

Towards a Computational Paradigm for Biomedical Structure

Stefan Schulz

Department of Medical Informatics
Universitätsklinikum Freiburg
Stefan-Meier-Str. 26, D-79104 Freiburg, Germany
stschulz@uni-freiburg.de

Udo Hahn

Text Knowledge Engineering Lab
Universität Freiburg
Werthmannplatz 1, D-79085 Freiburg, Germany
hahn@coling.uni-freiburg.de

Abstract

The symbolic representation of the physical structure of living organisms needs an ontologically well-founded and logically sound approach so that formal reasoning can adequately be supported. We describe a set of canonical relations and attributes necessary for the description of biological structures. Based on these epistemological primitives, we sketch how a broad range of organisms can be represented by cascading theories which are ordered by various dimensions, such as granularity, development, species and canonicity. We thus aim at a rational reconstruction and non-redundant representation of biological structure notions.

Keywords: *Biological Ontologies*

Introduction

Formally founded descriptions of the physical composition of biological entities have attracted increasing attention in the last few years, as their pivotal role in biomedical ontologies has been increasingly recognized [2, 13, 16, 14].

In order to achieve a comprehensive formal representation of living systems, the first step would be to construct a multi-purpose reference ontology of *biological structure*. Such an approach should ideally cross the boundaries between species, because even organisms with largely different phenotypes show surprising similarities at a genetic level. Hence, knowledge about one organism should be re-usable in order to understand other organisms [19]. In terms of sheer coverage, a large amount of biomedical terms are already represented by the UMLS [18], the Gene Ontology [5] and a continuously increasing number of “anatomies”, developed within the Open Biological Ontologies (OBO) framework. [17]. However, all of these systems are committed to a highly selective view of biological structure in terms of devel-

opmental stages, granularity and species-specific structure. Each species anatomy is being built from scratch, although the rough architecture of organisms exhibits considerable similarities between species and developmental stages.

Focusing on the anatomy of the heart, Fig. 1 shows a synopsis of several OBO models, together with the Foundational Model of Anatomy (FMA) [16, 15]. Abstracting away from terminological differences (e.g., *circulatory system* vs. *cardiovascular system*), we recognize a number of commonalities between diverse organisms. For instance, the *heart* is always part of the *circulatory system*. Except in the case of flies and in the early developmental stages of the mouse, *hearts* have *chamber(s)* and *valves*. The difference between *heart atriums* and *ventricles* exists in fish as well as in mice and humans.

With the exception of the FMA, which is based on strict principles and is moving towards a formally founded redesign, the anatomies of the other species, as well as the (theoretically) species-independent Gene Ontology, are no more than controlled vocabularies with thesaurus-like relations, which in some cases do not even make consistent use of the *part-of* relation and provide largely incomplete taxonomic links [1]. Consequently, the decision as to whether a deduction such as *cell has-part nucleolus* is valid or not and how it should be interpreted, assuming a model which asserts, e.g., *cell - has-part - cell nucleus* and *cell nucleus - has-part - nucleolus*, is left to the user, because there is no ontological commitment to either the algebraic properties of *has-part*, e.g. transitivity, or the dependency status of *has-part cell nucleus* (is every nucleolus part of a cell nucleus, or does every cell nucleus have a nucleolus as part?).

This may be acceptable when the use of these vocabularies is limited to manual, expert-level gene annotation or document retrieval tasks. How-

Mouse (embryonal stage TS11, source: MGI)

- cardiovascular system
- - heart
- - - cardiogenic plate



Mouse (embryonal stage TS18, source: MGI)

- cardiovascular system
- - heart
- - - atrio-ventricular canal
- - - atrium
- - - bulboventricular groove
- - - bulbus cordis
- - - endocardial cushion tissue
- - - mesentery
- - - outflow tract
- - - pericardium
- - - primitive ventricle
- - - sinus venosus



Mouse (embryonal stage TS26, source: MGI)

- cardiovascular system
- - heart
- - - aortic sinus
- - - atrio-ventricular canal
- - - atrio-ventricular cushion tissue
- - - atrium
- - - bulbar cushion
- - - endocardial cushion tissue
- - - endocardial tissue
- - - mesentery
- - - pericardium
- - - trabeculae carneae
- - - valve
- - - ventricle



Drosophila (adult, source: FlyBase)

- circulatory system
- - heart
- - - heart muscle
- - - adult aortic funnel
- - - adult ostia
- - - dorsal diaphragm
- - - heart chamber
- - - terminal opening



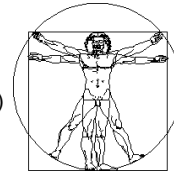
Zebrafish (adult, source: ZFIN)

- cardiovascular system
- - heart
- - - atrium
- - - bulbus arteriosus
- - - hypobranchial vessels
- - - sinus venosus
- - - ventricle



Human, Adult, (source: FMA)

- cardiovascular system
- - heart
- - - wall of heart
- - - right atrium
- - - left atrium
- - - right ventricle
- - - left ventricle
- - - right side of heart
- - - left side of heart
- - - fibrous skeleton of heart
- - - papillary muscle
- - - cardiac valve
- - - tricuspid valve
- - - mitral valve
- - - aortic valve
- - - pulmonary valve
- - - interatrial septum
- - - (...)



} is-a organ chamber

} is-a cardiac valve

Figure 1: Comparative Heart Anatomy (only part-whole links) from OBO Biological Ontologies and the Foundational Model of Anatomy

ever, anticipating their use for knowledge-intensive applications, the informal approach leads to an impasse, because semantically vague, and even inconsistent assertions about concepts may cause a broad range of arbitrary invalid and, thus, unwarranted deductions.

Therefore, we argue for a domain description in terms of a set of formal axioms which allow valid and correct inferences. Complex concept descriptions built from a set of formally founded conceptual relations may be a good starting point for a formally adequate treatment of biological structures. In the following sections, we focus on various aspects of conceptual modeling of biological structure in its broadest sense, aiming at a multi-purpose foundational ontology.

Relations and Attributes

An ontological analysis of any domain should be guided by generally shared principles. According to Gangemi *et al.* [4], this first requires selecting a set of foundational (formal) relations, then defining the ground axioms for these relations, establishing constraints across basic relations and defining a set of formal properties induced by these formal relations. Then a set of basic categories is introduced, and the relevant kinds of domain entities (concept classes and instances) are classified according to the basic categories. Finally, the dependencies and interrelations among basic categories have to be studied. In this paper, we limit ourselves to an overview of adequate foundational relations and attributes. Rather than proposing a single canonical formalism, we outline alternative axiomatizations, their consequences and intricacies. Our selection of relations comprises some of the (informal) relations provided by the UMLS semantic network, completed by additional ones, considered relevant for describing biological structure.

Foundational Relations

Domain entities can be ordered according to (strict) partial orders which are characterized by a set of ordering relations. Strict partial orders are transitive, antisymmetric and irreflexive, whereas partial orders are transitive, antisymmetric and reflexive. Additional constraints may stem from type restrictions on the domain and the range of a relation. Of paramount importance is the distinction between classes (universals, concepts, sorts of things, e.g., “*Left Hand*”) and individuals (particulars, instances, concrete things in the world, e.g., “*my left hand*”). Because we have found that relations (such as *part-of*) are commonly asser-

ted between concept classes – and have therefore a different semantics than their cognate relations between individuals – we stick to the following naming convention: lower case relation names are used for relations between individuals, and upper case names characterize relations between concept classes. Accordingly, we write concept (class) names with upper case initials, and instance names with lower case initials.

Taxonomy: The taxonomic *Is-A* relation, a partial order, [26] relates specific classes to conceptually more general classes, e.g. *Mitral Valve Is-A Atrioventricular Valve* or *Alanin Is-A AminoAcid*. More specific classes inherit all properties from more general classes. The definition of a class illuminates its distinctive characteristics in relation to already defined (more general) classes, following the Aristotelian principle of *genus* and *differentiae*. Whereas the genus assigns an entity to a class, the *differentiae* distinguish the entity from other entities also assigned to that class. For example, *Left Hand* has *Hand* as its genus and its laterality attribute *left* as *differentiae*. Taxonomies can have either a monohierarchical (single parent), or a polyhierarchical (multiple parent) structure. In the Foundational Model of Anatomy (FMA) [16, 11], e.g., huge taxonomies are represented as strict monohierarchies. The relation *Is-A* must not be mixed up with the relation *instance-of* which relates individuals with the classes they belong to, e.g., *my left hand instance-of Left Hand*. Unfortunately, the relation *instance-of* is often used inadequately in biomedical ontologies, e.g. *Muscle System instance-of Organ System* in the FlyBase vocabulary.

Mereology: At least for the life science domain, not only taxonomic relations (*Is-A* and *instance-of*) but also mereological relations (basically, *part-of* vs. *has-part*) are of outstanding and equal importance for the design of any ontology describing biological structure. In classical (i.e., axiomatic) mereology [22, 3] generic parthood is treated as a partial order. Common conceptualizations in the biological domain, however, suggest that the assumption that *part-of* be reflexive must be abandoned.¹ The most obvious distinction between *Is-A* and *part-of* relates to the fact that the first one is maintained between classes, whereas the second one is maintained only between individuals. As an example, *my left thumb* is *part-of my left hand*,

¹Otherwise, any instance of “stomach” would be an instance of “stomach part”, with the consequence that the class “*partial resection of stomach*” would include “*total resection of stomach*”.

but the class *Thumb* is certainly not *part-of* the class *Hand*. However, “being part-of a hand” is a property of any instance of *Thumb*. This distinction has been largely ignored in our domain. As a result, the meaning of mereological relations asserted between a pair of concepts, such as *Part-Of(CellNucleus, Cell)*, is ambiguous, allowing the possibility for conflicting interpretations to evolve: The Gene Ontology [5] interprets *Part-Of* as “can be a part of, not is always a part of” which frequently leads to unexpected conclusions [25]. In contrast, the Foundational Model of Anatomy (FMA) [16] conceptualizes *part-of* in a very strict manner: *A Part-Of B* means that any instance of *B* has an instance of *A* as part, and any instance of *A* is part of an instance of *B* [23]. This interpretation imposes a mutual dependency between parts and wholes and, therefore, may be too rigid in many cases. For example, we may want to express that any instance of a cell nucleus is part of a cell, but certainly not any instance of a cell has a cell nucleus. Certainly, we also may want to instantiate non-standard organisms which lack certain body parts. As far as other models of organisms referred to in the introductory section are concerned, especially mouse, zebrafish and drosophila anatomy, there is no commitment at all to the proper semantics of *Part-Of*.

A mereological relation between concepts (classes of individuals), therefore, cannot be interpreted unambiguously, unless we make clear statements on the existence of a whole with respect to its parts, as well as the existence of a part with respect to its whole. Taking into account the (supposed) intended meaning of mereological relations between concepts, we define, similar to [23], *Part-Of* and *Has-Part* on the basis of *part-of* and *has-part*, using *inst-of* as the membership relation between an individual and a class:

$$\begin{aligned} \text{Part-Of}(A,B) =_{\text{def}} \forall x : \text{inst-of}(x,A) \Rightarrow \\ \exists y : \text{inst-of}(y,B) \wedge \text{part-of}(x,y) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Has-Part}(A,B) =_{\text{def}} \forall x : \text{inst-of}(x,A) \Rightarrow \\ \exists y : \text{inst-of}(y,B) \wedge \text{has-part}(x,y) \end{aligned} \quad (2)$$

Location: The locative relation [3], characterized by the relation pair *location-of* vs. *has-location*, is another partial order between individuals. It relates a spatial entity with another spatial entity or a material object, e.g., *brain has-location cranial cavity*. Wherever locative relations are asserted between concept classes, we define *Location-Of* and *Has-Location* similar to

Part-Of and *Has-Part* in Formula (2) and (3). A crucial decision is whether to keep mereological and topological aspects separated, or to subscribe to a more simplified mereotopological view in which spatial objects coincide with the region they occupy. As an example, is a *bacterium* after being ingested by a *cell* (e.g., a *macrophage*) part of this *cell*? If not, do its components (e.g., molecules) become parts of the original structure after decomposition? Without any doubt, both the bacteria and its components are *located* within that cell. Similarly, is a hollow space a part of its host or part of the exterior space (cf. [20])? Is a boundary a part of the entity it bounds?

In a restricted domain such as biology, the distinction between mereology and topology may seem arbitrary and inconsistent. Here, *Part-Of* may imply *Has-Location*, and connection can be expressed in terms of mereology [3]. In this case, mereological relations would be mere subrelations of locative ones [21]. This may, however, complicate the conceptualization of detached parts, which one could still consider to be included in the notion of part. For example, a metastasis of a tumor may still be considered a part of the primary tumor which is, however, not located in the primary site (the alternative would be to consider it related to the primary tumor by a relation such as *has-origin*). This example makes clear how important it is in biology to clarify the meaning of part, where at least three conceptualizations co-exist: The locative one (a heart chamber is part of a heart), the functional one (an axon is part of a motor neuron), and the one motivated by origin (a metastasis is part of a tumor, an epithelium in a sputum sample is part of the respiratory mucosa).

Other Foundational Relations: **Branching** relations (*has-branch*, *branch-of*) define tree-like structures which typically describe pathways for the flow of matter or information in higher organisms (blood, lymphatic vessels and nerves), but which may also constitute the building principle of an organisms such as a plant or coral. There are several ways to conceptualize branching relations. In the FMA, a tree consists of a trunk and many generations of branches. Each branch is considered a subtree of a higher order tree, and each branch also has its own trunk. Thus *branch-of* can be interpreted either as a subtree or as a continuity relationship between two or more trunks. A subtree branch has a part relation to the higher order tree; two trunks have a branch relationship if they are continuous with one another end to side

or if a trunk terminates by bifurcating or trifurcating into subsidiary trunks. Consequently, branching relations cannot be subsumed by mereological relations because, generally speaking, a branch is not considered part of its trunk. To further illustrate this, any instance of *Aorta*, as the trunk of a systemic arterial tree, does not mereologically include any instance of its ramifications such as *Left Common Iliac Artery*, or *Femoral Artery*. Wherever branching relations are asserted between concept classes, we define *Has-Branch* and *Branch-Of* similar to *Part-Of* and *Has-Part* in Formula (2) and (3). The development of the individual (**ontogeny**) and the development of the species (**phylogeny**) accordingly form the relation pairs *has-developmental-form/ developmental-form-of*, and *Has-Descendent* vs. *Descends-From*. Both are strict partial orders. In an embryo, e.g., its *splanchnic mesenchyme* is a developmental form of its *cardiogenic cords*, which – across some other steps – is a developmental form of its *primitive heart*. According to the above comments, the inter-concept relations (*Has-Developmental-Form / Developmental-Form-Of*) have to be introduced when two concept classes are to be linked in terms of ontogeny.

All phylogeny relations, in contrast to the ontogeny relations, are maintained between concept classes, and not between individuals. As an example, *Homo Habilis Has-Descendant Homo Erectus* and *Homo Erectus Has-Descendant Homo Sapiens*. For any given instance of homo sapiens, there is no specific instance of any other hominid species, so there is no correlate of this relation at the level of individuals. Phylogenetic relations are maintained between organism concepts as well as between anatomical structure concepts (e.g., *Wing Descends-From Forelimb*).

There are other relations which are not partial orders but to which a foundational status can be equally ascribed. Topology provides, in addition to mereology, an important ontological organization principle. In formal approaches to topology, the basic relation, *connects*, is symmetric and relates two entities in space [3]. There are different kinds of connection, e.g. external connection (touching) or partial overlap [12]. Biological and common-sense notions of connection vary widely, so it may be advisable to talk about continuity, contiguity or attachment. If we stay closer to formal topology, we need the relation *externally connects*, which describes the touching of two objects without the sharing of parts, corresponding to the relation *continuous-with* in the

FMA. For example, an *endocardium* is externally connected to a *myocardium*. If we allow boundaries (see below), another important relation pair is *bounds* vs. *bounded-by* [10], which is irreflexive, intransitive and antisymmetric (e.g., a *heart* is bounded by a *surface-of-the-heart*). Again, all of these relations exclusively relate individuals. Therefore, new concept-to-concept relations have to be defined (e.g. *Connects*), wherever they occur, in concept class definitions, similar to formulae (2) and (3). Note that the algebraic properties of these relations may differ: *continuous-with* is symmetric, but *Continuous-With* is not: In an individual neuron, its *cell body* is connected to its *axon* and vice versa. This contrasts with what we observe at the level of concept classes: Although each *axon* is connected to some *cell body*, not every *cell body* is connected to an *axon*.

General Attributes

In contrast to relations (e.g., *has-part*, *Is-A*), ordinary attributes such as *has-dimension*, *has-inherent-shape* can only be filled once. Important attributes are the (geometric) dimension, the distinction between solid and holes, as well as the distinction of count, collection and mass entities. All biological structure (in a strict sense) has a spatial dimension, which ranges from volumes, surfaces, lines to points. Quite naturally, the notion of a boundary comes into play. Any boundary must have exactly one dimension less than the entity it bounds. This restricts the domain and the range of the bounding relation pair *bounds / bounded-by*. Upper-level concepts, such as *Volume*, *Surface*, *Line*, *Point*, divide the domain of spatially relevant biological concepts into four disjoint partitions, because each biological structure entity has exactly one defining dimension [10]. Bounding structures can also be divided into so-called fiat and bona fide boundaries. According to [24] and [9], *bona fide* boundaries are those which have a structural correlate, e.g. the surface of the body, or the inner surface of a cell membrane. *Fiat* boundaries are ‘artificial’ boundaries, e.g., the *Medioclavicular Line*, or the *Sagittal Plane* in gross anatomy.

The next fundamental ontological distinction between three-dimensional objects is between “hollow spaces” and “solids”. Examples for hollow spaces are the cranial cavity, the right atrium, the lumen of a bronchiole or the hollow space in a protein molecule. Nearly all biological objects have hollow space as parts (It is, therefore, not plausible to consider them as parts of the exterior space, such as in formal topology). A possible

axiomatization is that solids must have solids and may have hollow spaces as parts, whereas hollow spaces can only have hollow spaces and boundaries as parts [20].

Biological structures can occur as single, countable entities (e.g., a *liver*, a *tooth*, or a *cell*), but also as collections of uniform objects (e.g., *mitochondria*), or as stuff, e.g. a portion of *blood* or *water* [6]. Collection entities can be viewed either as sets of their constituents, or as their mereological sum. In the latter case, the relation between a collection and its elements boils down to a sort of *has-part*. As an example, the concept class *Leukocytes* denotes all possible mereological sums of individual *leukocytes*. Mass concepts can be treated as collections as well, because they are collections of small particles (cells, molecules, atoms). Whether to classify an item as a mass or collection is essentially a matter of perspective .

Non-foundational Relations between Concepts

There are some relations in the FMA, UMLS or in OPENGALLEN which do not have a foundational status: Subrelations of *part-of*, such as *segment-of*, *layer-of*, *shared-part-of*, *arbitrary-part-of*, or *constitutes*, for which transitivity no longer holds [14, 9], can often be derived from the foundational ones by domain or range restrictions. For instance, *layer-of* requires an anatomical layer as domain and a physical entity as range. Or, *constitutes* has a mass or material as domain. The relation *shared-part-of*, on the other hand, can be inferred from the fact that an entity is part of more than one other entity. E.g., an *aorta* is part of a *trunk* and part of a *systemic arterial tree*. Other relations that can be inferred are *innervation* (*nerve* whose *endings* are connected to a *muscle*) and *insertion* (*tendon* connected to a *bone*).

The relation *Is-Conceptually-Disjoint* relates two concept classes which do not have any instance in common. This is the default situation in strict monohierarchies where all classes which do not subsume one another are mutually disjoint. In polyhierarchic taxonomies a class may have more than one taxonomic parent. As an example, *Pancreas* may be modeled as being both an *Endocrine Organ* and an *Exocrine Organ*, and an *Amino Acid* both as an *Organic Acid* and an *Organic Amine*. Most pairs of concept classes, however, are mutually exclusive: An *organ* cannot be a *cell*, and a *nucleotide* cannot be a *lipid*. In order to prevent unintended models, these concepts (or any parent of them) must be linked via the relation *Is-*

Conceptually-Disjoint.

An analogous situation can be observed in a mereological ordering. Most arbitrary physical entities are spatially disconnected, e.g., there is no pair of respective instances that share any parts, e.g., a *hand* with a *foot*, or an *eye* with a *mouth*. Mereological disconnectedness between concepts can be asserted whenever the following condition is assumed to hold:

$$\begin{aligned} \text{MereologicallyDisconnected}(A, B) =_{\text{def}} \\ \forall x, y : \text{inst-of}(x, A) \wedge \text{inst-of}(y, B) \Rightarrow \\ \neg \exists z : \text{part-of}(z, x) \wedge \text{part-of}(z, y) \end{aligned}$$

Theories

The vast domain of life science requires a decomposition of the whole domain into local theories, both in terms of granularity [8] as well as scope [7]. We define a theory as a set of formal axioms which describe a restricted (local) domain. We propose a lattice of theories which is designed along four parameters, viz. *granularity* (G), *species* (S), *development* (D) and *canonicity* (C).

Granularity. The conceptualization of biology is coined by our cognition. Macroscopic anatomy is restricted to the naked eye's view, histology requires a light microscope, our notions of cell biology are formed by the electron microscope, and knowledge of molecular biology and genetics is gathered using chemical and physical techniques. Along these lines, granularity issues have a large impact on high-level properties. In a very coarse-grained view, one may even consider classifying a microscopically thin membrane, such as a *basement membrane*, as a two-dimensional boundary, thus completely neglecting its spatial extension. Besides the sortal difference (degrees of dimensionality are mutually disjoint), this also has an impact on the connection of neighboring structures. What may be defined as externally connected to the naked eye will appear disconnected under the microscope. A low granularity may also encompass abstractions in terms of neglecting structural differences of kinds of objects (concept classes). Cell populations, such as *Leukocytes*, e.g., may be further classified into *Lymphocytes*, *Granulocytes* and others. A distinction of *Lymphocytes* into *B-* and *T-Lymphocytes*, and the latter into *T4-* and *T8-lymphocytes* will be required only in fine-grained theories, e.g. needed for the description of the pathology of immunodeficiency. In a complete ontological account of living organisms, granularity ranges from populations, on the one hand, to atoms and subatomic particles, on the other hand.

Species. The universe of life includes millions of species. Hence, the domain of human anatomy is an extremely restricted one. Mediating domains are those of vertebrates or mammals. According to the classification of organisms, which is the prototype of a taxonomic order, properties can be introduced at any level of the classificatory tree and propagate across that tree. Under a simplifying view, *heart* is a muscular organ which has a cavity and is part of a *circulatory tract*. These properties hold true for *chordates*, *arthropods* and some other phyla. As far as the hearts of more specific organisms are concerned, additional properties are required, e.g., a certain number of *ventricles* and *valves*, the presence of *blood* or *hemolymph*, different locations of *pacemaker cells* (see Fig. 1). Additionally, we have to consider intra-species variations such as gender or race.

Development. Organisms traverse a life cycle from birth to death. Each developmental stage has its own characteristics. Even distantly related organisms, such as humans and flies, exhibit a high degree of similarity in the first embryologic stages. The existence of many parts of an organism is restricted to certain stages. For example, in mice embryos, an *ectoderm* exists only in the so-called Tanner stages TS9 – TS19, and there is no heart before the Tanner stage TS11. Other body parts (e.g. the heart, cf. Fig. 1) appear in a certain embryologic stage and perdure in all subsequent steps of the life cycle.

Canonicity. Here we introduce the notion of *canonicity*, as the well-formedness of biological structure, and define it as the degree by which a biological object corresponds to its canonical, i.e., idealized form. We suggest an ordinal scale with five levels of canonicity, cf. Table 1. The higher the canonicity level, the more axioms have to be applied. All axioms introduced in a lower level are propagated to all higher levels. Axioms which describe structural modifications specific to a concrete disorder, e.g. *Stomach Has-Part Ulcer* are not considered in this framework.

- *Level 1* introduces those axioms which hold even with lethal structural modifications or post-mortem degeneration, such as *Erythrocytes has-part Hemoglobin*, *Bone Has-Part Calcium Carbonate*, *Heart Ventricle Part-Of Heart*, *Leather Has-Part Collagen* (but not *Heart Valve Part-Of Heart* because it could be an isolated heart valve for transplantation);
- *Level 2* introduces, additionally, all those axioms which hold for the description of biological

structures organized in an *organism*, irrespective of living or dead, e.g., the axiom *Heart Valve Part-Of Heart* is introduced at this level, as well as *Cell Nucleus Part-Of Cell*;

- In *Level 3* all those axioms are added which hold in living organisms, in addition to dead organisms, e.g., *Aorta Location-Of Blood*, or *Vertebrate Body Has-Part Head* (but not yet *Gastrointestinal Tract Has-Part Stomach*, because most individuals survive a total remotion of the stomach).
- *Level 4* introduces, additionally, all those axioms which characterize a healthy organism, e.g. *Hand Has-Part Thumb* and *Gastrointestinal Tract Has-Part Stomach*. However, it still allows anatomical variations when they have no impact on the function of the organism.
- *Level 5* finally completes the set of axioms needed for the description of the “ideal” organism. Here enter, e.g., many cardinality constraints (e.g., in human: *32 teeth*), one *spleen*, three *lobes* of the *right lung*.

A theory can be expressed by a node in the lattice of the four axes *viz.* G, S, D, and C. Hereby the values of granularity (G), development (D), canonicity (C) are located on an ordinal scale, the values of species (S) are given by the nodes of the classificatory tree introduces properties which are inherited by its subsequent nodes. As an example, *Fish Heart Is-A Vertebrate Heart* or *Drosophila Eye Is-A Arthropod Eye*. This means that *Fish Heart* inherits all properties from *Vertebrate Heart*, and *Drosophila Eye* inherits all properties from *Arthropod Eye*. The same mechanism can be observed with canonicity. All properties that structures of low canonicity have in common (e.g., *Tissue* consisting of *Cells*) are inherited by the more canonic structures. No such inheritance rules apply to the variables *development* (D) and *granularity* (G).

Taking the *heart* as prototypical example, we will now demonstrate practical inferences which are supposed to be drawn from a biological ontology based on our assumptions:

- A heart with four chambers is not compatible, e.g., with any theory characterized by $S = \textit{fish}$, or by $S = \textit{human} \ \& \ D = \textit{4-week-embryo}$.
- Let us assume the relation *connects* which is maintained between the right and the left ventricles. We then may exclude most non-mammals

Level	1	2	3	4	5
Theory	any amount of matter, if of biological origin	any living or dead organism	any living organism	living organism without pathologic modifications	ideal organism
Set of Axioms	n_1	n_2 $n_1 \subset n_2$	n_3 $n_2 \subset n_3$	n_4 $n_3 \subset n_4$	n_5 $n_4 \subset n_5$

Table 1: Ordinal Scale of Canonicity

(since they have no right and left ventricle), but we may also exclude anatomical hearts of adult mammals (because they have a septum between the two ventricles). This scenario is compatible with the theories $D = \textit{embryo}$ & $S = \textit{mammal}$ as well as with $S = \textit{mammal}$ & $C = 3$.

- Given the theory $D = \textit{adult}$ & $S = \textit{vertebrate}$ & $C = 5$, every instance of heart implies the location of blood, and every instance of blood has an instance of erythrocytes as part. Assuming that *has-part* implies *location-of*, and that *location-of* is transitive, we are able to infer that in this theory every instance of *Heart* is the location of an instance of *Erythrocytes*, as well.

Conclusions

In this paper we defined requirements for ontologies of biological structure. We introduced a set of canonical relations and attributes required for the description of biological structure, and discussed their semantics as well as algebraic properties. Finally, we sketched an architecture by which terminological knowledge about the anatomy of a broad range of organisms, developmental stages as well as malformations and pathological modifications, can be expressed. A central element is the decomposition into theories, which help organize the hierarchies and the axioms in terms of granularity, developmental stage, species, and canonicity. We claim the following advantages in using this approach:

(i) Redundancies are avoided. As an example, most axioms that describe the species *mice*, *humans*, and *dogs* are identical and therefore can reasonably be encoded into a more general theory (such as the one of *vertebrates*). In turn, the more general theory inherits the shared properties of more specific theories, e.g., the ones pertaining to mice, humans or dogs. In a similar vein, attributes that healthy and pathologically modified organisms have in common are described in the non-canonical theory from which the canonical theory inherits the shared properties.

(ii) Adequate theories for a specific application can be selected. It is neither computationally tractable nor useful to export the whole knowledge of biology into a formalism in which logical operations can be performed, e.g., by a terminological reasoner. For example, if we need to reason about a TS12 mouse embryo, we select the adequate intersection of theories to access the axioms we really need. Some of these axioms are inherited from the *mammal* theory, others from the theory of the *vertebrates*, and still others come from the theory of the *chordates*. Some axioms are encoded in the subtheory of a developmental stage of the vertebrates, and, last but not the least, there are some axioms which are only specific to the TS12 mouse embryo.

(iii) The intersection of arbitrary theories has variable extensions. There are many cases with no extensions. The compatibility of theories can be checked by formal reasoning devices. As an example, a heart with one ventricle in a theory restricted by $S = \textit{human}$ and $D = \textit{adult}$ is not compatible with $C = \textit{canonical}$.

References

- [1] J. Aitken, B. Webber, and J. Bard. Part-of relations in anatomy ontologies: A proposal for RDFS and OWL formalisations. In *Proceedings of the Pacific Symposium on Biocomputing 2004*, pages 166–177. Hawaii, USA, January 6-10, 2004.
- [2] K. Campbell, A. Das, and M. Musen. A logical foundation for representation of clinical data. *Journal of the American Medical Informatics Association*, 1(3):218–232, 1994.
- [3] R. Casati and A. Varzi. *Parts and Places. The Structures of Spatial Representation*. Cambridge, MA: MIT Press, 1999.
- [4] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Understanding top-level ontological distinctions. In *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing*, pages 26–33. Seattle, USA, August 4-5, 2001.
- [5] Gene Ontology Consortium. Creating the Gene Ontology resource: Design and implementation. *Genome Research*, 11(8):1425–1433, 2001.
- [6] P. Gerstl and S. Pribbenow. Midwinters, end games and body parts: A classification of part-whole relations. *International Journal of Human-Computer Studies*, 43:865–889, 1995.
- [7] C. Ghidini and F. Giunchiglia. Local models semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence*, 127(2):221–259, 2001.
- [8] J. Hobbs. Granularity. In *Proceedings 9th Intl. Joint Conference on Artificial Intelligence*, pages 432–435. Los Angeles, CA, 18-23 August, 1985.
- [9] J. Mejino, A. Agoncillo, K. Rickard, and C. Rosse. Representing complexity in part-whole relationships within the foundational model of anatomy. In *Proceedings of the 2003 Annual Symposium of the American Medical Informatics Association*, pages 71–75. Washington, D.C., November 8-12, 2003.
- [10] P. Neal, L. Shapiro, and C. Rosse. The DIGITAL ANATOMIST structural abstraction: A scheme for the spatial description of anatomical entities. In *Proceedings 1998 AMIA Annual Fall Symposium*, pages 423–427. Orlando, FL, Nov. 7-11, 1998.
- [11] F. N. Noy, M. Musen, J. Mejino, and C. Rosse. Pushing the envelope: Challenges in a frame-based representation of human anatomy. Technical Report SMI-2002-0925, Stanford University, 2002.
- [12] D. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In *Principles of Knowledge Representation and Reasoning. Proceedings of the 3rd International Conference*, pages 165–176, 1992.
- [13] A. Rector, A. Gangemi, E. Galeazzi, A. Glowinski, and A. Rossi-Mori. The GALEN model schemata for anatomy: Towards a re-usable application-independent model of medical concepts. In *Proceedings of the 12th Conference of the European Federation for Medical Informatics*, pages 229–233. Lisbon, Portugal, 1994.
- [14] J. Rogers and A. Rector. GALEN’s model of parts and wholes: Experience and comparisons. In *Proceedings of the 2000 Annual Symposium of the American Medical Informatics Association*, pages 714–718. Los Angeles, CA, November 4-8, 2000.
- [15] C. Rosse and J. Mejino. A reference ontology for bioinformatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 2004. In press.
- [16] C. Rosse, J. Mejino, B. Modayur, R. Jakobovits, K. Hinshaw, and J. Brinkley. Motivation and organizational principles for anatomical knowledge representation: The DIGITAL ANATOMIST symbolic knowledge base. *Journal of the American Medical Informatics Association*, 5(1):17–40, 1998.
- [17] OBO. *Open Biological Ontologies (OBO)*. <http://obo.sourceforge.net/>, 2004.
- [18] UMLS. *Unified Medical Language System*. Bethesda, MD: National Library of Medicine, 2003.
- [19] XSPAN. *Cross Species Anatomy Network (XSPAN)*. <http://www.xspan.org>, 2004.
- [20] S. Schulz and U. Hahn. Mereotopological reasoning about parts and (w)holes in bio-ontologies. In *Formal Ontology in Information Systems. Collected Papers from the 2nd International FOIS Conference*, pages 210–221. Ogunquit, Maine, USA, October 17-19, 2001.
- [21] S. Schulz and U. Hahn. Parthood as spatial inclusion – evidence from biomedical conceptualizations. In *Principles of Knowledge Representation and Reasoning. Proceedings 9th International Conference*. Whistler, Canada, June 2-5, 2004.
- [22] P. Simons. *Parts: A Study in Ontology*. Oxford: Clarendon Press, 1987.
- [23] B. Smith and C. Rosse. The role of foundational relations in the alignment of biomedical ontologies. In *Proceedings 11th World Congress on Medical Informatics*. San Francisco, CA, Sept., 2004.
- [24] B. Smith and A. Varzi. Fiat and bona fide boundaries. *Philosophy and Phenomenological Research*, 60(2):401–420, 2000.
- [25] B. Smith, J. Williams, and S. Schulze-Kremer. The ontology of the Gene Ontology. In *Proceedings of the 2003 Annual Symposium of the American Medical Informatics Association*, pages 609–613. Washington, D.C., November 8-12, 2003.
- [26] C. Welty and N. Guarino. Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering*, 39(1):51–74, 2001.