

The Practical Impact of Ontologies on Biomedical Informatics

J. J. Cimino¹, X. Zhu^{1,2}

¹Department of Biomedical Informatics, Columbia University College of Physicians and Surgeons, New York, New York, USA

²Department of Biomedical Informatics, Philips Research North America, Briarcliff Manor, New York, USA

Summary

Objectives: To examine recent research work in the development and evaluation of controlled biomedical terminologies – especially, the representation of structured, controlled definitional knowledge about the terms themselves; such terminologies are often referred to as „ontologies“.

Methods: A review of the published literature using PubMed, as well as full-text searches of recent Medinfo and American Medical Informatics Association (AMIA) Symposia proceedings, searching for the terms “ontology” and “ontologies” and for articles discussing specific, prominent ontological work.

Results: We summarize the ontologic aspects of twelve current terminology projects: Galen, the Unified Medical Language System (UMLS), the Medical Entities Dictionary (MED), SNOMED-CT, LOINC, the Foundational Model of Anatomy (FMA), the Gene Ontology (GO), ISO Reference Terminology Model for Nursing Diagnosis, NDF-RT, RxNorm, the NCI Thesaurus, and DOLCE+. We discuss the origins, domain, and ontologic representation of each of these and attempt to summarize the impact that each has had on terminologic work and biomedical applications. We also note the contributions of the Protégé tool to many of these efforts.

Conclusion: Terminologic research and development have advanced significantly in the past 20 years, especially since the recent orientation toward controlled biomedical ontologies. This work has had significant impact on the development of terminologies themselves, their acceptance and dissemination as standards, and their use in supporting biomedical information systems.

Haux R, Kulikowski C, editors. *IMIA Yearbook of Medical Informatics 2006. Methods Inf Med 2006; 45 Suppl 1: Sxx-xx*

Keywords

Controlled Terminology, Ontology, Unified Medical Language System

1 Introduction

Controlled terminologies for health care have been in existence for centuries [1]. The first standard terminology, the International Classification of Diseases, can trace its lineage back at least to 1893. For most of that time, terminologies have been treated as a necessary element of health data collection, but the only attention paid to terminology development was, usually, to make sure that there were enough terms to cover the intended domain. Terminologies ranged, for example, from a set of options on multiple-choice surveys (e.g., gender and race), to long lists of causes of mortality and morbidity. In many cases, the controlled terminologies that were developed for use in applications such as clinical information systems, medical expert systems, and research databases seemed to barely deserve public comment.

Interest in the terminologies themselves grew slowly. Prior to 1966, articles containing the words “terminology” or “terminologies” numbered about 25 per year. In 1966, the National Library of Medicine added “Medical Nomenclature” to the Medical Subject Headings (MeSH – their list of terms used to index the biomedical literature); articles indexed with that term jumped to around 500 a year and continued to increase at about 4.5% per

year*, compared to an annual growth in the Medline database of about 3.9%, as shown in Figure 1.

Ten years ago, a review published in this Yearbook [2] identified a dozen or so important controlled terminologies. Each terminology was characterized as an abstraction terminology (intended for high-level indexing of medical records) or as a coding terminology (intended for capturing the details in such records). A brief survey of research issues was provided as well, based on the published biomedical literature of the time. A more in-depth review of research issues was published subsequently [3].

Interest in terminologies has continued to increase since then. New methods for representing terminologies have been explored, including those that make them more usable by computers. In particular, terminologies have been evolving to include knowledge about their terms, especially definitional knowledge. While this knowledge has taken many forms, there is a growing interest in representing the knowledge along the precepts of a branch of philosophy known as “ontology” (see Figure 1). Terminologies that follow these precepts are themselves often called “ontologies” [4, 5].

*In 1999, the index term was changed to „Terminology“

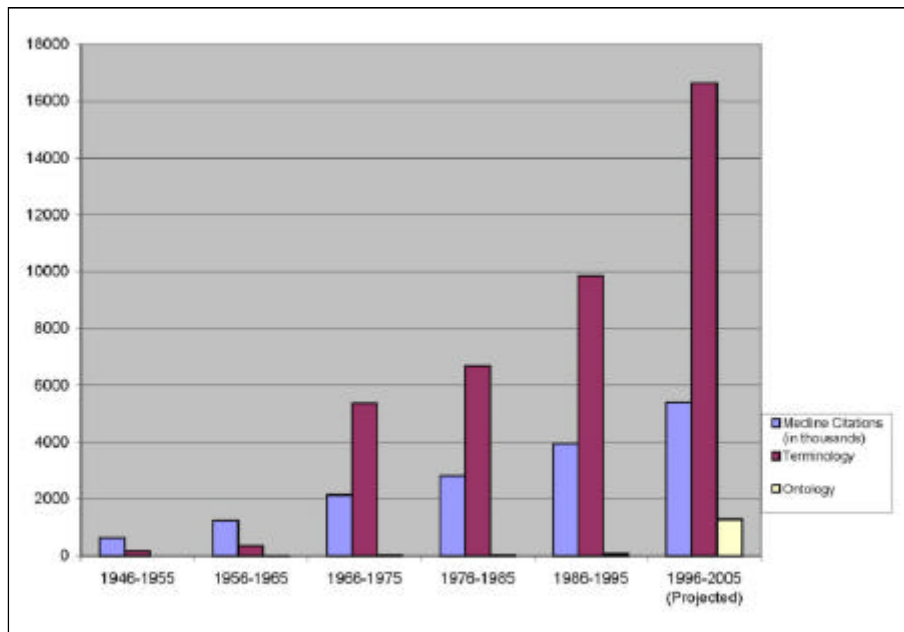


Fig. 1 Numbers of citations (1) in the Medline database overall (blue bars), (2) indexed with the MeSH term "Terminology" (or its predecessor "Medical Nomenclature"), or any more specific MeSH terms (or, prior to 1966, containing the word "terminology" or "terminologies"), and (3) the word "ontology" or "ontologies".

While ontologies have been constructed for many biomedical applications, it is their relationships to terminologies where they are currently having the most visible impact. The purpose of this paper is review some of the current work in the ontological approach to controlled biomedical terminologies. Particular attention is given to work that has resulted in practical impact on the terminology development or on the use of terminologies in health care.

2. Defining Ontology and Impact

The term "ontology" has long been used to refer to a branch of philosophy that deals with the study of being. When used in the context of knowledge representation, however, an "ontology" is "a formal specification of a conceptualization" [6]. In the biomedical informatics field, ontologies have been used for representing a variety of knowledge

bases; this paper focuses on terminology knowledge bases – that is, repositories of knowledge about the meanings of terms in terminologies. An alternate perspective is that ontologies are terminologies that contain some formal representation of definitional information.

The above definition covers a wide range of current work in the field. At one end of the spectrum are terminologies that are little more than manually created hierarchical arrangements of terms whose developers nevertheless consider them to be ontologies. At the other end of the spectrum are terminologies that contain sufficient knowledge that their hierarchies can be automatically inferred. Almost every important biomedical terminology fits somewhere in between these extremes. We are similarly inclusive in considering the form in which terminologies represent their ontological information. Common representations include frames, semantic networks, conceptual graphs and description logic. In some cases these

representations are interchangeable, while in other cases the names of the terms have a formal structure that conveys the ontologic information. A comprehensive review of biomedical ontologies is beyond the scope of this paper, especially given our broad definition. Therefore, our focus is on recent work that has produced, or has the potential to produce, some impact beyond the internal creation of the ontology itself. We consider ways in which the ontological representation of terminologies has impacted biomedical research and patient care, as well as ways in which advances in representation have influenced other work on knowledge representation.

3. Ontologies and Their Impact

3.1 GALEN

One of the most venerable ontology projects is GALEN (General Architecture for Languages, Encyclopedias and Nomenclatures in Medicine), which evolved from the knowledge-based terminology of Alan Rector's Pen&Pad electronic medical record system [7, 8]. Pen&Pad's terminology was represented using a formalism called Structured Meta Knowledge (SMK), in which terms were defined through relationships to other terms, and grammars were provided to allow combinations of terms into sensible statement.

A consortium of European universities, agencies, and vendors formed the GALEN project to develop standards for representing coded patient information [9]. GALEN is developing a reference model for representing medical concepts independently of the language being recorded and of the data model used by an electronic medical record

system. GALEN is working in cooperation with the Comité Européen de Normalisation Technical Committee 251 (CEN TC251) to develop terminologies using the reference model.

In 2000, an open source foundation called OpenGALEN was established to distribute the reference model free of charge and work with software vendors and terminology developers to support its extension and use [10]. Since then, the Galen model has been used to study nursing terminologies [11, 12], a pain terminology [13], decision support knowledge [14], surgical procedures [15], and anatomy [16], including the Foundational Model of Anatomy (see below) [17, 18].

3.2 Unified Medical Language System

The Unified Medical Language System (UMLS) was created by the US National Library of Medicine, under the leadership of Donald Lindberg, Betsy Humphreys, and Alexa McCray, as a kind of meta-terminology, subsuming the contents of other terminologies. It has grown to include over 100 terminologies, 1 million concepts, and 4 million names for those concepts [19]. The UMLS model identifies terminologic entities at three levels: the string (any name for a term in a terminology), the lexical group (to which strings of identical or near-identical lexical structure can be mapped) and the concept (to which strings of identical meaning can be mapped). Through this model, the UMLS attempts to merge terminologies such that synonymous terms from different terminologies map to identical concepts, while still retaining as much of each terminology's original information as possible.

Three major types of ontologic information are included in the UMLS. The

first is the UMLS Semantic Network, consisting of 135 semantic types, organized into a pair of non-overlapping strict hierarchies with 6,864 relations among them. The relations do not represent knowledge about how the semantic types relate directly to each other but rather, which types of concepts in the UMLS Metathesaurus can have the relations between them. The second type of knowledge is the assignment of semantic types to Metathesaurus concepts; that is, the concepts can be seen as being instances of the semantic types. Concepts are each assigned at least one semantic type and some have as many as 5 types assigned. The third type of knowledge is represented by actual instances of relations between actual concepts.

For example, the semantic type Disease or Syndrome has 22 types of relations to 109 other semantic types (a total of 392 unique relations), such as Disease or Syndrome *affects* Organism and Therapeutic or Preventive Procedure *treats* Disease or Syndrome. A total of 66,916 concepts in the Metathesaurus are assigned the semantic type Disease or Syndrome. One of these concepts is Pneumococcal Pneumonia, which is only assigned this one semantic type. Pneumococcal Pneumonia, in turn is related to 364 other concepts, including concept Lung, via the relation *has_location* and the concept Inflammation with the relation *has_associated_morphology*, and Streptococcus pneumoniae with the relation *has_causative_agent*.

Unlike other ontologies, the knowledge represented in the UMLS is not particularly definitional, nor is it particularly complete. It is merely the collection of knowledge obtained from the source terminologies. So, for example, most of the 364 terms associated with Pneumococcal Pneumonia are unrelated to

its definition, and there is no information in the UMLS that identifies which organisms are affected by pneumococcal pneumonia (e.g., Human), nor what treats it (e.g., Penicillin).

The UMLS has probably had greater impact on biomedical ontology work than any other terminology effort. This can be attributed to its long history, its early focus on knowledge representation, and its free availability. A review of all the work based on the UMLS is beyond the scope of this paper [20], but of the over 80 papers in the biomedical literature in 2005 alone, the topics have included identification of patient problems [21], clinical events [22], and patient identifiers [23] in electronic medical records, knowledge base construction [24, 25], terminology management [26], and representation of consumer information needs [27].

3.3 The Medical Entities Dictionary

The Medical Entities Dictionary (MED) is the controlled terminology developed by one of us (JJC) at Columbia University to provide a single unified coding system for data collected in the repository of the New York Presbyterian Hospital (NYPH) clinical information system. Begun in 1988, it is, in effect, a "local UMLS" that integrates terminologies from a variety of applications and ancillary departmental systems [28]. It is organized into a semantic network and, while it cannot properly claim to be an ontology [29], it contains ontological information for some of its constituent terminologies. For example, laboratory test terms are related to the substances they measure and the specimens in which they measure them, much as in LOINC (see below). Specimen terms, in turn, are related to anatomical terms representing the body parts sampled in the specimens. Simi-

larly, medication terms are related to terms representing the medication ingredients and drug physical forms (tablets, liquids, etc.).

The main purpose of the MED is as a coding system for data originating in various systems. However, the knowledge in the MED has been used to support a variety of functions, including summary reporting, automated decision support, terminology translation, and linking online health knowledge resources to clinical applications, including clinical applications at other institutions [30]. More recently, the model and content of the MED have been used for representing nursing diagnostic concepts [31], ophthalmology terms [32], and the Outcome and Assessment Information Set (OASIS) for home health care [33].

3.4 SNOMED-CT

SNOMED-CT is the result of a merger between two terminologies that each contained ontological information: SNOMED-RT (from Roger Côté and David Rothwell, at the College of American Pathologist, in the United States) and the Read Clinical Terms (from James Read, Michael O'Neill and Colin Price, at the National Health Service in the United Kingdom). Each of these terminologies has a long history of developing from simple lists of terms, through hierarchical representations, and finally (in the past 10 years or so) to include definitional knowledge about a subset of their terms; their merger, led by Kent Spackman, has brought the term sets and their knowledge together into a single ontology [34].

Informatics researchers have postulated that formal definitional information, such as that available in SNOMED-RT and Read, should be useful for

terminology management tasks; however, the initial merger between Read and SNOMED focused only on lexical and hierarchical information to identify synonymy and is-a relationships between concepts in the two terminologies [35]. Subsequently, the developers reached agreement on a joint terminology model, which then facilitated improved analysis of the merged product and subsequent improvements in content [36]. This model, and corresponding knowledge in SNOMED-CT, is represented using description logic (DL). For example, "Pulmonary Tularemia" is expressed in DL as:

```
Pulmonary
Tularemia ≡ Tularemia ∩
    Pneumonia due to aerobic bacteria ∩
    ∃Causative Agent.Francisella tularensis ∩
    ∃Associated morphology.Consolidation ∩
    ∃Associated morphology.Inflammation ∩
    ∃Finding site.Structure of interstitial tissue
    of lung
```

While SNOMED-CT's ontologic approach has had impact on its internal creation and maintenance, the ontologic information itself has not been widely used by users of SNOMED-CT or for reconciling SNOMED-CT with other, external terminologies. Other researchers are beginning to exploit SNOMED-CT's ontologic knowledge for their own terminologic work [37, 38]. However, SNOMED-CT's DL was not found useful when integrating SNOMED-CT with the Unified Medical Language System (UMLS) [39, 40]. Interest in SNOMED-CT has increased dramatically in recent years. This is in part because of its availability free of charge in the US, but it is also viewed as being a comprehensive, high-quality terminology, especially by advisory bodies such as the US National Committee for Vital and Health Statistics (NCVHS) [41].

SNOMED-CT has had significant impact through the stimulation of dialogue about the coding of clinical data. As familiarity with it increases, and pressure to standardize patient information is brought to bear, SNOMED-CT is in a position to have tremendous impact on health information systems and health care in general.

3.5 LOINC

Developed by a team led by Clem McDonald and Stan Huff as an ad hoc standard for coding clinical observations in HL7 messages, the ontological approach taken in the development of the Logical Observations, Identifiers, Names and Codes (LOINC) contrasts in several ways with the SNOMED-CT approach. While SNOMED-CT's representational model was developed based on decades worth of terms that it had accumulated from many different domains, LOINC started as a terminology model for a very limited domain and then accumulated terms that were described by the model [42]. Also, while SNOMED-CT openly embraced an ontological approach, the LOINC developers avoided explicit statements about formal definitional knowledge. The knowledge in LOINC is expressed through the use of a structured naming system, in which each name consists of 5 to 6 parts, with each part consisting of a term from a limited, controlled terminology. For example, a typical LOINC term name for a fingerstick glucose test is:

```
GLUCOSE | SCNC | PT | BLDC | QN
| GLUCOMETER
```

Each portion of the name corresponds to a specific property of the test, with the value of the property drawn from specific controlled terminologies. In

this example, the first part of the name, called the “component” of the test (the substance whose property is being measured), is Glucose, one of 16,000 allowable component values in LOINC. The fifth part of the name, “QN”, is one of the 12 allowable values for the test “scale”, in this case indicating that the test reports its results as a numeric (quantitative) value. It is structured, controlled definitional information such as this that qualifies LOINC as an ontology.

During its initial development, researchers used the logical structure of LOINC terms as a way to translate local terms into LOINC for the purpose of sharing patient data [43]. More recently, the analyte (or “component”) attribute in LOINC terms has been used to map laboratory test terms to the Medical Subject Headings (MeSH), in order to automate literature retrieval [44]. The LOINC model has also facilitated identifying LOINC codes for local laboratory test terms [45], merging LOINC terms with other terminologies [46], and guidelines [47, 48], as well as mapping other terminologies for nursing assessments [33] and home health care [49] into LOINC. However, the major impact of LOINC has been in its growing adoption as a standard terminology; like SNOMED-CT, the underlying ontologic foundation may be contributing to LOINC’s appeal and it has been similarly endorsed by the NCVHS. However, LOINC’s formal definitions are not being fully exploited in their own right.

3.6 FMA

Developed and maintained by Cornelius Rosse at the University of Washington, the Foundational Model of Anatomy (FMA) is a frame-based domain

ontology that represents declarative knowledge about human anatomy [50]. The FMA was specifically developed to provide concepts and relationships pertaining to human anatomical structures, with the intent of expanding the anatomical content of UMLS. The content of the FMA ranges from concepts at the molecular level up to macroscopic body parts, surfaces and spaces. The FMA consists of over 70,000 concepts and 110,000 anatomical terms. Its 168 relationship types yield over 1.5 million relations between its concepts [51, 52].

The content development of FMA flows logically from the representation process. As an evolving knowledge source, it is concerned with the representation of classes and relationships necessary for the symbolic modeling of the structure of the human body in a form that is understandable to humans and is also navigable by computer systems [53]. Instead of designing a terminology model to meet a particular purpose, three models were included in the FMA: an ontological model for representing classes of anatomical structures, a structural model for representing spatial and topological relationships, and a transformational model for representing morphological changes such as those that occur with development and aging [52].

The concepts included in the FMA are not intended to cover some external set of terms but rather the internal set of terms required to create the frame-based definitions of its own terminology. For instance, the definitions of many body parts cannot be accomplished precisely without characterizing the extent of the body part, establishing the other body parts to which it is adjacent, and defining the anatomical structures comprising the body part. Therefore, each definition

of an anatomical concept begets additional concepts that must, in turn, be defined by other structures [50].

For example, “Esophagus” is expressed in DL, by Shapiro and colleagues [54], as:

```
Name: Esophagus
Has physical state: Solid
Bounded by: External surface of esophagus
Part of: Upper gastrointestinal tract
Member of: Set of viscera
Contained in: Space of anterior compartment of neck
Adjacent to: Trachea, thoracic aorta, thoracic vertebral column
Continuous with: Pharynx, stomach
```

The hierarchical position of Esophagus is:

```
Anatomical entity
  Physical anatomical entity
    Material physical anatomical entity
      Anatomical structure
        Body
          Organ
            Cavitated organ
              Organ with organ cavity
                Esophagus
```

In addition to its frame-based representation, FMA has been represented with DL, using the Web Ontology Language (OWL) [55, 56]. OWL is a specification that is managed by the World Wide Web consortium, and which has gained a large following among the Semantic Web and ontology-development communities. One particular flavor of OWL, OWL-DL (Description Logic), captures all but a few of the syntactical and semantic features in the most widely used DL. By representing FMA in DL, researchers have been able to take advantage of generic reasoning tools [57, 58]. This has led to experiments with reasoning about the anatomical aspects of diseases (such as penetrating heart injury), using the classes and part-of relationships from the key FMA anatomic and physiological ontologies [59].

However, representing the frames of FMA in OWL illuminated some of the differences between description logics

and frames [57]. For example, some of the constructs in FMA could not be directly translated. In addition, a complete conversion of the FMA into DL overwhelmed the computational abilities of available reasoning systems. One solution has been the use of OWLFull, a superlanguage of OWL-DL, to create an intermediate representation of FMA. A case study has shown that this representation could be computed on a concept-by-concept basis, thus reducing the burden of an integral conversion [57].

The FMA has attracted some attention, as it addresses the need for high-quality anatomical terms in ways that were not satisfied by more general-purpose terminologies, such as SNOMED-CT. One example is the Biolucida system, which uses the FMA to deliver 3D virtual representations of human anatomical structures. The system allows an expert anatomist-author to create knowledge-based, customized, and interactive scenes and lessons for students of human macroscopic anatomy. The FMA knowledgebase facilitates content authoring, information presentation, automated student assessments, an injury propagation modeling environment, and a surgery simulator [60].

3.7 Gene Ontology

The Gene Ontology (GO) is a controlled biological terminology being created by a consortium of bioinformaticians, lead by Michael Ashburner and Judith Blake, involved in genomics and proteomics [61]. Although a relative newcomer, GO has helped popularize the term “ontology” more than any other effort, accounting for 40% of Medline citations containing the term in the past 6 years. This is due in part to its use of the word in its name,

but it is also due to the fact that it attempts to fill a controlled terminology void in the rapidly emerging field of bioinformatics.

Despite its name, GO is a fairly minimal ontology, in that it provides little in the way of logical representation of knowledge about the terms it contains. The GO Consortium has focused instead on the task of creating and agreeing on the “semantic concepts” of its domain [62]. GO started with terminologies from three genomic databases (Flybase, the Saccharomyces Genome Database and the Mouse Genome Database) and developed three hierarchies (or “name spaces”) of terms to describe biological processes, cellular components and molecular functions, respectively. Terms are related to each other within each hierarchy by is-a and part-of relationships. Annotations for the terms include original definitions and comments from the GO curators. For example:

```
id: GO:0043234 name: protein complex
namespace: cellular_component
def: "Any protein group composed of two or more subunits, which may or may not be identical. Protein complexes may have other associated non-protein prosthetic groups, such as nucleic acids, metal ions or carbohydrate groups."
GO:curators comment: "Note that although at some level almost all cellular components can be thought of as protein complexes, this term is intended to exclude structures composed of the same repeating subunit or subunits, for example microtubules. Protein complexes encompassed by this term are generally not structural, and usually have a defined set of subunits."
is_a: GO:0005575: cellular_component
```

```
id: GO:0000214 name: tRNA-intron endonuclease complex
namespace: cellular_component
def: "Catalysis of the endonucleolytic cleavage of pre-tRNA, producing 5'-hydroxyl and 2',3'-cyclic phosphate termini, and specifically removing the intron." [EC:3.1.27.9]
is_a: GO:0043234: protein complex
relationship: part_of GO:0005634: nucleus
```

A sample of the GO hierarchy is:

```
GO:0003673 gene_ontology
GO:0003674 molecular_function
GO:0005575 cellular_component
GO:0008150 biological_process
GO:0007582 physiological processes
GO:0016265 death
GO:0006915 apoptosis
GO:0006916 anti-apoptosis
GO:0006917 induction of apoptosis
GO:0008220 necrosis
GO:0009626 hypersensitive response
GO:0019835 cytolysis
...
```

The GO Consortium is working to make associations between the ontologies and the genes and gene products in the collaborating databases. In addition to providing the ontology itself, the Consortium also develops tools that facilitate its maintenance and use. GO's impact is evident in the published literature: when terminologies for genomic and proteomic data are discussed today, they are either represented in GO or in a terminology that is related to GO.

3.8 ISO Reference Terminology Model for Nursing Diagnosis

The nursing informatics community has been particularly active with respect to terminology development, with over twelve terminologies just for the domains of assessments, interventions, goals and outcomes [63]. Although the terminologies themselves often lacked formal models for representing meanings, a group of loosely affiliated nurse informaticians came together to create a model for the domain of nursing interventions [64]. This so-called “Loose Canon” model became the basis for the development, under the leadership of Virginia Saba and Suzanne Bakken, of an ISO standard for interventions (or “actions”) [65]. The model has since been expanded to include nursing assessments (or diagnoses) [66].

The ISO model for nursing actions contains six attributes (site, route, means, target, recipient of care, and timing), while the diagnosis model contains five (dimension, focus, judgment, site, and subject) that, in turn have attributes of their own. For example, the term “Spiritual Distress: Actual” is modeled as [31]:

Focus: Spiritual State
 Judgment: Distress
 Potentiality: Actual
 Subject of Information: Individual

Subsequent to its development, a number of studies have shown that the ISO model can be applied successfully to represent the various extant nursing terminologies [67]. The original intent of the model was to support the ability to translate terms between the various terminologies, and this has been partially borne out [65]. The model itself has also been used to support natural language processing [68] and as a model for developing new terminologies [69].

3.9 NDF-RT

The Veterans Health Administration (VHA) branch of the US Department of Veterans Affairs, as part of its Health Data Repository (HDR) project, has computerized a variety of clinical transactions, including physician orders and documentation [70]. VHA’s initial reference terminology project was National Drug File Reference Terminology (NDF-RT), a formalization of the original VHA National Drug File, led by Steven Brown and Michael Lincoln [71, 72]. The core of the NDF-RT model comes from the existing VHA database file, including approximately 400 VHA drug classes. The classes are legacy classes developed by the VA, informed by the US Food and Drug

Administration’s approved labeling [73]. After the VHA developed the NDF for coding drug terms in its clinical information systems, the NDF was expanded into a formal reference terminology (NDF-RT) that is being made freely available. NDF-RT is different from many existing drug coding systems in that drugs may appear in multiple classes; for example, the same drug can be classified as both an antihypertensive and a beta-blocker. NDF-RT uses a DL-based reference model, which includes a specified set of abstractions denoting levels of description for drug products and a set of hierarchical relationships and definitions. The model includes hierarchies for chemical structure, mechanism of action, physiologic effect, and therapeutic intent. The DL enables drug terms to inherit properties, such as their active ingredients, from their superclasses [74]. NDF-RT also has links to external terminologies, such as MeSH and LOINC. The representation of an example term in NDF-RT is as the follows:

Name: Atenolol preparation
 Kind: DRUG_KIND
 Code: C11290
 Pharmacologic_Action: Sympatholytics
 ID: 11291
 Some has_ingredient: Atenolol
 Some has_MoA: Adrenergic beta1-Antagonists
 Some has_PE: Negative Chronotropy
 Some has_PK: Renal Excretion
 Some may_prevent: Migraine
 Some may_treat: Angina Pectoris
 Some may_treat: Atrial Fibrillation
 Some may_treat: Atrial Flutter
 Some may_treat: Heart Failure, Congestive
 Some may_treat: Hypertension
 Some may_treat: Myocardial Infarction
 Some may_treat: Tachycardia, Ventricular

One of the VHA’s concerns is scalability of terminology development, including distribution, collaborative term definition, and speed of classifier algorithm performance. In the NDF-RT drug

domain, the highest rate of change is the level of specific actual products, at approximately 4000 updates per month. The goal is for a large portion of routine maintenance (e.g. adding a drug with a defined set of ingredients) to be automated via a series of electronic transactions, although it is expected that updates requiring changes to underlying reference taxonomies will get closer supervision from subject matter experts [75]. NDF-RT is exerting its impact largely through its contribution to the RxNorm project (see below).

3.10 RxNorm

RxNorm, developed by Stuart Nelson at the National Library of Medicine, is a standardized nomenclature for clinical drugs that addresses both the lack of an adequate standard for a national terminology for medications and the need for formal organization of medication terms within the UMLS [76]. RxNorm started with a terminology model, originally developed by the HL7 Vocabulary Technical committee, which was based on models used by four vendors of drug knowledge bases [77]. The model includes representation of drug ingredients, dosage strength, and physical form. Where possible, terms already in the UMLS were used for describing drug terms. Like LOINC, the terminology model also describes the format for standardized names, e.g.: *Acetaminophen 500 MG Oral Tablet*. RxNorm is fairly new terminology but has grown rapidly. Thanks to collaboration with the Veterans Affairs work on NDF-RT, the content of RxNorm has quickly reached over 14,000 terms, each of which has a full representation in the ontologic model. The principled, nonproprietary approach has paid off: RxNorm has been recommended by NCVHS as one of the standard termi-

nologies for the core patient medical record information [41]. In addition, publishers of on-line health information have begun to index their material with RxNorm codes, facilitating automated queries using patient data.

3.11 NCI Thesaurus

The US National Cancer Institute (NCI), led by Frank Hartel and others, has developed a DL-based terminology called the NCI Thesaurus (NCIT), to support cancer research based on current biomedical science, such as diseases and underlying biology [78, 79]. A major design goal of NCIT is to facilitate translational research, for example, to index clinical trials and document expert summaries [56].

The initial content for NCIT consisted of a collection of local terminologies in use for coding documents related to managing science (e.g., grants and reports), combined with the terminology used to index cancer clinical trials information in the Physician's Data Query (PDQ) system [80]. Today, the NCIT contains 100,000 terms and 34,000 concepts, covering chemicals, drugs and other therapies, diseases (more than 8,500 cancers and related diseases), genes and gene products, anatomy, organisms, animal models, techniques, biologic processes, and administrative categories, including definitions and synonyms [79].

NCIT was built using Apelon's Ontolog™ Data Model [81], and has since been converted into a version of OWL called OWL-Lite [82], which supports only classification hierarchies and simple constraints. Ontolog provides three basic components: concepts, kinds and roles. NCIT contains twenty different kinds, which are disjoint sets of concepts, including Anatomy, Biological Processes, Chemicals and Drugs, and

Diagnostic and Prognostic Factors [79]. NCIT uses approximately 50 types of role relationships to differentiate concepts within each kind. For example, the NCIT concept Chlorambucil is illustrated as [83]:

```
Kind (owl:Class):: Chemicals_and_Drugs_Kind
Concept (owl:Class): Chlorambucil (subclass of
Chemical and Drugs)
  Defining Concepts: Nitrogen Mustard
  Defining Roles: Chemical_or_Drug_FDA_
  Approved_for_Disease
Properties (owl:AnnotationProperty):
  Semantic_Type (property name)
  Pharmacologic Substance (property value)
```

Recently, the OWL-Lite form of NCIT was exploited to carry out a detailed analysis of NCIT content and structure. The analysis was able, through the use of the logical representation of NCIT terms, to detect numerous inconsistencies of the terms and their definitions [84]. As further adoption of NCIT by the cancer research community evolves, studies such as these will assist manual curation.

3.12 DOLCE+

Informatics researchers have occasionally drawn on developments from outside the health informatics field. The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [85], developed at the Institute of Cognitive Sciences and Technology of the Italian National Research Council, is a high-level, domain-independent conceptual framework for representing meaning. Pisanelli and colleagues extended DOLCE with a Descriptions and Situations Ontology to create DOLCE+, which they applied to the task of disambiguation of medical polysemy in the semantic web [86]. For example, they disambiguated the term "inflammation" into the DOLCE+ types of biologic process, feature, quality and situation.

3.13 Protégé

The software tool known as Protégé deserves special mention in any discussion of biomedical ontologies and their impact. Originally developed as a tool supporting the knowledge acquisition needed to represent cancer treatment protocols [87], Protégé evolved into a general-purpose knowledge representation environment [5, 88]. Recently, it has further developed into an open source platform that uses standards such as the Open Knowledge Base Connectivity (OKBC) knowledge model, XML and RDF. These standards, in turn, have supported the development of a library of useful plug-ins. The most important of these is the OWL plug-in [58], which enables the creation and maintenance of ontologies described in the OWL syntax. An additional application, the RACER system, is frequently used together with the Protégé-OWL environment, as it provides reasoning services such as consistency checking and automated classification of concepts [89].

Protégé has been used by hundreds of research groups around the world and the Protégé user community has more than 7000 members [90]. Besides its original protocol-based use, it has been used to represent clinical guidelines [91] and expert system rules [92, 93]. Early on, Protégé was also viewed as a means for constructing ontologies [5]. With the increasing interest in ontology-based terminologies, Protégé has been a welcome tool for researchers trying to handle sets of concepts that were far too large for traditional, commercial reasoning systems. Researchers have used it for managing GO [94, 95], and FMA [96], as well as for constructing ontologies in cardiovascular medicine [97], traditional Chinese medicine [98], hospital incidents [99], and congenital heart defects [100].

4 Discussion

Terminologies have come a long way from the simple lists and hierarchies of 20 years ago. The inclusion of formally represented knowledge is now becoming the rule, rather than the exception. In addition to those described above, many other ontologies are emerging – too many to describe here. But the motivation and results are usually similar: a terminology includes explicit information about the meanings of its terms; this information, in turn, is used to help humans and computers to recognize the intended meanings of the terms for proper coding of, retrieval of, and inferencing about biomedical data, as well as for maintenance of the terminology itself. The terminologies that have made the greatest strides with respect to full representation as ontologies are those that focus on particular domains. Not surprisingly, these are domains whose concepts are very well understood, either because they are in fields that have been closely studied for decades or centuries (such as anatomy) or because the terms represent human artifacts (such as medications, laboratory tests, and diagnostic or therapeutic procedures). The terminologies have grown rapidly to fill niches that were poorly served by previous domain-specific or broad coverage terminologies. Their growth is due in no small part to the principled approaches taken in their design and construction. Their acceptance as standards is similarly due to these approaches, which help users finally understand the meanings intended by terminology developers.

The orientation of terminology development toward ontologies is due, in part, to the evolution of logic-based representations, such as OWL and its variants. Previous attempts at standard representations of frame-based termi-

nologies, such as LOOM and KIF, failed to capture the rich semantics and constraints that were implicit in the frame-based reasoning systems. Description logic has some limitations as well: although superficially similar, these two approaches rely on fundamentally different modeling assumptions. However, solutions such as OWL-Full make the translation more acceptable [57]. OWL itself has proven to sufficiently expressive for representing frame-based ontologies, such as FMA and NCIT, in ways that support their comparison and integration [101, 102]. In addition to improved logical representation, the improvement of tools can be credited for the rapid evolution of terminologic ontologies. In the past, knowledge management systems were unable to support more than a couple thousand concepts. Today, tools such as Protégé and Ontolog can handle hundreds of thousands.

While the recent growth in ontologies has been encouraging, researchers and developers are now beginning to tackle the tougher domains, such as symptoms, physical findings, radiologic findings, and diseases. As adoption of these standard terminologies increases, the need will arise to standardize not just syntactic and functional aspects of knowledge (which handled by OWL). Work is still needed to unify the semantics of diverse terminology models. Broad-domain terminologies, such as SNOMED-CT, attempt to do this internally, but their sheer size and diversity are hindering success. The UMLS makes some attempt to standardize the semantics of terminologies through the assignments of semantic types to concepts, and representing the kinds of relationships in which concepts of a particular semantic type may participate. However, the UMLS Semantic Net does not attempt to model each terminology in

the Metathesaurus and has evolved only slowly as additional terminologies have been added.

Despite being, in some sense, in their infancy, biomedical ontologies have already had significant impact. Previously, terminology development was largely a manual process, with the attendant inefficiencies, inconsistencies, and delays. Now, terminologies such as the UMLS, the MED, SNOMED-CT, FMA and RxNorm supplement their manual development processes with automated, knowledge-based methods that provide more rapid, consistent and auditable results. For example, if diseases are being classified by their anatomic locations, the knowledge of which body parts are subsumed by others will assist in the proper assignment of disease classes; e.g., the inference that a disease of the lung alveoli is a lung disease will be possible, using a high-quality terminology such as FMA.

Ontologies have also made inroads into terminology standardization efforts that previously were mired politics and provincialisms. Today, agreement on terminology standards has progressed rapidly, thanks to Galen, SNOMED-CT, LOINC, GO, and RxNorm. The agreement on terminology standards is crucial for the success of standardized messaging and health information systems [103].

Improvements in biomedical research and health care through information technology have lagged behind the progress made in other fields. The lag has been generally attributed to the complex, conceptual nature of biomedicine. The problem of developing high-quality terminologies has similarly been ascribed to this complexity. The approach of solving the conceptualization problem through terminologies, rather than the other way around, has led to the ontologic approach, reviewed above.

Based on preliminary experiments with evolving biomedical ontologies (reviewed above), we expect the further development and adoption of standard terminologies to catalyze a wide range of advances in the use of health information. For example, we can anticipate that the sharing of expert system logic, through standards such as Medical Logic Modules, will be facilitated with standard terminologies that can help adapt generic logic to local systems (the so-called “curly braces” problem) [104]. We also expect that term translation will be facilitated by ontologic information, by mapping term descriptions from a source terminology into those that correspond to term descriptions in a target terminology. High quality translation, in turn, has implications for automated information retrieval (for example, searching literature databases using coded clinical data) and system integration (for example, when data need to be transmitted from one system to another for incorporation into a central database).

Twelve years ago, the US Government Accounting Office examined health information standards along a development spectrum of “Concept”, “Discussion”, “Drafts”, “Publication”, “Implementation”, “New Products” and “Market Acceptance”. At the time, the report listed “Vocabulary” as being at the “Concept” stage [103]. Today, terminology has outstripped other standards, such as health record structure, and is firmly in the “implementation stage. This advance can be directly attributed to careful ontologic research and development carried out by the international biomedical informatics community. The next decade will show us if the approach will pay off in terms of improved health information systems, improved translation of research into practice, and, ultimately, improved health outcomes.

References

1. World Health Organization. Manual of the International Classification of Diseases, Injuries, and causes of Death. Geneva, Switzerland; 1977.
2. Cimino JJ. Coding systems in health care. In: van Bemmel JH, McCray AT, editors. 1995 Yearbook of Medical Informatics. Stuttgart: Schattauer, 1995. p. 71-85. Reprinted in *Methods Inf Med* 1996; 35(4): 273-84.
3. Cimino JJ. Desiderata for controlled medical vocabularies in the Twenty-First Century. *Methods Inf Med* 1998; 37(4-5): 394-403.
4. Greenes RA, McClure RC, Pattison-Gordon E, Sato L. The findings—diagnosis continuum: implications for image descriptions and clinical databases. *Proc Annu Symp Comput Appl Med Care*. 1992:383-7.
5. Musen MA. Dimensions of knowledge sharing and reuse. *Comput Biomed Res* 1992 Oct; 25(5):435-67.
6. Gruber TR. A translation approach to portable ontologies. *Knowledge Acquisition* 1993; 5(2): 199-220.
7. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. *Methods Inf Med* 1991; 30(3): 179-86.
8. Nowlan WA, Rector AL, Kay S, Horan B, Wilson A. A patient care workstation based on user centred design and a formal theory of medical terminology: PEN&PAD and the SMK formalism. *Proceedings - the Annual Symposium on Computer Applications in Medical Care*. 1991: 855-7.
9. Rector AL, Nowlan WA. The GALEN project. *Comput Methods Programs Biomed* 1994 Oct; 45(1-2): 75-8.
10. Rector AL, Rogers JE, Zanstra PE, Van Der Haring E; OpenGALEN. OpenGALEN: open source medical terminology and tools. *AMIA Annu Symp Proc*. 2003: 982.
11. Hardiker NR. Mediating between nursing intervention terminology systems. *AMIA Annu Symp Proc*. 2001:239-43.
12. Hardiker NR, Rector AL. Structural validation of nursing terminologies. *J Am Med Inform Assoc* 2001; 8(3): 212-21.
13. ten Napel H, Rogers J. Assessment of the GALEN methodology on holistic classifications for professions allied to medicine. *Medinfo*. 2001:1369-73.
14. Karlsson D. A design and prototype for a decision-support system in the field of urinary tract infections—application of OpenGALEN techniques for indexing medical information. *Medinfo*. 2001: 479-83.
15. Trombert-Pavio B, Rodrigues JM, Rogers JE, Baud R, van der Haring E, Rassinoux AM, et al. GALEN: a third generation terminology tool to support a multipurpose national coding system for surgical procedures. *Int J Med Inform* 2000; 58-59:71-85.
16. Donnelly M. Containment relations in anatomical ontologies. *AMIA Fall Symposium*, 2005: 206-10.
17. Mork P, Pottinger R, Bernstein PA. Challenges in precisely aligning models of human anatomy using generic schema matching. *Medinfo*. 2004: 401-5.
18. Zhang S, Bodenreider O. Aligning representations of anatomy using lexical and structural methods. *AMIA Annu Symp Proc*. 2003:753-7.
19. http://www.nlm.nih.gov/research/umls/about_umls.html.
20. Selden CR, Humphreys BL. Unified Medical Language System. *Current Bibliographies in Medicine* 1997; 8: 96.
21. Meystre S, Haug PJ. Evaluation of medical problem extraction from electronic clinical documents using MetaMap Transfer (MMTx). *Stud Health Technol Inform* 2005;116: 823-8.
22. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc* 2005 Sep-Oct;12(5): 517-29.
23. Sweeney JP, Portell KS, Houck JA, Smith RD, Mentel JJ. Patient note deidentification using a find-and-replace iterative process. *J Healthc Inf Manag* 2005; 19(3): 65-70.
24. Sadeghi S, Barzi A, Smith JW. Ontology driven construction of a knowledgebase for Bayesian decision models based on UMLS. *Stud Health Technol Inform* 2005; 116: 223-8.
25. Perez-Rey D, Maojo V, Garcia-Remesal M, Alonso-Calvo R, Billhardt H, Martin-Sanchez F, et al. ONTOFUSION: Ontology-based integration of genomic and clinical databases. *Comput Biol Med*. In press 2005.
26. Bowman S. Coordinating SNOMED-CT and ICD-10. *J AHIMA* 2005 Jul-Aug; 76(7): 60-1.
27. Slaughter LA, Soergel D, Rindflesch TC. Semantic representation of consumer questions and physician answers. *Int J Med Inform*. In press 2005.
28. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J Am Med Inform Assoc* 1994 Jan-Feb; 1(1): 35-50.
29. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. Le rôle du lexique sémantique et de l'ontologie dans le traitement automatique de la langue médicale. In: Le Beux P, Burgun A, editors. *Actes du Colloque CRISTALS*, Saint-Malo, 1996.
30. Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *J Am Med Inform Assoc* 2000 May-Jun; 7(3): 288-97.
31. Hwang JI, Cimino JJ, Bakken S. Integrating nursing diagnostic concepts into the medical entities dictionary using the ISO Reference Terminology Model for Nursing Diagnosis. *J Am Med Inform Assoc* 2003 Jul-Aug; 10(4): 382-8.
32. Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. *Ophthalmology* 2005 Feb; 112(2): 175-83.
33. Choi J, Jenkins ML, Cimino JJ, White TM, Bakken S. Toward semantic interoperability in home health care: formally representing OASIS items for integration into a concept-oriented terminology. *J Am Med Inform Assoc* 2005 Jul-Aug; 12(4): 410-7.
34. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp*. 2001:662-6.

35. 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. *Proc AMIA Symp.* 2001: 741-5.
36. Wang AY, Sable JH, Spackman KA. The SNOMED clinical terms development process: refinement and analysis of content. *Proc AMIA Symp.* 2002:845-9.
37. Schulz S, Hahn U, Rogers J. Semantic clarification of the representation of procedures and diseases in SNOMED-CT. *Stud Health Technol Inform* 2005; 116: 773-8.
38. Burgun A, Bodenreider O, Mouglin F. Classifying diseases with respect to anatomy: a study in SNOMED-CT. *Proc AMIA Symp.* 2005: 91-5.
39. Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. *Proc AMIA Symp.* 2005: 266-70.
40. Fung KW, Hole WT, Nelson SJ, Srinivasan S, Powell T, Roth L. Integrating SNOMED-CT into the UMLS: an exploration of different views of synonymy and quality of editing. *J Am Med Inform Assoc* 2005; 12(4): 486-94.
41. Cohn SP. Seventh Annual Report to Congress on the Implementation Of the Administrative Simplification Provisions of the Health Insurance Portability and Accountability Act of 1996. September 8, 2005. www.ncvhs.hhs.gov/050908rpt.htm.
42. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, et al. Logical Observation Identifiers, Names and Codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem* 1996 Jan; 42(1): 81-90.
43. Baorto DM, Cimino JJ, Parvin CA, Kahn MG. Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC). *Int J Med Inform* 1998 Jul; 51(1): 29-37.
44. Cimino JJ, Meyer M, Lee NJ, Bakken S. Using patient data to retrieve health knowledge. *Proc AMIA Symp.* 2005: 136-40.
45. Khan AN, Russell D, Moore C, Rosario AC Jr, Griffith SP, Bertolli J. The map to LOINC project. *Proc AMIA Symp.* 2003: 890.
46. Dolin RH, Mattison JE, Cohn S, et al. Kaiser Permanente's Convergent Medical Terminology. *Medinfo.* 2004; 11(Pt 1): 346-50.
47. Dykes PC, Currie LM, Cimino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. *J Biomed Inform* 2003 Aug-Oct; 36(4-5): 313-25.
48. Parker CG, Rocha RA, Campbell JR, Tu SW, Huff SM. Detailed clinical models for sharable, executable guidelines. *Medinfo.* 2004; 11(Pt 1): 145-8.
49. Matney S, Bakken S, Huff SM. Representing nursing assessments in clinical information systems using the logical observation identifiers, names, and codes database. *J Biomed Inform* 2003 Aug-Oct; 36(4-5): 287-93.
50. Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003 Dec; 36(6): 478-500.
51. Rosse C, Shapiro LG, Brinkley JF. The digital anatomist foundational model: principles for defining and structuring its concept domain. *Proc AMIA Symp.* 1998:820-4.
52. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the Digital Anatomist symbolic knowledge base. *J Am Med Inform Assoc* 1998 Jan-Feb; 5(1): 17-40.
53. <http://sig.biostr.washington.edu/projects/fm/AboutFM.html>
54. Shapiro LG, Chung E, Detwiler LT, Mejino Jr JLV, Agoncillo AV, Brinkley JF, et al. Processes and problems in the formative evaluation of an interface to the Foundational Model of anatomy knowledge base. *J Am Med Inform Assoc* 2005; 12(1): 35-46.
55. WorldWideWeb Consortium. OWL Web Ontology Language Reference. In: W3C Recommendation 10 Feb. 2004.
56. Hartel FW, de Coronado S, Dionne R, Fragoso G, Golbeck J. Modeling a description logic vocabulary for cancer research. *J Biomed Inform* 2005; 38(2): 114-29.
57. Dameron O, Rubin DL, Musen MA. Challenges in converting frame-based ontology into OWL: the Foundational Model of Anatomy case-study. *Proc AMIA Symp.* 2005:181-5.
58. Knublauch H, Ferguson RW, Noy NF, Musen MA. The Protege OWL Plugin: An open development environment for Semantic Web applications. *Semantic Web - ISWC 2004, Proceedings* 2004; 3298: 229-43.
59. Rubin DL, Dameron O, Musen MA. Use of description logic classification to reason about consequences of penetrating injuries. *Proc AMIA Symp.* 2005: 649-53.
60. Warren W, Brinkley JF. Knowledge-based, interactive, custom anatomical scene creation for medical education: The Biolucida System. *Proc AMIA Symp.* 2005: 789-93.
61. 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.
62. The GO Consortium. Creating the Gene Ontology resource: design and implementation. *Genome Res* 2001 Aug; 11(8): 1425-33.
63. Coenen A, McNeil B, Bakken S, Bickford C, Warren JJ; American Nurses Association Committee on Nursing Practice Information Infrastructure. Toward comparable nursing data: American Nurses Association criteria for data sets, classification systems, and nomenclatures. *Comput Nurs* 2001 Nov-Dec; 19(6): 240-8.
64. Warren JJ, Mead CN, Button P, Henry SB, Androwich I. Development of the Loose Canon Model of nursing interventions represented using the Unified Modeling Language (UML). *Proc AMIA Symp* 1999: 1189.
65. Moss J, Coenen A, Mills ME. Evaluation of the draft international standard for a reference terminology model for nursing actions. *J Biomed Inform* 2003 Aug-Oct; 36(4-5): 271-8.
66. International Standards Organization. *Integration of a Reference Terminology Model for Nursing - FDIS.* Geneva, Switzerland: International Standards Organization; 2003.
67. Harris M, Kim H, Rhudy L, Savova G, Chute C. Testing the generalizability of the ISO model for nursing diagnoses. *AMIA Annu Symp Proc.* 2003: 274-8.
68. Bakken S, Hyun S, Friedman C, Johnson SB. ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. *Int J Med Inform* 2005 Aug; 74(7-8): 615-22.
69. Junttila K, Salantera S, Hupli M. Developing terminology for documenting perioperative nursing interventions. *Int J Med Inform* 2005 Jul; 74(6): 461-71.
70. Brown SH, Lincoln MJ, Groen P, Kolodner RM. VistA: The U.S. Department of Veterans Affairs national-scale Hospital Information System. *Int J Med Inform* 2003; 69(2/3): 135-56.
71. Carter JS, Brown SH, Erlbaum MS, Gregg W, Elkin PL, Speroff T, et al. Initializing the VA Medication Reference Terminology using UMLS Metathesaurus co-Occurrences. *Proc AMIA Symp* 2002: 116-20.
72. Rosenbloom S, Awad J, Speroff T, Elkin P, Rothman R, Spickard AI, et al. Adequacy of representation of the National Drug File Reference Terminology Physiologic Effects reference hierarchy for commonly prescribed medications. *Proc AMIA.* 2003: 569-73.
73. United States Pharmacopeia Medicare Model Guidelines Expert Committee. *United States Pharmacopeia Medicare Model Guidelines.* October 8, 2004.
74. Brown SH, Elkin PL, Rosenbloom ST, Husser C, Bauer BA, Lincoln MJ, et al. VA National Drug File Reference Terminology: a cross-institutional content coverage study. *Medinfo.* 2004; 11(Pt 1): 477-81.
75. Lincoln MJ, Brown SH, Nguyen V, Cromwell T, Carter J, Erlbaum M, et al. U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview. *Medinfo.* 11(Pt 1):391-5, 2004.
76. <http://www.nlm.nih.gov/research/umls/rxnorm/>
77. Cimino JJ, McNamara TJ, Meredith T, Broverman CA, Eckert KC, Moore M, et al. Evaluation of a proposed method for representing drug terminology. *J Am Med Inform Assoc* 1999; 6 (suppl.): 47-51.
78. <http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do>
79. de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus: using science-based terminology to integrate cancer research results. *Medinfo.* 2004; 11(Pt 1): 33-7.
80. Hubbard SM, Martin NB, Blankenbaker LW, Esterhay RJ Jr, Masys DR, Tingley DE, et al. The Physician Data Query (PDQ) cancer information system. *J Cancer Educ* 1986; 1(2): 79-87.
81. Apelon. <http://www.apelon.com/products/products.htm>
82. Smith MK, Welty C, McGuinness DL. *OWL Web Ontology Language Guide.* <http://www.w3.org/TR/owl-guide/2003>
83. Souripriya Das et al. Building Semantically Rich Life Science Applications using OWL Ontologies in Oracle. September, 2004.
84. Ceusters W, Smith B, Goldberg L. A terminological and ontological analysis of the NCI Thesaurus. *Methods Inf Med* 2005; 44: 498-507.
85. Masolo C, Borgo S, Gangemi A, Guarino N, Ultramari A, Schneider L. *The WonderWeb Library*

- of foundational ontologies. WonderWeb Deliverable 18, <http://wonderweb.semanticweb.org> (2003).
86. Pisanelli DM, Gangemi A, Battaglia M, Catenacci C. Coping with medical polysemy in the semantic web: the role of ontologies. *Medinfo*. 2004;11(Pt 1): 416-9.
 87. Tu SW, Eriksson H, Gennari JH, Shahar Y, Musen MA. Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: application of PROTEGE-II to protocol-based decision support. *Artif Intell Med* 1995; 7(3): 257-89.
 88. Gennari J, Musen MA, Ferguson RW, Grosso WE, Crubézy M, Eriksson H, et al. The evolution of Protégé: an environment for knowledge-based systems development. *Int J Hum Comput Stud* 2003; 58(1): 89-123.
 89. Volker Haarslev RM. Description of the RACER system and its applications. *International Workshop on Description Logics (DL-2001)*; 2001 August 1-3; Stanford, USA; 2001.
 90. Noy NF, Crubezy M, Ferguson RW et al. Protégé-2000: An open-source ontology-development and knowledge-acquisition environment. *Proc AMIA Symp*. 2003: 953.
 91. Shankar RD, Tu SW, Musen MA. A knowledge-acquisition wizard to encode guidelines. *Proc AMIA Symp*. 2003: 1007.
 92. Park JY, Musen MA. VM-in-Protege: a study of software reuse. *Medinfo*. 1998; 9 Pt 1: 644-8.
 93. Bang M, Eriksson H. Generation of development environments for the Arden Syntax. *Proc AMIA Symp*. 1997: 313-7.
 94. Yeh I, Karp PD, Noy NF, Altman RB. Knowledge acquisition, consistency checking and concurrency control for Gene Ontology (GO). *Bioinformatics* 2003; 19(2): 241-8.
 95. Lambrix P, Habbouche M, Perez M. Evaluation of ontology development tools for bioinformatics. *Bioinformatics* 2003; 19(12): 1564-71.
 96. Detwiler LT, Mejino Jr JV, Rosse C, Brinkley JF; Structural Informatics Group. Efficient web-based navigation of the Foundational Model of Anatomy. *AMIA Annu Symp Proc*. 2003;: 829.
 97. Jiang G, Ogasawara K, Endoh A, Sakurai T. Context-based ontology building support in clinical domains using formal concept analysis. *Int J Med Inform* 2003; 71(1): 71-81.
 98. Zhou X, Wu Z, Yin A, Wu L, Fan W, Zhang R. Ontology development for unified traditional Chinese medical language system. *Artif Intell Med* 2004; 32(1):15-27.
 99. Fan L, Smith A, Boenning D, Castro G, Champagne S, Loeb JM, et al. Building a hospital incident reporting ontology (HIRO) in the Web Ontology Language (OWL) using the JCAHO Patient Safety Event Taxonomy (PSET). *Proc AMIA Symp*. 2005: 952.
 100. Vishwanath K, Viswanath V, Drake W, Lee Y. OntoDiagram: automatic diagram generation for congenital heart defects in pediatric cardiology. *Proc AMIA Symp*. 2005: 754-8.
 101. Zhang S, Bodenreider O. Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference. *Proc AMIA Symp*. 2005: 864-8.
 102. Supekar K, Chute CG, Solbrig H. Representing lexical components of medical terminologies in OWL. *Proc AMIA Symp*. 2005:719-23.
 103. U.S. General Accounting Office. Automated Medical Records Leadership Needed to Expedite Standards Development. *GAO/IMTEX-93-17*, Washington, D.C.: April 30, 1993.
 104. Jenders RA, Sujansky W, Broverman CA, Chadwick M. Towards improved knowledge sharing: assessment of the HL7 Reference Information Model to support medical logic module queries. *Proc AMIA Annu Fall Symp*. 1997: 308-12.

Correspondence to:
 Dr. James Cimino
 Department of Biomedical Informatics
 Columbia University College of Physicians and Surgeons
 622 West 168th Street, VC-5
 New York, New York, 10032
 USA
 E-mail: jjc7@columbia.edu