

A top-level ontology of functions and its application in the Open Biomedical Ontologies

Patryk Burek^{1,3,†}, Robert Hoehndorf^{2,3,†}, Frank Loebe^{1,3,†}, Johann Visagie², Heinrich Herre^{1,3} and Janet Kelso^{2,*}

¹Department of Computer Science, Faculty of Mathematics and Computer Science, University of Leipzig, Augustusplatz 10–11, 04109 Leipzig, ²Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig and ³Research Group Ontologies in Medicine (Onto-Med), Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Härtelstraße 16–18, 04107 Leipzig

ABSTRACT

Motivation: A clear understanding of functions in biology is a key component in accurate modelling of molecular, cellular and organismal biology. Using the existing biomedical ontologies it has been impossible to capture the complexity of the community's knowledge about biological functions.

Results: We present here a top-level ontological framework for representing knowledge about biological functions. This framework lends greater accuracy, power and expressiveness to biomedical ontologies by providing a means to capture existing functional knowledge in a more formal manner. An initial major application of the ontology of functions is the provision of a principled way in which to curate functional knowledge and annotations in biomedical ontologies. Further potential applications include the facilitation of ontology interoperability and automated reasoning. A major advantage of the proposed implementation is that it is an extension to existing biomedical ontologies, and can be applied without substantial changes to these domain ontologies.

Availability: The Ontology of Functions (OF) can be downloaded in OWL format from <http://onto.eva.mpg.de/>. Additionally, a UML profile and supplementary information and guides for using the OF can be accessed from the same website.

Contact: bioonto@lists.informatik.uni-leipzig.de

1 INTRODUCTION

Ontologies play an increasingly important role in modern biology. Recent years have seen a significant expansion in the number of biomedical ontologies and controlled vocabularies. The Open Biomedical Ontologies (OBO)¹ project serves as an umbrella organization providing some basic criteria and guidelines for the standardization of biomedical ontologies.

The OBO project includes a large number of domain-specific ontologies such as the Gene Ontology (GO) (Ashburner *et al.*, 2000)—which provides information about processes, molecular functions and sub-cellular locations of genes and gene products—

and anatomical and developmental ontologies available for specific species.

Recently, several methodological approaches were discussed which aim to provide an ontological foundation for medical and biomedical domains by means of top-level ontologies (Heller and Herre, 2004b; Smith *et al.*, 2005). A top-level ontology explicitly provides domain-independent notions. According to the principles of ontological foundation as expounded in (Heller *et al.*, 2004; Heller and Herre, 2004b) and applied in (Herre and Heller, 2005), we pursue the idea of adding top-level layers to existing biomedical ontologies. These layers analyze and formalize general aspects of concepts occurring in these ontologies. The use of a top-level ontology potentially leads to fewer errors in the curation and creation of domain ontologies, a better understanding of the biological concepts and the means for data and ontology integration.

A number of top-level concepts used frequently in various OBO ontologies remain unanalyzed and undefined. Concepts like “role” (such as “oxygen accumulator”) or “function” (such as “to accumulate oxygen”) serve as examples of unanalyzed top-level categories in the OBO ontologies.

Nevertheless, the notion of function is widely used in biomedical ontologies. Most commonly, one of the three hierarchies in the GO is the molecular functions taxonomy. Although the GO provides a short definition for its notion of molecular function, an in-depth analysis is not provided. Further uses of the notion of function appear in the Chemical Entities of Biological Interest (ChEBI) Ontology (Brooksbank *et al.*, 2005) and in the Celltype (CL) Ontology (Bard *et al.*, 2005), equally without a strong theoretical basis concerning functions.

We believe that a theory of functions is useful for the development and application of biomedical ontologies. To date, criticisms of the use of the concept of function in biomedical ontologies either proposed no solution or implied extensive changes, or a complete refactoring of existing structures (Smith *et al.*, 2003). Considering the GO's molecular function taxonomy, for example, we realize that this poses problems for a resource under constant usage by the community. Therefore we propose to address this problem in another way.

*To whom correspondence should be addressed.

†These authors contributed equally to this work.

¹<http://obo.sourceforge.net>

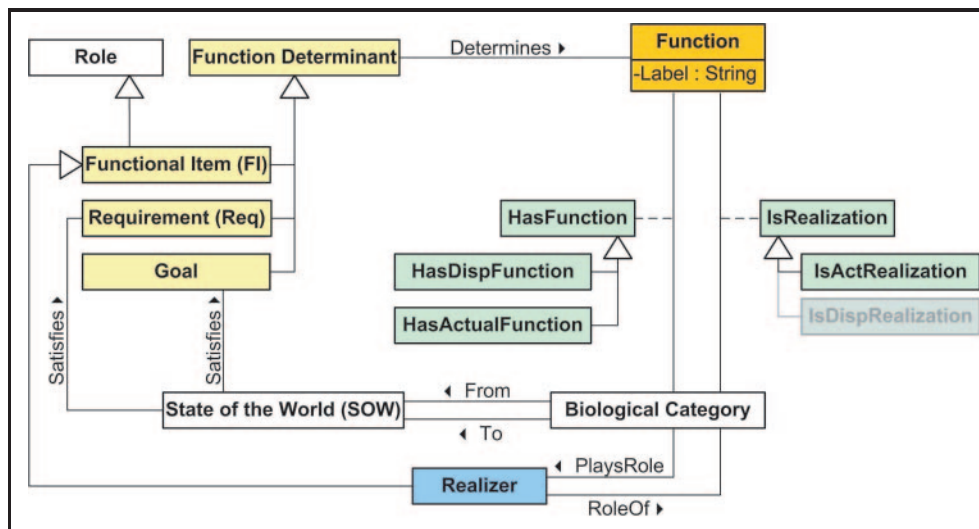


Fig. 1. A schematic representation of the concepts used and introduced by the OF (using the Unified Modeling Language (OMG, 2006)). Unlabelled relations indicate generalizations, where large arrowheads point at the more general concept. Functions (the orange box) are determined by entities indicated in yellow: a goal, requirements, and a functional item. A biological category may be related to a function in two ways (cf. the green boxes which provide labels for those relations connected to them by a dashed line): its instances may *realize* the function or they may *have* the function. A biological entity (such as a process) is a realization of a function if it mediates between two states of the world, one satisfying the requirements, the other satisfying the goal. A realizer in the OF, presented in blue, is the role played by an entity in a realization. In the function this role is determined by the functional item, hence realizer is generalized by functional item. Biological categories whose instances can play the role defined by the functional item have the function. The *HasFunction* relation relates biological categories with functions if every instance of this category has the actual or dispositional function.

We describe here a proposal for a top-level ontology of biological functions. This proposal introduces functions as an additional layer to the existing biomedical ontologies. We consider this ontology orthogonal to those currently in use. Although concepts relating to functions exist in biomedical ontologies, they are not yet adequately presented in an explicit, formal manner. Using our framework, this missing knowledge can be introduced in the existing biomedical ontologies while preserving their original structure.

For this purpose we introduce new relations such as *Has Function* and *IsRealization* in order to relate concepts of existing biomedical ontologies to functions as modelled in our approach. These relations and the specification of the structure of functions capture, in a separate ontology, information which is present at the stage of annotation.

We demonstrate the application of the proposed Ontology of Functions by showing how it can be used to systematically add explicit links between molecular functions and biological processes in the GO. We will further apply the formalism to the Celltype Ontology (CL), and will show how our proposal can serve to make definitions in CL precise, identify entities which are not yet covered by CL, and thereby contribute to CL's completeness.

Finally, we discuss advantages of our approach. In particular, we focus on the extent to which it may aid automated reasoning and data integration.

2 RESULTS

2.1 Introduction to the Ontology of Functions

We introduce here selected concepts of the *Ontology of Functions* (OF), which are presented in detail in (Burek, 2006). The OF will be

included as a module in the *General Formal Ontology* (GFO; cf. (Heller and Herre, 2004a; Herre *et al.*, 2006)), a top-level ontology developed and maintained by the research group Ontologies in Medicine (Onto-Med)². The OF aims to provide a domain-independent, conceptual framework for the representation of knowledge about functions. An overview of the main concepts and relations introduced by the OF is given in Figure 1.

In an adaption of (Searle, 1995; Sasajima *et al.*, 1995), we consider functions as the abstraction of biological processes or other entities towards a goal: when X has the function Y with the goal Z, then X is supposed to cause or otherwise bring about the state of the world Z, thus realizing Y.

For example, it may be the case that a red blood cell transports oxygen. But the statement that “the function of the red blood cell is to transport oxygen” adds a goal or purpose to this description: the red blood cell is *supposed* to transport oxygen – even if the red blood cell is in a condition where it cannot perform this transport, i.e., it is malfunctioning.

The OF addresses three major issues concerning functions:

- (1) *function structure*: how to represent and determine functions independently of their realizations
- (2) *realization*: the conditions under which a given entity realizes a function
- (3) *has-function relation*: the determination of the notion of an entity having a function

Two main assumptions underlie the OF: the separation of functional knowledge from non-functional and the top-level orientation.

²<http://www.onto-med.de>

Concerning the first, in the literature functional knowledge is often considered as providing information about *what* an entity does or what goal it serves, whereas non-functional knowledge describes the structure or behavior of entities and thus answers the question *how* an entity behaves, exists, or realizes functions (Iwasaki and Chandrasekaran, 1992; Rosenman and Gero, 1998). We consider these kinds of knowledge as highly independent, i.e., a function can be described independently of its realization and vice versa. Regarding the second aspect, we view the notions of function, realization, and the has-function relation as common to various domains. The OF therefore qualifies as a top-level ontology rather than a domain ontology.

These two aspects impact further on the application of the Ontology of Functions. The separation of functional and non-functional knowledge permits the application of the OF to domain ontologies without significant changes to them. The top-level orientation, on the other hand, allows for the reuse of the OF across various domains.

2.2 The structure of functions

The OF provides a formal way to represent functions independently of their realizations. The corresponding representation scheme is called a *function structure*. It consists of a set of *labels*, a set of *requirements*, a *goal*, and a *functional item*. Except for the labels, these form the *function determinants*.

Labels are natural language expressions which name the function. Most commonly, phrases of the form “to do something” serve as labels, e.g. “to transport oxygen”.

The *requirements* of the function contain all the necessary pre-conditions which must be met whenever the function is to be realized. For example, in case of the function “to transport oxygen from A to B”, oxygen must exist at location A.

Functions are *goal-oriented* entities—specifying a function requires providing the goal it serves. The goal of the function is a state or part of the world—temporally extended or not—which is intended to be achieved by any realization of the function. In the case of transporting oxygen, the location of the oxygen at B is the goal.

The goal specifies only the part of the world directly affected by the function. Often a goal is embedded in a wider context, called *final state*. A final state of a function contains the goal plus an environment for the goal, therefore making the goal more comprehensible.

Functions are dependent entities, in the sense that a function is always the function of some other entity. The *functional item* is a role played by this entity in any realization of the function. In the case of “to transport oxygen”, it would be an oxygen transporter. The notion of roles is required to explain the nature of a functional item more comprehensively.

We adopt the theory of roles developed in (Loebe, 2003, 2005) and incorporated into the GFO. Accordingly, roles are entities played by a *role player* in a *role context*. For example, “oxygen transporter” refers to a role in the role context of “oxygen transport”, and this role may be played by a red blood cell. This example can further be used to illustrate the dependence relationships of roles. Generally, roles and their role contexts are mutually dependent, i.e., one cannot exist without the other. In contrast, the dependence of roles on their players is one-sided because the player could

exist without playing a particular role. In our example, an oxygen transport necessarily involves a oxygen transporter and vice versa. A red blood cell may or may not transport oxygen, thus be playing or not playing the role of oxygen transporter. If it does not play that role, the cell still remains a red blood cell due to other properties such as its histology.

In OF, functional items are special roles which appear in the realization of functions. Note that usually there are more roles involved in the realization of a function than a single role given by the functional item. In a transport process, for example, in addition to the role “transporter” there is a “cargo” role, referring to that which is transported. Hence, the functional item singles out a particular role whose player is the entity *having* the function.

2.3 Realization and realizer

After introducing the structure of functions, their realization forms the second issue addressed in OF. The notion of realization refers to the question of how the goal of the function is to be accomplished. The *realization* is an entity which provides a transition from the state of the world in which the requirements of the function are fulfilled, to the state in which the goal of the function is fulfilled. This will usually be a process such as an—observed or measured—oxygen transport, but could be another kind of entity such as an instantaneous change.³

It is important to understand the difference between a function and a realization, in particular regarding their specification: to specify a function and its structure one has to state *what* is to be achieved; representing a realization usually means to specify *how* something is achieved.

Apart from individuals, it is even more relevant for biomedical ontologies to relate categories directly, such as the process category “transport” to the function “to transport”. The relation *IsRealization* is introduced for this purpose. If a process category stands in the *IsRealization* relation to some function, then all instances of this category are realizations of the function. For example, the category “oxygen transport” (a process) stands in the *IsRealization* relation to the function “to transport oxygen”.

Next, there is a counterpart of functional items on the level of realizations. A functional item is defined as a special role in every realization of a function. It is, in every case, a category (similarly to roles as discussed in (Guarino and Welty, 2000)). In the example of “to transport oxygen”, the role “oxygen transporter” is the functional item. Now consider an individual transport process, i.e., a realization, involving a single red blood cell. That cell has the role “oxygen transporter” within this realization. This fact gives rise to a new entity which mediates between the realization and the cell itself, namely the cell as an “oxygen transporter” (cell-qua-oxygen transporter). Such an entity is called the *realizer* of the function. Moving to the terminology of roles, we consider realizers to be *qu-individuals*, i.e., instances of roles (Masolo et al., 2005, 2004; Loebe, 2005).

³The full framework of OF distinguishes two types of realizations, actual and dispositional. Realizations as introduced here would be called “actual” in OF. Dispositional realizations are structurally similar to actual realizations in that they instantiate the same category. For a full discussion, see (Burek, 2006).

In summary, a realization corresponds to a function as a whole, whereas a realizer corresponds to the functional item of that function. The realizer is a qua-individual played by the entity which *has the function*. This leads us to the third major concept of the OF, the *HasFunction* relation.

2.4 Has-function

We address here the question under which conditions a function can be ascribed to an entity. In order to represent function ascription, a ternary relation *has-function* is introduced. This relation takes an individual, a function and a context as arguments. The connection between the first two arguments is such that the individual is involved in a realization of the function as the realizer (e.g., the red blood cell in an oxygen transport process realizing “to transport oxygen”).

The context argument reflects the intuition that a function is always ascribed in some context. That means, an individual does not necessarily have a given function in all contexts. For example, a hammer on a pile of papers on a desk may have the function of holding paper, while in the context involving a nail and a wall the function is different. It is out of scope of this paper to investigate the nature of contexts (McCarthy and Buvać, 1998; Akman and Surav, 1996) and we will not include it in this proposal but rather use the has-function relation as if it were a binary relation. However, the background theory surrounding the OF (Burek, 2006) allows for the use of a context argument in the function ascription.

The has-function relation appears in two versions: *actual has-function* and *dispositional has-function*. An entity has an actual function, if it is the role player of the realizer in a realization of the function. If an individual red blood cell is currently transporting oxygen, it has an actual function. If that red blood cell is not transporting oxygen, yet is structurally similar to red blood cells which have that function (by means of being an instance of the same category “red blood cell”), the non-transporting blood cell is said to have the dispositional function “to transport oxygen”.

Further, a relation between categories is derived from the has-function relation. A category stands in the *HasFunction* relation to a function, if every instance of the category has that function, actually or dispositionally. For example, “red blood cell” is in the *HasFunction* relation to the function “to transport oxygen”.

Having dealt with the three major issues in the OF—function (structure), realization, and function bearers—let us briefly return to the notion of a realizer, which is considered as a qua-individual. Entities of this kind are not present in the current biomedical ontologies, but they are required in order to link entities which can have functions to realizations. In order to remain consistent with already existing categories of biomedical ontologies we introduce a ternary relation among categories. $Realizes(E, F, R)$ represents the fact that entities of the category E can play the role of the realizer of the function F in realizations of type R . For instance, $Realizes(\text{“red blood cell”}, \text{“to transport oxygen”}, \text{“oxygen transport”})$ means that, intuitively speaking, red blood cells can realize the function “to transport oxygen” in an “oxygen transport” process.

The introduction of a ternary relation—*Realizes*—offers the highest degree of coherence and precision. $Realizes(E, F, R)$ entails $IsRealization(R, F)$ as well as $HasFunction(E, F)$, while one cannot conclude $Realizes(E, F, R)$ from $IsRealization(R, F)$, $HasFunction(E, F)$, and the fact that E can participate in R . To see why this is the case, consider the general function “to transport” (F). Red

blood cells (E) can be said to have this function if we think of an oxygen transport. However, consider a process in which red blood cells are transported, e.g. in the context of some experiment. This transport R is a realization of the function and red blood cells are involved in it. However, here they play the role of the cargo rather than that of the transporter. Accordingly, $Realizes(E, F, R)$ does not hold in this context and cannot be inferred, even given all other facts.

2.5 Relations between functions

Based on the framework developed in (Burek, 2006) we can introduce relations between functions. Some of the relations introduced are common ontological relations such as subsumption, instantiation, or the part-of relation. For example, the subsumption of functions is defined in terms of the subsumption between the appropriate function determinants.

We can also define new relations which are characteristic only for functions:

- *Support* – one function supports the other if its goal fulfills *partially* the second function’s requirements (the goal of the first function is a proper part of the requirements of the second function).
- *Enable* – one function enables the other if its goal fulfills *all* of the second function’s requirements (the requirements of the second function are a part of the goal of the first function).
- *Prevent* – one function prevents the other if its goal excludes the requirements of the second.

In (Burek, 2006), more relations between functions are defined, which affect the realizations of functions. For example, one function may *trigger* or *improve* the realization of other functions.

2.6 Application to OBO’s ontologies

We explore here potential applications of the Ontology of Functions, and investigate when and where it may be beneficial to use its framework.

2.6.1 Identifying links between processes and functions Our first application is the identification and explanation of relations between processes and functions. The Gene Ontology (Ashburner *et al.*, 2000) provides a prime example in this respect. There has been some controversy and discussion about whether the “Molecular Function” taxonomy of the Gene Ontology describes functions or activities, and how functions are related to processes (Smith *et al.*, 2003). To our knowledge, no practical or theoretical solution has yet been proposed. Functions and activities are usually considered different entities, and actions or activities may realize certain functions. Therefore, while the function of an enzyme may be “to catalyze” a reaction, the activity performed by the enzyme is the catalysis itself, which may be embedded in another process.

We assume that at least parts of the Molecular Function taxonomy refer to genuine functions in the sense of the OF, and the annotation relation for some of the gene products annotated to these terms corresponds to the *HasFunction* relation.

A general example is GO:0005215 (transporter activity), which we understand as referring to the function “to transport”. A more specific example is GO:0051119 (sugar transporter activity), which can be understood as the function “to transport sugar”.

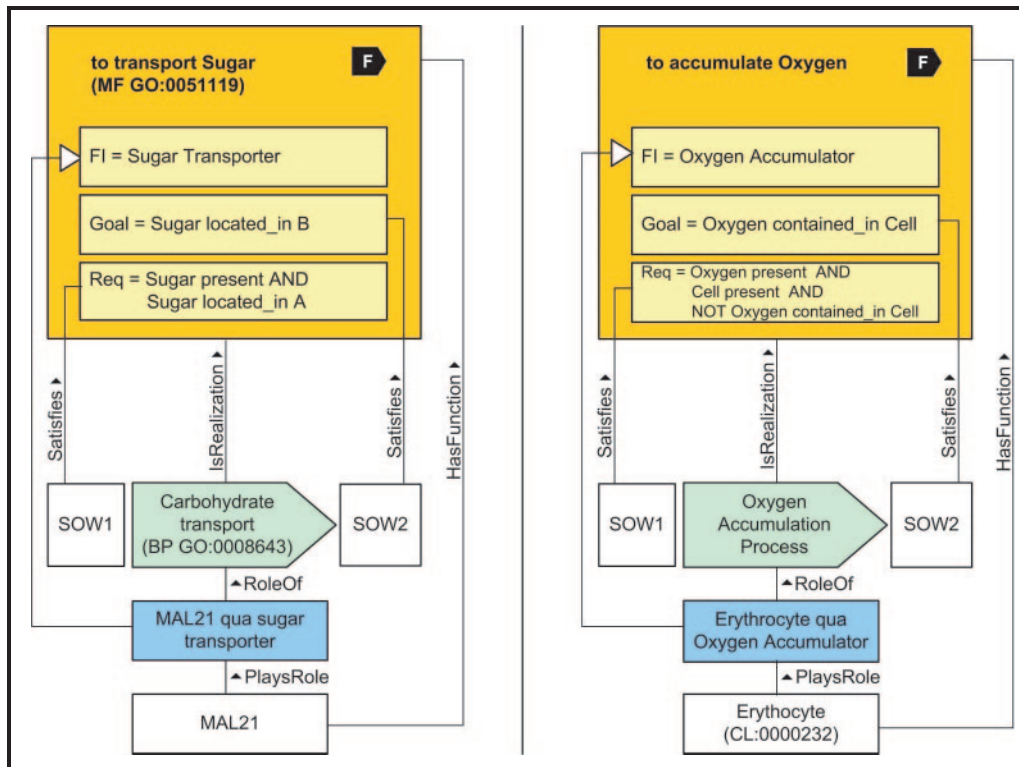


Fig. 2. Two exemplary models employing OF, instantiating the general model in Figure 1 (correspondences indicated by the coloring). On the left-hand side, a schematic version of the function “to transport sugar” together with its realization is shown. Processes of type “carbohydrate transport” realize this function, and an entity, in this case MAL21, has the function “to transport sugar”. Whenever applicable, the identifiers from the GO are used (for the function and process). MAL21 is currently annotated to the function and the process in the GO. In this model, the annotation relation is replaced by the *HasFunction* relation. On the right-hand side, the function “to accumulate oxygen” is modelled. This is a function taken from the Celltype Ontology. Except for erythrocyte, the entities involved in this model are not present in any of the OBO ontologies.

So let us investigate how the function “to transport sugar” can be modelled in the framework of the OF.

- As requirements, we assume that a sugar-molecule (CHEBI:25407 or CHEBI:25679) is located at some location.
- The goal is the location of the sugar molecule at a different location.
- The functional item is a role which we call “sugar transporter”.

We find that many of the gene products annotated with the “sugar transporter activity” in GO’s Molecular Function taxonomy are also annotated with some sub-category of the “transport” (GO:0006810) or “carbohydrate transport” (GO:0008643) categories in GO’s Biological Process taxonomy.

Also the names of the categories indicate a link, and of course there is an obvious one: gene products which have the function “to transport” may participate in a “transport” process. With the help of OF, we can make explicit some links between categories in GO’s Molecular Function and Biological Process taxonomies: Processes of type “carbohydrate transport” (GO:0008643) are realizations of the function “to transport sugar”; many of the gene products annotated with either carbohydrate transport or sugar transporter activity, such as MAL21 (maltose permease), can stand in the *HasFunction* relation to “to transport sugar”; new categories appear, namely gene products acting as (or “qua”) transporter, e.g. MAL21 *qua* transporter.

The left-hand side of Figure 2 demonstrates the full interconnections of this example by means of OF. In terms of the relations we introduced this is captured by *Realizes*(MAL21, GO:0051119, GO:0008643). What could be directly added to the GO are links of *IsRealization* and *HasFunction*: *IsRealization*(GO:0008643, GO:0051119) and *HasFunction*(MAL21, GO:0051119).

However, considering the GO’s definition of “sugar transporter activity” Enables the directed movement of a sugar into, out of, within or between cells. A sugar is any member of a class of sweet, water-soluble, crystallizable carbohydrates, which are the monosaccharides and smaller oligosaccharides.

It is possible to interpret this function differently: as the function “to enable F” or “to support F”, where F is the function “to transport sugar”.

Now the function “to support F” with F being “to transport sugar” would simply be a function where the goal of “to support F” would be part of the requirements of “to transport sugar”. So every realization of “to support F” would be a transition from a state of the world where some of the requirements for “to transport sugar” (the presence of a sugar molecule or its location) are not satisfied to a state where they are satisfied.

Many more relations between functions can be modelled and may be relevant in GO, such as “to trigger” or “to prevent”. Separating

these functions, which is made possible using OF, could lead to more accurate and comprehensive definitions.

2.6.2 Identifying implicit functions and processes The Ontology of Functions can be applied to existing taxonomies in order to make explicit functions and processes which are currently implied but not separately defined.

This kind of use of the concept of function occurs in the Celltype Ontology (Bard *et al.*, 2005) (CL) and the Ontology of Chemical Entities of Biological Interest (Brooksbank *et al.*, 2005) (ChEBI). We will only explore the Celltype Ontology, but the same argument can be applied to ChEBI.

CL uses the term function in the subtree “cell by function” which classifies cell types by the functions which they perform. A general example is “stuff accumulating cell” (CL:0000325), and more specifically “oxygen accumulating cell” (CL:0000329), of which a red blood cell or erythrocyte (CL:0000232) is a subcategory. The function “to accumulate oxygen (by a cell)” would be modelled as shown in the right-hand side of Figure 2:

- The presence of oxygen (ChEBI:25805) outside of a cell (CL:0000000) is the requirement of the function.
- The goal of the function is the cell’s accumulation of oxygen: The oxygen is contained in the cell.
- The functional item is called “oxygen accumulator”.

The subsumption of erythrocyte under oxygen accumulating cell in CL reflects the fact that erythrocytes have the function “to accumulate oxygen”, *HasFunction*(CL:0000232, “to accumulate oxygen”). Further, they may *act as* oxygen accumulators, a new category for CL, in the process of “oxygen accumulation”, *IsRealization* (“oxygen accumulation”, “to accumulate oxygen”). Again, the *Realizes* relation captures all these new relations appropriately: *Realizes*(CL:0000232, “to accumulate oxygen”, “oxygen accumulation”).

The analysis of erythrocyte in CL has led to the discovery of entities which are not yet part of CL or any other OBO ontology, but which contribute to the understanding of interactions among ontologies in cellular biology. Additionally, we can now define “oxygen accumulating cell” as a cell which has the function “to accumulate oxygen”.

3 DISCUSSION

3.1 Adding information systematically

The framework developed here and fully described in (Burek, 2006) can be used to provide additional information for existing biomedical ontologies such as the Gene Ontology (GO), without the need for modification of the existing structure of these ontologies. In general, we provide a *methodology* for defining functions and relating them to various other entities, such as processes, roles and even genes and gene products. This methodology may benefit the annotation and curation process and lead to improved definitions and completeness.

The advantage of the Ontology of Functions (OF) is *enhanced expressivity*. For example, the curators of the GO when annotating a gene product with the appropriate terms from the GO will have the information available that a certain protein is involved in some process and how it is related to a certain molecular function. They may also have more information about the protein, for exam-

ple the conditions under which it operates and other requirements which need to be satisfied for the protein to be active. By means of the OF, this information can be made explicit, and will not be lost as is currently the case.

The OF further allows for a refinement or *replacement of the annotation relation* in a number of cases by means of the *Has Function* relation. Note that the latter is an ontological relation, in contrast to the annotation relation, which is currently an arbitrary association relation introduced to link genes and gene products to the concepts of an ontology. Refined annotations do not only provide more information within ontologies themselves, but also with respect to the relation between categories of biomedical ontologies and genomic knowledge about biological reality.

Both, additional information due to enhanced expressivity as well as refined annotations may prove useful for the various statistical methods which have been applied to biomedical ontologies in order to detect biological correlations, such as (Subramanian *et al.*, 2005; Beissbarth and Speed, 2004; Berriz *et al.*, 2003).

It is interesting to consider to what extent and how the addition of information to existing biomedical ontologies can be automated. At present, we do not have an implemented solution for this issue. However, we expect that approaches to finding associations between categories using lexical and statistical analysis like (Bodenreider *et al.*, 2005; Burgun *et al.*, 2004) can be exploited and combined with the OF, in order to add categories and relations between them automatically. These could further be verified by existing natural language processing techniques (Mungall, 2004).

However, the rich formalism of the OF introduces another kind of new information which is less likely to be added automatically: *roles* and *qua-individuals*, the instances of roles. These concepts have mostly been neglected in the bio-ontology community, but ontological research has dealt with roles for a long time and rich theories of roles exist (Guarino and Welty, 2000, 2004; Masolo *et al.*, 2004, 2005; Poli, 1998; Loebe, 2003, 2005). We believe that they can prove useful in the explanation of biological phenomena. Making them explicit in biomedical ontologies can therefore serve to complete the coverage of these ontologies and enhance their conceptual modelling capabilities.

However, ontological theories must be applied cautiously. For instance, the theory of roles as proposed in (Guarino and Welty, 2004) defines constraints on the subsumption relation. Applied to an example from the Celltype Ontology, the subsumption link between “red blood cell” and “circulating cell” violates that constraint, if “circulating cell” is understood as a role. In this case “circulating cell” would refer to the role played by a red blood cell in the actual process of circulation. We, on the other hand, analyze “circulating cell” as a cell which has the actual or dispositional function “to circulate”, which would not violate a subsumption constraint in (Guarino and Welty, 2004).. This said, we want to emphasize that the application of formal ontological theories to domain ontologies must be done cautiously, and preferably in collaboration with domain experts.

3.2 Automated reasoning

The relation of our proposal to *automated reasoning* is highly relevant in the context of biomedical ontologies. Automated reasoning on biological data has been a goal of the bioinformatics and the bio-ontology community for some time (Wroe *et al.*, 2003). We believe

that much benefit can be gained from automated reasoning if a rich set of axioms is provided. The Ontology of Functions is equipped with a rich axiomatization (see (Burek, 2006)), which can be—for reasons of efficiency—adapted to description logic and used in conjunction with an automated reasoner such as FaCT (Horrocks et al., 1999).

Therefore, the OF can be seen as a formal and unambiguous specification framework for biological functions whose consistency can be verified, and in which implicit knowledge can be deduced.

3.3 Related work

To our knowledge, the only approach which in its aim is strictly similar to our proposal is that of Karp (2000). This proposal, however, is limited to a molecular granularity. Biological functions on the cellular, organismal or population level of granularity are not included. Moreover, functions are explicitly not context-dependent, while in the OF the *has-function* relation can, in principle, be dependent on a context. Furthermore, (Karp, 2000) attempted to create an ontology of functions as a module for EcoCyc⁴. The OF, on the other hand, is a top-level ontology of functions, and is therefore domain-independent and general. However, the view which (Karp, 2000) takes on functions is compatible with the OF.

The Gene Ontology (Ashburner et al., 2000) also provides a definition for a molecular function:

Molecular function describes activities, such as catalytic or binding activities, at the molecular level. GO molecular function terms represent activities rather than the entities (molecules or complexes) that perform the actions, and do not specify where or when, or in what context, the action takes place.

However, this definition does not separate activities and functions, as is the case in the OF which distinguishes functions and their realizations. Adding this distinction allows the capture of more information in the GO, while retaining GO's current structure.

In philosophy and ontology, many theories about biological functions have been developed (Searle, 1995; Johansson, 2004; Johansson et al., 2005; Kumar and Smith; Millikan, 1987; Melander, 1997). However, while these discussions provide valuable theoretical insight, they do not provide an immediate practical solution to the problem of conceptual modelling of functions in biology. We tried to learn from these discussions and develop the means for modelling function.

Many attempts to integrate the taxonomies of the GO have been made (Hill et al., 2002; Kumar et al., 2004; Aranguren, 2004; Wroe et al., 2003; Aranguren, 2005). However, none of these are based on a thorough ontological analysis of functions and their relation to other relevant biological entities such as processes.

4 CONCLUSION

The Ontology of Functions provides a framework for representing arbitrary functional knowledge in every domain of biology. This framework is used to define and specify functions, and relate them to other entities in biology. This helps to prevent errors, to clarify definitions and to support the integration of biological data and knowledge. We have shown how to use the OF to represent the relation between biological processes and functions in the Gene

Ontology, for which no ontologically founded representation formalism is currently available.

The introduced formalism requires no changes to the existing structure of the Gene Ontology, and could therefore be adopted gradually. Moreover, we have demonstrated how to analyze the annotation relation in the OF. Based on such analyses, the relation between genes or gene products and categories to which they are annotated can be made more precise. We have further shown how the OF framework can be used to identify and define functions of cells or chemicals.

The OF is a top-level ontology of functions which will be extended by including biological domain concepts. Statistical methods or text mining methods such as (Bodenreider et al., 2005; Burgun et al., 2004) could be used to extract the skeleton of a functional domain ontology from the existing ontologies. The OF can also be used to support the construction of a biological core ontology, which is a top-level ontology for the domain of biology (cf. also (Rector et al., 2006)) for an initial proposal of such an ontology).

Moreover, we are working on an implementation of this framework in the form of an annotation and curation tool, which will effectively guide the annotation and curation process by implementing the methodology defined by the OF to represent functional knowledge. We plan to integrate an automated reasoner with this tool in order to assist in maintaining consistency and to enable automated deduction.

ACKNOWLEDGEMENTS

We thank Michael Lachmann, Katrin Loebe, and Kay Prüfer for helpful discussions and critical reading of the manuscript. We thank the Max Planck Society, the German Federal Ministry of Education and Research, the Institute of Medical Informatics, Statistics and Epidemiology, and the Graduiertenkolleg Knowledge Representation of the German Research Foundation for financial support.

REFERENCES

- V. Akman and M. Surav. Steps toward formalizing context. *AI Magazine*, 17(3):55–72, 1996.
- M. E. Aranguren. Improving the structure of the gene ontology. Master's thesis, University of Manchester, 2004.
- M. E. Aranguren. Ontology design patterns for the formalisation of biological ontologies. Master's thesis, University of Manchester, 2005.
- M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.
- J. Bard, S. Y. Rhee, and M. Ashburner. An ontology for cell types. *Genome Biol*, 6(2):R21, 2005.
- T. Beissbarth and T. P. Speed. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, 20(9):1464–1465, Jun 2004.
- G. F. Berriz, O. D. King, B. Bryant, C. Sander, and F. P. Roth. Characterizing gene sets with FuncAssociate. *Bioinformatics*, 19(18):2502–2504, Dec 2003.
- O. Bodenreider, M. Aubry, and A. Burgun. Non-lexical approaches to identifying associative relations in the gene ontology. *Pac Symp Biocomput*, pages 91–102, 2005.
- C. Brooksbank, G. Cameron, and J. Thornton. The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res*, 33(Database issue):D46–D53, Jan 2005.
- P. Burek. *Ontology of Functions*. PhD thesis, University of Leipzig, Institute of Informatics (IfI), 2006. forthcoming.

⁴<http://ecocyc.org/>

- A. Burgun, O. Bodenreider, M. Aubry, and J. Mosser. Dependence relations in gene ontology: A preliminary study. In *Proceedings of the Workshop on The Formal Architecture of the Gene Ontology*, 2004.
- N. Guarino and C. A. Welty. A formal ontology of properties. In *Lecture Notes in Computer Science*, volume 1937, pages 97–112, 2000.
- N. Guarino and C. A. Welty. An overview of OntoClean. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, chapter 8, pages 151–172. Springer, 2004.
- B. Heller and H. Herre. Ontological categories in GOL. *Axiomathes*, 14(1):57–76, 2004a.
- B. Heller and H. Herre. Ontological foundations of medical information systems. In H. Fujita and V. Gruhn, editors, *New Trends in Software Methodologies, Tools and Techniques: Proceedings of the Third SoMeT_W04*, volume 111 of *Frontiers in Artificial Intelligence and Applications*, pages 3–17. Amsterdam, 2004b. IOS Press.
- B. Heller, H. Herre, K. Lippoldt, and M. Loeffler. Standardized terminology for clinical trial protocols based on top-level ontological categories. In K. Kaiser, S. Miksch, and S. Tu, editors, *Computer-based Support for Clinical Guidelines and Protocols: Proceedings of the Symposium on Computerized Guidelines and Protocols*, volume 101 of *Studies in Health Technology and Informatics*, pages 46–60. IOS Press, 2004.
- H. Herre and B. Heller. Ontology of time and situoids in medical conceptual modeling. In S. Miksch, J. Hunter, and E. T. Keravnou, editors, *Proceedings of the 10th Conference on Artificial Intelligence in Medicine (AIME 05), Aberdeen, Scotland, Jul 23–27*, volume 3581 of *Lecture Notes in Computer Science*, pages 266–275. Berlin, 2005. Springer.
- H. Herre, B. Heller, P. Burek, R. Hoehndorf, F. Loebe, and H. Michalek. General Formal Ontology (GFO) – a foundational ontology integrating objects and processes. Technical report, University of Leipzig, 2006.
- D. P. Hill, J. A. Blake, J. E. Richardson, and M. Ringwald. Extension and integration of the gene ontology (GO): combining GO vocabularies with external vocabularies. *Genome Res*, 12(12):1982–91, 2002.
- I. Horrocks, U. Sattler, and S. Tobies. Practical reasoning for expressive description logics. In *Proceedings of LPAR'99*, LNCS, 1999.
- Y. Iwasaki and B. Chandrasekaran. Design verification through function- and behavior-oriented representations. In J. S. Gero and F. Sudweeks, editors, *Proceedings Artificial Intelligence in Design Conference*, pages 597–616, 1992.
- I. Johansson. Functions, function concepts, and scales. *The Monist*, 87(1):96–114, 2004.
- I. Johansson, B. Smith, K. Munn, N. Tsikolia, K. Elsner, D. Ernst, and D. Siebert. Functional anatomy: A taxonomic proposal. *Acta Biotheoretica*, 53:153–166, 2005.
- P. D. Karp. An ontology for biological function based on molecular interactions. *Bioinformatics*, 16(3):269–285, Mar 2000.
- A. Kumar and B. Smith. The ontology of processes and functions: A study of the international classification of functioning, disability and health. <http://ontology.buffalo.edu/medo/ICF.pdf>. Draft.
- A. Kumar, B. Smith, and C. Borgelt. Dependence relationships between gene ontology terms based on TIGR gene product annotations. In *Proc CompuTerm*, pages 31–38, 2004.
- F. Loebe. An analysis of roles: Towards ontology-based modelling. Master's thesis, University of Leipzig, Institute of Informatics (IfI), 2003.
- F. Loebe. Abstract vs. social roles: A refined top-level ontological analysis. In G. Boella, J. Odell, L. van der Torre, and H. Verhagen, editors, *Proceedings of the 2005 AAAI Fall Symposium Roles, an Interdisciplinary Perspective: Ontologies, Languages, and Multiagent Systems*, Nov 3–6, Arlington, Virginia, number FS-05–08 in Fall Symposium Series Technical Reports, pages 93–100. Menlo Park (California), 2005. AAAI Press.
- C. Masolo, G. Guizzardi, L. Vieu, E. Bottazzi, and R. Ferrario. Relational roles and qua-individuals. In G. Boella, J. Odell, L. van der Torre, and H. Verhagen, editors, *Proceedings of the 2005 AAAI Fall Symposium 'Roles, an Interdisciplinary Perspective: Ontologies, Languages, and Multiagent Systems*, Nov 3–6, Arlington, Virginia, number FS-05–08 in Fall Symposium Series Technical Reports, pages 103–112. Menlo Park (California), 2005. AAAI Press.
- C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino. Social roles and their descriptions. In D. Dubois, C. Welty, and M.-A. Williams, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004)*, Whistler, Canada, June 2–5, pages 267–277. Menlo Park, 2004. AAAI Press.
- J. McCarthy and S. Buvač. Formalizing context (expanded notes). In A. Aliseda, R. J. van Glabbeek, and D. Westerståhl, editors, *Computing Natural Language*, volume 81 of *CSLI Lecture Notes*, pages 13–50. Center for the Study of Language and Information (CSLI), Stanford University, Stanford, 1998.
- P. Melander. *Analyzing Functions. An Essay on a Fundamental Notion in Biology*. Almqvist and Wiksell, 1997.
- R. G. Millikan. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. The MIT Press, 1987.
- C. J. Mungall. Obol: integrating language and meaning in bio-ontologies. *Comp Funct Genomics*, 5:509–520, 2004.
- OMG. Unified Modeling Language: Infrastructure. Specification v2.0, Object Management Group (OMG), Needham (Massachusetts), Mar 2006. <http://www.omg.org/docs/formal/05-07-05.pdf>.
- R. Poli. Qua-theories. In L. Albertazzi, editor, *Shapes of Forms*, pages 245–256. Kluwer, 1998.
- A. Rector, R. Stevens, and J. Rogers. Simple bio upper ontology. <http://www.cs.man.ac.uk/~rector/ontologies/simple-top-bio/>, 2006.
- M. A. Rosenman and J. S. Gero. Purpose and function in design. *Design Studies*, 2:161–186, 1998.
- M. Sasajima, Y. Kitamura, M. Ikeda, and R. Mizoguchi. FBRL: A function and behavior representation language. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95), Montréal, Québec, Canada, Aug 20–25*, volume 2, pages 1830–1836. Morgan Kaufmann, 1995.
- J. R. Searle. *The Construction of Social Reality*. Penguin Group, 1995.
- B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biol*, 6(5):R46, 2005.
- B. Smith, J. Williams, and S. Schulze-Kremer. The ontology of the gene ontology. *AMIA Annu Symp Proc*, pages 609–613, 2003.
- A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 102(43): 15545–15550, Oct 2005.
- C. J. Wroe, R. Stevens, C. A. Goble, and M. Ashburner. A methodology to migrate the gene ontology to a description logic environment using DAML+OIL. *Pac Symp Biocomput*, pages 624–635, 2003.