

# The application of an ontology design pattern for functional abnormalities to phenotype ontologies and the extraction of an ontology of anatomical functions

Robert Hoehndorf   Axel-Cyrille Ngonga Ngomo   Janet Kelso  
leechuck@acm.org   ngonga@informatik.uni-leipzig.de   kelso@eva.mpg.de

## Abstract

Functions play an important role throughout biology. Although molecular functions are covered in the Gene Ontology, there is currently no publicly available ontology of anatomical functions. Ontological considerations on the nature of functional abnormalities and their representation in current phenotype ontologies show that we can automatically extract a skeleton for such an ontology of anatomical functions by using a combination of process, phenotype and anatomy ontologies. We provide an ontological analysis of the nature of functions and functional abnormalities. From this analysis, we derive an approach to the automatic extraction of anatomical functions from existing ontologies using a combination of natural language processing, graph-based analysis of the ontologies and formal inferences. Alternatively, we introduce a new relation to relate material objects to processes that realize the function of the object to avoid a needless duplication of processes already present in the Gene Ontology in a new ontology of anatomical functions. We discuss several limitations of the current ontologies that still need to be addressed to ensure a consistent and complete representation of anatomical functions and functional abnormalities.

## 1 Introduction

The notion of *function* is of high importance throughout biology. It is used to describe genes (Eilbeck et al., 2005), gene products (Ashburner et al., 2000), cell types (Bard et al., 2005) and anatomical structures (Rosse and Mejino, 2003). Functions are also used in the description of phenotypes of *functionings*, i.e., observable phenomena regarding the functioning or malfunctioning of biological entities. These phenotypes are of high importance in the discovery of

gene functions, in the description of abnormalities, diseases, signs and symptoms.

### 1.1 Phenotype ontologies

A phenotype is any observable characteristic of an organism, and may include both structural and behavioral properties. Phenotypes of *functionings* (or *functional phenotype*) are either observable characteristics of a process that realizes a function of an organism or a part of the organism, or properties of an organism that involve its functions (such as having a function or lacking a function). Phenotype ontologies for mouse and human phenotypes were developed to annotate research databases of mouse and human phenotypes. The Mammalian Phenotype Ontology (MPO), focusing on mutant mouse phenotypes (Smith et al., 2005), and the Human Phenotype Ontology (HPO), focusing on Mendelian diseases in man (Robinson et al., 2008), are increasingly being used to describe human and mouse phenotypes.

Both the HPO and MPO refer to established standard anatomy ontologies for mice and humans. The HPO uses the Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003) to refer to anatomical entities in humans, and the MPO uses the Adult Mouse Anatomy Ontology (MA) (Hayamizu et al., 2005). These anatomy ontologies describe anatomical entities using, among others, part-whole relations, i.e., they focus on the anatomical *structure*.

Although the phenotype ontologies describe both structurally and functionally abnormal phenotypes, the anatomy ontologies do not include an elaborate description of the anatomical functions. As a consequence, although the classification of structural abnormalities in the phenotype ontologies follows well-defined principles, the classification of phenotypes of functionings is often unprincipled and sometimes ambiguous.

To address the issue of representing function phenotypes, we provide an ontology design pattern for functional abnormalities. This design pattern is applicable in phenotype ontologies, especially in the MPO and HPO. We discuss the benefits of the application of the design pattern and relate the design pattern to the composite names of the categories in the phenotype ontologies. Based on the category names, we apply a pattern-based approach to extract a skeleton for an ontology of anatomical functions from a combination of the anatomy and phenotype ontologies together with the Biological Process Ontology of the Gene Ontology (Ashburner et al., 2000).

## 1.2 Biological function

There is an ongoing discussion in philosophy of biology and theoretical biology as to the exact nature of a biological function. While functions of artifacts come into being due to the intentions of a designer, biological entities have evolved over time, and biological functions are not dependent on intentions the same way as artifacts are.

Philosophical theories of biological functions range from reductions to causality over social accounts of functions to the denial of the existence of biological functions. The first two are of major importance, i.e., the causal view of biological functions and the social view of biological functions. The major proponents of causal explanations of functionality are (Wright, 1973) and (Millikan, 1988), while the social view is defended by (Searle, 1997).

(Wright, 1973) gives the following definition of function:

**Definition 1.** *The function of  $X$  is  $Z$  means*

1.  $X$  is there because it does  $Z$ ,
2.  $Z$  is a consequence (or result) of  $X$ 's being there.

In the definition,  $X$  is a category of structures and  $Z$  is a process category, and instances of  $X$  are involved in instances of  $Z$ . Wright's definition assumes that an entity has only one function. However, as discussed in (Wright, 1973), the definition can be restated for entities having multiple functions by replacing "the function of  $X$  is  $Z$ " with "a function of  $X$  is  $Z$ ".

In the social view, functions are ascribed to *brute facts* by a conscious observer (Searle, 1997). A detailed analysis is provided by (Hartmann, 1966) and is illustrated in Figure 1. Hartmann distinguishes three elements to the ascription of a function: the setting of a goal in the future, the planning of how to achieve the goal, resulting in a *structure* that is

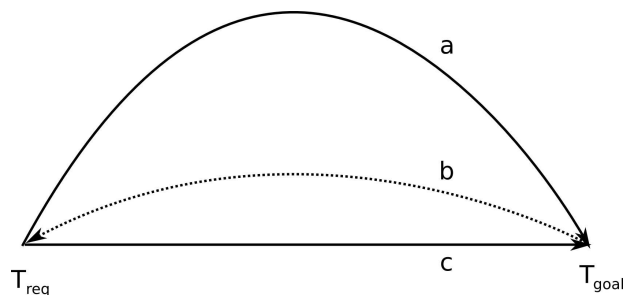


Figure 1: The figure shows the three conditions for the ascription of a *single* function to an entity. First, the goal of the function is established *in the future*. Second, the means for achieving the goal are selected or created. Finally, the goal can be realized by causal means, i.e., without the need for accessing or anticipating future states of the world.

capable of achieving the goal through causal means. Figure 1 shows how some entity obtains a single function. For an entity to have multiple functions, the same three steps are performed, yet the goal and the initial situation may change.

## Materials and methods

### Ontology of functions

For the current work, we do not choose a particular definition of biological function, and we do not add another definition to the literature. The method presented is compatible with most major views of *function*.

An analysis of how to represent the structure of a function has been provided by the Ontology of Functions (OF) (Burek et al., 2006; Burek, 2006). A function in the OF is described in terms of a requirement situation type, a goal situation type and a processual role. The requirement situation type serves as precondition for any function realization, the goal situation type is the postcondition, and the processual role (Loebe, 2007) is used to describe *how* a function bearer brings about the goal from the requirements. One major advantage of the treatment of functions in the OF is the explicit inclusion of preconditions for the function realizations, which serve to model the contexts in which a function can be realized.

### 1.3 Function realizations

Functions can be *realized* multiple times. Each realization of a function is a process, and in each realization of a function the function bearer achieves the goal of the function, starting at a situation satisfying the preconditions of the function.

While a function is an entity that is similar to a property in that it inheres in its bearer, a *functioning*

is a process that is a realization of a function. For example, while the function of the heart “to pump blood” is a property that the heart has in virtue of being a heart and in virtue of the evolutionary history of hearts. A *functioning* is the actual process of pumping blood which realized the function of the heart (Johansson, 2004). In particular, the function of the heart is “to pump blood” even when the heart is not *functioning*. This could be the case during a heart transplantation, or during a malfunctioning of the heart.

Furthermore, the function of the heart is “to pump blood” even when the heart cannot realize its function. Function realizations always require a *disposition* to realize the function in the function bearer, while the function itself can exist without such a disposition. For our present work, we use the simple conditional analysis of dispositions (Lewis, 1997) and use the term *disposition* interchangeably with *capability*: something  $x$  is disposed at time  $t$  to give response  $r$  to stimulus  $s$ , iff, if  $x$  were to undergo  $s$  at  $t$ ,  $x$  would  $r$ .

#### 1.4 Abnormal functionings

Abnormal functionings are processes which are similar to a functioning, but that are impaired in some way. Abnormal functionings are different from *malfunctioning* (Hoehndorf, 2009; Hoehndorf et al., 2009): in the case of a malfunctioning, the function bearer cannot cause the goal of its function although the preconditions for a function realization are given. An entity  $e$  has the property of being *malfunctioning* (with respect to the function  $f$ ), if  $e$  has a function  $f$ , but not a *disposition*  $d$  to realize the function  $f$ . Functions and dispositions are disjoint categories (i.e., neither is a subcategory of the other), yet they are related in a particular way (Hoehndorf et al., 2009).

While abnormal functionings are processes, *malfunctionings* are properties of the function bearer; in the case of a malfunctioning entity, no process of functioning occurs.

There are various kinds of abnormal functionings: functionings may be more or less effective, have unwanted side-effects or similar. We focus on the *malfunctioning* property here. A classification of kinds of abnormal functionings is out of the scope of this paper, and is subject to future work.

#### 1.5 Function and Structure

There is an important relationship between *function* and *structure*. At least biological functions are usually realized through causal processes (cf. (Hartmann, 1966) and Figure 1). In the case of biolog-

ical function bearers, the structure of the function bearer has developed through evolution to play a particular role in processes of a certain kind (e.g., the role of the heart as a *pump* in its function *to pump blood*). Therefore, if the heart – the function bearer – becomes *unable* to play this role in the function realization while everything else remains unchanged, this loss of capability is due to a change in the heart’s structure. In particular, the loss of a disposition or capability in the case of malfunctioning entities must go along with a change in the structure of the bearer of the disposition and function.

As a result, if  $e$  has the biological function  $f$ , and  $e$  is malfunctioning or functioning abnormally, then  $e$  must be abnormal, too. This pattern is already implied in the taxonomic backbones of the phenotype ontologies and reflected in the naming and the definitions of the phenotype ontologies’ categories.

#### 1.6 The functional abnormality pattern

The functional abnormality pattern is an ontology design pattern (Aranguren et al., 2008) for ontologies that classify both abnormal structural and functional phenotypes, such as both the Human and Mammalian Phenotype Ontology do.

According to the functional abnormality pattern, an abnormality of functioning (a property of a process) implies an abnormality of the function bearer. If multiple types of entities have the same kind of function, then an abnormality of the functioning implies a disjunction of the abnormalities of each possible kind of function bearer. A *malfunctioning* (a property of the function bearer), however, is a subcategory of (**is-a**) an abnormality of the function bearer, and if multiple types of entities have the same function, then a *malfunctioning* is a sub-category of a disjunction of the abnormalities of each possible kind of function bearer.

For example, an abnormality in *HearingP* processes<sup>1</sup> (which are functionings of the *HearingF* function) imply an abnormality of the ears, if the function of the ears is *HearingF*. If the function of both the left ear and the right ear was *HearingF*, then an abnormality of *HearingP* implies an abnormality of the left ear or an abnormality of the right ear. In this case, the category “abnormality of the left ear or abnormality of the right ear” should be named “ear abnormality” and defined as a disjunction of the two categories “abnormality of the left ear” and “abnormality of the right ear”, which are both subcategories of “ear abnormality”.

<sup>1</sup>We use *HearingF* to refer to the function, and *HearingP* to refer to the process realizing the function.

On the other hand, a *malfunctioning* of the ears with respect to their *HearingF* function is a property of the ears, and should be classified as a sub-category of *Ear abnormality*. A *malfunctioning* of the ears is defined as the absence of a dispositions which would normally be present (due to the ears' having a function whose realization requires the disposition), and the loss of a disposition entails a structural modification according to the theory of dispositions (Lewis, 1997). Therefore, a loss of a disposition is a special kind of structural change of the disposition's bearer.

### 1.7 Naming patterns in the phenotype ontologies

Our goal is to formally represent abnormal functionings and malfunctionings of anatomical functions. While there is no ontology of anatomical functions yet, such an anatomical function ontology is implied in the phenotype ontologies. These ontologies classify abnormal phenotypes, and in these phenotype ontologies, abnormal functionings are usually classified as a sub-category of abnormal structures which bear the function that is impaired. Therefore, the phenotype ontologies can serve as a seed for the construction of an ontology of anatomical functions.

However, as the phenotype ontologies rarely adequately define abnormal functionings formally, the challenge is to extract the information from the current structure, names and definitions of the phenotype ontologies. Such an approach will not suffice to create an exhaustive ontology of anatomical functions, because only few functions are addressed in the phenotype ontologies, nor will it provide a high-quality ontology that is suitable for use in applications. Instead, our goal is to extract functions that can be used as the backbone of an ontology of formally defined function categories after a manual review process.

The second major challenge in the extraction of anatomical functions is to provide an analysis and formal representation of the relations between anatomical functions, their bearers and the processes that realize the functions.

### 1.8 Formal representation of anatomical functions

In our formal analysis, we use the definition of the category *Deafness* in both the Mammalian and Human Phenotype Ontology as an example. The definition in the cross-products of both ontologies is the following statement in the OBO Flatfile Format (Golbreich and Horrocks, 2007):

```
[Term]
id: MP:0001967 ! deafness
```

```
intersection_of: PAT0:0000462 ! absent
intersection_of: inheres_in GO:0007605
```

In the OBO Flatfile Format, the definition of an ontological category is started with a [Term] statement, followed by a unique identifier of the category. Everything following an exclamation mark is considered to be a comment.

The GO category GO : 0007605 is named "sensory perception of sound" and has a synonym "hearing". The definition of *Deafness* in the two phenotype ontologies we use in our analysis claims that *Deafness* is a process of *HearingP* in which the quality *Absent* inheres. Inherence is a dependence relation between an instance of a quality and the bearer of the quality (Herre et al., 2006).

There are several problems with the analysis of *Deafness* in the phenotype ontologies. The first problem is that, according to the definition, *Deafness* is a process of *HearingP*. *Deafness* seems to be something different from a process, and certainly different from a *HearingP* process. An absence of hearing means that there is no *HearingP* process, no matter what properties such a process might have. In particular, *Absent* cannot inhere in an absent process, and, arguably, is not a quality at all.

The second problem is that there can be an absence of hearing without there being a case of *Deafness*. In a completely silent environment, both a human or a mouse will experience an absence of *HearingP* even when their *disposition* to hear is present. More precisely, according to the definition of *Deafness*, an absence of sound would also entail *Deafness*.

Therefore, to formally represent the phenotype *Deafness*, we are faced with two challenges: there is an absence of *HearingP* processes, and there is also an absence of the *disposition* to hear.

Using our ontological framework for representing malfunctionings, we can represent *Deafness* as a malfunctioning of the ears with respect to their *HearingF* function. However, a vital point is missing to apply our framework: an ontology of anatomical functions. The absence of such an ontology is one reason for the phenotype ontologies to model abnormal functionings using processes from the GO.

While the anatomical functions are not yet covered in an ontology, the processes that realize the anatomical functions are present in GO's Biological Process ontology. Therefore, we define a new relation that we call the **CC-has-function-realized-by** (*hfrb*) relation. This relation is based on the relations **CC-has-function** and **CC-realized-by**. The prefix *CC* indicates that the relation takes two ontological categories as argument. The relations between categories are defined using relations between

individuals (*II*-relations), following the pattern of defining *CC*-relations from the OBO Relationship Ontology (Smith et al., 2005) and (Hoehndorf et al., 2007). The definition of the relation **CC-has-function** is given in formula 1, where  $E$  denotes a category of *Presentials* (in GFO (Herre et al., 2006)), *Continuants* (in BFO (Grenon, 2003) or *Endurants* (in DOLCE (Masolo et al., 2003)):

$$\text{CC-has-function}(E, F) \iff \forall x(\text{instanceOf}(x, E) \rightarrow \exists y(\text{II-has-function}(x, y) \wedge \text{instanceOf}(y, F))) \quad (1)$$

According to this definition, the category  $E$  has the function (*CC*-relation)  $F$ , if and only if, for every instance  $x$  of  $E$  there is an instance  $y$  of  $F$  such that  $x$  has the function (*II*-relation)  $y$ .

While the relation **CC-has-function** follows the standard pattern for defining relations between categories (Golbreich and Horrocks, 2007), the relation **CC-realized-by** cannot follow the same pattern. Applying the same pattern would require that for every function, there is a process that realized the function. Yet not every function instance is realized, and according to our considerations above concerning malfunctionings, not every function *can* be realized. Therefore, we have to employ a different definition for the **CC-realized-by** relation, given in formula 2. In the formulation of the definition of the **CC-realized-by** relation, we consider that functions are not necessarily realized, but when they are realized, then always by processes of a certain kind<sup>2</sup>.

$$\text{CC-realized-by}(F, P) \iff \forall x(\text{instanceOf}(x, F) \wedge \exists y(\text{II-realized-by}(x, y) \rightarrow \text{instanceOf}(y, P))) \quad (2)$$

According to this definition, the function category  $F$  is realized by (*CC*-relation) the process category  $P$ , if and only if, whenever an instance  $x$  of  $F$  is realized by some  $y$ , then  $y$  is an instance of  $P$ .

With these definitions of the two relations **CC-has-function** and **CC-realized-by**, we can give a definition for the relation **CC-has-function-**

**realized-by**:

$$\text{CC-hfrb}(E, P) \iff \exists Z(\text{CC-hasFunction}(E, Z) \wedge \text{CC-realizedBy}(Z, P)) \quad (3)$$

This relation is a connection of the two previously defined relations with an implicit function as argument. The category  $E$  stands in the *CC-has-function-realized-by* relation to the category  $P$ , if and only if,  $E$  has the function  $Z$  and  $Z$  is realized by  $P$ .

The relation **CC-hfrb** is a relation between two categories. The relation can be defined in OWL2 as a connection between the two *CC*-relations using a property chain:

$$\text{CC-hfrb} = \text{CC-has-function} \circ \text{CC-realized-by} \quad (4)$$

Such a definition can be used in an OWL ontology in which ontological categories are in the domain of discourse (cf. (Hoehndorf et al., 2007; Herre et al., 2006; Hoehndorf et al., 2008)), i.e., in which there are OWL classes which have ontological categories as their instances.

A similar connection between the two relations **II-has-function** and **II-realized-by**, on the other hand, is very different from the relation between the categories: it is a relation between an entity with a function that is in fact (and currently) realized by a process:

$$\text{II-hfrb}(e, p) \iff \exists z(\text{II-hasFunction}(e, z) \wedge \text{II-realizedBy}(z, p)) \quad (5)$$

### 1.9 Application to anatomy and phenotype ontologies

We applied the framework for representing functional abnormalities to the automated extraction of anatomical functions from the HPO and MPO. For this purpose, we exploit the naming of the categories in the phenotype ontologies.

We make use of three types of ontologies in our approach:

1. the phenotype ontology that contains abnormal functional phenotypes, either the HPO or the MPO,
2. an anatomy ontology that contains the structures affected by the malfunctionings represented in the phenotype ontology, either the Adult Mouse Anatomy Ontology (Smith et al., 2005) or the Foundational Model of Anatomy (Rosse and Mejino, 2003), and
3. a process ontology, which contains the processes that realize an anatomical function.

<sup>2</sup>We recognize that this claim is controversial. There may be functions that can be realized by different kinds of processes. However, we assume that it is possible to find a super-category of these kinds of processes that include all and only those process categories that can realize the function. For example, a *Transport* function will always be realized by *Transport* processes (yet, arguably, not every *Transport* process is a realization of a *Transport* function), and these *Transport* processes can be of many different kinds, all of which are sub-categories of the *Transport* process category.

Because functional abnormalities are already classified as subclasses of structural abnormalities in the phenotype ontologies that we consider, we look for a pattern in the phenotype ontologies where

1. a category  $C$  in the phenotype ontology has a name  $name(C)$ ; e.g., *Abnormal blood coagulation* (MP:0002551),
2. in  $name(C)$ , the name or synonym  $name(D)$  of a GO BP category  $D$  occurs as a substring and  $name(D)$  is delimited by whitespaces in  $name(C)$ ; e.g., *Coagulation* (GO:0050817),
3. the category  $C$  is a sub-category of a category  $E$  with a name  $name(E)$ ; e.g., *Abnormal blood hemostasis* (MP:0009642),
4. the name  $name(E)$  contains the name or synonym  $name(F)$  of a category  $F$  from the anatomy ontology and  $name(F)$  is delimited by whitespaces in  $name(E)$ ; e.g., *Blood* (MA:0000059).

As a consequence, we find abnormalities of GO processes that are sub-categories of abnormalities of anatomical structures.

To exclude categories that are named after diseases or do not describe abnormalities, we only consider the categories of the phenotype ontologies which contain “abnormal”, “impaired”, “decreased” or “increased” in their name or synonyms and exclude the rest from our analysis. Furthermore, we excluded the GO categories GO:0032502 (developmental process), GO:0043473 (pigmentation) and GO:0001503 (ossification) from our analysis (see discussion section).

Figure 2 shows our implementation of the extraction pipeline<sup>3</sup>. To match the names of the categories, we stemmed all category labels and synonyms in the input ontologies using the PlingStemmer<sup>4</sup>. The PlingStemmer generates the singular forms of English words. Furthermore, all category labels were reduced to their lower case form before the matching was carried out.

## 2 Results

Using the HPO and the FMA, we could extract 25 structure-process pairs. These pairs and their evaluation are available from our project page.

Using the MPO and the MA ontologies, we extracted 331 structure-process pairs. A selection of the pairs we extracted that do stand in the **CC-hfrb** relation is shown in Table 2. In Table 2 we

<sup>3</sup>The implementation can be downloaded from <http://bioonto.de>.

<sup>4</sup>Available at <http://www.mpii.mpg.de/~suchanek>

Structure	Process
blood	morphogenesis of a branching structure
immune system	t cell apoptosis
trunk	biological regulation
pancreas	cell differentiation
blood vessel	endothelial cell differentiation

Table 1: Selection of false positives matches extracted from the Adult Mouse Anatomy Ontology and the MPO.

show pairs that do not stand in the **CC-hfrb** relation. An elaborate evaluation of our method in terms of its precision is ongoing work.

## 3 Discussion

### 3.1 Functions of parts

Although we successfully applied our proposed ontology pattern to harvest a basic ontology of anatomical functions from the phenotype ontologies by using naming patterns in the ontologies, there are cases in which our pattern fails. In particular the relation between functions and parts of structures remains a topic for further research.

We have argued that an abnormality of a function should be a sub-category of an abnormality of the bearer of the function. However, there may be cases where the bearer of the function is not included in the anatomy ontology or the abnormality of the function bearer is not included in the phenotype ontology. Instead, a structure of which the function bearer is a part, or an abnormality of such a structure is included.

The functional abnormality pattern is valid if we assume that the abnormality of the part is an abnormality of the whole. This assumption is supported by the phenotype ontologies. Nevertheless, to achieve completeness of both the anatomy ontologies and the phenotype ontologies, and to provide a principled way for building the phenotype ontologies, it is beneficial to include the abnormality of the function bearer whenever an abnormality of a function is included in the phenotype ontologies.

### 3.2 Text and naming problems

While processing the phenotype ontologies, we discovered several naming problems. First, plural forms are apparently randomly mixed with singular forms of the same term. For example, the label of MP:0003677 is “abnormal ear lobe” while the label of its subcategory MP:0003678 is “absent ear lobes” (plural). The same holds for HP:0000598 (abnormality of the ears) and HP:0000370 (abnormality of the

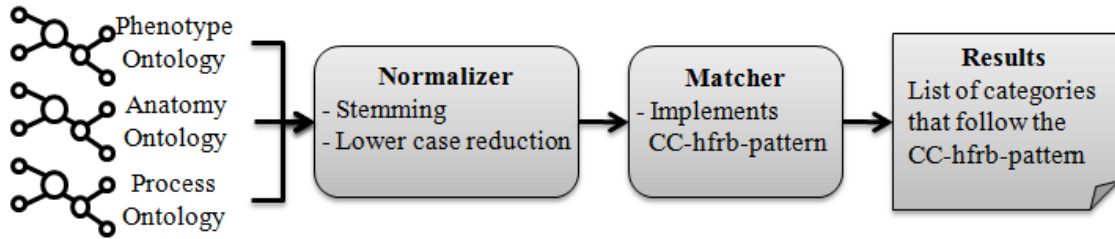


Figure 2: Processing sequence of the input categories

Structure	Process
cardiovascular system	anatomical structure morphogenesis
uterus	angiogenesis
blood vessel	cell migration
blood	coagulation
female reproductive system	diestrus
reproductive system	fertilization
pancreas	glucagon secretion
mammary gland	lactation

Table 2: Selection of true positives matches extracted from the Adult Mouse Anatomy.

middle ear). We suggest to use the plural form only in the case of explicitly disjunctively defined categories. For example, a category that is defined as the disjunction of the categories “abnormality of the left ear” and “abnormality of the right ear” may be called “abnormality of the ears”.

Another difficulty is the mixture of structural and functional abnormalities as category labels. For example, the category HP:0000251 is labeled “abnormality of tear glands OR tear production”. This name mixes structural and functional abnormalities: *tear glands* are an anatomical structure, while *tear production* is a process that realizes a function (the function “to produce tears”). To improve the usability and the possibilities for automatic processing of the phenotype ontologies, we suggest a separation of function and structure based abnormalities. For example, the category HP:0000251 should be split into two distinct categories, one labelled “abnormality of tear glands”, the other “abnormality of tear production”.

The third issue we found in the phenotype ontologies is the inconsistent use of category labels. The MPO contains the categories “abnormal hearing physiology” (MP:0001963), “hearing disability” and “hearing impairment” (exact synonyms for MP:0001965), “deafness” (MP:0001967) and “impaired hearing” (MP:0006325). The development of a naming convention would not only serve automatic processing of the ontologies, but help to improve the clarity of the phenotype ontologies.

The current use of polysemous words is one of the main drawback we face when trying to extract functions out of the phenotype ontology. *Ossification*, for example, can be understood both as the process of creating bone tissue and as a property of a bone (the outcome of the process). Thus, while bone ossification relates to the ossification process, skull ossification (HP:0002703) relates to the state of the skull, i.e., the result of the ossification process of the skull. Similarly, *pigmentation* is used widely as a property and not as the process of pigmentation.

Finally, a major problem for the phenotype ontologies is the use of “absent” as a property. In English, “absent” is used as an attributive adjective, and this is one reason why “absent” is present in some ontologies of qualities, in particular the phenotypic quality ontology PATO (PATO:0000462). In most ontologies, such as DOLCE (Masolo et al., 2003), GFO (Herre et al., 2006) or BFO (Grenon, 2003), qualities are dependent on a bearer, an entity *of which* they are a quality. The meaning of “absent”, however, entails that there is no such bearer. When “absent” is used in “absent appendix”, “absent nipple” or “absent hearing”, it does not correspond to an ontological quality (Hoehndorf et al., 2007). While this fact is increasingly being taken into consideration by the phenotype ontologies in the definition of categories pertaining to the absence of structures, “absent” is still used as a quality in the definition of categories of absent processes (or functions). These categories should be carefully examined and their

definition made clear. They can be formally defined using the functional abnormality pattern (Hoehndorf et al., 2009), which uses a form of the **lacks** relation (Ceusters et al., 2006) together with an ontological analysis of functions and dispositions.

The problem of “absent” is not a problem of the phenotype ontologies alone. The PATO ontology also includes “absent” as a quality, and it should be removed from the PATO.

### 3.3 Ontology problems

Our analysis is hindered by the lack of categories or synonyms for category names in GO’s Biological Process Ontology. For example, *tear production*, *cardiac conduction*, *hair pigmentation* or *taste sensation* are not in the GO, yet their existence is indicated by reference to these processes in the phenotype ontologies. Extension of the GO together with a consistent naming of the phenotype ontologies could improve our analysis and the clarity of the phenotype ontologies.

### 3.4 The need for an ontology of anatomical functions

One major problem in our analysis is the lack of an anatomical functions ontology. The phenotype ontologies imply that *HearingF* would be a function of the ears, by stating that an abnormality in *HearingF* is a sub-category of *Abnormality of the ears*. However, the ears can be normal and functioning normally, and still there may be an absence of *HearingP*. In particular, *Deafness* may be the result of an abnormality of the ears, or it may be the result of an abnormality in the nervous system. For example, an abnormality in the brain can impair *HearingF* just as well as an abnormality in the ears can. The ears only partially contribute to *HearingP*, and not every abnormality of *HearingF* is an abnormality of the ears.

Therefore, the ears and *HearingP* do not stand in the **CC-has-function-realized-by** relation according to our definition: the ears have some function which, if realized, is realized by processes that may be *part* of *HearingP* processes, but are not necessarily *HearingP* processes themselves, nor are they always part of *HearingP* processes. We could define another relation  $R(X, Y, Z)$  with the meaning that all instances of  $X$  have some function of the type  $Z$ , such that whenever a process of type  $Y$  occurs, an instance of  $Z$  is being realized. However, we believe that the relation **CC-has-function-realized-by** is currently more useful for application in the anatomy and phenotype ontologies, because it fits the current structure of both kinds of ontologies better.

One advantage of our introduction of the relation **CC-has-function-realized-by** is that needless duplication of the processes in the GO is avoided. In particular, many functions do not need to be explicitly named, because the processes in the GO are defined as processes that realize a given function. Nevertheless, as we have shown in the case of the function of the ears and *HearingP* processes, the functions of anatomical structures should be added to the anatomy ontologies. One option to do so would be to use the relation **CC-has-function-realized-by**, and extend the GO’s Biological Process Ontology with the appropriate processes.

## 4 Conclusions

We present an ontology design pattern for the representation of functional abnormalities. The design pattern is applicable to the Human Phenotype Ontology and the Mammalian Phenotype Ontology.

We show how to model anatomical functions using processes from the Gene Ontology that may realize these functions. For this purpose, we introduce a new relation between categories of anatomical structures and process categories. This relation states that an anatomical structure has some function that is realized by a process of a certain kind. Using this relation, functions can be specified without the explicit introduction of an ontology of anatomical functions.

We evaluated our method by exploiting the naming of categories from the phenotype ontologies to extract structure-process pairs that stand in the relation we introduce. We extracted several structure-process pairs from the Mammalian Phenotype Ontology together with the Adult Mouse Anatomy Ontology, and from the Human Phenotype Ontology together with the Foundational Model of Anatomy.

In our analysis, we find several problem with the phenotype ontologies. In particular, we found ambiguous namings of the categories and suggest the use of a naming convention for the categories in the phenotype ontologies. Additionally, we find a number of problematic formal definitions of categories in the phenotype ontologies. Most of these are categories of *malfunctionings*: the loss of the capability to perform a certain function. Use of our ontological framework permits an improved ontological representation of functional phenotypes and better capabilities for knowledge extraction from the phenotype ontologies.

## 5 Acknowledgements

We are grateful to Heinrich Herre, Roberto Poli and two anonymous reviewers for helpful comments on previous versions of this manuscript.



## References

- Mikel E. Aranguren, Erick Antezana, Martin Kuiper, and Robert Stevens. 2008. Ontology design patterns for bio-ontologies: a case study on the cell cycle ontology. *BMC Bioinformatics*, 9(Suppl 5):S1+.
- 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, Issel L. Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May.
- Jonathan Bard, Seung Y. Rhee, and Michael Ashburner. 2005. An ontology for cell types. *Genome Biology*, 6(2).
- Patryk Burek, Robert Hoehndorf, Frank Loebe, Johann Visagie, Heinrich Herre, and Janet Kelso. 2006. A top-level ontology of functions and its application in the Open Biomedical Ontologies. *Bioinformatics*, 22(14).
- Patryk Burek. 2006. *Ontology of Functions*. Ph.D. thesis, University of Leipzig, Institute of Informatics (IfI).
- Werner Ceusters, Peter Elkin, and Barry Smith. 2006. Referent tracking: The problem of negative findings. *Stud Health Technol Inform*, 124.
- Karen Eilbeck, Suzanna E. Lewis, Christopher J. Mungall, Mark Yandell, Lincoln Stein, Richard Durbin, and Michael Ashburner. 2005. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, 6(5).
- Christine Golbreich and Ian Horrocks. 2007. The OBO to OWL mapping, GO to OWL 1.1! In Christine Golbreich, Aditya Kalyanpur, and Bijan Parsia, editors, *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, Jun 6-7*, volume 258 of *CEUR Workshop Proceedings*, Aachen, Germany. CEUR-WS.org.
- Pierre Grenon. 2003. Bfo in a nutshell: A bi-categorical axiomatization of BFO and comparison with DOLCE. Technical report, University of Leipzig, Leipzig.
- Nicolai Hartmann. 1966. *Teleologisches Denken*. Walter de Gruyter.
- Terry F. Hayamizu, Mary Mangan, John P. Corradi, James A. Kadin, and Martin Ringwald. 2005. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biology*, 6(3).
- Heinrich Herre, Barbara Heller, Patryk Burek, Robert Hoehndorf, Frank Loebe, and Hannes Michalek. 2006. General Formal Ontology (GFO) – A foundational ontology integrating objects and processes [Version 1.0]. Onto-med report, Research Group Ontologies in Medicine, Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig.
- Robert Hoehndorf, Frank Loebe, Janet Kelso, and Heinrich Herre. 2007. Representing default knowledge in biomedical ontologies: Application to the integration of anatomy and phenotype ontologies. *BMC Bioinformatics*, 8(1).
- Robert Hoehndorf, Frank Loebe, Roberto Poli, Janet Kelso, and Heinrich Herre. 2008. Gfo-bio: A biological core ontology. *Applied Ontology*, 3(4):219–227.
- Robert Hoehndorf, Janet Kelso, and Heinrich Herre. 2009. Contributions to the formal ontology of functions and dispositions: an application of non-monotonic reasoning. In *Proceedings of the 11th Bio-Ontology Workshop*.
- Robert Hoehndorf. 2009. *Basic considerations for improving interoperability between ontology-based biological information systems*. Ph.D. thesis, University of Leipzig.
- Ingvar Johansson. 2004. Functions, function concepts and scales. *The Monist*.
- David Lewis. 1997. Finkish dispositions. *The Philosophical Quarterly*, 47(187):143–158.
- Frank Loebe. 2007. Abstract vs. social roles – Towards a general theoretical account of roles. *Applied Ontology*, 2(2).
- Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. 2003. Wonder-Web Deliverable D18: Ontology library (final). Technical report, Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy.
- Ruth G. Millikan. 1988. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
- Peter N. Robinson, Sebastian Köhler, Sebastian Bauer, Dominik Seelow, Denise Horn, and Stefan Mundlos. 2008. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *American journal of human genetics*, 83(5):610–615, November.
- Cornelius Rosse and José L. V. Mejino. 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36(6).
- John R. Searle. 1997. *The Construction of Social Reality*. Free Press, January.
- Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L. Rector, and Cornelius Rosse. 2005. Relations in biomedical ontologies. *Genome Biol*, 6(5).
- Larry Wright. 1973. Functions. *Philosophical Review*.