OXFORD

Briefings in Bioinformatics, 16(6), 2015, 1069–1080

doi: 10.1093/bib/bbv011
Advance Access Publication Date: 10 April 2015
Paper

The role of ontologies in biological and biomedical research: a functional perspective

Robert Hoehndorf, Paul N. Schofield and Georgios V. Gkoutos

Corresponding author. Robert Hoehndorf, Computational Bioscience Research Center, King Abdullah University of Science and Technology, 4700 KAUST, P.O. Box 2882, 23955-6900 Thuwal, Kingdom of Saudi Arabia. Tel.: +966-12-8081643; Fax: +966-12-8021344. Email: robert.hoehndorf@kaust.edu.sa

Abstract

Ontologies are widely used in biological and biomedical research. Their success lies in their combination of four main features present in almost all ontologies: provision of standard identifiers for classes and relations that represent the phenomena within a domain; provision of a vocabulary for a domain; provision of metadata that describes the intended meaning of the classes and relations in ontologies; and the provision of machine-readable axioms and definitions that enable computational access to some aspects of the meaning of classes and relations. While each of these features enables applications that facilitate data integration, data access and analysis, a great potential lies in the possibility of combining these four features to support integrative analysis and interpretation of multimodal data. Here, we provide a functional perspective on ontologies in biology and biomedicine, focusing on what ontologies can do and describing how they can be used in support of integrative research. We also outline perspectives for using ontologies in data-driven science, in particular their application in structured data mining and machine learning applications.

Key words: ontology; Semantic Web; data integration; data mining

Introduction

The past 15 years have seen a revolution in the volume and complexity of data created in the life sciences, and with the increase in available data, the need for data management, integration and analysis has become an increasingly important challenge. The use of ontologies began in the biological sciences around 1998 with the development of the Gene Ontology (GO) [1]. By 2007, there was sufficient interest and activity in the area to merit national and international coordination efforts such as the Open Biomedical Ontologies (OBO) Foundry [2] or the National Center for Biomedical Ontologies [3].

Many definitions of 'ontology' have been proposed in the literature [4–10], and classifications of different types of vocabularies, thesauri, ontologies and knowledge bases have been proposed, based on criteria such as their intended use, degree of formalization or philosophical interpretation [2, 11–15]. Independent of the actual definition of what an 'ontology' is, most artifacts labeled 'ontologies', as well as some 'vocabularies' and 'thesauri', provide several main features, and these features are used in almost all their applications (see Table 1):

i. classes and relations, referred to by an identifier such as an Internationalized Resource Identifier (IRI), a Uniform Resource Identifier (URI), or a database identifier string;

Robert Hoehndorf is an Assistant Professor in Computer Science at the King Abdullah University of Science and Technology in Thuwal. His research focuses on the applications of ontologies in biology and biomedicine, with a particular emphasis on integrating and analyzing heterogeneous, multimodal data.

Paul N. Schofield is Reader in Biomedical Informatics at the University of Cambridge. His interests are in mouse and human genetics, and building and applying ontologies for phenotypes and disease.

Georgios V. Gkoutos is Reader in Bioinformatics at the University of Aberystwyth. His research interests are in developing biomedical ontologies and applying them to the study of association between genotype and phenotype.

Submitted: 26 November 2014; **Received (in revised form):** 20 January 2015

Table 1. The main features provided by ontologies in support of biological and biomedical research

Ontology feature	Utility in research
Classes and relations	The use of standard identifiers for classes and relations in ontologies is what enables data integration across multiple databases because the same identifiers can be used across multiple, disconnected databases, files, or web sites.
Domain vocabulary	Through labels associated with classes and relations, ontologies provide a domain vocabulary that can be exploited for applications ranging from natural language processing, creation of user interfaces, etc.
Metadata and descriptions	Textual definitions, descriptions, examples and further metadata associated with classes in ontologies are what enable domain experts to understand the precise meaning of class in the ontology. The definitions and related metadata should allow consistent understanding of the meaning of classes in ontologies.
Axioms and formal definitions	Formal definitions and axioms enable automated and computational access to (some parts of) the meaning of a class or relation.

ii. a domain vocabulary, i.e. a list of terms associated with the ontology's classes and relations;
iii. textual definitions and descriptions that provide additional information about what kind of things a class or relation refers to,
iv. formal definitions and axioms that provide a computational counterpart to textual definitions and that can be accessed and exploited automatically using specialized software (i.e. automated reasoners) and axioms about a domain, i.e. statements that are considered to be true within that domain and which provide background knowledge about a domain.

Here, we discuss ontologies as artifacts containing these features, and we use these features to provide a 'functional' perspective on ontologies (as well as other artifacts such as thesauri, glossaries, semantic networks, or structured vocabularies that provide a similar functionality). We illustrate how these features can be exploited to enable or improve data analysis in biology and biomedicine, and how the combination of these features makes data integration and data analysis across traditional domain boundaries a reality.

A functional perspective on ontologies

Classes and relations

The principal components of ontologies are classes and relations. A 'class' is an entity that refers to a set of entities in the world, such as the class 'Protein' (referring to the set of all proteins), 'Apoptosis' (referring to the set of all apoptotic processes) or 'Red' (referring to the set of all red qualities). However, in contrast to sets that are defined by their extension (i.e. the entities that are part of the set), classes in ontologies are defined 'intensionally' by specifying the properties, features and relations that the entities belonging to a class must have [6, 9]. Relations are similar to classes but hold for two or more entities. Examples are the relations 'part of', 'participates in' or 'quality of'.

In ontologies, classes and relations are commonly referred to using a unique identifier. In the Semantic Web [16], this identifier is an IRI, which is a URI supporting Unicode characters. It is still common to use database identifier strings in biomedical databases to refer to classes and relation. For example, within the OBO [2] community, an identifier for a class or relation in an ontology consists of a prefix string, a colon and a series of digits [17]. In Figure 1, `PO:0009011` is an identifier for a class and `OBO_REL:0000002` an identifier for a relation, with the prefixes `PO` and `OBO_REL`, respectively. In communities in which database identifiers are still widely used, transformation policies that standardize how database identifiers are transformed into IRIs may be adopted. For example, within the OBO, `PO:0009011` would be translated to the IRI http://purl.obolibrary.org/obo/PO_0009011 [17].

Domain vocabulary

The second main feature that ontologies provide is a set of labels associated with the classes and relations in the ontology. Labels are strings that are used to refer to the kind of things a class or relation represents. In ontologies, labels may be provided in multiple languages, and multiple labels may be assigned to one class. Additionally, a primary label may be distinguished from secondary labels or synonyms. Such an assertion signifies that, within the context of an ontology, the primary label is what is used to refer to a class or relation, while the additional labels and synonyms are used to refer to the phenomena captured by a class or a relation in other contexts.

In some ontologies or structured vocabularies, the (primary) label of a class is also used as component of the class identifier (its IRI), but in the majority of ontologies the label and the class identifier are maintained as distinct features, as the label may change (in the simplest case owing to a misspelling) while the intended meaning of the class remains the same [18, 19]. The distinction between label and class identifiers caters for changing metadata associated with the class without having to modify data that are already characterized with the class identifier.

Provision of a domain vocabulary is a widely used feature of ontologies. If an ontology aims to cover a domain completely, the set of labels associated with the ontology classes and relations provide a large set of relevant terms within that domain. For example, an ontology for human anatomy such as the Foundational Model of Anatomy [20] will not only contain the classes and relations relevant to describe human anatomy, but also provide a large set of terms used to refer to human anatomical structures and the ways in which they may be related (as labels of the relations).

Textual definitions, descriptions and metadata

A third feature of ontologies is the provision of information about the kind of phenomena a class or relation is supposed to capture. The majority of ontologies contain two main kinds of additional information: the first is intended primarily for users of the ontology and provides textual definitions, examples and background information that makes the intended meaning of a class in the ontology as precise as possible to ontology users; the second is additional technical information that relates one class to entries in other databases, literature or other ontologies and vocabularies.

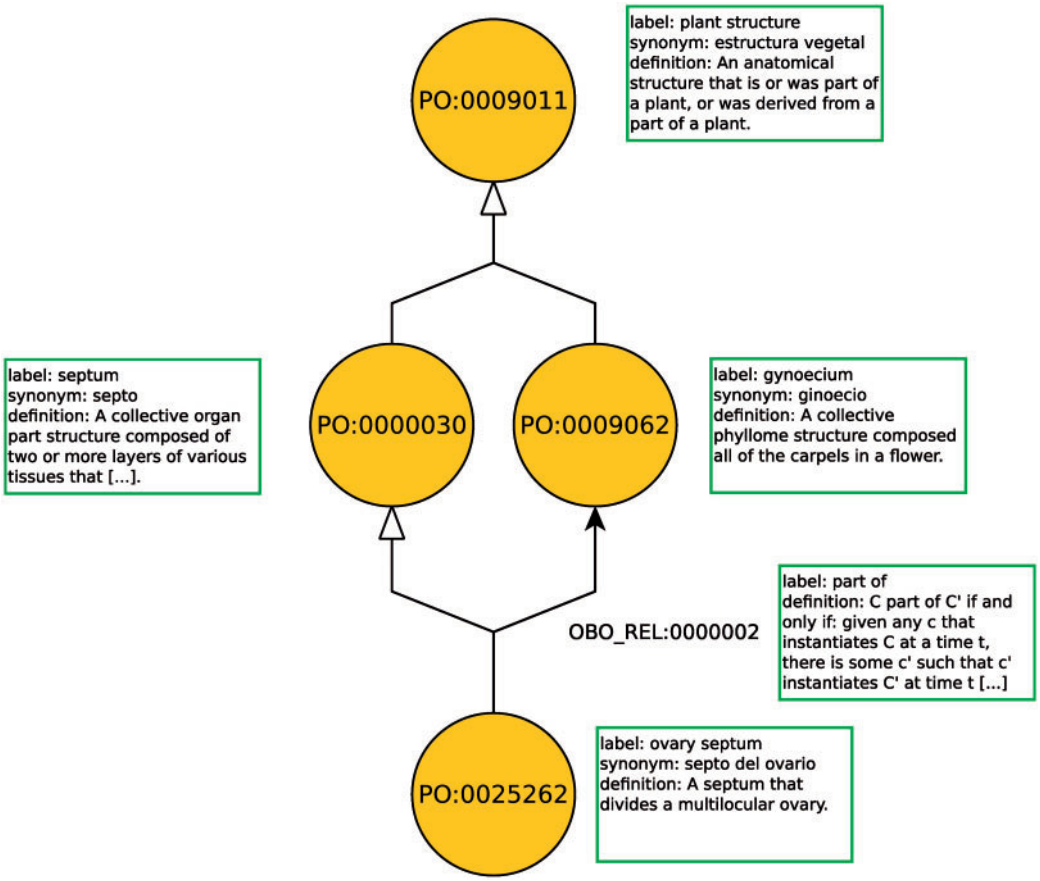

Figure 1. A part of the Plant Ontology. The figure shows classes as circles, labels and definitions in boxes and axioms as edges between classes. The label and definition of the relation `OBO_REL:0000002` is a label for an axiom pattern.

Most ontologies in biomedicine that are primarily intended for data annotation across multiple databases provide textual definitions for their classes. There has been some discussion about what constitutes a 'good' textual definition in ontologies [21]. In some domains, ontology users have opted to use Aristotelian definitions, i.e. definitions that state the general kind of thing that a class or relation represents, coupled with the properties that distinguish it from the general kind (the 'genus–differentia' model). For example, an 'ovary septum' can be defined as a 'septum' (the general kind) that 'divides a multilocular ovary' (the conditions or properties that separate it from others within the general kind). However, other types of textual definitions are widely used as well [22]. Ideally, the textual definitions are sufficient for an ontology user to understand exactly what kinds of phenomena a class in an ontology refers to, and a 'good' definition does exactly that: it is understandable to an ontology user and removes ambiguity in a term so that different ontology users can apply it consistently.

Formal definitions and axioms

Finally, ontologies provide 'formal' and 'machine-readable' definitions and axioms. These are some of the most valuable features of ontologies, as these may enable graph- and network-based analyses, 'fuzzy' matches in searches, verification of data consistency, as well as provide background knowledge about a domain and reveal new knowledge through deductive inference. The axioms and definitions of ontologies can be represented in many forms. In some cases, they are expressed directly as a graph structure that is intended to represent a taxonomy or a partonomy. In other cases, axioms and definitions are written in a formal language. For example, Figure 1 represents a part of an ontology as a graph in which the edges ending with a white triangle represent 'taxonomic' relationships and the edge labeled 'part of' represents a parthood relationship.

Ontologies are increasingly being expressed directly in a formal language, and graph representations of ontologies are being derived dynamically from this formal representation. Most commonly, ontologies in the biological and biomedical domain are represented in the Web Ontology Language (OWL) [23], a formal language based on description logics [24, 25], or a sub-language or profile of OWL such as OWL-EL [26]. The graph-based OBO Flatfile Format, which is still used by several ontologies (currently, in November 2014, 66 ontologies in the OBO library are represented natively in the OBO Flatfile Format, while 45 are represented natively in OWL), has now become a sub-language of OWL [27] and can be processed with the same tools and libraries used for OWL ontologies. Table 2 provides an overview over different representation and query languages for ontologies as well as key concepts centered around ontologies.

The construction of ontologies in a formal language often follows—explicitly or implicitly—the axiomatic method [28]. According to the axiomatic method, knowledge about a domain is formalized by first introducing a set of terms referring to classes and relations in the domain (the classes and relations of the ontology), and then explicitly defining these classes and relations by reference to other terms or relations, and possibly introducing new terms and relations. For example, the class

Table 2. Query and representation languages, and key concepts around ontologies in biology

Language	Description
Resource Description Framework (RDF)	RDF [120] is a graph-based language in which resources are identified through their IRI and statements take the form of triples (subject–predicate–object). Therefore, a set of RDF statements forms a labeled directed graph. RDF also comes with a predefined vocabulary that can be used to state the type of a resource (e.g. a class, or a literal) or represent relations between resources (e.g. labels of resources, subclass relations between resources).
Web Ontology Language (OWL)	OWL [23] is a language based on description logic and has a formal, model-theoretic semantics. Several sub-languages of OWL have been developed, including OWL-DL, OWL-EL, OWL-RL, OWL-QL and OWL Full, which support different language constructs, have different properties regarding decidability and complexity of reasoning tasks, and therefore different areas of application.
SPARQL Protocol and RDF Query Language (SPARQL)	SPARQL [121] is a standardized NoSQL query language, which can be used to query RDF databases and supports query federation (i.e. querying data distributed across multiple databases). SPARQL can also be used to query other kinds of data, including relational databases and flat files.
Linked Data	Linked Data [122] represents a method of publishing and sharing data on the web. When publishing Linked Data sets, data items are identified through a URI, and links to other data items are included in the data set by explicitly referring to the URI that denotes the other items. The URIs used to denote data items should be dereferencable, i.e. it should be possible to obtain additional information about the item through the URI (depending on the method used to access the URI, the information could be presented as HTML, RDF, JavaScript Object Notation or similar).
OBO Flatfile Format	The OBO Flatfile Format [27] is a graph-based knowledge representation language widely used for biological and biomedical ontologies. The majority of language constructs are compatible with OWL, and bi-directional transformations between the OBO Flatfile Format and OWL have been implemented.
Proprietary graph-based ontology representation formats	A number of graph-based representations of ontologies have been developed that primarily specify labeled graphs. Examples include the representation of the Medical Subject Headings thesaurus [123], the Unified Medical Language System [124] or the medical vocabulary SNOMED CT [125].

'ovary septum' (`PO:0025262`) in Figure 1 could be defined using the OWL language as:

```
'ovary septum' equivalentTo: septum and divides some
'multilocular ovary'
```

This definition states that the class on the left of `equivalentTo:` (i.e. 'ovary septum') is equivalent to the expression on the right of `equivalentTo:` (`septum and divides some 'multilocular ovary'`), making 'ovary septum' a shorthand form of the complex statement on the right (i.e. every occurrence of 'ovary septum' could be replaced with the expression on the right). A definition alone does not add any information about the intended meaning of a class: the meaning of 'ovary septum' now depends entirely on the meaning of 'septum', 'multilocular ovary' and the relation 'divides'. Following the axiomatic method, we can introduce further definitions for some of these terms. For example, 'multilocular ovary' could be further defined:

```
'multilocular ovary' equivalentTo: ovary and has-quality
some multilocular
```

Similarly, since this takes the form of an explicit definition (through the use of the `equivalentTo:` keyword), we can now replace every occurrence of 'multilocular ovary' with the expression on the righthand side. Applying this property of explicit definitions, we can rewrite the definition of 'ovary septum' as:

```
'ovary septum' equivalentTo: septum and divides some
  (ovary and has-quality some multilocular)
```

Now, the meaning of the class 'ovary septum' depends on the meaning of the classes 'septum', 'ovary', 'multilocular', as well as the relations 'divides' and 'has-quality'. We could continue defining these classes by introducing additional classes and relations. However, inevitably, we will come up with a set of classes and relations that we cannot further define.

As a second step in the axiomatic method, we use the ontology's classes and relations in statements that we consider to be true in the domain it is supposed to represent. These statements are the axioms, which form the features of ontologies that provide domain knowledge and fill the classes and relations with meaning. For example, we could state about the 'has quality' relation that, if an entity x has the quality q, and an entity y has the quality q, then x must be identical to y (i.e. a quality is always the quality of at most one entity). In OWL, we could state this simply as:

```
ObjectProperty: 'has quality'
Characteristics: InverseFunctional
```

Another kind of axiom is the 'subClassOf:' axiom in which one class is asserted to be a subclass of another class. A class X is a subclass of Y if and only if all instances of X are also instances of Y (i.e. all things satisfying the conditions for X also satisfy the conditions for Y). In Figure 1, these axioms are illustrated as arrows with white triangular pointers. Subclass axioms do not always take the form of simple assertions of a subclass relation between two named classes, but may involve more complex class expressions as well. For example, the 'part of' axiom in Figure 1 would be expressed in OWL as:

```
'ovary septum' subClassOf: 'part of' some gynoecium
```

Here, 'ovary septum' is a named class in the ontology while `'part of' some gynoecium` is a complex class expression involving the relation 'part of', the named class 'gynoecium' and the existential quantifier `some`.

Ontologies that are formalized in OWL may contain many more kinds of axioms [29], and some ontologies that are formalized in more expressive languages than OWL, such as first- and second-order predicate logic [30], may contain a large variety of axioms. Examples of ontologies that are formalized at least in parts in such expressive languages include the RNA Ontology [31], the Basic Formal Ontology [32] or parts of the Sequence Ontology [33, 34].

The axioms and definitions in ontologies can give rise to a graph structure that can be exploited using graph- and

network-based algorithms. In these graphs, nodes commonly represent classes, and edges represent types of axioms that hold between these classes [35]. In particular, ontologies give rise to 'taxonomic graphs', which represent the subclass relations between the named classes in the ontology. Another pattern that is frequently used in generating a graph structure from ontology axioms is the existential restrictions on the 'part of' relation to give rise to a partonomy [36]. Here, an edge labeled 'part of' is generated between classes X and Y if X is a subclass of `'part of' some Y`. Importantly, the label of the edge between classes (e.g. 'part of') is different from the relation 'part of' that holds between the instances of the class [37]; the label of the edge is a shortcut for the complex axiom pattern involving the two classes (or a relation between the two classes that is explicitly defined using such an axiom pattern).

Using ontologies

Several tools and methods have been developed that make use of ontologies and support their use. These tools often focus on one or two of the features of ontologies, and here we distinguish them by the main task they aim to support.

Annotation and data integration

The use of standard identifiers for classes and relations in ontologies is a key component in enabling data integration across multiple databases, because the same identifiers can be used across multiple, disconnected databases, files or web sites. Consequently, these identifiers are widely used in structured file formats, in knowledge bases and data repositories. In fact, one of the first applications for which biological ontologies were developed, notably the GO [1], was to make biological sense of the large data sets emerging from the new expression array technologies in the early 2000s. Differential expression screens and Serial Analyis of Gene Expression (SAGE) analyses generated data sets of often thousands of genes, which needed to be interpreted in terms of gene function. This provided the impetus behind the ongoing functional and structural annotation of gene products, which is now available through the GO database [38] and is a mainstay of modern bioinformatics. In particular, ontologies enabled the assignation of functions to gene products and the ability to compare these functions computationally within and across species; these features have become key tools in functional and comparative genomics.

At its core, an ontology-based annotation associates an entity and an ontology class, and combines this assertion with metadata that contains, among others, information about who created the annotation, the date at which the annotation was created or the evidence that was considered. The entity that is annotated can be represented by an identifier in a database, referred to by a word or phrase in text, or even visually represented in an image [39, 40]. Annotation tools are concerned with recording the annotation in standard formats, performing basic quality checks and providing the metadata for the annotations, as well as suggesting or inferring ontology-based annotations using custom algorithms. For example, when the annotations refer to entities mentioned in text, annotation tools may use natural language processing techniques, such as named entity recognition and relation extraction, and when annotations refer to entities represented in images, image processing techniques may be applied.

The majority of annotation tools allow for the inclusion of provenance information, such as the evidence for an ontology-based annotation as recorded using the Evidence Code Ontology [41] or the Provenance Ontology [42]. Tools such as Domeo [43], an annotation framework applied among others by the Neuroscience Information Framework and the OpenPhacts projects, uses the Annotation Ontology [39] to formally capture provenance information associated with ontology-based annotations. Furthermore, an increasing number of annotation tools use the W3C Open Annotation Data Model [44], or are able to import and export annotations in this format.

Annotation tools that support curators through markup of literature are widely used to suggest possible annotations [45]. For example, the Textpresso software tool [46] was one of the first tools developed to support literature curation for GO, and is still extensively used in model organism databases [47]. Some annotation tools come with additional functionality to allow interactions between curators of data sets and ontology developers. For example, the Phenex tool was designed to support the phenotype annotation of character matrices in the Phenoscape project [48]. Phenex contains workflow elements and inbuilt reliability algorithms that aim to reduce curator workload [49]. Furthermore, Phenex also allows feedback to ontology developers to request new ontology classes that are needed to capture data accurately. While Phenex is primarily an annotation tool relying on input from literature and experts, other tools can incorporate domain-specific algorithms to aid in the annotation process. For example, the GO consortium [38] applies the Phylogenetic Annotation and Inference Tool, which assists curators to infer annotations among members of a protein family based on sequence orthology [50], making GO an interesting example of the confluence of the use of manual assignment based on published evidence, and electronic inference (by orthology or structural motif) to fill the gaps in our knowledge concerning gene product function and location.

Data integration and annotation go hand-in-hand, and in particular for complex multimodal data sets, annotation with single ontologies is often not sufficient. A particularly complex use-case of annotation with multiple ontologies occurs in the domain of phenotype descriptions, as applied in large-scale mutagenesis projects. For example, in the Zebrafish Mutagenesis Project [51], much of the observed data is categorical and describes anatomical and physiological variation, and the phenotypic descriptions are based on anatomy and process ontologies [51]. The International Mouse Phenotyping Consortium (IMPC) [52], on the other hand, generates both categorical data, which are assigned by investigators directly based on a phenotype ontology, and quantitative data. The strategy adopted by the IMPC is to express phenodeviance by assigning a class from a phenotype ontology on the basis of predetermined statistical thresholds [53, 54]. This form of automated annotation, albeit on highly quality-controlled data, is time-efficient and facilitates data integration and mining across qualitative and quantitative information.

When it becomes necessary to use more than a single ontology for annotation, it is beneficial to fix the ontologies that are being used to annotate a data set. Ontology repositories (Table 3) can aid in finding ontologies suitable for annotating data within a domain.

Ontologies as vocabularies

Ontologies provide vocabularies of the terms used within a domain. Therefore, they can be used by a large variety of applications that rely on domain-specific terms. Example applications for the vocabulary component of ontologies include user interfaces for databases that contain ontology-based annotations, and natural language processing methods.

Table 3. Overview of main ontology repositories in the life science domain

Repository	Key features	URL
BioPortal	BioPortal [126] is the largest ontology repository for ontologies in biology and biomedicine. It contains >400 ontologies with a total of >6 million classes. BioPortal can be used to find ontologies based on the ontology name or the label of a class within the ontology. It further has a large number of web services and widgets that allow embedding of key BioPortal functions in web applications. The NCBO Annotator [127] is a part of BioPortal and can be used to find labels of ontology classes in text. BioPortal can also be accessed through a SPARQL endpoint.	http://bioportal.bioontology.org/
OntoBee	Ontobee [128] is an ontology repository in which ontologies are presented as Linked Data. Ontobee provides information about the classes and relations used by the OBO project.	http://www.ontobee.org/
Ontology Lookup Service	The Ontology Lookup Service [129] consists of a repository of ontologies represented in the OBO Flatfile Format, and enables search of single ontologies, lookup of terms across multiple ontologies and browsing and visualizing the ontology graph structures. The Ontology Lookup Service can be accessed through a web interface and a number of web services.	http://www.ebi.ac.uk/ontology-lookup/
OBO Library	The Open Biological and Biomedical Ontologies (OBO) library [2] consists of a number of ontologies that have been developed according to a set of agreed principles including complementarity and collaborative development.	http://obofoundry.org

Tools using the vocabularies associated with ontologies use them in two main ways. First, the labels of an ontology classes and relations enable access to data or text annotated with these ontologies. For this type of application, a link is established between a class and a user-readable name of that class. This link is then used to provide a way for human users of an ontology to access the information associated with the ontology class. Tools that use this feature include a wide range of browsers that enable access to ontology-based annotations through the class labels, such as the Amigo tool [55], which enables access to GO annotations, or GOPubMed [56], which enables access to scientific articles based on ontology classifications.

Second, the labels in an ontology can be used to identify whether the text mentions a phenomenon characterized by a class or relation in an ontology. Applications of this type typically require the utilization of natural language processing techniques [57]. One example of such application is the NCBO Annotator, a tool that can recognize the labels and synonyms of ontology classes in natural language texts [58]. The National Center for Biomedical Ontologies (NCBO) Annotator implements a basic concept recognition approach [59] that generalized well across multiple vocabularies and does not require additional training. However, more specialized approaches have been developed, in particular in the context of recognizing descriptions of gene functions and biological processes in text [60], which can then be used to develop software tools that assist domain experts in literature-based database curation.

The labels of classes in ontologies can also be used for large-scale text mining to identify system-wide associations between the phenomena to which they refer. Text mining based on ontologies has been used to identify the presence of disease modules based on phenotypes [61, 62], drug targets and drug indications [63, 74], drug–drug interaction [65] and candidate genes for diseases [66, 67]. The success of these methods depends on the coverage of terms used to refer to classes in the ontology.

The main challenge in relying on class labels to recognize the reference to an ontology class in text is that labels do not capture all of the possible linguistic variations around terms and phrases used to refer to an ontology class [68]. Recognizing ontology classes referenced in text poses a distinct set of challenges, in particular for semantically complex classes, or classes for which no common and widely used terms have been established [69–71].

Formalized definitions and axioms: reasoning with ontologies

Several tools and software libraries can make use of ontologies' axioms and formal definitions. The primary means to access and process ontologies semantically are automated reasoners, i.e. software tools that can directly infer knowledge from the axioms and definitions in ontologies using deductive inference. Automated reasoners can detect contradictions in the axioms and definitions of an ontology (consistency checking), infer the most specific subclasses and superclasses for all classes in an ontology (classification) and answer complex queries. A wide range of automated reasoners has been developed for different subsets of OWL, supporting different features and exhibiting different computational complexity for basic reasoning tasks such as answering queries (Table 4). Reasoners for subsets of OWL such as OWL-EL support less expressivity for axioms and queries in ontologies, but usually guarantee a lower computational complexity. For complex ontologies expressed in OWL, examples of commonly used reasoners include Pellet [72] owing to its support for a large number of features, and HermiT [73] owing to its high performance for complex ontologies. For ontologies expressed in the OWL-EL profile, the ELK reasoner [74] is widely used owing to its support for large ontologies and parallel reasoning. Recent developments include the Konklude reasoner [75], which outperforms most OWL-EL and OWL 2 reasoners even for large ontologies [76]. As reasoner technology is evolving rapidly, new optimization methods can lead to significant performance improvements. If a selected reasoner cannot perform a reasoning task over an ontology, it can pay off to review reasoner competitions such as the annual OWL Reasoner Evaluation workshops [76] to find another reasoner that is more adequate for an ontology and desired application. Alternatively, ontology modularization approaches [23, 77–79] can be applied to extract subsets of ontologies, which automated reasoners can process efficiently.

OWL reasoners are either implemented as stand-alone tools, or can be accessed through the OWL API [80] or the OWLLink protocol [81]. The OWL API is a reference implementation for creating and manipulating OWL ontologies and provides interfaces for automated reasoning that the majority of OWL reasoners implement. OWLLink is an HTTP-based protocol for communicating with OWL reasoners. Reasoners can also be accessed through ontology editors such as Protege [82]. Table 5 provides an overview of some common tools and software libraries used to process ontologies and interact with reasoners, and Table 6 shows some common analysis and visualization tools and libraries that use ontologies.

Most users of ontologies will not access ontologies directly through automated reasoners, but will either use the output of an automated reasoner (e.g. the inferred graph structure of an ontology) or interact with a reasoner indirectly (e.g. through a software tool that uses an automated reasoner as part of its operation). Nevertheless, in some approaches, automated reasoning has been applied directly to verify data consistency with respect to constraints in an ontology or reveal novel biological knowledge based on axioms in an ontology. The axioms in an ontology can be used to verify whether an entity described in a database is able to satisfy the conditions laid out for that kind of entity, and automated reasoning can be used to detect conflicts. For example, such an approach has been applied retrospectively to computational models in systems biology [83], but is increasingly being applied to ontology-based annotations at the time the annotation is made [84, 85]. Some data exchange standards are now being designed with data verification in mind, and a prime example is the BioPAX standard for pathway data sharing, which is based on formalized knowledge in OWL [86]. The axioms in an ontology can also be used to infer the class to which an entity belongs based on the features and descriptions of the class and the entity. An application of this is the inference of the protein family to which a protein belongs based on an ontology and automated reasoning [87].

More subtly, reasoning over ontologies can also be applied for integrating ontology-annotated data sets across different domain by systematically combining different ontologies using axioms or axiom patterns [88, 89]. In such applications, the relationship between classes in different ontologies is identified and expressed in the form of an axiom or axiom pattern that is systematically applied to several pairs of classes. Prime examples of this form of integration are species-specific anatomy and phenotype ontologies [90, 91]. Integrating data annotated with these ontologies relies on identifying homologous anatomical structures [92] and relating the classes that refer to these structures in different anatomy ontologies using axiom patterns [90, 93].

Mining and analyzing multimodal data with ontologies

The great potential in using ontologies for data analysis lies with the possibility of combining their different functional levels, and some exciting insights into the biological properties of whole systems have been achieved by combining data through ontologies. For example, one of the most widely used applications for ontologies is Gene Set Enrichment Analysis [94] or similar enrichment methods, which combine the graph structure of ontologies (axioms and definitions) with their potential for data integration (through ontology-based annotations) to provide a statistical interpretation of differences between two states with regard to the background knowledge provided by the ontology over which the enrichment analysis was performed. Another analysis method specifically relying on ontologies and their annotations is the use of similarity measures to determine the 'semantic' distance and proximity between data items [95]. In semantic similarity measures, the axioms and definitions of ontologies are exploited to define a similarity between annotated data items. Semantic similarity has widely been applied to computationally predict protein–protein interactions based on their functional similarity [96, 97], to the diagnosis of disease based on phenotypic similarity [98–100], or to the classification of chemicals based on structural similarity [101].

While statistical analysis of graphs or sets, or measures of semantic similarity, are well established methods that use ontologies for data mining, many machine learning and data mining algorithms that are applied to unstructured data are not yet widely used with ontologies and ontology-structured data. The challenges of using these methods occur both when using ontologies and ontology-annotated data as the target of a machine learning and data mining algorithm as well as when using ontologies and ontology-annotated data as features. When using ontologies as the target, i.e. when aiming to learn an ontology-based classification for some piece of data such as the functions of a protein, several challenges arise in relation to the adoption of these traditional algorithms to ontology-based data in the biological and biomedical domains. These challenges primarily relate to the 'multi-class' nature of the problem, as ontologies have often very large numbers of classes, the 'structured dependency relations' between these classes (i.e. the axioms in the ontology) and, in many cases, the 'multi-label' nature of the classification problem as data items are usually annotated to more than one ontology class. When using ontologies, or ontology-annotated data, as features in a machine learning task, challenges relate to the large number of classes that are often sparsely populated (more specific classes are usually present less frequently while more general classes are used more frequently), and again the dependency relations between classes (e.g. disjointness, subclass relations and axiom patterns that exist between classes).

Despite these challenges, progress is being made in incorporating ontologies and ontology-annotated data into machine learning and data mining algorithms. For example, in the area of prediction of protein functions, driven by the Critical Assessment of Function Annotation challenge [102], several approaches have been developed to predict GO annotations of proteins [103–106], some of which use ontologies as features as well [104, 107]. While these methods have been developed in the context of protein function prediction, parts of these can be transferred to other problems.

The use of ontologies can also help address a challenge that machine learning and data mining approaches face: the incorporation of different types of features for multimodal learning and classification [108]. Combining information from text, images, videos, molecular data or structured data in knowledge bases to improve classification can be facilitated through the use ontologies, by first extracting relevant features from each type of information and representing the results using a single ontology that combines the information used for training a classifier.

Perspective

There are now sufficient stable ontologies to permit routine reuse of classes from multiple ontologies in automated or semiautomated ontology construction algorithms [109]. With increasing size and number of ontologies, the ability to modularize ontologies to generate application-specific 'views' while maintaining interoperability with data sets in a domain

that are annotated with another module of the same ontology will become essential. A recent example of this is provided by the Bioassay ontology [110] or the automated generation of phenotype ontologies [111, 112]. To support these applications, coverage and quality of content in established ontologies must be further improved [113], a task that poses a serious challenge and requires the sustained engagement of domain experts.

One major application of exploiting multiple ontologies is to formalize the large, unstructured, multimodal and often distributed data from clinical records. It is now possible to capture information and knowledge related to diagnostic procedures, drugs, phenotypes, diseases and genotypes using existing ontologies, and there are efforts to create ontologies for capturing other environmental and behavioral data for patients. Such ontologies are now being applied in a clinical setting [114], but mainly for data mining from partially structured and legacy clinical records [115]. Incorporating ontologies directly in the electronic health record will lead to novel methods for patient classification and stratification, and the analysis and mining of large-scale patient data. With increasing numbers of whole exome and genome sequences in clinics, there is marked potential for using ontology-based enrichment algorithms or incorporating results from basic biological research into clinical decision making [116]. We expect to see further rapid developments in this important area.

From an algorithmic and methodological point of view, the next challenge we face is the development of new methods for applying ontologies in data mining and data analysis. These methods must be able to use the different features ontologies provide, combine them in meaningful ways and be applicable to large, complex and multimodal data sets. We also expect to see more complex ontology-based applications that combine the main features of ontologies in novel ways. For example, annotation tools will be developed that do not merely use the labels and class identifiers to associate entities with ontology classes, but use the ontology's axioms and formal definitions to preselect possible annotations (e.g. by eliminating possible process classes at places at which only annotations to material objects would be sensible), verify the consistency of an annotation [83], reveal the consequences of asserted information to users [87] and be applicable to multiple types of data (e.g. structured data, text and images).

Finally, to further improve ontology-based data integration and analysis, robust evaluation criteria need to be developed that are based on how ontologies are actually being used in research applications [117]. Recently, some exciting results have demonstrated that the GO accurately resembles modules found in experimentally derived gene and protein interaction networks, leading to a data-driven way for validating an ontology [118]. The increasing use of ontologies in scientific research will lead to improved methods for evaluating ontology quality based on their performance in scientific applications [117, 119]. A tighter integration between experimental results and the domain knowledge formalized in ontologies will not only lead to

Table 4. A selection of automated reasoners for OWL ontologies

Reasoner	OWL support	Description
Pellet [72]	OWL 2, OWL EL	General purpose OWL reasoner with a large set of features, including specialized OWL EL reasoning, support for rules, support of epistemic operators, integration in SPARQL, explanation of inferences, incremental reasoning.
HermiT [73]	OWL 2, OWL EL	General purpose, highly optimized OWL reasoner.
FacT++ [130]	OWL-DL, OWL 2 (partially)	Highly optimized reasoner implemented in C++.
Konklude [75]	OWL 2	Highly optimized OWL reasoner supporting parallel reasoning.
RacerPro 2.0 [131]	OWL 2 (partial)	Optimized OWL reasoner, with integration in the AllegroGraph [132] triple store.
TrOWL [133]	OWL 2	Scalable OWL reasoner with support for limited closed-world reasoning (negation as failure) and stream reasoning.
ELK [74]	OWL-EL	Optimized and feature-rich OWL EL reasoner with support for incremental and parallel reasoning.

Table 5. An overview over tools and software libraries for processing and interacting with ontologies

Tool	Description	Web site
Protege, WebProtege	Protege [82] is an OWL ontology editor with full support for OWL ontologies and a large number of plug-ins that provide integration of reasoners, export and import of various ontology representation formats, or ontology visualization. WebProtege is a web-based collaborative ontology editor, which provides similar functionality to Protege through a web interface.	http://protege.stanford.edu/, http://webprotege.stanford.edu/
OWL API	The OWL API [80] is a reference implementation and a *de facto* standard for processing OWL ontologies.	http://owlapi.sourceforge.net/
Owlcpp	owlcpp [134] is a C++ library for processing OWL ontologies. It includes support for querying ontologies through automated reasoners.	http://owl-cpp.sourceforge.net/
Brain	Brain [135] is a library based on the OWL API that provides convenience methods for processing and reasoning with ontologies, in particular biological and biomedical ontologies represented in the OWL-EL profile of OWL.	https://github.com/loopasam/Brain
Redland RDF API	An RDF library written in C. It provides a large set of commonly used command line tools to transform or collect basic statistics about an RDF file.	http://librdf.org/
Apache Jena	Jena is a Java library and collection of tools consisting of an RDF library, integration of SPARQL queries and support for OWL ontologies.	https://jena.apache.org/

Table 6. An overview over generic ontology analysis and visualization tools and libraries

Tool	Description	Web site
Gephi	Gephi [136] is a generic graph-visualization tool, and can be used to visualize classes and relations in ontologies. Gephi also supports a number of algorithms for basic graph analysis, including transitive inference over edges.	http://gephi.github.io/
Cytoscape	Cytoscape [137] is a tool for visualizing and analyzing interaction networks and other graphs including ontologies. Several Cytoscape plug-ins support using ontologies for visualization and analysis.	http://www.cytoscape.org/
Semantic Measures Library	The Semantic Measures Library and Toolkit [138] is a generic framework implementing a large variety of semantic similarity measures over ontologies.	http://www.semantic-measures-library.org/
GO enrichment analysis tools	Enrichment analysis uses the graph-structure underlying ontologies (usually the GO) together with transitive inference over the edges in the graph to statistically test a hypothesis. The graph structure is used to 'enrich' statistical power by propagating annotations transitively over the graph and performing a test at each level of the ontology hierarchy.	http://geneontology.org/page/go-enrichment-analysis
OntoFUNC	OntoFUNC [139] is a software tool to perform ontology enrichment analysis over arbitrary OWL ontologies.	http://phenomebrowser.net/ontofunc/

improved evaluation criteria and subsequently better ontologies, but is also a crucial step in making sense of large structured and unstructured data sets in biology and biomedicine.

Key Points

- Ontologies provide identifiers for classes and relations that represent phenomena within a domain, thereby enabling integration of data.
- Ontologies provide labels for classes and relations, thereby providing a domain vocabulary.
- Ontologies provide metadata associated with classes and relations that allows human users to understand their meaning and contribute to consistent use in annotation and other applications.
- Ontologies provide axioms and formal definitions that enable computational access to some aspects of the meaning of classes and relations.
- Combining the four main features of ontologies facilitates semantic integration of heterogeneous, multimodal data within and across domains, and enables novel data mining methods that span traditional boundaries between domains and data types.

Funding

This work has not received any dedicated funding.

References

1. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–29.
2. Smith B, Ashburner M, Rosse C, *et al*. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech* 2007;**25**:1251–5.
3. Musen MA, Noy NF, Shah NH, *et al*. The National Center for Biomedical Ontology. *J Am Med Inform Assoc* 2012;**19**:190–5.
4. Gruber TR, Olsen G. An ontology for engineering mathematics in principles of knowledge representation and reasoning. In: J Doyle, P Torasso, E Sandewall (eds). *Proceedings of the 4th International Conference (KR '94):Bonn, Germany, May 24–27, 1994*. Morgan Kaufmann Publishers, Burlington, Massachusetts, USA, 1994, 258–69.
5. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum Comput Stud* 1995;**43**(5/6):907–28.
6. Guarino N. Formal ontology and information systems. In: N Guarino (ed). *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems*. Amsterdam, Netherlands: IOS Press, 1998, 3–15.
7. Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. *AMIA Annu Symp Proc* 2003;**2003**:609–13.
8. Baader F, Calvanese D, McGuinness D, *et al*. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.
9. Heller B, Herre H. Ontological categories in GOL. *Axiomathes* 2004;**14**:57–76.
10. Herre H, Loebe F. A meta-ontological architecture for foundational ontologies. In: *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. Springer Verlag, Heidelberg, Germany, 2005, 1398–415.
11. Kumar A, Smith B. In: A Günter, R Kruse, B Neumann (eds). *KI2003: Advances in AI*. 2003, 135–48.
12. Bada M, Stevens R, Goble C, *et al*. A short study on the success of the Gene Ontology. *Web Semant* 2004;**1**:235–40.
13. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. *Brief Bioinform* 2006;**7**:256–74.
14. Smith B. Ontology (Science). In: *Proceeding of the 2008 conference on Formal Ontology in Information Systems: Proceedings of the Fifth International Conference (FOIS 2008)*. IOS Press, Amsterdam, The Netherlands, 2008, 21–35.
15. Merrill GH. Ontological realism: methodology or misdirection? *Appl Ontol* 2010;**5**:79–108.
16. Berners-Lee T, Hendler J, Lassila O, *et al*. The Semantic Web. *Sci Am* 2001;**284**:28–37.
17. Ruttenberg A, Courtot M, Mungall CJ. *OBO Foundry Identifier Policy*. 2013. http://www.obofoundry.org/id-_policy.shtml (2 November 2014 date last accessed).
18. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998;**37**:394–403.

19. Stevens R, Hull D. *Separating Concepts from Labels*. 2010. http://ontogenesis.knowledgeblog.org/818
20. Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the foundational model of anatomy. *J Biomed Inform* 2003;**36:**478–500.
21. Munn K, Smith B (eds). *Applied Ontology: An Introduction*, 1st edn. Ontos Verlag, Heusenstamm, Germany, 2009.
22. Burgun A, Bodenreider O, Jacquelinet C. Issues in the classification of disease instances with ontologies. *Stud Health Technol Inform* 2005;**116**:695–700.
23. Grau B, Horrocks I, Motik B, *et al*. OWL 2: the next step for OWL. *Web Semant* 2008;**6**:309–322.
24. Baader F. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.
25. Horrocks I, Kutz O, Sattler U. The even more irresistible SROIQ. In: P Doherty, J Mylopoulos, CA Welty (eds). *KR*. Palo Alto, California, USA: AAAI Press, 2006, 57–67.
26. Motik B, Grau BC, Horrocks I, *et al*. *OWL 2 Web Ontology Language: Profiles Recommendation*. World Wide Web Consortium (W3C), Cambridge, Massachusetts, USA, 2009.
27. Horrocks I. *OBO Flat File Format Syntax and Semantics and Mapping to OWL Web Ontology Language*. University of Manchester, 2007. http://www.cs.man.ac.uk/∼horrocks/obo/
28. Henkin L, Suppes P, Tarski A. The axiomatic method with special reference to geometry and physics. In: *Proceedings of an International Symposium Held at the University of California, Berkeley, December 26, 1957–January 4, 1958*. North-Holland Publishing Co, Amsterdam, 1959.
29. Stevens R, Stevens M. A family history knowledge base using OWL 2. In: C Dolbear, A Ruttenberg, U, Sattler. *OWLED,* **Vol. 432**. Aachen, Germany: CEUR-WS.org, 2008.
30. Barwise J. *Model-Theoretic Logics (Perspectives in Mathematical Logic)*. Springer, Heidelberg, Germany, 1985.
31. Hoehndorf R, Batchelor C, Bittner T, *et al*. The RNA Ontology (RNAO): an ontology for integrating RNA sequence and structure data. *Appl Ontol* 2011;**6**:53–89.
32. Grenon P. *BFO in a Nutshell: A Bi-categorial Axiomatization of BFO and Comparison with DOLCE. IFOMIS Report 06/2003*. Institute for Formal Ontology and Medical Information Science (IFOMIS), University of Leipzig, Leipzig, Germany, 2003.
33. Eilbeck K, Lewis SE, Mungall CJ, *et al*. The sequence ontology: a tool for the unification of genome annotations. *Genome Biol* 2005;**6**:R44.
34. Hoehndorf R, Kelso J, Herre H. The ontology of biological sequences. *BMC Bioinformatics* 2009;**10**:377+.
35. Hoehndorf R, Oellrich A, Dumontier M, *et al*. Relations as patterns: bridging the gap between OBO and OWL. *BMC Bioinformatics* 2010;**11**:441+.
36. Schulz S, Hahn U. Part-whole representation and reasoning in formal biomedical ontologies. *Artif Intell Med* 2005;**34**: 179–200.
37. Smith B, Ceusters W, Klagges B, *et al*. Relations in biomedical ontologies. *Genome Biol* 2004;**6**:R46.
38. The Gene Ontology Consortium. Gene ontology annotations and resources. *Nucleic Acids Res* 2013;**41**:D530–5.
39. Ciccarese P, Ocana M, Garcia Castro L, *et al*. An open annotation ontology for science on web 3.0. *J Biomed Semantics* 2011; **2**:S4.
40. Lingutla N, Preece J, Todorovic S, *et al*. AISO: annotation of image segments with ontologies. *J Biomed Semantics* 2014;**5**:50.
41. Chibucos MC, Mungall CJ, Balakrishnan R, *et al*. Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database* 2014;2014:pii: bau075.
42. Belhajjame K, Cheney J, Corsar D, *et al*. *PROV-O: The PROV Ontology*. 2012. http://www.w3.org/TR/prov-o/.
43. Ciccarese P, Ocana M, Clark T. Open semantic annotation of scientific publications using DOMEO. *J Biomed Semantics* 2012;**3**:S1.
44. Sanderson R, Ciccarese P, de Sompel HV. Designing the W3C open annotation data model. *CoRR* 2013;abs/1304.6709.
45. Hirschman L, Burns GAPC, Krallinger M, *et al*. Text mining for the biocuration workflow. *Database* 2012;2012:bas020.
46. Müller HM, Kenny EE, Sternberg PW. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol* 2004;**2**:e309+.
47. Dowell KG, McAndrews-Hill MS, Hill DP, *et al*. Integrating text mining into the MGI biocuration workflow. *Database (Oxford)* 2009;2009:bap019.
48. Dahdul WM, Balhoff JP, Engeman J, *et al*. Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. *PLoS One* 2010;**5**:e10708.
49. Balhoff J, Dahdul W, Dececchi T, *et al*. Annotation of phenotypic diversity: decoupling data curation and ontology curation using Phenex. *J Biomed Semantics* 2014;**5**:45.
50. Gaudet P, Livstone MS, Lewis SE, *et al*. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform* 2011;**12**:449–62.
51. Mabee P, Balhoff JP, Dahdul WM, *et al*. 500,000 fish phenotypes: the new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton. *J Appl Ichthyol* 2012;**28**:300–5.
52. Brown SDM, Moore MW. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis Model Mech* 2012;**5**:289–292.
53. Beck T, Morgan H, Blake A, *et al*. Practical application of ontologies to annotate and analyse large scale raw mouse phenotype data. *BMC Bioinformatics* 2009;**10**:S2+.
54. Koscielny G, Yaikhom G, Iyer V, *et al*. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res* 2014;**42**:D802–9.
55. Carbon S, Ireland A, Mungall CJ, *et al*. AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009;**25**:288–9.
56. Doms A, Schroeder M. GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Res* 2005;**33**:783–6.
57. Rebholz-Schuhmann D, Oellrich A, Hoehndorf R. Text-mining solutions for biomedical research: enabling integrative biology. *Nat Rev Genet* 2012;**13**:829–39.
58. Jonquet C, Shah NH, Musen MA. The open biomedical annotator. *Summit Translat Bioinforma* 2009;**2009**:56–60.
59. Ghazvinian A, Noy NF, Musen MA. Creating mappings for ontologies in biomedicine: simple methods work. *AMIA Annu Symp Proc* 2009;**2009**:198–202.
60. Mao Y, Van Auken K, Li D, *et al*. Overview of the gene ontology task at BioCreative IV. *Database* 2014;**2014**:bau086.
61. Van Driel MA, Bruggeman J, Vriend G, *et al*. A text-mining analysis of the human phenome. *Eur J Hum Genet* 2006;**14**: 535–42.
62. Zhou X, Menche J, Barabási A-L, *et al*. Human symptoms–disease network. *Nat Commun* 2014;**5**:4212.
63. Campillos M, Kuhn M, Gavin ACC, *et al*. Drug target identification using side-effect similarity. *Science* 2008;**321**:263–6.
64. Kuhn M, Campillos M, Letunic I, *et al*. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;**6**:343.
65. Percha B, Garten Y, Altman RB. Discovery and explanation of drug-drug interactions via text mining. *Pac Symp Biocomput* 2012;**2012**:410–21.

66. Oellrich A, Gkoutos GV, Hoehndorf R, *et al*. Quantitative comparison of mapping methods between Human and Mammalian Phenotype Ontology. *J Biomed Semantics* 2012; **3**:S1.
67. Hoehndorf R, Schofield PN, Gkoutos GV. An integrative, translational approach to understanding rare and orphan genetically based diseases. *Interface Focus* 2013;**3**: 20120055.
68. Rebholz-Schuhmann D, Kim JH, Yan Y, *et al*. Evaluation and cross-comparison of lexical entities of biological interest (LexEBI). *PLoS One* 2013;**8**:e75185.
69. Funk C, Baumgartner W, Garcia B, *et al*. Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinformatics* 2014; **15**:59.
70. Hunter L, Lu Z, Firby J, *et al*. OpenDMAP: an open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC Bioinformatics* 2008;**9**:78.
71. Shah N, Bhatia N, Jonquet C, *et al*. Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinformatics* 2009;**10(**Suppl 9**)**:S14.
72. Sirin E, Parsia B, Grau BC, *et al*. Pellet: a practical OWL-DL reasoner. *Web Semant* 2007;**5**:51–3.
73. Motik B, Shearer R, Horrocks I. Hypertableau reasoning for description logics. *J Artif Intell Res* 2009;**36**:165–228.
74. Kazakov Y, Krötzsch M, Simancik F. The incredible ELK. *J Autom Reason* 2014;**53**:1–61.
75. Steigmiller A, Liebig T, Glimm B. Konclude: system description. *Web Semant* 2014:27.
76. Bail S, Glimm B, Jiménez-Ruiz E, *et al*. (eds). *ORE 2014: OWL Reasoner Evaluation Workshop.CEUR Workshop Proceedings 1207*. CEUR-WS.org, Aachen, Germany, 2014.
77. Hoehndorf R, Dumontier M, Oellrich A, *et al*. A common layer of interoperability for biomedical ontologies based on OWL EL. *Bioinformatics* 2011;**27**:1001–8.
78. Rector AL. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL *in* K-CAP '03. In: *Proceedings of the 2nd International Conference on Knowledge Capture*. ACM Press, New York, NY, 2003, 121–8.
79. Jiménez-Ruiz E, Grau BC, Sattler U, *et al*. English. In: S Bechhofer, M Hauswirth, J Hoffmann, *et al*. (eds). *The Semantic Web: Research and Applications*. Springer, Berlin, Heidelberg, 2008, 185–199.
80. Horridge M, Bechhofer S, Noppens O. Igniting the OWL 1.1 Touch Paper: The OWL API. In: *Proceedings of OWLED 2007: Third International Workshop on OWL Experiences and Directions*. 2007. CEUR-WS.org, Aachen, Germany.
81. Noppens O, Luther M, Liebig T. The OWLlink API: Teaching OWL Components a Common Protocol. In: E Sirin, K Clark (eds). *Proceedings of the 7th International Workshop on OWL: Experiences and Directions (OWLED 2010), Vol. 614 of CEUR Workshop Proceedings*, 2010. http://ceur-_ws.org.
82. Noy NF, Sintek M, Decker S, *et al*. Creating semantic web contents with protege-2000. *IEEE Intell Syst* 2001;**16**:60–71.
83. Hoehndorf R, Dumontier M, Gennari JH, *et al*. Integrating systems biology models and biomedical ontologies. *BMC Syst Biol* 2011;**5**:124+.
84. Lenzerini M, Console M. Data quality in ontology-based data access: the case of consistency. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* 2014, 1020–6. AAAI Press, Palo Alto, California, USA.
85. Tao C, Song D, Sharma D, *et al*. Semantator: semantic annotator for converting biomedical text to linked data. *J Biomed Inform* 2013;**46**:882–93.
86. Demir E, Cary MP, Paley S, *et al*. The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 2010;**28**:935–42.
87. Wolstencroft K, Lord P, Tabernero L, *et al*. Protein classification using ontology classification. *Bioinformatics* 2006;**22**:e530–8.
88. Egaña M, Rector A, Stevens R, *et al*. Applying ontology design patterns in bio-ontologies. In: A Gangemi, J Euzenat (eds). *Knowledge Engineering: Practice and Patterns, Vol. 5268 of Lecture Notes in Computer Science*. 2008, 7–16.
89. Hoehndorf R, Dumontier M, Oellrich A, *et al*. Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning. *PLOS One* 2011;**6**:e22006.
90. Mungall C, Gkoutos G, Smith C, *et al*. Integrating phenotype ontologies across multiple species. *Genome Biol* 2010;**11**:R2. doi:10.1186/gb-2010-11-1-r2, R2+ (2010).
91. Washington NL, Haendel MA, Mungall CJ, *et al*. Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol* 2009;**7**:e1000247.
92. Mungall C, Torniai C, Gkoutos G, *et al*. Uberon, an integrative multi-species anatomy ontology. *Genome Biol* 2012;**13**:R5.
93. Hoehndorf R, Oellrich A, Rebholz-Schuhmann D. Interoperability between phenotype and anatomy ontologies. *Bioinformatics* 2010;**26**:3112–18.
94. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;**102**:15545–50.
95. Pesquita C, Faria D, Falcao AO, *et al*. Semantic similarity in biomedical ontologies. *PLoS Comput Biol* 2009;**5**:e1000443**.** DOI:10.1371/journal.pcbi.1000443, e1000443.
96. Guzzi PH, Mina M, Guerra C, *et al*. Semantic similarity analysis of protein data: assessment with biological features and issues. *Brief Bioinform* 2011;**13**:569–85.
97. Benabderrahmane S, Smail-Tabbone M, Poch O, *et al*. IntelliGO: a new vector-based semantic similarity measure including annotation origin. *BMC Bioinformatics* 2010;**11**:588.
98. Köhler S, Schulz MH, Krawitz P, *et al*. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 2009;**85**:457–64.
99. Hoehndorf R, Schofield PN, Gkoutos GV. PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Res* 2011;**39**:e119.
100. Schlicker A, Albrecht M. FunSimMat update: new features for exploring functional similarity. *Nucleic Acids Res* 2010;**38**: D244–8.
101. Ferreira JD, Couto FM. Semantic similarity for automatic classification of chemical compounds. *PLoS Comput Biol* 2010; **6**:e1000937.
102. Radivojac P, Clark WT, Oron TR, *et al*. A large-scale evaluation of computational protein function prediction. *Nat Meth* 2013;**10**:221–7.
103. Melvin I, Ie E, Kuang R, *et al*. SVM-Fold: a tool for discriminative multi-class protein fold and superfamily recognition. *BMC Bioinformatics* 2007;**8**:S2.
104. Sokolov A, Funk C, Graim K, *et al*. Combining heterogeneous data sources for accurate functional annotation of proteins. *BMC Bioinformatics* 2013;**14**:S10.
105. Cozzetto D, Buchan D, Bryson K, *et al*. Protein function prediction by massive integration of evolutionary analyses and multiple data sources. *BMC Bioinformatics* 2013;**14**:S1.

106. Lan L, Djuric N, Guo Y, *et al*. MS-kNN: protein function prediction by integrating multiple data sources. *BMC Bioinformatics* 2013;**14**:S8.
107. Sokolov A, Ben-Hur A. Hierarchical classification of gene ontology terms using the GOStruct method. *J Bioinform Comput Biol* 2010;**8**:357–76.
108. Skowron A, Wang H, Wojna A, *et al*. Multimodal classification: case studies. In: JF Peters, A Skowron (eds). *Transactions on Rough Sets V*. Springer-Verlag, Berlin, Heidelberg, 2006, 224–39.
109. Huang J, Dang J, Borchert GM, *et al*. OMIT: dynamic, semi-automated ontology development for the microRNA domain. *PLoS One* 2014;**9**:e100855.
110. Abeyruwan S, Vempati UD, Kucuk-McGinty H, *et al*. Evolving BioAssay Ontology (BAO): modularization, integration and applications. *J Biomed Semantics* 2014;**5**:S5.
111. Köhler S, Doelken SC, Ruef BJ, *et al*. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. *F1000Research* 2013;**2**:30.
112. Vos R, Biserkov J, Balech B, *et al*. Enriched biodiversity data as a resource and service. *Biodiversity Data J* 2014;**2**:e1125.
113. Pathak J, Kiefer RC, Bielinski SJ, *et al*. Applying semantic web technologies for phenome-wide scan using an electronic health record linked Biobank. *J Biomed Semantics* 2012;**3**:10.
114. Girdea M, Dumitriu S, Fiume M, *et al*. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat* 2013;**34**:1057–65.
115. Denny JC. Chapter 13: Mining electronic health records in the genomics era. *PLoS Comput Biol* 2012;**8**:e1002823.
116. Zemojtel T, Köhler S, Mackenroth L, *et al*. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med* 2014;**6**:252ra123.
117. Hoehndorf R, Dumontier M, Gkoutos GV. Evaluation of research in biomedical ontologies. *Brief Bioinform* 2013;**14**:696–712.
118. Dutkowski J, Kramer M, Surma MA, *et al*. A gene ontology inferred from molecular networks. *Nat Biotechnol* 2012;**31**:38–45.
119. Verspoor K, Dvorkin D, Cohen KB, *et al*. Ontology quality assurance through analysis of term transformations. *Bioinformatics* 2009;**25**:i77–84.
120. Manola F, Miller E (eds). *RDF Primer*. World Wide Web Consortium, Cambridge, Massachusetts, USA, 2004.
121. Seaborne A, Prud'hommeaux E. *SPARQL Query Language for RDF*. W3C Recommendation (W3C, 2008). http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/
122. Heath T, Bizer C. *Linked Data: Evolving the Web into a Global Data Space*, 1st edn. San Rafael, California, USA: Morgan & Claypool, 2011.
123. Nelson SJ, Schopen M, Savage AG, *et al*. The MeSH translation maintenance system: Structure, interface design, and implementation. In *Proceedings of the 11th World Congress on Medical Informatics*. IOS Press, Amsterdam, 2004, 67–9.
124. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;**32,** D267–70.
125. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008;**8**:S2.
126. Noy NF, Shah NH, Whetzel PL, *et al*. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009;**37**:W170–3.
127. Whetzel P, Team, N. NCBO technology: powering semantically aware applications. *J Biomed Semantics* 2013;**4**:S8.
128. Xiang Z, Mungall CJ, Ruttenberg A, *et al*. Ontobee: a linked data server and browser for ontology terms. In: *Proceedings of International Conference on Biomedical Ontology* 2011, 279–81. CEUR-WS.org, Aachen, Germany.
129. Cote R, Jones P, Apweiler R, *et al*. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* 2006;**7**:97+.
130. Tsarkov D, Horrocks I. FaCT++ description logic reasoner: system description. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Third International Joint Conference on Automated Reasoning, Seattle, USA. IJCAR 2006*, Vol. 4130 LNAI. Springer Verlag, Heidelberg, Germany. 2006, 292–7.
131. Haarslev V, Hidde K, Möller R, *et al*. The RacerPro knowledge representation and reasoning system. *Semantic Web J* 2012;**3:**267–77.
132. W3C. *AllegroGraph RDFStore Web 3.0's Database*. 2009. http://www.franz.com/agraph/allegrograph/
133. Thomas E, Pan JZ, Ren Y. TrOWL: tractable OWL 2 reasoning infrastructure. In: *Proceeding of the Extended Semantic Web Conference (ESWC2010)* Springer Verlag, Heidelberg, Germany, 2010.
134. Levin MK, Ruttenberg A, Masci AM, *et al*. owl_cpp, a C++ Library for Working with OWL Ontologies. In: O Bodenreider, ME Martone, A Ruttenberg (eds). *Proceedings of the 2nd International Conference on Biomedical Ontology* CEUR-WS.org, Aachen, Germany 2011, 40.
135. Croset S, Overington JP, Rebholz-Schuhmann D. Brain: biomedical knowledge manipulation. *Bioinformatics* 2013;**29**:1238–9.
136. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: E Adar, M Hurst, T Finin, *et al*. (eds). *International AAAI Conference on Weblogs and Social Media (ICWSM)*. Palo Alto, California, USA: The AAAI Press, 2009, 361–2.
137. Smoot ME, Ono K, Ruscheinski J, *et al*. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;**27**:431–2.
138. Harispe S, Ranwez S, Janaqi S, *et al*. The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. *Bioinformatics* 2014;**30**:740–2.
139. Hoehndorf R, Hancock JM, Hardy NW, *et al*. Analyzing gene expression data in mice with the Neuro Behavior Ontology. *Mamm Genome* 2014;**25**:32–40.