

The Role of Definitions in Biomedical Concept Representation

Joshua Michael, José L. V. Mejino Jr., M.D., and
Cornelius Rosse, M.D., D.Sc.

Structural Informatics Group, Department of Biological Structure,
University of Washington, Seattle, WA

ABSTRACT

The Foundational Model (FM) of anatomy, developed as an anatomical enhancement of UMLS, classifies anatomical entities in a structural context. Explicit definitions have played a critical role in the establishment of FM classes. Essential structural properties that distinguish a group of anatomical entities serve as the differentiae for defining classes. These, as well as other structural attributes, are introduced as template slots in Protégé, a frame-based knowledge acquisition system, and are inherited by descendants of the class. A set of desiderata has evolved during the instantiation of the FM for formulating definitions. We contend that 1. these desiderata generalize to non-anatomical domains and 2. satisfying them in constituent vocabularies of UMLS would enhance the quality of information retrievable through UMLS.

INTRODUCTION

Definitions are conventionally compiled in dictionaries. It has been advocated that definitions should also be incorporated in structured vocabularies, such as controlled medical terminologies (CMT).¹⁻³ The authors of some CMTs have adopted the definitions of established medical dictionaries (e.g., MeSH⁴), whereas in other CMTs definitions are implied by the hierarchical arrangement of terms without further explicit specification. In the process of establishing the Foundational Model (FM) of anatomy^{5,6} we encountered conflicts between dictionary definitions of anatomical concepts and the requirements for a logical and consistent structuring of a symbolic model. We were forced, therefore, to write our own definitions. This process led to the formulation of a set of desiderata for defining anatomical concepts. These requirements evolved over time, based on the insights we gained from extensive data entry, the iterative revision of definitions and the consequent rearrangement of classes within the Anatomy Ontology (AO) component of the FM. Our objective in this report is to illustrate the critical role of logical definitions in the principled representation of a knowledge domain. After contrasting the roles of definitions in dictionaries and ontologies, we illustrate the influence foundational principles have exerted on the development of requirements to be met by

definitions. These definitions, in turn, determined the semantic class structure of the FM.

ROLE OF DEFINITIONS

In dictionaries the unit of information is a term and definitions specify the meanings of terms. The sequence of terms is dictated by the alphabet and the definition of a given term bears no relationship to that of its neighbors. This arrangement satisfies the purpose of dictionaries, which is to define terms in isolation without any explicit relationship to other terms. A dictionary accommodates different meanings of a term (e.g., 'palm' considered in a botanical or anatomical context) by defining such homonyms individually in the contexts in which they are used.

Ontologies (i.e., true inheritance hierarchies) differ from dictionaries in both their nature and purpose. Thus the definitions supplied by dictionaries may be inadequate for the needs of ontology developers. In ontologies the unit of information is a concept, which is symbolically represented by one or more terms. One of these terms may be selected as the preferred name and the others may be associated with the concept as synonyms. The sequence of concepts forms a type hierarchy, which is dictated by the properties shared by groups of concepts. The soundness of this hierarchy depends on the explicit specification of the properties (attributes) that define the essence of concepts, providing the basis on which they may be grouped together or distinguished from one another. The role of definitions in an ontology is, therefore, to specify such defining attributes in a consistent manner, thus assuring their transitive inheritance through a type hierarchy. Consistency in definitions and, therefore, in the classification, requires that a unifying viewpoint (i.e., context) be also specified for concept representation. This context should hold true for the entire ontology. Provided such requirements are satisfied, the position of a concept will enrich its own definition by the definition of all of its parents within the hierarchy. Thus, unlike in a dictionary, a definition of a concept within an ontology is incomplete without that of all of its parents. It has been our objective to satisfy these requirements in the Foundational Model by establishing a rigorous ontology as the backbone of the model. Such a hierarchy, based solely on the *IS-A*

relationship within a consistent context will provide a logical semantic structure for a concept domain and will support inheritance of defining attributes of the domain's concepts.

FOUNDATIONAL PRINCIPLES AND DEFINITIONS

The Foundational Model is a conceptualization of the coherent body of anatomical knowledge. A set of principles declares how this knowledge should be structured and provides the basis for reasoning about anatomical entities⁶. The model is specified by the four-tuple

$$Fm = (AO, ASA, ATA, Mk)$$

where AO = anatomy ontology, which is a type hierarchy constituted by anatomical entities; ASA = Anatomical Structural Abstraction, which specifies the structural relationships of the concepts represented in AO; ATA = Anatomical Transformation Abstraction, which describes the morphological transformations of the concepts represented in AO during the human life cycle (including prenatal development, postnatal growth and aging); Mk = Metaknowledge, which comprises the principles, rules and definitions according to which relationships are to be represented in the model's other three components.

We illustrate how particular foundational principles exert a determining influence on concept definitions at different levels within the AO.

Constraint Principle. Application domains of anatomical knowledge influence explicit or implied definitions of anatomical entities. For instance, in clinical medicine anatomical entities are regarded as sites of disease. In some CMTs this context provides the basis for defining such fundamental anatomical attributes as part-whole relationships.^{7,8} Anatomy education, on the other hand, emphasizes a functional context. This is also the context predominantly reflected in dictionary definitions. Neither of these contexts can, however, support the establishment of a consistent and comprehensive inheritance hierarchy for anatomy. The constraint principle specifies a purely structural context for modeling anatomy: *The conceptualization should model the physical organization (structure) of the body.*⁶ This principle is consistent with the fact that the structure (anatomy) of biological organisms is the concept domain unique to the science of anatomy.⁵ Such structural knowledge provides the foundation for functionally or clinically-oriented biological knowledge domains, including a controlled terminology for the representation of concepts within these domains. Therefore we

formulated our definitions and established the AO in a structural context.

The constraint principle implies that the material objects that constitute the body should have primary importance in the FM (further discussed below). In addition, however, the AO must include the spaces enclosed within and among these objects, as well as the surfaces, edges and points that define the boundaries of the objects. Moreover, structural relationships that exist between these objects must also be represented in the AO as classes of ASA attributes,⁹ along with other attributes that pertain, for example, to the ATA.

Anatomical discourse in any application domain is concerned with much more specific and concrete classes and instances than these broad concept categories. Such high level classes at or near the root of an ontology are at best only implied by dictionaries and conventional knowledge sources. However, it is precisely these high level classes that must be defined explicitly in order to ensure consistency and a logical semantic structure in the ontology.

Formulation of Definitions. How are such definitions to be formulated? If the definitions are to provide the rationale for an "Aristotelian" hierarchy constructed on the basis of inheritance, then it is necessary to classify concepts according to 1. their *genus*, that is, the concept that subsumes the essential (defining) structural attributes shared by all of its descendants in the type hierarchy; 2. the *differentiae*, that is, the structural attributes that distinguish entities immediately descended from the genus. Therefore, the *essence* of a concept is constituted by two sets of attributes; one set necessary to assign the concept to a genus and the other set the *differentiae*, which distinguish it from other members of the genus. A group of entities that share the same set of essential characteristics constitutes a class of the ontology.

To identify the "essence" of the diverse kinds of concepts considered above is a challenging task; all must be regarded as 'anatomical entities'. A more restricted concept than 'entity' will not subsume material objects, spaces, surfaces, lines and points, as well as virtual and other abstract concepts that pertain to the structural organization of the body. Therefore we declared 'anatomical entity' as the root of the AO (Figure 1).

What are the essential characteristics that distinguish anatomical entities from non-anatomical entities? The essence of anatomical entities is that they can be conceptualized only in relation to biological

organisms; furthermore, they are unique among biological entities because they are restricted to the structural organization of these organisms. Non-anatomical biological entities relate to biological processes such as normal and abnormal functions. Consequently, the *genus* of 'anatomical entity' is 'biological entity' because it manifests the essence of all biological entities (namely that they pertain only to biological organisms), and the *differentia* is the restriction to structure. The definition may therefore be written as:

Anatomical entity
is a biological entity,
which *constitutes* the structural organization
of a biological organism, or
is an attribute of that organization.

In view of the fact that the constraint principle enforces a structural context, spatial dimension proved to be the "essence" on the basis of which anatomical entities could be subdivided into the two broadest classes:

1. Physical anatomical entity
is an anatomical entity
which *has* spatial dimension.

Examples: hemoglobin molecule, ribosome, hepatocyte, heart, head, blood, peritoneal cavity, diaphragmatic surface of heart, inferior margin of liver, apex of lung.

2. Conceptual anatomical entity
is an anatomical entity
which *has* no spatial dimension.

Examples: anatomical term, anatomical location, spatial adjacency, anterior.

Keeping the focus on 'physical anatomical entity', its descendants may be subdivided into two classes based on the essential attribute of 'mass':

1.1 Material physical anatomical entity
is a physical anatomical entity
which *has* mass.

Examples: hemoglobin molecule, ribosome, hepatocyte, heart, head, blood, urine.

1.2 Non-material physical anatomical entity
is a physical anatomical entity
which *has* no mass.

Examples: peritoneal cavity, inguinal canal, epiploic foramen, diaphragmatic surface of heart, inferior margin of liver, apex of lung, pterion.

Similarly, the descendants of 'Material physical anatomical entity' may be subdivided on the basis of the essential characteristic of "inherent 3D shape" into

classes of 'Anatomical structure', which have such a shape (e.g., hepatocyte, heart, head), and 'Body substance', which do not (e.g., osteoid, blood, urine, flatus). The essential characteristic of non-material physical anatomical entities is the number of spatial dimensions they possess. On this basis they may be subdivided into classes of anatomical point (0D), line (1D), surface (2D), and space (3D). Although spaces also have a 3D shape, it is not an inherent property; rather it is determined by the anatomical structures that surround the space.

Inherent 3D shape is a necessary, but not a sufficient, *differentia* for defining the class 'Anatomical structure'. Without specifying additional *differentiae*, a heart valve prosthesis or a uterine fibroma also manifests the definitional properties of anatomical structure and all of its parent classes, yet in an anatomical context they cannot be equated with the heart's own valve or the fundus of the uterus. Additional *differentiae* in the class definition must exclude such non-anatomical structures from the class. This is achieved by specifying *differentiae* that distinguish biological and non-biological material objects, and also biological objects generated by the coordinated expression of groups of the organism's structural genes from those that result from perturbed or abnormal biological processes. These *differentiae* are imposed on the definition of anatomical structure by the constitutive and spatial relationship principles.⁶ These principles, without restating them here, are reflected in the definition:

Anatomical structure
is a material physical anatomical entity
which *has* inherent 3D shape;
is generated by coordinated expression
of the organism's own structural genes;
consists of parts that
are anatomical structures;
spatially related to one another in patterns
determined by coordinated gene expression.

Examples: mitral valve, right ventricle, heart, myocardium, erythrocyte, hemoglobin molecule, cardiovascular system, thorax.

This definition implies that the largest anatomical structure is the organism itself (which is 'human body' in the current concept domain of the FM), and the smallest are biological molecules assembled from smaller non-biological molecules through the mediation of the organism's genes. The *differentiae* also specify the assembly of the parts of anatomical structures, including the whole organism, according to genetically predetermined patterns. Therefore a sediment of blood cells, for example, does not satisfy the definition. This definition also explains the

dominant role of the class 'Anatomical structure' in the AO, in that other high level classes of the AO (Body substance, Non-material anatomical entity) must be defined in reference to anatomical structures.

Definition and Organizational Unit Principles.

Consistent with the constraint principle, the definition principle puts a new set of constraints on defining anatomical structures: *Defining attributes of anatomical structures should be stated in terms of their constituent parts and in terms of the anatomical structures which they in turn constitute.* Among the class of anatomical relationships, this principle assigns a primary role to part-whole relationships for the classification of anatomical structures. The question of what is to be regarded as whole or part is specified by the organizational unit principle, which declares 'Organ' as the organizational unit of macroscopic anatomy. Other subclasses of 'Anatomical structure' either constitute organs (i.e., a hierarchy of organ parts, the most elementary of which is 'Tissue'), or are constituted by organs (i.e., organ systems and body parts). These two principles specify that organ systems (e.g., urinary system) and body parts (head, trunk, upper limb) should be defined in terms of the organs that constitute them, and therefore serve as differentiae for the subclasses of both the 'Organ system' and 'Body part' classes.

Consequently, the definition of 'Organ' plays a key role in the construction of the AO:

Organ

is an anatomical structure,

which consists of the maximal set of organ parts so connected to one another that together they constitute a self-contained unit of macroscopic anatomy morphologically distinct from other such units.

Together with other organs, an organ constitutes an **organ system** or a **body part**.

Examples: femur, biceps, liver, heart, skin, tracheobronchial tree, large intestine.

It follows, therefore, that differentiae for distinguishing organ subclasses must be based on the kinds of contiguous organ parts of which organs are constituted. Selecting as the differentia the presence of cavities within or among organ parts yields the structural classification of organs shown in Figure 1. Even without presenting definitions, further subclasses or instances, it may be appreciated that this classification supports the inheritance of essential structural characteristics, whereas a functional or clinical classification could not accommodate such diverse structures in a directed acyclic graph solely on the basis of the *IS_A* relationship. This inheritance

hierarchy results from the explicit definitional specification of essential structural characteristics of the kinds of parts that are shared by subclasses and instances of 'Organ' at successive levels of the AO.

It is at the level of organs and organ parts that traditional anatomy knowledge sources define anatomical structures. It is also at this level that function, location, site of disease, as well as other attributes (spatial adjacencies, attachments, landmarks) can be associated with anatomical entities. A selection of these attributes is usually enumerated in the various non-structural definitions. However, for the purpose of constructing an anatomical inheritance hierarchy (ontology), it is necessary to explicitly state those essential characteristics of anatomical entities that are usually taken for granted by the authors of non-structural definitions. The definitions in the FM meet this requirement. Therefore its AO can serve as a foundation to which other, non-defining attributes of anatomical entities can be linked in a systematic manner. This semantic structure requires the establishment of classes of entities that have not been defined in existing source of anatomical knowledge (Figure 1).

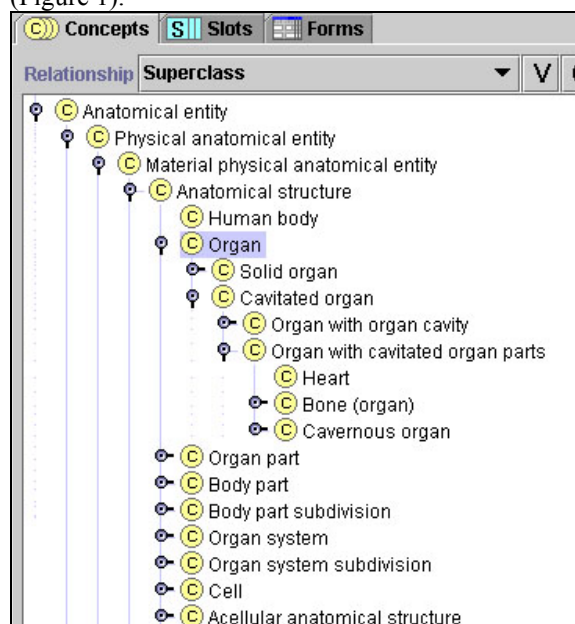


Figure 1. A part of the Anatomy Ontology of the Foundational Model viewed in Protégé. Subclasses of 'Organ', based on the presence of 'cavity' as a differentia, have been opened up in the ontology.

DESIDERATA FOR DEFINITIONS

The experience we have gained with the formulation of definitions may be synthesized as desiderata, which should generalize to the systematization of ontologies in other, non-anatomical concept domains.

1. The context for modeling a domain should be explicitly declared in the form of principles that can guide the formulation of an inheritance hierarchy.
2. All classes of the ontology should be explicitly defined.
3. Definitions should be consistent with the declared context and principles of the ontology.
4. Rather than stating the meaning of terms, definitions should state the essence of concepts in terms of their characteristics, consistent with the ontology's context.
5. The defining attributes of a class shall be all the essential characteristics shared by all members of the class.
6. The defining attribute/s shared by all concepts within the selected domain should specify the root of the ontology.
7. Immediate descendants of a class should be distinguished from one another on the basis of the same kind of defining attribute.
8. A definition should include the *genus* (that is, the class that manifests defining attributes of all its ontological descendants), and also the *differentiae* (that is the defining attributes that distinguish a class from its sibling classes).
9. To assure transitive inheritance of essential characteristics, classes of concepts should be defined that may not have been explicitly identified in existing sources of domain knowledge.
10. Definitions should be integrated in the implementation scheme of the machine-interpretable knowledge source.

IMPLEMENTATION OF DEFINITIONS

Definitions belong in the *Mk* component of the FM. The FM is implemented in Protégé, a frame-based knowledge acquisition system written in Java.¹⁰ A hierarchy of metaclasses provides templates of slots for the attributes that are inherited by descendants of a class. For example, the slot 'HAS_MASS' is introduced in the template of 'Physical anatomical entity' and has values 'YES' for 'Material physical anatomical entity' and 'NO' for 'Non-material physical anatomical entity'. Descendants of each class inherit the respective value. The frame of each metaclass contains a 'Definition' template slot.

DISCUSSION

The AO of the FM is being developed as an anatomical enhancement of UMLS. Its classes have been conceived largely as descendants of UMLS semantic types. However, the FM is fundamentally different in its purpose and design from UMLS. The goal of the FM is the unambiguous, systematic and comprehensive representation of a circumscribed

domain within a defined context. In contrast, the goal of UMLS is to unify the intended meaning of terms codified in a variety of terminologies compiled for diverse domains in varied or unspecified contexts. We hypothesize that representation of anatomy in a structural context will facilitate the reuse of the contents of the FM in the domains and contexts encompassed by UMLS source vocabularies. We further contend that the desiderata we have developed for anatomical definitions can be generalized to other domains. We, the developers of the FM, cannot test these hypotheses. Rather, we invite developers of controlled terminologies to make use of the FM when they require anatomical information and to develop definitions in their own fields in accord with the desiderata we propose. We believe that the quality of biomedical information retrievable through UMLS will be enhanced if these desiderata are incrementally met by its constituent vocabularies.

Acknowledgments

This work was supported in part by contract LM03528 and grant LM06822, National Library of Medicine.

References

1. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. *J Am Med Inform Assoc* 1994;1:218-32.
2. Campbell KE, Oliver DE, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. *J Am Med Inform Assoc* 1998;5:421-31.
3. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998;37:394-403.
4. National Library of Medicine. Medical Subject Headings. U.S. Department of Health and Human Services, Public Health Service; National Library of Medicine, Bethesda, MD, 1999.
5. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the Digital Anatomist symbolic knowledge base. *J Am Med Inform Assoc* 1998;5:17-40.
6. Rosse C, Shapiro LG, Brinkley JF. The Digital Anatomist Foundational Model: principles for defining and structuring its concept domain. *Proc AMIA Symp* 1998. 820-4.
7. Rogers J, Rector A. GALEN's model of parts and wholes: experience and comparisons. *Proc AMIA Symp* 2000;714-8.
8. Schulz S, Romacker M, Hahn U. Part-whole reasoning in medical ontologies revisited - introducing SEP triplets into classification-based description logics. *Proc AMIA Symp* 1998;830-4.
9. Mejino JLV, Fridman Noy N, Musen M, Rosse C. Representation of structural relationships in the Foundational Model of anatomy. *Proc AMIA Symp* 2001. Submitted.
10. <http://protege.stanford.edu>