



Published in final edited form as:

*Comput Biol Med.* 2006 ; 36(7-8): 674–693.

## Law and order: Assessing and enforcing compliance with ontological modeling principles in the Foundational Model of Anatomy

Songmao Zhang and Olivier Bodenreider\*

*U.S. National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

### Abstract

The objective of this study is to provide an operational definition of principles with which well-formed ontologies should comply. We define 15 such principles, related to classification (e.g., no hierarchical cycles are allowed; concepts have a reasonable number of children), incompatible relationships (e.g., two concepts cannot stand both in a taxonomic and partitive relation), dependence among concepts, and the co-dependence of equivalent sets of relations. Implicit relations—embedded in concept names or inferred from a combination of explicit relations—are used in this process in addition to the relations explicitly represented. As a case study, we investigate the degree to which the Foundational Model of Anatomy (FMA)—a large ontology of anatomy—complies with these 15 principles. The FMA succeeds in complying with all the principles: totally with one and mostly with the others. Reasons for non-compliance are analyzed and suggestions are made for implementing effective enforcement mechanisms in ontology development environments. The limitations of this study are also discussed.

### Keywords

Ontology; Anatomy; Foundational Model of Anatomy (FMA); Consistency; Hierarchical circular relations; Dependence; Relations equivalence

### 1. Introduction

Ontology modeling principles—specifying syntactic and semantic rules and constraints—are designed to ensure the soundness and consistency of the representation, conditions under which ontologies can serve the purpose of knowledge sharing and reuse. While some principles are general and therefore applicable to most ontologies, others are specific to the domain being represented. In some systems (e.g., description logic-based systems), axioms are used to implement such principles, thus enabling automatic classification and consistency checking.

---

\* Corresponding author. Tel.: +1 301 435 3246; fax: +1 301 480 3035. *E-mail addresses:* smzhang@math.ac.cn (S. Zhang), olivier@nlm.nih.gov (O. Bodenreider).

**Dr. Songmao Zhang** is currently a visiting scholar at the Lister Hill National Center for Biomedical Communications, US National Library of Medicine. She obtained her Ph.D. degree in computer science in 1992 from the Institute of Mathematics, Chinese Academy of Sciences where she is now an associate professor. Her research interests include ontology alignment, knowledge representation, data mining, AI-based automatic animation, and natural language understanding.

**Dr. Olivier Bodenreider** is a Staff Scientist in the Cognitive Science Branch of the Lister Hill National Center for Biomedical Communications at the National Library of Medicine. He obtained a M.D. degree from the University of Strasbourg, France in 1990 and a Ph.D. in Medical Informatics from the University of Nancy, France in 1993. Before joining NLM, he was an assistant professor of Biostatistics and Medical Informatics at the University of Nancy, France, Medical School. His research interests include terminology, knowledge representation, and ontology in the biomedical domain, both from a theoretical perspective and in their application to natural language understanding, reasoning, information visualization, and interoperability.

More generally, however, principles can be stated in natural language and used as guidelines by ontology developers.

Unlike concepts and relationships, the principles followed by ontology developers are rarely specified explicitly. Most ontology authoring tools do not even offer such capability. For example, an anatomical entity such as *Lobe of lung* may subsume *Lobe of left lung* and *Lobe of right lung*, but the classification criterion used here—laterality—is not represented explicitly. As a consequence, other developers contributing to the development of the same ontology may use a different spatial criterion—equally valid—for classifying *Lobe of lung* into *Upper lobe of lung* and *Lower lobe of lung*. This simple example illustrates that the lack of explicit articulation for classification criteria and, more generally, for modeling principles is likely to lead to inconsistent representations, especially in large ontologies. From the perspective of applications, consistent representations are required for tasks such as ontology integration and ontology mediation, crucial to the Semantic Web.

The objective of this study is to assess the degree to which an ontology complies with modeling principles and to investigate methods whereby compliance with these principles can be enforced. In practice, we first outline ontology modeling principles for which we provide operational definitions. Then, using these definitions, we assess the compliance of an ontology with these principles. Finally, we analyze reasons for non-compliance and suggest possible solutions for better enforcement.

The ontology under investigation is the Foundational Model of Anatomy (FMA). We selected the anatomical domain because it is central to biomedicine. While macroscopic anatomy is required for the representation of diseases and procedures, subcellular anatomy has become increasingly important for molecular biology. The FMA is a large-scale ontology of anatomy comprehensive enough to support clinical applications.

Several approaches to assessing consistency in ontologies have been suggested. Jones and Paton [1] presented five types of problems in the formal representation of hierarchical knowledge. The OntoClean methodology evaluates the nature of properties involved in taxonomic relationships based on a set of meta-properties originating from philosophical notions: identity, essence, unity, and dependence [2,3]. In the biomedical domain, we recently investigated how the description logic-based terminology SNOMED CT complies with principles of classification [4]. In another study of SNOMED CT, Ceusters et al. [5] used ontological and linguistic information to identify missing relations and improper assignment of relationships. As one of the few large-scale ontologies of anatomy, the FMA has been investigated from various perspectives. Relevant to our study is the work of Schulz et al. who transformed the FMA into a description logic-based representation (“Structure-Entirety-Part triplets”), by which some taxonomic and partonomic cycles were identified [6,7].

The major contribution of this study is to provide an operational definition of 15 principles related not only to classification, but also to various aspects of dependence (both among concepts and among sets of relations). As a case study, we investigate the degree to which the FMA complies with these principles. The algorithms we implemented for assessing and enforcing these principles are independent of the system in which the FMA was developed. However, validation mechanisms derived from these principles could be built in ontology development environments (e.g., as plug-ins in Protégé or axioms and rule extensions in description logic-based systems).

## 2. Materials

The FMA<sup>1</sup> is an evolving ontology that has been under development at the University of Washington since 1994 [8,9]. Its objective is to conceptualize the physical objects and spaces that constitute the human body. The FMA was developed around 10 foundational principles:

- *Unified context principle*: Structural anatomy is the only perspective considered (as opposed to functional or clinical anatomy).
- *Abstraction level principle*: The FMA represents canonical (not instantiated) anatomy.
- *Species specificity principle*: While currently focusing on human anatomy, high-level classes in the FMA are defined to represent the anatomy of vertebrates in general.
- *Definition principle*: Aristotelian definitions are provided for high-level classes.
- *Dominant concept principle*: Each class in the FMA is defined in reference to the dominant class (*Anatomical structure*).
- *Organizational unit principle*: Cell and Organ are the two organizational units and every other anatomical structure either constitutes or is constituted by cells and organs.
- *Content constraint principle*: The largest anatomical structure represented in the FMA is the whole organism; the smallest is *Biological macromolecule*.
- *Relationship constraint principle*: The three types of relationships among anatomical entities represented in the FMA are (1) class subsumption, (2) static physical relationships and (3) relationships indicating transformations between developmental stages.
- *Coherence principle*: The anatomy taxonomy is organized as a tree (single-inheritance class subsumption hierarchy), with *Anatomical entity* as its root.
- *Representation principle*: In addition to being an ontology of anatomy, the FMA is also concerned with terminology and collects the names (mostly in English and Latin) of the entities it represents.

The version of the FMA under investigation in this paper (February 9, 2004 version) comprises 69,889 classes covering the entire range of macroscopic, microscopic, and subcellular canonical anatomy. Concept names in the FMA are pre-coordinated, and, in addition to preferred names (one per concept), 40,683 synonyms are provided (up to 6 per concept).

The FMA is implemented in Protégé,<sup>2</sup> a frame-based ontology editing and knowledge acquisition environment developed at Stanford University [10]. Ontologies developed in Protégé are composed of classes and instances, classes being organized in taxonomic hierarchy. Slots and facets are another important component of frame-based systems: slots specify relationships between classes and describe class properties; facets express constraints on slots.

The FMA is modeled by the taxonomic relationship, *IS-A*. Additionally, seven kinds of partitive relationships are used: *PART OF*, *GENERAL PART OF*, *CLINICAL PART OF*, *CONSTITUTIONAL PART OF*, *REGIONAL PART OF*, *SYSTEMIC PART OF*, and *2D PART OF*. All the partitive relationships have inverses: *PART*, *GENERAL PART*, *CLINICAL PART*, *CONSTITUTIONAL PART*, *REGIONAL PART*, *SYSTEMIC PART*, and *2D PART*, respectively [11].

<sup>1</sup><http://fma.biostr.washington.edu/>

<sup>2</sup><http://protege.stanford.edu/>

There are also 74 associative relationships, e.g., *BRANCH OF*, *CONTAINED IN* and *NERVE SUPPLY OF*, of which 38 have inverses (e.g., *BRANCH/BRANCH OF* and *CONTAINS/CONTAINED IN*). *CONTINUOUS WITH* is its own inverse, and 35 associative relationships do not have inverses (e.g., *FASCICULAR ARCHITECTURE* and *HAS WALL*).

Taxonomic, partitive, and associative relationships link concepts to other concepts. In addition to such relationships, there are 121 slots in the FMA describing atomic properties of concepts, whose types are Boolean, Integer, Symbol, String and Instance.<sup>3</sup> For example, the slot *HAS MASS* accepts a Boolean value (true or false).

In addition to the properties represented through relations and slots, the classification principles defining the higher-level classes of the anatomy taxonomy are made explicit through textual definitions for these classes. In practice, each subdivision in the tree is motivated by a single property (e.g., *HAS MASS*). The number of subclasses for a given class is determined by the possible values for the corresponding property (e.g., *HAS MASS*: yes/no). The top-level of the anatomy taxonomy is represented in Fig. 1 along with the classification criteria for these classes.

### 3. Definitions

An ontology is defined as a theory of reality (in philosophy) or a conceptualization of what exists (in artificial intelligence). In practice, an ontology consists of categories of individuals organized in taxonomies and connected by various other relationships. This is the reason why a graph structure is often used for representing ontologies. In order to be able to assess and enforce the modeling principles for ontologies, we start by defining the following notions: graph structure, taxonomy, and ontology. Definitions of these notions focus on structural aspects and are not intended to capture all aspects of ontologies (for a formal definition of biological classes and ontological relations, see [12,13]).

A *graph*  $G$  consists of two sets  $N$  and  $E$ .  $N$  is a non-empty set of nodes, and  $E$  is a set of edges, an edge being a pair of nodes from  $N$ .  $G$  is directed if its edges are directed. The node from which a directed edge originates is called the source and the one in which it terminates is the target. A path in a directed graph is a sequence of nodes  $\langle x_0, x_1, \dots, x_n \rangle$  ( $n > 0$ ) where every two adjacent nodes  $x_i$  and  $x_{i+1}$  ( $0 \leq i \leq n - 1$ ) are source and target, respectively, of some edge. The path is direct if  $n = 1$ ; indirect otherwise. The path is called a cycle if  $x_0$  and  $x_n$  are the same node. A graph is acyclic if it has no cycles.

A *taxonomy* is a directed acyclic graph satisfying the following conditions:

1. The nodes in the graph are concepts (or classes or categories<sup>4</sup>).
2. An edge between  $x$  and  $y$  represents a direct taxonomic (*IS-A*) relationship from  $x$  to  $y$ .  $x$  is called a child (or subclass or subcategory) of  $y$  and  $y$  a parent (or superclass) of  $x$ . A concept–relationship–concept triple  $\langle x, \text{IS-A}, y \rangle$ , called a relation, can also be used to represent the edge between  $x$  and  $y$ .
3. A taxonomic (*IS-A*) relationship holds between concepts  $x$  and  $y$  (i.e.,  $\langle x, \text{IS-A}, y \rangle$ ) if (a)  $x$  is a child of  $y$ , or (b) there exists a concept  $z$  such that the two relations  $\langle x, \text{IS-A}, z \rangle$  and  $\langle z, \text{IS-A}, y \rangle$  hold. If  $\langle x, \text{IS-A}, y \rangle$  holds,  $x$  is called a descendant of  $y$  and  $y$  an ancestor of  $x$ ; in such cases,  $x$  is more specific than  $y$  (or is subsumed by  $y$ ) and  $y$  is more general than  $x$ .
4. There is one and only one concept, called the root of the taxonomy, which has no parents. Every concept except the root has at least one parent.

<sup>3</sup>Instances in the FMA correspond to special types of slot values, not to the realization of anatomical concepts as it is generally understood.

<sup>4</sup>The three terms *concept*, *class*, and *category* are used interchangeably throughout this paper.

5. The concepts  $x_1, x_2, \dots, x_n$  ( $n > 1$ ) are called siblings if they all have the same parent.

A concept is called a leaf if it has no children in the taxonomy. Single inheritance characterizes a taxonomy where every concept except the root has one and only one parent. Conversely, multiple inheritance characterizes a taxonomy where at least one concept has more than one parent.

An *ontology* is composed of at least one taxonomy and may comprise several distinct taxonomies. Concepts across taxonomies do not stand in a taxonomic relation. Concepts in an ontology represent categories of things existing in reality or abstractions generated for classification purposes. Each category or abstraction is represented exactly by one concept. Additionally, an ontology may satisfy the following:

1. In addition to the *IS-A* relationship, partitive (meronomic) relationships may hold between concepts, denoted by *PART-OF*. Every *PART-OF* relationship is irreflexive, asymmetric and transitive. *IS-A* and *PART-OF* are also called hierarchical relationships.
2. In addition to hierarchical relationships, associative relationships may hold between concepts. Some associative relationships are domain-specific (e.g., the branching relationship between arteries in anatomy and rivers in geography).
3. Relationships  $r$  and  $r'$  are inverses if, for every pair of concepts  $x$  and  $y$ , the relations  $\langle x, r, y \rangle$  and  $\langle y, r', x \rangle$  hold simultaneously. A symmetric relationship is its own inverse. Inverses of hierarchical relationships are called *INVERSE-IS-A* and *HAS-PART*, respectively.
4. Every non-taxonomic relation of  $x$  to  $z$ ,  $\langle x, r, z \rangle$ , is either inherited ( $\langle y, r, z \rangle$ ) or refined ( $\langle y, r, z' \rangle$ ) where  $z'$  is more specific than  $z$  by every child  $y$  of  $x$ . In other words, every child  $y$  of  $x$  has the same properties ( $z$ ) as its parent or more specific properties ( $z'$ ).
5. In addition to inter-concept relationships, concepts may have various properties, some of which are constrained by type (e.g., Boolean and integer), value range, cardinality, etc.

Hierarchical relationships are generally considered to be partial ordering relationships, i.e., reflexive, antisymmetric, and transitive [14]. In practice, however, the convention used in most ontologies is that the *IS-A* relationship is irreflexive. Hence we define a taxonomy as acyclic. Similarly, irreflexive partitive relationships define the so-called proper parts in some systems or theories [15]. In this paper, we refer to *IS-A* and *PART-OF* as hierarchical relationships.

## 4. Ontological principles

In addition to the structural aspects of ontologies defined earlier, there are principles of good ontology modeling. In this section we discuss such principles related to taxonomy, relationships, and dependence. Some of these principles are independent of the domain and others are specific to anatomy, the domain under investigation in this paper. The compliance of the FMA with the principles presented in this section will be investigated in the next section.

### 4.1. Principles related to taxonomy and relationships

We use  $x_0 - r_1 \rightarrow x_1 - r_2 \rightarrow x_2 \dots - r_n \rightarrow x_n$  to denote a path between the concepts  $x_0$  to  $x_n$  in the graph structure of an ontology, where  $n > 0$  and  $r_1, r_2, \dots$  and  $r_n$  are relationships (hierarchical or associative). According to the definition given for a taxonomy, there should be no *IS-A* cycles in an ontology. Analogously, according to the definition of *PART-OF* relationships, there should be no *PART-OF* cycles in an ontology. This means that for every concept  $x$ , there should be no paths such as  $x - \text{IS-A} \rightarrow \dots - \text{IS-A} \rightarrow x$  or  $x - \text{PART-OF} \rightarrow \dots - \text{PART-OF} \rightarrow x$ , where *PART-OF* represents any partitive relationship. Cycles containing a mix of *IS-A* and *PART-OF* relationships are not allowed either.

In addition to concepts, relationships can be organized in taxonomies. Relationship  $r_1$  being a child of relationship  $r_2$  implies that for every pair of concepts  $x$  and  $y$ ,  $\langle x, r_2, y \rangle$  holds if  $\langle x, r_1, y \rangle$  holds. Some relationships are incompatible with other relationships. If relationships  $r_1$  and  $r_2$  are incompatible, there is no pair of concepts  $x$  and  $y$  for which  $\langle x, r_1, y \rangle$  and  $\langle x, r_2, y \rangle$  hold simultaneously. For example, IS-A (representing class subsumption relationship) and PART-OF (representing part-whole relationship) are incompatible and should be clearly distinguished within an ontology [16]. Incompatibility can also occur among various kinds of partitive relationships. The FMA, for example, defines seven different partitive relationships, four of which are incompatible with each other (CONSTITUTIONAL PART OF, SYSTEMIC PART OF, REGIONAL PART OF and 2D PART OF) [11]. Finally, incompatibility can also occur between hierarchical and associative relationships, or among associative relationships themselves. For example, the relationships BRANCH OF and TRIBUTARY OF in the FMA describe the spatial connections (subdivision and confluence) of linear portions of tree-like structures such as veins, arteries, and nerves. These two associative relationships are therefore incompatible with both IS-A and PART-OF relationships.

Taxonomies can be thought of as the backbone of an ontology. Each taxonomy is organized according to classification principles. A classification criterion defines the difference between a given concept and its children. The classification criteria associated with top-level classes in the anatomy taxonomy of the FMA (spatial dimension, mass, inherent three-dimensional shape, dimensionality) have been presented in Fig. 1. Other major criteria in the anatomical domain include the presence of a cavity (cavitated vs. solid organ), the internal architecture (parenchymatous vs. non-parenchymatous), the organizational pattern (lobar, corticomedullary, etc.) Additional criteria used to further classify anatomical entities include spatial criteria (e.g., right/left, upper/lower, anterior/posterior), temporal criteria (e.g., permanent/deciduous), shape (e.g., long/short), and size (e.g., large/small). Ideally, a classification satisfies the so-called “jointly exhaustive and pairwise disjoint” rule. This implies that every concept should be classified by only one criterion, although several classification criteria may be used at different levels in the classification. Additionally, every child should differ from its parent by at least one difference (e.g., different values for a given slot). Moreover, each non-leaf concept should have at least two children. All children of a given parent have distinct characteristics (differentiae) compared to their parent (genus) and all differ from one another.

#### 4.2. Principles related to dependence

In an ontology, some entities are dependent on other entities. For example, concept  $x$  is dependent on concept  $y$  when  $x$  cannot exist unless  $y$  exists. The multiple aspects of dependence among concepts in an ontology include physiological, causal, logical, functional, and practical dependence [17]. Ref. [18] investigated the ontological, lexical and statistical dependence relations in Gene Ontology. In the following, we discuss two types of dependence relations: concept dependence and concept–relation dependence, both deriving from terms (i.e., concept names). Unlike terminology, ontology is not directly concerned with naming entities. In practice, however, terms often reflect dependence relations among concepts. Concepts generally have at least one name. Multiple names for a concept are synonyms and concepts may be identified by one unique preferred name.

*Concept dependence* reflects the necessary co-existence of entities in reality. For example, in anatomy, any entity whose name contains the word “wall” (e.g., *Thoracic wall*) indicates an ontological dependence on the entity bounded by this wall (here, the *Thorax*). Similarly, *Nail wall* is not expected to be present if *Nail* is not defined in the ontology. More generally, terms such as *Subdivision of  $x$*  or *Organ component of  $x$*  should not exist in an ontology unless  $x$  exists. Analogously, entities such as *Primary  $x$*  and *Secondary  $x$*  are not expected to exist independently of each other or independently of  $x$ .

Beyond dependence relations, terms may also embed various relations, of which concept dependence is the byproduct. We call this kind of dependence between concept names and relations *concept–relation dependence*. In anatomy, concept names such as *Thoracic wall* not only reflect the dependence of *Thoracic wall* to *Thorax*, but also indicate that the *Thorax* has *Thoracic wall* as one of its parts. Other types of relationships (e.g., *IS-A*) can be embedded in concept names. For example, *Sweat gland* indicates that *Sweat gland* is a kind of *Gland*.

#### 4.3. Principles related to co-dependence of equivalent relations

*Relation equivalence* is exemplified by the reification of *PART-OF* relationships. It consists of using a concept named *Part of W* to subsume a concept *P* instead of using a *PART-OF* relationship between the concept *P* (the part) and *W* (the whole). The two representations  $\langle P, \text{IS-A}, \text{Part of } W \rangle$  and  $\langle P, \text{PART-OF}, W \rangle$  are equivalent for most purposes. In addition, these two relations are co-dependent, i.e., they cannot be modified or removed independently of each other. Concept names reifying *PART-OF* relationships in the FMA include *Subdivision of W* and *Organ component of W*, where *W* is a concept present in the ontology. For example, *Subdivision of hand* suggests that all of its descendants, e.g., *Hand proper*, stand in part of relation with *Hand*.  $\langle \text{Hand proper}, \text{REGIONAL PART OF}, \text{Hand} \rangle$  is also present in the FMA. Therefore, the two relations  $\langle \text{Hand proper}, \text{IS-A}, \text{Subdivision of hand} \rangle$  and  $\langle \text{Hand proper}, \text{REGIONAL PART OF}, \text{Hand} \rangle$  are equivalent and co-dependent.

*Relation dependence* reflects logical or semantic connections among relations. For example, the two relations  $\langle x, \text{IS-A}, y \rangle$  and  $\langle y, \text{PART-OF}, z \rangle$  can be combined logically and their combination  $\langle x, \text{IS-A}, y \rangle \wedge \langle y, \text{PART-OF}, z \rangle$  implies  $\langle x, \text{PART-OF}, z \rangle$  [16]. In other words, the relation  $\langle x, \text{PART-OF}, z \rangle$  is equivalent to the combination of relations  $\langle x, \text{IS-A}, y \rangle \wedge \langle y, \text{PART-OF}, z \rangle$  and the two sets of relations  $\{\langle x, \text{PART-OF}, z \rangle\}$  and  $\{\langle x, \text{IS-A}, y \rangle, \langle y, \text{PART-OF}, z \rangle\}$  are co-dependent. When the three co-dependent relations are explicitly represented in an ontology, the modification of anyone of them requires that the other two be checked for validity.

In a previous study [19,20], we noted that relations are not always explicitly present in the FMA, and we proposed methods for making such relations explicit for the purpose of aligning anatomical ontologies. More precisely, we investigated augmentation methods for acquiring relations embedded in concept names (i.e., reified relations) and inference methods for generating new relations from a combination of existing relations.

Examples of relations acquired by augmentation include  $\langle \text{Finger}, \text{PART-OF}, \text{Hand} \rangle$ . Unlike *Hand proper*, *Finger* does not have an explicit *PART-OF* relation to *Hand*. Instead, *Finger* is a child of *Subdivision of hand*. Therefore, we created the relation  $\langle \text{Finger}, \text{PART-OF}, \text{Hand} \rangle$  from  $\langle \text{Finger}, \text{IS-A}, \text{Subdivision of hand} \rangle$ , where *Subdivision of hand* reifies a partitive relation to hand.

Relations can also be captured by various other linguistic phenomena such as nominal modification and prepositional attachment. The former often represents a hyponymic relation involving the head of the noun phrase (e.g., *Sweat gland* is a kind of *Gland*). In anatomical terms, prepositional attachment using “of” (*P of W*) often denotes a partitive relation between *P* and *W* (e.g., *Neck of femur* is part of *Femur*).

Inference consists of generating new relations from combinations of existing relations. The inference rules we used generate a partitive relation between a specialized part and the whole and between a part and a more generic whole. For example, the relation  $\langle \text{Atrioventricular valve}, \text{PART-OF}, \text{Heart} \rangle$  can be inferred from  $\langle \text{Atrioventricular valve}, \text{IS-A}, \text{Cardiac valve} \rangle$  (explicitly represented) and  $\langle \text{Cardiac valve}, \text{PART-OF}, \text{Heart} \rangle$  (generated by augmentation).

Augmentation and inference can be seen as a corollary of the dependence relations presented above. In both cases, the implicit relations are expected to be consistent with the relations

explicitly represented. In what follows, we take advantage of the methods developed for identifying implicit relations for checking the consistency of dependence relations.

## 5. Operational definition of ontological principles

We want to study the degree to which the FMA complies with 15 principles selected from the ontological principles presented in Section 4. The 15 principles listed in Table 1 are related to hierarchical cycles, classification, incompatible relationships and dependence. Additionally, some principles are applied to relations represented implicitly.

## 6. Methods and results

In order to study the degree to which the FMA complies with the principles in Table 1, we first acquire the terms, concepts, and relations explicitly represented in the FMA. In addition, we extract implicit relations and study their consistency.

### 6.1. Acquiring terms, concepts, and explicit and implicit relations

Acquiring terms consists of extracting both the preferred name and synonyms (if any) of every concept. Acquiring relations consists of extracting the relations explicitly represented. Then, these relations are complemented with missing inverse relations. For example,  $\langle \textit{Right lung}$ , CLINICAL PART OF,  $\textit{Apex of right lung (viewed clinically)}$   $\rangle$  is an explicit relation in the FMA but its inverse relation is missing. In order to make the relations extracted from the FMA complete in terms of inverse relations, we generated the missing ones (e.g.,  $\langle \textit{Apex of right lung (viewed clinically)}$ , CLINICAL PART OF,  $\textit{Right lung}$   $\rangle$ ), but still consider them as explicit relations. As shown in the upper part of Table 2, 367,224 relations are explicitly represented between concepts in the FMA, 93% of which are hierarchical relations. After complementation, the number of relation increased by about 25%.

As mentioned earlier, many relations are not explicitly represented in the FMA. Instead, the FMA-based applications such as OQAFMA [21], Emily [22] and GAPP [23] communicate directly with the FMA knowledge base and explore paths among classes dynamically. We are not suggesting that all possible relations should be represented explicitly. However, we argue that all relations FMA-based applications generate should be compatible among themselves and with the relations represented explicitly. Implicit relations can be acquired, for example, by augmentation and inference (see Section 4.3). As shown in the bottom part of Table 2, 198,330 hierarchical relations were created by augmentation in the FMA (excluding the relations represented explicitly), 39% of which came from reified PART-OF relations and 61% from other linguistic phenomena (prepositional attachment and nominal modification). A total of 11,581,584 relations were generated by inference (excluding the relations represented explicitly or acquired by augmentation) [20]. Inference resulted in many more relations because the inference rules we applied perform like the transitive closure of the combination of IS-A and PART-OF.

### 6.2. Hierarchical cycles ( $H_1$ and $H_2$ )

**Principle  $H_1$** —For every concept  $x$ , we created the set of all the concepts reachable from  $x$  through the IS-A relationship, directly or indirectly. This set constitutes the IS-A transitive closure for  $x$ , denoted by  $isaTransClosure(x)$ . If  $x \in isaTransClosure(x)$ , an IS-A hierarchical cycle is identified. No IS-A cycles were found in the FMA.



**Principle H<sub>2</sub>**—Similarly, we created the PART-OF transitive closure<sup>5</sup> for every concept  $x$ , i.e., the set of all the concepts reachable from  $x$  through PART-OF relationships, directly or indirectly, denoted by  $partofTransClosure(x)$ . If  $x \in partofTransClosure(x)$ , a PART-OF hierarchical cycle is identified. Thirty-two PART-OF cycles were identified. One such cycle is direct ( $\langle Skeletal\ muscle, \text{SYSTEMIC PART OF}, Skeletal\ muscle \rangle$ ), while the others are indirect (e.g.,  $Right\ conjunctival\ sac - \text{PART OF} \rightarrow Right\ conjunctiva - \text{SYSTEMIC PART OF} \rightarrow Right\ conjunctival\ sac$ ).

### 6.3. Number of children per concept ( $T_1$ and $T_2$ )

We counted the number of children of every concept, which ranged from 0 (for leaf nodes) to 221. There are 23,368 non-leaf concepts in the FMA, accounting for about 33% of all FMA concepts. Of these non-leaf concepts, a vast majority (23,111) have at least two children, i.e., comply with principle  $T_1$ . For example, *Limb* has two children, *Upper limb* and *Lower limb*. However, 257 of non-leaf concepts (about 1%) have a single child. Examples of such concepts include *Alveolus* (single child: *Pulmonary alveolus*), *Intercostal artery* (single child: *Supreme intercostal artery*), and *Organ component of rectum* (single child: *Non-striated muscle fiber of rectum*).

The distribution of the number of children per concept having more than one child is shown in Fig. 2 (cumulative frequency). About 79% of these concepts have two children, 18% have between 3 and 10 children, and 2.7% have between 11 and 50 children. Only 0.1% of these concepts have more than 50 children. Overall, 95% of the concepts have seven children or less. Principle  $T_2$  does not specify precisely the higher bound for the number of children per concept [24]. Intuitively, however, an unusually large number of children—relatively to other concepts in the same ontology—is likely to reflect inadequate classification. Table 3 shows some examples of concepts with a large number of children.

### 6.4. Existence of differences among siblings ( $T_3$ )

For every concept having more than one child, we created a list of all the relations in which each child was involved, and we identified the children having exactly the same sets of relations. In other words, siblings sharing identical relations with other concepts were grouped together.

Eleven thousand three hundred and nineteen such groups of siblings were identified, corresponding to 11,199 distinct parent concepts. As illustrated in Fig. 3, 48% of the 23,111 concepts having more than one child have at least two children sharing identical relations to other concepts. For example, *Ligament of right wrist* and *Ligament of left wrist* are siblings (children of *Ligament of wrist*) and exhibit no differences in their relations to other concepts. In practice, the two siblings—as represented in the FMA—differ only by their names.

Not surprisingly, this phenomenon is more frequent in the children of concepts having a large number of children. Among the 31 concepts having from 50 to 221 children, there is only one (*Set of joints*) whose children can be distinguished from each other by their relations to other concepts. In contrast, the children of the other 18 concepts exhibit no differences among most of their siblings.

### 6.5. Existence of a single classification criterion for each concept ( $T_4$ )

Unlike many other biomedical ontologies, the FMA uses explicit classification criteria for its top- and mid-level concepts. These criteria are expressed through textual, Aristotelian definitions (e.g., *Material physical anatomical entity*: physical anatomical entity which has mass) and, sometimes, properties (e.g., HAS INHERENT 3D SHAPE is true for *anatomical structure* and

<sup>5</sup>The limitations of such a transitive closure where several kinds of partitive relations are mixed is discussed in Section 7.1.

false for *body substance*). However, when present, classification criteria are not systematically identifiable through properties in the FMA. Moreover, no classification criterion is recorded for concepts close to the leaf level. Therefore, checking compliance with principle  $T_4$  cannot be systematic and automatic. In a limited number of cases, however, a classification criterion can be identified from the names of the children of a given concept. Laterality (right/left) offers an example of such criterion, where a concept  $x$  has children *Left x* and *Right x*. In this case, according to principle  $T_4$ ,  $x$  should have no other children than these two.<sup>6</sup> For simplicity reasons, and because it is a frequent feature in anatomy, we limited our study to laterality.

Sixteen thousand one hundred and eighty-one concepts in the FMA have children whose names exhibit a laterality marker (right/left), accounting for 70% of the 23,111 concepts with multiple children in the FMA. For example, *Base of patella* has two children: *Base of right patella* and *base of left patella*, and the two children of *cardiac chamber* are *Right cardiac chamber* and *Left cardiac chamber*. Laterality is the single classification criterion for nearly 95% of the 16,181 concepts, including, for example, *Base of patella*. Such cases are in compliance with principle  $T_4$ .

Eight hundred and fifty-eight (5%) of the 16,181 concepts exhibit multiple classification criteria. For example, as illustrated in Fig. 4, in addition to *Right cardiac chamber* and *Left cardiac chamber*, children of *Cardiac chamber* also include *Cardiac atrium* and *Cardiac ventricle*. *Cardiac chamber* is classified by two criteria: laterality (*Right cardiac chamber* and *Left cardiac chamber*) and morphology<sup>7</sup> (*Cardiac atrium* and *Cardiac ventricle*). Interestingly, the two children classified by laterality are further classified by morphology (e.g., *Right atrium* and *Right ventricle* for *Right cardiac chamber*). Similarly, the two other children of *Cardiac chamber* corresponding to the morphology criterion should be further classified by laterality (dotted lines in Fig. 4). However, this would make concepts such as *Left ventricle* hybrid concepts, inheriting from both *Cardiac ventricle* and *Left cardiac chamber*, which is not allowed in the FMA where single inheritance is the rule.

## 6.6. Incompatible relationships ( $R_1$ , $R_2$ , and $R_3$ )

**Principle  $R_1$** —The transitive closure created for  $IS-A$  and for  $PART-OF$  (see Section 6.2) is used to check the presence of incompatible relations  $\langle x, IS-A, y \rangle$  and  $\langle x, PART-OF, y \rangle$ . For every pair of concepts  $x$  and  $y$ , the presence of both  $y \in isaTransClosure(x)$  and  $y \in partofTransClosure(x)$  is checked.

Among the 1,105,164  $IS-A$  relations and 972,612  $PART-OF$  relations—direct and indirect—in the FMA, only 309 pairs of concepts stand in both  $IS-A$  and  $PART-OF$  relation. For example, *Auricle of atrium* has  $IS-A$  and  $REGIONAL\ PART\ OF$  relationships to *Auricle of heart*. Thirty concepts include all their children also as parts (e.g., *Auricle of heart*). Fig. 5 shows an example where the incompatible  $IS-A$  and  $PART-OF$  relations are both indirect and therefore more difficult to detect by manual review.

**Principle  $R_2$** —For each concept  $x$ , a set of all concepts having a path to  $x$ —direct or not—through the relationship  $CONSTITUTIONAL\ PART\ OF$  is created. A similar set of concepts is created for each of the three other relationships:  $SYSTEMIC\ PART\ OF$ ,  $REGIONAL\ PART\ OF$ , and  $2D\ PART\ OF$ . The presence of a concept in more than one set indicates incompatible relations.

Among the 222,994 pairs of concepts standing in at least one of the four kinds of incompatible  $PART-OF$  relations, 123,353 (about 55%) stand in more than one kind of  $PART-OF$  relation. For example,

<sup>6</sup>There are a few exceptions to this principle. For example, *Leaflet of pulmonary valve* has three children: *Right leaflet of pulmonary valve*, *Left leaflet of pulmonary valve*, and *Anterior leaflet of pulmonary valve*.

<sup>7</sup>The myocardium of the atrium and ventricle differ, for example, in structure and thickness.

*Prostatic lobule* stands in CONSTITUTIONAL PART OF, SYSTEMIC PART OF, and REGIONAL PART OF to *Prostatic gland*. The large number of incompatible partitive relations in the FMA comes from the strategy adopted during the modeling.<sup>8</sup>

**Principle R<sub>3</sub>**—For each pair of concepts  $x$  and  $y$  standing in either BRANCH OF OR TRIBUTARY OF relation, the transitive closure created for IS-A and for PART-OF is used to check the presence of incompatible relations (e.g.,  $\langle x, \text{BRANCH OF}, y \rangle$  and  $\langle x, \text{IS-A}, y \rangle$ ). Since BRANCH OF and TRIBUTARY OF are not transitive, only direct BRANCH OF and TRIBUTARY OF relations need to be searched.

Among the 6989 pairs of concepts linked by direct BRANCH OF OR TRIBUTARY OF relationships, 21 were identified as also standing in IS-A relation. For example, *Lingual branch of vagus nerve* has IS-A and BRANCH OF relationship to *Pharyngeal branch of vagus nerve*. Four hundred and thirty pairs of concepts linked by direct BRANCH OF OR TRIBUTARY OF relationships (about 6%) were identified as also standing in PART-OF relation. For example, *Internal iliac artery* is in both PART OF and BRANCH OF relationship with *Common iliac artery*.

### 6.7. Dependence ( $D_1$ and $D_2$ )

**Principle D<sub>1</sub>**—For each term in the form of *Subdivision of  $x$*  or *Organ component of  $x$* , a systematic check was made for a concept with name  $x$ . Preferred names and synonyms were both used in this process.

One thousand nine hundred and eighty terms in the FMA contain either “subdivision of” or “organ component of”. In 1809 cases (91%), the concept on which they depend was present in the FMA, e.g., *Temporal part of head* for *Subdivision of temporal part of head*. However, in 171 cases (9%), this concept was not defined in the FMA. For example, the concept *Upper premolar tooth* is absent although *Subdivision of upper premolar tooth* exists.

**Principle D<sub>2</sub>**—The existence of a direct PART-OF relationship is checked between each concept whose name contains “wall” and some other concept. Again, both preferred names and synonyms were used in this process.

There are 1321 terms containing “wall” in the FMA, corresponding to 1068 concepts. Six hundred and eighty-two of these (64%) exhibit the expected PART-OF relation, e.g.,  $\langle \text{Wall of gut, REGIONAL PART OF, Gut} \rangle$ . For the remaining 386 (36%), however, there is no explicit PART-OF relationship to any other concept (e.g., *Nail wall*). This is precisely what the techniques we developed for acquiring implicit relations are meant to address. In all but three cases (1%), these techniques were able to generate the missing partitive relations. In 75% of the cases, the missing relation was obtained by augmentation (i.e., by analyzing concept names). For example,  $\langle \text{Nail wall, PART-OF, Skin} \rangle$  can be created from  $\langle \text{Nail wall, IS-A, Organ component of skin} \rangle$  (reified PART-OF relation). In 24% of the cases, the missing relation was generated by inference. For example, the relation  $\langle \text{Wall of portal vein proper, PART-OF, Portal vein} \rangle$  can be inferred from  $\langle \text{Wall of portal vein proper, IS-A, Wall of portal vein} \rangle$  and  $\langle \text{Wall of portal vein, PART-OF, Portal vein} \rangle$ .

### 6.8. Co-dependence of equivalent relations ( $C_1$ and $C_2$ )

Some relations in the FMA are redundant, such as  $\langle x, \text{IS-A}, \text{subdivision of } y \rangle$  and  $\langle x, \text{PART-OF}, y \rangle$ , when both are explicitly represented. A relation which is explicitly represented may also be redundant with a combination of other explicit relations. As mentioned earlier, it is important that the co-dependence between these equivalent relations (principle  $C_1$ ) or sets of relations (principle  $C_2$ ) be clearly identified.

<sup>8</sup>Whenever a PART-OF relation is added to a concept, all the PART-OF slots are populated simultaneously, leaving the burden of removing the non-relevant ones to the author. This issue is being addressed.

Among the 147,077 direct PART-OF relations explicitly represented in the FMA, we identified 5546 relations (4%) equivalent to some IS-A relations. For example,  $\langle \textit{Fascia of muscle},_{\text{PART-OF}}, \textit{Muscle (organ)} \rangle$  is equivalent to  $\langle \textit{Fascia of muscle},_{\text{IS-A}}, \textit{Organ component of muscle (organ)} \rangle$ . We also identified 45,800 relations (31%) equivalent to some combinations of relations. For example,  $\langle \textit{Upper lobe of lung},_{\text{PART-OF}}, \textit{Lung} \rangle$  is equivalent to  $\langle \textit{Upper lobe of lung},_{\text{IS-A}}, \textit{Lobe of lung} \rangle$  and  $\langle \textit{Lobe of lung},_{\text{PART-OF}}, \textit{Lung} \rangle$ .

### 6.9. Consistency of implicit relations ( $I_1$ and $I_2$ )

A large number of implicit relations were made explicit by the augmentation and inference techniques presented earlier. These relations are expected to be consistent with themselves (principle  $I_1$ ) and with the relations explicitly represented (principle  $I_2$ ). The technique used to check the consistency of implicit relations is similar to that used for principles  $H_1$  and  $H_2$ , (i.e., checking the transitive closures of IS-A and PART-OF relationships, respectively, for each concept), with the difference that the sets of relations now include implicit relations.

The implicit relations are generally consistent with themselves except in two cases. For example, the explicit relation  $\langle \textit{Surface of umbilicus},_{\text{IS-A}}, \textit{Subdivision of surface of umbilicus} \rangle$  implicitly represents a direct PART-OF cycle  $\langle \textit{Surface of umbilicus},_{\text{PART-OF}}, \textit{Surface of umbilicus} \rangle$ . Combining IS-A and PART-OF relations contributed to the identification of 37 cycles. Examples include  $\textit{Cardiac muscle of atrium} -_{\text{IS-A}} \textit{Cardiac muscle(tissue)} -_{\text{CONSTITUTIONAL PART OF}} \rightarrow \textit{Cardiac muscle of atrium}$ , and  $\textit{Smooth muscle} -_{\text{SYSTEMIC PART OF}} \rightarrow \textit{Muscularis mucosae of stomach} -_{\text{IS-A}} \rightarrow \textit{Muscularis mucosae} -_{\text{IS-A}} \rightarrow \textit{Smooth muscle}$ .

In all but two cases, the implicit relations are consistent with the relations represented explicitly. For example, the explicit relation  $\langle \textit{Apex of urinary bladder},_{\text{IS-A}}, \textit{Subdivision of urinary bladder} \rangle$  implicitly represents  $\langle \textit{Apex of urinary bladder},_{\text{PART-OF}}, \textit{Urinary bladder} \rangle$ . This conflicts with the explicit relation  $\langle \textit{Apex of urinary bladder},_{\text{REGIONAL PART}}, \textit{Urinary bladder} \rangle$ . Note that when implicit and explicit relations conflict, a human review is needed to determine which relation is valid. In this particular case, the explicit relation  $\langle \textit{Apex of urinary bladder},_{\text{REGIONAL PART}}, \textit{Urinary bladder} \rangle$  is wrong.<sup>9</sup>

## 7. Discussion

### 7.1. Limitations of this study

The 15 principles presented in Section 5 for which we provide an operational definition are not always theoretically sound, are far from complete, and rely in part on terminological features.

**Soundness issues**—As mentioned in Section 6.2, one obvious limitation of this approach is the use of a transitive closure of PART-OF where several kinds of partitive relationships are mixed, some of which are not transitive.<sup>10</sup> As a consequence, some of the partitive relations present in such a transitive closure may not hold and conflicts detected between such relations and other relations may be false positives. In practice, however, most of the partitive relations generated in this process are indeed valid and the benefit of over-generating relations outweighs the risk of falsely identifying conflicts. Analogously, some relations generated from a combination of IS-A and PART-OF relations may not hold. For example,  $\langle \textit{Left eyeball},_{\text{PART-OF}}, \textit{Eye} \rangle$  is not true (not every *Eye* has a *Left eyeball* as its part), although it was generated from  $\langle \textit{Left$

<sup>9</sup> $\langle \textit{Apex of urinary bladder},_{\text{REGIONAL PART OF}}, \textit{Urinary bladder} \rangle$  is accurate (the relationships REGIONAL PART and PART-OF go in opposite directions).

<sup>10</sup>Such a transitive closure was originally created for the purpose of aligning two ontologies of anatomy. Because the specific partitive relationships defined in each ontology were different, we relied on their least common subsumer—a generic PART-OF relationship—for comparing the two ontologies.

*eyeball*, <sub>IS-A</sub>, *Eyeball*) and  $\langle \textit{Eyeball}$ , <sub>PART-OF</sub>, *Eye*  $\rangle$ . Therefore, the practical limitation of this approach is the need for a human review of the problems identified. In other words, when implemented in an ontology development environment, such mechanisms should be used to draw the attention of ontology developers, not for the automatic resolution of conflicts.

**Completeness issues**—To some extent, a similar objection can be made about taxonomic relations. Although there is only one <sub>IS-A</sub> relationship in the FMA, the relations  $\langle \textit{Right cardiac chamber}$ , <sub>IS-A</sub>, *Cardiac chamber*  $\rangle$  and  $\langle \textit{Cardiac atrium}$ , <sub>IS-A</sub>, *Cardiac chamber*  $\rangle$  actually exhibit different kinds of <sub>IS-A</sub> (topological and morphologic <sub>IS-A</sub>, respectively). The uncontrolled use of <sub>IS-A</sub> to signify different sorts of relations results in what Guarino has called “<sub>IS-A</sub> overload”, which is often associated with examples of incorrect subsumption [25]. In contrast, the use of explicit subsumption links would enable a large taxonomy such as the FMA to be subdivided in *partitions*—in this example, a topological (left/right) partition and a morphologic (atrium/ventricle) partition—within which taxonomic reasoning can be more reliably performed. The various partitions would yield complementary views on different aspects of one and the same reality. The formal theory underlying such partitions is presented in [26].

In addition to explicit subsumption links, additional principles could be based on concepts properties not expressed through inter-concept relationships (e.g., constraints on slot values); the incompatibility among some associative relationships could also be exploited. More generally, the principles defined in this study largely rely on the structural features of an ontology. Formal-ontological principles defined, for example, in [12] would of course constitute a powerful complement to our approach.

**Terminology issues**—Our method for assessing the dependence among concepts in an ontology mostly relies on the lexico-syntactic features of concept names. However, collecting the names of entities is the objective of terminology, not ontology. Although the FMA provides a rather extensive list of names for each concept, other ontologies of anatomy do not (e.g., GALEN<sup>11</sup>). Therefore, the applicability of our dependence principles is likely to vary among ontologies.

## 7.2. Why do ontologies fail to comply with modeling principles?

In this analysis, the FMA complies overall with all 15 principles to a large extent. However, the compliance is complete only with principle **H**<sub>1</sub>: the absence of <sub>IS-A</sub> cycles; non-compliance with the other principles ranges from a few occurrences (e.g., for principle **I**<sub>2</sub>: consistency between implicit and explicit relations) to about 50% (e.g., for principle **T**<sub>3</sub>: presence of explicit differences among siblings). There are many reasons why compliance of the FMA with the other 14 principles is less than complete. The most obvious is that the FMA is considered by its authors and curators to be incomplete and still evolving, with large segments of the ontology awaiting population with relationships. We believe, however, that the single most important factor is the lack of support for modeling principles in ontology development environments. This is the case not only in frame-based systems such as Protégé—used for developing the FMA—but also, as noted by Golbreich et al. [27] in description logic-based systems, including OWL editors.

In the absence of built-in mechanisms for enforcing consistency, the burden of assessing it lies with the developers of the ontology. In a large ontology such as the FMA, maintaining consistency without automatic mechanisms is a considerable challenge and the relatively small number of conflicts identified in this study reflects the careful work of its team of developers. Although ontology development environments such as description logic-based systems

<sup>11</sup><http://www.opengalen.org/>

provide some support for checking consistency, their use does not necessarily ensure compliance with all principles [4]. The approach proposed in this paper, although limited, is independent of any formalism or environment and is therefore applicable to (or adaptable by) most ontologies.

### 7.3. Explicit articulation of modeling principles and dependence relations

Making explicit the classification criterion used in a subsumption link—as suggested earlier—would result not only in reliable reasoning, but also in more consistent ontologies. Labeling *IS-A* relations with the underlying classification criterion would, for example, enable the identification of gaps along hierarchical paths for a particular criterion. By representing some classification criteria through slots (e.g., *HAS MASS*), the FMA exposes the most salient properties of anatomical entities. However, other classification criteria are described solely in textual definitions, preventing computers from reasoning automatically about these criteria.

Similarly, we believe that the explicit representation of dependence relations among concepts and co-dependence between sets of equivalent relations constitutes an important step towards more consistent ontologies. Such a representation would make it possible to alert the developers of ontologies when any of the dependent concepts or co-dependent relations are about to be modified. Additionally, the ability to establish equivalence between relations across ontologies constitutes a key factor for aligning ontologies developed under different principles.

These measures would of course be complementary to existing techniques such as OntoClean and other formal ontological approaches to assessing and enforcing consistency.

## 8. Conclusions

The creation and maintenance of large ontologies require not only the appropriate domain expertise but also well-defined organizational principles. In addition, maintaining ontologies in a consistent state also requires that such principles be actually enforced in the editing environments used to develop ontologies. Few environments, however, currently provide such capabilities. We proposed a set of 15 general principles based on classification, dependence and co-dependence of equivalent relations, for which we provide operational definitions. These definitions can be used as quality assurance tools for ontologies.

In the domain of anatomy, the Foundational Model of Anatomy (FMA) offers a good example of a large ontology (about 70,000 classes) based on a set of modeling principles. We investigated the degree to which the FMA complies with our 15 principles. The FMA succeeds in complying with all the principles: totally with one and mostly with the others. While demonstrating the effectiveness of our principles for detecting inconsistency in ontologies, we also offer suggestions for implementing effective enforcement mechanisms in ontology development environments.

### Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine. This work was done while Songmao Zhang was a visiting scholar at the Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Department of Health and Human Services.

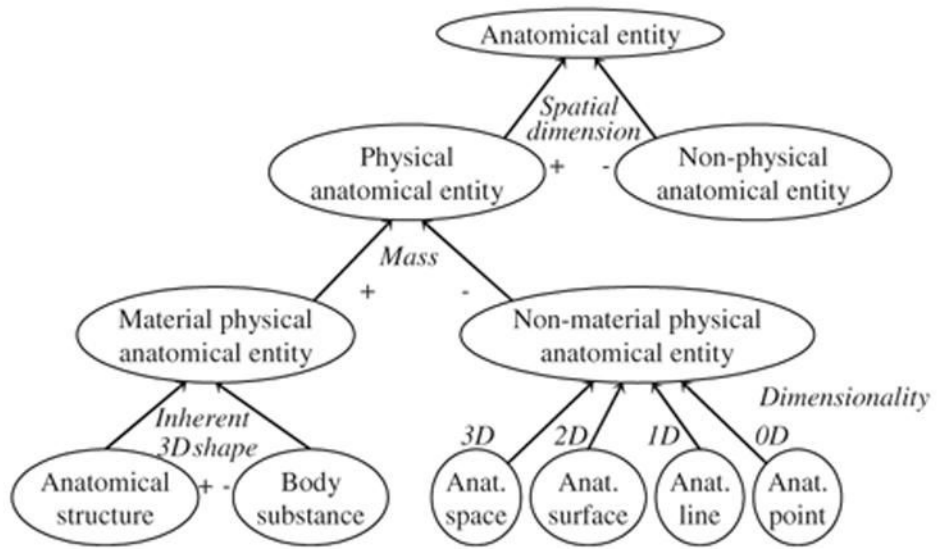
The authors would like to thank Anand Kumar for suggesting this study and for helpful discussions. Thanks also to Cornelius Rosse, José Mejino and Todd Detwiler, the authors and curators of the FMA, for their support and encouragement.

## References

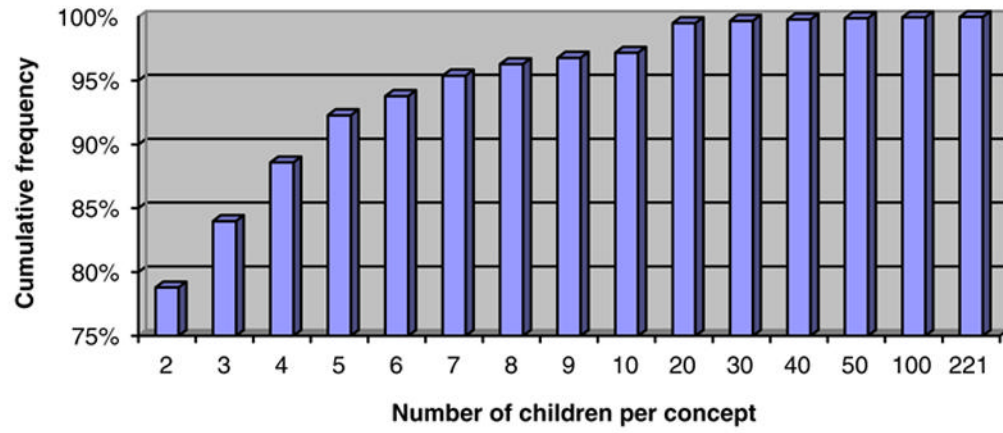
1. Jones DM, Paton RC. Toward principles for the representation of hierarchical knowledge in formal ontologies. *Data Knowl Eng* 1999;31:99.
2. Guarino N, Welty C. Evaluating ontological decisions with ontoclean. *Commun ACM* 2002;45:61.
3. Welty C, Guarino N. Supporting ontological analysis of taxonomic relationships. *Data Knowl Eng* 2001;39:51.
4. Bodenreider, O.; Smith, B.; Kumar, A.; Burgun, A. Investigating subsumption in DL-based terminologies: a case study in SNOMED CT, First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004). 2004. p. 12
5. Ceusters, W.; Smith, B.; Kumar, A.; Dhaen, C. Mistakes in medical ontologies: where do they come from and how can they be detected?. In: Pisanelli, D., editor. *Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies*. IOS Press; Burke, VA: 2004. p. 145
6. Beck, R.; Schulz, S. Logic-based remodeling of the digital anatomist foundational model; *Proceedings of AMIA Symposium*; 2003. p. 71
7. Schulz, S.; Hahn, U. A knowledge representation view on biomedical structure and function; *Proceedings of AMIA Symposium*; 2002. p. 687
8. 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 Inf Assoc* 1998;5:17.
9. Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the foundational model of anatomy. *J Biomed Inf* 2003;36:478.
10. Noy NF, Musen MA, Mejino JLV, Rosse C. Pushing the envelope: challenges in a frame-based representation of human anatomy. *Data Knowl Eng* 2004;48:335.
11. Mejino, JL.; Agoncillo, A.; Richard, KL.; Rosse, C. Representing complexity in part-whole relationships within the foundational model of anatomy; *Proceedings of AMIA Symposium*; p. 2003-450.
12. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. *Genome Biology* 2005;6:R46. [PubMed: 15892874]
13. Smith B, Rosse C. The role of foundational relations in the alignment of biomedical ontologies. *Medinfo* 2004:444. [PubMed: 15360852]
14. Sowa, JF. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole; Pacific Grove: 2000.
15. Aitken JS, Webber BL, Bard JB. part-of relations in anatomy ontologies: a proposal for RDFS and OWL formalisations. *Pac Symp Biocomput* 2004:166. [PubMed: 14992501]
16. Winston ME, Chaffin R, Herrmann D. A taxonomy of part-whole relations. *Cogn Sci* 1987;11:417.
17. Simons, PM. *Parts: A Study in Ontology*. Clarendon Press, Oxford University Press; Oxford, New York: 1987.
18. Burgun, A.; Bodenreider, O.; Aubry, M.; Mosser, J. Dependence Relations in Gene Ontology: A Preliminary Study. 2004 [checked December 2, 2004]. [http://mor.nlm.nih.gov/pubs/pdf/2004-go\\_workshop-ab.pdf](http://mor.nlm.nih.gov/pubs/pdf/2004-go_workshop-ab.pdf)
19. Zhang, S.; Bodenreider, O. Aligning representations of anatomy using lexical and structural methods; *Proceedings of AMIA Symposium*; 2003. p. 753
20. Zhang, S.; Bodenreider, O. Investigating implicit knowledge in ontologies with application to the anatomical domain. In: Altman, RB.; Dunker, AK.; Hunter, L.; Jung, TA.; Klein, TE., editors. *Pacific Symposium on Biocomputing 2004*. World Scientific; Singapore: 2004. p. 250
21. Mork P, Brinkley JF, Rosse C. OQAFMA querying agent for the foundational model of anatomy: a prototype for providing flexible and efficient access to large semantic networks. *J Biomed Inf* 2003;36:501.
22. Detwiler LT, Chung E, Li A, Mejino J, Agoncillo A, Brinkley J, Rosse C, Shapiro L. A relation-centric query engine for the foundational model of anatomy. *Medinfo* 2004;2004:341. [PubMed: 15360831]

23. Distelhorst, G.; Srivastava, V.; Rosse, C.; Brinkley, JF. A prototype natural language interface to a large complex knowledge base, the foundational model of anatomy; AMIA Annual Symposium Proceedings; 2003. p. 200
24. Noy, NF.; McGuinness, DL. Ontology Development 101: A Guide to Creating your First Ontology. Stanford University; Knowledge Systems Laboratory: [(checked December 2, 2004).]. (<http://www.ksl.stanford.edu/numberindex.html>)
25. Guarino, N. Some ontological principles for designing upper level lexical resources. In: Rubio, A.; Gallardo, N.; Castro, R.; Tejada, A., editors. Proceedings of First International Conference on Language Resources and Evaluation. ELRA—European Language Resources Association; Granada, Spain: 1998. p. 527
26. Bittner, T.; Smith, B. A theory of granular partitions. In: Duckham, M.; Goodchild, MF.; Worboys, MF., editors. Foundations of Geographic Information Science. Taylor & Francis; London: 2003. p. 117
27. Golbreich C, Dameron O, Gibaud B, Burgun A. Web ontology language requirements w.r.t expressiveness of taxonomy and axioms in medicine. Semantic Web—ISWC 2003 2003;2870:180.

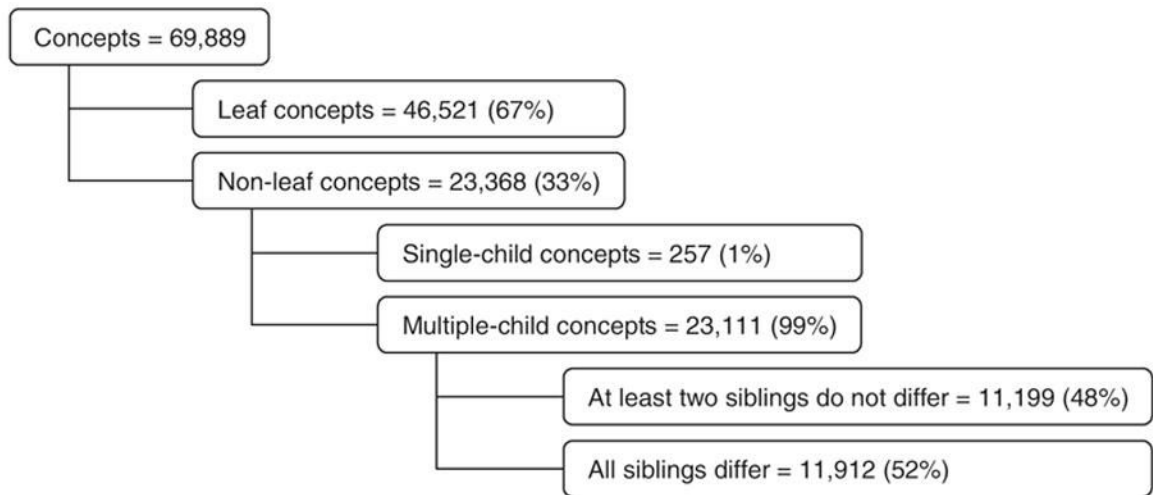




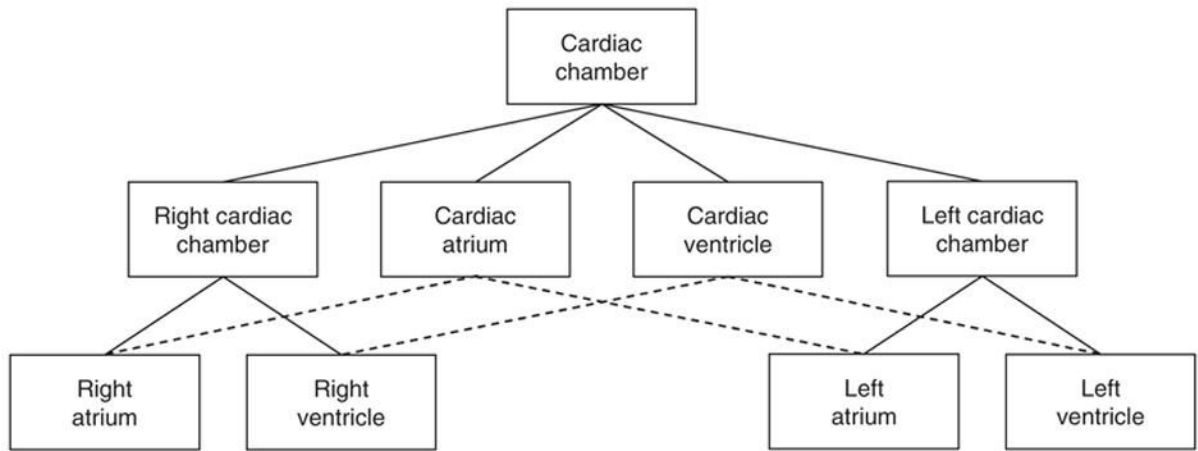
**Fig. 1.**  
Top-level classes of the anatomy taxonomy of the FMA.



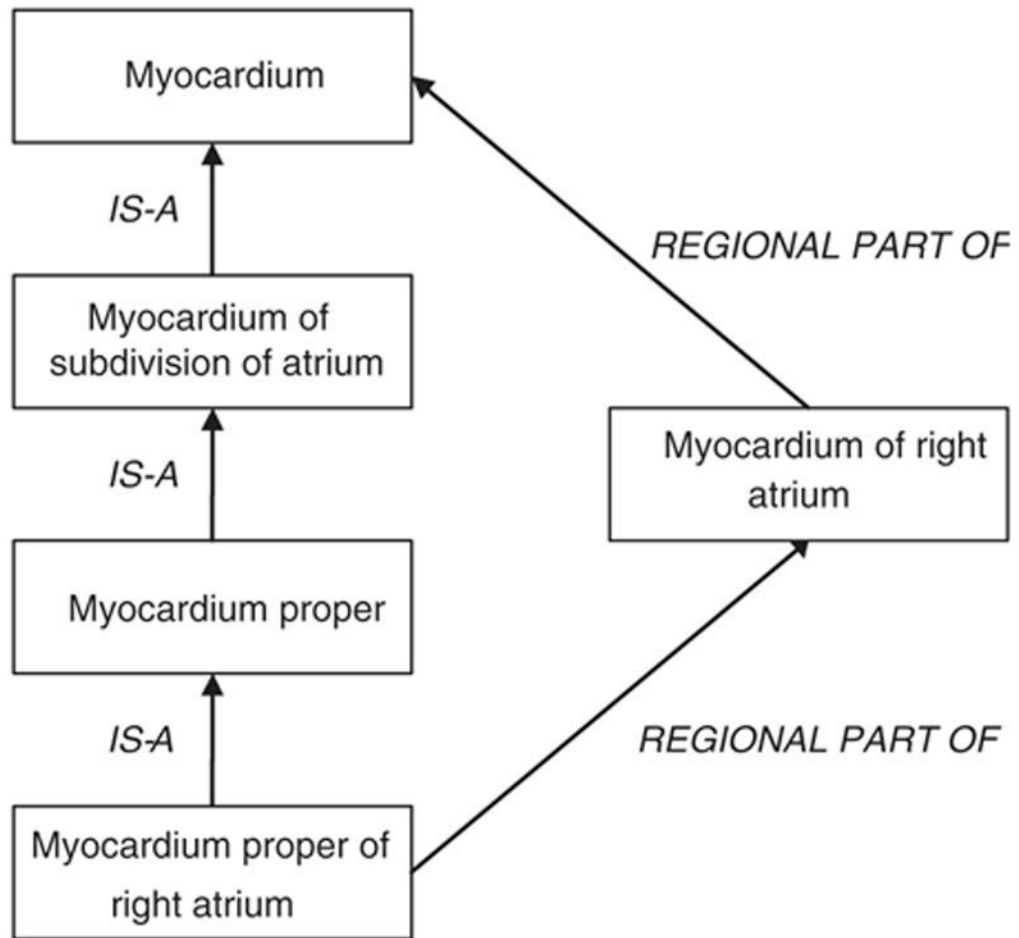
**Fig. 2.** Distribution of the number of children per concept (cumulative frequency).



**Fig. 3.**  
Existence of differences among siblings.



**Fig. 4.**  
Classification of *Cardiac chamber* in the FMA.



**Fig. 5.**  
Indirect *IS-A* and part of incompatible relations in the FMA.

Table 1

## Operational definition of the principles

Principle		Definition
Hierarchical cycles	<b>H<sub>1</sub></b>	No IS-A hierarchical cycles are allowed.
	<b>H<sub>2</sub></b>	No PART-OF hierarchical cycles are allowed.
Classification	<b>T<sub>1</sub></b>	Every non-leaf concept has at least two children.
	<b>T<sub>2</sub></b>	Every concept has a reasonable number of children (relative to other concepts in the same ontology).
	<b>T<sub>3</sub></b>	In every group of siblings, each concept has specific properties or relations to other concepts.
	<b>T<sub>4</sub></b>	Every non-leaf concept is classified according to a single criterion.
Incompatible relationships	<b>R<sub>1</sub></b>	For every pair of concepts $x$ and $y$ , $x$ and $y$ do not have both IS-A and PART-OF relationships.
	<b>R<sub>2</sub></b>	For every pair of concepts $x$ and $y$ , $x$ and $y$ have at most one kind of the four PART-OF relationships: CONSTITUTIONAL PART OF, SYSTEMIC PART OF, REGIONAL PART OF and 2D part of.
	<b>R<sub>3</sub></b>	For every pair of concepts $x$ and $y$ , $x$ and $y$ do not have both BRANCH OF (OR TRIBUTARY OF) and hierarchical relationships (IS-A OR PART-OF).
Dependence	<b>D<sub>1</sub></b>	Concept <i>Subdivision of <math>x</math></i> (or <i>Organ component of <math>x</math></i> ) does not exist unless concept $x$ exists.
	<b>D<sub>2</sub></b>	Term containing the word “wall” indicates that the corresponding concept has PART-OF relationship to some larger concept.
Co-dependence of equivalent relations	<b>C<sub>1</sub></b>	The co-dependence between equivalent relations $\langle x, \text{IS-A}, \text{Subdivision of } y \rangle$ (or $\langle x, \text{IS-A}, \text{Organ component of } y \rangle$ ) and $\langle x, \text{PART-OF}, y \rangle$ must be identified.
	<b>C<sub>2</sub></b>	The co-dependence between equivalent sets of relations $\{\langle x, \text{IS-A}, y \rangle, \langle y, \text{PART-OF}, z \rangle\}$ (or $\langle x, \text{PART-OF}, y \rangle, \langle y, \text{IS-A}, z \rangle$ ) and $\langle x, \text{PART-OF}, z \rangle$ must be identified.
Implicit relations	<b>I<sub>1</sub></b>	The implicit relations are consistent within themselves.
	<b>I<sub>2</sub></b>	The implicit relations are consistent with explicit relations.

**Table 2**  
Number of explicit and implicit relations in the FMA

---

Explicit relations (direct)	Explicitly represented	367,224
	Complemented	92,480
	Total	459,704
Unique implicit relations	Augmented	198,330
	Inferred	11,581,584
	Total	11,779,914

---

**Table 3**  
Examples of concepts with a large number of children (#)

Concept name	#	Children (partial list)
Set of arteries	221	<ul style="list-style-type: none"> <li>• Ventral branches of aorta</li> <li>• Set of short posterior ciliary arteries</li> <li>• Set of long posterior ciliary arteries</li> </ul>
Organ component of muscle (organ)	218	<ul style="list-style-type: none"> <li>• Fascia of muscle</li> <li>• Organ component of muscle of arm</li> </ul>
General anatomical term	186	<ul style="list-style-type: none"> <li>• Right side of jaw</li> <li>• Left side of jaw</li> <li>• Muscle</li> </ul>
Set of veins	164	<ul style="list-style-type: none"> <li>• Ocular muscle</li> <li>• Set of veins of heart</li> <li>• Set of right ventricular veins</li> <li>• Set of left ventricular veins</li> </ul>
Anatomical line	125	<ul style="list-style-type: none"> <li>• Groove</li> <li>• Groove for transverse sinus</li> <li>• Crest of round window</li> </ul>