bend the rules of term-formation involves paying a price of unsure coding on the part of those who are then left with no clear rules to follow.

Thus, as is shown in [7], from

> bud tip *is_a* site of polarized growth (*sensu* Saccharomyces).

and

> site of polarized growth (*sensu* Saccharomyces) *is-a* site of polarized growth (*sensu* Fungi),

we can infer logically either (a) that every instance of non-Saccharomyces Fungus polarized growth is co-localized with an instance of Saccharomyces polarized growth or (b) that there is Fungus polarized growth only in Saccharomyces. (a) we take to be biologically false; (b), however, implies that the terms 'site of polarized growth (*sensu* Saccharomyces)' and 'site of polarized growth (*sensu* Fungi)' in fact refer, confusingly, to the same class, and thus that the latter should be removed from GO's cellular component ontology. The lesson to be learned from this example is that there is a logic to which one becomes committed when using terms like 'site of' – a logic which may stand in conflict with the logic to which one becomes committed when one uses '*is_a*' and '*part_of*' to form assertions by combining terms.

Problems with '*Sensu*'

Part of the problem just referred to derives from GO's use of what is perhaps the most important operator used by GO in the formation of new terms, namely *sensu*. *Sensu* terms are introduced to cope with those cases where a word or phrase has different meanings when applied to different organisms, as for example in the case of *cell wall*. (Cell walls in bacteria and in fungi have a completely different composition.)

Since it is a primary goal of the GO Consortium to provide an ontology of gene products applying to all species, GO insists that *sensu* terms be introduced sparingly. In consequence, such terms often have non-*sensu* terms as children, as in:

R7 differentiation *is_a* eye photoreceptor differentiation (*sensu* Drosophila).

GO's interpretation of *is_a* sanctions the inference from *A is_a B* to: every instance of *A* is an instance of *B*. If this is correct, however, then this statement carries the implication that R7 differentiation occurs only in Drosophila, which seems to stand in conflict with the fact that such differentiation occurs also for example in crustaceans. Analogous problems involving *sensu* and non-*sensu* terms arise also in connection with GO's *part_of* relation. Thus we have

larval fat body development *part_of* larval development (*sensu* Insecta)

which seems to tell us that every instance of larval fat body development occurs in insects, which ignores for example the presence of fat bodies in crustaceans and worms.

GO has responsed to these concerns by pointing to special features of its reading of '*sensu*':

> by adding *sensu* the idea was not to exclude certain taxa from using a *sensu* term, but rather to give a user an idea of what sense a term should be used in. For example, if another flying insect were to be annotated to GO, we would hope that the '*sensu Drosophila*' terms could be used for this new species.
>
> An example where you might want to annotate a gene product from a taxon outside that specified in the *sensu* designation is 'fruiting body formation (*sensu* Dictyosteliida)'. If you were annotating a gene from the taxon Myxogastria (the true slime moulds, Dictyosteliida are the cellular slime moulds) you would still use this term because the process in both taxa is identical. – Jane Lomax (personal communication):

Note that

larval fat body development *part_of* larval development (*sensu* Insecta)

is an example of a *sensu* term which has a non-*sensu* term as child. Such child-parent relations might at first seem counter-intuitive, given that the purpose of '*sensu*' is precisely to allow a non-*sensu* term to be modified in such a way that it can refer to entities marked by special features which precisely do not arise in the entities referred to by the term in its original form. Closer inspection reveals, however, that there may be disadvantages to including the *sensu* designation in all children of *sensu* terms. Thus the term 'cell wall (*sensu* Fungi)' has the *part_of* child 'hyphal cell wall'. Because hyphae are only ever found in fungi it would then be confusing to add the *sensu* qualifier to the term 'hyphal cell wall' since this would suggest precisely that there were hyphal cell walls of other, non-fungal types. Against this, however, it is to be pointed out that the current rule, whereby the '*sensu* X' operator can be applied even to terms relating to taxa disjoint from the taxon X creates one more barrier to the automatic retrieval of information. This is because a term like 'Y (*sensu* X)' identifies only indirectly what features are shared in common by all the Ys at issue – in a way that requires the intervention of a human biologist with the relevant specialist knowledge. A better solution, therefore, would be to replace such terms with terms of the form 'Y which is Z', where 'Z' would then contain in explicit form the relevant positive information about the peculiar features at issue, rather than providing this information in coded form via a (somewhat indeterminate) linkage to a taxon.

Problems with Syntactic Operators

GO also employs a series of purely syntactic operators, such as ‘,’, ‘/’, and ‘:’, in ways which seem to contravene the underlying idea of a controlled vocabulary. Many of the terms involving ‘,’ (for example *1,4 lactonase activity*) are standard IUPAC names. Others, however, are problematic. Does the comma in *hydrolase activity, acting on acid anhydrides* mean ‘while’ or ‘of the type which is’? Here the definition helps to resolve the issue in favor of the latter, though again: the information contained in GO’s definitions is not formulated in such a way as to be accessible to software tools.

Problems arise also with GO’s inconsistent use of ‘/’. In some cases GO’s ‘/’ means ‘and’, for example in GO:0005954 *calcium/calmodulin-dependent protein kinase complex*. In others it means ‘or’, as in GO:0001539 *ciliary/flagellar motility*. In yet other cases it means ‘and/or’, as in GO:0045798 *negative regulation of chromatin assembly/disassembly*. In GO:0008608 *microtubule/kinetochore interaction*, it means ‘between’. In GO:0000082 *G1/S transition of mitotic cell cycle* ‘/’ it means ‘from … to …’. And in GO:0001559 *interpretation of nuclear/cytoplasmic to regulate cell growth* it means ‘with respect to’. It may be that human biologists find no difficulty in keeping control over these and a range of other different meanings of a single piece of syntax. What is certain, however, is that the information that is currently coded by means of such operators is to a large degree masked to automatic tools for information extraction, and we are thus gratified to see that reforms are currently under way by virtue of which the treatment of syntactical and other operators will be standardized through the imposition of a set of rules governing the use of these operators in different ontologies.

In Table 1 we provide a list (which complements the discussion in [11]) of the more important syntactic operators in GO, in order to give some idea of the scale of the problems at issue – problems which are currently being addressed by the GO consortium under the auspices of its OBOL project (see Mungall, C. *et al*. The OBOL Ontology Language, unpublished). In the left-hand column are the terms or syntactic operators which contribute to the compositional character of GO – they are, as it were, the standard sorts of linking expressions in terms of which complex terms are built up out of simpler parts. Examples in the next column are selected to illustrate how these linking expressions are characteristically used. The remaining columns give information as to the number of uses of the expressions in question in GO's three ontologies.

Conclusion

As the GO consortium has recognized (Mungall, *op. cit.*), many of the problems connected with GO's departure from compositionality can be resolved by preparing a canonical list of admissible operators and providing strict usage rules for each. The terms involving such operators currently receive a significantly lower number of annotations than do other terms in GO. This, we believe, provides some indication that the meanings conveyed by the terms in question are not only inaccessible to software tools but that they pose difficulties to understanding also on the part of human biologists. The examples here treated thus suggest a more general lesson as concerns the development and curation of systems like GO in the future: that terminologies are likely to be less susceptible to error and also more susceptible to integration with other terminologies if they are subjected to robust principles for handling syntax and for formulating terms and definitions.

Operator	**Examples**	**Component ontology**	**Function ontology**	**Process ontology**
with	GO: 0010483 conjugation **with** cellular fusion	1	17	36
within	GO: 0045153 electron transporter, transferring electrons **within** CoQH2-cytochrome c reductase complex activity	1	5	8
without	GO: 0000748 conjugation **without** cellular fusion	0	0	10
from	GO: 0019285 betaine biosynthesis **from** choline	0	2	139
during	GO: 0042074 cell migration **during** gastrulation	0	0	73
and	GO: 0016743 carboxyl**- and** carbamoyltransferase activity	2	42	136
in	GO: 000014 G1-specific transcription **in** mitotic cell cycle	0	18	32
acting on	GO: 0016684 oxidoreductase activity, **acting on** peroxide as acceptor	0	145	1

resulting in	GO: 0000077 DNA damage response, signal transduction **resulting in** cell cycle arrest	0	1	7

regulator;
regulatory;
regulated;
regulation
	GO: 0042754 negative **regulation of** circadian rhythm
	GO:0045055 **regulated** secretory pathway
		0	66	1260

dependent	GO: 0004692 cGMP-**dependent** protein kinase activity	14	90	78
constituent, constitutive	GO: 0030280 structural **constituent of** epidermis	0	28	1
response	GO: 0000751 cell cycle arrest in **response** to pheromone	0	5	261
sensu	GO: 0000143 actin cap (***sensu*** Saccharomyces)	140	14	315
site of	GO: 0016366 site of polarized growth	3	0	1
complex	GO: 0015667 proteasome activator complex	518	11	34
: (colon)	GO: 0015296 anion**:**cation symporter activity	3	170	4
/ (slash)	GO: 0000871 pilin fibrilin exporter activity	12	112	162
, (comma)	GO: 0002279 cyclin-dependent protein kinase, intrinsic regulator activity	71	724	411

Acknowledgements: This paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the European Union Network of Excellence on Medical Informatics and Semantic Data Mining, and the Volkswagen Foundation under the auspices of the project "Forms of Life". We should like to thank also Bert R. E. Klagges, Jacob Köhler, Jane Lomax and Steffen Schulze-Kremer.

Literature

[1] Gene Ontology Consortium. Creating the Gene Ontology resource: Design and implementation. Genome Res. 2001; 11: 1425-1433.

[2] Grenon P, Smith B: SNAP and SPAN: Towards dynamic spatial ontology, Spatial

Cognition and Computation, 2004: 4; 69-103.

[3] Bada M, Turi D, McEntire R, Stevens, R: Using reasoning to guide annotation with Gene Ontology terms in GOAT. SIGMOD Record, 2004:33(2).

[4] Kumar A, Smith B, Borgelt C. Dependence Relationships between Gene Ontology Terms based on TIGR Gene Product Annotations. CompuTerm 2004: 3rd International Workshop on Computational Terminology (In Press)

[5] Kumar, A. and Smith, B. The Unified Medical Language System and the Gene Ontology: Some Critical Reflections, in A. Günter, R. Kruse and B. Neumann (eds.), KI 2003: Advances in Artificial Intelligence (Lecture Notes in Artificial Intelligence 2821), Berlin: Springer, 2003; 135–148.

[6] Smith, B., Williams, J., Schulze-Kremer, S.: The Ontology of the Gene Ontology. In: Proc. Annual Symposium of the American Medical Informatics Association. 2003;609-613.

[7] Smith, B. and Rosse, C.: The Role of Foundational Relations in Biomedical Ontology Alignment. Proc Medinfo 2004. (In press)

[8] Schulz, S., Hahn, U.: Mereotopological reasoning about parts and (w)holes in bio-ontologies. Formal Ontology and Information Systems (FOIS 2001); 210-221.

[9] Hahn U., Schulz S., Markó K.: Mereological semantics for bio-Ontologies. AAAI 2004: 257-262.

[10] Smith, B. Köhler J and Kumar A.: On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In: Proceedings of DILS 2004 (Data Integration in the Life Sciences), (Lecture Notes in Bioinformatics 2994), Berlin: Springer, 2004; 79-94.

[11] Ogren, P. V., Cohen, K. B., Acquaah-Mensah, G. K., Eberlein, J., Hunter, L. T.: The Compositional Structure of Gene Ontology Terms. Pacific Symposium on Biocomputing (PSB 2004). 9; 214-225.