

DRAFT COPY

[The definitive version of the article can be purchased at <http://www.eurekaselect.com/109232/article>]

Ontologies in Medicinal Chemistry: Current Status and Future Challenges

Asunción Gómez-Pérez¹, [Marcos Martínez-Romero](#)^{2,*}, [Alejandro Rodríguez-González](#)³,
Guillermo Vázquez⁴ and [José M. Vázquez-Naya](#)²

¹Artificial Intelligence Department, Computer Science Faculty, Polytechnic University of Madrid, 28660 Boadilla del Monte, Madrid, Spain, ²Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain, ³Centre for Plant Biotechnology and Genomics UPM-INIA, Polytechnic University of Madrid, Parque Científico y Tecnológico de la U.P.M., Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain, ⁴Institute of Biomedical Research of A Coruña (INIBIC), Xubias de Arriba 84, Hospital Materno Infantil (1ª planta), 15006 A Coruña, Spain.

E-mail addresses:

asun@fi.upm.es, marcosmartinez@udc.es, alejandro.rodriguez@upm.es, Guillermo.Vazquez.Gonzalez@sergas.es,
jmvazquez@udc.es

* Address correspondence to this author at the Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña, Campus de Elviña, S/N, 15071 A Coruña, Spain; Tel: +34 981 167 000 Ext. 2667; Fax: +34 981 167 172; E-mail: marcosmartinez@udc.es

Abstract: Recent years have seen a dramatic increase in the amount and availability of data in the diverse areas of medicinal chemistry, making it possible to achieve significant advances in fields such as the design, synthesis and biological evaluation of compounds. However, with this data explosion, the storage, management and analysis of available data to extract relevant information has become even a more complex task that offers challenging research issues to Artificial Intelligence (AI) scientists. Ontologies have emerged in AI as a key tool to formally represent and semantically organize aspects of the real world. Beyond glossaries or thesauri, ontologies facilitate communication between experts and allow the application of computational techniques to extract useful information from available data. In medicinal chemistry, multiple ontologies have been developed during the last years which contain knowledge about chemical compounds and processes of synthesis of pharmaceutical products. This article reviews the principal standards and ontologies in medicinal chemistry, analyzes their main applications and suggests future directions.

Key words: drug design, drug discovery, medicinal chemistry, ontologies, Semantic Web

1 INTRODUCTION

The past several years have seen a sudden increase in the amount of publicly available data throughout different fields of the natural sciences. In areas that involve aspects of multiple disciplines, such as medicinal chemistry, this flood of data and knowledge is considered crucial to enable research through complex data-mining and knowledge discovery methods and develop new biologically active compounds for therapeutic use. However, this massive amount of data are frequently distributed over the Web and organized using different terminologies and formats, which constitutes an important barrier to process and analyze them in an integrated manner. As a consequence, retrieving relevant information from the available sources has become much more difficult. This situation demands the development and application of innovative strategies addressed to properly organize, retrieve and share the huge amount of available data in medicinal chemistry and its related fields.

The word ontology is derived from Greek and refers to the philosophical study of the nature of being and existence. Ontologies were originally used by ancient Greek philosophers such as Aristotle in his *Metaphysics* [1] to name and classify the things they saw in the Universe and the relationships between them. More recently, in the 1990s, and especially since the Semantic Web was conceived [2], computer scientists have acquired the principles of philosophical ontology to represent the knowledge of some areas of interest (e.g. biology, medicine, chemistry, etc.) in a standard and machine-processable manner [3]. In computer science, ontologies are models for describing the world that consist of a set of types, properties, and relationships. An ontology can be graphically represented as a directed graph where nodes correspond to concepts, also known as entities or classes (e.g. “Biological agent”, “Specimen processing”, “Central nervous system deficit”), and arcs correspond to relationships between concepts (e.g. “is a”, “part of”). This graph-oriented representation enables the application of powerful techniques from the field of topological graph theory and other areas of mathematics that have been demonstrated to be useful in fields such as Medical Chemistry, Oncology, Neurology, Proteomics, Genomics, Microbiology, Technology, Sociology or Laws [4-8]. One example in Medicinal Chemistry are the 2D/3D protein graphs, used to predict protein biological functions [9-12]. Apart from its graphical representation, different languages have been proposed to represent ontologies in a machine-readable form [13]. Currently, the most popular and widely used ontology language is the Web Ontology Language (OWL) [14, 15], proposed by the World Wide Web Consortium (W3C) and based on the Resource Description Framework (RDF) [16]. OWL is the recommended language for building ontologies for the Semantic Web and provides computational reasoning capabilities across ontologies that can be used to infer new knowledge from existing data.

In areas such as biology, medicine, chemistry and pharmacology, the necessity of structuring, sharing and reusing the huge amount of data about genes, proteins, diseases, molecular functions, etc. that has been generated during the last years, has led to the development of numerous ontologies that have demonstrated to be useful to support a variety of applications, such as data annotation [17] and integration [18], semantic search [19], terminology unification and standardization [20], information extraction [21], automated reasoning [22], development of expert systems [23], data mining [24] and prediction of protein-protein interactions [25]. In these fields, as the number and variety of available ontologies

continues steadily growing, so does the number of initiatives aimed to index, classify and organize them for researchers and knowledge engineers. Some of the largest and most referenced ontology repositories are the NCBO's BioPortal [26], which contains more than 300 well-structured and controlled biomedical ontologies and terminologies, the EBI Ontology Lookup Service (OLS) [27] and the OBO Foundry portal [28], which contains a set of reference biomedical ontologies that adhere to the good ontology design principles proposed by the OBO Foundry initiative [29].

This review is intended to provide researchers and developers with a summary of the current ontology landscape for medicinal chemistry. To do so, we will present the most popular ontologies and reference terminologies in chemistry and drug development. We will also provide some future trends in the field.

2 ONTOLOGIES IN MEDICINAL CHEMISTRY

Medicinal chemistry is a multidisciplinary area that began to emerge about 150 years ago and that has been defined as *"the applied science that is focused on the design and discovery of new chemical entities and their optimization and development as useful drug molecules for the treatment of disease processes"* [30]. Traditionally most of the known information on the diverse areas of medicinal chemistry has been stored in books, journals and commercial databases. Nevertheless, during the last decade this situation changed dramatically. The discovery and application of new information and communication technologies and innovative computing strategies has dramatically increased the volume and availability of data about molecules, structures, proteins, sequences, drug targets and drug actions. Ontologies have been shown to be extremely useful to assist in the classification and organization of these data, to unify definitions and to facilitate data exchange, analysis and searching.

Medicinal chemistry is a discipline that involves aspects of different areas such as chemistry, medicine or drug discovery and development. In fact, most medicinal chemists work with a team of scientists from different disciplines, including biologists, toxicologists, pharmacologists, theoretical chemists, microbiologists and biopharmacists [31]. As a consequence, a medicinal chemistry researcher who wants to use ontologies to enhance the usability of his data or a knowledge engineer aimed to build a system that manages medicinal chemistry information, should take into account a variety of ontologies from different fields.

Keeping in mind that this review is intended to report the most relevant ontologies in medicinal chemistry, we will start examining the information provided by existing ontologies for the term "medicinal chemistry". This term can be found in two popular biomedical ontologies: Medical Subject Headings (MeSH) and the National Cancer Institute (NCI) Thesaurus. MeSH [32] is a classification of terms created and maintained by the U.S. National Library of Medicine (NLM) for purposes of indexing and classifying medical literature. The NCI Thesaurus [33], developed by the U.S. National Cancer Institute, is an ontology providing broad coverage of the cancer domain, including cancer-related diseases, findings and abnormalities. Fig. 1 shows the location of "medicinal chemistry" both in the hierarchy of classes of MeSH and the NCI Thesaurus.

In MeSH, the concept is represented using the term "Chemistry, Pharmaceutical", which is classified as a children of "Pharmacology" and "Chemistry" categories, and as a superclass of "Drug Discovery". The definition provided by MeSH is *"Chemistry dealing with the composition and preparation of agents having pharmacologic actions or diagnostic use"*.

The NCI Thesaurus provides a more comprehensive definition of this term. In this ontology, “Medicinal Chemistry” is a leaf node classified under the “Pharmacology” category that is defined as “A chemistry-based discipline, also involving aspects of biological, medical and pharmaceutical sciences. It is concerned with the invention, discovery, design, identification and preparation of biologically active compounds, the study of their metabolism, the interpretation of their mode of action at the molecular level and the construction of structure-activity relationships”.

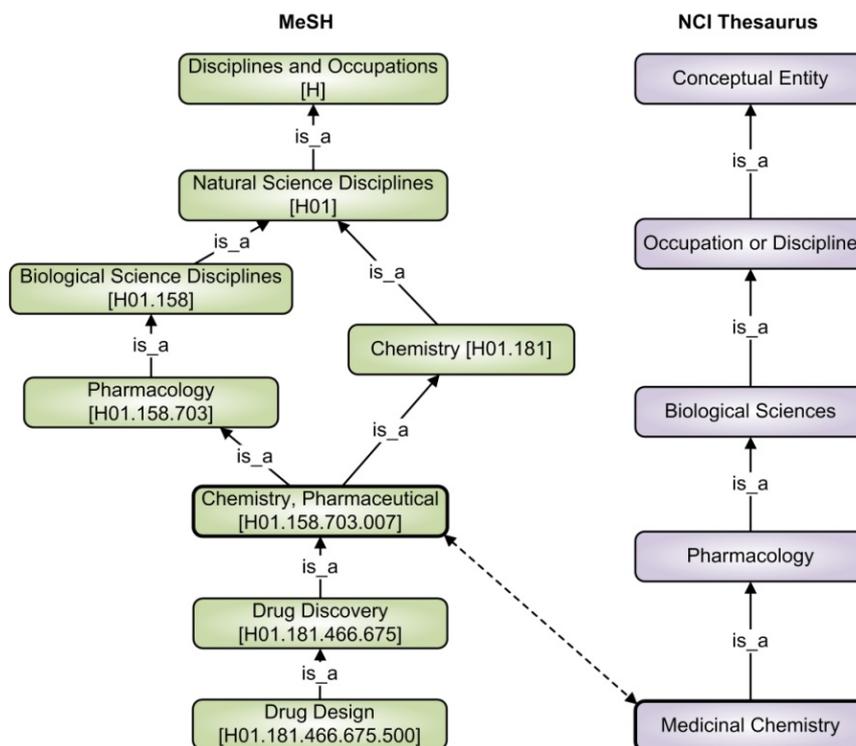


Fig. 1. Location of the concept “Medicinal Chemistry” in the hierarchy of classes of MeSH and the NCI Thesaurus.

In the following sections, we will firstly present the most popular ontologies about chemical compounds. Then, we will review the ontologies addressed to classify pharmaceutical agents, or drugs. Finally, ontologies aimed to represent the essential entities and techniques in drug discovery, drug design, drug development and other related disciplines will be reported.

2.1 Classification of chemical compounds

Extracting, identifying, naming, classifying and exchanging information on compounds have been a major process of evolution of human civilization over tens of thousands of years. Initially, elements were classified by properties related to taste and smell, like “salty”, “sweet” or “bitter”. With the passing of time and in parallel with the discoveries made in chemistry, these classifications have become increasingly complex and detailed, giving rise to huge hierarchies of compounds and relationships between them.

One of the first examples of using ontologies to classify chemical compounds was **Medical Subject Headings (MeSH)** [32], already mentioned in the previous section. The first official list of subject headings published by the NLM appeared in 1954 under the title Subject Heading Authority List. MeSH was initially developed for purposes of indexing and classifying medical literature, and nowadays it is used to facilitate searches in many resources. For example, the most popular query interfaces to MedLine

(e.g. PubMed [34]) use MeSH to support the retrieval of MedLine records in ways which improve the use of simple string searches. As it is shown in Fig. 2, MeSH is structured under 16 top-level categories that represent knowledge from different domains related to medicine. The “D” subtree (Chemicals and Drugs) contains chemical classes, individual compounds and biological concepts. Originally in English, MeSH has been translated into German, Japanese, Portuguese, Spanish, French, Russian, Polish, Romanian, Arabic, and other languages. The translations enable users not facile in English to find documents that are of sufficient potential interest to warrant reading.

1. + Anatomy [A]
2. + Organisms [B]
3. + Diseases [C]
4. - Chemicals and Drugs [D]
 - o [Inorganic Chemicals \[D01\]](#) +
 - o [Organic Chemicals \[D02\]](#) +
 - o [Heterocyclic Compounds \[D03\]](#) +
 - o [Polycyclic Compounds \[D04\]](#) +
 - o [Macromolecular Substances \[D05\]](#) +
 - o [Hormones, Hormone Substitutes, and Hormone Antagonists \[D06\]](#) +
 - o [Enzymes and Coenzymes \[D08\]](#) +
 - o [Carbohydrates \[D09\]](#) +
 - o [Lipids \[D10\]](#) +
 - o [Amino Acids, Peptides, and Proteins \[D12\]](#) +
 - o [Nucleic Acids, Nucleotides, and Nucleosides \[D13\]](#) +
 - o [Complex Mixtures \[D20\]](#) +
 - o [Biological Factors \[D23\]](#) +
 - o [Biomedical and Dental Materials \[D25\]](#) +
 - o [Pharmaceutical Preparations \[D26\]](#) +
 - o [Chemical Actions and Uses \[D27\]](#) +
5. + Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. + Psychiatry and Psychology [F]
7. + Phenomena and Processes [G]
8. + Disciplines and Occupations [H]
9. + Anthropology, Education, Sociology and Social Phenomena [I]
10. + Technology, Industry, Agriculture [J]
11. + Humanities [K]
12. + Information Science [L]
13. + Named Groups [M]
14. + Health Care [N]
15. + Publication Characteristics [V]
16. + Geographicals [Z]

Fig. 2. Detail of the “Chemical and Drugs” subtree in MeSH. Screenshot from the MeSH Browser [35].

ChEBI (Chemical Entities of Biological Interest) [36-39] is the currently most comprehensive chemical compounds ontology. It is a freely available dictionary and ontological classification of molecular entities and groups, developed and maintained by the European Bioinformatics Institute (EBI) and focused on “small” chemical compounds. The term “molecular entity” refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity [40]. A group is a defined linked collection of atoms or a single atom within a molecular entity [41]. ChEBI provides a high quality, systematically annotated controlled vocabulary to promote the correct and consistent use of unambiguous biochemical terminology throughout the molecular biology databases at the EBI. It contains not only biochemical compounds, but also pharmaceuticals, agrochemicals, laboratory reagents, isotopes and subatomic particles. Since its first public release (21 July 2004), ChEBI has evolved to represent more than 30,000 molecular entities, groups and classes. The ontology structure is essentially a directed acyclic graph (DAG) and comprises three separate sub-ontologies: (1) Chemical Entity, which classifies molecular entities according to structure; (2) Role, which classifies entities on the basis of their role within an application, biological or

chemical context (e.g. pesticide, placebo, antioxidant, etc.); and (4) Subatomic Particle, which classifies particles smaller than atoms. An interesting application of ChEBI is ARISTO (Automatic Reduction of Ion Spectra To Ontology), a webtool that provides assignments to a subset of ChEBI using a mass spectrum of compounds as input [42]. Chemical ontologies such as ChEBI also facilitate new ways of knowledge discovery, for example, by extracting relationships between compound classes and related data from other domains, which are traditionally known as quantitative structure-activity relationships (QSAR) or quantitative structure–property relationships (QSPR) [5, 43-47].

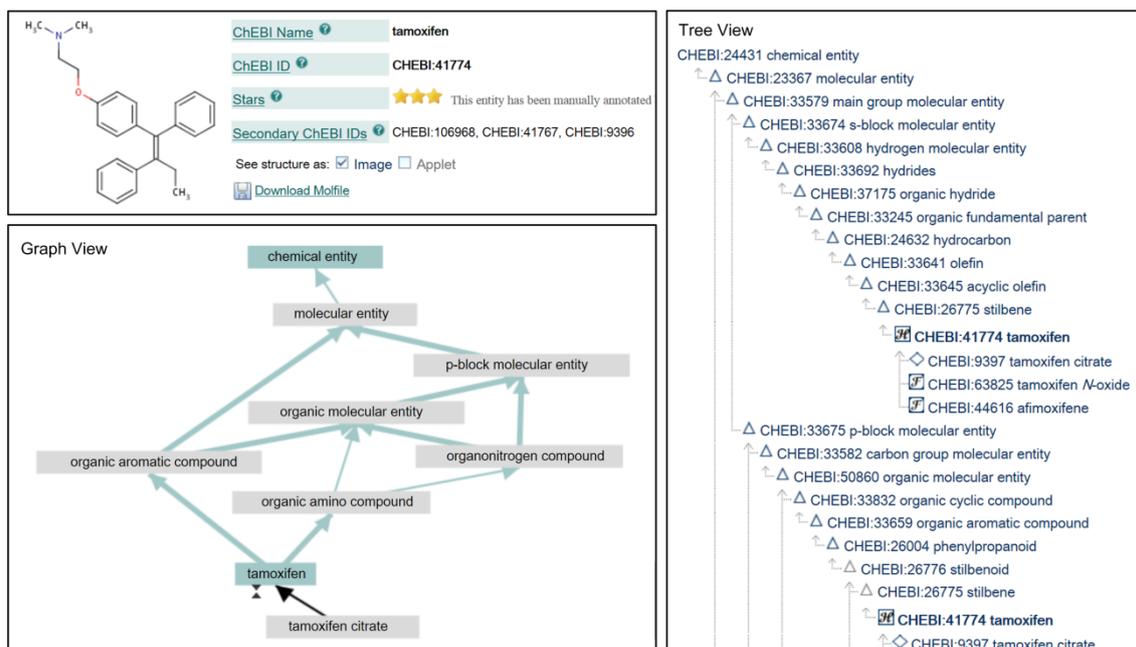


Fig. 3. Information about the drug “tamoxifen” in ChEBI. The figure shows a graphical representation of the drug and two different views (graph and tree) of its location in the hierarchy of classes.

Fig. 3 shows some of the information provided by ChEBI for the chemical entity “tamoxifen”, which is a drug commonly used in breast cancer treatment. In ChEBI, each time a new compound is added, it must be assigned to the ontology by hand, which is a very labour-intensive and time-consuming process.

A well-known effort aimed to automatically categorize small molecules instead of doing it by hand is the **Chemical Ontology (CO)** [48], based on the idea that small molecules can be classified by identifying their chemical functional groups using objective and computable criteria. CO has been built according to the functional groups assigned by the Checkmol software [49], a command-line utility program that is able to analyze an input molecule for the presence of various functional groups and structural elements. CO can be used to search chemical databases and identifying key functional groups responsible for biological activities.

Another effort addressed to represent chemical structure and properties is the **Chemical Information Ontology (CHEMINF)**. It has been conceived by several groups who have independently developed terminologies in the domain of chemical information. They have created a semantic chemistry working group [50] to develop CHEMINF, which is a unified, coherent OWL ontology to encode the terms, definitions, and logical axioms of chemical information entities [51]. The basic content of CHEMINF can be divided into named descriptors, named algorithms which calculate descriptors, and software libraries,

which contain software modules that implement algorithms. Fig. 4 shows the subclasses provided by CHEMINF ontology for the class "chemical descriptor".

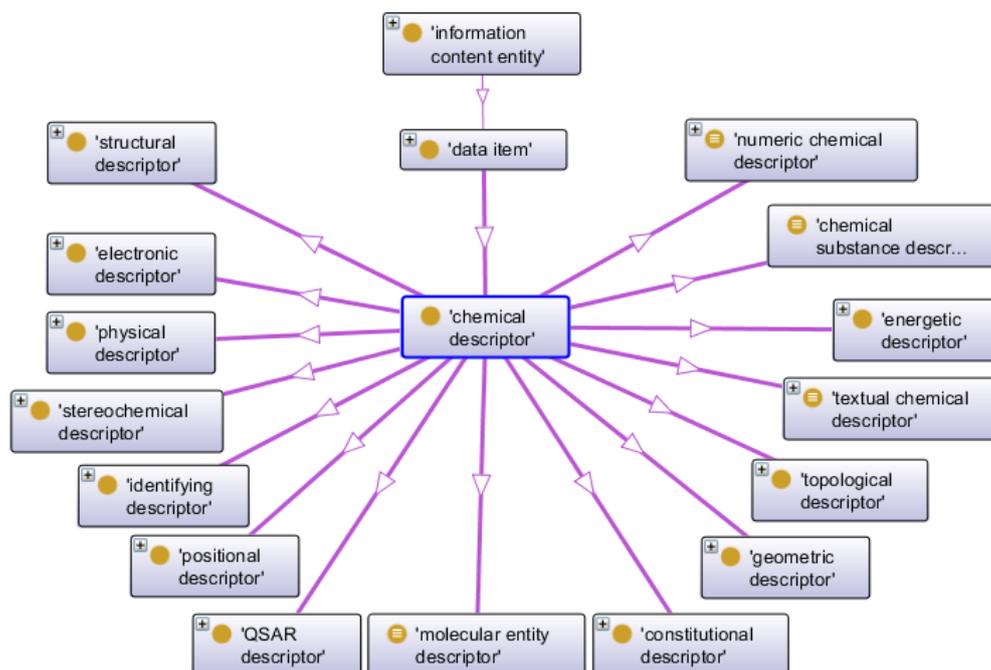


Fig. 4. Fragment of CHEMINF showing the subclasses of "chemical descriptor". Arrows represent hierarchical relationships between classes.

An **Ontology for Pharmaceutical Ligands** has also been reported [52], which allows to bridge cheminformatics and bioinformatics by linking ligands to sequences. This ontology enables annotation-based searching of ligands and sequence similarity-based identification of ligands of next homologous receptors, and is key for further systematic chemogenomics exploration of previously well explored target families.

Another relevant initiative is **ChemTop** [53], a top-domain, OWL-DL ontology that was created as an extract of BioTop [54] and that describes the foundational entities needed to characterize phenomena in the domain of chemistry and to interface both top and domain ontologies (e.g. EntireMolecularEntity, FattyAcid, NucleicAcid, etc.). ChemTop is based on the top ontology BFO (Basic Formal Ontology) [55] and on the relationships defined in RO (the Relation Ontology) [56]. As a consequence, it can be used as a top-level ontology for domain ontologies such as ChEBI.

Other notable ontologies in the chemistry domain are **REX** [57] and **FIX** [58]. REX models physico-chemical processes occurring in the course of time. It includes both microscopic processes, involving molecular entities or subatomic particles, and macroscopic ones. Some biochemical processes from the popular Gene Ontology [59, 60] (GO Biological process) can be described as instances of REX. FIX complements REX and is an ontology of physico-chemical methods and properties.

There are also some chemical ontologies that have been developed to represent specific subfields of chemistry. An example of ontology for representing lipid structures is **LIPID MAPS** [61, 62]. Another example is **GlycO (the Glycomics Ontology)** [63], which focuses on the glycoproteomics domain and models the structure and functions of glycans and glycoconjugates, the enzymes involved in their biosynthesis and modification, and the metabolic pathways in which they participate.

2.2 Classification of drugs

Hospitals, pharmacies, research centers and other organizations use computer systems to record and process drug information. In order to ensure that these systems can exchange drug-related information efficiently and unambiguously, it is necessary to use a unified set of drug names. Many clinical information tasks can benefit from the use of a standard terminology for representing drug information, such as creation and management of electronic health records (EHR), communication between groups of experts, supporting interoperation between information systems, healthcare research, etc. [64].

RxNorm [65, 66] is a popular source created by the U.S. National Library of Medicine (NLM) to provide a single, standard, multipurpose terminology for representing medications, which has been built by aggregating, organizing and reconciling content from various source vocabularies. The first release of RxNorm date back to November 2004. Since then, it has been frequently used to navigate across drug vocabularies, exchange drug names and drug codes and as a reference to provide decision support in clinical environments. RxNorm provides links from clinical drugs to their active ingredients, drug components and related brand names. It also contains normalized names for clinical drugs and links their names to many drug terminologies regularly used in pharmacy management and drug interaction software.

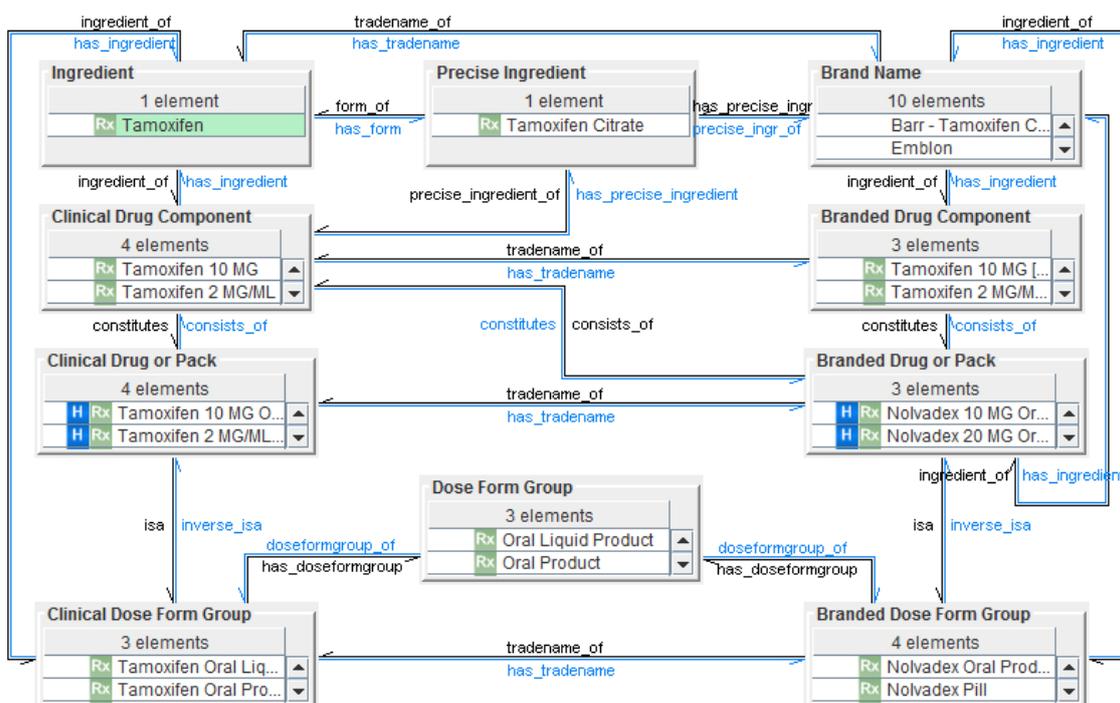


Fig. 5. Partial screenshot of RxNav showing the information provided by RxNorm for the drug Tamoxifen.

To facilitate and enrich the use of RxNorm, the NLM provides several services which can be helpful to understand the RxNorm content, to facilitate the access to the RxNorm data and to make it easier to integrate the RxNorm vocabulary into specific applications or systems. One of these services is RxNav [67, 68], the RxNorm navigator, which is a Java application that allows to browse RxNorm and other sources, such as RxTerms and the National Drug File-Reference Terminology (NDF-RT) in a visually friendly and interactive manner. RxNav displays links from clinical drugs, both branded and generic, to their active ingredients, drug components and related brand names. Fig. 5 shows a screenshot of RxNav for the drug Tamoxifen. The NLM also provides an Application Programming Interface (API) for RxNav,

which allows to develop new applications based on data retrieved from the RxNorm database. Other useful services provided for RxNorm are RxTerms [69], MyMedicationList [70], MyRxPad [71] and RxMix. Current version of RxNorm is based on 11 different external source vocabularies, which are listed in Table 1. Apart from being part of RxNorm, some of these sources are widely-recognised vocabularies that are being used in a variety of medical applications around the world.

Table 1. RxNorm data sources.

Source	Refs.
Gold Standard Drug Database	[72]
Medi-Span Master Drug Data Base	[73]
Medical Subject Headings (MeSH)	[32]
Multum MediSource Lexicon	[74]
Micromedex RED BOOK	[75]
FDA National Drug Code Directory	[76]
FDA Structured Product Labels	[77]
FDB MedKnowledge (formerly NDDF Plus)	[78]
Veterans Health Administration National Drug File - Reference Terminology	[79]
SNOMED Clinical Terms (drug information)	[80]
Veterans Health Administration National Drug File	[81]

One of the sources that are part of RxNorm is Medical Subject Headings (MeSH) [32]. As explained in section 2.1, MeSH contains a subtree specifically addressed to classify “Chemicals and Drugs” into 16 categories, such as: inorganic chemicals, organic chemicals, polycyclic compounds, enzymes and coenzymes, complex mixtures, etc. Fig. 6 shows the location of the drug Tamoxifen in the MeSH hierarchy.

[Organic Chemicals \[D02\]](#)

[Hydrocarbons \[D02.455\]](#)

[Hydrocarbons, Cyclic \[D02.455.426\]](#)

[Hydrocarbons, Aromatic \[D02.455.426.559\]](#)

[Benzene Derivatives \[D02.455.426.559.389\]](#)

[Benzylidene Compounds \[D02.455.426.559.389.150\]](#)

[Stilbenes \[D02.455.426.559.389.150.700\]](#)

[4-Acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic Acid \[D02.455.426.559.389.150.700.050\]](#)

[Bibenzyls \[D02.455.426.559.389.150.700.075\]](#)

[Chlorotrianisene \[D02.455.426.559.389.150.700.100\]](#)

[Clomiphene \[D02.455.426.559.389.150.700.125\]](#)

[Diethylstilbestrol \[D02.455.426.559.389.150.700.175\] +](#)

[4,4'-Diisothiocyanatostilbene-2,2'-Disulfonic Acid](#)

[\[D02.455.426.559.389.150.700.200\]](#)

[Stilbamidines \[D02.455.426.559.389.150.700.550\]](#)

▶ [Tamoxifen \[D02.455.426.559.389.150.700.900\]](#)

[Raloxifene \[D02.455.426.559.389.150.700.900.775\]](#)

[Toremifene \[D02.455.426.559.389.150.700.900.900\]](#)

Fig. 6. Location of Tamoxifen in the MeSH tree. The root of the tree (“Organic Chemicals [D02]”) is one of the 16 children of the top-level category “Chemicals and Drugs [D]”. Screenshot from the MeSH Browser [35].

A popular source that must also be also taken into account when looking for classifications of drugs is the **National Drug File (NDF)** [81] from the U.S. Veterans Health Administration (VHA), an organization that provides health care to more than 8 million veterans and dependants. The NDF is a nationally maintained medication terminology used to support the VHA information systems that includes information about drug costs, ingredients and inventory management [82]. It has been implemented as a

single-inheritance hierarchy of drug classes, drug products and national drug codes. A cross-reference file is used to list the ingredients of each product.

After the VHA developed the NDF for representing drug terms in its clinical information systems, they released the **National Drug File Reference Terminology (NDF-RT)** [79], which is an enhancement of NDF that organizes the drug list into a formal representation, so that each drug has a computer-usable definition supporting data aggregation and retrieval. NDF-RT combines the NDF hierarchical drug classification with a multi-category reference model that includes orthogonal hierarchies for chemical structure, mechanisms of action, physiologic effects, clinical kinetics and therapeutic diseases, while preserving the existing VHA drug classes. NDF-RT was developed using description logics (DL) and is available for download in several formats, including XML and OWL. NDF-RT is part of the Federal Medication Terminologies (FMT) initiative and has studied within numerous academic and industry publications.

Another relevant effort to classify drugs is the **Anatomical Therapeutic Chemical (ATC) Classification System** [83], controlled by the World Health Organization Collaborating Centre for Drug Statistics Methodology (WHOCC). In the ATC classification system, drugs are divided into 14 main groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Each group represents a top-level category and is identified by an uppercase letter. Some of the ATC groups are: "A: Alimentary tract and metabolism", "B: Blood and blood forming organs", "C: Cardiovascular system" and "D: Dermatologicals". The ATC categories are used in multiple databases related to different areas of medicinal chemistry, such as KEGG [84] and PharmGKB [85].

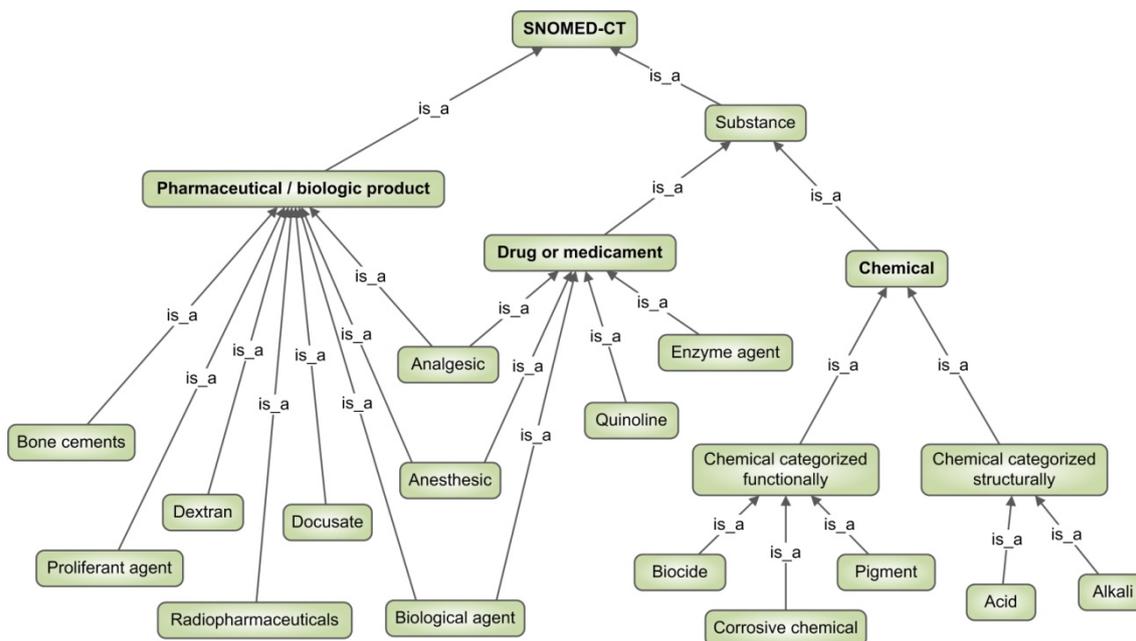


Fig. 7. Part of the hierarchy of classes of SNOMED-CT, showing some of the relevant classes in medicinal chemistry research. Note that there are some concepts that belong to more than one category (e.g. Analgesic is classified both as a "Pharmaceutical / biologic product" and a "Drug or medicament").

Another source that contains relevant information for medicinal chemistry professionals is the **Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)** [80, 86], which is a comprehensive terminology system developed by the College of American Pathologists. It contains over 395,000 terms and was formed by the merger, expansion, and restructuring of SNOMED RT (SNOMED

Reference Terminology) and the United Kingdom National Health Service Clinical Terms (also known as the Read Codes). SNOMED-CT contains a top-level category called “Pharmaceutical / biologic product” (see Fig. 7) that classifies pharmaceutical products into 93 different categories, such as: analgesic, anti-infective agent, dietary product, gastrointestinal drug, nasal dosage form product, etc. In addition, it also contains a category “Chemical”, which classifies chemical substances both functionally and structurally, and a category “Drug or medicament”, with 53 subcategories that represent different kind of drugs (e.g. alcohol agent, central nervous system agent, nasal agent, renal agent, etc.). SNOMED-CT also contains terms to represent knowledge related to drug action and interactions (“Drug action” and “Drug interaction” classes).

Apart from these general drug ontologies and reference terminologies, there are also some sources addressed to classify drugs used to treat specific diseases. An example in the cancer domain is the **National Cancer Institute Thesaurus (NCI Thesaurus)** [33], which is an ontology addressed to integrate molecular and clinical cancer-related information with a controlled terminology, providing a structured representation of key cancer-related concepts in fields such as drugs, therapies, pathways, cellular processes, etc. It is published monthly and provides information (definition, synonyms, etc.) on nearly 10,000 cancers and related diseases, 8,000 single agents and combination therapies, and a wide range of other topics related to cancer and biomedical research. The evolution and characterization of the NCI Thesaurus has been investigated by Gonçalves et al. [87]. Under the concept “Pharmacologic Substance” (id: C1254351), the NCI Thesaurus provides a wide set of technical definitions and synonyms for drugs or agents used to treat patients with cancer or conditions related to cancer. For an overview of the most relevant ontologies of drug discovery and design for complex diseases such as cancer, the reader is referred to previous work [88].

Another effort that requires attention is the research by Richesson et al. [89], who have studied the requirements and strategy for using an extraction of the NDF-RT terminology as a **Pediatric Drug Ontology**. They have stated that the NDF-RT ontology is too complex for routine tasks at the pediatric environment and that a subset of its concepts and relationships would be sufficient to support pediatric research analyses involving classes and properties of medications.

2.3 Drug discovery, design and development

Data produced with disparate methods from distinct organizations is heterogeneous. Standardization through knowledge engineering techniques enables the integration of diverse resources, increasing the total volume of data for analysis. In drug discovery, design and development, ontologies play an important role to provide a unified vision of different data sources, making it possible to integrate disparate data on a unified infrastructure and enable networked and collaborative research. In the previous sections, we have reviewed the ontologies that contain information about chemical compounds in general, and also those aimed to categorize drugs. Below, we will summarize recent efforts made to formalize knowledge related to drug discovery and design processes, paying special attention to those ontological efforts made to prevent adverse events caused by chemical substances such as drugs and vaccinations.

One of the best examples of ontology to support research in drug discovery is the **Drug Discovery Investigations (DDI)** ontology [90], which is aimed to add value to the information generated in the research and development stages of the drug pipeline by making the information easier to reuse, integrate,

curate, retrieve, and reason with. DDI is an application of the **Ontology for Biomedical Investigations (OBI)** [91], which is an integrated ontology for the description of investigations in the area of biology and medicine, to the area of drug discovery. DDI has been built on the basis of **EXPO** [92], an ontology of scientific experiments, and **LABORS** [93], which is a customized version of EXPO expressed in OWL-DL. One of the main applications of DDI is the support for the Robot Scientist Eve [94], which is designed to run automatic drug discovery investigations.

Another successful example of bio-ontology that plays an essential role in drug development processes and applications is the well-known **Gene Ontology (GO)** [59, 60], whose aim is to address the need for consistent descriptions of gene products across species and databases. GO allows to describe attributes of gene products in three non-overlapping domains of molecular biology. It is composed by three ontologies, that describe gene products in terms of their related biological processes, cellular components and molecular functions in a species-independent way. An application of GO to the drug discovery domain is SuperTarget [95]. This resource integrates drug-related information associated with medical indications, adverse drug effects, drug metabolism, pathways and GO terms for target proteins. It contains a core dataset of more than 333,000 drug-target interactions, of which about 310,000 have binding affinity data.

Yao et al. analyzed novel opportunities in drug discovery and emphasized the importance of conceptual standardization. They also suggested an overall design for an ontology to model drug discovery processes [96]. According to their work, the development of ontologies and terminologies enables the standardization of text mined information so that it can be used to model wider biological processes and inspire new drug targets and strategies. Andronis et al. reviewed the application of ontologies to the area of drug repurposing (DR) [97], which emerged as an alternative to the traditional drug development process. The authors explained that among the multitude of ontologies capturing biomedical information GO is among the most referenced ontologies in drug repurposing.

The **Translational Medicine Ontology (TMO)** [98] is a high-level, patient-centric ontology that extends existing domain ontologies to bridge diverse areas of translational medicine including hypothesis management, discovery research, drug development and formulation, clinical research, and clinical practice. TMO has been developed by participants in the W3C Semantic Web for Health Care and Life Sciences Interest Group and member of the National Center for Biomedical Ontology (NCBO).

Another area related to drug discovery and design that is increasingly being supported by ontologies is the early detection of unwanted or adverse effects of drugs, or adverse drug reactions (ADRs), which are a common cause for the rejection of drug candidates during clinical trial or withdrawal after approval. WHO's definition of an adverse drug reaction, which has been in use for about 30 years, is "*a response to a drug that is noxious and unintended and occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease, or for modification of physiological function*" [99]. The standard classification for ADRs is the **Medical Dictionary for Regulatory Activities (MedDRA)** terminology [100], which classifies adverse event information associated with the use of biopharmaceuticals and other medical products (e.g. medical devices and vaccines). MedDRA contains more than 90,000 terms structured in five levels of hierarchy and is employed by regulatory organizations worldwide (e.g. the U.S. Food and Drug Administration). Fig. 8 shows the top-level classes of MedDRA. A widely

recognized terminology in the field of the ADRs that was recently replaced by MedDRA is **COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms)** [101], developed by the FDA for coding, filing, and retrieving post-marketing adverse drug and biologic experience reports. MedDRA also provides the basis for other projects and innovative ideas. As an example, Zhichkin et al. believe that it is necessary to build an **Adverse Reactions and Mechanisms ontology (ARM)** [102], which would be a framework founded on the MedDRA terminology and consistent with the MedDRA general classification that would be addressed to link ADRs with biological mechanisms and functions, establishing correlations between them and allowing statistically meaningful inference from one to the other.

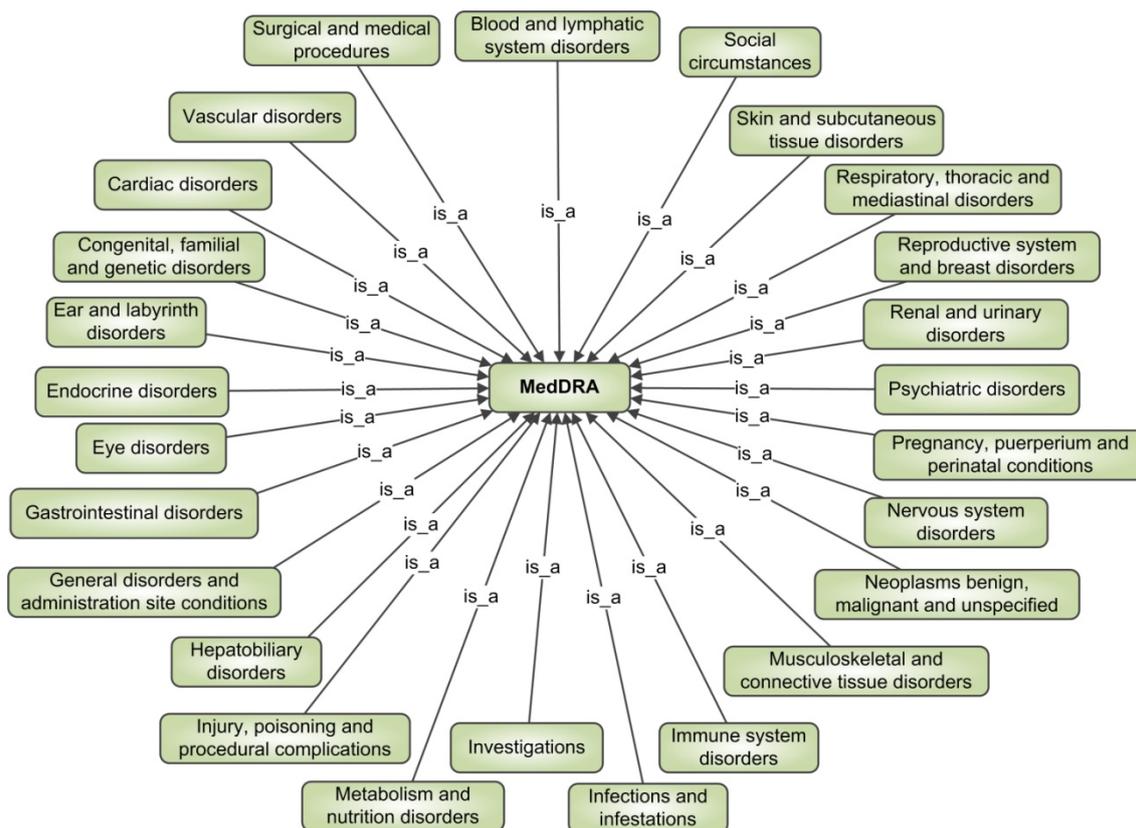


Fig. 8. Top-level classes of MedDRA.

The **WHO-ART (World Health Organization - Adverse Reaction Terminology)** [103] is a terminology that has been developed to serve as a basis for rational coding of adverse reaction terms and is widely used by both regulatory agencies and pharmaceutical manufacturers. Having into account that new drugs and new indications produce new adverse reaction terms, the structure of the terminology has been designed to be flexible enough to allow new entries to be incorporated whilst maintaining its structure and without losing previous information. Alecu et al. have reported a method to group ADR terms by means of SNOMED-CT [104]. They proposed a strategy to improve the WHO-ART structure by integrating the associative relationships included in SNOMED-CT.

Another ontology that contains relevant knowledge about ADRs is the **Ontology of Adverse Events (OAE)** [105], previously named Adverse Event Ontology (AEO). It is a community-driven ontology that has been developed to standardize and integrate data on biomedical adverse events induced by different medical interventions such as administrations of vaccines, drugs, medical devices and nutritional products. OAE has been designed according to the OBO Foundry guidelines [29] such as openness,

collaboration, and use of a common shared syntax, and it is aligned with the Basic Formal Ontology (BFO) [55] and the Relation Ontology (RO) [56]. An ontology to improve adverse events reporting has also been developed. It is called the **Adverse Event Reporting Ontology (AERO)** [106] and is aimed at supporting clinicians at the time of data entry, increasing quality and accuracy of reported adverse events. As explained in [107], the main difference between AERO and AEO is that in AERO, the causality between the reported event and the medical intervention is not established. It describes the event that happened at the reporting time, when no association other than temporal can be established. However, in AEO, adverse events are defined as being induced by the medical intervention, where induced means that there is a chain of processes starting with the intervention and resulting in the adverse event.

Current classifications of drugs and drug reactions have also been used to conceive semantic approaches and build new systems to facilitate drug prescription. Koutkias et al. have designed a knowledge-based framework to support adverse drug event prevention [108]. A decision support system for drug prescription in diabetes has also been reported [109]. Another example is the SemMed approach [110], aimed at reducing the risks caused by human factors in medication management.

Finally, another noteworthy standard that can be useful in medicinal chemistry activities is **Logical Observation Identifier Names and Codes (LOINC)** [111], widely known as a universal code system for identifying laboratory and clinical observations. LOINC was initiated in 1994 by the Regenstrief Institute, an internationally respected non-profit medical research organization associated with Indiana University. Currently, it is considered the reference code set for laboratory test names in transactions between health care facilities, laboratories, laboratory testing devices, and public health authorities. In addition, LOINC is widely recognized and recommended for transmitting laboratory and clinical observations in HL7 messages [112].

Table 2 summarizes the ontologies and reference terminologies related to medicinal chemistry that have been reviewed in this work. For each source, the table contains its acronym, name, main scope, release version, number of terms and the most representative bibliographic references.

Table 2. Main reference ontologies and terminologies in medicinal chemistry. A hyphen (-) means undetermined or not applicable.

Acronym	Name	Main scope	Version	Terms
AERO [106]	Adverse Event Reporting Ontology	Adverse event reports	12/18/2012	388
ARM [102]	Adverse Reactions and Mechanisms	Adverse drug reactions	-	-
ATC Classification [83]	Anatomical Therapeutic Chemical Classification System	Drugs	2013	5,771
ChEBI [36-39]	Chemical Entities of Biological Interest	Chemical entities	2013/01/07 (r99)	30,797
CHEMINF [51]	Chemical Information Ontology	Chemical information entities	1.0	622
ChemTop [53]	ChemTop	Chemistry	2009	110
CO [48]	Chemical Ontology	Chemical entities	-	-
COSTART [101]	Coding Symbols for Thesaurus of Adverse Reaction Terms	Adverse drug reactions	1995	1,641
DDI [90]	Drug Discovery Investigations	Drug discovery research	0.9	645
EXPO and LABORS [92, 93]	EXPO and LABORS	Scientific experiments	1.0	218
GlycO [63]	Glycomics Ontology	Glycoproteomics	4.6.0.0	235
FIX [58]	FIX	Physico-chemical methods and properties	1.2	1,163
GO [59, 60]	Gene Ontology	Gene products	2013/02/03	38,942
LIPID MAPS [61, 62]	LIPID MAPS	Lipid structures	2010	10,774
LOINC [111]	Logical Observation Identifier Names and Codes	Laboratory and clinical observations	236	171,399
MedDRA [100]	Medical Dictionary for Regulatory Activities	Adverse drug reactions	12.0	69,389
MeSH [32]	Medical Subject Headings	Medicine	2012-11-09-09	229,698
NCI-T [33]	National Cancer Institute Thesaurus	Cancer	12.04e	93,411
NDF [81]	National Drug File	Drugs	2009/08/07	26,866
NDF-RT [79]	National Drug File - Reference Terminology	Drugs	2011/09/06	45,757
AEO [105]	Ontology of Adverse Events	Adverse events	1.1.88	1,687
OBI [91]	Ontology for Biomedical Investigations	Investigations in biology and medicine	2012/07/01	3,689
OPL [52]	Ontology for Pharmaceutical Ligands	Pharmaceutical ligands	-	-
PDO [89]	Pediatric Drug Ontology	Drugs in Pediatrics	-	-
REX [57]	REX	Physico-chemical processes	1.13	552
RxNorm [65, 66]	RxNorm	Drugs	11AB_120305F	242,979
SNOMED-CT [80]	Systematized Nomenclature of Medicine-Clinical Terms	Medicine	2011-07-31	395,036
TMO [98]	Translational Medicine Ontology	Translational medicine	1.0.2	300
WHO-ART [103]	World Health Organization - Adverse Reaction Terminology	Adverse drug reactions	December 2005	2,085

3 FUTURE CHALLENGES

In spite of the dramatic advances that have been made during the last decade, the application of ontologies to the field of medicinal chemistry is far from being fully developed. There are many areas and problems in medicinal chemistry that can still benefit from the application of Artificial Intelligence

techniques such as ontologies and ontology-based systems. In addition, one of the main challenges which will have to be faced in the following years is the development of a solid infrastructure that allows to realise the Semantic Web vision in medicinal chemistry.

One of the tasks the community will have to continue undertaking is ontology development. Currently, there are numerous ontologies and reference terminologies whose knowledge can be helpful to improve research processes in drug discovery and design. Nevertheless, they are still in an early phase. One of the major future challenges in pharmacology, medicinal chemistry and toxicology is the development of new ontologies to cover all the numerous subareas of medicinal chemistry and the intersection between fields such as chemistry and biology [21]. In order to ensure an adequate interoperability and scalability, these ontologies should be developed according to good ontology development principles [29]. The existence of multiple ontologies will also foster the development of ontology selection methods addressed to automatically choose the best ontology for a specific context or problem, instead of selecting it by hand, making it possible to build ontology-based systems that require a step of automatic knowledge reusing [113].

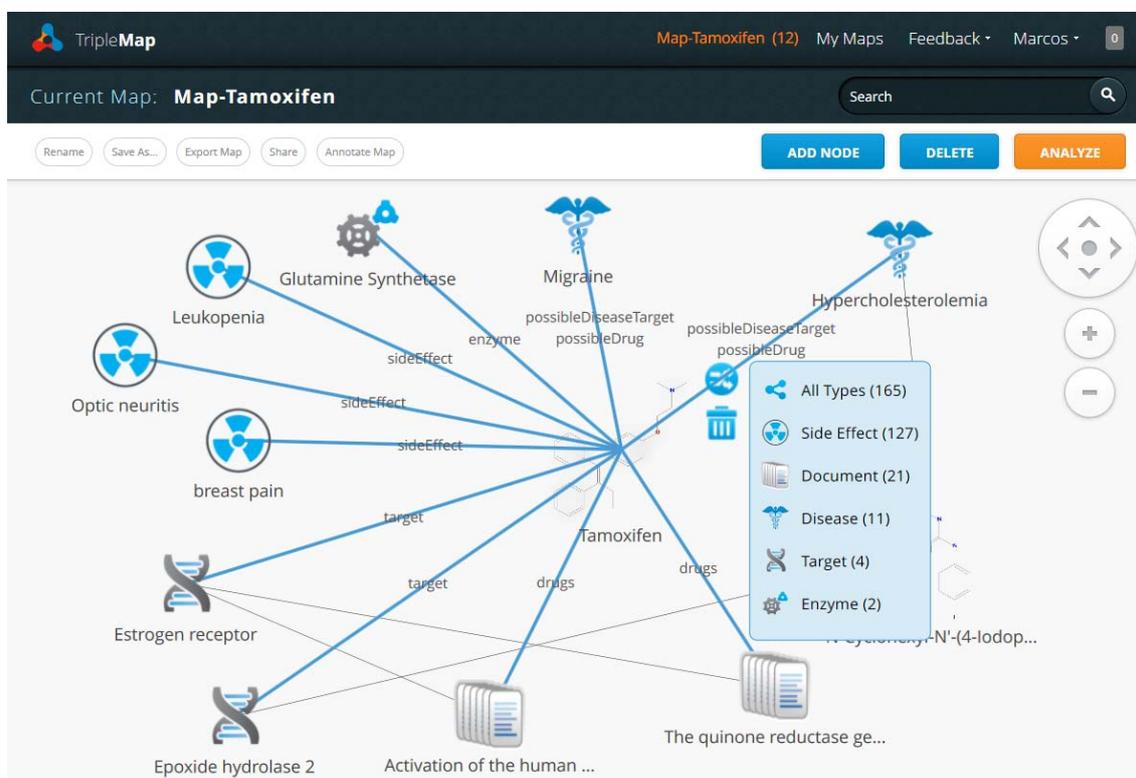


Fig. 9. Screenshot of TripleMap showing some of the information provided by LOOD resources for the drug Tamoxifen.

Another strand of work will focus on the development and application of ontology-based techniques and the coordination of both academic and industry partners to connect all the available sources of drug data together to answer questions in medicinal chemistry. From the perspective of a researcher, the ideal data infrastructure should make it easy to search across different data sources containing data about drugs, clinical trials, diseases, pharmaceutical companies, traditional Chinese medicine, etc. in order to identify novel and meaningful correlations. As a consequence, initiatives aimed to connect all the isolated information related to medicinal chemistry available on the Web, such as the Linked Open Drug Data (LODD) project [114], are expected to attract attention over the next years. LODD is a task force within

the World Wide Web Consortium's (W3C) Health Care and Life Sciences Interest Group (HCLS IG) addressed to provide recommendations for the best practices of exposing pharmaceutical data in the Web by means of a set of technologies and conventions that are commonly referred to as Linked Data [115]. An example of application that provides a rich visual interface to explore and query current LOD resources is TripleMap [116]. It is a web-based application that allows researchers to explore the relationships between biomedical entities that are stored in different datasets but connected through the LOD infrastructure. Following with the previous examples about the drug Tamoxifen (see figures 3, 5 and 6), Fig. 9 shows a screenshot of TripleMap with a graphical representation of some of the biomedical entities connected to Tamoxifen, such as side effects (leukopenia, optic neuritis, breast pain), enzymes (glutamine synthetase), possible disease targets (migraine, hypercholesterolemia), targets (estrogen receptor, epoxide hydrolase 2) and related research documents.

A research area that it is also expected to play an essential role to link all the available semantic information about medicinal chemistry is ontology matching [117, 118], which is known as the process of finding the semantic relations between elements belonging to different ontologies. Given that different tasks or different points of view usually require different conceptualizations, utilizing a single ontology is neither always possible nor advisable. This can lead to the usage of different ontologies that might contain overlapping information (e.g. the same chemical compounds represented in two different ontologies). Ontology matching allows to find the correspondences between the elements of the different ontologies being used, ensuring an adequate exchange information between people, systems and organizations that use different standards to represent the same knowledge.

Personalized medicine is another area where ontologies are expected to be crucial over the next years. There is an increasing interest within the medical community on the development and administration of personalized drugs, requiring the patient to receive the right drug at the right dose. Ontologies will facilitate the integration and analysis of traditionally separate data sets from early drug discovery through to patients in the clinical setting.

4 CONCLUSIONS

An important body of work has been done during the last decade regarding the formal representation of entities in chemistry and pharmacology. As a result, numerous ontologies and reference terminologies are available that allow to formalize knowledge about chemical compounds, drugs, drug discovery processes, adverse drug reactions, etc., facilitating the communication between groups of experts and also between information systems. Future prospects in this rapidly evolving arena look very promising. However, further work needs to be done in areas such as ontology development, ontology matching, ontology selection and ontology-based systems to facilitate the adoption of knowledge sources by the community and to foster the availability of innovative systems able to solve problems for medicinal chemistry professionals, according to the Semantic Web vision of people and computers working in cooperation.

ACKNOWLEDGMENTS

Work supported by the Carlos III Health Institute (grant FIS-PI10/02180), the Ibero NBIC Network (ref. 209RT0366) funded by CYTED, and grants CN2012/217 (REGICC), CN2011/034 ("Programa de consolidación y estructuración de unidades de investigación competitivas") and CN2012/211

(“Agrupación estratégica”) from the Xunta de Galicia. Work also co-funded by FEDER (European Union).

REFERENCES

- [1] Aristotle's Metaphysics. <http://plato.stanford.edu/entries/aristotle-metaphysics/> (Accessed January 10, 2013).
- [2] Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web, *Scientific American*, **2001**, 284, 34-43.
- [3] Studer, R.; Benjamins, V. R.; Fensel, D. Knowledge Engineering: Principles and Methods, *IEEE Transactions on Data & Knowledge Engineering*, **1998**, 25, 161-197.
- [4] Gonzalez-Diaz, H.; Perez-Montoto, L. G.; Duardo-Sanchez, A.; Paniagua, E.; Vazquez-Prieto, S.; Vilas, R.; Dea-Ayuela, M. A.; Bolas-Fernandez, F.; Munteanu, C. R.; Dorado, J.; Costas, J.; Ubeira, F. M. Generalized lattice graphs for 2D-visualization of biological information, *J Theor Biol*, **2009**, 261, 136-47.
- [5] Perez-Bello, A.; Munteanu, C. R.; Ubeira, F. M.; De Magalhaes, A. L.; Uriarte, E.; Gonzalez-Diaz, H. Alignment-free prediction of mycobacterial DNA promoters based on pseudo-folding lattice network or star-graph topological indices, *J Theor Biol*, **2009**, 256, 458-66.
- [6] Aguiar-Pulido, V.; Seoane, J. A.; Rabunal, J. R.; Dorado, J.; Pazos, A.; Munteanu, C. R. Machine learning techniques for single nucleotide polymorphism--disease classification models in schizophrenia, *Molecules*, **2010**, 15, 4875-89.
- [7] Riera-Fernández, P.; Munteanu, C. R.; Pedreira-Souto, N.; Martín-Romalde, R.; Duardo-Sanchez, A.; González-Díaz, H. Definition of Markov-Harary Invariants and Review of Classic Topological Indices and Databases in Biology, Parasitology, Technology, and Social-Legal Networks, *Curr Bioinf*, **2011**, 6, 94-121.
- [8] Fernandez-Blanco, E.; Rivero, D.; Rabunal, J.; Dorado, J.; Pazos, A.; Munteanu, C. R. Automatic seizure detection based on star graph topological indices, *J Neurosci Methods*, **2012**, 209, 410-9.
- [9] Fernandez-Blanco, E.; Aguiar-Pulido, V.; Munteanu, C. R.; Dorado, J. Random Forest classification based on star graph topological indices for antioxidant proteins, *J Theor Biol*, **2012**, 317, 331-7.
- [10] Aguiar-Pulido, V.; Munteanu, C. R.; Seoane, J. A.; Fernández-Blanco, E.; Pérez-Montoto, L. G.; González-Díaz, H.; Dorado, J. Naïve Bayes QSDR classification based on spiral-graph Shannon entropies for protein biomarkers in human colon cancer, *Molecular BioSystems*, **2012**, 8, 1716-1722.
- [11] Gonzalez-Diaz, H.; Muino, L.; Anadon, A. M.; Romaris, F.; Prado-Prado, F. J.; Munteanu, C. R.; Dorado, J.; Sierra, A. P.; Mezo, M.; Gonzalez-Warleta, M.; Garate, T.; Ubeira, F. M. MISS-Prot: web server for self/non-self discrimination of protein residue networks in parasites; theory and experiments in Fasciola peptides and Anisakis allergens, *Mol Biosyst*, **2011**, 7, 1938-55.
- [12] Gonzalez-Diaz, H.; Munteanu, C. R.; Postelnicu, L.; Prado-Prado, F.; Gestal, M.; Pazos, A. LIBP-Pred: web server for lipid binding proteins using structural network parameters; PDB mining of human cancer biomarkers and drug targets in parasites and bacteria, *Mol Biosyst*, **2012**, 8, 851-62.
- [13] Gomez-Perez, A.; Fernández-López, M.; Corcho, O. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer Verlag **2004**.
- [14] OWL Web Ontology Language overview. W3C recommendation. 10 February, 2004. <http://www.w3.org/TR/owl-features/> (Accessed January 4, 2013).
- [15] OWL 2 Web Ontology Language document overview (second edition). W3C recommendation 11 December 2012. <http://www.w3.org/TR/owl2-overview/> (Accessed January 27, 2013).
- [16] Resource Description Framework (RDF) model and syntax specification. W3C recommendation. 22 February, 2009. <http://www.w3.org/TR/REC-rdf-syntax/> (Accessed January 3, 2013).
- [17] Ciccarese, P.; Ocana, M.; Garcia Castro, L. J.; Das, S.; Clark, T. An open annotation ontology for science on Web 3.0, *Journal of Biomedical Semantics*, **2011**, 2, S4.
- [18] Schriml, L. M.; Arze, C.; Nadendla, S.; Chang, Y. W. W.; Mazaitis, M.; Felix, V.; Feng, G.; Kibbe, W. A. Disease Ontology: a backbone for disease semantic integration, *Nucleic Acids Research*, **2012**, 40, D940-D946.
- [19] Dietze, H.; Schroeder, M. GoWeb: a semantic search engine for the life science web, *BMC Bioinformatics*, **2009**, 10, S7.

- [20] Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Research*, **2004**, *32*, 267-270.
- [21] Zimmermann, M.; Fluck, J.; Thi, L. T.; Kolarik, C.; Kumpf, K.; Hofmann, M. Information extraction in the life sciences: perspectives for medicinal chemistry, pharmacology and toxicology, *Current Topics in Medicinal Chemistry*, **2005**, *5*, 785-796.
- [22] Lamy, J. B.; Soualmia, L. F.; Kerdelhué, G.; Venot, A.; Duclos, C. Validating the semantics of a medical iconic language using ontological reasoning, *Journal of Biomedical Informatics*, **2013**, *46*, 56-67.
- [23] Martínez-Romero, M.; Vázquez-Naya, J. M.; Pereira, J.; Pereira, M.; Pazos, A.; Baños, G. The iOSC3 system: using ontologies and SWRL rules for intelligent supervision and care of patients with acute cardiac disorders, *Computational and Mathematical Methods in Medicine* **2013**, *2013*, 1-13.
- [24] Mohammadi, A.; Saraee, M. H.; Salehi, M. Identification of disease-causing genes using microarray data mining and Gene Ontology, *BMC Medical Genomics*, **2011**, *4*, 12.
- [25] De Bodt, S.; Proost, S.; Vandepoele, K.; Rouzé, P.; Van de Peer, Y. Predicting protein-protein interactions in arabidopsis thaliana through integration of orthology, gene ontology and co-expression, *BMC Genomics*, **2009**, *10*, 288.
- [26] 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. BioPortal: ontologies and integrated data resources at the click of a mouse, *Nucleic Acids Research*, **2009**, *37*, W170-W173.
- [27] Côté, R.; Reisinger, F.; Martens, L.; Barsnes, H.; Vizcaino, J. A.; Hermjakob, H. The Ontology Lookup Service: bigger and better, *Nucleic Acids Research*, **2010**, *38*, W155-W160.
- [28] The Open Biological and Biomedical Ontology (OBO) Foundry site. <http://www.obofoundry.org/> (Accessed December 7, **2012**).
- [29] Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L. J.; Eilbeck, K.; Ireland, A.; Mungall, C. J. The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration, *Nature Biotechnology*, **2007**, *25*, 1251-1255.
- [30] Nogrady, T.; Weaver, D. F. *Medicinal chemistry: a molecular and biochemical approach*, Oxford University Press, USA **2005**.
- [31] Nadendla, R. R. *Principles of organic medicinal chemistry*, New Age International **2007**.
- [32] Lipscomb, C. E. Medical Subject Headings (MeSH), *Bulletin of the Medical Library Association*, **2000**, *88*, 265.
- [33] Sioutos, N.; 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, *Journal of Biomedical Informatics*, **2007**, *40*, 30-43.
- [34] PubMed. <http://www.ncbi.nlm.nih.gov/pubmed> (Accessed December 17, **2012**).
- [35] MeSH Browser Website. <http://www.nlm.nih.gov/mesh/MBrowser.html> (Accessed January 15, **2013**).
- [36] Degtyarenko, K.; De Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; Mcnaught, A.; Alcántara, R.; Darsow, M.; Guedj, M.; Ashburner, M. ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic Acids Research*, **2008**, *36*, D344-D350.
- [37] De Matos, P.; Alcántara, R.; Dekker, A.; Ennis, M.; Hastings, J.; Haug, K.; Spiteri, I.; Turner, S.; Steinbeck, C. Chemical entities of biological interest: an update, *Nucleic Acids Research*, **2010**, *38*, D249-D254.
- [38] De Matos, P.; Ennis, M.; Darsow, M.; Guedj, M.; Degtyarenko, K.; Apweiler, R. ChEBI-chemical entities of biological interest, *Nucleic Acids Research*, **2006**.
- [39] ChEBI Home. <http://www.ebi.ac.uk/chebi/> (Accessed January 9, **2013**).
- [40] Degtyarenko, K.; Ennis, M.; Garavelli, J. S. Good annotation practice for chemical data in biology, *In Silico Biology*, **2007**, *7*, 45-56.
- [41] McNaught, A. D.; Wilkinson, A. *Compendium of chemical terminology*, Blackwell Science Oxford, UK **1997**.
- [42] Askenazi, M.; Linial, M. ARISTO: ontological classification of small molecules by electron ionization-mass spectrometry, *Nucleic Acids Research*, **2011**, *39*, W505-W510.
- [43] Gonzalez-Diaz, H.; Prado-Prado, F.; Sobarzo-Sanchez, E.; Haddad, M.; Maurel Chevalley, S.; Valentin, A.; Quetin-Leclercq, J.; Dea-Ayuela, M. A.; Teresa Gomez-Munos, M.; Munteanu, C. R.; Jose Torres-Labandeira, J.; Garcia-Mera, X.; Tapia, R. A.; Ubeira, F. M. NL MIND-BEST: A web server for ligands and proteins discovery-Theoretic-experimental study of proteins of *Giardia lamblia* and new compounds active against *Plasmodium falciparum*, *J Theor Biol*, **2011**, *276*, 229-249.

- [44] [Munteanu, C. R.; Gonzalez-Diaz, H.; Borges, F.; de Magalhaes, A. L. Natural/random protein classification models based on star network topological indices, *J Theor Biol*, **2008**, *254*, 775-83.](#)
- [45] [Garcia, I.; Munteanu, C. R.; Fall, Y.; Gomez, G.; Uriarte, E.; Gonzalez-Diaz, H. QSAR and complex network study of the chiral HMGR inhibitor structural diversity, *Bioorg Med Chem*, **2009**, *17*, 165-75.](#)
- [46] [Vazquez, J. M.; Aguiar, V.; Seoane, J. A.; Freire, A.; Serantes, J. A.; Dorado, J.; Pazos, A.; Munteanu, C. R. Star Graphs of Protein Sequences and Proteome Mass Spectra in Cancer Prediction, *Current Proteomics*, **2009**, *6*, 275-288.](#)
- [47] [Rodriguez-Soca, Y.; Munteanu, C. R.; Dorado, J.; Rabuñal, J.; Pazos, A.; González-Díaz, H. Plasmod-PPi: a web-server predicting complex biopolymer targets in Plasmodium with entropy measures of protein-protein interactions, *Polymer*, **2010**, *51*, 264-273.](#)
- [48] [Feldman, H. J.; Dumontier, M.; Ling, S.; Haider, N.; Hogue, C. W. V. CO: A chemical ontology for identification of functional groups and semantic comparison of small molecules, *FEBS letters*, **2005**, *579*, 4685-4691.](#)
- [49] [The `checkmol/matchmol` homepage. <http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm.html> \(Accessed December 4, **2012**\).](#)
- [50] [Semantic chemistry working group. <http://code.google.com/p/semanticchemistry/> \(Accessed January 3, **2013**\).](#)
- [51] [The CHEMINF ontology, OWL version. <http://semanticchemistry.googlecode.com/> \(Accessed January 3, **2013**\).](#)
- [52] [Schuffenhauer, A.; Zimmermann, J.; Stoop, R.; van der Vyver, J. J.; Lecchini, S.; Jacoby, E. An ontology for pharmaceutical ligands and its application for in silico screening and library design, *Journal of Chemical Information and Computer Sciences*, **2002**, *42*, 947-955.](#)
- [53] [Stenzhorn, H.; Schulz, S.; Beißwanger, E.; Hahn, U.; Van Den Hoek, L.; Van Mulligen, E. Biotop and chemtop-top-domain ontologies for biology and chemistry. 7th International Semantic Web Conference \(ISWC\), 2008.](#)
- [54] [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*, **2008**, *3*, 205-212.](#)
- [55] [Arp, R.; Smith, B. Function, role, and disposition in basic formal ontology, *Nature Preceedings*, **2008**, 1-4.](#)
- [56] [Smith, B.; Ceusters, W.; Klagges, B.; Köhler, J.; Kumar, A.; Lomax, J.; Mungall, C.; Neuhaus, F.; Rector, A. L.; Rosse, C. Relations in biomedical ontologies, *Genome Biology*, **2005**, *6*, R46.](#)
- [57] [The REX Ontology. <http://www.obofoundry.org/cgi-bin/detail.cgi?id=rex> \(Accessed December 4, **2012**\).](#)
- [58] [The FIX ontology. <http://obofoundry.org/cgi-bin/detail.cgi?id=fix> \(Accessed December 4, **2012**\).](#)
- [59] [Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T. Gene Ontology: tool for the unification of biology, *Nature Genetics*, **2000**, *25*, 25-29.](#)
- [60] [Harris, M.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C. The Gene Ontology \(GO\) database and informatics resource, *Nucleic Acids Research*, **2004**, *32*, D258.](#)
- [61] [Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R. H.; Shimizu, T.; Spener, F.; Van Meer, G.; Wakelam, M. J. O.; Dennis, E. A. Update of the LIPID MAPS comprehensive classification system for lipids, *Journal of Lipid Research*, **2009**, *50*, S9-S14.](#)
- [62] [LIPID MAPS - Lipidomics Gateway. <http://www.lipidmaps.org/> \(Accessed December 28, **2012**\).](#)
- [63] [Thomas, C. J.; Sheth, A. P.; York, W. S. Modular ontology design using canonical building blocks in the biochemistry domain, *Frontiers in Artificial Intelligence and Applications*, **2006**, *150*, 115.](#)
- [64] [Ackerman, M.; Ball, M.; Clayton, P.; Frisse, M.; Gardner, R.; Greenes, R. Standards for medical identifiers, codes and messages needed to create an efficient computer stored medical record, *Journal of the American Medical Informatics Association*, **1994**, *1*, 1-7.](#)
- [65] [Liu, S.; Ma, W.; Moore, R.; Ganesan, V.; Nelson, S. RxNorm: prescription for electronic drug information exchange, *IT professional*, **2005**, *7*, 17-23.](#)
- [66] [Nelson, S. J.; Zeng, K.; Kilbourne, J.; Powell, T.; Moore, R. Normalized names for clinical drugs: RxNorm at 6 years, *Journal of the American Medical Informatics Association*, **2011**, *18*, 441-448.](#)
- [67] [RxNav Website. <http://rxnav.nlm.nih.gov/> \(Accessed January 16, **2013**\).](#)

- [68] Zeng, K.; Bodenreider, O.; Kilbourne, J.; Nelson, S. RxNav: towards an integrated view on drug information. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems, 2007.
- [69] Fung, K. W.; McDonald, C.; Bray, B. E. RxTerms—a drug interface terminology derived from RxNorm. AMIA Annual Symposium Proceedings, 2008.
- [70] Zeng, K.; Bodenreider, O.; Nelson, S. J. Design and implementation of a personal medication record—MyMedicationList. AMIA Annual Symposium Proceedings, 2008.
- [71] Nelson, S. J.; Zeng, K.; Kilbourne, J. Building a standards-based and collaborative e-prescribing tool: MyRxPad, *International Journal of Data Mining and Bioinformatics*, **2011**, *5*, 252-265.
- [72] Gold Standard Drug Database. <http://www.goldstandard.com/product/gold-standard-drug-database/> (Accessed January 9, **2013**).
- [73] Medi-Span Electronic Drug File™ (MED-File) v2. <http://www.medispans.com/medi-span-electronic-drug-file.aspx> (Accessed January 9, **2013**).
- [74] Multum MediSource Lexicon. <http://www.multum.com/Lexicon.htm> (Accessed January 9, **2013**).
- [75] Micromedex Red Book. <http://www.micromedex.com/products/redbook/windows/> (Accessed January 9, **2013**).
- [76] National Drug Code Directory. <http://www.accessdata.fda.gov/scripts/cder/ndc> (Accessed January 9, **2013**).
- [77] Structured Product Labeling Resources. <http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/> (Accessed January 9, **2013**).
- [78] MedKnowledge. <http://www.fdbhealth.com/fdb-medknowledge/> (Accessed January 9, **2013**).
- [79] Erlbaum, M. S. T. US Department of Veterans Affairs enterprise reference terminology strategic overview. Medinfo 2004: Proceedings of the 11th World Congress on Medical Informatics, 2004.
- [80] Price, C.; Spackman, K. SNOMED clinical terms, *British Journal of Healthcare Computing and Information Management*, **2000**, *17*, 27-31.
- [81] National Drug File. <http://www.pbm.va.gov/NationalFormulary.aspx> (Accessed January 9, **2013**).
- [82] Carter, J. S.; Brown, S. H.; Erlbaum, M. S.; Gregg, W.; Elkin, P. L.; Speroff, T.; Tuttle, M. S. Initializing the VA medication reference terminology using UMLS metathesaurus co-occurrences. Proceedings of the AMIA Symposium, 2002.
- [83] Anatomical Therapeutic Chemical (ATC) Classification System. <http://www.whocc.no/atc/> (Accessed January 31, **2013**).
- [84] Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Research*, **2010**, *38*, D355-D360.
- [85] Hewett, M.; Oliver, D.; Rubin, D.; Easton, K.; Stuart, J.; Altman, R.; Klein, T. PharmGKB: the pharmacogenetics knowledge base, *Nucleic Acids Research*, **2002**, *30*, 163-165.
- [86] Stearns, M. Q.; Price, C.; Spackman, K. A.; Wang, A. Y. SNOMED Clinical Terms: Overview of the Development Process and Project Status. AMIA Annual Symposium, 2001.
- [87] Gonçalves, R. S.; Parsia, B.; Sattler, U. Analysing the evolution of the NCI thesaurus. Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on, 2011.
- [88] Vázquez-Naya, J. M.; Martínez-Romero, M.; Porto-Pazos, A. B.; Nóvoa, F.; Valladares-Ayerbes, M.; Pereira, J.; Munteanu, C. R.; Dorado, J. Ontologies of drug discovery and design for neurology, cardiology and oncology, *Current Pharmaceutical Design*, **2010**, *16*, 2724-2736.
- [89] Richesson, R.; Pathak, J.; McLeod, W.; Blackmon, G.; Vehik, K. Requirements and strategy for the development of a pediatric drug ontology. International Conference on Biomedical Ontology (ICBO), 2012.
- [90] Qi, D.; King, R.; Hopkins, A.; Bickerton, G.; Soldatova, L. An ontology for description of drug discovery investigations, *Journal of Integrative Bioinformatics*, **2010**, *7*, 126.
- [91] Peters, B. Ontology for Biomedical Investigations. International Conference on Biomedical Ontology, 2009.
- [92] Soldatova, L.; King, R. An ontology of scientific experiments, *Journal of the Royal Society Interface*, **2006**, *3*, 795-803.
- [93] King, R.; Rowland, J.; Oliver, S.; Young, M.; Aubrey, W.; Byrne, E.; Liakata, M.; Markham, M.; Pir, P.; Soldatova, L.; Sparkes, A.; Whelan, K.; Clare, A. The automation of science, *Science*, **2009**, *324*, 85-89.

- [94] Sparkes, A.; Aubrey, W.; Byrne, E.; Clare, A.; Khan, M.; Liakata, M.; Markham, M.; Rowland, J.; Soldatova, L.; Whelan, K. Towards robot scientists for autonomous scientific discovery, *Automated Experimentation*, **2010**, *2*.
- [95] Hecker, N.; Ahmed, J.; von Eichborn, J.; Dunkel, M.; Macha, K.; Eckert, A.; Gilson, M. K.; Bourne, P. E.; Preissner, R. SuperTarget goes quantitative: update on drug–target interactions, *Nucleic Acids Research*, **2012**, *40*, D1113-D1117.
- [96] Yao, L.; Evans, J. A.; Rzhetsky, A. Novel opportunities for computational biology and sociology in drug discovery, *Trends in Biotechnology*, **2009**, *27*, 531-540.
- [97] Andronis, C.; Sharma, A.; Virvilis, V.; Deftereos, S.; Persidis, A. Literature mining, ontologies and information visualization for drug repurposing, *Briefings in Bioinformatics*, **2011**, *12*, 357-368.
- [98] Dumontier, M.; Andersson, B.; Batchelor, C.; Denney, C.; Domarew, C.; Jentzsch, A.; Luciano, J.; Pichler, E.; Prud'hommeaux, E.; Whetzel, P. The Translational Medicine Ontology: Driving personalized medicine by bridging the gap from bedside to bench. Proceedings of the 13th Annual Bio-Ontologies Meeting, Boston, USA, 2010.
- [99] Edwards, I. R.; Aronson, J. K. Adverse drug reactions: definitions, diagnosis, and management, *The Lancet*, **2000**, *356*, 1255-1259.
- [100] Brown, E. G.; Wood, L.; Wood, S. The medical dictionary for regulatory activities (MedDRA), *Drug Safety*, **1999**, *20*, 109-117.
- [101] *COSTART. Coding Symbols for Thesaurus of Adverse Reaction Terms*, Department of Health and Human Services, Food and Drug Administration, Center for Drugs and Biologics: Washington **1989**.
- [102] Zhichkin, P.; Athey, B.; Avigan, M.; Abernethy, D. Needs for an expanded ontology-based classification of adverse drug reactions and related mechanisms, *Clinical Pharmacology & Therapeutics*, **2012**, *91*, 963-965.
- [103] WHO. *International monitoring of adverse reactions to drugs: adverse reaction terminology*, WHO Collaborating Centre for International Drug Monitoring: Uppsala, Sweden **1992**.
- [104] Alecu, I.; Bousquet, C.; Jaulent, M. C. A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms, *BMC Medical Informatics and Decision Making*, **2008**, *8*, S4.
- [105] He, Y.; Xiang, Z.; Sarntivijai, S.; Toldo, L.; Ceusters, W. AEO: a realism-based biomedical ontology for the representation of adverse events. Proceedings of the Adverse Event Representation Workshop, International Conference on Biomedical Ontologies (ICBO), University at Buffalo, NY, 2011.
- [106] Courtot, M.; Brinkman, R. R. Towards an Adverse Event Reporting Ontology. International Conference on Biomedical Ontology (ICBO), 2011.
- [107] Adverse Event Reporting Ontology (AERO) site in Google Code. <http://code.google.com/p/adverse-event-reporting-ontology/> (Accessed January 27, **2013**).
- [108] Koutkias, V.; Kilintzis, V.; Stalidis, G.; Lazou, K.; NièS, J.; Mcnair, P.; Beuscart, R.; Maglaveras, N. Knowledge engineering for adverse drug event prevention: On the design and development of a uniform, contextualized and sustainable knowledge-based framework, *Journal of Biomedical Informatics*, **2012**.
- [109] Chen, R.-C.; Chiu, J.-Y.; Bau, C.-T. The recommendation of medicines based on multiple criteria decision making and domain ontology—An example of anti-diabetic medicines. Machine Learning and Cybernetics (ICMLC), 2011 International Conference on, 2011.
- [110] Rodríguez-González, A.; García-Crespo, Á.; Colomo-Palacios, R.; Gómez-Berbís, J. M.; Jiménez-Domingo, E. Using ontologies in drug prescription: the SemMed approach, *International Journal of Knowledge-Based Organizations (IJKBO)*, **2011**, *1*, 1-15.
- [111] McDonald, C. J.; Huff, S. M.; Suico, J. G.; Hill, G.; Leavelle, D.; Aller, R.; Forrey, A.; Mercer, K.; DeMoor, G.; Hook, J. LOINC, a universal standard for identifying laboratory observations: a 5-year update, *Clinical Chemistry*, **2003**, *49*, 624-633.
- [112] Dolin, R. H.; Alschuler, L.; Boyer, S.; Beebe, C.; Behlen, F. M.; Biron, P. V.; Shvo, A. S. HL7 clinical document architecture, release 2, *Journal of the American Medical Informatics Association*, **2006**, *13*, 30-39.
- [113] Martínez-Romero, M.; Vázquez-Naya, J. M.; Munteanu, C. R.; Pereira, J.; Pazos, A. An approach for the automatic recommendation of ontologies using collaborative knowledge, *Lecture Notes in Artificial Intelligence*, **2010**, *6277*, 74-81.
- [114] Samwald, M.; Jentzsch, A.; Bouton, C.; Kallesøe, C.; Willighagen, E.; Hajagos, J.; Marshall, M.; Prud'hommeaux, E.; Hassanzadeh, O.; Pichler, E. Linked open drug data for pharmaceutical research and development, *Journal of Cheminformatics*, **2011**, *3*, 19.

- [115] [Bizer, C.; Heath, T.; Berners-Lee, T. Linked data-the story so far, *International Journal on Semantic Web and Information Systems \(IJSWIS\)*, **2009**, 5, 1-22.](#)
- [116] TripleMap. www.triplemap.com (Accessed January 27, **2012**).
- [117] [Shvaiko, P.; Euzenat, J. Ontology matching: state of the art and future challenges, **2012**.](#)
- [118] [Martínez-Romero, M.; Vázquez-Naya, J. M.; Pereira, J.; Ezquerro, N. Ontology alignment techniques. In *Encyclopedia of Artificial Intelligence*, **2008**; Vol. 3, pp. 1290-1295.](#)