16

CARO — The Common Anatomy Reference Ontology

Melissa A. Haendel, Fabian Neuhaus, David Osumi-Sutherland, Paula M. Mabee, José L.V. Mejino Jr., Chris J. Mungall, and Barry Smith*

Summary. The Common Anatomy Reference Ontology (CARO) is being developed to facilitate interoperability between existing anatomy ontologies for different species, and will provide a template for building new anatomy ontologies. CARO has a structural axis of classification based on the top-level nodes of the Foundational Model of Anatomy. CARO will complement the developmental process sub-ontology of the GO Biological Process ontology, using the latter to ensure the coherent treatment of developmental stages, and to provide a common framework for the model organism communities to classify developmental structures. Definitions for the types and relationships are being generated by a consortium of investigators from diverse backgrounds to ensure applicability to all organisms. CARO will support the coordination of cross-species ontologies at all levels of anatomical granularity by cross-referencing types within the cell type ontology (CL) and the Gene Ontology (GO) Cellular Component ontology. A complete cross-species CARO could be utilized by other ontologies for cross-product generation.

16.1 Necessity of a Common Anatomy Reference Ontology

Genomes are modified over evolutionary time to produce a diversity of anatomical forms. Understanding the relationship between a genome and its phenotypic outcome requires an integrative approach that synthesizes knowledge derived from the study of biological entities at various levels of granularity, encompassing gene structure and function, development, phylogenetic relationships, and ecology.

Many model organism databases collect large amounts of data on the relationship between genetic/genomic variation and morphological phenotypes in databases. Model organism databases standardize the description of morphological phenotypes and gene expression patterns by using types from anatomy ontologies that are specific to their focus species of interest. These ontologies have allowed the model organism databases to group phenotypic and gene expression data pertaining to partic-

* Melissa Haendel, Fabian Neuhaus, and David Osumi-Sutherland contributed equally to this chapter.

ular anatomical types.[2] Methods of phenotype curation are being extended and standardized as part of the work of the National Center for Biomedical Ontology, which aims to provide data-mining tools that can be applied across all species. In particular these tools will facilitate queries relating to anatomical structures and associated genes. Currently, however, there is no system for standardizing the representation of anatomy in ontologies.

Cross-species standardization among anatomy ontologies would bring a number of benefits. First, it would allow the development of standardized tools for grouping and querying anatomy-linked data. Second, it is a prerequisite for inference of anatomically based phenotypic and gene expression data within and across species. Third, if anatomy ontologies were standardized, then a method for representing homology between anatomical types in different anatomy ontologies could be devised. Fourth, standardization would allow better interoperability between anatomy ontologies and other ontologies.

In this chapter, we propose a common anatomy reference ontology (CARO), which is designed to serve as a standardized, generic structural classification system for anatomical entities. We also propose a standardized set of relations for use in building anatomy ontologies, extending the set of relations already defined as part of the OBO Relations Ontology [17]. By necessity, this proposal also begins to address the key issue of representation of homology between anatomical types in the context of anatomy ontologies.

This chapter summarizes progress on creating CARO, drawing on conclusions reached during an anatomy ontology workshop held in Seattle, WA, in September of 2006 sponsored by the National Center for Biomedical Ontology.[3]

16.2 What is CARO?

CARO is an ontology of common anatomy. At its core is a single, structural classification scheme based on that developed by the Foundational Model of Anatomy (FMA), a well established ontology of human anatomy [11] – see also Rosse and Mejino, in this volume – which adheres to the principles laid out by the OBO Foundry.[4] CARO has also adopted the FMA policy of single inheritance. This policy is based principally on the empirical observation that ontologies that allow multiple

[2] In keeping with the nomenclature of Smith et al. [18], we prefer the term 'type' to 'class'. Ontologies contain terms that refer to types of things in the real world. A type should not be confused with its instances. For example, a human anatomy ontology might contain the term 'foot'. This refers to the type *human foot*, of which your left foot is an instance. The collection of all such instances is the extension of the corresponding type.

[3] http://bioontology.org/wiki/index.php/Anatomy_Ontology_Workshop

[4] http://www.obofoundry.org/

inheritance, while easier to build, are marked by characteristic errors, which generally result from the use of multiple classification schemes within a single ontology, leading to what has been called '*is_a* overloading'. This can be avoided by utilizing *genus-differentia* definitions of the terms in ontologies, in which each type is specified as a refinement (via some *differentia*) of an existing more general type (the *genus*, i.e. the corresponding parent type, in the *is_a* hierarchy). Definitions following this form are typically written along the lines of 'An S *is_a* G which D'. This provides unambiguous definitions that can be applied consistently and leads, if done properly, to clean classification hierarchies in which all types have a single (*is_a*) parent and all children of a given type are disjoint (so that nothing can be an instance of both a type and its sibling).

CARO provides relations and the definitions for high-level anatomical types for canonical anatomies. A canonical anatomy gives an account of the 'prototypical' composition of the members of a given species.[5] This simplifies the task of constructing anatomy ontologies, because information captured in them, for example pertaining to part and location relationships, can differ radically in non-canonical types. Scientific communities have different perspectives on what constitutes canonical anatomy. Biologists working on model organisms generally have a standard strain or strains that are considered 'wild-type' for their chosen species. Within medicine, canonical anatomy is a generalization deduced from qualitative observations that are implicitly sanctioned by their accepted usage by anatomists [12, 18]. Defining canonical anatomy is even more problematic in the context of evolutionary biology, where natural variation within a species is often the object of study. Taxonomists therefore utilize voucher or 'type' specimens to define what is representative for a given species.[6] Extensions of CARO to enable integration with the disease ontology (DO) or other ontologies representing pathology or non-canonical anatomy can be accomplished in due course; but such integration will be unfeasible except on the basis of a foundation of canonical anatomy in relation to which relevant deviations can be defined.

CARO includes structural definitions of many generic anatomical types such as cell, portion of tissue, complex organ, anatomical system, and multicellular organism (see appendix for a complete list), organized in an *is_a* hierarchy. *Part_of* and other relations between these types will also be represented. CARO thereby provides a standardized reference ontology on which to build single or multi-species anatomy ontologies or from which to reorganize existing ontologies. This can be achieved by using a clone of CARO to create upper-level types for a single or multi-species ontology. As part of a single or multi-species ontology, the cloned types will refer to anatomical types in the species or taxon in question. Each of these types cloned

[5] For a more detailed analysis see Chapter 14.

[6] International Commission on Zoological Nomenclature, International Code of Zoological Nomenclature online, chapter 13: The type concept in nomenclature, Article 61. Principles of Typification. http://www.iczn.org/iczn/index.jsp

from CARO will have an *is_a* relationship to the corresponding CARO type, and will inherit from the latter its definition.

The CARO types *cell* and *cellular component* are potential root nodes for two existing non-species-specific anatomy ontologies: GO *cell component* and OBO *cell type*. Work is already under way to coordinate definitions and type names that are common to CARO and the latter ontologies, and definitions in all three ontologies will cross-reference each other.

A structural classification alone is not sufficient for the complete representation of anatomy. Other classification systems required for this task include an ontology of functions applicable to anatomical structures and an ontology of phenotypic qualities such as shape (see Figure 16.1). Types from ontologies of function and quality can be used in conjunction with CARO types to build combined anatomy ontologies for single species with multiple inheritance 'views'. For example, components of the immune system are grouped based on the function 'body defense'; they are not part of some single structure or group that can be structurally defined in CARO. Some suitable ontologies of functions are already in existence or are planned (GO Molecular Function [5]; FMP [3]). However, it may be necessary to supplement these ontologies with others still to be created.

Anatomical types classified under CARO can also be linked to types representing biological processes in which they participate, such as those found in the Biological Process Ontology (GO) or in developmental stage ontologies (see Section 16.6). The formalism for combining definitions of types from different parent ontologies in a definition follows the genus and differentia methodology described earlier.

CARO is an ontology of independent anatomical continuants. Continuants have a continuous existence through time. Dependent continuant entities are things that inhere in independent continuant entities such as qualities and functions. Occurrents (processes) have temporal parts which unfold in time (every occurrent depends on one or more independent continuants as its participant or bearer). The prefixes shown in parentheses in Figure 16.1 refer to ontologies that are either under development (FMP, RnaO, PrO) or are available at OBO web site.[7]

16.3 CARO Structure and Definitions

At time of publication, the first version of CARO is under active development. A CARO listserve and wiki track discussion of the ontology and related subjects. CARO can be downloaded in obo and owl formats.[8]

The CARO types and definitions are based on the topmost nodes of the FMA (see [11]; and also Rosse and Mejino, elsewhere in this volume). The top levels of

[7] http://obo.sourceforge.net/browse.html
[8] http://obo.sourceforge.net/cgi-bin/detail.cgi?caro

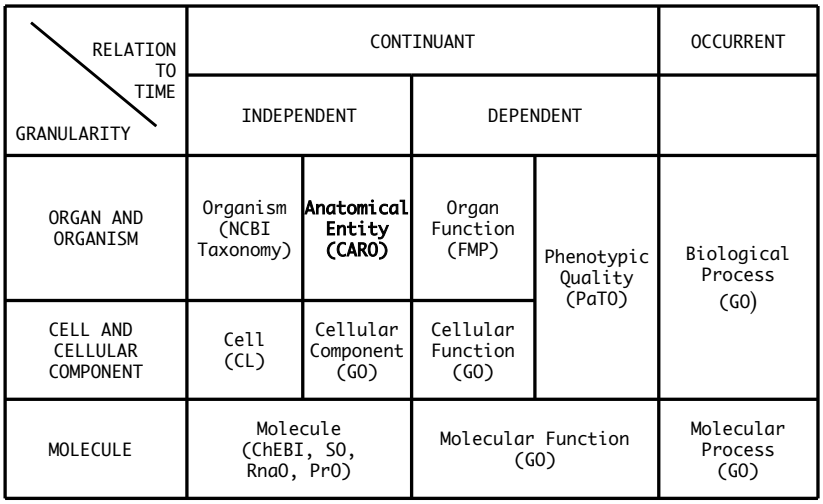

Fig. 16.1. Coverage of species-independent ontologies relevant to biology

the FMA provide a rich set of abstract structural classifications that take into account qualities such as dimensions and contiguity and cover many levels of granularity from whole organism down to cell parts. All of these characteristics have made the FMA an ideal starting point for CARO. However, many of the FMA type definitions are not applicable to all species; some are mammal-specific, some are human-specific, and some are specific to only adult humans. The definitions of these types have been generalized in CARO to be inclusive of more species. Organismal domain specialists will be required to validate the CARO types, in much the same way that human anatomists were required to build and validate the FMA. In addition, the FMA is incomplete in its treatment of developmental structures and developmental relations. Because the representation of developmental anatomy in ontologies is central to the functioning of multiple model organism databases, we have begun to extend the CARO classification scheme to fill this gap. Figure 16.2 shows the taxonomy of the types in CARO. At the end of this chapter we have appended a table that lists all types of CARO including their definitions. Definitions which have been modified from those used by the FMA for use in CARO are discussed below.

16.3.1 Representing Granularity

In order to represent different levels of granularity in CARO, the appropriate types must be specified in such a way as to be applicable across all taxa. The FMA has a well developed system for classifying structural types according to a hierarchy of granularity. Each level of the hierarchy defines the basic building blocks for the level above; for example, portions of tissues are defined as aggregates of cells. However, because the FMA applies only to human anatomy, the FMA developers have used

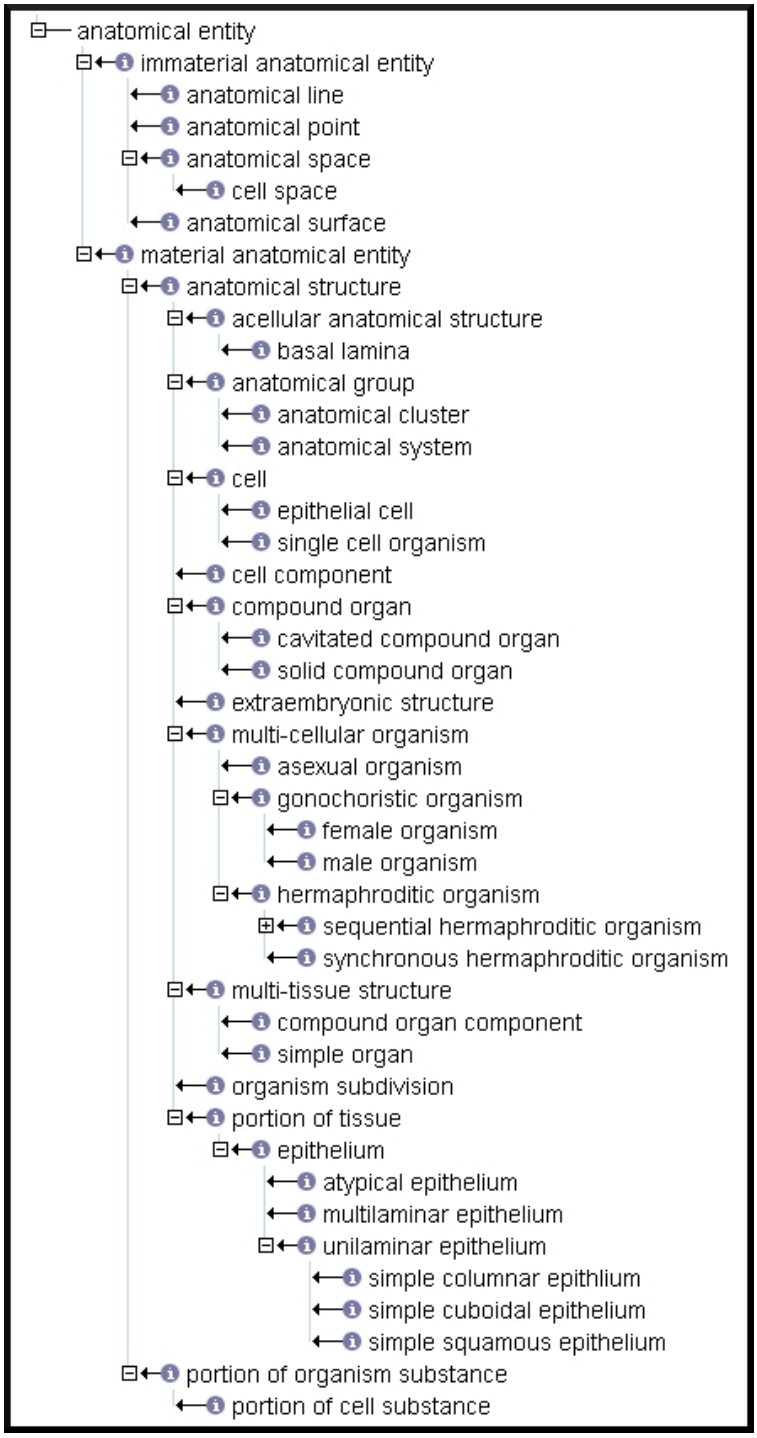

Fig. 16.2. The taxonomy of CARO.

both this bottom up definition of structural types along with a human-specific top down naming system: a cardinal organ part is made up of multiple portions of tissues and an organ is made up of multiple cardinal organ parts. The term 'organ' in the FMA scheme is therefore restricted to structures with a high level of granularity. We have retained this scheme, but have renamed 'cardinal organ part' as *multi-tissue structure* and redrafted the definition so that it also applies to aggregates of portions of tissue that are not themselves part of compound organs. This results in two subtypes of multi-tissue structure. The first, *simple organ*, is representative of many structural units in anatomically simpler organisms and during the development of more anatomically complex organisms. The second, *compound organ component*, refers to discrete multi-tissue structures found within compound organs.

In order to accommodate anatomical structures which are comprised of other anatomical structures of varying levels of granularity, we propose the type *anatomical group*. The subtypes of anatomical group are *anatomical cluster* and *anatomical system*, which permit classification of structures connected either directly or distally. In contrast to an anatomical cluster, the major elements of an anatomical system are discrete, localized anatomical structures of any granularity, or anatomical clusters of varying granularity, distributed across an organism. It has components that while connected, are not adjacent to each other and are separated by intervening structures that are not part of the system. Particularly illustrative examples are the nervous system, the vascular system of vertebrates and the tracheal tree of arthropods. In these examples, the system is in the form of trees or networks that are woven into the fabric of other tissues and organs. The type *anatomical group* and its children allow representation of systems or clusters of anatomical structures for all organisms, where the component parts may vary in their degree of granularity.

Portion of tissue: The term 'tissue' is used sometimes as a mass noun (compare: 'luggage', 'sugar') in such a way as to refer ambiguously to indeterminate amounts of cellular material. We prefer *portion of tissue* (a count noun analogous to 'suitcase' or 'sugar-lump') to make it clear that the term refers unambiguously to a single discrete structure. In addition, we have altered the definition to make 'cells of one or more types spatially arranged in a characteristic pattern' one of the defining features of tissue, rather than 'similarly specialized cells' as we believe this to be more inclusive of different taxa and of developing structures. 'Characteristic' is used to signify that each type of portion of tissue is marked by a distinctive pattern of organization of cells of distinctive types.

16.3.2 Defining Organism Subdivisions

Definitions based on the level of granularity are not sufficient to define all types of anatomical structure. Some types need to be defined as divisions of a whole organism. The segmental organization of the anterior-posterior body axis in arthropods and annelids provides a particularly clear example. Segments are not defined by their level of granularity (e.g. portions of tissue, multi-tissue structures, etc.), but by

morphological boundaries distributed along the anterior-posterior axis of the animal. However, within a particular taxonomic group, it may be possible to develop specific definitions of divisions of the whole organism that specify the granularity of these regions as well as defining them in relation to other such divisions. For example, the FMA's definition of cardinal body part subtypes (head, neck, trunk and limbs) is defined relative to the skeletal system. Because the particular ways that organisms are divided up differs between taxonomic groupings, we have added a generic node in place of 'cardinal body part', *organism subdivision*. This can be used as a parent term for more detailed definitions, including specification of granularity if appropriate, in more taxonomically restricted anatomy ontologies.

16.3.3 Cross-ontology Coordination of CARO Types

A number of types in CARO are present in other ontologies, such as the Gene Ontology Cellular Component (GO CC), and the Cell Type ontology (CL) (see Table 16.1). Specifically, these types represent integration of different levels of anatomical granularity. Coordination of definitions between the GO CC, the CL, and CARO ontologies has begun, and these types will be linked via cross-references.

Table 16.1. CARO types and their corresponding types in other OBO ontologies

CARO	other OBO ontologies
acellular anatomical structure	GO:0044421 *extracellular region part*
cell	GO:0005623 *cell* and CL:0000000 *cell*
epithelial cell	CL:0000066 *epithelial cell*
cell component	GO:0044464 *cell part*
basal lamina	GO:0005605 *basal lamina*

16.3.4 The Organism Types

We include the whole organism as an anatomical structure to allow the formulation of part relations of sexually dimorphic anatomical structures. For example, humans have as parts gonads, but only male humans have testes. Different life strategies for reproduction have different corresponding anatomical structures, requiring that these organism types be defined in CARO.

16.4 Developing Structure Types

Prior to extensive morphogenesis and differentiation, most developing structures are sufficiently simple that they can be defined as a subtype of the CARO type *portion of tissue*. In some cases, types originally defined for adult structures are clearly applicable to developing structures. For example, the regions of the imaginal discs of

Drosophila that will develop into adult appendages have a structure consistent with our definition of *columnar epithelium*. However, other developing tissues share many but not all of the qualities of mature tissues. For example, many tissues of the early *Drosophila* embryo fit the definition of *epithelium* except that they lack a basal lamina. For this reason, the number of generic structural types will be expanded in future versions of CARO to ensure applicability to developing tissues.

Our system also allows the gradual increases in granularity that occurs during development to be captured in a consistent fashion. As development proceeds, developing structures of different granularity levels are formed. As they do, such structures can be reclassified from *portion of tissue* to *multi-tissue structure*, etc. Use of structurally classified developmental types to curate gene expression and phenotypic data will make it possible to look for genes common to the development and maintenance of particular structural types and to the transitions from one structural type to another.

These generic structural types will provide a basic structural classification of developing structures. However, many important details of structural types specific to a single species or taxonomic group will need to be captured in the relevant leaf nodes (the lowest nodes) of species-specific anatomy ontologies. These details can be formalized by referencing structural qualities specified in the Phenotype Attribute and Trait Ontology.[9]

Structural classification is limited in its ability to capture some of the dynamic structural changes which are important to developmental biologists. Specifically, they are interested in defining and classifying portions of developing tissue. CARO cannot provide terms that refer to specific regions of portions of tissue that do not have a structural differentia, but we think it important to specify how this might best be achieved in species-specific or multi-species anatomy ontologies built using CARO as a template. In the following we will first discuss an example and afterwards present the template that allows us to define structures by shared cell fate.

Developmental biologists traditionally define and name portions of tissue, at least in part, on the basis of some shared fate: lens placode, limb field, limb bud, fat-body primordium, and so on. The boundaries of these regions delimit groups of cells that are precursors of some specific type or types of anatomical structure. For example, each of the pair of heart primordia in a zebrafish embryo consists of all the members of a connected group of heart precursor cells, and the *Drosophila* wing pouch consists of all members of a connected group of cells that give rise to the wing. This can be made explicit by the following definition:

d 13 *x is a wing pouch if and only if:*

[9] http://www.bioontology.org/wiki/index.php/PATO:Main_Page

1. *x is a portion of columnar epithelium such that some cells that are part of x are ancestors of some cells that are part of some instance of the type wing; and*
2. *for all y, z: if y is a cell that is part of x and y is the ancestor of the cell z, then there is some type C and some instance c such that c is an instance of C, z is part of c and (either C is identical with the type wing or wing develops_from C).*

The underlying template of this definition is:

d 14 *x is a P if and only if:*

1. *x is an instance of Q such that some cells that are part of x are ancestors of some cells that are part of some instance of the type D; and*
2. *for all y, z: if y is a cell that is part of x and y is the ancestor of the cell z, then there is some type C and some instance c such that c is an instance of C, z is part of c and (either C is identical with the type D or D develops_from C).*

In our example P is the developing type *wing pouch*, Q is the structurally defined supertype *columnar epithelium*, and D is the 'mature' type *wing*. The details of this formalization ensure that it is compatible with the apoptosis of cells that are part of precursor structures during development and can apply to precursor anatomical structures where cell division has ceased but which have yet to differentiate.

In order to apply this approach to structures that are the precursors of multiple later types we need to generalize the definition. Let P again be the developing type, Q the structurally defined supertype, and let S be a set of types of compound organs, multi-species structures, and (maximal) portions of tissue. (S is the set of types of entities that the instances of P develop into.) We now define:[10]

d 15 *x is a P if and only if:*

1. *x is an instance of Q such that for every element D of S the following holds: some cells that are part of x are ancestors of some cells that are part of some instance of D; and*
2. *for all y, z: if y is a cell that is part of x and y is the ancestor of the cell z, then there is some type C and some instance c such that c is an instance of C, z is part of c and (either C is an element of S or there is some element D of S such that D* develops_from *C).*

Note that the differentia of this definition schema distinguishes precursor tissues from other portions of developing tissues that do not consist of a group of cells sharing some fate. Hensen's node in the chicken embryo, for example, contains different precursors at different stages of gastrulation, and does not delimit a connected group of cells sharing some particular fate [14].

The definition schema 15 provides a template for definitions of types of precursor tissues, which can be used in species specific ontologies. As mentioned above,

[10] Definition schema 15 is a generalization of schema 14, since schema 14 is the consequence of schema 15 if we assume that $S = \{D\}$.

this approach is especially useful in cases where developing types cannot be defined on a purely structural bases, because the precursor tissues are not yet mophologically distinct from their surroundings, but have been experimentally defined. The approach also provides a way to define germ-layers, mesoderm, ectoderm and endoderm according to the classes of mature structure whose precursor cells they contain. Finally, as mature structures are named in these definitions, it is possible to use this information to group developing structures according to what they will develop into.

16.5 Relations in CARO

An ontology is a controlled vocabulary that encapsulates the meanings of its terms in a computer parsable form. An anatomy ontology consists of statements composed of two kind of terms, denoting types and relations, respectively. Typically such statements involve two type terms A and B, so that they are of the form: A *rel* B. Relations commonly encountered in anatomical ontologies include the *is_a* relation, indicating that one type is a subtype of another, and the *part_of* relation, indicating that every instance of the first type is, on the instance level, a part of some instance of the second type. Examples of use include *pancreas is_a lobular organ* in the FMA and *cell nucleus part_of cell* in the GO Cellular Component ontology. However, anatomical ontologies are by no means limited to these two relations; the FMA employs a large number of spatial relations [11][11] and ontologies that encompass entities at various developmental stages typically link types using relations such as *develops_from*, as in the OBO Cell Type ontology (CL) and in anatomical ontologies for model organisms such as fly and zebrafish.

Relations play an essential role in ontologies, since they are the primary bearer of semantic content (see Chapter 14). To ensure a consistent use of terms that denote relationships within and across ontologies, it is important to agree on shared, unambiguous definitions of these terms. These definitions utilize the dependence of relationships between types (e.g. *cell nucleus* and *cell*) on the relationships between instances of these types (e.g. concrete cell nuclei and the cells which contain them), as is discussed in detail in the Chapters 14 and 15 of this book. In this section, we will discuss the extension of the OBO Relations Ontology [17] to provide relations that are necessary for CARO and species-specific anatomies. This extension comes in different flavors: (a) in some cases, we need to add new relations to capture important aspects of anatomical entities, (b) in other cases, we need to add new relations that further specify existing ones in order to better represent the dynamic changes within developing organisms, and (c) we need to consider relations that link anatomy ontologies to other ontologies.

[11] Also see 'spatial association relationship' at:
http://fme.biostr.washington.edu:8089/FME/index.html

16.5.1 Defining *develops_from*

The OBO Relations Ontology covers the most important relationships for anatomy ontologies, but lacks explicit definitions of many spatial relations that it would be desirable to include. Some of these are discussed in chapter 15 of this book. Further, for CARO to provide a representation of developmental anatomy, we need to define a relationship that represents the various ways that anatomical structures change through development. We lack a single, transitive relationship that can represent the transformation, fission and fusion of developing structures over time. Here we outline the relationship *develops_from*, which fulfills these criteria. In order to define *develops_from* we need to distinguish two cases. In the first case, some entity changes its properties but remains numerically identical; for example, if an adult develops from a child, then the adult will have different properties (e.g. a different weight and height) but it will be still the same individual. In contrast, if a zygote develops from a sperm cell and an ovum, then the zygote is not identical with either; but the zygote arises from the sperm cell and the ovum. These two relations are used to define the type level relationships *transformation_of* and *derives_from*[12] in the OBO Relations Ontology. Since it is often unknown during development whether one structure arising during development is a transformation of another or whether some portion of a structure arises from another one, we need a *develops_from* relation which covers both cases.

More formally, the *develops_from* relationship is defined as follows:[13]

d 16 *C develops_from D if and only if, for any x and any time t, the following holds: if x instantiates C at time t, then*

1. EITHER *for some time t_1, x instantiates D at t_1 and t_1* **precedes** *t, and there is no time interval t_2 such that x instantiates C at t_2 and x instantiates D at t_2;*
2. OR *for some time t_1, there is some y such that y instantiates D at t_1 and x* **arises_from** *y.*

The relation **succeeds** is defined with the help of the relations **buds_from** and **arises_from**. Note while *develops_from* is a relationship between types, **precedes**, **buds_from**, **succeeds**, and **arises_from** hold between instances.

d 17 *x* **arises_from** *y is defined recursively in the following way:*

1. *if x* **succeeds** *y, then x* **arises_from** *y;*
2. *if x* **buds_from** *y, then x* **arises_from** *y;*
3. *if x* **arises_from** *y and y* **succeeds** *z, then x* **arises_from** *z;*

[12] To avoid confusion with the very different meaning of 'derives from' in an evolutionary context, we plan to rename this type level relationship '*arises_from*'. *The corresponding instance level relationship is referred to as '**arises_from**' in the following text.*

[13] These definitions, and the definitions below, are provided for the sake of technical completeness. They will not play any role in the actual use of CARO in day-to-day annotation and information retrieval purposes.

4. *if x* **arises_from** *y and y* **buds_from** *z, then x* **arises_from** *z;*
5. *x* **arises_from** *y holds only because of (1)-(4).*

With other words **arises_from** is the transitive closure of **buds_from** and **succeeds**. The relations **succeeds** and **buds_from** are defined in the following way.[14]

d 18 *x* **succeeds** *y if and only if*

1. *x and y are instances of the type anatomical entity; and*
2. *x begins to exist at the same instant of time at which y ceases to exist; and*
3. *there is some anatomical structure z such that z is* **part_of** *y when y ceases to exist and z is* **part_of** *x when x begins to exist.*

d 19 *x* **buds_from** *y if and only if*

1. *x and y are anatomical entities; and*
2. *at no time t, x is* **part_of** *y* **at** *t; and*
3. *there is some anatomical structure z such that z is* **part_of** *y immediately before x begins to exist, and x* **succeeds** *z; and*
4. *x continues to exist for some interval of time from the point when y begins to exist.*

16.5.2 Defining Time-Restricted Part Relationships

The parthood relations as defined in the OBO Relations Ontology [17] do not adequately represent some dynamic aspects of developmental anatomy. In particular, the relationships *has_part* and *part_of*, both apply at all stages: C *has_part* D means that every C, regardless of stage, has some D as instance-level part. The *Drosophila* anatomy ontology, however, contains types of neuroblasts that are part of the ventral nerve cord primordium (VNC). As these neuroblasts divide, more types become identifiable – at stage 9 there are 10 types but by stage 11 there are 34 [1]. We cannot capture the part relationship between these cell types and the VNC primordium using the *has_part* relation, because this would imply that all instances of the VNC have instances of each of these neuroblast types as a part at all stages. Similarly, the relation *part_of* also applies irrespective of stage. We can solve this dilemma by defining versions of *part_of* and *has_part* which are applicable only during the stages in which both partners in the relationship exist. The formal definitions of these relationships are:

d 20 *C time_restricted_part_of D if and only if the following holds for any x and any time t: if x instantiates C at time t, then there is a y such that*

1. *for some time* t_1*, y instantiates D at* t_1 *and x* **part_of** *y* **at** t_1*; and*
2. *for all times* t_2*: if x* **exists_at** t_2 *and y* **exists_at** t_2*, then x is* **part_of** *y* **at** t_2.

[14] The observant reader will notice that these definitions are less rigorous than the previous ones. For a full logical analysis of '**buds_from**' and '**succeeds**' we would need to spell out the underlying temporal theory; which is beyond the scope of this chapter.

d 21 *C time_restricted_has_part D if and only if the following holds for any x and any time t: if x instantiates C at time t, then there is a y such that*

1. *for some time* t_1*, y instantiates D at* t_1 *and y* **part_of** x **at** t_1*; and*
2. *for all times* t_2*: if x* **exists_at** t_2 *and y* **exists_at** t_2*, then y is* **part_of** x **at** t_2.

16.5.3 Relationships Linking Separate Ontologies

As mentioned above, the structural classification of anatomical entities in CARO is separate from the treatment of functional classification and of homology between anatomical entities across different species. In order to record function and homology information, the anatomical types within a species-specific anatomy ontology need to be linked to types in other ontologies, and the necessary relations – including *has_function* and *homologous_to* – will be added to the OBO Relations Ontology in due course. We discuss relations between developmental stage and anatomical types in the following section. Note that the spatial relations and the *develops_from* relation mentioned above are relations that are used within a given anatomical ontology. In contrast relations such as *has_function*, *homologous_to*, *starts_during* and *ends_during* are relationships that link types across different ontologies. Similarly, *is_a*, too, can link types across different ontologies, as for instance when we make the assertion that *mouse compound organ is_a* CARO*:compound organ*.

16.6 Representing Stages

Development can be considered a process that *has_participant* [17] whole organism. For any single species, events during development occur in a predictable order. However, the precise timing of these events is dependent on environmental conditions. Developmental biologists traditionally measure progress through (the process of) development relative to the occurrence of some standard series of events which can be easily and reliably scored [2, 10]. A standard table of development divides the process of development into stages, each delimited by a pair of events, and it describes key events occurring within each stage.

For some organisms, not only is the order of events consistent, but under standard laboratory conditions their timing relative to a reference event (e.g. fertilization) shows little variation. In these cases it is possible to define stages in terms of the period of time that elapsed since the reference event. This method of defining stages is particularly useful if no easily score-able morphological stage criteria are available. For example, in the zebrafish, early stages are often referred to either by morphological criteria or by time since fertilization, while the later stages are referred to exclusively by time since fertilization [8].

As stage series are necessarily species-specific, ontologies representing individual stage series have to be constructed for each species. Minimally, a stage ontology

will contain types for the stages that make up a standard table of development. The relative timing of these stages can be recorded using the relation *preceded_by* [17]. Stages can be grouped together into super-stages, or divided into sub-stages, with the latter having a *part_of* relationship to the stages themselves, which are in turn *part_of* super-stages. While stage series are species-specific, many of the developmental processes described in standard tables of development are not. Information about the relative timing of developmental processes described in each standard table of development can be captured within species-specific stage ontologies. The relative timing of these processes to each other and to stage boundaries can be recorded using the relations *part_of*, *preceded_by* and an additional relationship *simultaneous_with*[15]. Linking these to relevant GO types such as cellularization (see Figure 16.3) will facilitate reasoning between species-specific stage ontologies.

We propose that these species-specific stage ontologies be used to record the periods of development during which anatomical entities exist by using the relationships *starts_during* and *ends_during* (a formalized version of the strategy used by ZFIN). These relationships link anatomy ontology types to appropriate types in the stage ontology. This will give a crude resolution to records of timing: the existence of X begins some time during stage N and ends some time during stage N′. The temporal resolution of these links could be improved, as data allows, in two ways. Where some standard system of substages has been defined, we can simply make *starts_during* and *ends_during* links to these substages. Alternatively, we can refine our record of the timing of the beginning or end of existence of an anatomical entity by instantiating these as events within the stage ontology and using *preceded_by* relations to processes beginning or ending within a stage (see Figure 16.3).

16.7 CARO Depth and Application

The question of CARO depth is closely related to its utility in building new anatomy ontologies. The top-level types in CARO together with the relationships defined above can be used to structure application anatomy ontologies. However, the types in CARO are very generic relative to the types commonly defined within a species-specific anatomy ontology. This is because it is very difficult to further subtype CARO and remain within the bounds of disjoint structural definitions. For example, the compound eye of a *Drosophila* and the camera-lens eye of a human have little in common structurally, making it unlikely that the type *eye* would be included in CARO (though these types might be grouped, outside of CARO, using the function ‘to see’). However, it may be possible to achieve a disjoint set of structural definitions for particular monophyletic groups within multi-species anatomy ontologies.

[15] To be defined in a future publication.

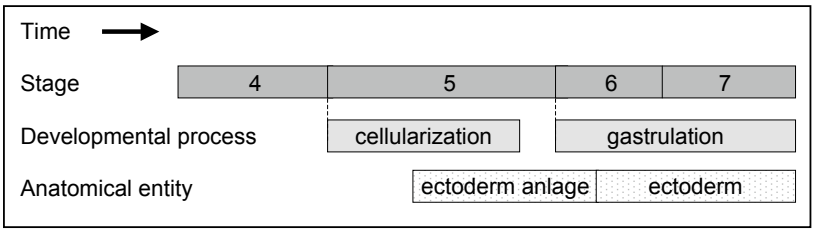

Fig. 16.3. Relationship between anatomical entities, stage, and process. For each species, an ontology will be constructed containing types for stage and developmental process in a single ontology of occurents. Anatomical entities are contained in a separate ontology of continuants. The ends of each bar represent events for which relative timing can be recorded using the relations *preceded_by* and *simultaneous_with*. These ordering relations will be used in conjunction with *starts_during* and *ends_during* to define the period during which an anatomical entity exists. This example illustrates ectoderm development in the *Drosophila* embryo, wherein the ectoderm anlage *starts_during* stage 5, the ectoderm anlage *ends_during* stage 6, the ectoderm *starts_during* stage 6, the process gastrulation *preceded_by* cellularization, and gastrulation *simultaneous_with* stage 6 and stage 7.

A number of projects aim to generate anatomy ontologies of multiple taxa. In particular, the Cypriniformes Tree of Life (CToL)[16], the plant ontology[17], as well as the amphibian[18], and Hymenoptera[19] anatomy ontologies. As in the case of species-specific anatomy ontologies, multi-species anatomy ontologies can also clone the CARO types for use as their topmost nodes. Within a multi-species anatomy ontology, a type that satisfies the definition of a CARO type will have an *is_a* relation to the CARO type with the *differentia* of a taxon rather than a species. For example, for the cypriniform fish anatomy ontology, the cypriniform type *compound organ is_a* CARO:*compound organ*, with the differentia being that it is a compound organ of a type found in Cypriniformes. CARO can in this way be used as a template for multi-species anatomy ontologies as well as for species-specific ones.

Currently, many ontology developers use an existing ontology when building a new one (as CARO itself is modeled on the FMA). For example, the zebrafish anatomy ontology has been used as a template for both fish and amphibian multi-species ontologies. This is because the zebrafish anatomy ontology refers to anatomical structures that evolved within chordates – a post-anal tail evolved at the level of Chordata, the lateral line system evolved at the level of Craniata, jaws evolved at the level of Gnathostomata, and bone at the level of Vertebrata (Figure 16.4).

Within multi-species anatomy ontologies it is necessary to specify in which organisms the anatomical entities are applicable. This can be accomplished with

[16] http://www.nescent.org/wg_fishevolution
[17] http://www.plantontology.org
[18] http://www.morphologynet.org
[19] http://ceb.scs.fsu.edu/ronquistlab/ontology/wiki/index.php/Main_Page

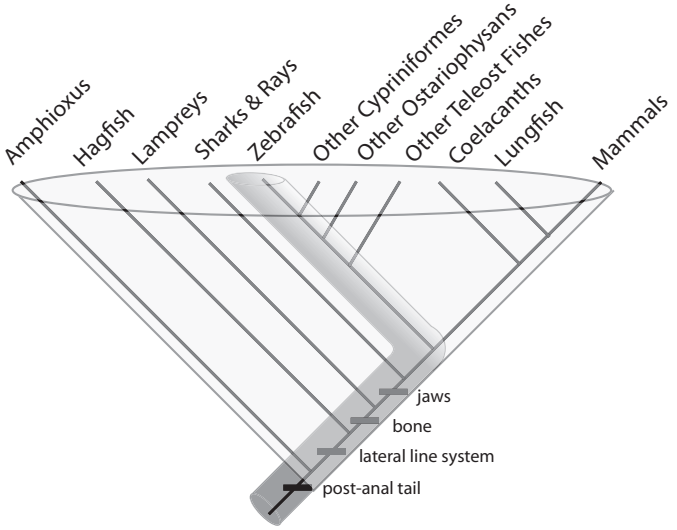

Fig. 16.4. Species-specific anatomy ontologies contain types applicable to more diverse taxa. The zebrafish anatomy ontology (inner lighter cylinder) includes terms referring to features that evolved at various times in the chordate lineage. This ontology could be expanded to include anatomical structures found in all vertebrates (entire cone).

the relation, *part_of_organism*, proposed by the CToL-ZFIN working group to link anatomical entities to taxa within a taxonomy ontology. Similarly, the types in CARO are not applicable to all organisms. For example, diploblastic animals such as cnidarians (a phylum that includes jellyfish and sea anemones) lack compound organs (a proposed CARO term) while sponges may have no distinct multi-tissue structures at all [6]. CARO classes could also be linked to a taxonomy ontology to indicate which classes are applicable at various taxonomic levels. The purpose of cross-referencing multi-species anatomy ontologies and CARO to a taxonomic ontology would be to provide a user with choice of appropriate types. A similar method has been proposed to limit classes to specific taxa in other species-independent ontologies such as the GO or the CL (Waclaw Kusnierczyk, personal communication). It is important to note that cross-referencing anatomy and taxonomy ontologies in this manner does not specify homology.

16.8 Representing Homology

Methods for recording homology between types in anatomy ontologies are extremely important both to provide resources for evolutionary biologists and for the development of tools for inter-species inference regarding the molecular basis of morphological phenotypes or traits. Structures (including genes) are homologous if they evolved from some structure in a common ancestor, and homology implies genealogical descent as the vehicle of transfer of information. Homology must be addressed within

the context of multi-species anatomy ontologies because of the very nature of how anatomical structures evolve. The reason anatomical types are structurally or functionally similar, and therefore classified together in some ontology, may be because they are evolutionarily related. However, many well documented counter examples exist. For example, both zebrafish and humans have a skull bone named the parietal bone, and another named the frontal bone. These could be grouped in an ontology on the basis of position within the skull and name; but there is good evidence that the parietal bone in humans is homologous to the frontal bone in zebrafish [7, 13]. Thus, one cannot assume homology based on structural similarity or name.

We propose that homology information be captured independently of both structure and function information. Specifically, statements of homology are hypotheses and require evidence (codes) and attribution. This is particularly important to evolutionary biologists creating phylogenies, where different evidence is often used to generate different phylogenetic views. In light of this need to capture homology, a new relationship, *homologous_to*, is proposed to be included in the OBO Relations Ontology, but its definition is still under discussion. The ontological implications for this new relationship are as yet untested. For instance, if two structures are deemed homologous, is this information transitive down *is_a* chains? Can two structures be homologous if none of their parts are homologous? Erwin and Davidson [4] have suggested that the regulatory processes that underlie development may be homologous, whereas the creation of gross anatomical structures is specific to phyla or classes (and may not be homologous). In this respect, it is the processes or functions that are homologous whereas the structures are not.

To establish a homology relation between sister anatomical entities may require the determination of an evolutionary precursor in order to create sister subtypes within a multi-species anatomy ontology. It may prove difficult in some cases to define an evolutionary precursor purely on a structural basis and will require domain experts whose expertise spans large branches of the tree of life. However, it is possible that a function ontology used in combination with homology statements could overcome this difficulty. Multi-species anatomy ontologies will have to reconcile these homology issues with maintenance of disjoint definitions based on structure. It is important to note that even though one intended use for CARO is as a template for building multi-species anatomy ontologies, no homology between types is implied by common treatment within CARO, since CARO types are classified purely on the basis of structural criteria and not on evolutionary history.

16.9 Long Term CARO Goals

One of the long-term goals of CARO is to provide the source of standardized representations of anatomical types used in creating composite types of the kind found in ontologies such as the GO's Biological Process ontology. Like CARO, GO is cross-species, describing types of biological process that occur across a wide variety

of species, encompassing types such as *heart development* and *neural tube closure*. Like CARO, GO is also canonical – it describes the features of typical, wild-type instances. At the present time, GO does not contain explicit references to types from an anatomical entity ontology. Instead, rough definitions of types such as *heart* and *neural tube* are 'embedded' inside the definitions of the corresponding GO types. This leads to redundancy, duplication of effort, inconsistency and a poor basis for cross-domain inference.

Once CARO is in use as a template for species-specific or multi-species anatomy ontologies, types from these ontologies along with their taxonomic reference can be referenced by the GO. GO will retain types such as *neural tube closure*, but the corresponding definitions can refer to definitions taken from CARO or from one of the multi-species or single-species anatomy ontologies created in a way which will allow the ontologies to be kept synchronized [9].

While the primary axis of classification in CARO is structural, not functional, this does not mean that CARO ignores function. Rather, CARO insists that function be treated as a separate *orthogonal* ontology. Instead of stating that *verterbrate eye is_a sense organ* as we may do in a mixed classification, we instead state that *vertebrate eye has_function visual perception*, with the *is_a* parent of verterbrate eye being the appropriate structural supertype (i.e. cavitated compound organ). Separating structure from function in this way leads to cleaner ontology design, with each type having a single *is_a* parent. At the same time, this methodology still allows for cross-ontology queries, such as 'find all genes active in seeing structures'. The organismal function ontology that will be used in conjunction with CARO or other anatomy ontologies is yet to be developed. Like CARO, this ontology will adhere to OBO Foundry principles and be itself placed in the OBO Foundry.[20] Many of these functions will be realized in biological processes of the kind found in the GO, so this ontology will be developed in coordination with the Gene Ontology Consortium.

One final consideration is that CARO compliance can be exploited to help build phylogenetic views of a given set of taxa. Since all species-specific and multi-species anatomy ontologies will have *is_a* links to CARO nodes, it will be possible to view an assembly of anatomical structures by limiting the taxonomic level. In combination with a set of homology statements, one could build different phylogenies based on different evidence. This is not unlike the current method of creating phylogenies, except that the anatomical structures are named and assigned to taxa in a standardized manner thereby providing links to other relevant data. For example, the development and function of homologous structures in two different species are likely to retain at least some of the molecular mechanisms present in the ancestral structure in their most recent common ancestor. CARO should in this way prove a useful organizational tool to facilitate the inference of molecular mechanisms underlying morphology.

[20] http://obofoundry.org/

Acknowledgements

This work was funded in part by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028. Information on the National Centers for Biomedical Computing can be found at `http://nihroadmap.nih.gov/bioinformatics`. The CToL-ZFIN working group has been supported by the National Evolutionary Synthesis Center. Additional funding came from a grant to FlyBase, `http://flybase.org`, from the National Human Genome Research Institute (USA), Grant 5 P41 HG 000 739. Thanks go to Michael Ashburner, Rachel Drysdale, Monte Westerfield, and Cornelius Rosse for their valuable comments.

References

1. C. Berger, J. Urban, and G.M. Technau. Stage-specific inductive signals in the Drosophila neuroectoderm control the temporal sequence of neuroblast specification. *Development*, 128: 3243-3251, 2001.
2. J.A. Campos-Ortega and V. Hartenstein. *The Embryonic Development of Drosophila melanogaster*. (Second Edition), Springer-Verlag, Berlin, 1999.
3. D.L. Cook, J.L.V. Mejino, and C. Rosse. Evolution of a foundational model of physiology: symbolic representation for functional bioinformatics. *Medinfo*, 11(1): 336-340, 2004.
4. D.H. Erwin and E.H. Davidson. The last common bilaterian ancestor. *Development*, 129: 3021-3032, 2002.
5. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, 34: D322-6, 2006.
6. P.W.H. Holland. Major transitions in animal evolution: a developmental genetic perspective. *American Zoologist*, 38(6): 829-842,1998.
7. M. Jollie. *Chordate Morphology*. New York, Reinhold Books, 1962.
8. C.B. Kimmel, W.W. Ballard, S.R. Kimmel, B. Ullmann, and T.F. Schilling. Stages of embryonic development of the zebrafish. *Dev. Dyn.*, 203: 253-310, 1995.
9. C. Mungall. Obol: integrating language and meaning in bio-ontologies. *Comparative and Functional Genomics*, 5 (6-7), 509-520, 2004.
10. P.D. Nieuwkoop and J. Faber. *Normal Table of Xenopus laevis*. 3rd Ed, 1994.
11. C. Rosse and J.L.V. Mejino. A reference ontology for bioinformatics: The Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36: 478-500, 2003.
12. C. Rosse, J.L.V. Mejino, B.R. Modayur, R. Jakobovits, K.P. Hinshaw, J.F. 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.
13. H. P. Schultze and M. Arsenault. The panderichthyid fish Elpistostege: A close relative of tetrapods? *Paleontology*, 28: 293-309, 1985.
14. M.A. Selleck and C.D. Stern. Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development*, 112(2): 615-626, 1991.
15. J. Slack. *From egg to Embryo* (2nd Ed). Cambridge University Press, 1991.
16. B. Smith. Fiat objects. *Topoi*, 20(2): 131-148, 2001.

17. B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(5):r46, 2005.
18. B. Smith, W. Ceusters, and C. Rosse. On carcinomas and other pathological entities. *Comparative and Functional Genomics*, vol 6, 7-8, 379-387, 2005.

Appendix

The following table contains the types of CARO and their definitions in the order they appear in Figure 16.2.

CARO Definitions

anatomical entity	Biological entity that is either an individual member of a biological species or constitutes the structural organization of an individual member of a biological species.
immaterial anatomical entity	Anatomical entity that has no mass.
anatomical line	Non-material anatomical entity of one dimension, which forms a boundary of an anatomical surface or is a modulation of an anatomical surface.
anatomical point	Non-material anatomical entity of zero dimension, which forms a boundary of an anatomical line or surface.
anatomical space	Non-material anatomical entity of three dimensions, that is generated by morphogenetic or other physiologic processes; is surrounded by one or more anatomical structures; contains one or more organism substances or anatomical structures.
cell space	Anatomical space that is part of a cell.
anatomical surface	Non-material anatomical entity of two dimensions, that is demarcated by anatomical lines or points on the external or internal surfaces of anatomical structures.
material anatomical entity	Anatomical entity that has mass.
anatomical structure	Material anatomical entity that has inherent 3D shape and is generated by coordinated expression of the organism's own genome.
acellular anatomical structure	Anatomical structure that consists of cell parts and cell substances and together does not constitute a cell or a tissue.
basal lamina	Acellular anatomical structure that consists of a thin sheet of fibrous proteins that underlie and support the cells of an epithelium. It separates the cells of an epithelium from any underlying tissue.
anatomical group	Anatomical structure consisting of at least two non-overlapping organs, multi-tissue aggregates or portion of tissues or cells of different types that does not constitute an organism, organ, multi-tissue aggregate, or portion of tissue.

CARO Definitions

anatomical cluster	Anatomical group that has its parts adjacent to one another.
anatomical system	Anatomical group that is has as its parts distinct anatomical structures interconnected by anatomical structures at a lower level of granularity.
cell	Anatomical structure that has as its parts a maximally connected cell compartment surrounded by a plasma membrane.
epithelial cell	Cell which has as its part a cytoskeleton that allows for tight cell to cell contact and which has apical-basal cell polarity.
single cell organism	Cell that is an individual member of a species.
cell component	Anatomical structure that is a direct part of the cell.
compound organ	Anatomical structure that has as its parts two or more multi-tissue structures of at least two different types and which through specific morphogenetic processes forms a single distinct structural unit demarcated by bona fide boundaries from other distinct anatomical structures of different types.
cavitated compound organ	Compound organ that contains one or more macroscopic anatomical spaces.
solid compound organ	Compound organ that does not contain macroscopic anatomical spaces.
extraembryonic structure	Anatomical structure that is contiguous with the embryo and is comprised of portions of tissue or cells that will not contribute to the embryo.
multi-cellular organism	Anatomical structure that is an individual member of a species and consists of more than one cell.
asexual organism	Multi-cellular organism that does not produce gametes.
gonochoristic organism	Multi-cellular organism that has male and female sexes.
female organism	Gonochoristic organism that can produce female gametes.
male organism	Gonochoristic organism that can produce male gametes.
hermaphroditic organism	Multi-cellular organism that can produce both male and female gametes.
sequential hermaphroditic organism	Hermaphroditic organism that produces gametes first of one sex, and then later of the other sex.
synchronous hermaphroditic organism	Hermaphroditic organism that produces both male and female gametes at the same time.
multi-tissue structure	Anatomical structure that has as its parts two or more portions of tissue of at least two different types and which through specific morphogenetic processes forms a single distinct structural unit demarcated by bona-fide boundaries from other distinct structural units of different types.
compound organ component	Multi-tissue structure that is part of a compound organ.
simple organ	Multi-tissue structure that is not part of a compound organ.

CARO Definitions

organism subdivision	Anatomical structure which is a primary subdivision of whole organism. The mereological sum of these is the whole organism.
portion of tissue	Anatomical structure, that consists of similar cells and intercellular matrix, aggregated according to genetically determined spatial relationships.
epithelium	Portion of tissue, that consists of one or more layers of epithelial cells connected to each other by cell junctions and which is underlain by a basal lamina.
atypical epithelium	Epithelium that consists of epithelial cells not arranged in one ore more layers.
multilaminar epithelium	Epithelium that consists of more than one layer of epithelial cells.
unilaminar epithelium	Epithelium that consists of a single layer of epithelial cells.
simple columnar epithlium	Unilaminar epithelium that consists of a single layer of columnar cells.
simple cuboidal epithelium	Unilaminar epithelium that consists of a single layer of cuboidal cells.
simple squamous epithelium	Unilaminar epithelium that consists of a single layer of squamous cells.
portion of organism substance	Material anatomical entity in a gaseous, liquid, semisolid or solid state; produced by anatomical structures or derived from inhaled and ingested substances that have been modified by anatomical structures as they pass through the body.
portion of cell substance	Portion of organism substance located within a cell.