A Report to the Board of Scientific Counselors
April 2010

Quality Assurance in Biomedical Terminologies and Ontologies

Olivier Bodenreider, M.D., Ph.D.

Table of Contents

---

1 Background

Biomedical terminologies and ontologies are enabling resources for clinical decision support systems and data integration systems for translational research [1-2]. Therefore, the quality of these resources has a direct impact on healthcare and biomedical research. In the past few years, quality assurance (QA) of biomedical terminologies and ontologies has become a key issue in the development of standard terminologies, such as SNOMED CT, and has emerged as an active field of research. Approaches to quality assurance include the use of lexical, structural, semantic and statistical techniques applied to particular biomedical terminologies and ontologies, as well as techniques for comparing and contrasting biomedical terminologies and ontologies [3].

In this report, we review 36 studies performed in our research group over the past twelve years having some quality assurance component [4-39]. About half of these studies have a primary focus on quality assurance in terminologies. In the other half, quality assurance is generally an application of the method. As it is not possible or desirable to report each study in detail, we first present an overview of the 36 studies, using the analytical framework presented in [3]. Then, we selected four studies representative of the range of methods developed and present them in more detail.

For the purpose of this report, we use "biomedical terminologies and ontologies" as a generic term for the various kinds of artifacts available for representing the names, meaning and usage of biomedical entities. Ontologies typically define types of entities and their relations (e.g., the Foundational Model of Anatomy (FMA)); terminologies tend to focus on naming (e.g., the list of official gene names and symbols established by the HUGO Gene Nomenclature Committee); thesauri organize entities for a given purpose (e.g., the Medical Subject Headings – MeSH – created for indexing the biomedical literature); classifications allow users to place entities in non-overlapping classes (e.g., the International Classification of Diseases); and knowledge bases incorporate assertional knowledge (e.g., "quinine treats malaria") in addition to the definitional knowledge found in ontologies (e.g., "pneumonia has location lung"). In many cases, however, the distinction among these categories of artifacts is not so sharp. For example, some ontologies also collect names for the entities they represent (e.g., the FMA collects synonyms and names in languages other than English). Conversely, most terminologies are not mere collections of terms, but are organized into hierarchies denoting relations among entities. Finally, the very name of some of these artifacts is misleading. For example, despite its name, the Gene Ontology is mostly a controlled vocabulary for the annotation of gene products. For these reasons, we do not attempt to make a difference between terminologies and ontologies when we refer to the artifacts we analyzed.

2 Project Objectives

The overall objective of the *Medical Ontology Research* project is to develop methods whereby ontologies can be acquired from existing resources and validated against other knowledge

sources, including the Unified Medical Language System (UMLS). Investigating the quality of biomedical terminologies has been an early goal of this project and has remained an important driving force. Over time, we have investigated a variety of terminologies, turning our attention to new ontologies as they became available (e.g., the Gene Ontology), or at the request of their developers (e.g., our investigation of the NCI Thesaurus in 2006). We also investigated quality assurance issues as part of other projects. For example, aligning ontologies to the Foundational Model of Anatomy prompted us to examine its compliance with ontological principles. More recently, we investigated the possible synergies between SNOMED CT and other terminologies including MedDRA and LOINC. These studies also raised such issues as coverage and differences in representation, which are elements of quality. Although not the only goal of the *Medical Ontology Research* project, the theme of quality assurance has been central to our activities.

3 Project Significance

This project is significant from several points of view. By reporting errors to the developers of biomedical terminologies and terminologies, we contributed to increasing their quality. Despite the best efforts of the human editors and the use of formalisms such as description logics, content errors remain frequent in biomedical ontologies and terminologies, which justified the development of multiple approaches to identifying these problems. By sharing our methods with the community, we made it possible for the developers of terminologies and ontologies to implement into their systems some of the techniques we have developed and to integrate them as part of the life cycle of their products. Finally, the experience we acquired while working on quality issues benefits organizations such was the International Health Terminology Standard Development Organization (IHTSDO), of which we have been a member of the Quality Assurance Committee since its inception in 2007.

4 Methods and Procedures

In a recent review of auditing methods for biomedical terminologies and ontologies, Zhu and colleagues [3] provide an elegant analytical framework for analyzing such methods, which we borrowed and adapted liberally. This framework includes the following elements: what is audited; which knowledge sources are used for auditing; and which methods are used for auditing. We first present an overview of our research on quality assurance from the perspective of Zhu's analytical framework illustrated with specific examples. A complete analysis of the 36 research studies is summarized in Table 1 (at the end of this manuscript). Then we present four quality assurance studies in more detail.

4.1 Overview

4.1.1 What is audited

The first element of this framework is the type of entity being audited: Terms/concepts, relations and categorization. Coverage studies evaluate the presence of concepts (e.g., for genomics [31]) and sometimes of particular terms within concepts (e.g., lay synonyms for technical biomedical

---

terms [36]). Relations can be inspected for completeness (e.g., [31]) and for consistency (e.g., [30]). In this review, we do not distinguish between hierarchical and associative relations. In the UMLS, where concepts are categorized with semantic types from the Semantic Network, the categorization of the concepts can be audited as well (e.g., [31]). As shown in Figure 1, our main focus has been on terms/concepts and relations, including the interplay between these two kinds of entities.
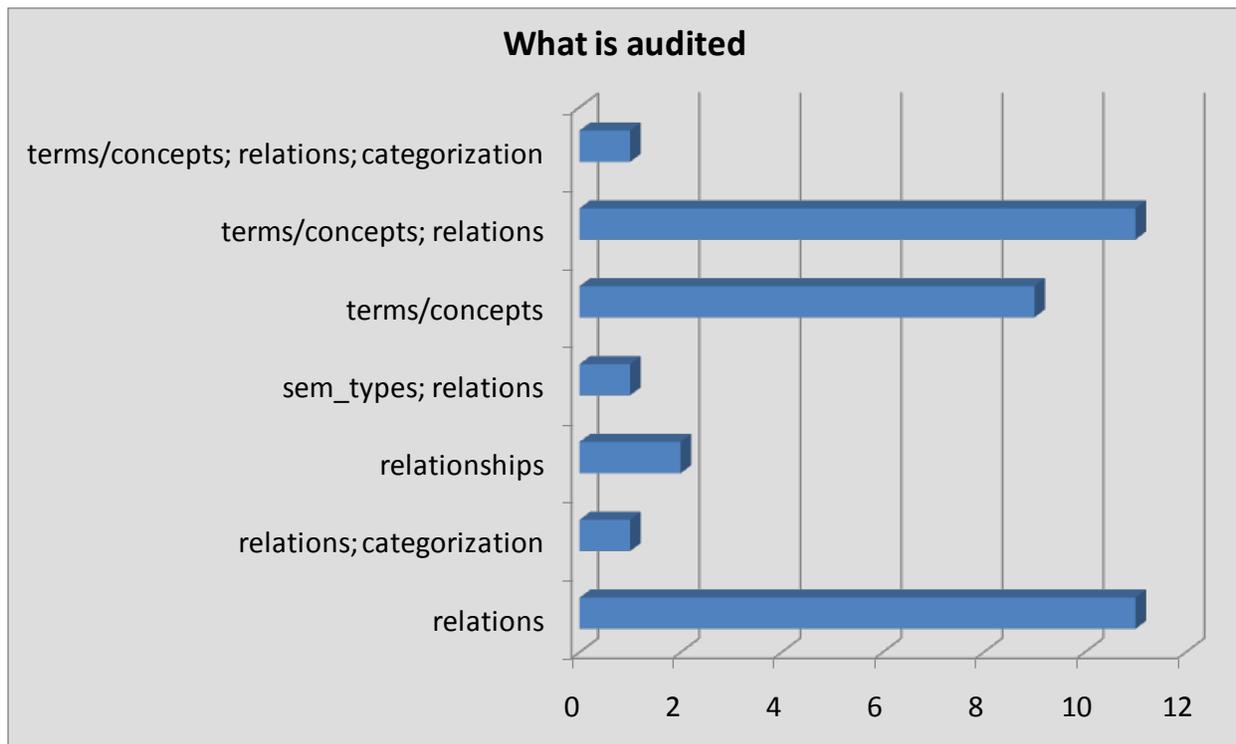
Figure 1. Type of entity being audited in the 36 studies reviewed

4.1.2 *Knowledge sources used for auditing*

The second element characterizes the sources of knowledge used in the auditing. In addition to terms/concepts, relations and categorization, i.e., elements intrinsic to the terminology itself, extrinsic knowledge sources can be used, including corpora and mappings. Terms and concepts can be used for auditing relations (e.g., because adjectival modifiers generally create more specific terms, the concept acute pharyngitis is more specific than pharyngitis [35]). Relations and categorization can be used jointly in order to assess their consistency (e.g., [33]). Extrinsic resources include text corpora (e.g., assessing UMLS strings against the Medline corpus [38]) and corpora of annotations (e.g., functional annotations of gene products used for identifying relations among concepts from the Gene Ontology [24]). Existing mappings between two terminologies enable the comparison of their terms and relations (e.g., comparing SNOMED CT and MedDRA through their mapping to the UMLS Metathesaurus [6]). The distribution of the type of knowledge sources used is shown in Figure 2. Several knowledge sources are often used in combination, the most frequent combination being the use of terms/concepts and relations.
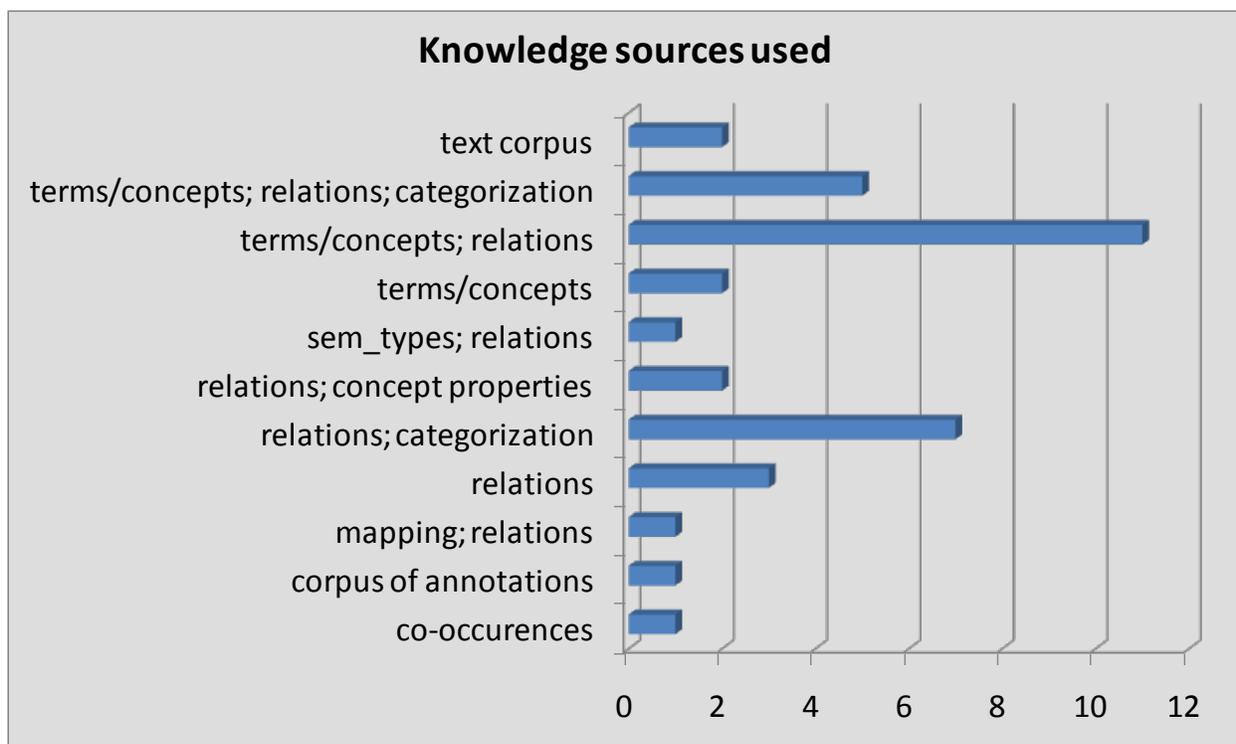
Figure 2. Type of knowledge sources used in the 36 studies reviewed

4.1.3 Methods used for auditing

In terms of methods, approaches to auditing terminologies include lexical, structural, semantic and statistical techniques. These methods can be used for assessing coverage and consistency. Lexical approaches are based on the properties of terms, such as compositionality (e.g., consistency of the relations between compositional terms [30]). Structural approaches are based on the properties of the structures used for the representation of terminologies, including trees or directed acyclic graphs (e.g., testing the UMLS Metathesaurus for the presence of hierarchical cycles [34]). Semantic methods generally rely on the categorization of UMLS Metathesaurus concepts with semantic types (e.g., analysis of polysemous concepts in the Metathesaurus [9]). Statistical approaches are used more rarely but can help detect relations among concepts which co-occur in a corpus (e.g., co-occurring Gene Ontology terms in a corpus of functional annotations of gene products [24]). Other approaches include transforming the formalism of representation of an ontology (e.g., converting the Foundational Model of Anatomy from its native frame-based representation to description logics [21]), and evaluating the compliance of an ontology with ontological principles (e.g., the relationships isa and part of are mutually exclusive [23]). Finally terminologies and ontologies can also be evaluated through their use in an application (e.g., [39]) and by comparison to other ontologies to which there exists a mapping (e.g., [6, 36]). As shown in Figure 3, the approach we utilized the most frequently is a combination of structural and semantic methods, but we explored a variety of methods, isolated and in combination.
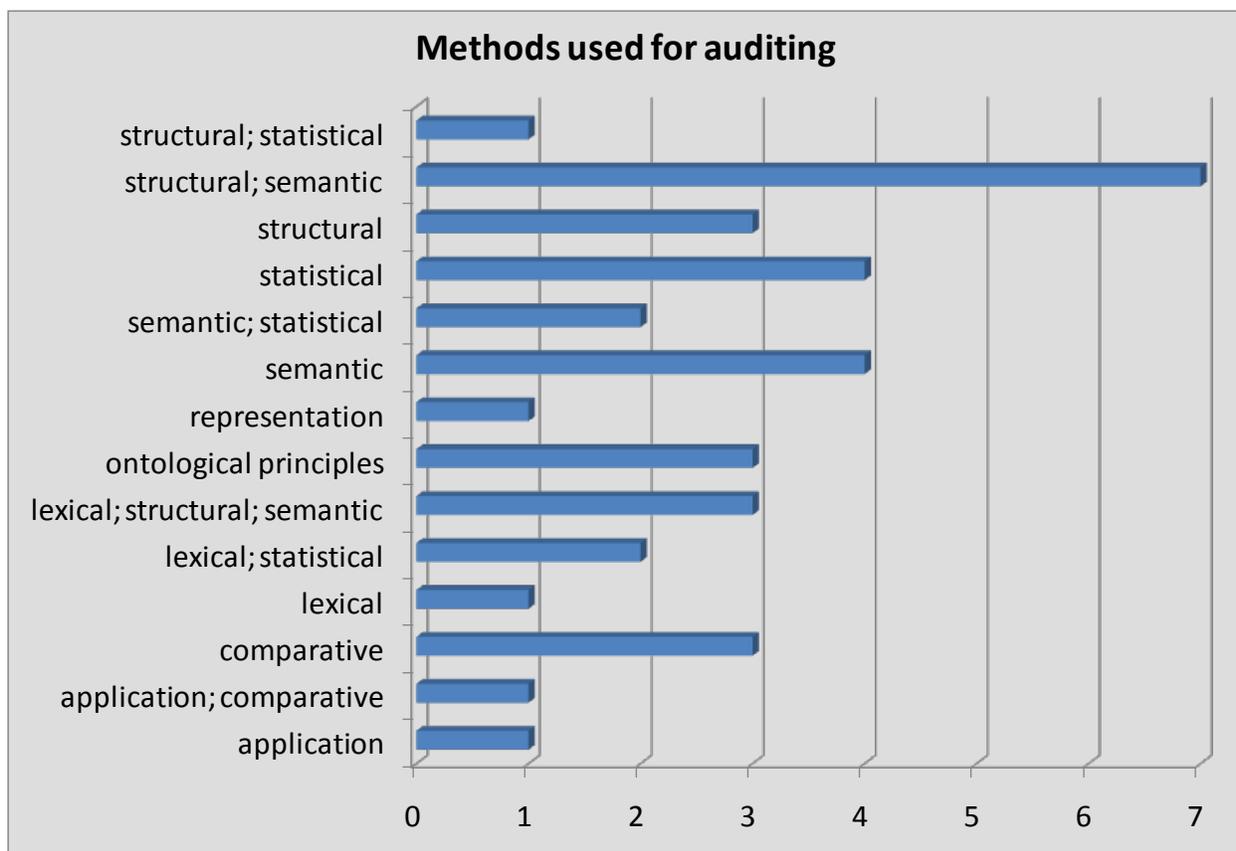
Figure 3. Type of methods used in the 36 studies reviewed

4.1.4 Terminologies and ontologies audited

We investigated twelve biomedical terminologies and ontologies, including anatomical ontologies (Foundational Model of Anatomy (FMA) and the anatomical portion of GALEN), drug terminologies (RxNorm), adverse event terminologies (MedDRA), clinical terminologies (International Classification of Diseases (ICD-10), LOINC, the NCI Thesaurus, SNOMED CT), and specialized terminologies, such as the International Classification of Diseases for Oncology (ICD-O). Additionally, we investigated UMLS Metathesaurus, in which most of these terminologies are integrated, as well as its companion upper-level ontology, the UMLS Semantic Network. Finally, we also explored a generic thesaurus of English, WordNet, while investigating consumer health terminology. Figure 4 recapitulates the various terminologies and ontologies investigated in the 36 studies reviewed.
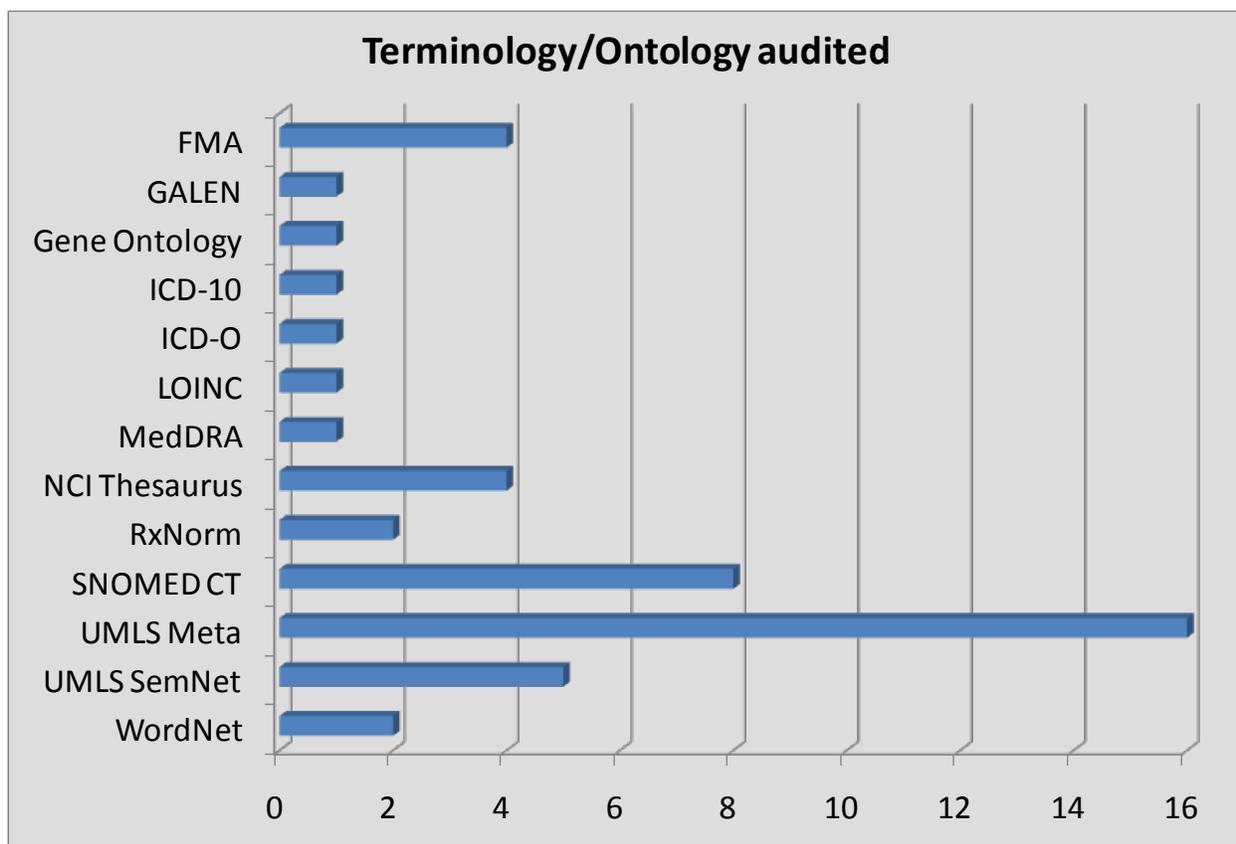
Figure 4. Terminologies and ontologies investigated in the 36 studies reviewed

4.2 Four examples

In this section, we present four particular quality assurance studies to illustrate a variety of methods and terminologies to which they were applied.

4.2.1 *Identifying polysemous concepts in the UMLS [4]*

The objective of this study was to quantify semantic inconsistency in UMLS concepts from the perspective of their hierarchical relations and to show how semantically-inconsistent concepts can help reveal erroneous synonymy relations. Inconsistency is defined in reference to concepts from the UMLS Metathesaurus. Each concept is categorized with semantic types from the Semantic Network and semantic types are partitioned into a 15 groups, called semantic groups. A concept and its parent concept(s) are expected to be in the same semantic group.

Consistency was evaluated by comparing the semantic groups of the two concepts in each pair of hierarchically-related concepts. Answer Set Programming (ASP), a type of logic programming, was used to implement the testing of all hierarchical pairs from the 2M concepts in the UMLS. A limited number of inconsistent concepts was inspected manually.

We identified 81,512 concepts as inconsistent due to the differences in semantic groups between the concept and its parent. A vast majority of inconsistent hierarchical relations were not indicative of any real errors, but rather denoted "navigational relations", i.e., pseudo-hierarchical rela-

tions created for the purpose of navigating within a terminology. Examples of such navigational links include relations between the sign hepatomegaly (child) and the anatomical structure liver (parent), from OMIM, and between the test quantitative dosage of the concentration of ketoconazole in plasma or serum (child) and the drug ketoconazole (parent), in LOINC.

Besides navigational relations, we identified a limited number of real errors in the UMLS Metathesaurus. More specifically, we discovered an interesting semantic pattern along hierarchies, which seems associated with wrong synonymy. Figure 5 depicts an example of wrong synonymy identified through this approach. The concept capsule (pharmacologic) (C1181304) belongs to the semantic group Chemicals & Drugs. Its parents include anatomical concepts such as Membranous layer (C2338391), as well as drug concepts (e.g., Pill (C0994475)). Analogously, mixed semantics is found among its children, with anatomical concepts such as Capsule of adrenal gland (C1181304) and drug concepts including Oral Capsule (C0991533). In order to address the wrong synonymy in capsule (pharmacologic), a distinct concept should be created for the anatomical capsule, with a semantic type from the semantic group Anatomy. Three additional errors of the same type were identified and reported to the UMLS team.

In conclusion, while most of the inconsistencies identified did not correspond to real errors, we were able to uncover four instances of wrong synonymy through a specific pattern of semantic rupture along hierarchical relations.
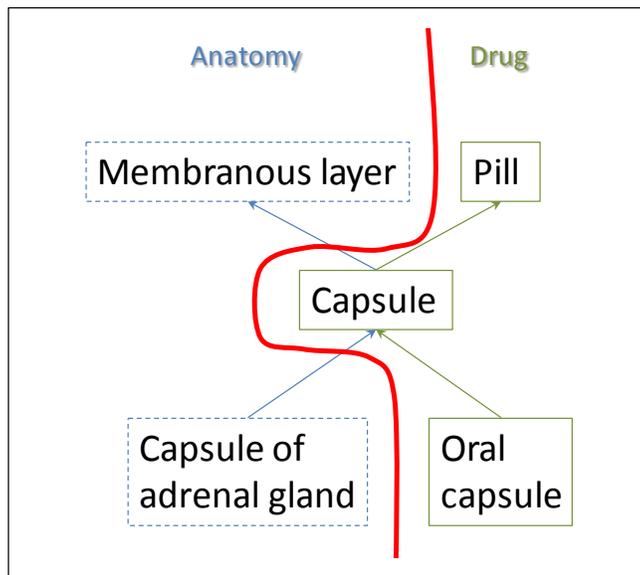
Figure 5. Example of wrong synonymy in the UMLS Metathesaurus. The two senses of capsule – from anatomy and drug – are not represented as distinct concepts.

4.2.2 Identifying errors in RxNorm [7]

RxNorm is a standardized nomenclature for clinical drug entities developed by the National Library of Medicine. As shown in Figure 6, the various kinds of drug entities (ingredient, brand name, clinical drug, branded drug, etc.) form a graph and are linked by specific relationships

(e.g., tradename_of, between brand name and ingredient). In this study, we audited relations in RxNorm for consistency and completeness through the systematic analysis of the graph of its concepts and relationships. One of the premises for this work is that, in the RxNorm graph, there exist alternate paths between a start node and an end node, and that alternate paths are expected to be equivalent.
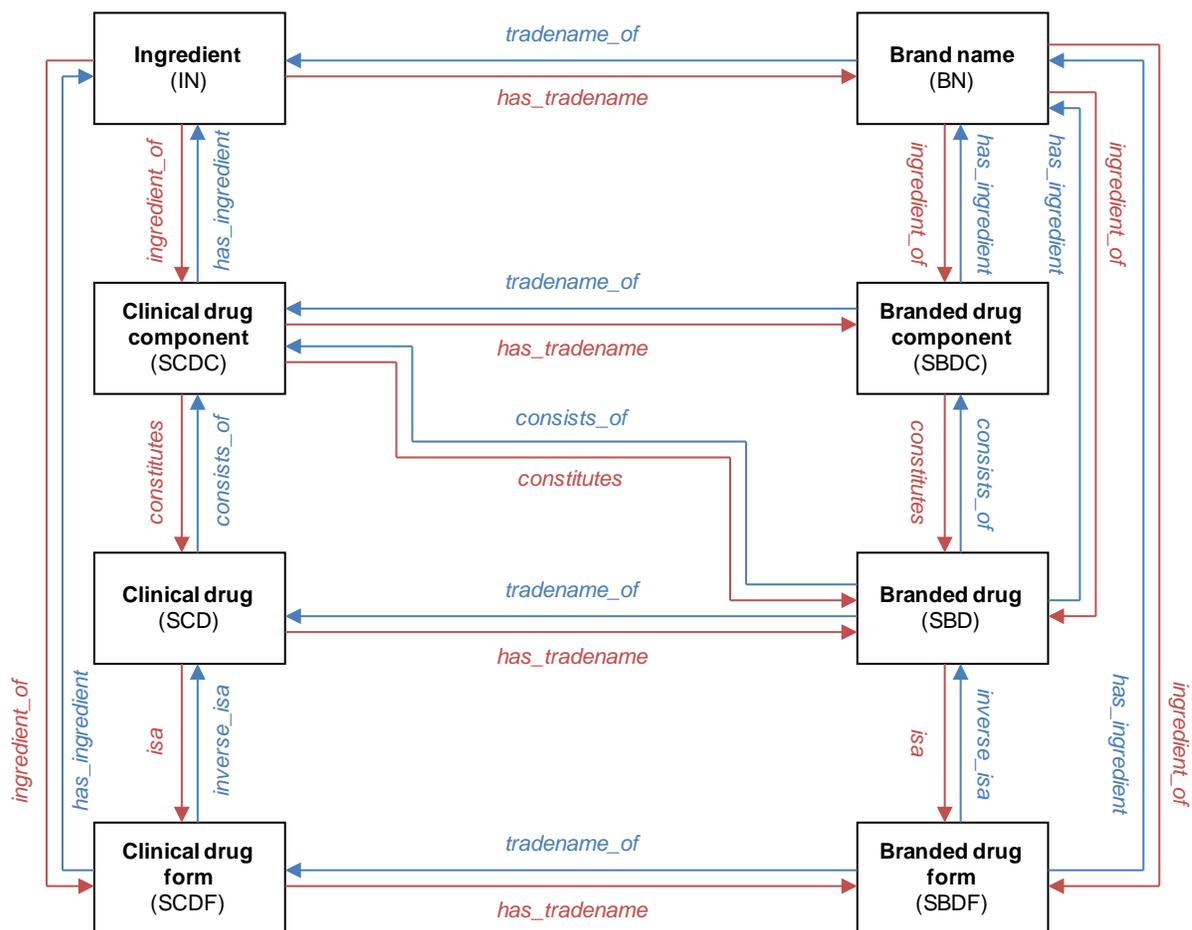
Figure 6. Graph of the eight major categories in RxNorm and their interrelations

We started by normalizing representation of multi-ingredient drugs in order to make it compatible with that of single-ingredient drugs. That is, we created new "multi-ingredients" composed of several ingredients, so that the corresponding brand name points to one multi-ingredient rather than several ingredients. All meaningful paths between two nodes in the type graph were computed and instantiated. Alternate paths were automatically compared and manually inspected in case of inconsistency.

The 115 meaningful paths identified in the type graph can be grouped into 28 groups with respect to start and end nodes. Of the 19 groups of alternate paths (i.e., with two or more paths) between the start and end nodes, 9 (47%) exhibited inconsistencies. Overall, 28 (24%) of the 115 paths were inconsistent with other alternate paths. A total of 348 inconsistencies were identified

in the April 2008 version of RxNorm and reported to the RxNorm team, of which 215 (62%) had been corrected in the January 2009 version of RxNorm.

Here is one particular type of error identified though our analysis. We found cases where the direct path between brand name and ingredient was inconsistent with alternate paths, such as the indirect path through branded drug and clinical drug. In 17 cases, a brand name entity has no relation to any ingredient entities. Examples include *Sochlor*, not linked directly to its ingredient, *Sodium chloride*. All 17 cases had been corrected in the January 2009 version of RxNorm.

In conclusion, the inconsistencies identified involved missing nodes (93), missing links (17), extraneous links (237) and one case of mix-up between two ingredients. Our auditing method proved effective in identifying a limited number of errors that had defeated the quality assurance mechanisms currently in place in the RxNorm production system. We provided some recommendations for the development of RxNorm.

4.2.3   Non-lexical approaches to identifying relations in the Gene Ontology [24]

The Gene Ontology (GO) is a controlled vocabulary widely used for the annotation of gene products. GO is organized in three hierarchies for molecular functions, cellular components, and biological processes but no relations are provided among terms across hierarchies. The objective of this study was to investigate three non-lexical approaches to identifying such associative relations in GO and compare them among themselves and to lexical approaches. The motivation for this work was that GO was at the time limited to isa and part_of relationships within a given hierarchy and did not explicitly represent any relations across hierarchies.

The three approaches were: computing similarity in a vector space model, statistical analysis of co-occurrence of GO terms in annotation databases, and association rule mining. Five annotation databases (FlyBase, the Human subset of GOA, MGI, SGD, and WormBase) were used in this study. These methods were evaluated against several baselines including one lexical approach based on terms compositionality (i.e., when one term is included in another term), the existence of relations in the UMLS, and the existence of co-occurrence relations in Medline.

A total of 7,665 associations were identified by at least one of the three non-lexical approaches. Of these, 12% were identified by more than one approach. While there are almost 6,000 lexical relations among GO terms, only 203 associations were identified by both non-lexical and lexical approaches. Specific examples of relations identified through these approaches are listed in Table 2.

In conclusion, the three non-lexical approaches to identifying relations in the Gene Ontology were productive and complementary to the lexical approach. The associations identified in this study could serve as the starting point for adding associative relations across hierarchies to GO, but would require manual curation. We suggested that this approach, primarily designed for the acquisition of new relations, could also find application to quality assurance of annotation databases by suggesting related terms to curators.

Table 2 – Examples of association identified simultaneously by several methods (VSM: vector space model; COC: co-occurrences; ARM: association rule mining; LEX: lexical approach; REL: existing relations in the UMLS; MDL: co-occurrence in Medline)

| Association			VSM	COC	ARM	LEX	REL	MDL
MF:	*potassium channel activity*	[GO:0005267]	X	X	X			
BP:	*potassium ion transport*	[GO:0006813]						
MF:	*chemokine activity*	[GO:0008009]		X	X			
BP:	*immune response*	[GO:0006955]						
CC:	*hemoglobin complex*	[GO:0005833]	X	X				
BP:	*oxygen transport*	[GO:0015671]						
MF:	*taste receptor activity*	[GO:0008527]	X		X			
BP:	*perception of taste*	[GO:0050909]						
MF:	*metal ion transporter activity*	[GO:0046873]	X		X	X		
BP:	*metal ion transport*	[GO:0030001]						
CC:	*transport vesicle*	[GO:0030133]				X	X	
BP:	*transport*	[GO:0006810]						
CC:	*gap junction*	[GO:0005921]	X	X				X
BP:	*cell communication*	[GO:0007154]						

4.2.4 *Lexical approaches to assessing the consistency of relations in SNOMED [30]*

We investigated the use of adjectival modification as a way of assessing the systematic use of linguistic phenomena to represent similar lexical or semantic features in the constituent terms of a vocabulary. For a pair of adjectival modifiers often associated with the same context (e.g., primary adrenal insufficiency, secondary adrenal insufficiency), we wanted to assess whether the two concepts are siblings and children of the same context parent (adrenal insufficiency).

Terms consisting of one or more adjectival modifiers followed by a head noun were selected from disease and procedure terms in SNOMED. Frequently co-occurring adjectival modifiers were systematically combined with the contexts (i.e., terms minus modifier) of each modifier. The existence of these combinations was checked in both SNOMED and the entire UMLS Metathesaurus; the term corresponding to the context alone was similarly checked. Relationships among terms sharing a context and between each of these terms and their context were studied.

Four pairs of modifiers were studied: (acute, chronic), (unilateral, bilateral), (primary, secondary), and (acquired, congenital). The numbers of contexts studied for each pair ranged from 73 to 974. The percentage of contexts associated with both modifiers ranged from 5% to 50% in SNOMED and from 10% to 60% in UMLS. The presence of the context term varied from 31% to 64% in SNOMED and from 43% to 79% in UMLS. Finally, 172 occurrences (9%) of synonymy between a modified term and the context term were found in SNOMED. 145 such occurrences (8%) were found in the entire Metathesaurus.

The pair (acquired, congenital) will be used to illustrate the results. A global evaluation of the consistency of the representation of terms modified by acquired or congenital in SNOMED is shown in Figure 7. 974 contexts are associated with either modifier of the pair. Both modified terms are present in SNOMED in 52 cases, and in the UMLS in 97 cases (e.g., acquired spondylolisthesis, congenital spondylolisthesis). Terms modified by congenital only (e.g., congenital bronchiectasis) are more frequent (822 in SNOMED) than those modified by acquired only (e.g., acquired epidermolysis bullosa, 100 in SNOMED). Their contexts (e.g., epidermolysis bullosa) are present in SNOMED in 306 cases and in the UMLS in 418 cases. The terms modified by acquired and congenital are not frequently represented as siblings in SNOMED (10 cases). For example, acquired keratoderma (D0-22310) and congenital keratoderma (D4-40130) are represented in two separate branches of the disease hierarchy in SNOMED. Moreover, the relationships between modified terms and their context also contribute to the characterization of a pair of modifiers. Most terms modified by acquired and congenital do not have any paradigmatic relationship represented with their context. For example, although keratoderma exists as a concept in the Metathesaurus, there is no relationship between acquired keratoderma or congenital keratoderma and keratoderma. In 44 cases, the relationship is hierarchical (e.g., between congenital porphyria and porphyria). In 18 cases, the modified term and its context are siblings (e.g., congenital Addison's disease and Addison's disease). Finally, in 99 cases, they are considered synonyms in SNOMED (e.g., acquired polycythemia and polycythemia).
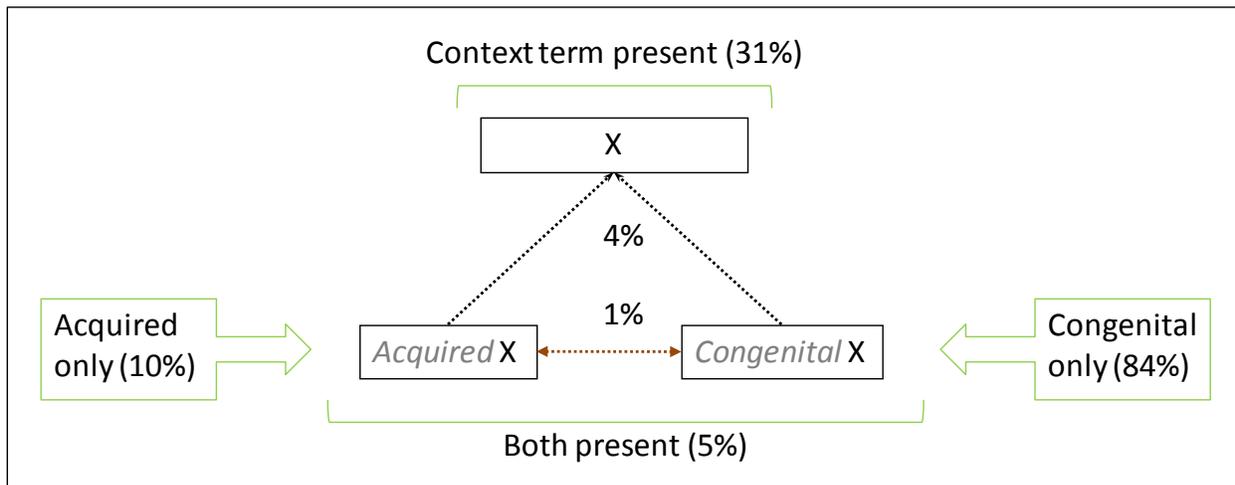
Figure 7. Evaluation of the consistency of the representation of terms modified by acquired or congenital in SNOMED

In conclusion, we showed that there was limited consistency in the representation of disease terms in SNOMED and, to a smaller extent, in the UMLS Metathesaurus. The use of lexical knowledge, namely adjectival modification, was effective for assessing the systematic use of linguistic phenomena to represent similar lexical or semantic features in the constituent terms of a vocabulary.

5 Project Status

This project is an ongoing project, under the umbrella of the *Medical Ontology Research* project. As shown in Figure 8, our commitment to assessing the quality of biomedical terminologies and ontologies has been sustained over the past twelve years. The 36 studies listed in this report represent 28% of our publications. Other research efforts in our project have included ontology alignment (with focus on anatomical ontologies), development of terminology services and visualization tools (e.g., RxNav and application programming interfaces for RxNorm), the use of Semantic Web technologies for information integration in biomedicine, and applications of semantic similarity among biomedical concepts (e.g., to information retrieval).
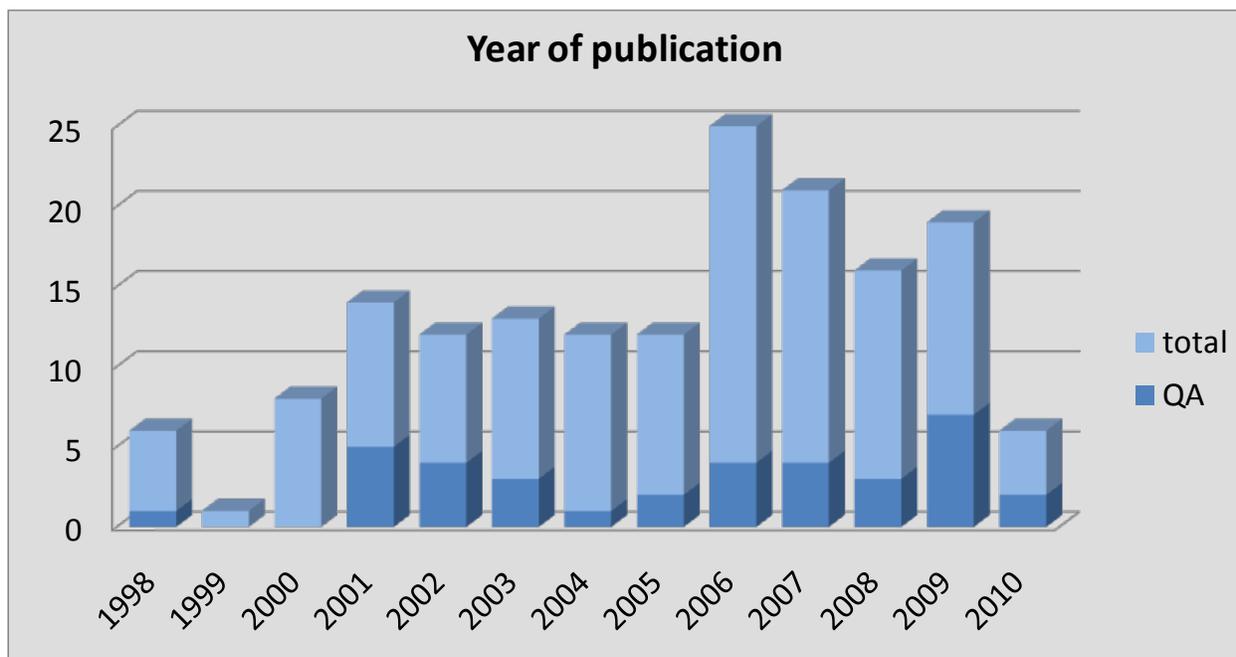
Figure 8. Proportion of publications from our group devoted to quality assurance

6 Evaluation Plan

Each quality assurance research study is different and includes elements of evaluation. One typical evaluation schema is manual review of the results by subject matter experts. Ideally, the evaluation is performed in collaboration with the developers of the terminology or ontology, who analyze the potential errors identified and correct them in the next version of the product. In our experience, this has happened, for example, with the Foundational Model of Anatomy ([21]) and with RxNorm ([7]). In both cases, our methods had detected a limited number of errors that had defeated the quality assurance mechanisms in place in terminology development systems. In other cases, the quality assurance process is mostly automated. This is the case, for example, of the automatic process we developed for creating a directed acyclic graph from hierarchical relations in the UMLS Metathesaurus ([34]), which we use routinely to find the closest MeSH term from an arbitrary UMLS concept by traversing this graph.

7   Project Schedule and Resources

Some of these research studies have involved exclusively Lister Hill Center personnel, including project staff members and other Lister Hill researchers. (The *Medical Ontology Research* project has benefited from the programming support of one or two staff members over time.) This is the case of our work on RxNorm [7, 15]), some comparative work between terminologies ([6, 12-13]), and earlier work with the UMLS ([27, 29, 33-34]). Other projects have been carried out by summer students (e.g., [5]) and post-doctoral students (e.g., [5, 19, 22]). Finally, a significant part of our work is done in collaboration with research groups outside NIH, including former visiting scientists who spent some time in our group (e.g., [17, 21]). In addition to Lister Hill colleagues, our collaborators have included academic groups from several sites in the US (Harvard, University of Utah, University at Buffalo and New Jersey Institute of Technology), as well as in countries including France, Germany, the Netherlands, Turkey and China.

8   Summary and Future Plans

As part of the *Medical Ontology Research* project, we have explored quality assurance issues in biomedical terminologies and ontologies including SNOMED CT, the NCI Thesaurus, the Foundational Model of Anatomy (FMA), and RxNorm. We have also exploited the organizational structure of the Unified Medical Language System (UMLS) for quality assurance purposes. In our work, we put special emphasis on the development of principled, automated, scalable methods, applied systematically to the entire content of a terminology by independent researchers, as opposed to manual review of subsets by domain experts.

The outcome of our work is twofold. From an academic perspective, we have developed effective quality assurance processes, which we have shared with the community through scientific publications and presentations at conferences. The practical impact of our work is our contribution to the improvement of the quality of the terminologies and ontologies we investigated. While only a limited number of errors have been identified – which is a testament to the high quality of these artifacts – these errors had defeated the quality assurance mechanisms in place in terminology development systems. A number of such inconsistencies and errors have been reported to their developers of the biomedical terminologies and ontologies we investigated, resulting in incremental improvement of these terminological resources.

In the future, we want to develop the use of Semantic Web technologies (RDF / SPARQL and OWL) to support quality assurance in biomedical terminologies and ontologies. Preliminary work has shown that the use of these technologies can help reduce the amount of *ad hoc* programming necessary for investigating the quality of ontologies. We have also started to investigate terminologies of clinical interest, such as NDF-RT, in order to evaluate their capacity to support clinical decision. We are especially interested in evaluating quality issues through applications (e.g., usefulness of the therapeutic classes present in NDF-RT for supporting the identification of drug-drug interactions). Finally, we remain committed to improving the quality of SNOMED CT through our participation in the activities of the IHTSDO.

9 References

1. **Bodenreider, O**., *Biomedical ontologies in action: role in knowledge management, data integration and decision support.* Geissbuhler A, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2008. Methods Inf Med, 2008. **47**(Suppl 1): p. 67-79 [PubMed: 18660879].

2. **Bodenreider, O**. and R. Stevens, *Bio-ontologies: Current trends and future directions.* Briefings in Bioinformatics, 2006. **7**(3): p. 256-274 [PubMed: 16899495].

3. Zhu, X., J.W. Fan, D.M. Baorto, C. Weng, and J.J. Cimino, *A review of auditing methods applied to the content of controlled biomedical terminologies.* J Biomed Inform, 2009. **42**(3): p. 413-25 [PubMed: 19285571].

4. Erdogan, H., E. Erdem, and **O. Bodenreider**, *Exploiting UMLS semantics for quality assurance purposes.* Medinfo, 2010: p. (in press).

5. Wei, D. and **O. Bodenreider**, *Using the Abstraction Network in complement to Description Logics for quality assurance in biomedical terminologies - A case study in SNOMED CT.* Medinfo, 2010: p. (in press).

6. **Bodenreider, O**., *Using SNOMED CT in combination with MedDRA for reporting signal detection and adverse drug reactions reporting.* AMIA Annu Symp Proc, 2009: p. 45-49.

7. **Bodenreider, O**. and L.B. Peters, *A graph-based approach to auditing RxNorm.* Journal of Biomedical Informatics, 2009. **42**(3): p. 558-570 [PubMed: 19394440].

8. Cimino, J.J., T.F. Hayamizu, **O. Bodenreider**, B. Davis, G.A. Stafford, and M. Ringwald, *The caBIG terminology review process.* Journal of Biomedical Informatics, 2009. **42**(3): p. 571-580 [PubMed: 19154797].

9. Mougin, F., **O. Bodenreider**, and A. Burgun, *Analyzing polysemous concepts from a clinical perspective: Application to auditing concept categorization in the UMLS.* Journal of Biomedical Informatics, 2009. **42**(3): p. 440-451 [PubMed: 19303057].

10. Schulz, S., E. Beisswanger, L. van den Hoek, **O. Bodenreider**, and E.M. van Mulligen, *Alignment of the UMLS Semantic Network with BioTop: Methodology and assessment.* Bioinformatics, 2009. **25**(12): p. i69-i76 [PubMed: 19478019].

11. Vizenor, L.T., **O. Bodenreider**, and A.T. McCray, *Auditing associative relations across two knowledge sources.* Journal of Biomedical Informatics, 2009. **42**(3): p. 426-439 [PubMed: 19475724].

12. **Bodenreider, O**., *Comparing SNOMED CT and the NCI Thesaurus through Semantic Web technologies.* Proceedings of the Third International Conference on Knowledge Representation in Medicine (KR-MED 2008), 2008: p. 37-43 (electronic proceedings: http://ceur-ws.org/Vol-410/).

13. **Bodenreider, O**., *Issues in mapping LOINC laboratory tests to SNOMED CT.* AMIA Annu Symp Proc, 2008: p. 51-55 [PubMed: 18999311].

14. Mougin, F. and **O. Bodenreider**, *Auditing the NCI thesaurus with Semantic Web technologies.* AMIA Annu Symp Proc, 2008: p. 500-504 [PubMed: 18999265].

15. Peters, L. and **O. Bodenreider**, *Using the RxNorm web services API for quality assurance purposes.* AMIA Annu Symp Proc, 2008: p. 591-595 [PubMed: 18999038].

16. **Bodenreider, O**., B. Smith, A. Kumar, and A. Burgun, *Investigating subsumption in SNOMED CT: An exploration into large Description Logic-based biomedical terminologies.* Artificial Intelligence in Medicine, 2007. **39**(3): p. 183-195 [PubMed: 17241777].

17. Burgun, A. and **O. Bodenreider**, *Issues in integrating epidemiology and research information in oncology: Experience with ICD-O3 and the NCI Thesaurus.* AMIA Annu Symp Proc, 2007: p. 85-89 [PubMed: 18693803].

18. Fung, K.W., **O. Bodenreider**, A.R. Aronson, W.T. Hole, and S. Srinivasan, *Combining lexical and semantic methods of inter-terminology mapping using the UMLS.* Stud Health Technol Inform, 2007. **129**: p. 605-609 [PubMed: 17911788].

19. Zhang, S. and **O. Bodenreider**, *Experience in aligning anatomical ontologies.* International Journal on Semantic Web and Information Systems, 2007. **3**(2): p. 1-26 [PubMed: 18974854].

20. **Bodenreider, O**. and S. Zhang, *Comparing the representation of anatomy in the FMA and SNOMED CT.* AMIA Annu Symp Proc, 2006: p. 46-50 [PubMed: 17238300].

21. Golbreich, C., S. Zhang, and **O. Bodenreider**, *The Foundational Model of Anatomy in OWL: Experience and perspectives.* Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 2006. **4**(3): p. 181-195 [PubMed: 18360535].

22. Vizenor, L., **O. Bodenreider**, L. Peters, and A.T. McCray, *Enhancing biomedical ontologies through alignment of semantic relationships: Exploratory approaches.* AMIA Annu Symp Proc, 2006: p. 804-808 [PubMed: 17238452].

23. Zhang, S. and **O. Bodenreider**, *Law and order: Assessing and enforcing compliance with ontological modeling principles.* Computers in Biology and Medicine, 2006. **36**(6-7): p. 674-693 [PubMed: 16144698].

24. **Bodenreider, O**., M. Aubry, and A. Burgun, *Non-lexical approaches to identifying associative relations in the Gene Ontology*, in *Pacific Symposium on Biocomputing 2005*, R.B. Altman, et al., Editors. 2005, World Scientific. p. 91-102.

25. Mougin, F. and **O. Bodenreider**, *Approaches to eliminating cycles in the UMLS Metathesaurus: Naïve vs. formal.* AMIA Annu Symp Proc, 2005: p. 550-554 [PubMed: 16779100].

26. **Bodenreider, O**. and A. Burgun, *Aligning knowledge sources in the UMLS: Methods, quantitative results, and applications.* Medinfo, 2004: p. 327-331 [PubMed: 15360828].

27. **Bodenreider, O**., *Strength in numbers: Exploring redundancy in hierarchical relations across biomedical terminologies.* AMIA Annu Symp Proc, 2003: p. 101-105 [PubMed: 14728142].

28. **Bodenreider, O**., A. Burgun, and J.A. Mitchell, *Evaluation of WordNet as a source of lay knowledge for molecular biology and genetic diseases: A feasibility study.* Stud Health Technol Inform, 2003. **95**: p. 379-384 [PubMed: 14664016].

29. **Bodenreider, O**. and A.T. McCray, *Exploring semantic groups through visual approaches.* Journal of Biomedical Informatics, 2003. **36**(6): p. 414-432 [PubMed: 14759816].

30. **Bodenreider, O**., A. Burgun, and T.C. Rindflesch, *Assessing the consistency of a biomedical terminology through lexical knowledge.* International Journal of Medical Informatics, 2002. **67**(1-3): p. 85-95 [PubMed: 12460634].

31. **Bodenreider, O**., J.A. Mitchell, and A.T. McCray, *Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics.* Proc AMIA Symp, 2002: p. 61-65 [PubMed: 12463787].

32. **Bodenreider, O**., T.C. Rindflesch, and A. Burgun, *Unsupervised, corpus-based method for extending a biomedical terminology.* Proceedings of the ACL'2002 Workshop "Natural Language Processing in the Biomedical Domain", 2002: p. 53-60.

33. McCray, A.T. and **O. Bodenreider**, *A conceptual framework for the biomedical domain*, in *The semantics of relationships: an interdisciplinary perspective*, R. Green, C.A. Bean, and S.H. Myaeng, Editors. 2002, Kluwer Academic Publishers: Dordrecht; Boston. p. 181-198.

34. **Bodenreider, O**., *Circular hierarchical relationships in the UMLS: Etiology, diagnosis, treatment, complications and prevention.* Proc AMIA Symp, 2001: p. 57-61 [PubMed: 11825155].

35. **Bodenreider, O**., A. Burgun, and T.C. Rindflesch, *Lexically-suggested hyponymic relations among medical terms and their representation in the UMLS.* Proceedings of TIA'2001 "Terminology and Artificial Intelligence", 2001: p. 11-21.

36. Burgun, A. and **O. Bodenreider**, *Comparing terms, concepts and semantic classes in WordNet and the Unified Medical Language System.* Proceedings of the NAACL'2001 Workshop, "WordNet and Other Lexical Resources: Applications, Extensions and Customizations", 2001: p. 77-82.

37. Burgun, A. and **O. Bodenreider**, *Methods for exploring the semantics of the relationships between co-occurring UMLS concepts.* Medinfo, 2001. **10**(Pt 1): p. 171-175 [PubMed: 11604727].

38. McCray, A.T., **O. Bodenreider**, J.D. Malley, and A.C. Browne, *Evaluating UMLS strings for natural language processing.* Proc AMIA Symp, 2001: p. 448-452 [PubMed: 11825228].

39. **Bodenreider, O**., A. Burgun, G. Botti, M. Fieschi, P. Le Beux, and F. Kohler, *Evaluation of the Unified Medical Language System as a medical knowledge source.* J Am Med Inform Assoc, 1998. **5**(1): p. 76-87 [PubMed: 9452987].

Table 1. Overview of characteristics of the Quality Assurance studies reported (based on Zhu's criteria)

Reference		[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[37]	[36]	[38]	[39]
What is audited	Concepts	•	•	•		•	•	•		•	•				•	•	•	•	•		•					•	•	•	•	•					•		•
	Relations		•		•	•		•	•			•	•	•	•	•			•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•		•
	Categorization																														•				•		
Knowledge sources used	Concepts				•		•	•	•	•					•	•	•	•	•	•	•					•	•	•	•				•		•		•
	Relations	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•		•	•	•	•			•			•	•	•		•		•
	Categorization	•	•				•			•				•						•						•	•	•					•		•		•
	Corpus																					•									•					•	
	Mapping			•																																	
Method used for auditing	Lexical														•		•	•											•			•				•	
	Structural	•	•		•		•						•		•	•	•	•					•	•	•	•						•					
	Semantic	•			•		•		•	•	•	•	•		•	•	•	•		•				•			•					•					
	Statistical									•										•		•			•		•	•		•					•	•	
	Other		•			•		•						•					•		•						•			•					•		•
Terminology / Ontology audited	FMA															•	•	•			•																
	GALEN																				•																
	GO																					•															
	ICD-10																•																				
	ICD-O																			•																	
	LOINC										•																										
	MedDRA			•																																	
	NCI Thesaur.				•					•		•			•																						
	RxNorm				•								•																								
	SNOMED CT		•	•						•	•			•		•	•	•																			
	Metathesaurus	•					•		•											•				•	•	•					•	•	•	•	•	•	•
	Semantic Net.					•															•	•		•													
	WordNet																													•					•		