

# Bio-Medical Ontologies Maintenance and Change Management

Arash Shaban-Nejad<sup>†</sup> and Volker Haarslev<sup>\*</sup>

<sup>†</sup>arash\_sh@cs.concordia.ca, <sup>\*</sup>haarslev@cs.concordia.ca

Department of Computer Science and Software Engineering, Concordia University,  
Montreal, Quebec H3G 1M8, Canada

**Abstract** Things change. Words change, meanings and context change. To manage a large volume of evolving bio-medical data of various types, one needs to employ several techniques from areas such as knowledge representation, semantic web and databases. Many of these techniques require a formal description of a part of the real world. Ontologies can provide a set of shared and precisely defined terms in various degrees of formality to describe a particular domain of interest. When the knowledge changes, then the related definitions will be altered. Changes to ontologies may occur for many reasons. The issues arising from ontological change can affect the validity of information in applications that are tightly bound to concepts in a particular ontological context. Many knowledge-based systems are now reaching a stage where they need a change management strategy to update their ontological knowledge. This area is becoming increasingly important in science as high throughput techniques frequently necessitate updates to existing scientific 'truths'. In this chapter, we survey and review state of the art change management in bio-ontologies as well as some of the available tools and techniques in this area. We also survey various potential changes in biomedical ontologies, with actual examples from some of the most popular ontologies in the biomedical domain. In addition we investigate the potential of some of the advanced formalisms in this context by proposing our formal method for analyzing and supporting ontology evolution and change management.

## 1 Introduction

Employing clinical terminologies has a long history in medicine and life sciences (Bodenreider and Stevens 2006). However, it is a new trend to use ontology as it is defined by Gruber (1993), the “specification of conceptualization” that aims to provide an underlying discipline of sharing knowledge and modeling biomedical applications by defining concepts, properties and axioms. Ontologies are extensively being employed in biomedical systems to share common terminologies, provide annotation, and organize and extract knowledge from a domain of interest. Ontologies are constantly

evolving to fix errors, reclassify the taxonomy, and add/remove concepts, attributes, relations and instances. Due to the importance of ontology as the conceptual backbone of modern decision support systems in life sciences, the maintenance phase is crucial in an ontology life cycle to preserving the validity and consistency of ontological knowledge. Ontology maintenance traditionally includes two main activities: ontology integration and change management. However, there is not always a clear line to distinguish between these activities. In many cases, the ontology change management process requires one to perform some data or semantic integration, and it is also very unlikely to keep all the integrated ontologies unchanged. The maintenance cost for evolving ontologies and knowledge bases is relatively expensive, and may range up to 90% (Jones 1998) of the total cost, depending on the size, complexity and domain of an ontology.

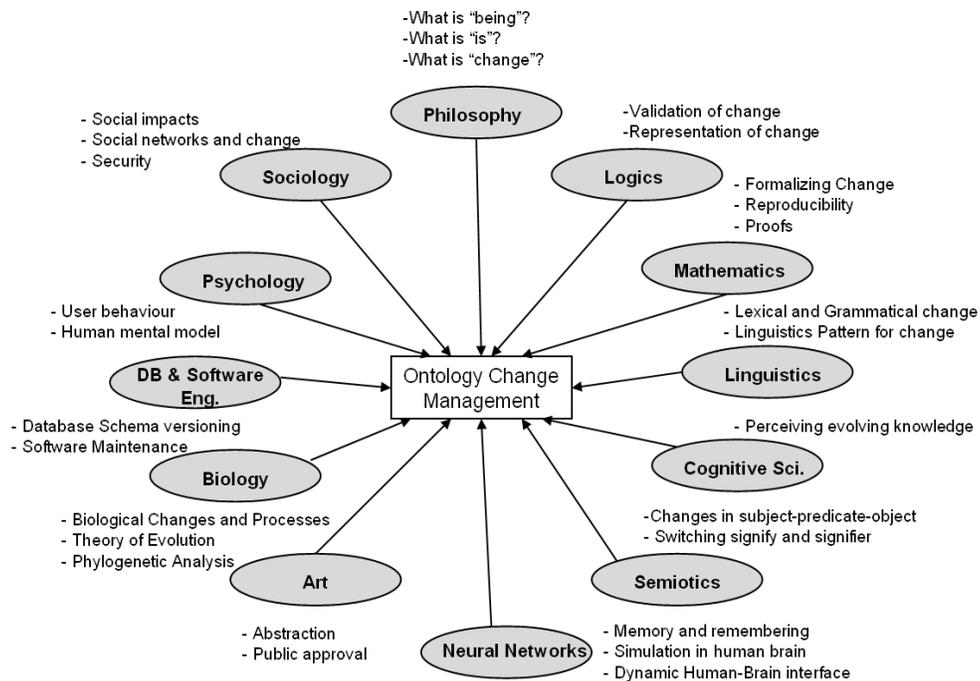
Ontologies evolve all the time, and each change in the ontological structure or nomenclature can have a crucial impact on the inferred knowledge. Especially in a heterogeneous environment, like the Web, with vast numbers of interdependencies, even simple changes to ontological elements can trigger a domino effect, and it is very hard—sometimes impossible—to guess all the effects of a simple change. Different versions of an ontology behave differently in response to posed queries. If one works with a system based on frequently changing ontologies, how can s/he even ask queries and be sure of the logical and scientific correctness of the answer? The issues arising from ontology evolution can affect the validity of information in applications that are tightly bound to concepts in a particular ontological context.

In the last decade, several studies (Cimino and Clayton 1994, Oliver et al. 1999, Klein and Noy 2003, Stojanovic 2004, Noy et al. 2006, Flouris 2006) on ontology evolution have been reported. Despite worldwide efforts, the topic of ontology evolution is still a source of much debate, as it brings together various issues that are central to philosophy, logic, artificial intelligence, cognitive science, neural nets, linguistics and physics, including identity, persistence and time.

This chapter is structured as follows. First, Section 2 describes a brief philosophical foundation of change. Section 3 reviews state of the art change management in some selected biomedical ontologies. In Section 4, we look at different types of common alterations with some actual examples from popular bio-ontologies. We survey some of the tools and approaches other researchers take to handle ontology evolution in Section 5. Section 6 describes our proposed framework for autonomous ontology change management in an agent-based environment grounded in category theory by using the FungalWeb ontology (Baker et al. 2006) as the application scenario. This manuscript is partially based on papers (Shaban-Nejad and Haarslev 2007(a), 2007(b), 2008) and other conferences, with many additions, including a variety of new materials and examples.

## 2 Philosophical Foundations

Designing a framework for ontology evolution by using available methods in the area of knowledge representation (KR) is the main strategic plan in the Semantic Web community. However, since the problem of change management is not completely computational, it seems necessary to incorporate complementary techniques from other disciplines such as philosophy, mathematics, biology, neural networks, semiotics, linguistics, psychology (to study the behavioral affects), etc. (Fig. 1) for the ontology evolution process. The topic of change—and particularly changing ontology (as the study of “being”)—brings together various issues that are central to philosophy, including identity, persistence and time (Wasserman 2006).



**Fig. 1.** Multi-disciplinary nature of research on ontology change management.

Discussion about change is as old as philosophy itself. Heraclitus (535–475 BCE), for example, argued that “All is flux,” and everything is changing all the time, so that it is impossible to step into the same river twice. Parmenides (b. 510 BCE) and Zeno

of Elea (490–430 BCE) were not in agreement with Heraclitus’s statement; they believed in the constancy and stability of the world. Parmenides had stated that “reality is one, and this one, which only is, is unchanging” (Magee 1999). Zeno of Elea also believed all changes and motions are in fact illusions of the senses (Hardie and Gaye 2004), and to show the paradoxical nature of change and motion, he summarized his philosophy into several paradoxes, including The Dichotomy, Achilles and the Tortoise and The Arrow (Kemerling 2006). Plato (427–347 BCE) in his allegory of the Cave tried to overcome this issue by separating the world into the visible world, which is uncertain and changes frequently, and the intelligible or real world, which is stable, arose from reason and includes the timeless unchanging “Forms”. Husserl (born 1859) tried to define the concept of changes by considering the notion of time, saying, “Things are always intended *toward* something, and are always ‘about’ something,” which shifts the notion of ontology from studying “being” towards studying “becoming”.

It has been commonly acknowledged that a change happens in relation to time. However, Aristotle (384–322 BCE) in his book *Physics IV* (10) argued that since change, unlike time, occurs at different rates, it is distinct from time (Hardie and Gaye 2004). The nature of change may appear contradictory and a source of inconsistency, as “it requires both sameness and difference” in parts and attributes (Wasserman 2006) and deals with contrary facts about the identity of things. See Leibniz’s Law at (Wasserman 2006), Theseus’s paradox at (Cohen 2004) and The Heap (Sorites) paradox at (Zalta 2005) for more information on change, persistence and identity.

Due to the paradoxical nature of change, change in a thing causes various problems, including the problem of the consistency of change. Some have said that the only way to make sense of change is through inconsistency (Varzi 2005). Many philosophers believe that studying and reasoning out change only make sense when things extend through “time”. It means the temporal parts of a changing “concept” can have different properties at different times (Varzi 2005). In order to talk about the identity of objects, ontologists need to distinguish between Continuants/Occurrents and Dependent/Independent and Universals/Particulars (Smith et al. 2003). According to Smith et al. (2003), Continuants (objects) are things that continue to exist through time and their identities remain unchanged. Occurrents (processes) are time dependent entities, whose identities unfold at different points in time. The existence of a “Dependent” depends on the existence of other things (i.e., a bodily injury is dependent upon the injured organ), in contrast to an “Independent”, whose existence does not necessarily depend on other things (i.e., atoms, molecules). Also, “Universals” can be considered classes or groups of things (i.e., “student”) while “Particulars” are “instances” of those classes (i.e., a specific student).

In Section 6, we will consider “time” as a primary factor in our approach to analyzing changes in temporal biomedical ontologies.

### **3 Biomedical Ontologies and the Editorial Procedure – State of the Art**

There are currently a growing number of ontologies and controlled vocabularies in various areas of life sciences. In this section, we review the state of the art of change management in some available bio-ontologies. It is not a surprise that many of them do not sufficiently meet the requirements to be considered a formal ontology (Guarino 1995). Most ontologies in the biomedical domain are recognized to be acutely defective from both terminological and ontological perspectives (Kumar and Smith 2003, Smith et al. 2003, Kumar et al. 2004, Grenon et al. 2004, Ceusters et al. 2004, Smith and Rosse 2004, Ceusters and Smith 2006, Smith and Ceusters 2006, Smith 2006). A list of open-source ontologies used in life sciences can be found on the Open Biological Ontologies (OBO) website (<http://obo.sourceforge.net/>). Many of the available ontologies are still under active development, revision and improvement, and are subject to frequent changes. The following ontologies and controlled vocabularies have been selected for a study of their change management mechanism based on several criteria, such as availability, popularity, and complexity of and accessibility to the source and documentation. The Gene Ontology (GO) (Ashburner et al. 2000) is a community standard and the Unified Medical Language System (UMLS) (Humphreys et al. 1998) is quite popular, with its rich collection of biomedical terminologies. Clinical Terms Version 2 (Cimino 1996, Bailey and Read 1999) deals with actual patient care records. We also look at HL7, FMA and Terminologia Anatomica (TA) (Whitmore 1999) to see different examples of potential changes.

#### **3.1. The Gene Ontology (GO)**

The Gene Ontology (GO) is a collaborative project (Ashburner et al. 2000) that intends to provide a controlled vocabulary to describe gene and gene product attributes in existing organisms based on their associated biological processes, cellular components and molecular functions. GO has been modeled and implemented based on three distinct ontologies, represented as directed acyclic graphs (DAGs) or networks consisting of a number of terms, represented by nodes within the graph, connected by relationships that are represented by edges (Lord et al. 2003). The current GO term count as of Dec 18, 2008 at 14:00 (PST) (<http://www.geneontology.org/GO.download.s.shtml>) is 26475 terms with 1362 obsolete terms. The GO consortium makes cross-links between the ontologies and the genes and gene products in the collaborating databases (Sklyar 2001). The Gene Ontology is currently available in Flat File, FASTA, MySQL, RDF-XML, OBO-XML and OWL formats. Members of the consortium contribute to updates and revisions of the GO. Changes in GO occur on a daily basis and a

new version of GO is published monthly, a snapshot of the current status of the database (Klein 2004). As GO becomes larger and complexity arises, it also becomes more difficult to control and maintain. To ensure consistency of the modified ontology, all changes are coordinated by a few biologists in the GO editorial office staff, who have write access to the Concurrent Versions System (CVS) (Cederqvist 1993) repository in which GO files are maintained. The users can make requests for modifications through an online system that tracks the suggestions and manages the change requests. All tracking information about requests and changes are archived and several curator interest groups have been established with associated actively archived mailing lists (Harris 2005). The GO editorial staff notifies others of the changes via monthly reports (<http://www.geneontology.org/MonthlyReports>) to the users (by email), or at the GO site. Different sources of suggested changes in GO, as described by Harris (2005), are advances in biology that alter the knowledge of gene and protein roles in cells, joining new groups that require new terms and relations, fixing errors, completing unfinished parts of the ontology, updating legacy terms and improving the formal representation of the ontology by identifying missing or misplaced relationships and terms. One of the problems in Gene Ontology maintenance is related to the versioning tool. CVS repositories, which currently handle versioning in GO, work based on syntactic differences between ontologies. For instance, CVS is not able to differentiate class versions, being able only to differentiate text/file differences (Volkel et al. 2005). The research on conceptualization change over time (Volkel et al. 2005) is still promising.

### ***3.2 UMLS Semantic Network***

The Unified Medical Language System (UMLS) (McCray and Nelson 1995) is a composite of about 100 source vocabularies that contain 870,853 concepts and 2.27 million terms (UMLS documentation, 2008). It was created by the National Library of Medicine (NLM) to facilitate the development of computer systems that behave as if they "understand" the meaning of the biomedicine/health language. To that end, the NLM produces and distributes the UMLS knowledge sources (databases) and associated software tools (programs) to system developers for use in informatics research and in building or enhancing electronic information systems that create, process, retrieve, integrate, and aggregate biomedical/health data and information. The UMLS Knowledge Sources are multi-purpose, and can utilize a variety of data and information, such as patient records, scientific literature, guidelines and public health data (UMLS documentation 2008). Due to the popularity and multi-purpose nature of the UMLS, it seems to be a perfect candidate to study change management. The UMLS Semantic Network covers different levels of granularities, which have a key effect on interpreting the meaning that has been assigned to the Metathesaurus concepts (Fact sheet of UMLS 2006). Changes in the UMLS are usually recommended by the UMLS

contractors and others who have experimented with the previous versions of the ontology. UMLS terms that share the same conceptual meaning are linked by a concept unique identifier (CUI) (Campbell et al. 1995). Two files called DELETED.CUI, which lists deleted concepts, and MERGED.CUI, which lists all pairs of CUIs that were merged, are associated with each new release of the UMLS (Olson et al. 1996). These files help users to determine whether a CUI that is no longer present in the new version was removed due to a deletion of the concept, or due to a merger of the concept with another concept (Oliver et al. 1999).

### ***3.3 Clinical terms version 3 (The Read Codes)***

The Clinical Terms Version 3 (CTV3) (<http://www.nhsia.nhs.uk/terms/pages/>) (NHS Information Authority 2000.a) or Read Codes are a set of coded terms arranged in a hierarchical structure for use in clinical practice, with such applications as viewing a patient's record from different perspectives (e.g., clinical audit, producing reports, meeting central returns, research, etc.). The CTV3 classifies chemicals by their name, i.e., alphabetically. The first version of Read Codes (CTV1) was initially developed to provide a terminology for describing relevant clinical summaries and administrative data for general practice. It is known as the 4-Byte Set since each code is four characters long. In the next version (CTV2), the codes were subsequently adapted for use in hospitals, and were extended to allow more detail. To hold more detailed information, a supplementary alphanumeric character was included in the Read Codes (5-Byte Sets) (NHS Information Authority 2000.b). CTV2 uses the code to specify a class and its unique place within the taxonomy, which has a limited, fixed number of levels. The CTV3, with its flexible structure unlike the previous versions, allows more changes in terminology (Jurisica et al., 2004). The Read Codes have been changed in each version (based on strict protocol under central control of NHS) by adding terms and codes to fix the errors and reflect the newly discovered knowledge (mostly to enrich the descriptions). Further alterations include changes to qualifiers and atoms (semantic definitions), the hierarchical structure and the mapping files (NHS Information Authority 2000.a). CTV1 and CTV2 changed relatively little between releases, due to their rigid file structure that was limited to five levels of offspring, and about 60 siblings. The CTV3 "Description Change File" (DCF) (NHS Information Authority 2000.a) shares the entire change management procedure between "terminology providers" and "terminology users" (i.e., clinicians). The DCF starts by recommending a new code for any terminology discovered to be incorrectly classified and suggesting that the user replace it. The process continues by labeling the obsolete concepts as "extinct". An example from (NHS Information Authority 2000.a) describes the deletion of the relation between the terms 'Cardiac rupture' and 'Myocardial infarct', which turned out to

have the same code in CTV2, and the addition of a new code to ‘Cardiac rupture’ in CTV3.

We also consider some other popular controlled vocabularies in life science including:

**Health Level 7 (HL7):** HL7 (<http://www.hl7.org/>) is an ontology that aims to provide a UML-based standard for the exchange, management, and integration of data to support clinical patient care and the management, delivery, and evaluation of healthcare services.

**The Foundational Model of Anatomy (FMA):** FMA (<http://sig.biostr.washington.edu/projects/fm/AboutFM.html>) represents a coherent body of explicit declarative knowledge about human anatomy and claims to be the most complete ontology of canonical human anatomy in a high granularity from the macromolecular to the macroscopic levels (Rosse and Mejino 2003).

**Terminologia Anatomica (TA):** Terminologia Anatomica (Whitmore 1999) is a standard controlled vocabulary on human anatomical terminology, developed by the Federative Committee on Anatomical Terminology (FCAT).

## 4 Different Types of Changes in Biomedical Ontologies

Based on our research of the literature, observations of different releases of ontologies, surveys, and interviews with several domain experts and ontology engineers, we distinguished 74 different types of changes that frequently occur in the life cycles of existing bio-ontologies. These changes are usually classified under ten general terms: addition, deletion, retirement (obsolescence), merging, splitting, replacement (editing or renaming), movement, importing, integration, or changes to file structure.

- **Addition:** This is one of the basic operations in ontology evolution, and aims to improve ontological structure by adding one or more components to the available structure. The most common additions in the observed bio-ontologies are of the following elements: namespace, identifier code, concept, attribute, abbreviation, super/subclass, attribute value, synonym, constraint (cardinality, type, min/max, inverse roles, and default value), associative relationships, annotation description and instance. As an example from Gene Ontology, in order to update the biological structure of GO for the annotation of genes, the GO maintenance team needs to improve the content of the ontology regularly. For instance, as described in (GO Newsletter, May 2006), the curators at MGI, who were reviewing the existing terms for the comprehensive annotation of “mammalian genes involved in the regulation of blood pressure”, realized that the existing GO terms were not sufficient to annotate these genes. They then proposed 43 new terms

(<http://www.informatics.jax.org/searches/GO.cgi?id=GO:0008217>), which were accepted after refinement in the GO discussion forum, and new annotations for the mouse genes involved in blood pressure regulation came into existence. Another example from GO (GO Newsletter, Feb 2007) is the addition of a new evidence code (<http://www.geneontology.org/GO.evidence.shtml>) used for the annotation of gene products, "Inferred from Genomic Context" (IGC).

- **Deletion:** Deletion in ontology evolution process, refers to the erasure of a selected element of an ontology when it does not reflect the ontological “truth” anymore. The most common deletions in the bio-ontologies are of the following elements: namespace, identifier code, concept, synonym, abbreviation, annotation (description), constraint (cardinality, type and min/max, inverse roles, default value), attribute value, super/subclass, associative relationships, and instance. For example, the GO terms must characterize biological entities. As stated in (GO Newsletter, May 2006), the terms classified as “Unknown” violated this principle, so the decision was made to delete the following terms, biological process unknown; GO:0000004, molecular function unknown; GO:0005554 and cellular component unknown; GO:0008372, from the ontology. The new annotations signify that a given gene product should have a molecular function, biological process, or cellular component, but that no information was available as of the date of annotation.
- **Retirement (Obsolescence):** Retirement or, as it is referred to in (Cimino 1996), obsolescence means diminishing the role of an older ontological element when a newer, more functional element or meaning supersedes it. In this situation, the older version can be kept somewhere for future use, but its usage will be discouraged. The retirement process usually involves concepts, attributes, identifier codes, instances and associative relationships. For example in Release 2.0 of HL7, the certain components, ClinicalDocument.copyTime, MaintainedEntity, CodedEntry, inkHtml.name, table.border, table.cellspacing and table.cellpadding, are retained for backwards compatibility with HL7 Clinical Document Architecture (CDA), Release 1.0, and have been retired. Further use of these components is discouraged (Dolin 2004).
- **Merging:** This is defined as the process of creating a consistent and coherent ontological element that includes information from two or more basic elements. The process of merging can be seen in this format: merging two or more concepts/associative relationships/identifier codes/attributes into one of the same or into a new concept/relationship/code/attribute (Oliver et al. 1999). For instance, in HL7, the purpose of the header is to enable clinical document management, compilation and exchange across institutions (Dolin 2004). In HL7's Clinical Document Architecture (CDA), Release 2.0, two concepts in the header (service\_actor and service\_target) have been merged (Dolin 2004).

- **Splitting:** An ontological element may be split into two or more new elements. This means that a concept, associative relationship, identifier code, or attribute can be split into two or more of the same. For example, in Terminologia Anatomica (TA) (Whitmore 1999), terms that share an identifier code are considered synonyms. However, this is not valid for sexually dimorphic anatomical parts, such as ‘Ovarian artery’ and ‘Testicular artery’, as they share the same TA code (A12.2.12.086). The two arteries are not synonymous but distinct, and are involved in different relationships. Therefore, they need to be treated as two separated concepts, meaning the code A12.2.12.086 can be split into A12.2.12.086-1 for ‘Ovarian artery’ and A12.2.12.086-2 for ‘Testicular artery’ (Whitmore 1999).
- **Replacement (Edit, Renaming):** This process is for editing available labels and values. Editing mainly occurs to change namespace, concept name/definition/role, and attribute value/name/definition. A typical scenario (Dolin, 2004) from HL7, Release 2.0, is a simple replacement of

ClinicalDocument.id "1.2.345.6789.123"

With

ClinicalDocument.id "1.2.345.6789.266"

Another example from Terminologia Anatomica (TA): TA renames the ‘Angiologia’ chapter of Nomina as ‘Cardiovascular system’. This new name resolves the inconsistency in Nomina Anatomica (the older anatomy classification), which assigned the heart to the ‘system of vessels’ (Angiologia), rather than to viscera (Splanchnologia) (Rosse 2000).

- **Movement (Transition):** This is defined as the transition of one or more ontological elements across the ontological hierarchy. This transition can happen to identifier codes, concepts, attributes, super/subclass, associative relationships, and instances. For example, GO terms representing transporter activity in the Molecular Function sub-ontology are gradually being overtaken to better represent current scientific knowledge. A new high-level term called “transmembrane transporter activity” (GO:0022857) was introduced. In addition, the related child terms and subclasses have been moved and organized under GO terms that describe the activity of the transporters, such as channel, active transporter, symporter, antiporter and uniporter activities (GO Newsletter, Aug 2007).
- **Importing:** This refers to the process of bringing an existing ontology (a tree) or parts of an existing ontology (a sub-tree) into another ontological structure. As an example from Gene Ontology, in 2001, the GO developers imported the first pass annotation from SWISS-PROT (<http://ca.expasy.org/sprot/>), trEMBL (<http://www.ebi.ac.uk/trembl/>), and Ensembl (<http://www.ensembl.org/index.html>) (GO

Meeting collected notes, 2001). Also, 7,316 GO annotations were imported from Proteome and literature associations (GO Meeting collected notes, 2001).

- **Integration:** In data integration, scattered process data is extracted from different sources with different data formats, and then “normalized into a consistent syntactic representation and semantic frame of reference” (Buttler et al. 2002). Semantic integration is more complex than data integration. For example (Martin et al. 2001), in the Foundational Model of Anatomy (FMA), to have a comprehensive knowledgebase in neuroinformatics, the FMA developers have integrated several terminologies of NeuroNames and Terminologia Anatomica into the FMA. They have enhanced the FMA to support unique information on neuronal structures (Martin et al. 2001).
- **Changes to release file (File structure):** By the advancement of technology for storing and retrieving data files and the emergence of new standards, the format of file structures can be changed. As an example, in the first version of Read Codes, four-character alphanumeric codes determined the position of a term in a hierarchy; this version is known as the 4-Byte Set (Bentley 1996). The restrictions imposed by only four levels of hierarchy led to the development of a 5-Byte Set, which expanded the set to support secondary and tertiary care. As stated in (Robinson et al. 1997) “this set was released in two structurally different versions, and has increased content and provided a more structured mechanism for representing synonyms”. Ver. 1.0 has shorter terms and keys than Version 2.0. Ver. 3.0 with its rich structure supports the character structures of both Versions 1.0 and 2.0 (Robinson et al. 1997).

A comprehensible survey on different types of ontological changes and relations between change management practices with others overlapping research disciplines can be found in (Flouris et al. 2008).

## 5 Tools and Methods to Support Ontology Change Management

There are a few tools (Haase and Sure 2004, Stojanovic 2004) to manage changes in ontologies. Some of the available tools are simply ontology editors, such as Protege (Noy et al. 2000), OntoEdit (Sure et al. 2003) and TopBraid Composer (<http://www.topquadrant.com/topbraid/composer/index.html>). Despite their differences, they all assist users in implementing, updating and managing elementary changes in ontologies. According to (Stojanovic 2004, Stojanovic and Motik 2002), the most critical requirements for ontology editors in order to be more robust in a changing environment are related to functionality, customizability, transparency, reversibility, auditing, refinement and usability. Other available tools include but are not limited to

Concurrent Version System (CVS) (Cederqvist 1993), CONCORDIA (Oliver and Shahar 2000), KAON (Maedche and Staab 2003, Gabel et al. 2004) management tool, OntoView (Klein et al. 2002), OntoManager (Stojanovic et al. 2003) and TextToOnto (Maedche and Volz 2001). Table 1 represents some of the popular ontology editors and management tools with their descriptions.

**Table 1.** Some of the ontology editors and management tools

Tool	Description
<b>Protege</b> (Noy et al. 2000) (ver. 4.0 beta with web 2.0 support)	A popular ontology design environment with support for RDF and OWL ontologies. It provides some editing facilities such as: adding/deleting/renaming ontological elements, undo/redo of changes and version archiving (Liang et al. 2005). Protege also includes plug-ins such as PROMPT for managing multiple ontologies. It can compare versions of the same ontology, merge two ontologies into one and extract part of an ontology (Noy and Musen 2003). PromptDiff (Noy and Musen 2004) also can determine the changes between two versions.
<b>TopBraid Composer</b> (TopBraid Composer Guide 2007)	A commercial ontology editor that supports editing RDF Schemas and OWL Ontologies, as well as executing rules and queries in the SPARQL Protocol and RDF Query Language (SPARQL) (Beckett 2006) and the Semantic Web Rule Language (SWRL) within a multi-user environment. It manages multiple versions of ontologies by using the following set of rules. Any changes to the statements are written into the source ontology. If the change is “overtyping” an entry, it will be saved in the original ontology as an update. In case of the “deletion” of an entry and then the “addition” of a new one, the deletion would be done in the original file and the new triple would be saved in the existing file. Also, by changing any class, the composer scans to see if there are any other ontologies that import this class. It keeps a log of the changes that is accessible from the Change History view. Unsaved changes can be undone. To prevent accidental changes, a file can be defined as “read only”.
<b>Concurrent Version System (CVS)</b> (Cederqvist 1993) (www.nongnu.org/cv)	Supports basic version control functionality and maintains a history of the changes. CVS can reveal syntactical and textual differences between two files. It mostly works on the syntactic level. Since ontology versioning and change management need operations on the conceptual level rather than the syntactic level, CVS might not seem an appropriate tool for ontology change management (Völkel and Groza 2006). However, CVS can provide basic support for managing structural changes in RDF and OWL files.

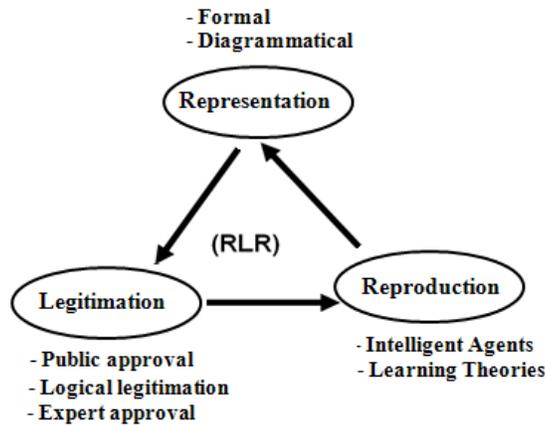
Bio-Medical Ontologies Maintenance and Change Management

<p><b>CONCORDIA</b> (Oliver and Shahar 2000)</p>	<p>A model for managing divergence in concept-based terminologies, developed to facilitate the study of synchronization in health care terminologies. CONCORDIA uses the models of Medical Subject Headings (MeSH) (Nelson et al. 2001), ICD-9-CM (Cimino 1996), and ICD-10. It enables one to manage 27 different kinds of changes, such as adding, deleting, retiring, or merging concepts, terms or attributes (Oliver and Shahar 2000). CONCORDIA does not provide any services to log motivations for the changes (Ceusters and Smith 2006).</p>
<p><b>KAON</b> (Maedche and Staab 2003), ( Gabel et al. 2004). <a href="http://kaon.semanticweb">http://kaon.semanticweb</a></p>	<p>An integrated open-source ontology management system targeted at semantics driven business applications, KAON components can be divided into 3 layers: (i) The applications/services layer realizes user interface applications and provides interfaces to non-human agents; (ii) The API, which is the major part of KAON, checks the validity of change sequences, and also requests user approval for performing a change, justifies the necessity of a particular change, executes the modifications, reverses the effect of some undesirable changes and keeps a history of changes; (iii) The data and remote services layer provides data storage facilities. See (Gabel et al. 2004) for more information.</p>
<p><b>OntoView</b> (Klein et al. 2002)</p>	<p>A web-based system that assists users in handling ontology evolution. The system helps to keep different versions of web-based ontologies interoperable by maintaining the transformations between ontologies and the relations between concepts in different versions. OntoView was inspired by and can be considered a Web interface for CVS. OntoView compares ontologies at a conceptual level, analyzes effects of changes (e.g., by checking consistency and highlighting the places in the ontology where conceptually changed concepts or properties are used) (Klein et al. 2002)) and utilizes changes.</p>
<p><b>OntoManager</b> (Stojanovic et al. 2003)</p>	<p>Has been designed to assist ontology managers in managing ontologies according to the users' requirements. The technique used to evaluate users' needs depends on the information source by tracking user interactions with the application in a log file. The OntoManager consists of three modules: (i) The data integration module, which aggregates, transforms, and correlates the usage data; (ii) The visualization module that presents the integrated data in a comprehensible visual form; and (iii) The analysis module, as the major part of the change management, provides guidance for adapting and consistently improving the ontology with respect to the users' requirements. This module keeps track of the changes and has the ability to undo any action taken upon the ontology.</p>
<p><b>TextToOnto</b> (Maedche and Volz 2001)</p>	<p>A tool suite built upon KAON in order to support the ontology engineering process by text mining techniques. Since TextToOnto does not keep any references between the ontology and the text documents it has been extracted from, it does not allow for mapping textual changes to the ontology. Therefore data-driven change discovery is not supported by this tool.</p>

As can be seen from the current state-of-the-art change management in existing ontologies in life sciences, the current biomedical ontologies do not follow any standard, consistent, formal change models with clear semantics. Most of the available tools are just simple ontology editors with a few extended features. Some parts of ontology evolution, such as the change representation and conceptualization change, are not satisfactorily managed by existing tools and they are left to be handled by the users. The major issues in available ontology management tools can be summarized as: (i) Too much reliance on human decisions due to lack of fully automatic ontology change management tools and too much dependency of the existing systems on the human factor (Haase and Sure 2004), which both give rise to several issues relating to complexity, accuracy, security and reproducibility (Flouris 2006); (ii) Representation and tracking of complex changes using available technologies are limited; (iii) Lack of formal evaluation methods, which makes the comparison and evaluation of different algorithms extremely difficult (Flouris 2006); (iv) Little or no support for conceptualization change management; and (v) Change models that have been designed based on time/space independent ontologies.

## 6 The RLR Framework for Ontology Change Management

Here we propose our approach, based on the RLR framework (Shaban-Nejad and Haarslev 2008), for recruiting intelligent agents to capture the pattern of changes, track a change, predict the rate and direction of changes and validate the results. The RLR framework will be used to Represent, Legitimate and Reproduce the changes and their effects (Fig. 2).



**Fig.2.** A general view of the RLR framework.

In this structure, all changes are to be captured, represented and then validated logically and approved publicly and by experts. To support reproducibility of the outcomes and automate the ontology evolution process, agents can be used to learn the patterns of changes and their consequences in the Reproduction phase. The Representation phase updates the representations (either formal or diagrammatical) of the new knowledge in a consistent way. For formal representation, we use Description Logic (DL) (Baader et al. 2003), and for diagrammatical representation, we employ a method based on discrete state model and category theory (Shaban-Nejad and Haarslev 2007(a)). The Legitimation stage determines the consistency and validity of a change. This phase assesses the impact of a potential change before the change is actually made. Software agents can search, notify, perceive and collect information about different actions from multiple sources in a frequently changing environment. Intelligent agents perform rationally in dynamic environments and provide prompt responses. In our framework, we have employed various types of agents: the change capture, learner, reasoning and negotiation agents.

The change capture agents perceive and capture different changes and accumulate the related information within the change logs. The learner agent can use these records of changes, which occur over and over in a change management procedure, to design a change pattern. This means that after several changes—most likely in different releases—the direction and rate of change can be predicted with reasonable accuracy. The learner agent and the change patterns are essential in the Reproduction phase. A reasoning agent controls the logical validity of the ontological structure and reveals inconsistencies, hidden dependencies, redundancies and misclassifications. RACER (Haarslev and Möller 2001) has been used in RLR as a Description Logic reasoning agent. The negotiation agent acts as a mediator to assist the ontology engineer and other autonomous agents in negotiations concerning the realization of particular change. The negotiation outcome will be confirmed, deleted, or modified by human experts in accordance to the purpose of the system.

We have used category theory for mathematical notation, which is independent of a specific choice of ontology language and any particular implementation, as the main formalism of the RLR framework. Furthermore, we will demonstrate the applicability of category theory for representing changes in the ontological structure.

### ***6.1 Representation of Evolving Ontologies Using Category Theory***

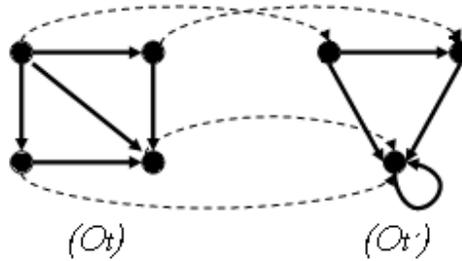
We noticed that the main issues in change management strategies are simply problems about the representation of change. Category theory is an intuitive domain of mathematics, introduced in 1945 (Eilenberg and MacLane 1945). Category theory is grounded in logic and algebra, and allows an ontology engineer to define different

states of conceptual models to represent the reality. Categories can be used for specifying different objects, capturing their interactions and the relations between them, determining patterns of interacting objects and decomposing a composite object into its basic elements (Ehresmann and Vanbreemsersch 2006). Categorical representations consist of diagrams with arrows (i.e.,  $f: X \rightarrow Y$ ). A typical category includes classes of objects and morphisms (arrows), and for each morphism  $f$ , there is one object as the domain of  $f$  (i.e.,  $X$ ) and one object as the codomain (i.e.,  $Y$ ). In addition, for each object,  $X$ , we can define an identity morphism ( $\text{id}_X$ ) which has domain and codomain  $X$ . Also, for each pair of morphisms  $f: X \rightarrow Y$  and  $g: Y \rightarrow Z$ , (in other words,  $\text{cod}(f) = \text{dom}(g)$ ) a *composite morphism* ( $g \circ f: X \rightarrow Z$ ) exists. The representation of a category can be formalized using the notation of a diagram. Ontologies in their simplest form can be seen as the categorization of things in the real world. It is quite common to use diagrammatical notations to analyze ontological structures. For example, the UMLS Semantic Network (SN) has been modeled as a graph, whose nodes denote concepts (McCray 2003). Category theory, with its analytical characteristics, can act as the prospective knowledge representation language, particularly for modeling dynamic abstract models, where ontologies can be defined as an interconnected hierarchy of categories (Healy and Caudell 2006) with directed ontological relationships as “morphisms” (Krötzsch 2005). As a knowledge representation language, category theory improves the incremental analysis of changes in ontological structures. Some of the primitive constructors of category theory, which we use in our framework for ontology change management, are as follows: primary categorical objects and morphisms, functors, pushouts, pullbacks, natural transformations and isomorphisms.

*Functors* are a special type of mapping between categories, defined as morphisms in the category of all small categories (classes are defined as categories) (Awodey 2006). As defined in (Van Oosten 2002), assume  $A, B$  are two categories, so a functor  $F: A \rightarrow B$  is joint functions:  $F_0: A_0 \rightarrow B_0$  and  $F_1: A_1 \rightarrow B_1$ , such that:

$$\begin{aligned} \forall f: \alpha \rightarrow \beta \in A_1 &\Rightarrow F_1(f): F_0(\alpha) \rightarrow F_0(\beta) \in B_1 \\ \forall \alpha \in A_0 &\Rightarrow F_1(\text{id}_\alpha) = \text{id}_{F_0(\alpha)} \\ \forall g \circ f \in A_1 &\Rightarrow F_1(g \circ f) = F_1(g) \circ F_1(f) \in B_1 \end{aligned}$$

A categorical model can represent transitions between different states of an ontological structure. To represent and track this kind of transition, we represent the conceptualization of things indexed by time (i.e., from the FungalWeb Ontology (Baker et al. 2006), “enzyme has KM value at  $t$ ” is rendered as “enzyme-at- $t$  has KM value”). Then we represent a set of time-indexed categories using functors to capture different states of ontological structure at different points in time. The category  $O$  at time  $t$  ( $O_t$ ) models the state of the ontologies and all related interactions at this time. A functor can represent the transition from  $O_t$  to  $O_{t'}$  (Fig. 3), where the time changes from  $t$  to  $t'$ . Also, each sub-ontology  $A$  can be modeled by the series of its successive states,  $A_t$ , from its creation to its destruction (Ehresmann and Vanbreemsersch 2006).



**Fig. 3.** Using Functor for studying model transition between different states.

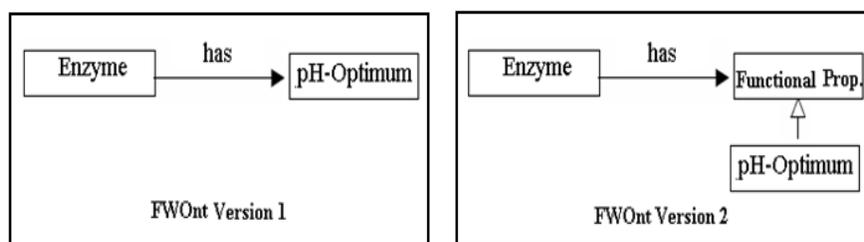
Also, categorical arrows (morphisms) can be used to describe different conditional changes and to measure coupling as the extent of the connections between elements of a system (Shaban-Nejad and Haarslev 2007(a)). Coupling identifies the complication of a changing system. Measuring coupling is useful for predicting and controlling the scope of changes to an ontological application. Often a change in one class can cause some changes to the dependent classes. When the coupling is high, it indicates the existence of a large number of dependencies in an ontological structure, which must be checked to analyze and control the chain of changes. Coupling for ontological elements can be described by the number of connections and links between them. Therefore, we focus on arrows in category theory to study these connections. For analyzing a conditional change, we followed the formal model described in (Whitmire 1997) by identifying three types of arrows in our category: precondition, post-condition and message-send arrows for an existing category (Whitmire 1997). The type of message is determined by the types of changes caused by a method.

## 6.2 Application Scenario

For the application scenario, we have applied our method for managing changes to the FungalWeb Ontology (Baker et al. 2006). The FungalWeb Ontology is a formal ontology in the domain of fungal genomics, which provides a semantic web infrastructure for sharing knowledge using four distinct sub-ontologies: enzyme classification based on their reaction mechanism, fungal species, enzyme substrates and industrial applications of enzymes. The ontology was developed in OWL-DL by integrating numerous online textual resources, interviews with domain experts, biological database schemas (e.g., NCBI (Wheeler et al., 2000), EC, NEWT (Phan et al. 2003), SwissProt (Bairoch 2000), Brenda (Schomburg et al. 2004)) and reusing some existing bio-ontologies, such as GO and TAMBIS (Baker et al. 1998). Fungi are widely used in industrial, medical, food and biotechnological applications. They are also related to many human,

animal and plant diseases. It is estimated that there are about 1.5 million fungal species (Heywood 1995) on the earth, but only about 10% of those are known and only a few of them have an identified usage. The fungal taxonomy is frequently changing. Most of the alterations are changes in names and taxonomic structure and relationships.

- **Changes in Names:** Fungal names reflect information about the organisms. Therefore, as our knowledge increases, names need to be changed when they no longer express the correct information (Crous 2005). These changes may cause misunderstanding and miscommunication, and affect the soundness of different queries. Crous et al. (2003) describe an example of eyespot disease in cereals and the issues related to naming its associated fungi. Most fungi names are currently based on their visible characteristics and appearances. To manage this process of constant evolution, one solution is employing ontologies, where names are only meaningful once linked to descriptive information that was extracted from trustworthy data sources. The incorporation of DNA data is also crucial to ensure stability in names and uniquely distinguish the species. At this time, only about 1.1% of the estimated 1.5 million species are represented by DNA sequence data, meaning very few have been preserved from change (Hawksworth 2004).
- **Changes in Taxonomic Structure:** By advancing our knowledge in life sciences, fundamental changes in taxonomical structure and relationships can be foreseen. For instance, by studying some molecular, morphological and ecological characteristics, *Glomeromycota* was discovered in 2001 (Schüßler et al. 2001) as a new fungal phylum. Another example is the sedge parasite *Kriegeria eriophori*, which has never been satisfactorily classified. Recently, ribosomal RNA gene sequences and nucleus-associated ultrastructural characters were analyzed separately and combined to define the new subclass, *Microbotryomycetidae* (Swann et al. 1999). Fig. 4 represents how the place of the concept “pH optimum” has been changed within the FungalWeb taxonomy (ver. 2.0) by adding the new concept “Functional Property”.



**Fig 4.** A simple change in taxonomical structures of two consecutive versions of the FungalWeb Ontology (FWOnt).

Fig. 5 demonstrates a portion of the FungalWeb application in categorical representation.

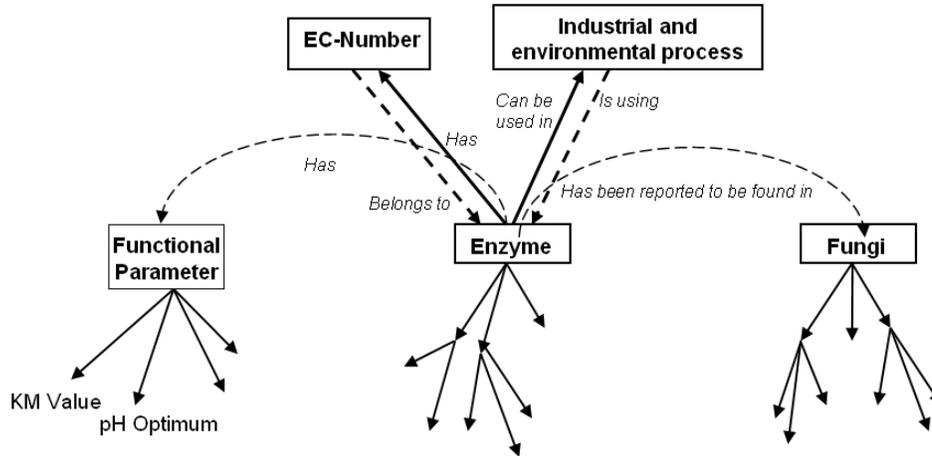


Fig. 5. A category model of portion of the FungalWeb application.

Most common operations during ontology evolution, such as adding or deleting a class/property/relationship, combining two classes into one, or adding a generalization/association relationship, can be represented using category theory (Shaban-Nejad and Haarslev 2007(b)). Based on our application, we designed our class diagrams following Whitmire (1997) to track different states of our ontological structure (Fig. 6). The  $Op_i$  arrows in this figure represent the operations, which cause an ontological state to change to another state. For instance, in Fig. 11, the operation  $Op_1$  causes the transition of an object from state  $St_1$  to state  $St_2$ . The  $Op_r$  (reverse operation) returns the system to its initial state ( $St_1$ ). For more details see (Shaban-Nejad and Haarslev 2007, 2008)

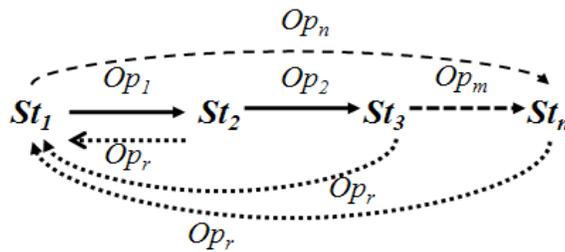


Fig. 6. A class diagram to track different operations and states in an evolving structure.

## Summary

Biology is known as a field with continuous evolution. As the knowledge about biology rapidly grows and new methods become available, one can anticipate a fundamental change in the way we capture the knowledge of the real world. One of the important activities in knowledge representation and bioinformatics is properly responding to changes and coping with the ontological evolution. Research on ontology change management is an ongoing effort that is still in its early stages. Many tools and techniques are available to assist ontology engineers to implement changes. However, they still have long road ahead to be considered for practical usage due to following issues:

- Lack of formal change models with clear semantics
- Inconsistencies among change models and log models
- Too much reliance on human decisions
- Reproducibility of the results cannot be guaranteed
- Little or no support for the representation of complex changes
- Lack of formal evaluation methods
- Little support for handling changes in conceptualization

The proposed RLR framework can be used to Represent, Legitimate and Reproduce the changes and their effects. Using this framework can help capture, track, represent and manage the changes in a formal and consistent way, which enables the system to create reproducible results. In this framework, all changes should be represented in either formal or diagrammatical representations. The change then should be legitimated and validated logically, by a reasoning agent, and it should be approved publicly and by experts. In order to reproduce the results of changes and automate the change management process, agents can be used to learn the patterns of changes, using learner agents, while the negotiation agent acts as a mediator that allows the ontology engineer and other autonomous agents to negotiate the proper implementation of a specific change. We are currently using the RLR framework to manage the evolving structure of the FungalWeb Ontology. Due to importance of the Representation phase, we concentrated more on this issue in this manuscript. For diagrammatical representation, we proposed using category theory, which has significant potential as a supplementary tool for capturing and representing the full semantics of ontology-driven applications. Category theory can provide a formal basis for analyzing complex evolving biomedical ontologies. Our future research will be focused on improving the reasoning process and generalizing the usage of category theory along with other formalisms to improve ontological conceptualization change management. Due to the multidisciplinary nature of research on ontology change management, any advances in this field would be highly dependent on advances in the research of various related topics, such as ontology integration, translation, merging or alignment, and computer-human interactions.

## References

- Ashburner M, Ball CA, Blake JA, Botstein D et al (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25–29.
- Awodey S (2006) *Category Theory*. Oxford University Press.
- Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (2003) *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Bailey PS, Read J (1999) Software implementation of clinical terminologies: The use of component technology (tutorial), AMIA '99 Annual Symposium. Washington, DC.
- Bairoch A (2000). The ENZYME database in 2000. *Nucleic Acids Res* 28: 304-305.
- Baker PG, Brass A, Bechhofer S, Goble C, Paton N, Stevens R (1998) TAMBIS-Transparent Access to Multiple Bioinformatics Information Sources. *Proc Int Conf Intell Syst Mol Biol*; 6:25-34.
- Baker CJO, Shaban-Nejad A, Su X, Haarslev V, Butler G (2006) Semantic Web Infrastructure for Fungal Enzyme Biotechnologists. *Journal of Web Semantics*, 4(3):168-180.
- Beckett D (2006) SPARQL RDF Query Language Reference v1.8. <http://www.dajobe.org/2005/04-sparql/SPARQLreference-1.8.pdf>. Accessed 10 Jan 2009.
- Bentley TE, Price C, Brown PJB (1996) Structural and lexical features of successive versions of the Read Codes. in *Proceedings of the Annual Conference of the Primary Health Care Specialist Group*. Worcester, 91-103. <http://www.phcsg.org/main/pastconf/camb96/readcode.htm>.
- Bodenreider O, Stevens R (2006) Bio-ontologies: current trends and future directions. *Brief Bioinform* 7(3):256-274.
- Buttler D, Coleman M, Critchlow T, Fileto R et al. (2002) Querying Multiple Bioinformatics Data Sources: Can Semantic Web Research Help? *SIGMOD Record* 31(4):59-64.
- Campbell KE, Oliver DE, Spackman KA, Shortliffe EH (1998) Representing Thoughts, Words, and Things in the UMLS. *J Am Med Inform Assoc* 5(5): 421–431.
- Cederqvist P, Version Management With CVS Copyright 1993-2005 Free Software Found Inc. <http://ftp.gnu.org/non-gnu/cvs/source/stable/1.11.22/cederqvist-1.11.22.pdf>. Accessed 10 Jan 2009.
- Ceusters W, Smith B (2005) A Terminological and Ontological Analysis of the NCI Thesaurus. *Method Inform Med* 44:498-507.
- Ceusters W, Smith B (2006) A Realism-Based Approach to the Evolution of Biomedical Ontologies. in *Proceedings of AMIA Annual Symposium*.
- Ceusters W, Smith B, Kumar A, Dhaen C (2004) Ontology-Based Error Detection in SNOMED-CT. In *Proc. of the 11<sup>th</sup> World Congress on Medical Informatics; MEDINFO04*, IOS Press; 482-486.
- Cimino JJ (1996) Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. *Method Inform Med* 35(3):202–210.
- Cimino JJ (1996) An approach to coping with the annual changes in ICD9-CM. *Method Inform Med* 35(3):220.
- Cimino JJ, Clayton PD (1994) Coping with changing controlled vocabularies. In *Proc. Annu Symp Comput Appl Med Care*. 1994(Nov):135-139
- Cohen SM (2004) Identity, Persistence, and the Ship of Theseus. Department of philosophy, University of Washington ( <http://faculty.washington.edu/smcohen/320/theseus.html>)
- Crous W (2005) Plant pathology is lost without taxonomy. *Outlooks on Pest Management* 16:119-123.
- Crous PW, Groenewald JZ, Gams W (2003) Eyespot of cereals revisited: ITS phylogeny reveals new species relationships. *European J. Plant Pathol* 109:841–50.
- Crous PW, Groenewald JZ (2005) Hosts, species and genotypes: opinions versus data. *Australas Plant Path* 34(4):463–470.
- Dolin RH, Alschuler L, Boyer S, Beebe C et al (2004) HL7 Clinical Document Architecture, Release 2.0. Published on 12/12/2004 ([http://www.e-ms.ca/documents/pdf\\_v3ballotCDA\\_2005Jan.pdf](http://www.e-ms.ca/documents/pdf_v3ballotCDA_2005Jan.pdf))
- Ehresmann AEC, Vanbreemsch JP (2006) The Memory Evolutive Systems as a Model of Rosen's Organism-(Metabolic, Replication) Systems. *Axiomathes* 16:137–154.

- Eilenberg S, Mac Lane S (1945) General Theory of Natural Equivalences. *T Am Math Soc* 58:231–294.
- Fact Sheet of Unified Medical Language System (UMLS) Semantic Network (2006) last updated on 28 Mar 2006. <http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html>. Accessed 10 Jan 2009.
- Flouris G (2006) On Belief Change and Ontology Evolution. Ph.D. thesis in the Department of Computer Science, University of Crete.
- Flouris G, Manakanatas D, Kondylakis H, Plexousakis D, Antoniou G (2008) Ontology change: classification and survey. *Knowl Eng Rev* 23(2):117-152.
- Gabel T, Sure Y, Voelker J (2004). KAON – ontology management infrastructure. SEKT informal deliverable 3.1.1.a, Institute AIFB, University of Karlsruhe. <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/SEKT-D3.1.1.a.pdf>. Accessed 10 Jan 2009.
- GO Browser representing blood pressure terminologies. <http://www.informatics.jax.org/searches/GO.cgi?id=GO:0008217>. Accessed 10 Jan 2009.
- GO Meeting collected notes, July 14-15,2001, Hosted by Judy Blake and the Jackson Lab in Bar Harbor, ME. compiled by L. Reiser. [http://www.geneontology.org/minutes/collected\\_minutes.txt](http://www.geneontology.org/minutes/collected_minutes.txt)
- GO Newsletter, No. 1 May 2006. <http://www.geneontology.org/newsletter/archive/200605.shtml#bp>
- GO Newsletter, No. 4 February 2007. <http://www.geneontology.org/newsletter/archive/200702.pdf>
- GO Newsletter, No. 5 May 2007. <http://www.geneontology.org/newsletter/archive/200705.pdf>
- GO Newsletter, No. 6 Aug 2007. <http://www.geneontology.org/newsletter/current-Newsletter.shtml>
- Grenon P, Smith B, Goldberg L (2004) Biodynamic ontology: Applying BFO in the Biomedical Domain, in Pisanelli DM (ed). *Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies*, Rome, IOS Press, *Studies in Health Technology and Informatics*, vol 102:20-38.
- Gruber TR (1993) A translation approach to portable ontologies. *Knowl Acquis* 5(2):199-220.
- Guarino N (1995) Formal Ontology, Conceptual Analysis and Knowledge Representation. *Int J Hum-Comput St* 43(5/6):625–640.
- Haarslev V, Möller R (2001) RACER System Description. In *Proceedings of the First International Joint Conference on Automated Reasoning (IJCAR 2001)*, 701-706.
- Haase P, Sure Y (2004) D3.1.1.b State-of-the-Art on Ontology Evolution. Institute AIFB, University of Karlsruhe, Deliverable D3.1.1.b, EU-IST Project IST-2003-506826 SEKT. <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/SEKT-D3.1.1.b.pdf>. Accessed 10 Jan 2009.
- Hardie RP, Gaye RK (translators) (2004) *Physics by Aristotle*. eBooks@Adelaide, Adelaide University: (<http://etext.library.adelaide.edu.au/a/aristotle/a8ph/index.html>)
- Harris MA (2005) Why GO there? Ensuring that the Gene Ontology Meets Biologists' Needs. The Gene Ontology Consortium and EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, UK.
- Hawksworth DL (2004) Fungal diversity and its implications for genetic resource collections. *Stud Mycol* 50:9–17.
- Healy MJ, Caudell TP (2006) Ontologies and Worlds in Category Theory: Implications for Neural Systems. *Axiomathes* 16:165–214.
- Heywood VH (ed.) (1995) *Global Biodiversity Assessment*. Cambridge University Press.
- Humphreys BL, Lindberg DAB, Schoolman HM, and Barnett GO (1998) The Unified Medical Language System An Informatics Research Collaboration. *J Am Med Inform Assoc* 5(1): 1–11.
- Jones TC (1998) *Estimating Software Costs*; McGraw-hill.
- Jurisica I, Mylopoulos J, Yu ESK (2004) Ontologies for Knowledge Management: An Information Systems Perspective. *Knowl Inf Syst* 6(4): 380-401.
- Kemerling G (2006) The Origins of Western Thought, Kemerling philosophy page, last update on Dec 20, 2006. (<http://www.philosophypages.com/hy/2b.htm#hera>)
- Klein M (2004) Change Management for Distributed Ontologies. Ph.D thesis, Vrije University.
- Klein MCA, Fensel D, Kiryakov A, Ognyanov D (2002) Ontology Versioning and Change Detection on the Web. In *proceedings of EKAW 2002*, LNCS: 197-212.
- Klein MCA, Noy NF (2003) A Component-Based Framework for Ontology Evolution. In *Proceedings of the IJCAI-03*, CEUR-WS, vol. 71.

## Bio-Medical Ontologies Maintenance and Change Management

- Krötzsch M, Hitzler P, Ehrig M, Sure Y (2005) Category Theory in Ontology Research: Concrete Gain from an Abstract Approach. Technical Report, AIFB, U of Karlsruhe.
- Kumar A, Schulze-Kremer S, Smith B (2004) Revising the UMLS Semantic Network. In Proceedings of the 11<sup>th</sup> World Congress on Medical Informatics MEDINFO2004. IOS Press, 1700-4.
- Kumar A, Smith B (2003) The Unified Medical Language System and the Gene Ontology, KI2003: Advances in Artificial Intelligence, LNCS 2821:135-48.
- Liang Y, Alani H, Shadbolt N (2005) Ontology Change Management in Protege. In Proceedings of the 1st AKT Doctoral Symposium.
- Lord PW, Stevens RD, Brass A, Goble CA (2003) Semantic similarity measures as tools for exploring the Gene Ontology. Pacific Symposium on Biocomputing 8:601-612.
- Maedche A, Staab S (2003) KAON-The Karlsruhe Ontology and Semantic Web Meta Project. *Künstliche Intelligenz (KI)* 17(3): 27-30.
- Maedche A, Volz R (2001) The ontology extraction and maintenance framework text-to-onto. In Proceedings of the ICDM'01 Workshop on Integrating Data Mining and Knowledge Management, San Jose, California, USA.
- McCray AT (2003) An upper-level ontology for the biomedical domain. *Comp Funct Genom* 4: 80-84
- McCray AT, Nelson SJ (1995) The representation of meaning in the UMLS. *Method Inform Med* 34 (1/2):193-201.
- MacLane S (1971) Categories for the Working Mathematician. (corrected 1994), Springer.
- Martin RF, Mejino J LV, Bowden DM et al. (2001) A Next Generation Knowledge Source: Foundational Model of Neuroanatomy. in Proc. Society for Neuroscience Annual Meeting, pages 23-48.
- Magee J (1999) The Problem of change. In the website of Thomistic Philosophy at the Center for Thomistic Studies at the University of St. Thomas, Houston, Texas, last update on 1999. <http://www.aquinasonline.com/Topics/change.html>. Accessed 10 Jan 2009.
- Nelson S J, Johnston, D, Humphreys B L (2001) Relationships in Medical Subject Headings. In: Bean, Carol A.; Green, Rebecca, editors. Relationships in the organization of knowledge. New York: Kluwer Academic Publishers,171-184.
- NHS Information Authority (2000.a) The Clinical Terms Version 3 (The Read Codes): Managing Change: Description Change File. Reference Number1999-IA-173 v1.0.
- NHS Information Authority (2000.b) The Clinical Terms Version 3 (The Read Codes): Introduction and Overview. Reference Number 1999-IA-166 v1.0. . <http://www.connectingforhealth.nhs.uk/systemsandservices/data/readcodes/docs/chap1.pdf>. Accessed 10 Jan 2009.
- Noy NF, Chugh A, Liu W, Musen M (2006) A Framework for Ontology Evolution in Collaborative Environments. In proceedings of the 5th International Semantic Web Conference (ISWC'06), Athens, GA, USA, LNCS: 544-558.
- Noy NF, Fergerson RW, Musen MA (2000) The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility. In proceedings of 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2000), French Riviera, LNCS: 17-32.
- Noy NF, Musen MA (2003) The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. *Int J Hum-Comput St* 59(6): 983-1024.
- Noy NF, Musen MA (2004) Ontology versioning in an ontology management framework. *IEEE Intelligent Systems* 19(4) 6-13.
- Oliver DE (2000) Change management and synchronization of local and shared versions of a controlled vocabulary, Ph.D. thesis, Stanford University.
- Oliver DE, Shahar Y (2000) Change management of shared and local versions of health-care terminologies. *Method Inf Med* 39(4/5):278-290.
- Oliver D, Shahar Y, Shortliffe EH, Musen MA (1999) Representation of change in controlled medical Terminologies. *Artif Intell Med* 15:53-76.
- Olson NE, Erlbaum MS, Tuttle MS et al (1996) Exploiting the metathesaurus update model. in Proc. of the 18<sup>th</sup> Symp. on Computer App. in Medical Care. Philadelphia: Hanley and Belfus, 902.

- Phan IQH, Pilbout SF, Fleischmann W, Bairoch A (2003) NEWT, a new taxonomy portal, *Nucleic Acids Res* 31(13):3822-3823.
- Robinson D, Comp D, Schulz E, Brown P et al. (1997) Updating the Read Codes: User-interactive Maintenance of a Dynamic Clinical Vocabulary. *J Am Med Inform Assoc* 4(6):465-472.
- Rosse C (2000) Terminologia Anatomica; Considered from the Perspective of Next-Generation Knowledge Sources. *Clinical Anatomy* 14:120-133.
- Rosse C, Mejino Jr JLV (2003) A reference ontology for bioinformatics: the Foundational Model of Anatomy. *J Biomed Inform* 36:478-500.
- Rosen R (1958) The Representation of Biological Systems from the Standpoint of the Theory of Categories, *Bulletin of Mathematical Biophysics* 20:245-260.
- Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 32(DB issue):431-433.
- Schüßler A, Schwarzott D, Walker C (2001) A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol Res* 105(12):1413-1421.
- Shaban-Nejad A, Haarslev V (2007 (a)) Managing Conceptual Revisions in a Temporal Fungi Taxonomy, in proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2007), Maribor, Slovenia, 624-632.
- Shaban-Nejad A, Haarslev V (2007 (b)) Categorical Representation of Evolving Structure of an Ontology for Clinical Fungus. In: Bellazzi, R., Abu-Hanna, A., Hunter, J. (eds.) AIME 2007. LNCS, vol. 4594, Springer, Heidelberg, 277-286.
- Shaban-Nejad A, Haarslev V (2008) Incremental Biomedical Ontology Change Management through Learning Agents, in Proceedings of the 2nd International Symposium on Agent and Multi-Agent Systems (KES-AMSTA'08), Incheon, Korea, LNCS:526-535.
- Sklyar N (2001) Survey of existing Bio-ontologies, Technical Report 5/2001, Department of Computer Science, University of Leipzig.
- Smith B (2003) Realism, Concepts and Categories or: how realism can be pragmatically useful for information. Talk in OntoQuery at Copenhagen Business School, May 18-22, 2003.
- Smith B (2006) From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. *J Biomed Inform* 39(3):288-298.
- Smith B, Rosse C (2004) The role of foundational relations in the alignment of biomedical ontologies. In Proc. of 11<sup>th</sup> World Congress on Medical Informatics; MEDINFO2004, IOS Press, 444-448.
- Smith B, Ceusters W (2006) HL7 RIM: An Incoherent Standard. *Studies in Health Technology and Informatics* 124:133-138.
- Smith B, Williams J, Schulze-Kremer S (2003) The ontology of the gene ontology. in Proceedings of AMIA2003 Annual Symposium; 609-13.
- Stojanovic L (2004) Methods and Tools for Ontology Evolution, Ph.D. thesis University of Karlsruhe.
- Stojanovic L, Maedche A, Motik B, Stojanovic N (2002) User-Driven Ontology Evolution Management. In Proceedings of the 13<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management (EKAW02), Sigüenza, Spain, LNCS: 285-300.
- Stojanovic L, Maedche A, Stojanovic N, Studer R (2003) Ontology evolution as reconfiguration-design problem solving. In Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP03), Sanibel Island, FL, USA. ACM:162-171.
- Stojanovic L, Stojanovic N, Gonzalez J, Studer R (2003) OntoManager - A System for the Usage-Based Ontology Management. In Proceedings of CoopIS/DOA/ODBASE 2003, Catania, Sicily, Italy, LNCS: 858-875.
- Stojanovic L, Motik B (2002) Ontology Evolution within Ontology Editors. In Proceedings of the International Workshop on Evaluation of Ontology-based Tools (EON'02), CEUR-WS Vol-62.
- Sure Y, Angele J, Staab S (2003) OntoEdit: Multifaceted Inferencing for Ontology Engineering. *Journal on Data Semantics (LNCS)* (1):128-152.
- Swann EC, Frieders EM, McLaughlin DJ (1999) Microbotryum, Kriegeria, and the changing paradigm in basidiomycete classification. *Mycologia* 91:51-66.

## Bio-Medical Ontologies Maintenance and Change Management

- TopBraid Composer, Getting Started Guide Version 2.0. TopQuadrant, Inc. (2007) July 27<sup>th</sup>.  
<http://www.topbraidcomposer.com/docs/TBC-Getting-Started-Guide.pdf>. Accessed 10 Jan 2009.
- UMLS documentation (2008) last reviewed on 17 Nov 2008.  
[http://www.nlm.nih.gov/research/umls/umlsdoc\\_intro.html#s1\\_0](http://www.nlm.nih.gov/research/umls/umlsdoc_intro.html#s1_0). Accessed 10 Jan 2009.
- Van Oosten J (2002) Basic Category Theory. Lecture Notes (83 pp). BRICS Lecture Series LS-95-01, last update 2002. Available at (<http://www.math.uu.nl/people/jvoosten/syllabi/catsmoeder.ps.gz>)
- Varzi AC (2005) Change, Temporal Parts and the Argument from Vagueness. *Dialectica* 59(4):485–498.
- Völkel M, Groza T (2006) SemVersion: An RDF-based Ontology Versioning System. In Proceedings of IADIS International Conference on WWW/Internet, vol(1): 195-202.
- Völkel M, Enguix CF, Kruk SR, Zhdanova AV, Stevens R, Sure Y (2005) SemVersion–Versioning RDF & Ontologies. EU-IST Network of Excellence (NoE) KWEB Deliverable D2.3.3.v1(WP2.3)
- Wand YA (1989) A Proposal for a Formal Model of Objects in Object-Oriented Concepts, Databases, and Applications. In Kim W and Lochovsky F, (eds.), ACM Press Frontier Serie:537-559.
- Wasserman R (2006) The Problem of Change. *Philosophy Compass* 1:48-57.
- Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL et al (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28(1):10-4.
- Whitmire SA (1997) Object Oriented Design Measurement, John Wiley & Sons.
- Whitmore I (1999) Terminologia Anatomica: new terminology for the new anatomist. *Anat Rec (New Anat.)* 257:50-53.
- Zalta EN (ed.) (2005) Sorites Paradox. Stanford encyclopedia of philosophy. First published on Jan 17, 1997; substantive revision on Aug 15, 2005.

## Abbreviations

- CDA: Clinical Document Architecture  
CUI: Concept Unique Identifier  
CVS: Concurrent Versions System  
DAG: Directed Acyclic Graphs  
DL: Description Logic  
FMA: The Foundational Model of Anatomy  
FWOnt: FungalWeb Ontology  
GO: Gene Ontology  
HL7: Health Level 7  
KR: Knowledge Representation  
MeSH: Medical Subject Headings  
NLM: The National Library of Medicine  
OBO: Open Biological Ontologies  
OWL: The Web Ontology Language  
RACER: Renamed Abox and Concept Expression Reasoner  
RLR: Representation, Legitimation and Reproduction  
SPARQL: SPARQL Protocol and RDF Query Language  
SWRL: The Semantic Web Rule Language  
TA: Terminologia Anatomica  
UMLS: The Unified Medical Language System