Lister Hill National Center for
Biomedical Communications
An Intramural Research Division of the U.S. National Library of Medicine

Evaluating the Quality and Interoperability of Biomedical Terminologies

April 2018

Olivier Bodenreider, MD, PhD

Cognitive Science Branch

U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894

Table of Contents

Executive Summary

Biomedical terminologies and ontologies are enabling resources for clinical decision support systems and data integration systems for translational research and health analytics. Therefore, the quality of these resources has a direct impact on healthcare and biomedical research. In the past decade, quality assurance (QA) of biomedical terminologies has become a key issue in the development of standard terminologies and has emerged as an active field of research. Approaches to quality assurance include the use of lexical, structural and semantic techniques applied to biomedical terminologies, as well as techniques for comparing and contrasting these resources.

As part of the Medical Ontology Research project, we have explored quality assurance and interoperability issues in a variety of biomedical terminologies including drug terminologies, clinical terminologies, and specialized terminologies, such as HPO – the Human Phenotype Ontology and the Orphanet terminology for rare diseases. In this report, we review 32 investigations performed in our research group since this project was last reviewed by the BSC in 2010. About half of these investigations have a primary focus on quality assurance, for which we developed novel methods. In the other half, we applied existing techniques to assess interoperability among terminologies or some aspect of quality (e.g., coverage) in a terminology. 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 QA processes we developed have proved effective in identifying a limited number of errors that had defeated the quality assurance mechanisms in place in terminology development systems. We have shared our findings and techniques with the scientific community through scientific publications and presentations at conferences. Whenever possible, we have also reported these issues to the developers of the biomedical terminologies we investigated.

This work is also a contribution to the LHC Training Program, since 21 of the 32 studies listed in this report (66%) have involved post-doctoral fellows or summer (graduate and undergraduate) students.

1 Background

Biomedical terminologies and ontologies are enabling resources for clinical decision support systems and data integration systems for translational research and health analytics [1, 2]. Therefore, the quality of these resources has a direct impact on healthcare and biomedical research [3]. In the past decade, quality assurance (QA) of biomedical terminologies and ontologies has become a key issue in the development of standard terminologies and has emerged as an active field of research [4, 5].

In this context, we do not make a distinction between biomedical terminologies (typically collections of terms for biomedical entities, often arranged hierarchically) and ontologies (typically collections of types of biomedical entities with their definitions and interrelations). Many artifacts, such as SNOMED CT, share features from both. We refer to them collectively as terminologies for simplicity.

In addition to "quality assurance", several terms are used to refer to the notion of identifying issues in and improving the quality of biomedical terminologies, including "auditing" and "evaluating" these resources. We will not make a strong distinction among these terms and generically refer to quality assurance methods. Approaches to quality assurance include the use of lexical, structural and semantic techniques applied to biomedical terminologies, as well as techniques for comparing and contrasting these resources [4, 5]. Of note, the techniques used for quality assurance are often also used for ontology alignment [6]. Aspects of quality assurance include completeness (of the concepts, terms, relations and definitions) and consistency, as well as conciseness (non-redundancy) and accuracy [4]. These features can be evaluated within a terminology (intrinsic evaluation) or between terminologies (extrinsic evaluation) [5]. Beyond the quality of individual terminologies, interoperability among them is a prerequisite to effective data integration.

In this report, we review 32 investigations performed in our research group since this project was last reviewed by the BSC in 2010 [7-38]. About half of these studies have a primary focus on quality assurance in terminologies, for which we developed novel methods. In the other half, we applied existing techniques to assess interoperability among terminologies or some aspect of quality (e.g., coverage) in a terminology.

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). In practice, the activities of the *Medical Ontology Research* project are organized into three major components: 1) Quality and interoperability of biomedical terminologies; 2) Integration, dissemination, quality assurance and applications of drug terminologies; and 3) Biomedical Linked Open Data.

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, Human Phenotype Ontology, Disease Ontology) and to terminologies directly supported by NLM (e.g., RxNorm, SNOMED CT). Some aspects of our research have been driven by the development of services (e.g., interoperability among pharmacologic classes was investigated for the development of *RxClass*, and investigation of the quality of value sets to support the development of the NLM *Value Set Authority Center*).

3 Project Significance

Despite the best efforts of human editors and the use of formalisms, such as description logics, content errors remain frequent in biomedical terminologies, which justifies the development of multiple approaches to identifying these problems. Moreover, many other QA efforts merely focus on identifying subsets of a terminology where errors tend to occur more frequently (e.g., [39-41]). While these methods help focus the effort of human reviewers, their precision is generally insufficient to be truly effective. In contrast, the methods we have developed for QA of SNOMED CT not only identify errors with precision, but also suggest remediation.

By sharing our investigations with the community, we make it possible for the developers of terminologies not only to fix the errors, but also to implement into their systems some of the techniques we have developed, contributing to the life cycle of their products.

Our investigation of the coverage of human phenotypes and rare diseases in standard clinical terminologies has provided evidence for integrating the Human Phenotype Ontology into the UMLS Metathesaurus and has demonstrated the limited added value of the Disease Ontology compared to SNOMED CT.

The experience we have acquired while working on quality issues benefits organizations, such as SNOMED International – the developers of SNOMED CT. More specifically, we have been a member of their Quality Assurance Committee (2007-2013). We also work in close collaboration with the RxNorm content development team at NLM. Several drug terminologies (namely ATC and DrugBank), while primarily used in research projects, were added as sources to RxNorm after we demonstrated the role RxNorm could play in the interoperability between clinical and research datasets.

Finally, as mentioned earlier, there are strong synergies between our research activities and the development of services. Our experience with interoperability issues and the methods we have developed to support it have had a direct influence on the services we develop to support the distribution and use of drug terminologies at NLM (e.g., the techniques for comparing pharmacologic classes implemented in *RxClass* originated in a research project).

4 Methods and Procedures

In a recent review of auditing methods for biomedical terminologies, Zhu and colleagues [5] 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 review the terminologies we investigated and present an overview of our research on quality assurance from the perspective of Zhu's analytical framework illustrated with specific examples. Then, we present three investigations in more detail to illustrate the range of methods we have developed.

4.1 Overview

4.1.1 Terminologies investigated

We investigated a wide variety biomedical terminologies, including drug terminologies (RxNorm, ATC – the Anatomical-Therapeutic-Chemical drug classification, DrugBank, NDF-RT – the National Drug File-Reference terminology, commercial drug knowledge bases), clinical terminologies (ICD – the International Classification of Diseases, LOINC, SNOMED CT), and specialized terminologies, such as HPO –

the Human Phenotype Ontology and the Orphanet terminology for rare diseases. Additionally, we investigated UMLS Metathesaurus, in which most of these terminologies are integrated. Finally, we also explored the Disease Ontology, a new terminology for diseases used in many research projects. As shown in Figure 1, most of our efforts were devoted to SNOMED CT and drug terminologies.

4.1.2 What is audited

The main types of features investigated for QA include terms/concepts, relations and definitions. Occasionally, we also investigated value sets and terminology services. As shown in Figure 2, our main focus has been on terms/concepts and relations, including the interplay between these two kinds of entities. We also leveraged the lexical features of terms to assess the coverage and alignment between HPO and SNOMED CT [13, 14]. Relations can be inspected for completeness (e.g., [29]) and for consistency (e.g., [31]).

4.1.3 Knowledge sources used for auditing

The same features used as the object of a QA investigation (e.g., terms/concepts, relations and definitions) can also be used to support this investigation. For example, our recent investigations of SNOMED CT [10, 11] identify missing concepts and hierarchical relations based on patterns found in existing terms and relations. In some cases, the investigation relies on intrinsic features, within the terminology under investigation (e.g., [10, 31]). In other cases, the investigation of a terminology leverages features from another terminology. This is often the case in coverage studies (e.g., [23, 30]) and when several terminologies are integrated for QA purposes (e.g., [7]). Of note, the logical definitions provided by terminologies based on description logic formalisms, such as SNOMED CT, provide a rich source of information for QA (e.g., [13, 31]). As shown in Figure 3, terms/concepts and relations were used predominantly in our investigations, alone or in combination.

4.1.4 Methods used for auditing

In terms of methods, approaches to auditing terminologies primarily include lexical, structural and semantic techniques. Lexical approaches are based on the properties of terms, such as compositionality (e.g., leveraged to derive partial mappings [13]). Structural approaches are based on the organizational structure of concepts in terminologies. Of particular interest are lattices (e.g., leveraged to identify potential errors in SNOMED CT [11]). Semantic methods generally rely on the logical definitions of concepts in description logic-based terminologies, such as SNOMED CT (e.g., [14]). Other approaches include transforming the representation of a terminology to a different formalism (e.g., converting the LOINC to OWL [7] or MeSH to RDF [32]), and evaluating the compliance of terminologies with desired principles (e.g., desiderata for drug classification systems [17]). Finally terminologies can also be evaluated by comparison to other ontologies to which a mapping can be derived (e.g., comparing the Disease Ontology to SNOMED CT [23]). As shown in Figure 4, we explored a variety of methods, isolated and in combination.

As mentioned earlier, these methods can be used to investigate various aspects of quality assurance. Our research includes a mix of basic and applied research (Figure 5). Basic research corresponds to the development of novel QA methods (e.g., [11, 28]), while applied research focuses on the application of existing techniques to assess the coverage of a terminology for a specific domain (e.g., coverage of phenotypes in SNOMED CT [23]) or the alignment between two terminologies (e.g., between the Disease Ontology and SNOMED CT [23]).

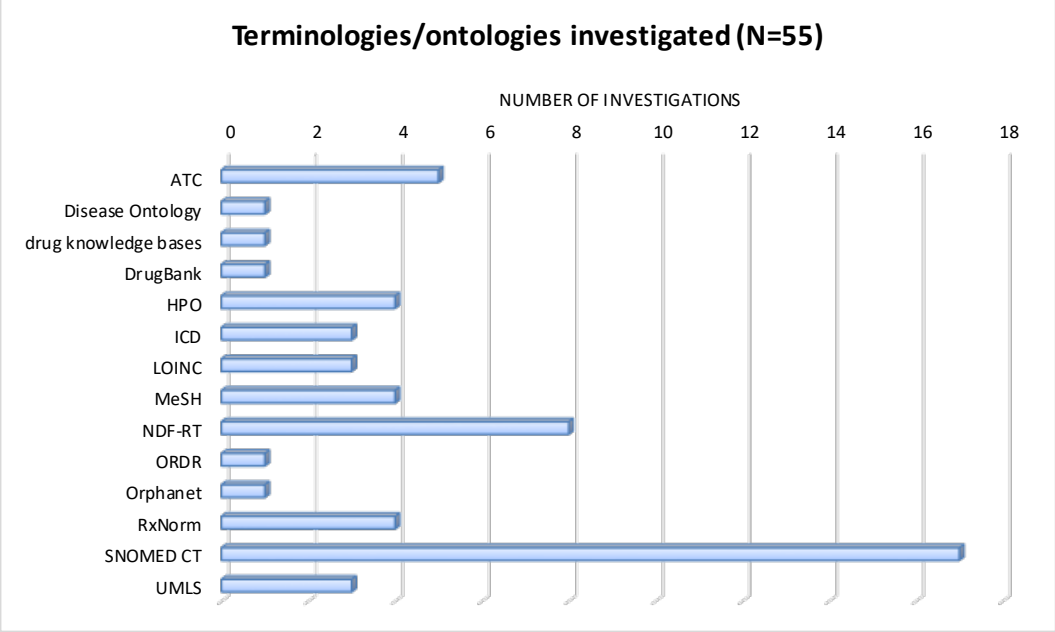

Figure 1. Terminologies and ontologies investigated

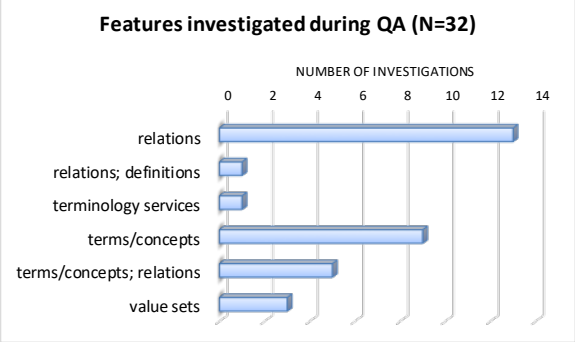

Figure 2. Types of features investigated during QA

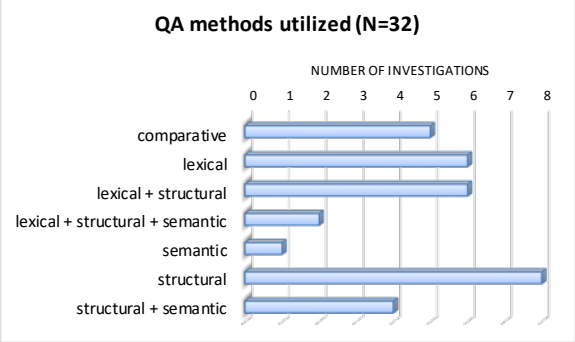

Figure 4. Types of methods utilized for QA

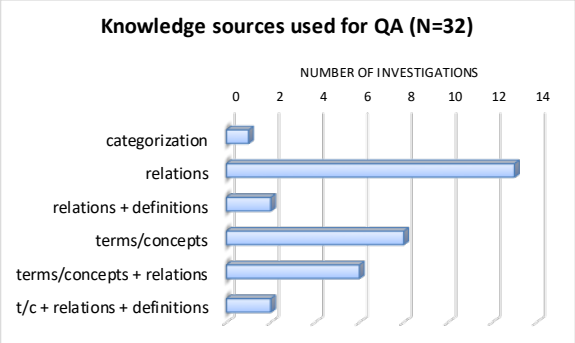

Figure 3. Types of knowledge sources used for QA

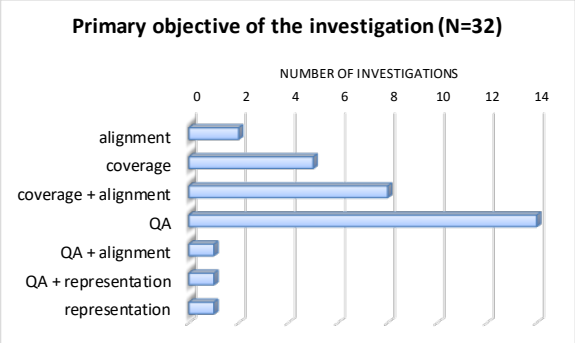

Figure 5. Primary objective of the QA investigation

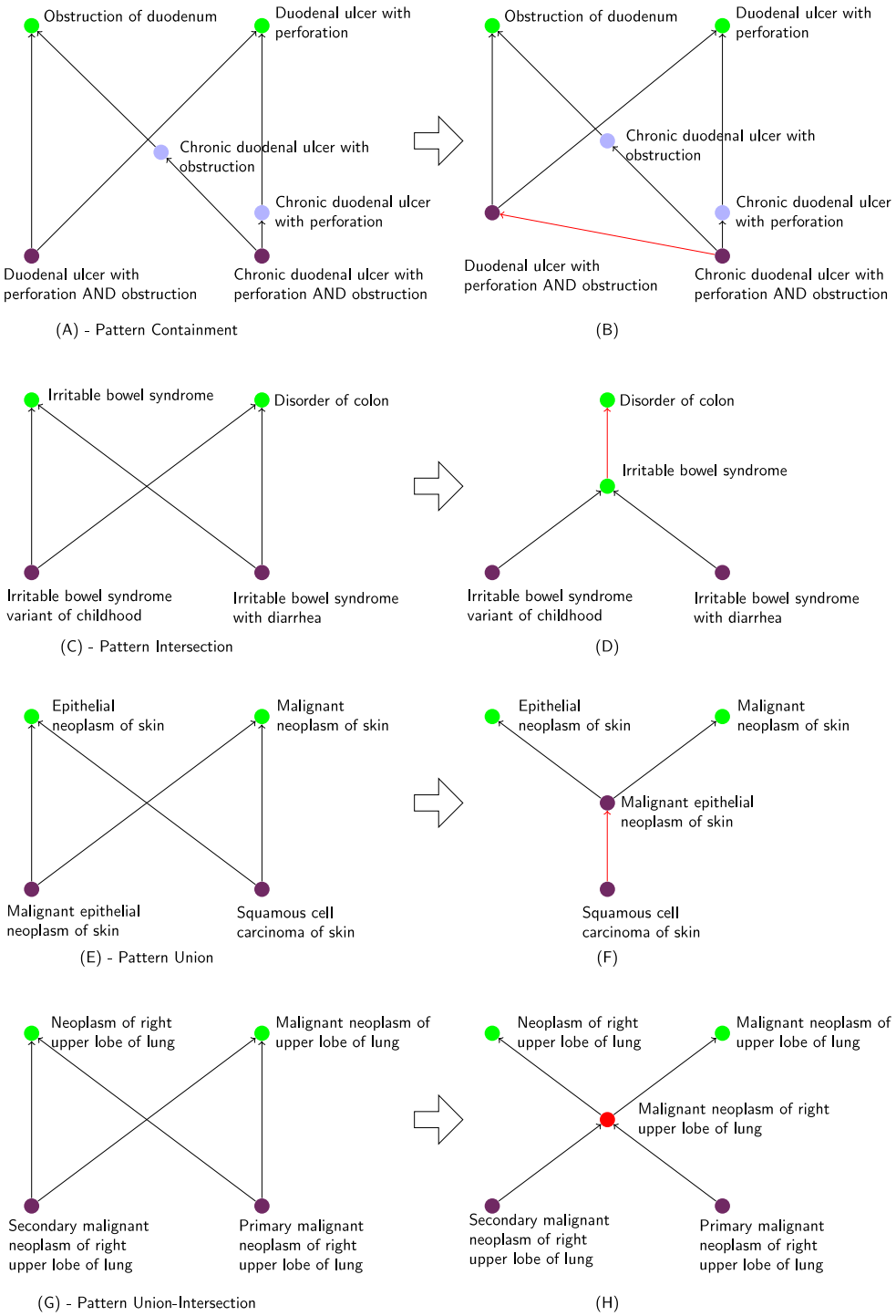

Figure 6. Examples of the four lexical patterns used to analyze non-lattice subgraphs (right-hand side) and suggest remediation (left-hand side)

4.2 Three examples

In this section, we present three quality assurance investigations to illustrate the variety of methods and terminologies to which they were applied.

4.2.1 Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT [11]

Objective: Quality assurance of large ontological systems such as SNOMED CT is an indispensable part of the terminology management lifecycle. We introduce a hybrid structural-lexical method for scalable and systematic discovery of missing hierarchical relations and concepts in SNOMED CT.

Material and Methods: All non-lattice subgraphs (the structural part) in SNOMED CT are exhaustively extracted using a scalable MapReduce algorithm. Four lexical patterns (the lexical part) are identified among the extracted non-lattice subgraphs. Non-lattice subgraphs exhibiting such lexical patterns are often indicative of missing hierarchical relations or concepts. Each lexical pattern is associated with a potential specific type of error. Examples of the four lexical patterns are shown in Figure 6.

Results: Applying the structural-lexical method to SNOMED CT (September 2015 U.S. edition), we found 6,801 non-lattice subgraphs that matched these lexical patterns, of which 2,046 were amenable to visual inspection. We evaluated a random sample of 100 small subgraphs, of which 59 were reviewed in detail by domain experts. All the subgraphs reviewed contained errors confirmed by the experts. The most frequent type of error was missing is-a relations due to incomplete or inconsistent modeling of the concepts.

Conclusions: Our hybrid structural-lexical method is innovative and proved effective not only in detecting errors in SNOMED CT, but also in suggesting remediation for these errors.

This investigation is part of an effort started in 2010 to leverage specific structural properties (non-lattice subgraphs) for QA purposes with application to SNOMED CT. It is a long-term collaboration Dr. GQ Zhang's team, now at the University of Kentucky, Lexington. One significant aspect of this work is the demonstration that structural features alone have limited precision in identifying errors [37]. So do lexical features alone [9]. In contrast, the combination of non-lattice subgraphs and lexical patterns has proved effective in identifying errors precisely. Moreover, this approach is not only a diagnostic tool, but can also suggest remediation for the errors identified. Therefore, it is effective in assisting the developers of terminologies. We very recently published a follow-up investigation generalizing this approach [10].

4.2.2 Interoperability between phenotypes in research and healthcare terminologies – Investigating partial mappings between HPO and SNOMED CT [13]

Background. Identifying partial mappings between two terminologies is of special importance when one terminology is finer-grained than the other, as is the case for the Human Phenotype Ontology (HPO), mainly used for research purposes, and SNOMED CT, mainly used in healthcare.

Objectives. To investigate and contrast lexical and logical approaches to deriving partial mappings between HPO and SNOMED CT.

Methods. 1) Lexical approach - We identify modifiers in HPO terms and attempt to map demodified terms to SNOMED CT through UMLS; 2) Logical approach - We leverage subsumption relations in HPO

to infer partial mappings to SNOMED CT; 3) Comparison - We analyze the specific contribution of each approach and evaluate the quality of the partial mappings through manual review.

Results. There are 7358 HPO concepts with no complete mapping to SNOMED CT. We identified partial mappings lexically for 33% of them and logically for 82%. We identified partial mappings both lexically and logically for 27%. The clinical relevance of the partial mappings (for a cohort selection use case) is 49% for lexical mappings and 67% for logical mappings.

Conclusions. Through complete and partial mappings, 92% of the 10,454 HPO concepts can be mapped to SNOMED CT (30% complete and 62% partial). Equivalence mappings between HPO and SNOMED CT allow for interoperability between data described using these two systems. However, due to differences in focus and granularity, equivalence is only possible for 30% of HPO classes. In the remaining cases, partial mappings provide a next-best approach for traversing between the two systems. Both lexical and logical mapping techniques produce mappings that cannot be generated by the other technique, suggesting that the two techniques are complementary to each other. Finally, this work demonstrates interesting properties (both lexical and logical) of HPO and SNOMED CT and illustrates some limitations of mapping through UMLS.

Extended example. To illustrate the main steps of our partial mapping approach, we consider the HPO concept Recurrent bronchitis [HP:0002837], for which there is no complete lexical mapping to SNOMED CT (Figure 7).

- *Partial lexical mapping*. The lexico-syntactic profile of this term is [MOD-HEAD], in which the head noun bronchitis is modified by the adjective Recurrent. We demodified this term by removing its sole modifier, Recurrent, resulting in the bare head noun, bronchitis. According to the UMLS, bronchitis is equivalent to three SNOMED CT concepts, Bronchitis (disorder) [SCTID:32398004], Acute bronchitis (disorder) [SCTID:10509002], and Acute tracheobronchitis (disorder) [SCTID:35301006]. Therefore, we identified a level-1 partial lexical mapping for Recurrent bronchitis [HP:0002837] to three target concepts in SNOMED CT.
- *Partial logical mapping*. The concept Recurrent bronchitis [HP:0002837] has three direct ancestors in the subsumption hierarchy of HPO, Abnormality of the bronchi [HP:0002109], Bronchitis [HP:0012387] and Recurrent upper respiratory tract infections [HP:0002788]. According to the UMLS, the concept Abnormality of the bronchi [HP:0002109] has no equivalent in SNOMED CT. The concept Bronchitis [HP:0012387] is equivalent to the same three concepts identified as a mapping for the demodified term bronchitis. Finally, the concept Recurrent upper respiratory tract infections [HP:0002788] is equivalent to two SNOMED CT concepts: Upper respiratory infection (disorder) [SCTID:54150009] and Recurrent upper respiratory tract infection (disorder) [SCTID:195708003]. Therefore, we inferred a partial logical mapping for Recurrent bronchitis [HP:0002837] to five target SNOMED CT concepts, three from Bronchitis [HP:0012387] and two from Recurrent upper respiratory tract infections [HP:0002788]. Of note, since a partial mapping was found through a direct ancestor of Recurrent bronchitis [HP:0002837], we did not explore its more distant ancestors.
- *Overall*. A partial mapping to SNOMED CT can be derived for the HPO concept Recurrent bronchitis [HP:0002837] both lexically and logically, at the first level (of demodification or subsumption) in both cases. Moreover, all the target concepts from the lexical mapping were also identified by the logical mapping, which also identified two additional target concepts.

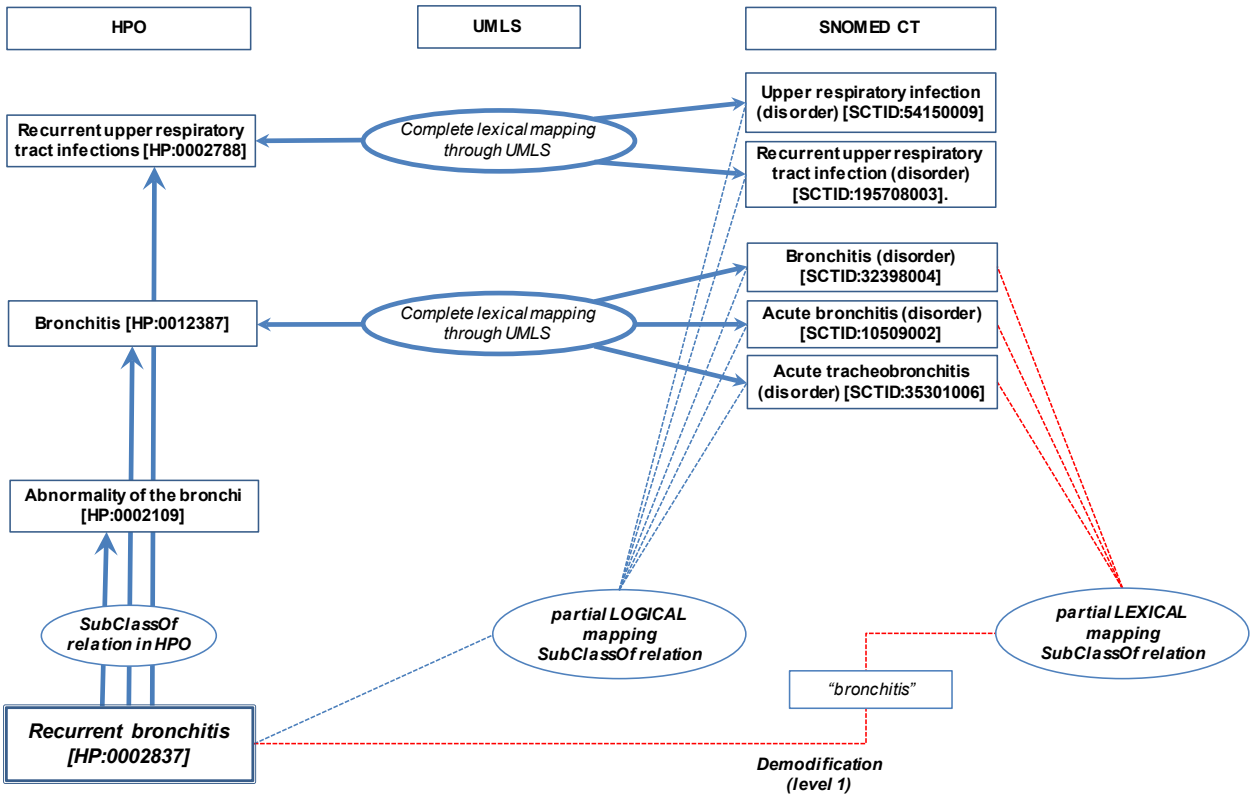

Figure 7. Example of partial logical mapping derived between HPO and SNOMED CT

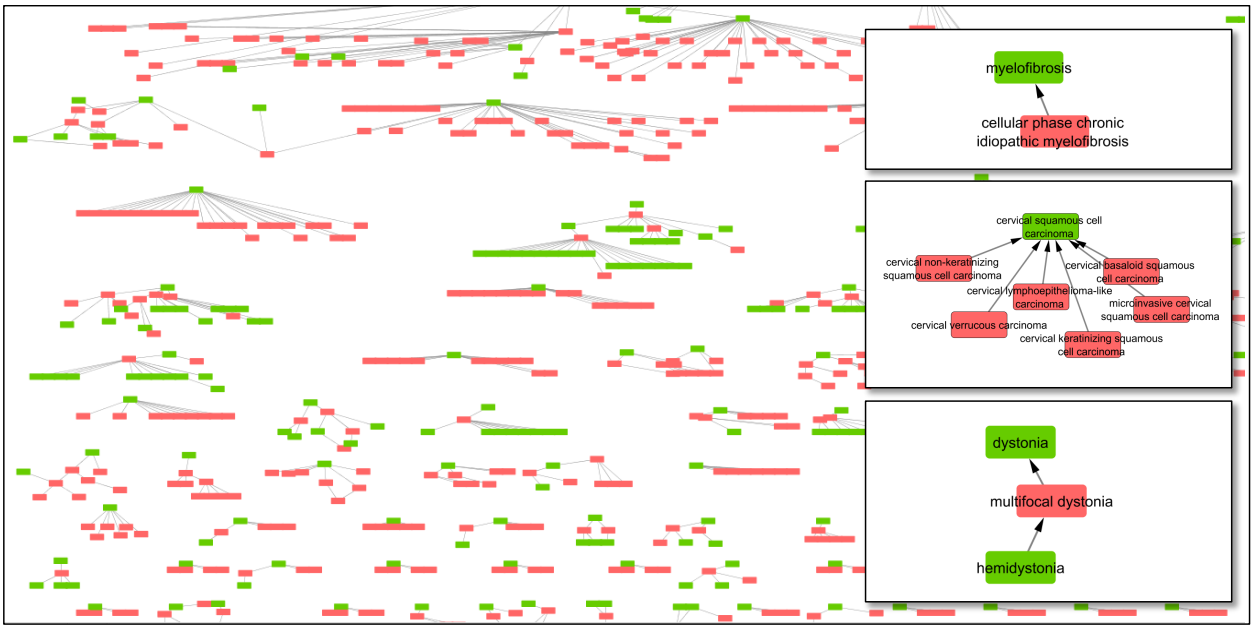

Figure 8. Characterization of the Disease Ontology concepts unmapped to SNOMED CT (red) through their hierarchical relations to shared concepts between the two terminologies (green) [overview visualization with Cytoscape]

4.2.3 Interoperability of disease concepts in clinical and research ontologies – Contrasting coverage and structure in the Disease Ontology and SNOMED CT [23]

Objectives. To contrast the coverage of diseases between the Disease Ontology (DO) and SNOMED CT, and to compare the hierarchical structure of the two ontologies.

Methods. We establish a reference list of mappings. We characterize unmapped concepts in DO semantically and structurally. Finally, we compare the hierarchical structure between the two ontologies.

Results. Overall, 4478 (65%) the 6931 DO concepts are mapped to SNOMED CT. The cancer and neoplasm subtrees of DO account for many of the unmapped concepts. The most frequent differentiae in unmapped concepts include morphology (for cancers and neoplasms), specific subtypes (for rare genetic disorders), and anatomical subtypes. Unmapped concepts usually form subtrees, and less often correspond to isolated leaves or intermediary concepts (Figure 8).

Conclusion. This detailed analysis of the gaps in coverage and structural differences between DO and SNOMED CT contributes to the interoperability between these two ontologies and will guide further validation of the mapping.

There are several aspects to the significance of this work. First, it was an opportunity to independently curate the mapping to SNOMED CT provided by the Disease Ontology. We eliminated wrong mappings and found mappings that had been missed by the DO developers. Second, it yielded an analysis of the differential coverage between DO and SNOMED CT at two different levels of granularity. Our overview revealed that the vast majority of DO concepts unmapped to SNOMED CT correspond to specialized concepts (i.e., descendants of existing SNOMED CT concepts) and represented specific types of tumors. Our detailed analysis of the differentiae revealed editorial differences between DO and SNOMED CT, namely that DO creates pre-coordinated concepts in cases where SNOMED CT prefers post-coordination. For example, both DO and SNOMED have a concept for ganglioneuroblastoma. However, only DO has a more specific concept for adrenal gland ganglioneuroblastoma. The expression of the same concept in SNOMED CT would require the use of post-coordination to combine ganglioneuroblastoma with a finding site (location) of Adrenal structure. Of note, SNOMED CT has pre-coordinated concepts for other locations of ganglioneuroblastoma (e.g., Ganglioneuroblastoma of central nervous system), but not for the adrenal gland.

5 Project Status

This project is an ongoing project, under the umbrella of the *Medical Ontology Research* project. As shown in Figure 9, our commitment to assessing the quality of biomedical terminologies has been sustained over the past twenty years and has intensified in the past few years. Excluding posters and abstracts, the 32 studies listed in this report represent 46% (32/69) of our publications since this project was last reviewed in 2010, and 35% (68/196) of our publications overall. Other research efforts in our project have included ontology alignment (with focus on anatomical ontologies), integration and dissemination of drug terminologies (e.g., *RxNav* and application programming interfaces for RxNorm), the use of Semantic Web technologies for information integration in biomedicine. Our Google Scholar citation profile, shown in Figure 10, offers a proxy for the impact of our work (3398 citations since 2009).

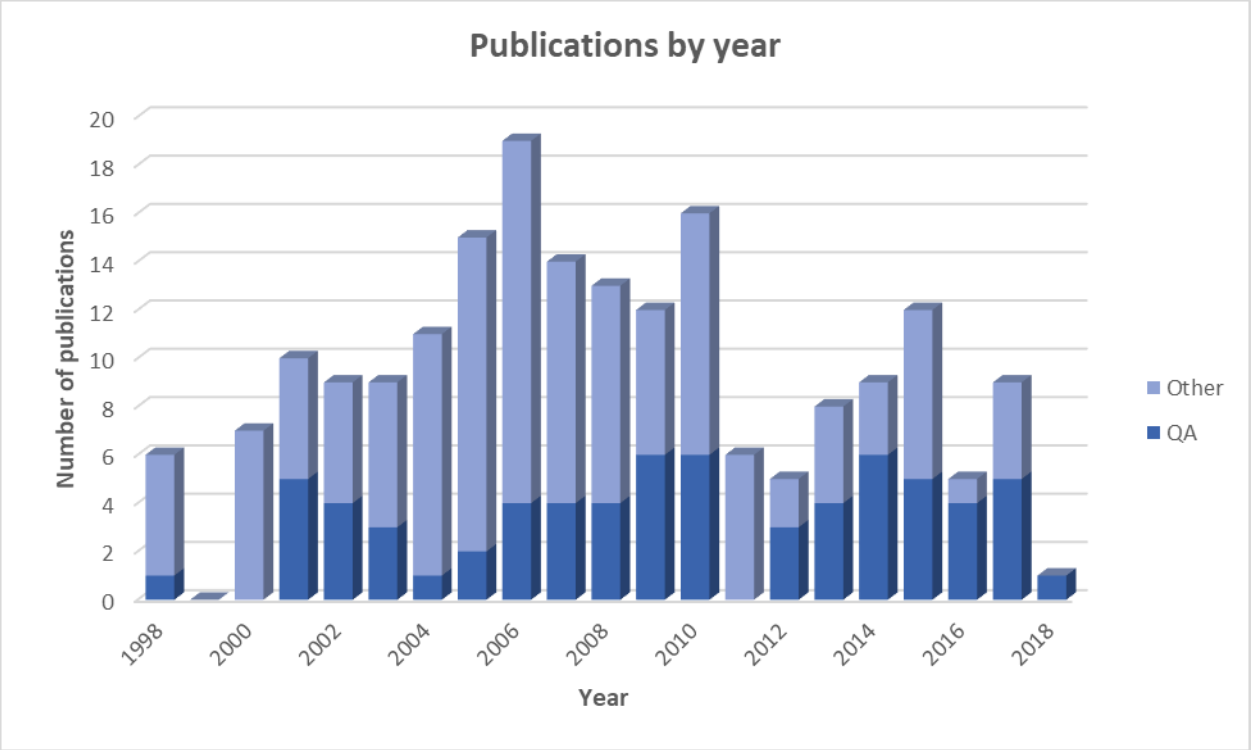

Figure 9. Number of publications per year

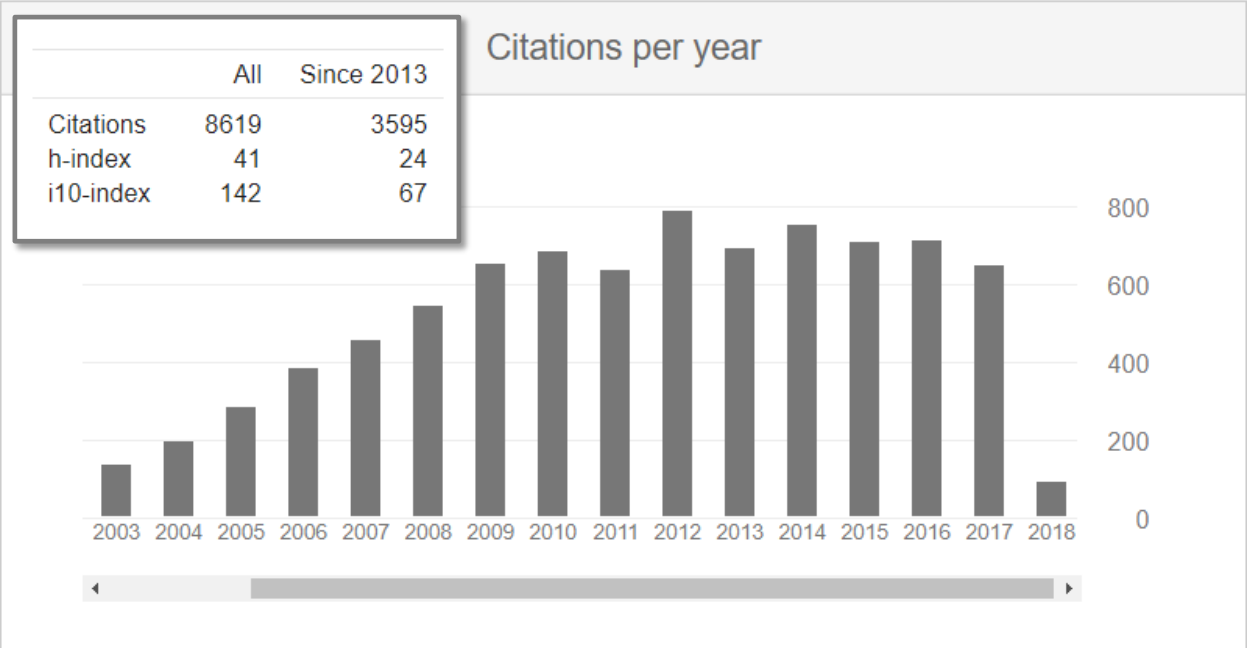

Figure 10. Google Scholar citations for Olivier Bodenreider as of March 2018

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 under investigation, 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 SNOMED CT ([11, 14]). In both cases, our findings were evaluated by a key member of the SNOMED CT development team. In other cases, the quality assurance or interoperability techniques are mostly automated. This is the case, for example, of the automatic process we developed for assessing the coverage of phenotypes [30], drug classes [18] or rare diseases [25]. In such investigations, two terminologies are simply compared against each other, but there is no gold standard per se.

7 Project Schedule and Resources

Some of these research studies have involved exclusively Lister Hill Center (LHC) personnel, including project staff members and other LHC researchers. (The *Medical Ontology Research* project has benefited from the programming support of two staff members over time, but they mostly support the development of our services, i.e., our drug terminology graphical and programming interfaces.)

Of note, 21 of the 32 studies listed in this report (66%) have involved post-doctoral fellows or summer (graduate and undergraduate) students. Here are some of the accomplishments of the five post-doctoral fellows:

- Bastien Rance (September 2010-June 2013) worked peripherally on the representation of rare diseases in biomedical terminologies [25] and created fingerprints to characterize UMLS vocabularies [24]. Bastien helped mentor summer student Thai Le.
- Tomasz Adamusiak (October 2011-March 2012) designed and implemented our QA analysis of LOINC using description logics [7].
- Rainer Winnenburg (January 2012-October 2014) contributed to multiple QA investigations (Meaningful Use value sets, pharmacologic classes, phenotypes, MeSH in RDF) and authored 12 publications on this topic [8, 14, 20, 27-35]. Rainer helped mentor summer student Nathan Bahr.
- Ferdinand Dhombres (November 2014-August 2016) investigated the representation of human phenotypes in SNOMED CT [12-14]. Of note, one of Dhombres' paper won best paper at Medinfo 2015.
- Satya Raje (June 2016-May 2017) investigated interoperability between the Disease Ontology, used in research projects, and SNOMED CT, used in healthcare settings [23].

Most of our investigation of hierarchical relations in SNOMED CT was performed in collaboration with Dr. GQ Zhang, currently at the University of Kentucky, Lexington. This collaboration started in 2010, when Dr. Zhang was a visiting scientist at the Lister Hill Center, with the identification of non-lattice fragments in SNOMED CT, yielding two papers [36, 37]. It intensified recently after we overcame some technical hurdles we encountered initially [10, 11, 38].

8 Summary and Future Plans

As part of the *Medical Ontology Research* project, we have explored quality assurance and interoperability issues in biomedical terminologies including the main standard clinical terminologies, SNOMED CT, LOINC and RxNorm, as well as emerging ontologies (e.g., HPO, Disease Ontology). 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. Moreover, we believe that efficient QA processes should have high precision and suggest remediation for the errors they identify.

The significance 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 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. Whenever possible, we have reported these issues to the developers of the biomedical terminologies we investigated.

In the future, we want to keep developing the use of Semantic Web technologies, including RDF (Resource Description Framework), SPARQL (SPARQL Protocol and RDF Query Language) and OWL (Web Ontology Language) to support quality assurance in biomedical terminologies, as these technologies can help reduce the amount of *ad hoc* programming necessary for investigating the quality of ontologies. In addition to methods to precisely identify errors, we want to work on methods for identifying the root cause of errors. In particular, we expect to develop strategies for "repairing" ontologies (i.e., fixing the logical definitions of concepts, such that the appropriate inferences can be produced). We have also started to investigate new terminologies of clinical interest, such as MED-RT, the successor of NDF-RT. We are especially interested in evaluating quality issues through applications (e.g., use of clinical terminologies for health analytics in OHDSI, the Observational Health Data Sciences and Informatics, and other networks of clinical data repositories). Finally, we remain committed to improving the quality of SNOMED CT through our participation in the activities of SNOMED International, as well as RxNorm through our long-term collaboration with its developers at NLM and its worldwide user community.

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. Rector, A.L., S. Brandt, and T. Schneider, *Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications.* J Am Med Inform Assoc, 2011. **18**(4): p. 432-40 [PubMed: 21515545].
4. Amith, M.F., Z. He, J. Bian, J. Antonio Lossio-Ventura, and C. Tao, *Assessing the Practice of Biomedical Ontology Evaluation: Gaps and Opportunities.* J Biomed Inform, 2018 [PubMed: 29462669].
5. 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].
6. Euzenat, J. and P. Shvaiko, *Ontology matching*. 2007, Berlin ; New York: Springer. ix, 333 p.
7. Adamusiak, T. and **O. Bodenreider**, *Quality assurance in LOINC using description logic.* AMIA Annu Symp Proc, 2012: p. 1099-1108 [PubMed: 23304386].

8. Bahr, N.J., S.D. Nelson, R. Winnenburg, and **O. Bodenreider**, *Eliciting the intension of drug value sets – Principles and quality assurance applications.* Stud Health Technol Inform (Proc Medinfo), 2017: p. 843-847 [PubMed: 29295218].
9. **Bodenreider, O.**, *Identifying missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names.* Proceedings of the 6th International Conference on Biomedical Ontology (ICBO 2016), 2016: p. (electronic proceedings: http://ceur-ws.org/Vol-1747/IT601_ICBO2016.pdf).
10. Cui, L., **O. Bodenreider**, J. Shi, and G.-Q. Zhang, *Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs.* Journal of Biomedical Informatics, 2018. **78**: p. 177-184 [PubMed: 29274386].
11. Cui, L., W. Zhu, S. Tao, J.T. Case, **O. Bodenreider**, and G.Q. Zhang, *Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT.* J Am Med Inform Assoc, 2017. **24**(4): p. 788-798 [PubMed: 28339775].
12. Dhombres, F. and **O. Bodenreider**, *Investigating the lexico-syntactic properties of phenotype terms – Application to interoperability between HPO and SNOMED CT.* Proceedings of the Joint Bio-Ontologies and BioLINK ISMB'2015 SIG session "Phenotype Day", 2015: p. 8-11.
13. Dhombres, F. and **O. Bodenreider**, *Interoperability between phenotypes in research and healthcare terminologies – Investigating partial mappings between HPO and SNOMED CT.* J Biomed Semantics, 2016. **7**(3) [PubMed: 26865946].
14. Dhombres, F., R. Winnenburg, J.T. Case, and **O. Bodenreider**, *Extending the coverage of phenotypes in SNOMED CT through post-coordination.* Stud Health Technol Inform (Proc Medinfo), 2015: p. 795-799 [PubMed: 26262161].
15. Fung, K.W., J.E. Kapusnik-Uner, J. Cunningham, S. Higby-Baker, and **O. Bodenreider**, *Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support.* J Am Med Inform Assoc, 2017. **24**(4): p. 806-812 [PubMed: 28339701].
16. Fung, K.W., R.L. Richesson, and **O. Bodenreider**, *Coverage of rare disease names in standard terminologies and implications for patients, providers, and research.* AMIA Annu Symp Proc, 2014: p. 564-570 [PubMed: 25954361].
17. Kury, F.S.P. and **O. Bodenreider**, *Desiderata for drug classification systems for their use in analyzing large drug prescription datasets.* Proceedings of the 3rd Workshop on Data Mining for Medical Informatics (DMMI 2016), 2016.
18. Mortensen, J. and **O. Bodenreider**, *Comparing pharmacologic classes in NDF-RT and SNOMED CT.* Proceedings of the Fourth International Symposium on Semantic Mining in Biomedicine (SMBM 2010), 2010: p. 116-121 (electronic proceedings: http://ceur-ws.org/Vol-714/).
19. Mougin, F., A. Burgun, and **O. Bodenreider**, *Comparing drug-class membership in ATC and NDF-RT.* Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI 2012), 2012: p. 437-443.
20. Nelson, S.D., J. Parker, R. Lario, R. Winnenburg, M.S. Erlbaum, M.J. Lincoln, and **O. Bodenreider**, *Interoperability of medication classification systems: Lessons learned mapping established pharmacologic classes (EPCs) to SNOMED CT.* Stud Health Technol Inform (Proc Medinfo), 2017: p. 920-924.
21. Pathak, J., L.B. Peters, C.G. Chute, and **O. Bodenreider**, *Comparing and evaluating terminology services APIs: RxNav, UMLSKS, and LexBIG.* J Am Med Inform Assoc, 2010. **17**(6): p. 714-719 [PubMed: 20962136].
22. Peters, L.B., N. Bahr, and **O. Bodenreider**, *Evaluating drug-drug interaction information in NDF-RT and DrugBank.* J Biomed Semantics, 2015. **6**(19) [PubMed: 25964850].
23. Raje, S. and **O. Bodenreider**, *Interoperability of disease concepts in clinical and research ontologies – Contrasting coverage and structure in the Disease Ontology and SNOMED CT.* Stud Health Technol Inform (Proc Medinfo), 2017: p. 925-929 [PubMed: 29295235].
24. Rance, B., T. Le, and **O. Bodenreider**, *Fingerprinting biomedical terminologies – Automatic classification and visualization of biomedical vocabularies through UMLS semantic group profiles.* Stud Health Technol Inform (Proc Medinfo), 2015: p. 771-775 [PubMed: 26262156].
25. Rance, B., M. Snyder, J. Lewis, and **O. Bodenreider**, *Leveraging terminological resources for mapping between rare disease information sources.* Stud Health Technol Inform (Proc Medinfo), 2013. **192**: p. 529-533 [PubMed: 23920611].

26. Tao, S., L. Cui, W. Zhu, M. Sun, **O. Bodenreider**, and G.-Q. Zhang, *Mining relation reversals in the evolution of SNOMED CT.* Proceedings of the AMIA Summit on Translational Bioinformatics, 2015: p. 46-50 [PubMed: 26306232].
27. Winnenburg, R. and **O. Bodenreider**, *Issues in creating and maintaining value sets for clinical quality measures.* AMIA Annu Symp Proc, 2012: p. 988-996 [PubMed: 23304374].
28. Winnenburg, R. and **O. Bodenreider**, *Metrics for assessing the quality of value sets in clinical quality measures.* AMIA Annu Symp Proc, 2013: p. 1497-1505 [PubMed: 24551422].
29. Winnenburg, R. and **O. Bodenreider**, *Exploring pharmacoepidemiologic groupings of drugs from a clinical perspective.* Stud Health Technol Inform (Proc Medinfo), 2013. **192**: p. 827-831 [PubMed: 23920673].
30. Winnenburg, R. and **O. Bodenreider**, *Coverage of phenotypes in standard terminologies.* Proceedings of the Joint Bio-Ontologies and BioLINK ISMB'2014 SIG session "Phenotype Day", 2014: p. 41-44.
31. Winnenburg, R. and **O. Bodenreider**, *Evaluating the consistency of inferred drug-class membership relations in NDF-RT.* Proceedings of the 17th ISMB'2014 SIG meeting "Bio-Ontologies", 2014: p. (electronic proceedings: https://drive.google.com/file/d/0B6AZUdKw1IVHeGVtTVZCRXJhbFk/).
32. Winnenburg, R. and **O. Bodenreider**, *Desiderata for an authoritative representation of MeSH in RDF.* AMIA Annu Symp Proc, 2014: p. 1218-1227 [PubMed: 25954433].
33. Winnenburg, R. and **O. Bodenreider**, *A framework for assessing the consistency of pharmacological classes across sources.* J Biomed Semantics, 2014. **5**(30) [PubMed: 25101165].
34. Winnenburg, R., J. Mortensen, and **O. Bodenreider**, *Using description logics to evaluate the consistency of drug-class membership relations in NDF-RT.* J Biomed Semantics, 2015. **6**(13) [PubMed: 25866612].
35. Winnenburg, R., L.M. Rodriguez, F. Callaghan, A. Sorbello, A. Szarfman, and **O. Bodenreider**, *Aligning pharmacologic classes between MeSH and ATC.* Proceedings of the International Workshop on Vaccine and Drug Ontology Studies (VDOS 2013), 2013: p. (electronic proceedings: http://www2.unb.ca/csas/data/ws/semantic-trilogy-workshops/papers/ICBO-Workshops.pdf).
36. Zhang, G.-Q. and **O. Bodenreider**, *Using SPARQL to test for lattices: application to quality assurance in biomedical ontologies*, in *Proceedings of the 9th International Semantic Web Conference (ISWC 2010), Shanghai, China, November 7-11, 2010*, P.F. Patel-Schneider, et al., Editors. 2010, Springer: Berlin, Heidelberg. p. 273-288 [PubMed: 25699294].
37. Zhang, G.-Q. and **O. Bodenreider**, *Large-scale, exhaustive lattice-based structural auditing of SNOMED CT.* AMIA Annu Symp Proc, 2010: p. 922-926 [PubMed: 21347113].
38. Zhang, G.-Q., W. Zhu, M. Sun, S. Tao, **O. Bodenreider**, and L. Cui, *MaPLE: A MapReduce Pipeline for Lattice-based Evaluation and its application to SNOMED CT.* Proceedings of the IEEE International Conference on Big Data 2014 (IEEE BigData 2014), 2014: p. 754-759 [PubMed: 25705725].
39. Ochs, C., J. Geller, Y. Perl, Y. Chen, J. Xu, H. Min, J.T. Case, and Z. Wei, *Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies.* J Am Med Inform Assoc, 2015. **22**(3): p. 507-18 [PubMed: 25336594].
40. Ochs, C., Y. Perl, J. Geller, M. Halper, H. Gu, Y. Chen, and G. Elhanan, *Scalability of abstraction-network-based quality assurance to large SNOMED hierarchies.* AMIA Annu Symp Proc, 2013. **2013**: p. 1071-80 [PubMed: 24551393].
41. Zheng, L., H. Yumak, L. Chen, C. Ochs, J. Geller, J. Kapusnik-Uner, and Y. Perl, *Quality assurance of chemical ingredient classification for the National Drug File - Reference Terminology.* J Biomed Inform, 2017. **73**: p. 30-42 [PubMed: 28723580].