

# Knowledge Representation and Ontologies

**Kin Wah Fung and Olivier Bodenreider**

*National Library of Medicine, Bethesda, Maryland, USA*

Ontologies have become important tools in biomedicine, supporting critical aspects of both health care and biomedical research, including clinical research [1]. Some even see ontologies as integral to science [2]. Unlike terminologies (focusing on naming) and classification systems (developed for partitioning a domain), ontologies define the types of entities that exist, as well as their interrelations. And while knowledge bases generally integrate both definitional and assertional knowledge, ontologies focus on what is always true of entities, i.e., definitional knowledge [3]. In practice, however, there is no sharp distinction between these kinds of artifacts and ‘ontology’ has become a generic name for a variety of knowledge sources with important differences in their degree of formality, coverage, richness and computability [4]. In this chapter, we focus on those ontologies of particular relevance to clinical research. After a brief introduction to ontology development and knowledge representation, we present the characteristics of some of these ontologies. We then show how ontologies are integrated in and made accessible through knowledge repositories, and illustrate their role in clinical research.

## **Ontology development**

Ontology development has not yet been formalized to the same extent as, say, database development has, and there is still no equivalent for ontologies to the entity-relationship model. However, ontology development is guided by fundamental ontological distinctions and supported by the formalisms and tools for knowledge representation that have emerged over the past decades. Several top-level ontologies provide useful constraints for the development of domain ontologies and one the most recent trends is increased collaboration among the creators of ontologies for coordinated development.

### *Important ontological distinctions*

A small number of ontological distinctions inherited from philosophical ontology provide a useful framework for creating ontologies. The first distinction is between types and instances. Instances correspond to individual entities (e.g., my left kidney, the patient identified by 1234), while types represent the common characteristics of sets of instances (e.g., a *kidney* is a bean-shaped, intra-abdominal organ – properties common to all kidneys) [5]. Instances are related to the corresponding types by the relation *instance of*. For example, my left kidney is an *instance of kidney*. (It must be noted that most biomedical ontologies only represent types in reference to which the instances recorded in patient records and laboratory notebooks can be annotated). Another fundamental distinction is between continuants and occurrents [6]. While continuants exist (endure) through time, occurrents go through time in phases. Roughly speaking, objects (e.g., a liver, an endoscope) are continuants and processes (e.g., the flow of blood through the mitral valve) are occurrents. One final distinction is made between independent and dependent continuants. While the kidney and its shape are both continuants, the shape of the kidney “owes” its existence to the kidney (i.e., there cannot be a kidney shape unless there is a kidney in the first place). Therefore, the kidney is an independent continuant (as most objects are), whereas its shape is a dependent continuant (as are qualities, functions and dispositions, all dependent on their bearers). These distinctions are important for ontology developers, because they help organize entities in the ontology and contribute to consistent ontology development, both within and, more importantly for interoperability, across ontologies.

### *Building blocks: Top-level ontologies and Relation Ontology*

These ontological distinctions are so fundamental that they are embodied by top-level ontologies such as BFO [7] (Basic Formal Ontology) and DOLCE [8] (Descriptive Ontology for Linguistic and Cognitive Engineering). Such upper-level ontologies are often used as building blocks for the development of domain ontologies. Instead of organizing the main categories of entities of a given domain under some artificial root, these categories can be implemented as specializations of types from the upper-level ontology. For example, a protein is an independent continuant, the catalytic function of enzymes is a dependent continuant, and the

activation of an enzyme through phosphorylation is an occurrent. Of note, even when they do not leverage an upper-level ontology, most ontologies implement these fundamental distinctions in some way. For example, the first distinction made among the semantic types in the UMLS Semantic Network [9] is between *Entity* and *Event*, roughly equivalent to the distinction between continuants and occurrents in BFO. While BFO and DOLCE are generic upper-level ontologies, Bio-Top [10] – itself informed by BFO and DOLCE – is specific to the biomedical domain and provides types directly relevant to this domain, such as *Chain Of Nucleotide Monomers* and *Organ System*. BFO forms the backbone of several ontologies from the Open Biomedical Ontologies (OBO) family and Bio-Top has also been reused by several ontologies. Some also consider the UMLS Semantic Network, created for categorizing concepts from the UMLS Metathesaurus, an upper-level ontology for the biomedical domain [9]. In addition to the ontological template provided for types by upper-level ontologies, standard relations constitute an important building block for ontology development and help ensure consistency across ontologies. The small set of relations defined collaboratively in the Relation Ontology [5], including *instance of*, *part of* and *located in*, has been widely reused.

### *Formalisms and tools for knowledge representation*

Many ontologies use description logics for their representation. Description logics (DLs) are a family of knowledge representation languages, with different levels of expressiveness [11]. The main advantage of using DL for ontology development is that DL allows developers to test the logical consistency of their ontology. This is particularly important for large biomedical ontologies. Ontologies including CTO, OCRE, OBI, SNOMED CT, NDF-RT and the NCI Thesaurus, discussed later in this chapter, all rely on some sort of DL for their development.

Ontologies are key enabling resources for the Semantic Web, the “web of data”, where resources annotated in reference to ontologies can be processed and linked automatically [12]. It is therefore not surprising that the main language for representing ontologies, OWL – the Web Ontology Language, has its origins in the Semantic Web. OWL is developed under the auspices of the World Wide Web Consortium (W3C). The current version of the OWL specification is OWL 2, which offers several profiles (sublanguages) corresponding to different levels of

expressivity and support of DL languages [13]. Other Semantic Web technologies, such as RDF/S (Resource Description Framework Schema) [14] and SKOS (Simple Knowledge Organization System) [15] have also been used for representing taxonomies and thesauri, respectively.

The OWL syntax can be overwhelming to biologists and clinicians, who simply want to create an explicit specification of the knowledge in their domain. The developers of the Gene Ontology created a simple syntax later adopted for the development of many ontologies from the Open Biomedical Ontologies (OBO) family. The so-called OBO syntax [16] provides an alternative to OWL, to which it can be converted [17].

The most popular ontology editor is Protégé, developed at the Stanford Center for Biomedical Informatics Research for two decades [18, 19]. Originally created for editing frame-based ontologies, Protégé now supports OWL and other Semantic Web languages. Dozens of user-contributed plugins extend the standalone version (e.g., for visualization, reasoning services, support for specific data formats) and the recently-developed web version of Protégé supports the collaborative development of ontologies. Originally created to support the development of the Gene Ontology, OBO-Edit now serves as a general ontology editor [20, 21].

Simpler than Protégé, OBO-Edit has been used to develop many of the ontologies from the Open Biomedical Ontologies (OBO) family. Rather than OWL, OBO-Edit uses a specific format, the OBO syntax, for representing ontologies. Both Protégé and OBO-Edit are open-source, platform independent software tools.

Other ontology editors related to some of the ontologies presented in this chapter include Apelon's proprietary Terminology Development Environment (TDE), based on the description logics KRSS and used for the development of NDF-RT, and the IHTSDO Workbench, an open-source, freely-available editing environment created for the collaborative development of SNOMED CT.

### *OBO Foundry and other harmonization efforts*

Two major issues with biomedical ontologies are their proliferation and their lack of interoperability. There are several hundreds of ontologies available in the domain of life sciences, some of which overlap partially but do not systematically cross-reference equivalent entities in other ontologies. The existence of multiple representations for the same entity makes it difficult for ontology users to select

the right ontology for a given purpose and requires the development of mappings between ontologies to ensure interoperability. Two recent initiatives have offered different solutions to address the issue of uncoordinated development of ontologies.

The OBO Foundry is an initiative of the Open Biomedical Ontologies (OBO) consortium, which provides guidelines and serves as coordinating authority for the prospective development of ontologies [22]. Starting with the Gene Ontology, the OBO Foundry has identified kinds of entities for which ontologies are needed and have selected candidate ontologies to cover a given subdomain, based on a number of criteria. Granularity and fundamental ontological distinctions form the basis for identifying subdomains. For example, independent continuants (entities) at the molecular level include proteins (covered by the protein ontology), while macroscopic anatomical structures are covered by the Foundational Model of Anatomy. In addition to syntax, versioning and documentation requirements, the OBO Foundry guidelines prescribe that OBO Foundry ontologies be limited in scope to a given subdomain and orthogonal. This means, for example, that an ontology of diseases referring to anatomical structures as the location of diseases (e.g., *mitral valve regurgitation has location mitral valve*) should cross-reference entities from the reference ontology for this domain (e.g., the Foundational Model of Anatomy for *mitral valve*), rather than redefine these entities. While well adapted to coordinating the prospective development of ontologies, this approach is extremely prescriptive and virtually excludes the many legacy ontologies used in the clinical domain, including SNOMED CT and the NCI Thesaurus.

The need for harmonization, i.e., making existing ontologies interoperable and avoiding duplication of development effort, has not escaped the developers of large clinical ontologies. The International Health Terminology Standard Development Organization (IHTSDO), in charge of the development of SNOMED CT, is leading a similar harmonization effort in order to increase interoperability and coordinate the evolution of legacy ontologies and terminologies, including Logical Observation Identifiers Names and Codes (LOINC, for laboratory and clinical observations), the International Classification of Diseases (ICD) and the International Classification for Nursing Practice (ICNP, for nursing diagnoses) [23].

## **Ontologies of particular relevance to clinical research**

Broadly speaking, clinical research ontologies can be classified into those that model the characteristics (or metadata) of the clinical research and those that model the data contents generated as a result of the research. [24] Research metadata ontologies center around characteristics like study design, operational protocol and methods of data analysis. They define the terminology and semantics necessary for formal representation of the research activity and aim to facilitate activities such as automated management of clinical trials and cross-study queries based on study design, intervention or outcome characteristics. Ontologies of data content focus on explicitly representing the information model of and data elements (e.g. clinical observations, laboratory test results) collected by the research, with the aim to achieve data standardization and semantic data interoperability. Some examples of the two types of ontology will be described in more detail. Finally, examples of ontology-driven knowledge bases for translational research will be presented briefly.

### *Research metadata ontology*

A survey of the public repository of ontologies in the Open Biomedical Ontologies (OBO) library hosted by the National Center of Biomedical Ontology (see below) yielded three ontologies that fit the description of research metadata ontology. These are the Epoch Clinical Trial Ontologies (CTO), Ontology of Clinical Research (OCRe) and Ontology for Biomedical Investigations (OBI).

### **Epoch Clinical Trial Ontologies**

CTO is a suite of ontologies that encodes knowledge about clinical trials. The use of this ontology is demonstrated in the integration of software applications for the management of clinical trials under the Immune Tolerance Network. [25] By building an ontology-based architecture the disparate clinical trial software applications can share essential information to achieve interoperability for efficient management of the trials and analysis of trial data. CTO is made up of the following component ontologies:

1. Clinical trial ontology – the overarching ontology that covers protocol specification and operational plan
2. Protocol ontology – the knowledge model of the clinical trial protocol
3. Organization ontology – supports the specification of study sites, laboratories and repositories
4. Assay ontology – models characteristics of tests (e.g. specimen type, workflow of specimen processing)
5. Labware ontology – models the laboratory entities (e.g. specimen containers)
6. Virtual trial data ontology – models the study data being collected (e.g. participant clinical record, specimen workflow log)
7. Constraint expression ontology – models logical and temporal constraints
8. Measurement ontology – models physical measurements and units of measurement

There are three stated goals of CTO: to support tools which help acquire and maintain knowledge about protocol and assay designs, to drive data collection during a trial, and to facilitate implementation of querying methods to support trial management and *ad hoc* data analysis. A clinical trial protocol authoring tool has been developed based on CTO. [26] The ability to map from CTO to the Biomedical Research Integrated Domain Group (BRIDG) information model has been demonstrated. [27]

### **Ontology of Clinical Research**

While the main use case of CTO is in the automation of design and workflow management of clinical research, the primary aim of OCR<sub>e</sub> is to support the annotation and indexing of human studies to enable cross-study comparison and synthesis. [28] Developed as part of the Trial Bank Project, OCR<sub>e</sub> provides terms and relationships for characterizing the essential design and analysis elements of clinical studies. Domain specific concepts are covered by reference to external vocabularies. Workflow related characteristics (e.g. schedule of activities) and data structure specification (e.g. schema of data elements) are not within the scope of OCR<sub>e</sub>.

The three core modules of OCR<sub>e</sub> are:

1. Clinical module – the upper-level entities (e.g. clinician, study subject)
2. Study design module – models study design characteristics (e.g. investigator assigned intervention, external control group)
3. Research module – terms and relationships to characterize a study (e.g. outcome phenomenon, assessment method)

OCRe entities are mapped to the Basic Formal Ontology (BFO).

### **Ontology for Biomedical Investigations**

Unlike CTO and OCRe whose creations are rooted in clinical research, the origin of OBI is in the molecular biology research domain. [29] The forerunner of OBI is the MGED Ontology developed by the Microarray Gene Expression Data Society for annotating microarray data. Through collaboration with other groups in the ‘OMICS’ arena such as the Proteomics Standards Initiative (PSI) and Metabolomics Standards Initiative (MSI), MGED Ontology was expanded to cover proteomics and metabolomics and was subsequently renamed Functional Genomics Investigation Ontology (FuGO). [30] The scope of FuGO was later extended to cover clinical and epidemiological research and biomedical imaging, resulting in the creation of OBI, which aims to cover all biomedical investigations [31].

Another difference between OBI and the other two ontologies is the collaborative approach to its development. As OBI is an international, cross-domain initiative, the OBI Consortium draws upon a pool of experts from many fields, including even fields outside biology such as environmental science and robotics. The goal of OBI is to build an integrated ontology to support the description and annotation of biological and clinical investigations, regardless of the particular field of study. OBI also uses the BFO as its upper-level ontology and all OBI classes are a subclass of some BFO class. OBI covers all phases of the experimental process, and the entities or concepts involved, such as study designs, protocols, instrumentation, biological material, collected data and their analyses. OBI also represents roles and functions which can be used to characterize and relate these entities or concepts. Specifically, OBI covers the following areas:

1. Biological material – e.g. blood plasma
2. Instrument – e.g. microarray, centrifuge



3. Information content – e.g. electronic medical record, biomedical image
4. Design and execution of an investigation – e.g. study design, electrophoresis
5. Data transformation – e.g. principal components analysis, mean calculation

For domain-specific entities, OBI makes reference to other ontologies such as Gene Ontology (GO) and Chemical Entities of Biological Interest (ChEBI). The ability of OBI to adequately represent and integrate different biological experimental processes and their components has been demonstrated in examples from several domains, including neuroscience and vaccination.

### *Data content ontology*

While there are relatively few metadata ontologies, there is a myriad of ontologies that cover research data contents. Unlike metadata ontologies, in this group the distinction between ontologies, terminologies, classifications and code sets often gets blurred. Three ontologies are chosen for more detailed discussion here: the National Cancer Institute Thesaurus (NCIT), Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) and National Drug File Reference Terminology (NDF-RT). These are chosen because they are arguably closer to the ontology end of the ontology-vocabulary continuum than most other artifacts in this category, and their content areas are most relevant to clinical research. All of them have concept-based organization with a rich network of inter-concept relationships and use Description Logic formalism in content creation and maintenance. All three ontologies are available through the Unified Medical Language System (UMLS) and the BioPortal ontology repositories (see below).

### **National Cancer Institute Thesaurus**

NCIT is developed by the U.S. National Cancer Institute (NCI). It arose initially from the need for an institution-wide common terminology to facilitate interoperability and data sharing by the various components of NCI. [32-34] NCIT covers clinical and basic sciences as well as administrative areas. Even though the content is primarily cancer-centric, since cancer research spans a broad area of biology and medicine, NCIT can potentially serve the needs of other research communities. Due to its coverage of both basic and clinical research,

NCIT is well positioned to support translational research. NCIT is the reference terminology for the NCI's Cancer Biomedical Informatics Grid (caBIG) and other related projects. It is also one of the U.S. Federal standard terminologies designated by the Consolidated Health Informatics (CHI) initiative.

NCIT contains about 80,000 concepts organized into 19 disjoint domains. A concept is allowed to have multiple parents within a domain. NCIT covers the following areas:

1. Neoplastic and other diseases
2. Findings and abnormalities
3. Anatomy, tissues and subcellular structures
4. Agents, drugs and chemicals
5. Genes, gene products and biological processes
6. Animal models of disease
7. Research techniques, equipment and administration

NCIT is updated monthly. It is in the public domain under an open content license and is distributed by the NCI in OWL format.

### **SNOMED Clinical Terms**

SNOMED CT was originally developed by the College of American Pathologists. Its ownership was transferred to the International Health Terminology Standards Development Organisation (IHTSDO) in 2007 to enhance international governance and adoption. [35] There are currently 15 member countries including U.S, United Kingdom, Canada, Australia, Netherlands, Sweden and Spain.

SNOMED CT is the most comprehensive clinical terminology available today, with almost 300,000 active concepts. The concepts are organized into 19 disjoint hierarchies. Within each hierarchy, a concept is allowed to have multiple parents. Additionally, SNOMED CT provides a rich set of associated relations (across hierarchies), which form the basis for the logical definitions of its concepts. The principal use of SNOMED CT is to encode clinical information (e.g. diseases, findings, procedures). It also has comprehensive coverage of drugs, organisms and anatomy. SNOMED CT is a CHI-designated U.S. Federal terminology standard. It is also one of the named terminology standards for the Problem List in the "meaningful use" criteria for the Electronic Health Record published by the

U.S. Department of Health and Human Services [36, 37]. SNOMED CT is updated twice yearly. The use of SNOMED CT is free in all IHTSDO member countries, in low-income countries as defined by the World Bank, and for qualified research projects in any country. SNOMED CT is available in proprietary release format from the National Release Center of the IHTSDO member countries.

### **National Drug File Reference Terminology**

NDF-RT is developed by the U.S. Veteran Health Administration (VA) as an extension to their National Drug File, which is the master list of drugs prescribed to VA patients. In addition to drug names, ingredients, dose forms and strengths, NDF-RT contains hierarchies for the chemical structure, mechanism of action, physiologic effect and therapeutic intent of drugs. There is also a disease hierarchy to which drugs may be linked through roles such as *may\_treat*, *may\_prevent* and *may\_diagnose*. NDF-RT contains about 4,000 drugs at the ingredient level. The coverage of NDF-RT has been evaluated using data outside of the VA system and found to be adequate. [38, 39] NDF-RT is in the public domain and is updated monthly. [40] It is available in XML and OWL formats. NDF-RT has recently been integrated with RxNorm and is now available through RxNav and its application programming interfaces (APIs) [41].

### *Ontology-driven knowledge bases for translational research*

Several ontology-driven knowledge bases have been developed in the past few years for translational research purposes. On the one hand, there are traditional data warehouses created through the Clinical and Translational Science Awards (CTSA) program and other translational research efforts. Such warehouse include BTRIS [42], based on its own ontology, the Research Entity Dictionary, and STRIDE [43], based on standard ontologies, such as SNOMED CT and RxNorm. On the other hand, several proof-of-concept projects have leveraged Semantic Web technologies for translational research purposes. In the footsteps of a demonstration project illustrating the benefits of integrating data in the domain of Alzheimer's disease [44], other researchers have developed knowledge bases for cancer data (leveraging the NCI Thesaurus) [45] and in the domain of nicotine dependence (using an ontology developed specifically for the purpose of

integrating publicly-available datasets) [46]. The Translational Medicine Knowledge Base, based on the Translational Ontology, is a more recent initiative developed for answering questions relating to clinical practice and pharmaceutical drug discovery [47].

### **Ontology repositories**

Because most biomedical terminologies and ontologies are developed by different groups and institutions independently of each other and made available to users in heterogeneous formats, interoperability among them is generally limited. In order to create some level of semantic interoperability among ontologies and facilitate their use, several repositories have been created. Such repositories provide access to integrated ontologies through powerful graphical and programming interfaces. This section presents the two largest repositories: the Unified Medical Language System (UMLS) and the BioPortal.

#### *Unified Medical Language System (UMLS)*

The U.S. National Library of Medicine (NLM) started the UMLS project in 1986. One of the main goals of UMLS is to aid the development of systems that help health professionals and researchers retrieve and integrate electronic biomedical information from a multitude of disparate sources. [48-51] One major obstacle to cross-source information retrieval is that the same information is often expressed differently in different vocabularies used by the various systems and there is no universal biomedical vocabulary. Knowing that to dictate the use of a single vocabulary is not realistic, the UMLS circumvents this problem by creating links between the terms in different vocabularies. The UMLS is available free of charge. Users need to acquire a license because some of the UMLS contents are protected by additional license requirements. [52] Currently, there are over 3,000 UMLS licensees in more than 50 countries. The UMLS is released twice a year.

#### **UMLS knowledge sources**

The Metathesaurus of the UMLS is a conglomeration of a large number of terms that exist in biomedical vocabularies. All terms that refer to the same meaning (i.e. synonymous terms) are grouped together in the same UMLS concept. Each UMLS concept is assigned a permanent unique identifier (the Concept Unique

Identifier, CUI), which is the unchanging pointer to that particular concept. This concept-based organization enables cross-database information retrieval based on *meaning*, independent of the lexical variability of the terms themselves. In the 2010AB release, the UMLS Metathesaurus incorporates 153 source vocabularies and includes terms in 20 languages. There are two million biomedical concepts and eight million unique terms. The Metathesaurus also contains relationships between concepts. Most of these relationships are derived from relationships asserted by the source vocabularies. To edit the Metathesaurus, the UMLS editors use a sophisticated set of lexical and rule-based matching algorithms to help them focus on areas that require manual review.

The Semantic Network is another resource in the UMLS. The Semantic Network contains 133 semantic types and 54 kinds of relationship between the semantic types. The Semantic Network is primarily used for the categorization of UMLS concepts [9]. All UMLS concepts are assigned at least one semantic type. The semantic relationships represent the possible relationships between semantic types, which may or may not hold true at the concept level. A third resource in the UMLS is the SPECIALIST Lexicon and the lexical tools. The SPECIALIST Lexicon is a general English lexicon that includes over 450,000 lexical items. Each lexicon entry records the syntactic, morphological and orthographic information that can be used to support activities such as natural language processing of biomedical text. The lexical tools are designed to address the high degree of variability in natural language words and terms. Normalization is one of the functions of the lexical tools that helps users to abstract away from variations involving word inflection, case and word order [53].

### **UMLS tooling**

The UMLS is distributed as a set of relational tables that can be loaded in a database management system. Alternatively, a web-based interface and an application programming interface (API) are provided. The UMLS Terminology Services (UTS) is a web-based portal that can be used for downloading UMLS data, browsing the UMLS Metathesaurus, Semantic Network and SPECIALIST Lexicon, and for accessing the UMLS documentation. Users of the UTS can enter a biomedical term or the identifier of a biomedical concept in a given ontology,

and the corresponding UMLS concept will be retrieved and displayed, showing the names for this concept in various ontologies, as well as the relations of this concept to other concepts. For example, a search on “addison’s disease” retrieves all names for the corresponding concept (C0001403) in 56 ontologies (version 2010AB, as of April 2011), including SNOMED CT, the NDF-RT and several translations of the International Classification of Primary Care. Each ontology can also be navigated as a tree. In addition to the graphical interface, the UTS also offers an application programming interface (API) based on SOAP (Simple Object Access Protocol) web services. This API provides access to the properties and relations of Metathesaurus concepts, as well as semantic types and lexical entries. Most functions of the UTS API require UMLS credentials to be checked in order to gain access to UMLS data. Support for user authentication is provided through the UTS API itself.

### **UMLS applications**

The UMLS provides convenient one-stop access to diverse biomedical vocabularies, which are updated as frequently as resources allow. One important contribution of the UMLS is that all source vocabularies are converted to a common schema of representation, with the same file structure and object model. This makes it much easier to build common tools that deal with multiple vocabularies, without the need to grapple with the native format of each. Moreover, this also enhances the understanding of the vocabularies as the common schema abstracts away from variations in naming conventions. For example, a term may be called ‘preferred name’, ‘display name’ or ‘common name’ in different vocabularies, but if they are determined to mean the same type of term functionally they are all referred to as ‘preferred term’ in the UMLS.

One common use of the UMLS is inter-terminology mapping. The UMLS concept structure enables easy identification of equivalent terms between any two source terminologies. In addition to mapping by synonymy, methods have been reported that create inter-terminology mapping by utilizing relationships and lexical resources available in the UMLS. [54] Natural language processing is another important use of the UMLS making use of its large collection of terms, the SPECIALIST Lexicon and the lexical tools. MetaMap is a publicly available tool

developed by NLM which aims to identify biomedical concepts in free text. [55, 56] This is often the first step in data-mining and knowledge discovery. Other uses of the UMLS include terminology research, information indexing and retrieval, and terminology creation. [57]

### *BioPortal*

BioPortal is developed by the National Center for Biomedical Ontology (NCBO), one of the National Centers for Biomedical Computing, created in 2004. The goal of NCBO is “to support biomedical researchers in their knowledge-intensive work, by providing online tools and a Web portal enabling them to access, review, and integrate disparate ontological resources in all aspects of biomedical investigation and clinical practice.” BioPortal not only provides access to biomedical ontologies, but it also helps link ontologies to biomedical data [58].

### **BioPortal ontologies**

The current version of BioPortal integrates over 250 ontologies for biomedicine, biology and life sciences, and includes roughly 5 million terms. A number of ontologies integrated in the UMLS are also present in BioPortal (e.g., Gene Ontology, LOINC). However, BioPortal also provides access to the ontologies from the Open Biomedical Ontologies (OBO) family, an effort to create ontologies across the biomedical domain. In addition to the Gene Ontology, OBO includes ontologies for chemical entities (e.g., ChEBI), biomedical investigations (OBI), phenotypic qualities (PATO) and anatomical ontologies for several model organism, among many others. Some of these ontologies have received the “seal of approval” of the OBO Foundry (e.g., Gene Ontology and ChEBI). Finally, the developers of biomedical ontologies can submit their resources directly to BioPortal, which makes BioPortal an open repository, as opposed to the UMLS. Examples of such resources include the African Traditional Medicine Ontology and the Electrocardiography Ontology and the Ontology of Clinical Research. BioPortal supports several popular formats for ontologies, including OWL, OBO format and the Rich Release Format (RRF) of the UMLS.

### **BioPortal tooling**

BioPortal is a web-based application allowing users to search, browse, navigate, visualize and comment on the biomedical ontologies integrated in its repository. For example, a search on “addison’s disease” retrieves the corresponding entries in 19 ontologies (as of April 2011, restricted to exact matches, including synonyms), including SNOMED CT, the Human Phenotype Ontology and DermLex. Visualization as tree or graph is offered for each ontology. The most original feature of BioPortal is to support the addition of marginal notes to various elements of an ontology, e.g., to propose new terms or suggest changes in relations. Such comments can be used as feedback by the developers of the ontologies and can contribute to the collaborative editing on ontologies. Users can also publish reviews of the ontologies. In addition to the graphical interface, BioPortal also offers an application programming interface (API) based on RESTful web services and is generally well integrated with Semantic Web technologies, as it provides URIs for each concept, which can be used as a reference in linked data applications.

### **BioPortal applications**

As the UMLS, BioPortal identifies equivalent concepts across ontologies in its repositories (e.g., between the term *listeriosis* in DermLex and in Medline Plus Health Topics). The BioPortal Annotator is a high-throughput named entity recognition system available both as an application and a web service. The Annotator identifies the names of biomedical concepts in text using fast string matching algorithms. While users can annotate arbitrary text, BioPortal also contains a list of textual resources, which have been preprocessed with the Annotator, including several gene expression data repositories, ClinicatTrials.gov and the Adverse Event Reporting System from the Food and Drug Administration (FDA). In practice, BioPortal provides an index to these resources, making it possible to use terms from its ontologies to search these resources.

### ***Approaches to ontology alignment in ontology repositories***

Apart from providing access to existing terminologies and ontologies, the UMLS and BioPortal also identify bridges between these artifacts, which will facilitate inter-ontology integration or alignment. For the UMLS, as each terminology is added or updated, every new term is comprehensively reviewed (by lexical



matching followed by manual review) to see if they are synonymous with existing UMLS terms. If so, the incoming term is grouped under the same UMLS concept. In the BioPortal, equivalence between different ontologies is discovered by a different approach. For selected ontologies, possible synonymy is identified through algorithmic matching alone (without human review). It has been shown that simple lexical matching works reasonably well in mapping between some biomedical ontologies in BioPortal, compared to more advanced algorithms [59]. Users can also contribute equivalence maps between ontologies.

### **Ontology in action – Uses of ontologies in clinical research**

To facilitate discussion, the use of ontologies and ontology-based technology in clinical research is classified into three major areas: workflow management, data integration and computer reasoning [1]. However, these are not meant to be water-tight categories (e.g. the ontological modeling of the research design can facilitate workflow management, as well as data sharing and integration).

#### *Research workflow management*

In most clinical trials, knowledge about protocols, assays and specimen flow is still stored and shared in textual documents and spreadsheets. The descriptors used are neither encoded nor standardized. Standalone computer applications are often used to automate specific portions of the research activity (e.g. trial authoring tools, operational plan builders, study site management software). These applications are largely independent and rarely communicate with each other. Integration of these systems will result in more efficient workflow management, improve the quality of the data collected and simplify subsequent data analysis. However, the lack of common terminology and semantics to describe the characteristics of a clinical trial impedes efforts of integration. Ontology-based integration of clinical trials management applications is an attractive approach. One such effort of integration resulted in the creation of CTO (described above) which has been applied successfully in the Immune Tolerance Network, a large distributed research consortium engaged in the discovery of new therapy for immune-related disorders.

Another notable effort in the use of ontology in the design and implementation of clinical trials is the Advancing Clinical Genomic Trials on Cancer (ACGT) Project in Europe. [60] ACGT is a European Union co-funded project that aims at developing open-source, semantic and grid-based technologies in support of post genomic clinical trials in cancer research. One component of this project is the development of a tool called Trial Builder to create ontology-based case report forms (CRF). The Trial Builder allows the researcher to build CRFs based on a master ontology called ACGT Master Ontology (ACGT-MO). [61] During this process, the metadata of the research is also captured which can be used in the automatic creation of the ontology-based data management system. The advantage of this approach is that the alignment of research semantics and data definition is achieved early in the research process, which guarantees easy downstream integration of data collected from disparate data sources. The early use of a common master ontology obviates the need of a *post hoc* mapping between different data and information models, which is time-consuming and error-prone.

### *Data integration*

In the post-genomic era of research, the power and potential value of linking data from disparate sources is increasingly recognized. A rapidly developing branch of translational research exploits the automated discovery of association between clinical and genomics data. [62] Ontologies can play important roles at different strategic steps of data integration. [63]

For most existing data sources, data sharing and integration only occurs as an after-thought. To align multiple data sources to support activities such as cross-study querying or data-mining is no trivial task. The classical approach, warehousing, is to align the sources at the **data** level (i.e. to annotate or index all available data by a common ontology). When the source data are encoded in different vocabularies or coding systems, which is sadly a common scenario, data integration requires alignment or mapping between the vocabularies. Resources like the UMLS and BioPortal are very useful in such mapping activity.

Another approach to data integration is to align data sources at the **metadata** level, which allows effective cross database queries without actually pooling data in a common database or warehouse.

OCRe (described above) is specifically created to annotate and align clinical trials according to their design and data analysis methodology. Another effort is BIRNLex which is created to annotate the Biomedical Informatics Research Network (BIRN) data sources. [64] The BIRN sources currently include image databases ranging from magnetic resonance imaging of human subjects, mouse models of human neurologic disease to electron microscopic imaging. BIRNLex not only covers terms in neuroanatomy, molecular species and cognitive processes, it also covers concepts such as experimental design, data types and data provenance. BIRN employs a mediator architecture to link multiple databases. The mediator integrates the various source databases by the use of a common ontology. The user query is parsed by the mediator, which issues database-specific queries to the relevant data sources each with their specific local schema. [65]

Other innovative approaches of using ontologies to achieve data integration have also been described. One study explored the possibility of tagging research data to support real-time meta-analysis. [66] Another described a prototype system for ontology-driven indexing of public data sets for translational research. [67]

One particular form of data integration supported by ontologies is represented by what has become known as “Linked Data” in the Semantic Web community [68]. The foundational idea behind linked data and the Semantic Web is that resources semantically annotated to ontologies can be interrelated when they refer to the same entities. In practice, datasets are represented as graphs in RDF, the Resource Description Framework, in which nodes (representing entities) can be shared across graphs, enabling connections among graphs. Interestingly, a significant portion of the datasets currently interrelated as Linked Data consists of biomedical resources, including PubMed, KEGG and DrugBank. For privacy reasons, very few clinical datasets have been made publicly available, and no such datasets are available as Linked Data yet. However, researchers have illustrated the benefits of Semantic Web technologies for translational research [44-47]. Moreover, the

development of personal health records will enable individuals to share their clinical data and effective de-identification techniques might also contribute to the availability of clinical data, which could enable knowledge discovery through the mining of large volume of data. Ontologies support Linked Data in three important ways. Ontologies provide a controlled vocabulary for entities in the Semantic Web; integrated ontology repositories, such as the UMLS and BioPortal, support the reconciliation of entities annotated to different ontologies; finally, relations in ontologies can be used for subsumption and other kinds of reasoning. An active community of researchers is exploring various aspects of biomedical linked data as part of the Semantic Web Health Care and Life Sciences interest group [69], with particular interest in the domain of drug discovery through the Linking Open Drug Data initiative [70].

### *Computer reasoning*

To harness the reasoning power of computers is another important reason to use ontologies in clinical research. The use of ontologies to support reasoning is not new. The Foundational Model of Anatomy (FMA) has been used to predict the anatomic consequences of penetrating injuries and the physiological consequences of injury to the arteries supplying the heart. [71-73]

The ready availability of enabling tools and utilities like Protégé, Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL) makes it easier to implement computer reasoning through the use of ontologies. One example is the use of Protégé and the accompanying SWRL Temporal Built-in Library in a study of quality standards in the management of hypertension by family practitioners. [74] Clinical research often involves chronic patients with multiple co-morbidities. Hierarchical and temporal types of queries are often necessary. Traditional data stored in relational databases cannot easily support queries involving hierarchical entities (e.g. all patients with codes related to hypertension) or temporal concepts (e.g. all patients with a lapse in anti-hypertension therapy during a certain period). This kind of queries are often necessary in clinical trials (e.g. identifying subjects that are eligibility for a particular study). As illustrated in this study, an ontology-based approach using readily available tools turned out to be a better solution.

Another area of the use of computer reasoning in clinical medicine is clinical decision support systems (CDSS). As CDSS become more widely used, it is not uncommon to find CDSS to be an important component in clinical research. CDSS often rely on ontologies to enable them to do logical reasoning. One example is ATHENA, which is an ontology-based inferencing system that encourages blood pressure control and recommends guideline-concordant choice of drug therapy in relation to co-morbid diseases. [75] The ATHENA ontology specifies eligibility criteria, risk stratification, blood pressure targets, relevant co-morbidities and preferred drugs within each drug class. One special feature of ATHENA is that clinical experts themselves can customize the knowledge base to incorporate new evidence or to reflect local interpretation of guideline ambiguities.

Looking forward, it is encouraging that the value of ontologies in clinical research becomes more recognized. This is evidenced by the increase in the number of research making use of ontologies. At the same time, this is also accompanied by an increase in the number of ontologies, which in itself is a mixed blessing. Many researchers still tend to create their own ontologies to suit their specific use case. Re-use of existing ontologies is only a rarity. If left unchecked, this tendency has the potential of growing into the very problem that ontologies are created to solve – the multitude of ontologies will itself become the barrier to data interoperability and integration. *Post hoc* mapping and alignment of ontologies is often difficult (if not impossible) and an approximation at best (with inherent information loss). The solution is to coordinate the development and maximize the re-use of existing ontologies, which will significantly simplify things downstream.

To facilitate re-use of ontologies, resources like the UMLS and BioPortal are indispensable. They enable users to navigate the expanding sea of biomedical ontologies. In addition to listing and making these ontologies available, what is still lacking is a better characterization of these ontologies to help users decide whether they are suitable for the tasks at hand. In case there are multiple candidate ontologies, some indicators of quality (e.g. user base, ways in which they are used, user feedback and comments) will be very useful to help users decide on the best choice.

## Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM).

## References

1. Bodenreider, O.: Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform*, 67-79 (2008)
2. Smith, B.: Ontology (Science). *Nature Precedings* Available from *Nature Precedings* (<http://hdl.handle.net/10101/npre.2008.2027.2>) (2008)
3. Bodenreider, O., Stevens, R.: Bio-ontologies: current trends and future directions. *Brief Bioinform* 7, 256-274 (2006)
4. Cimino, J.J., Zhu, X.: The practical impact of ontologies on biomedical informatics. *Yearb Med Inform*, 124-135 (2006)
5. Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L., Rosse, C.: Relations in biomedical ontologies. *Genome Biol* 6, R46 (2005)
6. Simmons, P., Melia, J.: Continuants and occurrents. *Proceedings of the Aristotelian Society, Supplementary Volumes* 74, 59-75+77-92 (2000)
7. BFO, <http://www.ifomis.org/bfo/>
8. DOLCE, <http://www.loa-cnr.it/DOLCE.html>
9. McCray, A.T.: An upper-level ontology for the biomedical domain. *Comp Funct Genomics* 4, 80-84 (2003)
10. Beißwanger, E., Schulz, S., Stenzhorn, H., Hahn, U.: BioTop: An upper domain ontology for the Life Sciences - A description of its current structure, contents, and interfaces to OBO ontologies. *Applied Ontology* 3, 205-212 (2008)
11. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): *The description logic handbook : theory, implementation, and applications*. Cambridge University Press, Cambridge New York (2007)
12. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* 284, 34-43 (2001)
13. OWL 2 Web Ontology Language Document Overview, <http://www.w3.org/TR/owl2-overview/>
14. RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>
15. SKOS Simple Knowledge Organization System Reference, <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
16. The OBO Flat File Format Specification, [http://www.geneontology.org/GO.format.obo-1\\_2.shtml](http://www.geneontology.org/GO.format.obo-1_2.shtml)
17. Golbreich, C., Horridge, M., Horrocks, I., Motik, B., Shearer, R.: OBO and OWL: leveraging semantic web technologies for the life sciences. *Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference*. Springer-Verlag, Busan, Korea (2007) 169-182

18. Noy, N., Tudorache, T., Nyulas, C., Musen, M.: The ontology life cycle: Integrated tools for editing, publishing, peer review, and evolution of ontologies. *AMIA Annu Symp Proc* 2010, 552-556 (2010)
19. Protégé, <http://protege.stanford.edu/>
20. Day-Richter, J., Harris, M.A., Haendel, M., Lewis, S.: OBO-Edit--an ontology editor for biologists. *Bioinformatics* 23, 2198-2200 (2007)
21. OBO-Edit, <http://oboedit.org/>
22. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 25, 1251-1255 (2007)
23. Harmonization, <http://www.ihtsdo.org/about-ihtsdo/harmonization/>
24. Richesson, R.L., Krischer, J.: Data standards in clinical research: gaps, overlaps, challenges and future directions. *J Am Med Inform Assoc* 14, 687-696 (2007)
25. Shankar, R.D., Martins, S.B., O'Connor, M., Parrish, D.B., Das, A.K.: An ontology-based architecture for integration of clinical trials management applications. *AMIA Annu Symp Proc*, 661-665 (2007)
26. Shankar, R., Arkalgud, S., Connor, M., Boyce, K., Parrish, D., Das, A.: TrialWiz: an Ontology-Driven Tool for Authoring Clinical Trial Protocols. *AMIA Annu Symp Proc*, 1226 (2008)
27. Tu, S.W., Fridsma, D.B., Shankar, R., Connor, M., Das, A., Parrish, D.: Bridging Epoch: Mapping Two Clinical Trial Ontologies. 10th International Protege Conference (2007)
28. Tu, S.W., Carini, S., Rector, A., Maccalum, P., Toujilov, I., Harris, S., Sim, I.: OCR: Ontology of Clinical Research. 11th International Protege Conference (2009)
29. The OBI Consortium, <http://obi-ontology.org/page/Consortium>
30. Whetzel, P.L., Brinkman, R.R., Causton, H.C., Fan, L., Field, D., Fostel, J., Fragoso, G., Gray, T., Heiskanen, M., Hernandez-Boussard, T., Morrison, N., Parkinson, H., Rocca-Serra, P., Sansone, S.A., Schober, D., Smith, B., Stevens, R., Stoeckert, C.J., Jr., Taylor, C., White, J., Wood, A.: Development of FuGO: an ontology for functional genomics investigations. *Omics* 10, 199-204 (2006)
31. Brinkman, R.R., Courtot, M., Derom, D., Fostel, J.M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Soldatova, L.N., Stoeckert, C.J., Jr., Turner, J.A., Zheng, J.: Modeling biomedical experimental processes with OBI. *J Biomed Semantics* 1 Suppl 1, S7 (2010)
32. de Coronado, S., Haber, M.W., Sioutos, N., Tuttle, M.S., Wright, L.W.: NCI Thesaurus: using science-based terminology to integrate cancer research results. *Medinfo* 11, 33-37 (2004)
33. Sioutos, N., de Coronado, S., Haber, M.W., Hartel, F.W., Shaiu, W.L., Wright, L.W.: NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 40, 30-43 (2007)
34. Fragoso, G., de Coronado, S., Haber, M., Hartel, F., Wright, L.: Overview and Utilization of the NCI Thesaurus. *Comp Funct Genomics* 5, 648-654 (2004)

35. SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms)  
<http://www.ihtsdo.org/our-standards/>
36. Blumenthal, D., Tavenner, M.: The "meaningful use" regulation for electronic health records. *N Engl J Med* 363, 501-504 (2010)
37. Office of the National Coordinator for Health Information Technology (ONC) - Department of Health and Human Services: Standards & Certification Criteria Interim Final Rule: Revisions to Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology. *Federal Register* 75, 62686-62690 (2010)
38. Brown, S.H., Elkin, P.L., Rosenbloom, S.T., Husser, C., Bauer, B.A., Lincoln, M.J., Carter, J., Erlbaum, M., Tuttle, M.S.: VA National Drug File Reference Terminology: a cross-institutional content coverage study. *Stud Health Technol Inform* 107, 477-481 (2004)
39. Rosenbloom, S.T., Awad, J., Speroff, T., Elkin, P.L., Rothman, R., Spickard, A., 3rd, Peterson, J., Bauer, B.A., Wahner-Roedler, D.L., Lee, M., Gregg, W.M., Johnson, K.B., Jirjis, J., Erlbaum, M.S., Carter, J.S., Lincoln, M.J., Brown, S.H.: Adequacy of representation of the National Drug File Reference Terminology Physiologic Effects reference hierarchy for commonly prescribed medications. *AMIA Annu Symp Proc*, 569-578 (2003)
40. National Drug File Reference Terminology,  
<ftp://ftp1.nci.nih.gov/pub/cacore/EVS/NDF-RT/>
41. National Library of Medicine: RxNav.
42. Cimino, J.J., Ayres, E.J.: The clinical research data repository of the US National Institutes of Health. *Stud Health Technol Inform* 160, 1299-1303 (2010)
43. Lowe, H.J., Ferris, T.A., Hernandez, P.M., Weber, S.C.: STRIDE--An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009, 391-395 (2009)
44. Ruttenberg, A., Clark, T., Bug, W., Samwald, M., Bodenreider, O., Chen, H., Doherty, D., Forsberg, K., Gao, Y., Kashyap, V., Kinoshita, J., Luciano, J., Marshall, M.S., Ogbuji, C., Rees, J., Stephens, S., Wong, G.T., Wu, E., Zaccagnini, D., Hongsermeier, T., Neumann, E., Herman, I., Cheung, K.H.: Methodology - Advancing translational research with the Semantic Web. *Bmc Bioinformatics* 8, - (2007)
45. McCusker, J.P., Phillips, J.A., Gonzalez Beltran, A., Finkelstein, A., Krauthammer, M.: Semantic web data warehousing for caGrid. *Bmc Bioinformatics* 10 Suppl 10, S2 (2009)
46. Sahoo, S.S., Bodenreider, O., Rutter, J.L., Skinner, K.J., Sheth, A.P.: An ontology-driven semantic mashup of gene and biological pathway information: application to the domain of nicotine dependence. *J Biomed Inform* 41, 752-765 (2008)
47. Translational Medicine Ontology and Knowledge Base,  
<http://www.w3.org/wiki/HCLSIG/PharmaOntology>
48. Humphreys, B.L., Lindberg, D.A., Hole, W.T.: Assessing and enhancing the value of the UMLS Knowledge Sources. *Proc Annu Symp Comput Appl Med Care*, 78-82 (1991)
49. Humphreys, B.L., Lindberg, D.A., Schoolman, H.M., Barnett, G.O.: The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc* 5, 1-11 (1998)



50. Lindberg, D.A., Humphreys, B.L., McCray, A.T.: The Unified Medical Language System. *Methods Inf Med* 32, 281-291 (1993)
51. Bodenreider, O.: The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res* 32 Database issue, D267-270 (2004)
52. Unified Medical Language System (UMLS), <http://www.nlm.nih.gov/research/umls/>
53. McCray, A.T., Srinivasan, S., Browne, A.C.: Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care*, 235-239 (1994)
54. Fung, K.W., Bodenreider, O.: Utilizing the UMLS for semantic mapping between terminologies. *AMIA Annu Symp Proc*, 266-270 (2005)
55. Aronson, A.R.: Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp*, 17-21 (2001)
56. Aronson, A.R., Lang, F.M.: An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 17, 229-236 (2010)
57. Fung, K.W., Hole, W.T., Srinivasan, S.: Who is using the UMLS and how - insights from the UMLS user annual reports. *AMIA Annu Symp Proc*, 274-278 (2006)
58. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 37, W170-173 (2009)
59. Ghazvinian, A., Noy, N.F., Musen, M.A.: Creating mappings for ontologies in biomedicine: simple methods work. *AMIA Annu Symp Proc* 2009, 198-202 (2009)
60. Weiler, G., Brochhausen, M., Graf, N., Schera, F., Hoppe, A., Kiefer, S.: Ontology based data management systems for post-genomic clinical trials within a European Grid Infrastructure for Cancer Research. *Conf Proc IEEE Eng Med Biol Soc* 2007, 6435-6438 (2007)
61. ACGT Master Ontology, [http://www.ifomis.org/wiki/ACGT\\_Master\\_Ontology\\_%28MO%29](http://www.ifomis.org/wiki/ACGT_Master_Ontology_%28MO%29)
62. Genome-Wide Association Studies, <http://grants.nih.gov/grants/gwas/>
63. Bodenreider, O.: Ontologies and data integration in biomedicine: Success stories and challenging issues. In: Bairoch, A., Cohen-Boulakia, S., Froidevaux, C. (eds.): *Proceedings of the Fifth International Workshop on Data Integration in the Life Sciences (DILS 2008)*, Vol. LNBI 5109, pp. 1-4. Springer, Berlin Heidelberg New York (2008)
64. Biomedical Informatics Research Network, <https://xwiki.nbirn.org:8443/xwiki/bin/view/+BIRN-OTF-Public/About+BIRNLex>
65. Rubin, D.L., Shah, N.H., Noy, N.F.: Biomedical ontologies: a functional perspective. *Brief Bioinform* 9, 75-90 (2008)
66. Cook, C., Hannley, M., Richardson, J.K., Michon, J., Harker, M., Pietrobon, R.: Real-time updates of meta-analyses of HIV treatments supported by a biomedical ontology. *Account Res* 14, 1-18 (2007)
67. Shah, N.H., Jonquet, C., Chiang, A.P., Butte, A.J., Chen, R., Musen, M.A.: Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics* 10 Suppl 2, S1 (2009)

68. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. *Int J Semant Web Inf* 5, 1-22 (2009)
69. HCLS: Semantic Web Health Care and Life Sciences (HCLS) Interest Group.
70. Linking Open Drug Data, <http://www.w3.org/wiki/HCLSIG/LODD>
71. Rosse, C., Shapiro, L.G., Brinkley, J.F.: The digital anatomist foundational model: principles for defining and structuring its concept domain. *Proc AMIA Symp*, 820-824 (1998)
72. Rubin, D.L., Dameron, O., Bashir, Y., Grossman, D., Dev, P., Musen, M.A.: Using ontologies linked with geometric models to reason about penetrating injuries. *Artif Intell Med* (2006)
73. Rubin, D.L., Dameron, O., Musen, M.A.: Use of description logic classification to reason about consequences of penetrating injuries. *AMIA Annu Symp Proc*, 649-653 (2005)
74. Mabotuwana, T., Warren, J.: An ontology-based approach to enhance querying capabilities of general practice medicine for better management of hypertension. *Artif Intell Med* 47, 87-103 (2009)
75. Goldstein, M.K., Hoffman, B.B., Coleman, R.W., Tu, S.W., Shankar, R.D., O'Connor, M., Martins, S., Advani, A., Musen, M.A.: Patient safety in guideline-based decision support for hypertension management: ATHENA DSS. *Proc AMIA Symp*, 214-218 (2001)