ISSN 2079-0597, Russian Journal of Genetics: Applied Research, 2016, Vol. 6, No. 7, pp. 749–758. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © N.L. Podkolodnyy, O.A. Podkolodnaya, 2015, published in Vavilovskii Zhurnal Genetiki i Selektsii, 2015, Vol. 19, No. 6, pp. 652–660.

Ontologies in Bioinformatics and Systems Biology

N. L. Podkolodnyy[a, b, c] and O. A. Podkolodnaya[a]

[a] *Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia*
[b] *Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia*
[c] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: pnl@bionet.nsc.ru

Received September 28, 2015; in final form, October 29, 2015

Abstract—Computer simulation is now becoming a central scientific paradigm of systems biology and a basic tool for the theoretical study and understanding of the complex mechanisms of living systems. The increase in the number and complexity of these models leads to the need for their collaborative development, reuse of models, their verification, and the description of computational experiments and their results. Ontological modeling is used to develop formats for knowledge-oriented mathematical modeling of biological systems. In this sense, ontology associated with the entire set of formats that support research in systems biology, in particular, computer modeling of biological systems and processes, can be regarded as a first approximation to the ontology of systems biology. This review summarizes the features of the subject area (bioinformatics, systems biology, and biomedicine), the main motivation for the development of ontologies and the most important examples of ontological modeling and semantic analysis at different levels of the hierarchy of knowledge: molecular genetic, cellular, tissue, organs, and the body. Bioinformatics and systems biology is an excellent ground for testing the technologies and efficient use of ontological modeling. Several dozens of verified basic reference ontologies now represent a source of knowledge for the integration and development of more complex domain models aimed at addressing specific issues in biomedicine and biotechnology. Further formalization and ontological accumulation of knowledge and the use of formal methods of analysis can take the entire cycle of research in systems biology to a new technological level.

Keywords: ontological modeling, bioinformatics, systems biology
DOI: 10.1134/S2079059716070091

INTRODUCTION

The emergence of qualitatively new opportunities for research, based on the use of high-performance experimental technologies of massive parallel DNA sequencing, multilocus genotyping, multiparameter gene expression profiling using DNA chips, ChiP-on-chip technologies, proteomic and metabolomic technologies, etc., have led to the accumulation of unprecedentedly large amounts of experimental data and knowledge.

The huge amount of molecular biological information, its complexity, and the large number of barriers—technological, informational, resource, etc.—have complicated its analysis, organization, and application in specific problems of bioinformatics, biotechnology, pharmacology, personalized medicine, etc. To assimilate, organize and effectively use this kind of information, one requires new approaches to processing large data (BIG DATA), in particular, automated methods of semantic integration of heterogeneous data, one of the main stages of which is the coordination of domain concepts, methods of their description and use (comparison, data processing, etc.). Such a consistent description of a particular domain is called ontology.

Ontology development is a complex and costly process. The first stage of this process is the ontological analysis and modeling of the domain, including the creation of a glossary of terms, their precise definitions and relationships between them, as well as rules and regulations, according to which the introduced terminology is used to form credible allegations describing the state of the object under study.

Why do we need ontologies? Ontologies make it possible to present concepts in a way that they become suitable for machine processing and consequently are used as an intermediary between the user and the information system, or between members of the scientific community during data exchange. It is important for a molecular biologist to be able to describe molecular events, interacting components, and the roles played by these components in molecular events and processes, as well as to evaluate hypotheses. Bioinformaticists are interested in data integration, computer annotation, and modeling processes and systems. The general demand is the use of ontologies in education.

As a scientific discipline, systems biology emerged with the advent of possibilities to construct portrait models of biological systems and processes through the integration and joint computer analysis of a large amount of these fundamentally new experimental data describing the behavior of molecular-genetic systems as a whole. Systems biology studies biological objects and their complex, hierarchically organized networks of interactions, controlled by the information encoded in the genomes (Kitano, 2002).

In this regard, at present ontological analysis is one of the basic tools of bioinformatics and systems biology, used for the semantic integration of experimental data and knowledge in order to build "a unified picture of the world" (Podkolodnyy, 2011).

FORMAL REPRESENTATION OF ONTOLOGIES

In computer science, the term ontology is a conceptual model of representation of objects, properties of objects, and relations between them (Chandrasekaran et al., 1999). The ontology includes a set of concepts (terms) of the domain, their definitions and attributes, as well as an associated set of axioms and rules of inference (Gruber, 1995).

Thus, the formal ontology model is an ordered triple of finite sets $O = \langle T, R, F \rangle$, where T is a finite and nonempty set of classes and concepts (concepts, terms) of the domain as part of the real world, considered within a given context (in our case, bioinformatics and systems biology), which is described by ontology O; R is a finite set of relationships between the concepts of a given domain; F is a finite set of interpretation functions given by the concepts and/or relationships of ontology O or axioms used for modeling statements that are always true, which limits the interpretation and ensures correct use of the concepts.

One of the most productive approaches to describe and use knowledge about the domain is descriptive logics (DL) that define a formal language for the description of concepts (term, class, category, or entity) and their relationships (called roles), statements of facts, and requests to them. In addition, DLs include constructors (operations) for conceptual expressions, including conjunction, disjunction, and definition of relationships.

In terms of descriptive logic, domain knowledge bases are divided into general knowledge about a variety of classes of domain concepts, properties, and relationships between them (terminological knowledge, or T-Box) and knowledge about individual objects (class instances), their properties, and relationships with other objects (assertional knowledge, or A-Box); i.e., they describe the domain at the level of specific data (database). Both the components in the knowledge base are interconnected.

In general, the creation of applied ontologies focused on a specific domain can be greatly accelerated by using previously developed canonical (reference) ontologies for the construction of ontological classes and relationships between them.

Specifically, such a reference ontology may be an upper level ontology or basic knowledge ontology that describes the most general concepts (space, time, material, object, system, state, behavior, event, process, action, structure, function, etc.) and relationships (part–whole, general–specific, taxonomy, influence, cause, result, regulation, association, similarity, as well as spatial and temporal relationships, etc.). As these concepts do not depend on a particular issue or domain, it seems reasonable to unify them for large user communities.

OPEN BIOLOGICAL ONTOLOGIES

The project Open Biological Ontologies (OBO) is aimed at developing unified approaches to create ontologies, methods of their integration, and the tools to work with them (Bada and Hunter, 2007; Smith et al., 2007). The OBO contains information about ontologies and projects that are carried out in the field of biology (http://obofoundry.github.io/).

Currently, OBO describes over 70 ontologies in various fields, including anatomy, biochemistry, biological processes, functions, and sequences, diseases, environment, experimental evidence, phenotypes, proteins, and taxonomy (Schober et al., 2009).

To ensure the compatibility of biomedical ontologies developed within the OBO project, there are recommendations on standardization and the used ontological relationships. The formal properties of relationships are set, which can be used for inference of new assertions. In particular, it is assumed that relationships *part_of* and *is_a* are transitive, reflexive, and antisymmetric.

However, in fact, depending on the further clarification of the semantics of relationships and the specifics of their application, the properties of these relationships may not be met. Even such popular relationships as *part_of* and *is_a* have different practical interpretations. In this case, the transitive property may be violated. This was the problem for developers of the project Gene Ontology (GO), when they started formal verification of its ontology (Smith et al., 2003).

Below are some of the problems that have arisen in GO during the interpretation of the relationship *part_of*:

P1. *A part_of B* means *A* is sometimes part of *B*, that is, for every *A* at a certain time *t A* is part of *B*.

Example: "replication fork" part_of "nucleoplasm" ("replication fork" is observed in a particular phase of the cell cycle).

P2. *A part_of B* means *A* may be part of *B*. Class *A* is part of class *B* if and only if there is subclass $C \subset B$,

Types of *part_of* relationships and their defining properties (by: Winston et al., 1987)

Types of relationship *part_of*	Properties		
	"Functional"	"Homeomerous"	"Separable"
Component/Integral-Object	+	–	+
Member/Collection	–	–	+
Portion/Mass	–	+	+
Stuff/Object	–	–	–
Feature/Activity	+	–	–
Place/Area	–	+	–

where all instances of *A* are included as part of the instances of *C* and all instances of class *C* include the instances of class *A*.

Example: "flagellum" part_of "cell" (some cell types include flagella as their part).

P3. *A part_of B* means: *A* is always part *B*.

Example: "membrane" *part_of* "cell" (membrane is a part of every cell).

Similarly, GO faced problems of interpreting relationship *is_a*. A well-structured classification can be obtained by replacing relationship *is_a* for specific types, for example: *has_role*, *is_dependent_on*, *is_involved_in*, *contributes_to*, and *is_located_in*, as well as adding different categories of entities: *sites*, *constituents*, *roles*, *functions*, and *qualities*.

In general, to solve these problems one needs to clarify the semantics of these relationships, which is used in the development of a specific ontology.

There are various attempts to clarify the semantics of relationships *part of* and *is_a* and their classification for the resolution of typical conflicts and violations of the properties of these relationships. The paper of Winston et al. (1987) introduced six different types of relationships *part_of* (table) based on the following criteria or properties:

• "Functional" is met when parts in specific spatial or temporal localization perform the same functional role as a whole.

• "Homeomerous" is met when every part is similar to every other part and to the whole to which it belongs.

• "Separable" is met when parts are not physically connected and at least in principle are separated from the whole which they constitute.

Problems with the transitivity of relationship *part_of* arise when different types of the relationship *part_of* are combined. In general, the transitivity should be assumed at least when using relationships *part_of* of the same type, that is, when they have the same properties.

Depending on the characteristics of the domain, one can use different sets of properties that define classes of the relationship *part_of*, for example, configurational, encapsulated, exchangeable, functional, homeomerous, homogeneous, mandatory, canonically necessary, removable, segmental, separable, and shareable.

Bioinformatics and systems biology traditionally widely use ontology representation in OBO. In recent years, many ontologies are translated into OWL (Ontology Web Language) (Stevens et al., 2007). The main problem with these transformations consists in errors, contradictions, and irregularities of the interpretation of relationships. The use of formal methods to search for inconsistencies and incompleteness can significantly improve the quality of the descriptions.

DEVELOPMENT OF ONTOLOGIES IN BIOINFORMATICS

Currently, hundreds of ontologies in the field of biology have been developed, which can be used for the description and integration of knowledge, as well as inference of new knowledge.

In particular, bioinformatics resources and ontologies have been developed and widely used, making it possible to describe molecular structures, functions, processes, and gene networks (GO).

Ontologies MIAPE (Minimum Information About a Proteomics Experiment) (Taylor et al., 2007) and MIMIx (Minimum Information required for reporting a Molecular Interaction eXperiment) (Orchard et al., 2007) were proposed by the working group Human Proteome Organization (HUPO) to describe proteomic studies and experiments on molecular interactions, respectively.

Small upper-level ontology BFO (Basic Formal Ontology) is designed to develop ontologies oriented for the search and integration of scientific data. BFO has already been used for the development of more than 130 ontologies in different domains (http://ifomis.uni-saarland.de/bfo/).

The knowledge base ChEBI (Chemical Entities of Biological Interest) includes the ontology of molecular objects—natural compounds or synthetic products that affect processes in living organisms, including any constitutionally or isotopically distinct atoms, molecules,

ions, ion pairs, radicals, radical ions, complexes, and conformers (https://www.ebi.ac.uk/chebi/). ChEBI currently (release 131) includes 46 477 fully annotated molecular objects.

The ontology of cell types CL (Cell Type Ontology), in fact, is a structured controlled vocabulary that includes the description of cell types of different species of organisms, from prokaryotes to mammals (http://www.obofoundry.org/cgi-bin/detail.cgi?id=cell).

Ontologies developed in the KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al., 2004) are focused on the wide range of molecular biology, from genes and proteins to metabolic and gene networks. Knowledge base TAMBIS (Transparent Access to Multiple Bioinformatics Information Source, http://www.cs.man.ac.uk/~stevensr/tambis/) provides users-biologists with a single point of access to global sources of biological information, which in this system is integrated through an ontological description (Stevens et al., 2000). EcoCyc is a scientific database that accumulates information obtained as a result of annotating the scientific publications on the genome of *E. coli*, regulating its gene transcription, and transport and metabolic pathways (Karp et al., 2014).

Sequence Ontology (SO) includes a number of concepts and controlled vocabularies used to describe the properties and initial annotation of nucleotide or protein sequences, the structural representation of these annotations in genomic databases, the mutations in both types of sequences, and at a higher level (Cunningham et al., 2015).

The Mouse Atlas Project develops a digital atlas and a database on mouse gene expression and cell lines with the description of the anatomical localization of cells (Graham et al., 2015).

Ontology MGED (Microarray and Gene Expression Data) is used to describe experiments and gene expression data (Whetzel et al., 2006).

MIAME (Minimum Information About a Microarray Experiment) is used to describe the expression data (Brazma et al., 2001).

The knowledge base PharmGKB (https://www.pharmgkb.org/) provides information on pharmacogenetics (Klein et al., 2001).

The project Cell Cycle Ontology (Antezana et al., 2009) aims at the extension of the existing ontologies associated with the cell cycle for the integration and knowledge management of its components and regulatory aspects. The sources of this knowledge include already existing resources (GO, UniProt, IntAct, BIND, NCBI taxonomy, etc.). The integration and combination of this knowledge makes it possible to provide the most complete picture of cell division processes.

The anatomical and morphological ontologies for model organisms are important examples of ontological modeling, which uses a wide variety of spatial and temporal relationships.

In particular, the basic anatomical model FMA (Foundational Model of Anatomy ontology) is a reference ontology which includes concepts and relationships used to describe the structural organization of the human body at various levels, from macromolecules, cells, tissues, and organs to the body, taking into account ontogeny, and designed for the symbolic computer modeling of anatomical structures (Rosse et al., 2003; Rosse et al., 2007).

FMA ontology knowledge is represented as frames and is stored in a relational database. FMA (http://sig.biostr.washington.edu/projects/fm/) includes about 75000 anatomy classes, more than 130000 unique terms, more than 205000 frames, and 174 unique slots, which are used to represent different types of relationships, attributes, and attribute relationships. The FMA relationships network contains more than 2.5 million instances of relationships, more than 1 million instances of classes, and about 450000 relationships between classes.

GENE ONTOLOGY

GO (http://www.geneontology.org/) is an example of one of the most successful projects of ontology creation.

The structure of GO includes three sections:

• Molecular function, an elementary activity/task or a role performed by a gene or a gene product in certain biological processes, for example, "catalytic activity" or "Toll receptor binding."

• Biological processes describe a series of events that implement one or more organized ensembles of molecular functions. Unlike the function, the process must have several distinct stages. For example, "pyrimidine metabolic process."

• Cellular components as part of the anatomical structure which describes the localization of a gene or its product in the body at levels of cell structures and macromolecular complexes (for example, "nucleus," "membrane") or groups of gene products (for example, "ribosome," "proteasome," or "protein dimer").

In fact, GO makes it possible to describe the knowledge about what is the function of the gene or its product (RNA, protein) in a particular biological process or a particular cellular structure.

GO contains more than 40000 concepts (The Gene Ontology Consortium, 2015), including the following concepts:

• Biological process, about 30000;

• Molecular function, more than 10000;

• Cellular component, 3758.

GO was used to develop the resource GOA (Gene Ontology Annotation, http://www.ebi.ac.uk/GOA), which is used for the annotation of proteins from UniProtKB (UniProt Knowledgebase). Currently GOA

contains 368 million GO annotations for nearly 54 million proteins from 480000 taxonomic groups (Huntley et al., 2015).

The basic relationships between the concepts that are used in GO are *is_a*, *part_of*, and *regulates*.

• *is_a* is a simple relationship"class–subclass," where *A* is_a *B* means that *A* is a subclass of *B*.

• *part_of* is a relationship"part–whole." Expression A *part_of B* means that if *A* exists, *A* is always part *B*.

• *Regulates*, *positively_regulates*, and *negatively_regulates* describe relationships between biological processes, molecular functions, or biological properties.

GO describes hierarchical relationships, however, the relationship graph is not a tree. One concept may have several ancestors. The transitivity property of relationships used in GO makes it possible to build a lattice of relationships between concepts and carry out logical inference about the properties of concepts and their relationships (Srinivas, 2009).

ONTOLOGY OF SYSTEMS BIOLOGY

Computer models have become a central scientific paradigm of systems biology and a main tool for theoretical research and understanding of mechanisms of the functioning of complex living systems. The increase in the number and size of these models leads to a greater need for collaborative development, reuse of models, their verification, and description of the computational experiment and its results.

Knowledge representation formats for the mathematical modeling of biological systems are being developed through the active use of the ontological modeling of the domains. In this sense, ontology associated with the entire set of formats that support research in systems biology, in particular computer modeling of biological systems and processes, can be considered as a first approximation to the ontology of systems biology.

In this section, we discuss ontologies of systems biology aimed at the description of genetic systems and their models.

Many ontologies are not developed from scratch but integrate (or compile) previously developed partial ontologies used as basic sources of knowledge. These include the formalization of the structure of mathematical models (SBML), standardized description of model components from the kinetic and biological points of view (SBO, GO, and UniProt). However, just the formalization of the structure of the computer model is not sufficient to support the entire process chain of the computing experiment and computer simulation in systems biology. It is necessary to formalize pragmatic and dynamic aspects of the modeling process.

Systems Biology Ontology (SBO) (http://www.ebi.ac.uk/sbo/) is an example of a specialized project. This project is aimed at the development of controlled vocabularies and ontologies focused on solving problems of systems biology, especially in the context of computer modeling.

SBO includes six orthogonal controlled vocabularies, including a description of the roles of the reaction participants (for example, "reactant," "product," or "modifier"), the values of quantitative parameters of the reactions' models (for example, "kinetic constant"), exact classification of the mathematical expressions that describe the system (for example, "mass action rate law"), the type of the used modeling environment (for example, "logical framework" or "discrete framework"), types of system entities (for example, "macromolecule," "enzyme," or "ligand"), and interactions therein (for example, "process," "biochemical reaction," "genetic interaction," or "relationship"). The quantitative parameters and mathematical expressions are described in the MathML 3.0 language (http://www.w3.org/TR/MathML3/).

The mathematical model can be annotated in SBO at any stage of the life cycle (from creation to the time of expansion and modification of the model) through the successive expansion of its semantics.

There are examples of other knowledge representation formats and their related systems biology ontologies:

• Systems Biology Markup Language (SBML, http://sbml.org), a format for the representation of the structure of biological models (Hucka et al., 2003).

• Biological Pathway Exchange Language (BioPAX, http://www.biopax.org), a format for the description and integration of information about molecular interactions and biological processes (Demir et al., 2010). This approach is used to represent knowledge in the existing databases (BioCyc, BIND, WIT, aMAZE, KEGG, Reactome, and so on) that are important for the description of gene expression mechanisms.

• Minimal Information Requested In the Annotation of biochemical Models (MIRIAM) (Le Novère et al., 2005), a format for the standardization of a minimal set of information required to annotate the model and allow collective annotation, curation, and development, as well as the reuse of models.

• Simulation Experiment Description Markup Language (SED-ML, http://sed-ml.org/), a format for the description of experiments on the simulation and exchange of modeling results regardless of the used language of model specification and simulation environment (Waltemath et al., 2011).

• TErminology for the Description of DYnamics (TEDDY, http://www.ebi.ac.uk/compneursrv/teddy) (Chelliah and Endler, 2009), ontology for the description of the dynamic behavior of a biological system or a dynamic phenomenon, control of elements of a biological model and a system in systematic and synthetic biology. In particular, TEDDY makes it possible to quantitatively describe characteristics of the biological

system or model: oscillation type (chaotic, periodic, quasi-periodic, etc.), oscillation fields, periods, points of stationarity or instability, parameter-dependent bifurcations, functional motifs (for example, negative feedback), etc.

• Kinetic Simulation Algorithm Ontology (KiSAO, http://biomodels.net/kisao/), ontology for the description of algorithms for modeling kinetic processes (Chelliah and Endler, 2009). The classification of algorithms for modeling biological models in KiSAO is built using a variety of categories and taking into account algorithm versions. For example, deterministic or stochastic rules, spatial and non-spatial approaches, discrete or continuous variables, and fixed or adaptive time steps.

• Systems Biology Graphical Notation (SBGN, http://www.sbgn.org), a format for the graphical representation of biological systems and processes (Dada et al., 2010).

• SBRML (www.comp-sys-bio.org/tiki-index.php?page=SBRML), a format for saving simulation results (Orchard et al., 2007).

• CellML (www.cellml.org), a format for the description of mathematical models of biological systems and processes (Lloyd et al., 2004). Mathematical expressions are represented in CellML through the use of the language MathML.

Reference ontologies, GO, SO, Chemical Entities of Biological Interest (ChEBI), FMA, FMP, CPRO, PaTO, Pro, RnaO, and CARO, which describe biological systems at different levels, can be integrated into a common ontology to describe the object studied by systems biology (organism, organ, tissue, or cell).

USE OF ONTOLOGIES IN SYSTEMS BIOLOGY AND BIOINFORMATICS

In general, the use of ontologies gives a tangible effect when solving the following problems in bioinformatics and systems biology (Bodenreider and Stevens, 2006; Bodenreider, 2008; Beck et al., 2009; Noy et al., 2009; Podkolodnyy, 2011).

1. Interpretation of molecular genetic knowledge, semantic interpretation of data analysis methods and models in systems biology. In particular, gene enrichment analysis using terms of GO (GO Enrichment Analysis) is used to interpret the data (for example, the functional description of the set of genes), quality control, as well as ordering and selection of the data.

2. Prioritization of genes, proteins, biomarkers, etc.

3. Similarity analysis and clustering of objects. For example, the analysis of expression levels of tens of thousands of genes in different cell situations, different conditions, and at different stages of development of a cell, tissue, organ, or body. After detecting a group of genes with similar expression patterns (coexpressed genes), it becomes necessary to describe these groups. The use of GO makes it possible to describe functions of the genes comprising the cluster (Khatri and Draghici, 2005). In fact, using ontologies, one can quantify the semantic similarity of the domain objects.

4. Support of interoperability and knowledge exchange:

—unified access to multiple heterogeneous data sources;

—the search for relevant information in documents. In this case, ontology sets the context for the annotation of the document content using semantic information, as well as indexing and binding the facts described in the databases (Shah et al., 2009);

—integration of information from different sources and the creation of large knowledge bases;

—combining the experimental data and knowledge from the ontology to form a knowledge base;

—interoperability, support of communication (between people and organizations), and knowledge exchange (among people and/or systems);

—text mining and semantic analysis (Chapman and Cohen, 2009);

—knowledge acquisition, extraction of knowledge, as well as implicit and explicit relationships between entities in annotated sources, and analytics.

5. Creating new ontologies based on the reuse of basic canonical ontologies and various types of operations with them, including ontology matching, ontology merging, ontology mapping, and ontology alignment.

6. Ensuring the consistency and correctness of knowledge representation. Support for the process of ontology construction, including all types of automatic output to find errors and identify new relationships. In modern ontologies, there are hundreds of thousands of concepts and relationships, so manual verification is not possible. In this case, an expert checks the contradictions and the results obtained by formal inference in the ontologies (Livingston et al., 2015).

7. Support of inductive inference to extract additional knowledge from a variety of facts and hypothesis testing. For example, the work of Podkolodnyy et al. (2012) presents approaches to the ontological modeling of the regulation mechanisms of gene transcription and shows examples of the reconstruction of the hypothetical mechanisms of the transcriptional regulation, taking into account information about the structure of the regulatory regions of genes and functions of the regulatory proteins present in the given cells or tissues at a certain stage of development.

8. Better argumentation of bioinformatics techniques, including a precise description of the biomedical experimental protocols, data analysis methods, and modeling of biological processes and systems (Chen et al., 2007).

There are examples of the successful application of the descriptive logic and inference for formal ontologies in the biomedical field. A review of Keith et al. (Keet et al., 2007) provides a list of common scenarios obtained by the analysis and synthesis of examples of

the use of inference means. It was noted that the standard Racer, Pellet, and FaCT++ tools do not make it possible to perform many of the scenarios on real biomedical ontologies because of their large size and complexity. This is why the development of new effective software inference means and approaches for the formalization of biological knowledge remains highly relevant.

SOFTWARE FOR ONTOLOGY APPLICATION

Currently, GO is the most widely used ontology in the field of bioinformatics and systems biology. To this end, a large number of software tools have been developed:

—AMIGO (http://amigo.geneontology.org/amigo), QuickGO (http://www.ebi.ac.uk/QuickGO/), Protein2GO, and Ontology Lookup Service (http://www.ebi.ac.uk/ontology-lookup/)—search and view of GO and annotation data (Carbon et al., 2009).

—OBO-Edit, view and edit of ontological descriptions (Day-Richter et al., 2007).

—Blast2GO (https://www.blast2go.com/), functional annotation of unknown sequences by a homology search using BLAST and the analysis of the GO annotation of the obtained results (Conesa and Gotz, 2008).

—GoPubMed, search for biological texts based on GO and Medical Subject Headings (MeSH, http://www.nlm.nih.gov/mesh/introduction.html). GoPubMed connects GO with the abstracts database PubMed and answers the questions: What? Who? Where? When? (http://www.gopubmed.org/web/gopubmed/).

—Onto-Tools (http://vortex.cs.wayne.edu/), a set of services including Onto-Express, Onto-Compare, Onto-Design, Onto-Translate, Onto-Miner, Pathway-Express, Promoter-Express, nsSNPCounter, TAQ, and OE2GO—profiling of multiple genes, comparison of expression data, computer support of the design of DNA chips, including selecting a set of genes based on their functions, the processes in which these genes are involved, or cellular components where these genes are expressed.

—GOToolBox, analysis of the results of DNA chip experiments (http://genome.crg.es/GOToolBox/) (Martin et al., 2004).

GO is most often used for the analysis of the enrichment of the description of genes with gene ontology terms. To solve this problem, several dozens of software systems have been developed, among which the most popular are (ftp://ftp.geneontology.org/pub/go/www/GO.tools.microarray.shtml):

• Gorilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool) (http://cbl-gorilla.cs.technion.ac.il/), a web-application for the identification of enriched GO terms in the ranked list of genes.

• PANTHER (http://pantherdb.org/), classification of proteins (and their genes) in accordance with information from GOA, evolutionary relationships, interactions, participation in the metabolic and gene networks, etc. As part of the GO project, this system is regularly updated and uses the most current data (Thomas et al., 2003; Huaiyu et al., 2003; Mi et al., 2005).

• DAVID (http://david.abcc.ncifcrf.gov/), a web-service for the annotation and analysis of expression and proteomic data obtained with the help of high-throughput experimental technologies, as well as the enrichment of gene description. The main drawback of DAVID is the use of outdated data of GO (lagging 3–4 years).

• Web Service WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) is used in functional genomics, proteomics, and large-scale genetic research for functional annotation of large groups of genes, for example, groups of differentially expressing and coexpressing genes.

The rapid development of the experimental techniques in molecular biology has led to ontological modeling becoming the basic method in bioinformatics and systems biology for the integration and analysis of heterogeneous experimental data and their use to construct mathematical models of molecular-genetic systems and processes.

Bioinformatics and systems biology is an excellent testing ground for technologies and the efficient use of ontological modeling. The creation of several dozen basic reference ontologies and their verification make it possible to use these ontologies as a source of knowledge for the integration and construction of more complex domain models aimed at solving specific problems of biomedicine. As an example we can mention the knowledge base KaBOB (Livingston et al., 2015), which integrates knowledge from 14 heterogeneous sources—databases of genes and their homologs, proteins, drugs, genetic associations, regulatory sequences, protein-protein interactions, etc. The integration of data and the search for information in this knowledge base are based on the ontology that combines the following knowledge sources:

1. Basic Formal Ontology (BFO) (http://purl.obolibrary.org/obo/bfo.owl);

2. BRENDA Tissue/Enzyme Source (BTO) (http://purl.obolibrary.org/obo/bto.owl);

3. ChEBI (http://purl.obolibrary.org/obo/chebi.owl);

4. Cell Type Ontology (CL) (http://purl.obolibrary.org/obo/cl.owl);

5. Gene Ontology (GO) (http://purl.obolibrary.org/obo/go.owl);

6. Information Artifact Ontology (IAO) (http://purl.obolibrary.org/obo/iao.owl);

7. Protein-Protein Interaction Ontology (MI) (http://purl.obolibrary.org/obo/mi.owl);

8. Mammalian Phenotype Ontology (MP) (http://purl.obolibrary.org/obo/mp.owl);

9. NCBI Taxonomy (http://purl.obolibrary.org/obo/ncbitaxon.owl);

10. Ontology for Biomedical Investigation (OBI) (http://purl.obolibrary.org/obo/obi.owl);

11. Protein Modification (MOD) (http://purl.obolibrary.org/obo/mod.owl);

12. Protein Ontology (PR) (http://purl.obolibrary.org/obo/pr.owl);

13. Relation Ontology (RO) (http://purl.obolibrary.org/obo/ro.owl);

14. Sequence Ontology (SO) (http://purl.obolibrary.org/obo/so.owl).

The knowledge was represented with the OWL language. The description of the ontology KaBOB contains more than 13 million RDF-triples. The SPARQL 1.1 is used as the query language. Two versions of the KaBOB knowledge base are implemented: (i) only for humans and (i) for humans and a number of model organisms. As a result, the system makes it possible to perform queries on the set of heterogeneous data, comparing the knowledge at different levels of the description of biological systems. "What human genes or gene products localized in mitochondria are involved in oxidative phosphorylation and are targets of drugs? What kind of drugs?" is an example of a query. The KaBOB system architecture and applied information technologies make it possible to expand the used ontologies and databases and in the future to solve complex bioinformatics tasks.

Thus, further formalization and accumulation of ontological knowledge, as well as the use of formal methods of their analysis can take the entire cycle of research in the field of systems biology to a new technological level.

ACKNOWLEDGMENTS

The work was supported by the Russian Science Foundation (grant no. 14-2400123).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Antezana, E., Egaña, M., Blondé, W., Illarramendi, A., Bilbao, I., De Baets, B., Stevens, R., Mironov, V., and Kuiper, M., The cell cycle ontology: An application ontology for the representation and integrated analysis of the cell cycle process, *Genome Biol.*, 2009, vol. 10, no. 5, p. R58. doi 10.1186/gb-2009-10-5-r58

Bada, M. and Hunter, L., Enrichment of OBO ontologies, *J. Biomed. Inf.*, 2007, vol. 40, pp. 300–315.

Beck, T., Morgan, H., Blake, A., Wells, S., Hancock, J.M., and Mallon, A.-M., Practical application of ontologies to annotate and analyze large-scale raw mouse phenotype data, *BMC Bioinf.*, 2009, vol. 10, no. 5, p. S2. doi 10.1186/1471-2105-10-S5-S2

Bodenreider, O. and Stevens, R., Bio-ontologies: Current trends and future directions, *Brief Bioinf.*, 2006, vol. 7, no. 3, pp. 256–274.

Bodenreider, O., Biomedical ontologies in action: Role in knowledge management, data integration and decision support, in *IMIA Yearbook Medical Informatics*, 2008.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, et al., Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nat. Genet.*, 2001, vol. 29, pp. 365–371.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S., Amigo GO: Online access to ontology and annotation data, *Bioinformatics*, 2009, vol. 25, no. 2, pp. 288–289.

Chandrasekaran, B., Josephson, J.R., and Benjamins, V.R., What are ontologies and why do we need them?, *IEEE Intell. Syst.*, 1999, vol. 14, no. 1, pp. 20–26. doi 10.1109/5254.747902

Chapman, W.W. and Cohen, K.B., Current issues in biomedical text mining and natural language processing, *J. Biomed. Inf.*, 2009, vol. 42, pp. 757–759. doi 10.1016/j.jbi.2009.09.001

Chelliah, V., Endler, L., Juty, N., Laibe, C., Li, C., Rodriguez, N., and Le Novère, N., Data integration and semantic enrichment of systems biology models and simulations. Data integration in the life sciences, *Lect. Notes Comput. Sci.*, 2009, vol. 5647, pp. 5–15.

Chen, Q., Chen, Y.-P.P., and Zhang, C., Detecting inconsistency in biological molecular databases using ontologies, *Data Min. Knowl. Disc*, 2007, vol. 15, pp. 275–296. doi 10.1007/s10618-007-0071-0

Conesa, A. and Gotz, S., Blast2GO: A comprehensive suite for functional analysis in plant genomics, *Int. J. Plant Genomics*, 2008, vol. 2008, p. 619832. doi 10.1155/2008/619832

Cunningham, F., Moore, B., Ruiz-Schultz, N., Ritchie, G.R., and Eilbeck, K., Improving the sequence ontology terminology for genomic variant annotation, *J. Biomed. Semantics*, 2015, vol. 6, p. 32. doi 10.1186/s13326-015-0030-4

Dada, J.O., Spasić, I., Paton, N.W., and Mendes, P., SBRML: A markup language for associating systems biology data with models, *Bioinformatics*, 2010, vol. 26, no. 7, pp. 932–938. doi 10.1093/bioinformatics/btq069

Day-Richter, J., Harris, M.A., Haendel, M., Gene ontology OBO-edit working group, and Lewis, S., OBO-Edit – an ontology editor for biologists, *Bioinformatics*, 2007, vol. 23, no. 16, pp. 2198–2200.

Demir, E., Cary, M.P., Paley, S., et al., The BioPAX community standard for pathway data sharing, *Nat. Biotechnol.*, 2010, vol. 28, pp. 935–942. doi 10.1038/nbt.1666

Gene Ontology Consortium, Gene Ontology Consortium: Going forward, *Nucl. Acids Res.*, 2015, vol. 43, pp. D1049–D1056. doi 10.1093/nar/gku1179

Graham, E., Moss, J., Burton, N., Armit, C., Richardson, L., and Baldock, R.D., The atlas of mouse development eHistology resource, *Development*, 2015, vol. 142, pp. 1909–1911. doi 10.1242/dev.124917

Gruber, T.R., Toward principles for the design of ontologies used for knowledge sharing, *Int. J. Human-Comput. Stud.*, 1995, vol. 43, nos. 5/6, pp. 907–928.

Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J., and O'Donovan, C., The GOA database: Gene ontology annotation updates for 2015, *Nucleic Acids Res.*, 2015, vol. 43, pp. D1057–D1063. doi doi 10.1093/nar/gku1113

Huaiyu, M.I., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H., and Thomas, P.D., The PANTHER database of protein families, subfamilies, functions and pathways, *Nucleic Acids Res.*, 2005, vol. 33, no. 1, pp. D284–D288. doi 10.1093/nar/gki078

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., SBML Forum. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models, *Bioinformatics*, 2003, vol. 19, no. 4, pp. 524–31.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M., The KEGG resource for deciphering the genome, *Nucleic Acids Res.*, 2004, vol. 32, pp. D277—D280.

Karp, P.D., Weaver, D., Paley, S., Fulcher, C., Kubo, A., Kothari, A., Krummenacker, M., Subhraveti, P., Weerasinghe, D., Gama-Castro, S., Huerta, A.M., Muñiz-Rascado, L., Bonavides-Martinez, C., Weiss, V., Peralta-Gil, M., et al., The EcoCyc database, *Ecosal Plus*, 2014, vol. 2014. doi doi 10.1128/ecosalplus

Keet, C.M., Roos, M., and Marshall, M.S., A survey of requirements for automated reasoning services for bio-ontologies in owl, *Workshop on OWL: Experiences and Directions*, Innsbruck, Austria, 2007.

Khatri, P. and Draghici, S., Ontological analysis of gene expression data: Current tools, limitations, and open problems, *Bioinformatics*, 2005, vol. 21, no. 18, pp. 3587–3595.

Kitano, H., Systems biology: A brief overview, *Science*, 2002, vol. 295, pp. 1662–1664.

Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E., Rubin, D.L., Shafa, F., Stuart, J.M., and Altman, R.B., Integrating genotype and phenotype information: An overview of the PharmGKB project, *Pharmacogenomics J.*, 2001, vol. 1, pp. 167–170.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J.L., Spence, H.D., and Wanner, B.L., Minimum information requested in the annotation of biochemical models (MIRIAM), *Nat. Biotechnol.*, 2005, vol. 23, no. 12, pp. 1509–1515.

Livingston, K.M., Bada, M., Baumgartner, W.A., Jr., and Hunter, L.E., KaBOB: Ontology-based semantic integration of biomedical databases, *BMC Bioinf.*, 2015, vol. 16, p. 126. doi 10.1186/s12859-015-0559-3

Lloyd, C.M., Halstead, M.D.B., and Nielsen, P.F., CellML: Its future, present and past, *Prog. Biophys. Mol. Biol.*, 2004, vol. 85, nos. 2/3, pp. 433–450.

Martin, D., Brun, C., Remy, E., Mouren, P., Thieffry, D., and Jacq, B., GOToolBox: Functional analysis of gene datasets based on Gene Ontology, *Genome Biol.*, 2004, vol. 5, no. 12, p. R101.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H., and Thomas, P.D., The PANTHER database of protein families, subfamilies, functions and pathways, *Nucleic Acids Res.*, 2005, vol. 1, no. 33, pp. D284–288.

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., and Musen, M.A., Bioportal: Ontologies and integrated data resources at the click of a mouse, *Nucleic Acids Res.*, 2009, vol. 37, pp. W170–W173. doi 10.1093/nar/gkp440

Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stmpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., Salama, J.J., Moore, S., Wojcik, J., Bader, G.D., Vidal, M., et al., The minimum information required for reporting a molecular interaction experiment (MIMIx), *Nat. Biotechnol.*, 2007, vol. 25, no. 8, pp. 894–898.

Podkolodnyi, N.L., Ontological modeling in bioinformatics and systems biology, in *Ontologicheskoe modelirovanie* (Ontological Modeling), IPI RAN, 2011, pp. 233–269.

Podkolodnyi, N.L., Ignat'eva, E.V., Podkolodnaya, O.A., and Kolchanov, N.A., Information support of research of transcriptional regulation mechanisms: Ontological approach, *Vavilovskii Zh. Genet. Sel.*, 2012, vol. 16, no. 4/1, pp. 742–755.

Rosse, C. and Mejino, J.L., A reference ontology for biomedical informatics: The foundational model of anatomy, *J. Biomed. Inf.*, 2003, vol. 36, no. 6, pp. 478–500.

Rosse, C. and Mejino, J.L.V., *The Foundational Model of Anatomy Ontology. Anatomy Ontologies for Bioinformatics: Principles and Practice, Burger, A.*, Davidson, D., and Baldock, R., Eds., New York: Springer, 2007.

Schober, D., Smith, B., Lewis, S.E., Kusnierczyk, W., Lomax, J., Mungall, C., Taylor, C.F., Rocca-Serra, P., and Sansone, S.A., Survey-based naming conventions for use in OBO foundry ontology development, *BMC Bioinf.*, 2009, vol. 27, no. 10, p. 125. doi 10.1186/1471-2105-10-125

Shah, N.H., Jonquet, C., Chiang, A.P., Butte, A.J., Chen, R., and Musen, M.A., Ontology-driven indexing of public datasets for translational bioinformatics, *BMC Bioinf.*, 2009, vol. 10, no. 2, p. S1. doi 10.1186/ 1471-2105-10-S2-S1

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Gold¬berg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., et al., The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration, *Nat. Biotech.*, 2007, vol. 25, no. 11, pp. 1251–1255.

Smith, B., Williams, J., and Schulze-Kremer, S., The ontology of the gene ontology, *AMIA Annual Symp. Proceedings*, 2003.

Srinivas, K., OWL reasoning in the real world: Searching for godot, in *Proc. of the 22nd Intern. Workshop on Description Logics (DL 2009)*, Oxford, UK, 2009.

Stevens, R., Aranguren, M.E., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., and Rector, A., Using OWL to model biological knowledge, *Int. J. Hum.-Comput. Stud.*, 2007, vol. 65, no. 7, pp. 583–594. http://portal.acm.org/citation.cfm?id=1247774.

Stevens, R., Baker, P., Bechhofer, S., Ng, G., Jacoby, A., Paton, N.W., Goble, C.A., and Brass, A., TAMBIS: Transparent access to multiple bioinformatics information sources, *Bioinformatics*, 2000, vol. 16, no. 2, pp. 184–185.

Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.A., Julian, R.K.Jr., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., Dunn, M.J., Heck, A.J., Leitner, A., Macht, M., Mann, M., et al., The minimum information about a proteomics experiment (MIAPE), *Nat. Biotechnol.*, 2007, vol. 25, no. 8, pp. 887–893.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A., PANTHER: A library of protein families and subfamilies indexed by function, *Genome Res.*, 2003, vol. 13, no. 9, pp. 2129–2141.

Waltemath, D., Adams, R., Bergmann, F.T., Hucka, M., Kolpakov, F., Miller, A.K., Moraru, I.I., Nickerson, D., Snoep, J.L., and Le Noère, N., Reproducible computational biology experiments with SED-ML – the simulation experiment description markup language, *BMC Syst. Biol.*, 2011, vol. 5, p. 198.

Whetzel, P.L., Parkinson, H., Causton, H.C., Fan, L., Fostel, J., Fragoso, G., Game, L., Heiskanen, M., Morrison, N., Rocca-Serra, P., Sansone, S.A., Taylor, C., White, J., and Stoeckert, C.J., The MGED ontology: A resource for semantics-based description of microarray experiments, *Bioinformatics*, 2006, vol. 22, no. 7, pp. 866–873.

Winston, M.E., Chaffin, R., and Herrman, D., A taxonomy of part-whole relations, *Cognit. Sci.*, 1987, vol. 11, pp. 417–444.

Translated by K. Lazarev