Health Concept and Knowledge Management: Twenty-five Years of Evolution

R. Cornet[1, 2], C. G. Chute[3]

[1] Academic Medical Center – University of Amsterdam, Department of Medical Informatics, Amsterdam, The Netherlands

[2] Linköping University, Department of Biomedical Engineering, Linköping, Sweden

[3] Johns Hopkins Medicine, Johns Hopkins University, Division of General Internal Medicine, Baltimore, MD 21287, USA

Summary

Objectives: The fields of health terminology, classification, ontology, and related information models have evolved dramatically over the past 25 years. Our objective was to review notable trends, described emerging or enabling technologies, and highlight major terminology systems during the interval.

Methods: We review the progression in health terminology systems informed by our own experiences as part of the community involved in this work, reinforced with literature review and citation.

Results: The transformation in size, scope, complexity, and adoption of health terminological systems and information models has been tremendous, on the scale of orders of magnitude.

Conclusion: The present "big science" era of inference and discovery in biomedicine would not have been possible or scalable absent the growth and maturation of health terminology systems and information models over the past 25 years.

Keywords

Terminology, classification, ontology, information model

Yearb Med Inform 2016;Suppl1:S32-41
http://dx.doi.org/10.15265/IYS-2016-s037
Published online August 2, 2016

Prologue

The authors of a 1965 JAMA editorial, "What's In a Name? [1]," observed "… the 'names' of diseases represent concepts very different from those of even a dozen years ago." They open with deference to Socrates, and his challenges around discovering clear meanings of words among his Athenian colleagues. It seems fitting, then, that modern description logics arise from the logical reasoning of Socrates' student, Aristotle. With this 25th anniversary issue of the IMIA Yearbook, we examine some of the trends in terminology, ontology, and concept representation over these past 25 years, or approximately since 1990. We consider less the semantic drift of specific names, and more the underpinning informatics of terminology representation and concept science.

Some understanding of terms, not unseemly for an article about terminology, can frame our considerations. While many definitions abound, we opt for more generalized, informal, and flexible understandings of our key terms. We assert the global assumption that all of these terms pertain to health care, biomedicine, or the clinical sciences.

Terminology: A system of concepts with assigned identifiers and human language terms, typically involving some kind of semantic hierarchy. Some systems may support the assignment of multiple terms, or synonyms, to a given concept; these may include terms in multiple natural languages, such as English or Dutch.

Ontology: A terminology invoking formal semantic relationships between and among concepts, typically manifest as a type of Description Logic.

Classification: A terminology system intended to exhaustively describe a domain or topic, typically invoking the judicious placement of residual categories, such as Unspecified or Not Elsewhere Classified, to achieve comprehensiveness.

Statistical Classifications: A classification where all concepts are mutually exclusive to avoid counting things twice. This is typically achieved using a mono-hierarchy, where each concept has one and only one parent.

A word about personalities and credit throughout this history is in order. Obviously the sweep of twenty-five years engaged countless persons in these efforts. With few exceptions, we have chosen to imply credit to persons for major work and advances through their inclusion by citation, otherwise this document would be larger still with enumerations of the names who made these advances possible.

In this historical retrospect, we address the major changes over the last 25 years. We first address different modes of medical concept representation, which have evolved from being predominantly based on classifications to more or less formal ontologies, and the development of thesauri that focus on descriptions rather than concepts, supporting use of human languages and natural language processing. Then we address the information models, which not only provide a structure in which terminologies can be used, but are also themselves increasingly based on terminologies. Then we briefly touch on the developments in formalisms underlying medical concept representation. We conclude with a brief outlook on the impact of the merger of the various technological advances that were made largely in

relative isolation, but that are now maturing to become a basis for the imminent wave of (big) data analytics.

Medical Concept Representation

The science of unambiguously representing biomedical concepts has a long history [2]. The past 25 years have been particularly transformative. What has changed is the introduction of computer-driven ontological reasoners, virtually unknown prior to 1990. While the principles of first order predicate logics have been well understood for centuries, their computable subsets as Description Logics [3], and the reasoners that go with them, are a distinctly recent innovation. These innovations and their recent history are considered in the second part of this review.

Distinguishing statistical classifications from terminologies is important, as they serve different purposes. The most visible and often reviled statistical classification is the multiple versions of the International Classification of Diseases (ICDs), described below. Criteria or desiderata for one--say, terminology--should not be applied to the other (statistical classification). Cimino's famous Desiderata paper [4] correctly asserted that poly-hierarchy and "no residual categories" were among the criteria for a good terminology. However, statistical classifications by definition cannot adhere to those precepts in particular and remain mutually exclusive and exhaustive. This does not mean that statistical classifications are bad terminologies, rather they are a use-case specific subset of terminology that must be distinguished with respect to criteria for a "good terminology."

Classifications

The ICDs

The International Classification of Diseases (ICD) has a venerable history, arguably going back to the 16th century London Bills of Mortality [5]. However, for the purposes of this review, the 10th revision of the ICD (ICD-10) was published at the start of our 25-year window in 1990. This may surprise American readers, as the United States did not choose to adopt ICD-10 until the end of our 25-year window, in 2015. Architecturally, the ICD has not fundamentally changed from the 16th century model of the London Bills, in that each new code is added as a new row in a single list; there was no post-coordination of terms. Additionally, at the time that ICD-10 was introduced, it remained a paper artifact; for all practical purposes, ICD-10 was not authored, edited, or published in electronic format.

The Australians, with their National Center for Classifications in Health at the University of Sydney, became the first group to migrate ICD-10 into an electronic format. They created and curated the Australian adaptation, ICD-10-AM, with computer assistance beginning in the mid-1990s. They were followed almost immediately by the Nordic countries, who jointly created their adaptation of ICD-10 using electronic media. By 2005, the WHO-FIC, an organization chartered by the World Health Organization (WHO) comprising national centers for classification around the world, created an international forum to advise on the content and evolution of WHO's Family of International Classifications (WHO-FIC) which includes ICD and the International Classification of Functioning (ICF).

Currently, the WHO manages an electronic revision and update platform with WHO-FIC as a webpage. Additionally, the publication and distribution of ICD content in many cases adheres to an ISO standard, the Classification Markup Language (ClaML) [6], an XML schema specifically developed for the purpose of systematically annotating features unique to ICD structures, such as *excludes*, *code as, code also*, and *includes* notations. ClaML, while human language- independent, is still largely oriented toward preserving annotations needed for accurately printing ICD content in a paper book format; there is less attention given to the consistency of underlying semantics and none given to terminology linkages outside of the ICD family.

Evolution Through ICD-11 Thinking

Beginning in 2006, the WHO commenced a strategic rethinking of how the ICD should be produced, maintained, and integrated with emerging electronic environments. While fidelity with the historical scaffolding and content rubrics of earlier revisions of the ICD, such as ICDs 9 and 10, were a major priority to minimize disorientation of users and disruption of statistical trends for morbidity or mortality, a fundamental rethinking of how the ICD can and should connect to related terminologies, such as SNOMED CT, or should be used with Electronic Health Records (EHRs) drove the architecture of ICD-11.

ICD-11 is a suite of terminology and classification artifacts, with a multiple-inheritance network of terms and concepts called the Foundation Component of ICD-11 [7] as its semantic core. The WHO and the International Health Terminology Standards Development Organisation (IHTSDO) gave great attention to aligning this Foundation Component with SNOMED CT [8]. A subset of SNOMED CT was designated as the Common Ontology, and provides the semantic backbone for the Foundation Component, which, while not yet fully populated, has demonstrated methodology and utility in the circulatory system domain[9].

However, such a multi-hierarchy semantic artifact as the Foundation Component violates two core principles for statistical classifications: mutual exclusivity and exhaustiveness. The first requires that a classification architecture be a mono-hierarchy, so that the is no chance that a specific instance can appear in more than one place on the tree—be double counted—for statistical purposes. This obligation forces a violation of one of Cimino's desiderata, multiple inheritance, but as mentioned, those desiderata should not be applied to statistical classifications. This also introduces arbitrary parenting structures, where for example gastric cancer is "allowed" to be in the cancer hierarchy, but therefore cannot be in the GI disease hierarchy. The reality that gastric cancer is also a disease of the gastrointestinal system must be suppressed to achieve mutually exclusive coding. The exhaustiveness obligation, that every diagnosis have a place where it can be coded, requires the introduction of residual categories, such as *Not Elsewhere Classified* (NEC) or *Not Otherwise Specified* (NOS); again, these fly in the face of Cimino's desiderata. Nevertheless, they are necessary for a statistical classification, which remains the primary use case for the ICD.

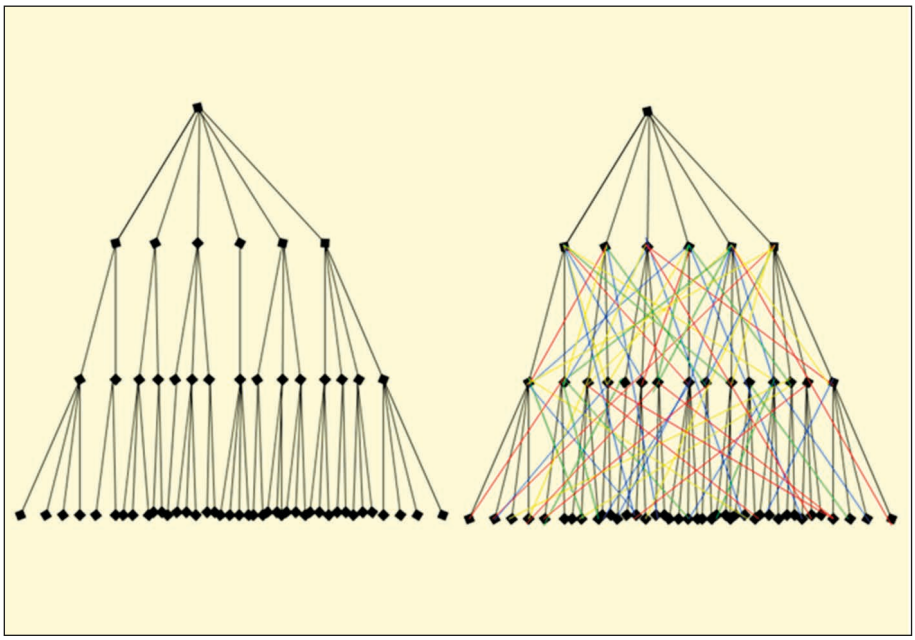

Fig 1 Cartoon illustrations of a mono-hierarchy (left) compared to the enriched linkages of a poly-hierarchy (right) exhibiting multiple parents for many nodes.

To solve the conundrum of having a well-formed semantic core (the Foundation Component) which does adhere to Cimino's Desiderata, while at the same time delivering a workable statistical classification, required the introduction of another layer of ICD-11 architecture. To address the mutually exclusive requirement, one can "walk the semantic network tree" in a continuous line to create a mono-hierarchy, or a linear serialization of the semantic network that can be printed in a book. ICD11 calls this a linearization, which forces each concept to have one and only one parent. However, a mono-hierarchy, as outlined above, forces arbitrary selections for parenting which tend to follow historical patterns of disease classification—gastric cancer will likely always be a cancer rather than a GI disease in the mortality linearization of the ICD.

Creating a mono-hierarchy is only one half of the requirements for a statistical classification. The other is the algorithmic addition of residual categories such as NEC and NOS to codable rubrics in the ICD-11 tree of linearizations. While these are sometimes pejoratively dismissed as "waste-basket" categories, they are crucial for satisfying the exhaustive criteria, where every health disease or condition must have a codable rubric, even if that is a residual category, to contain it.

The major linearization of ICD-11 is the Joint Linearization for Morbidity and Mortality Statistics (JLMSS). As might be inferred by now, if ICD-11 can generate one serialization of the Foundation Component, it can generate an arbitrary number of linearizations with varying choices for parenting and depth of the network linearized before being subsumed into residual categories (called "creating a linearization shoreline"). At present a separate linearization is being developed for Primary Care, which is more shallow than the JLMSS. Similarly, sub-specialty linearizations are contemplated, which go deeper into the Foundation Component before invoking residual categories, while selectively pruning hierarchies outside the specialty domain.

The application of the ICD to morbidity has historically had some major shortcomings. As the ICD was developed to tabulate mortality statistics, the notion of disease severity was moot—once one has died, how dead one might be makes little sense. However, disease severity is a major concern in morbidity. Consider two men with prostate cancer, where one has indolent, low grade, organ confined disease discovered incidentally during a routine trans-urethral resection of the prostate to treat BPH, which the other man has widely metastatic disease. To not be able to distinguish between these men for purposes of clinical guidelines, quality metrics, or discovery research is problematic. ICD-11 explicitly accommodates severity, and a great many other axes of qualification such as temporality, acuity, anatomy, and extent, by introducing post-coordination. Diseases can also be administratively qualified by attributes such as *present_on_admission, rule out, family_history*, or *history_of* .

Case Mix and Severity Classifications

With the advent of capitated payments came the inevitable claim "our patients are sicker than everyone else's, and we should get paid more." The question was how to objectively determine severity of illness, in order to appropriately adjust capitated payments, or case mix. As outlined above, traditional disease classifications such as the ICD did not enjoy explicit severity of illness parameters; all that could be done was to infer disease severity on the basis of co-morbidity. Thus, if one had congestive heart failure, and also had liver disease, lung disease, and perhaps neurological disease, then one's congestive heart failure was inferred to be severe relative to patients who did not have those co-morbidities. The fact that there may be no evidence demonstrating causality between the condition of interest, such as congestive heart failure, and the co-morbid conditions is notwithstanding. Case mix required some objective metrics, and co-morbidity was it.

The most ubiquitous set of measures for co-morbidity has been the Diagnosis-Related Groups (DRGs) [6]. While several versions have continued to emerge since their inception, architecturally they have changed little over the past 25 years, combining demographic, diagnoses, and procedures into several hundred categories of care. These categories can in turn be considered to have, or not have, "complications." An entire software industry has arisen, combing

electronic health records and tweaking the order of diagnoses in order to maximize the reimbursement categories for patients, particularly hospitalized patients.

The Johns Hopkins Adjusted Clinical Groups (ACG®) System is one of the most widely adopted case-mix systems internationally [10]. In the United States, it is the preferred severity adjustment system among health services researchers and other academics due to the high reliability and validity of its risk scores [11]. ACG was generated using data from commercially managed patient populations and Medicaid data, which reflect relatively healthy populations. Furthermore, ACG's adjustments are not impacted by encounter frequency, but rather clinically relevant information, thus isolating these adjustments from variations in practice style and patient visit patterns. The ACG in recent years has begun to explicitly incorporate clinical data from EHRs, thus deepening the scope of information brought to bear for severity assignation.

Functional Classification

Many scholars in the rehabilitation medicine field long recognized [12] that the entire axis of patient functioning, a strong predictor of patient outcomes, was completely missing from major classification efforts, although there were efforts by the early 1990s in this direction [13]. Nevertheless, an ambitious and comprehensive effort to create a multi-axial classification of functioning convened by the WHO published the International Classification of Impairments, Disabilities, and Handicaps [14], followed by a second edition in 1997 [15]. However, a major shift in philosophy occurred in the functional classification community, focusing on a biopsychosocial model of functioning and human capacities, rather than emphasizing impairments. Thus emerged a radically revised International Classification of Functioning (ICF) in 2001 [16], with a genuine multi-axial and multidisciplinary perspective. The ICF enjoys a large and passionate development community, with recent innovations being the merger of the previously separate ICF-CY (for Children and Youth) into the parent ICF in 2012.

Terminology Systems

Bioinformatics and Basic Sciences

No treatment of terminology and classification around the turn of the 21st century would be complete without mention of innovations over the past 25 years in the biology and biomedical research community. A cohesive alignment around the OBO-Foundry [17], which coordinates the development of interlocking ontologies sharing a common "upper ontology" and invoking description logics such as OWL, has greatly advanced the quality and quantity of biomedical ontologies, predominantly among the biomedical research community. By far the highest profile and highest impact classification in this family is the Gene Ontology (GO) [18], although it preceded the formation of the OBO community. The GO's characterization of "gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner" [19] enables the annotation of model organism databases, analytic datasets, and many products of biological research. As such, the GO has been perhaps the single greatest enabler of the emergence of big science, invoking GO-annotated "big data" in the early 21st century.

SNOMED

The use of terminologies to capture data at such a level of detail that it is useful in clinical practice has long been a goal. Already in 1928 the National Conference on Nomenclature of Disease was organized in the USA to present a logical clinical nomenclature with two axes: topography and etiology [20]. This resulted in the Standard Nomenclature of Diseases and Operations (SNDO), which was published between 1933 and 1961. After the discontinuation of SNDO, the College of American Pathologists (CAP) published the Systematized Nomenclature of Pathology (SNOP) in 1965, which became the international standard within pathology. SNOP evolved and extended to all of medicine and in 1974 became SNOMED (Systematized Nomenclature of Medicine). In 1979 SNOMED II was published [21]. It remained the current version until 1992. This 7-axis nomenclature contained about 50,000 nomina [22]. It was superseded by SNOMED International [23] (SNOMED 3) in 1993 and SNOMED 3.5 in 1998. The latter included 12 axes (Chemicals, Drugs and Biological Products; Diseases/Diagnoses; Function; General Linkage/Modifiers; List of Pharmaceutical Companies; Living Organisms; Morphology; Occupations; Physical Agents; Forces; and Activities; Procedures; Social Context; and Topography) and vastly increased the number of concepts to approximately 114,000 involving 165,000 unique names. Size mattered, as demonstrated in a comprehensive content coverage study [24] of clinical classifications and terminologies in 1996, which showed SNOMED 3 to have significantly greater coverage of medical record content compared to all other clinical nomenclatures.

Table 1 A sample of expressive examples for appendicitis from SNOMED International. Adapted from Evans et al. [25] In fact, the Canon group found 17 distinct combinations of SNOMED concepts to represent appendicitis.

Combination Set	SNOMED International Codes	Term Expansion
1	D5-46210	Acute appendicitis, NOS
2	Df-46100 G-A231	Appendicitis, NOS Acute
3	M-41000 G-C006 T-59200	Acute inflammation, NOS In Appendix, NOS
4	G-A231 M-40000 G-C006 T-59200	Acute Inflammation, NOS In Appendix, NOS

A feature of a rich, compositional terminology such as SNOMED International, was that some concepts (for example, appendicitis; see Table 1) could be expressed in myriad ways using compositional axes. However, as the Canon group pointed out [25], this and many related complexities created unanticipated problems, such as attempting to retrieve appendectomy required that one anticipate all compositional forms of the

single composite concept. In part to address this, Mayo Clinic and Kaiser Permanente partnered in acquiring a grant jointly funded by NIH and AHRQ to craft a logic-based terminology, spawning the Convergent Medical Terminology (CMT) Project in the late 1990s. CMT was a major breakthrough with regard to the underlying formalism [26, 27]. It moved away from a multi-axial representation which had been used from SNDO through SNOMED International towards a logic-based representation, in which concepts were formally defined by use of attributes and attribute values, in a representation language called Ontylog [28], an early description logic language. The use of the Ontylog reasoner helped in overcoming two important drawbacks of multi-axial systems: 1) the capability of detecting semantic equivalence of syntactically different expressions, and 2) the automated classification of concepts in a hierarchy. The CMT project led directly to the creation of SNOMED RT (for Reference Terminology), released in 2000. It is interesting to note that SNOMED CT, to the present day, does not invoke a conventional description logic such as OWL but rather an EL++ [29] dialect of description logic that supports complete reasoning in polynomial time, inheriting this heritage from its Ontylog roots [28].

It had long been recognized that the multi-axial system allowed for different ways of expressing the same meaning (Table 1). As another example, "Acute interstitial pneumonitis" could be encoded as a single concept, as a combination of "acute" and "interstitial pneumonitis" by specifying all constituent parts "acute" + "interstitial" + "lungs" + "inflammation", or other combinations It is important to be able to recognize that these multiple representations bear the same semantics. Similarly, it is important to determine which expressions are mere refinements of other expressions. To compute these distinctions, it is essential to be able to distinguish between concept definitions that contain only necessary properties (called primitive concept definitions) and definitions based on necessary and sufficient properties (called fully defined concepts).

Meanwhile, the UK National Health System V3 Clinical Terms project [30, 31], evolving from the Read Codes described below, had adopted a logic basis [32] similar to SNOMED RT basis. By late 1996, the major personalities behind SNOMED (Roger Côté) and the UK Read Codes (James Read) had, for very different reasons, withdrawn from controlling interest in their respective projects. Their withdrawal enabled the people actually working on the projects to consider partnership, recognizing the enormous overlap in content and method between them. One of us (CGC) acted as a neutral intermediary between the College of American Pathologists (CAP) in Chicago, IL, who owned SNOMED RT, and the NHS Centre for Coding and Classification in Leicestershire, UK, who managed the Clinical Terms V3 content on behalf of the UK NHS, ultimately brokering an agreement to merge. Thus was born the joint project, SNOMED Clinical Terms (CT), harmonizing the content between them into a union of concepts and relationships [33, 34].

Since 2002 SNOMED CT has been released twice a year. Initially, CAP owned and maintained SNOMED CT, but in 2007 ownership was transferred to a newly formed organization, the International Health Terminology Standards Development Organization (IHTSDO). This organization consists of regional members (predominantly countries) and currently has more than 25 members. In 2013 a study was published in which the researchers managed to determine seven production systems (of which one was deactivated) and three development / implementation projects [35]. This shows that the uptake of SNOMED CT in clinical practice is slow. This conclusion is supported by a literature review [36], which showed that the majority of published studies on SNOMED CT implementation has a theoretical or pre-development/design focus.

Read Codes

In 1983 James Read, a UK primary care physician, developed his own coding system, geared toward general practitioners. This system aimed at providing a compact way of capturing detailed clinical information, using 4-byte codes. It evolved into the comprehensive and much larger Read V2 in 1988 [37], after which it was acquired by the UK National Health Services (NHS) in 1990 becoming Crown copyright. Read 2 is still used through many parts of the UK.

While SNOMED was moving to SNOMED International in 1993, in the United Kingdom the clinical coding was moving to Clinical Terms version 3 [30, 31] based upon the Read codes and initially led by James Reed. Realizing that maintenance of two separate but similar terminology systems on both sides of the Atlantic is a time- and resource-intensive endeavor, negotiations started to merge Clinical Terms version 3 and SNOMED RT, resulting in SNOMED CT as described above.

LOINC

Whereas terminology systems such as SNOMED CT aim at enabling unique identification of detailed clinical information, LOINC (Logical Observation Identifiers, Names and Codes) [38] aims at providing a means of uniquely identifying the information elements in electronic health records. LOINC is remarkable for being the first completely open clinical terminology, making all content available without royalties or charges; this was driven by its creator Clem McDonald. For example, a serum sodium test will generate a numerical result, for which no terminology system is needed. But to uniquely identify the test name "serum sodium," LOINC code 2951-2 can be used. Each LOINC code contains its own nested, six-axis information model (e.g. sodium analyte, measured on serum, defined as a serum concentration, at a point in time, on a quantitative scale, etc.), and a logical structure to its creation of composite concepts such as 2951-2. It is often said that if SNOMED is the answer (result), then LOINC is the question (test name and code). More recently, Stan Huff has led the parallel effort to create Clinical LOINC, focusing on vital signs and related clinical measurements, such as height and weight.

Thesauri

Thesauri are neither classifications nor terminologies. In health informatics they are compilations of component classifications and terminologies indexed by shared con-

cepts among them. Concepts are distinct from lexical terms. For example, "heart attack" and "myocardial infarction" are distinct lexical terms, but are most often collapsed into the same logical concept. Thesauri often support multiple human languages, such as English or Dutch, and thus can sometimes function as a crude adjunct to language translation tasks. In this role, thesauri can be regarded as interface terminologies, providing close-to-user descriptions which are mapped to concepts in a reference terminology, which can in turn be mapped to classes in classifications (also called aggregation terminologies).

UMLS Metathesaurus

By far the largest and best curated biomedical thesaurus today is the Unified Medical Language Metathesaurus [39]. It includes 3.2M concepts, comprising 12.8M lexical terms from nearly 200 source terminologies and classifications. The Metathesaurus content is approximately 70% English language, but contains terms in 24 additional human languages. Approximately half of the source providers impose no intellectual property restrictions whatsoever, and most of the rest permit free use for research and development.

The UMLS dates back to 1988 with its first release [40, 41], and has been glibly credited with enabling 25 years of SCAMC/AMIA papers [42] more than anything else. Many naïve informatics researchers attempt to use the Metathesaurus as one large biomedical vocabulary, particularly as an ontology for entity recognition in natural language processing (NLP). While the Metathesaurus is undoubtedly the largest and richest thesaurus of concepts and synonyms in the world, and thus highly attractive for NLP use, it is by design not ontologically consistent. The UMLS purports to accept and combine terminology and classifications from its myriad sources without imposing editorial correction. Thus if one source asserts that A is the parent of B, while a second source asserts that B is the parent of A, the Metathesaurus will contain both relationships, knowingly introducing a logic-cycle conflict. This is a deliberate decision of the UMLS developers, and not an oversight. Nevertheless, these inconsistencies have profound implications for users who attempt to invoke the Metathesaurus as a well-formed ontology.

UMLS Specialist Lexicon

In the case of matching words and phrases to a lexical term, the presence of lexical and syntactic variants may impede success. Most of these variants can be predicted or asserted, which is exactly what the Specialist Lexicon [43] purports to do. Specifically, the Specialist Lexicon includes approximately 20,000 terms drawn from the UMLS Metathesaurus, Dorland Medical Dictionary, the most frequent general English terms drawn from The American Heritage Word Frequency Book, and words used in definitions from Longman's Dictionary of Contemporary English [44]. For each of these terms, the Lexicon asserts how to normalize a term into its root form as a function of syntactic variation due to verb tense, noun case, plural forms, gender, person, possession, and types of agreement, complement, and inflection, among other attributes. Further, the Lexicon indicates whether the term conforms to normal English rules of inflection, or is irregular, and if irregular, explicitly enumerates the grammatical variants.

The Specialist Lexicon is a unique resource, clearly designed to permit the normalization of grammatical and syntactic variants into a normalized form to facilitate dictionary matching to terms; it is a more sophisticated and accurate approach than the conventional alternative of word "stemming [45]" or truncation. As such, it is accompanied within the UMLS by a rich suite of tooling [46] that invokes the Lexicon to either normalize [47] the term or to generate specified variants [48].

Bioportal

The National Center for Biomedical Terminology (NCBO) [49] was established in 2005 by NIH as part of the Biomedical Information Science and Technology Initiative, a forerunner of the current Big Data to Knowledge (BD2K) program at NIH today. The NCBO had many activities, such as hosting major symposium dedicated to terminology and ontology. However, the most enduring and visible effort within the NCBO was the creation of BioPortal [50, 51] , a thesaurus of major biomedical terminologies and ontologies, with many associated tools and annotations. Today, BioPortal contains over 500 source terminologies, nearly 6M concepts, and links to many bioscience resources that have been annotated against domain appropriate subsets of BioPortal ontologies.

Information Models

Whereas terminologies and classifications have a long history, the field of information modeling is relatively young. In 1968 Larry Weed wrote [52]: "One solution is to orient data around each problem. Each medical record should have a complete list of all the patient's problems, including both clearly established diagnoses and all other unexplained findings that are not yet clear manifestations of a specific diagnosis, such as abnormal physical findings or symptoms." However, it took many years until efforts were undertaken to agree upon biomedical information models. HL7v3, based on a reference information model (RIM), wasn't published until 1997. Around the same time, work was undertaken on the development of the European (now international) standard CEN/ISO 13606 - Electronic health record communication (EHRcom). Both standards aim at describing a general structure for exchanging information, as well as specifications for the exact pieces of information that can be transferred in a given context, so called archetypes.

These archetypes (templates for clinical data element structure) specify, for example, that exchange of blood pressure information incorporates systolic and diastolic blood pressure, measured in mmHg, and the circumstances of the measurement, such as position of the patient, measurement method, cuff size, etc. Although archetypes are in principle implementation-independent pieces of information models, the existence of various similar standards required either a choice of implementation paradigm (e.g., HL7v3 or CEN/ISO 13606), or a specification for each of the standards. To overcome this burden, in 2007 specification

of Detailed Clinical Models started within ISO[53], aiming at providing a means of specifying archetypes in an implementation-independent fashion.

The Clinical Information Modeling Initiative (CIMI) [54] was started in 2011, with the aim to bring together stakeholders from all relevant standardization organizations, including ISO, CEN, HL7, OpenEHR, CDISC, the UK NHS, the US Department of Defense (DOD), the US Veterans' Administration (VA), and others to develop a specifications for the representation of health information content that facilitates creating and sharing semantically interoperable information in health records, messages, and documents. Members recognized that their work was highly overlapping, redundant, and at risk of siloed fragmentation of clinical archetypes. CIMI has largely been successful in coordinating compromise and harmonization among the major archetype developers in healthcare.

A popular complement to CIMI standards for information modeling is FHIR (Fast Healthcare Interoperability Resources) [55]. FHIR aims at providing a lightweight alternative for the extensive, hence complex, model-based approach of HL7v3. Its approach is based on the definition of "resources," which are similar to archetypes. CIMI and FHIR are highly complementary, and have mutually evolved to where CIMI provides additional specificity and bound value sets to the deliberately flexible FHIR resources.

A challenge that remains to be solved is the binding of information model and terminology. Similar to nomenclatures, where different composite codes could represent the same meaning, there is often no clear demarcation between what is represented in the information model and what is represented in the terminology. Post-coordination is a kind of simple information model; choosing pre- or post-coordination representation schema, which in the end are iso-semantic, illustrates the simplest case of this challenge. Efforts are underway to come to solutions, among others in the TermInfo-project that was initiated by HL7 in 2004, and through the alignment of LOINC and SNOMED CT, which has been catalyzed by US Meaningful Use [56, 57] mandates.

Biomedical Data Types

The notion of a data type in computer science is restricted to things like integer, floating point, or character. Health informatics has overloaded this term by adding another layer of aggregation atop traditional computer science data types, including items such as date-time, coded value (a lexical term, coded value, and code system), person name (parsing out first, last, and component elements), and address (with discrete components), or physical quantity (numeric value with a unit of measure.) These are microschemas above the level of computer science data types, but not at the level of clinical data elements or archetypes; archetypes are typically built from biomedical data types. There have been many standards for clinical data types, though there is a relatively direct lineage from two revisions of HL7 data types [58], to the ISO 21090 standard [59], to the more spare subset invoked for FHIR [60].

Applications

The ultimate goal of standards for terminologies and classifications is to achieve comparability and consistency of biomedical information that can sustain analysis and inferencing [61]. Absent these traits, clinical decision support rules cannot be linked to the data, and NLP algorithms have no semantic framework to extract and map information. It is neither practical nor reproducible to generate value sets that can be bound to archetypes or clinical models without clearly defined parent terminologies and classifications. The challenge of clinical data interoperability, where information can be exchanged between electronic health records, and understood by people and computers, depends fundamentally on the problem of syntactic and semantic consistency. Many national health information technology programs have declared sanctioned terminologies and classifications that will be used for specific use cases, such as ICD, LOINC, and SNOMED CT within the United States Meaningful Use specifications [56, 57]. There are corresponding efforts among most countries of the world. An interesting recent effort is the beginning of a cooperative process between the United States and the European Union, the Trillium Bridge project [62], to define a consensus specification that would facilitate trans-Atlantic clinical information exchange.

Knowledge Representation

In our anchoring year 1990, representation of terminological knowledge was predominantly realized using frame-based systems, which were introduced in 1975 by Minsky [63], who passed away in 2016. Frames intend to provide a means to capture relevant information, similar to archetypes, so a frame "student" enables capture of student id, programs in which the student is enrolled, courses the student takes, and grades obtained. A variety of systems and frame-based ontologies were developed in the 1980s and 1990s. LOOPS, developed by Xerox in 1983, was the first commercial frame system [64]. In 1984 the CYC project was initiated at Microelectronics and Computer Technology Corporation (MCC), intending to construct an ontology of common-sense knowledge, including medical knowledge. This work was evaluated and applied in part to healthcare by Cleveland Clinic in 2007 [65]. The LOOM system and knowledge representation language, introduced in 1987, provided an engine to perform classification for a frame-based language [66].

From the above description, it is clear that frames do not cater to providing definitions of concepts. E.g., what can be captured about a "student" is different from a formal definition of what makes someone a student. To fill this gap and to enable classification, or more formally subsumption testing, Ronald Brachman *et al*. introduced KL-ONE (Knowledge Language One), which can be considered the predecessor of contemporary description logic, in 1978 [67]. In the late 1980s, NIKL was developed as an implementation of KL-ONE [68]. These may be considered the first description-logic-like reasoners.

Description Logics

In 1991 Schmidt-Schauß and Schmolka introduced ALC [69], and analyzed the

complexity of subsumption and consistency testing. This has initiated extensive research on description logics with varying levels of expressivity, aiming at pushing the limits of complexity in relation to expressivity. From a mathematical perspective, the focus was put on determining, and where possible reducing, the computational (worst-case) complexity of more expressive description logics. In the 1990s this has been one of the main areas of research regarding description logics.

During the 1990s the GALEN and GALEN-IN-USE projects addressed the development of terminologies and terminology services for healthcare [70]. Within these projects, a formal language for representation of medical concepts, the GALEN Representation and Integration Language (GRAIL), was developed. The GRAIL language can be considered an inexpressive but useful language for medical concept representation. The development of GRAIL paralleled that of the Knowledge Representation System Specification (KRSS) [71], which was developed as a standard syntax for description logics, and of Ontylog, the language underlying the predecessor of SNOMED CT, SNOMED RT [72].

In the late 1990s the syntax of description logics got increasing attention from researchers in the semantic web area, and attempts were made to enable XML-based representation of description logic statements. Two prominent approaches were the DARPA Agent Markup Language (DAML), and the Ontology Interchange Language (OIL), which in the early 2000s merged as DAML+OIL [73]. In 2004 DAML+OIL was superseded by the Web Ontology Language OWL [74].

Around 2005 the elements were in place that have since gained increasing importance in the area of health concept management: the XML-based OWL syntax, the clear semantics of the various description logics, and software that performed inferences based on these logics. This combination has sparked research on systems to support modeling health concept systems and performing reasoning.

At Stanford since the mid-1980s work had been ongoing on the development of OPAL, Protégé, and Protégé-2000, various generations of frame-based ontology editors, based on the principle that "formal semantics are irrelevant [75]." This principle was evidently reconsidered when in 2004 an OWL-plugin for Protégé was developed. This ultimately led to Protégé version 3.5 (released in April 2013) which was frame-based, and Protégé 4+, providing native OWL-support.

Existence of very large owl-based ontologies inspired the computer science community to address reasoning with large ontologies with relatively inexpressive languages. Whereas until the early 2000's focus was on ever more expressive logics, being applied to very small ontologies. Around 2004 the interest in reasoning with large but inexpressive logics emerged, among others in the workshop on description logics and reasoning about patient data in Saarbrücken [76]. This, together with computers still keeping up with Moore's law, ultimately resulted in reasoners being able to perform classification of SNOMED CT in seconds instead of hours [77].

The feasibility of modeling and classifying large ontologies using OWL has led to the development of a wide range of biomedical ontologies. NCBO BioPortal currently contains over 300 OWL-based ontologies, of which currently 28 contain over 10,000 concepts [78].

Synthesis of Traditions

Information Models as Context

Not only have terminologies developed to adopt formal semantics, also information models have moved towards explicitly representing context. Whereas these formal representations have not yet reached implementation level, research in this area is ongoing. For example, ISO EN 13606 has been evaluated by means of an XML implementation [79], and LOINC concepts have been represented in OWL and merged with SNOMED CT [80]. The importance of agreeing upon representation of information models can also be seen by the work on the LOINC-SNOMED mapping, which provides, among others, and OWL-representation [81].

The Era of Big Data and Big Science

One of the big challenges remaining is the actual implementation of formalized information models and terminologies and the exploitation of the data that is represented using such information models and terminologies. The steps from using plain codes to taking into account hierarchical ordering, supporting and processing more complex expressions, and detecting iso-semantic representations (i.e., various ways of conveying the same information) are only the beginning of moving towards a true big-data approach, or rather, a linked-data approach, as is promoted in the semantic web community [82]. Furthermore, enormous amounts of legacy data exist that will benefit from post-hoc formal representation of context and contents using natural language processing [83].

Conclusions

The past twenty-five years have witnessed an ineffable impact of computing, computational capacity, and algorithmic sophistication on our notion of tractable classifications, terminologies, and knowledge systems. While not unimaginable 25 years ago, our present state of sophistication, specification, and scale was certainly unachievable. The pace of these transitions, as has been the case across most industries, is accelerating. For those of us who have experienced firsthand the challenges, achievements, and progress in biomedical semantics and knowledge organization over the past 25 years, we await the developments and potentially unimaginable achievements of the next 25.

References

1. Editors. What's In a Name? JAMA 1965;193(10):831-2.
2. Chute CG. Clinical classification and terminology: some history and current observations. J Am Med Inform Assoc 2000;7(3):298-303.
3. Baader F. The description logic handbook : theory, implementation, and applications. Cambridge, UK ; New York: Cambridge University Press; 2003.
4. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods

Inf Med 1998;37(4-5):394-403.
5. Graunt J, Petty W. Natural and political observations mentioned in a following index, and made upon the bills of mortality. London,: Printed by Tho. Roycroft, for John Martin, James Allestry, and Tho. Dicas; 1662.
6. Vladeck BC. Diagnostic-related groups. JAMA 1982;247(24):3314-5.
7. Chute C, Ustun B. ICD-11 Preview. In: Giannangelo K, ed. Healthcare Code Sets, Clinical Terminologies, and Classification Systems, 3rd Edition. Chicago: AHIMA; 2014:128-37.
8. Rodrigues JM, Robinson D, Della Mea V, Campbell J, Rector A, Schulz S, et al. Semantic Alignment between ICD-11 and SNOMED CT. Stud Health Technol Inform 2015;216:790-4.
9. Rodrigues JM, Schulz S, Rector A, Spackman K, Millar J, Campbell J, et al. ICD-11 and SNOMED CT Common Ontology: circulatory system. Stud Health Technol Inform 2014;205:1043-7.
10. Chang HY, Weiner JP. An in-depth assessment of a diagnosis-based risk adjustment model based on national health insurance claims: the application of the Johns Hopkins Adjusted Clinical Group case-mix system in Taiwan. BMC medicine 2010;8:7.
11. Haas LR, Takahashi PY, Shah ND, Stroebel RJ, Bernard ME, Finnie DM, et al. Risk-stratification methods for identifying patients for care coordination. Am J Manag Care 2013;19(9):725-32.
12. Meyboom-de Jong BM, Smith RJ. How do we classify functional status? Fam Med 1992;24(2):128-33.
13. Functional status measures in general practice. WONCA Classification Committee. Aust Fam Physician 1991;20(6):846, 848, 850-1.
14. World Health Organization. World Health Assembly. International classification of impairments, disabilities, and handicaps : a manual of classification relating to the consequences of disease. Geneva: World Health Organization; 1980.
15. World Health Organization. ICIDH-2 : International classification of impairments, activities and participation : a manual of dimensions of disablement and functioning : Beta-1 draft for field trials June 1997 : includes basic Beta-1 field trial forms. Geneva: The Organization; 1997.
16. World Health Organization. International classification of functioning, disability and health : ICF. Geneva: World Health Organization; 2001.
17. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;25(11):1251-5.
18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25(1):25-9.
19. Gene Ontology Consortium. Documentation. 2016; http://geneontology.org/page/documentation.
20. Cote RA. Ending the classification versus nomenclature controversy. Med Inform (Lond) 1983;8(1):1-4.
21. Côté RA, College of American Pathologists. Systematized nomenclature of medicine. 2d ed. Skokie, Ill.: The College; 1979.
22. Cote RA, Robboy S. Progress in medical information management. Systematized nomenclature of medicine (SNOMED). JAMA 1980;243(8):756-62.
23. Côté RA, College of American Pathologists. SNOMED International : the systematized nomenclature of human and veterinary medicine. 3rd ed. Northfield, Ill. Schaumberg, Ill.: College of American Pathologists ; American Veterinary Medical Association; 1993.
24. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. J Am Med Inform Assoc 1996;3(3):224-33.
25. Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. The Canon Group. J Am Med Inform Assoc 1994;1(3):207-17.
26. Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Galapagos: computer-based support for evolution of a convergent medical terminology. Proceedings : a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium. 1996:269-73.
27. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic-based convergent medical terminology. Methods Inf Med 1998;37(4-5):426-39.
28. Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED. Proceedings / AMIA ... Annual Symposium. AMIA Symposium. 2002:712-6.
29. World Wide Web Consortium (W3C). EL. 2007; https://www.w3.org/2007/OWL/wiki/EL.
30. Symmons DP, Dawes PT. The clinical terms project. Br J Rheumatol 1992;31(11):723-4.
31. Stannard CF. Clinical terms project: a coding system for clinicians. Br J Hosp Med 1994;52(1):46-8.
32. Brown PJ, Price C. Semantic based concept differential retrieval & equivalence detection in clinical terms version 3 (Read Codes). Proceedings / AMIA ... Annual Symposium. AMIA Symposium. 1999:27-31.
33. Wang AY, Barrett JW, Bentley T, Markwell D, Price C, Spackman KA, et al. Mapping between SNOMED RT and Clinical terms version 3: a key component of the SNOMED CT development process. Proceedings / AMIA ... Annual Symposium. AMIA Symposium. 2001:741-5.
34. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. Proceedings / AMIA ... Annual Symposium. AMIA Symposium. 2001:662-6.
35. Lee D, Cornet R, Lau F, de Keizer N. A survey of SNOMED CT implementations. J Biomed Inform 2013;46(1):87-96.
36. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. J Am Med Inform Assoc 2014;21(e1):e11-19.
37. Green LA. Read Codes: a tool for automated medical records. J Fam Pract 1992;34(5):633-4.
38. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clin Chem 1996;42(1):81-90.
39. National Library of Medicine. Unified Medical Language System® (UMLS®): Metathesaurus. 2016; https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/.
40. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-91.
41. Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc 1993;81(2):170-7.
42. Lindberg DA, Humphreys BL. "You have to be there": twenty-five years of SCAMC/AMIA symposia. J Am Med Inform Assoc 2002;9(4):332-45.
43. National Library of Medicine. Unified Medical Language System® (UMLS®): SPECIALIST Lexicon. 2012; https://www.nlm.nih.gov/pubs/factsheets/umlslex.html.
44. McCray AT, Srinivasan S. Automated access to a large medical dictionary: online assistance for research and application in natural language processing. Comput Biomed Res 1990;23(2):179-98.
45. Wikipedia. Stemming. 2016; https://en.wikipedia.org/wiki/Stemming.
46. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care 1994:235-9.
47. National Library of Medicine. The NORM program. 2014; https://www.nlm.nih.gov/research/umls/new_users/online_learning/LEX_005.html.
48. National Library of Medicine. Lexical Variant Generation (LVG). 2014; https://www.nlm.nih.gov/research/umls/new_users/online_learning/LEX_004.html.
49. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, et al. The National Center for Biomedical Ontology. J Am Med Inform Assoc 2012;19(2):190-5.
50. National Center for Biomedical Ontology. BioPortal. 2016; http://bioportal.bioontology.org/.
51. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 2009;37(Web Server issue):W170-173.
52. Weed LL. Medical Records That Guide and Teach. New Engl J Med 1968;278(11):593-600.
53. ISO. Health informatics -- Detailed clinical models, characteristics and processes. 2015; http://www.iso.org/iso/catalogue_detail.htm?csnumber=62416.
54. Clinical Information Modeling Initiative. CIMI. 2015; http://opencimi.org/b.
55. HL7. Fast Healthcare Interoperability Resources (FHIR). 2015; http://www.hl7.org/fhir/.
56. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363(6):501-4.
57. Office of the National Coordinator for Health Information Technology Interoperability. Interoperability Standards Advisory (ISA). 2016; https://www.healthit.gov/standards-advisory.
58. Health Level Seven. HL7 Version 3 Standard: Data Types - Abstract Specification, Release 2. 2016; http://www.hl7.org/implement/standards/product_brief.cfm?product_id=264.

59. ISO. Health informatics -- Harmonized data types for information interchange. 2011; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35646.
60. Health Level Seven. Value Set Codes for FHIR Datatypes. 2016; https://www.hl7.org/fhir/DSTU1/data-types.html.
61. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-Based Patient Records Institute Working Group on Codes and Structures. J Am Med Inform Assoc 1998;5(6):503-10.
62. HL7 and the European Commission. Trillium Bridge. 2016; http://www.trilliumbridge.eu/.
63. Minsky M. A framework for representing knowledge. In: Winston P, editor. The Psychology of Computer Vision. New York: McGraw-Hill; 1975. P. 211-77.
64. Stefik M, Bobrow DG, Mittal S, Conway L. Knowledge programming in LOOPS. The AI Magazine 1983:3-13.
65. Ogbuji C, Blackstone E, Pierce C. Case Study: A Semantic Web Content Repository for Clinical Research. 2007; https://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/. Accessed 2016-02-20, 2016.
66. Schulz S, Hahn U. Medical knowledge reengineering - converting major portions of the UMLS into a terminological knowledge base. Int J Med Inform 2001;64(2-3):207-21.
67. Woods W. The KL-ONE Family. Computers & Mathematics with Applications. 1992;23:133-77.
68. Haimowits IJ, Patil RS, Szolovits P. Representing Medical Knowledge in a Terminological Language is Difficult. Paper presented at: Twelfth Symposium on Computer Applications in Medical Care1988; Los Angeles.
69. Schmidt-Schauß M, Smolka G. Attributive concept descriptions with complements. Artif Intell 1991;48(1):1-26.
70. Rector AL, Nowlan WA. The GALEN project. Comput Methods Programs Biomed 1994;45(1-2):75-8.
71. Patel-Schneider P, Swartout B. Description-Logic Knowledge Representation System Specification from the KRSS Group of the ARPA Knowledge Sharing Effort. KRSS Group of the ARPA Knowledge Sharing Effort; 1 november 1993; 1993.
72. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. Paper presented at: Proceedings of the 1997 AMIA Annual Fall Symposium1997; Nashville, TN, USA.
73. Wroe C, Stevens R, Goble C, Ashburner M. An Evolutionary Methodology To Migrate The Gene Ontology To A Description Logic Environment Using DAML+OIL. Paper presented at: Pro 8th Pacific Symposium on Biocomputing (PSB)2003; Hawaii.
74. Patel-Schneider PF, Hayes P, Horrocks I. OWL Web Ontology Language Semantics and Abstract Syntax. 2004; https://www.w3.org/TR/2004/REC-owl-semantics-20040210/syntax.html. Accessed 2016-02-20, 2016.
75. Grosso WE, Eriksson H, Fergerson RW, Gennari JH, Tu SW, Musen MA. Knowledge Modeling at the Millennium (The Design and Evolution of Protégé 2000). Paper presented at: 12th International Workshop on Knowledge Acquisition, Modeling and Management (KAW'99)1999; Banff, Canada.
76. Smith B. Workshop on description logics and reasoning about patient data. 2004; http://ontology.buffalo.edu/04/DLs_and_Medicine/. Accessed 2016-02-20, 2016.
77. Dentler K, Cornet R, Teije At, Keizer Nd. Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. Semantic Web 2011;2(2):71-87.
78. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011;39(Web Server issue):W541-545.
79. Austin T, Sun S, Hassan T, Kalra D. Evaluation of ISO EN 13606 as a result of its implementation in XML. Health Informatics J 2013;19(4):264-80.
80. Adamusiak T, Bodenreider O. Quality assurance in LOINC using Description Logic. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium. 2012;2012:1099-108.
81. Vreeman D. Draft LOINC-SNOMED CT Mappings and Expression Associations Now Available. 2014; https://loinc.org/news/draft-loinc-snomed-ct-mappings-and-expression-associations-now-available.html/. Accessed 2016-02-20.
82. Janowicz K, Hitzler P, Adams B, Kolas D, II CV. Five Stars of Linked Data Vocabulary Use. Semantic Web Journal 2014;5(3):173-6.
83. Zeng Z, Shi H, Wu Y, Hong Z. Survey of Natural Language Processing Techniques in Bioinformatics. Comput Math Methods Med 2015;2015:674296.

Correspondence to:
Ronald Cornet, PhD
Visiting Associate Professor, Linköping University
Assistant Professor, Academisch Medisch Centrum
Medical Informatics, J1b-115
P.O. Box 22700
1100 DE Amsterdam
The Netherlands
E-Mail: r.cornet@amc.uva.nl

Christopher G. Chute, MD DrPH
Bloomberg Distinguished Professor of Health Informatics
Professor of Medicine, Public Health, and Nursing
Chief Research Information Officer, Johns Hopkins Medicine
Johns Hopkins University, Division of General Internal Medicine
2024 E Monument St, Suite 1-200
Baltimore, MD 21287
USA
E-Mail: chute@jhu.edu