Metamodeling Integration Architecture for Open Biomedical Ontologies: *The GO extensions' Case Study*

Marie-Noelle Terrasse[1], Marinette Savonnet[1], Eric Leclercq[1], George Becker[2], Eric Fourmentin[3], Damien Lariviere[3], Pierre Grenon[4], and Magali Roux-Rouquie[5]

(1) LE2I UMR CNRS 5158 - University of Burgundy - France

firstname.lastname@u-bourgogne.fr

(2) gbecker@nerim.net

(3) Scientific Foundation Fourmentin-Guilbert, 2 avenue du Pavé Neuf, 93160 Noisy-Le-Grand, France

{Eric.Fourmentin,Damien.Lariviere,fondation}@fourmentinguilbert.org

(4) IFOMIS, University of Saarland - Saarbrücken, Germany

pierre.grenon@ifomis.uni-saarland.de

(5) LIP6 CNRS-UPMC - 104, avenue du Pt Kennedy - 75016 Paris, France

magali.roux@lip6.fr

Abstract

New technologies used in biology are generating huge quantities of data; up to two petabytes of overall data are to be expected by the end of the decade. Modern biology also has to deal with multitude of information representations that are driven by the continuous evolution of biological knowledge. Conceptual modeling allows layered descriptions whose abstractions are rather stable while their concrete descriptions may evolve: formal ontologies versus domain ontologies, metamodels versus models, etc.

In this paper we introduce an incremental integration methodology that is based on the Open Biomedical Ontology (OBO) umbrella. The proof of concept for our methodology is carried out using the Gene Ontology's (GO) case study in which GO extensions are viewed as manifestations of the domain knowledge evolution. Based on this conceptual foundation, we present a metamodeling architecture that encompasses metamodels and models of GO, as well as those of GO extensions. Our architecture can be used for ensuring semantical quality of biological knowledge integration.

Keywords: conceptual modeling, metamodeling architecture, ontology, integrative biology.

1 Introduction

Recent technological developments allow high-dimensional biology, generally known as omics sciences (genomics, transcriptomics, proteomics, metabolomics, etc.). Such technological developments concern the systematic characterization of temporal and conditional changes at molecule, cell and organism levels. One of the main challenges when striving to take full advantage of such new experimental technologies consists in developing computational tools that allow integration of automatically produced biological data into a consistent organized knowledge representation.

1.1 State of the art

Possible approaches to integration (e.g., ontologies, federation, data warehouses, peer data management systems) correspond to fairly stable technologies. Such technologies are widely used in several domains in which the data volume and the need for sharing existing data explode, such as in e-business or web services.

Adaptation of these approaches to the management of information systems in biology were discussed in [8, 17]. Data warehouses that present serious updating challenges were nevertheless found more suited for datasets focused on narrow areas of research [31]; by limiting application domains it is possible to design precise data representations that can be then regularly updated by merely reflecting source schema changes. Database federations were found more suitable when the most recent specific information was requested [15], while mediated schemas that support modularity were more efficient in the case of integration of complex data [9]. New approaches (e.g., peer data management [27], collaborative construction of ontologies [2] have recently been applied in biology in order to cope with the challenge of building comprehensive knowledge of high-dimensional biology.

Efforts to organize vast domains of the biological knowledge have started with the design of Gene Ontology (GO) [18] and controlled vocabularies [3]. OBO was created to gather all of the ontologies of the biomedical disciplines under a common umbrella. At this time, over 900 databases [1] and over 60 ontology participants of the OBO group keep reporting on day-by-day advances in molecular biology. In addition to this huge data volume, high-dimensional biology has to face, on the one hand, the practical challenges of evolving concepts, and on the other hand, those of evolving technologies. In addition to such problems, extra complexity is introduced when different biological granularity levels are treated with the same technology, and, conversely, when different methodologies

are used to analyze a single biological object, thus introducing multiple representations of that object.

The conceptual modeling widely uses description methods based on ontologies and metamodeling architectures. They both introduce descriptions that rely on concepts and relations. Ontologies emphasize the generalization/specialization relation. Each ontology has a predefined set of relations for modeling concept dependencies [24]. Spear [26] defines two dimensions (i.e., relevance and granularity) "*along which the appropriate contents of a domain specific ontology need to determined*". Ontologies provide modelers with description languages having wide-spectrum granularities (thus allowing to describe both general and specific features within a single ontology). Metamodeling architectures neither limit the set of usable relations nor emphasize the use of a particular type of relation. Metamodeling provides description languages with wide-spectrum relevance (allowing to describe large groups of inter-related elements). By combining ontologies and metamodeling architectures, modelers can benefit from both of them: consensual descriptions of narrow-focused domains are devoted to ontologies while comprehensive descriptions of wide-scope domains are provided by metamodeling architectures. Furthemore, models, metamodels and transformations [23] offer a theoretical framework for data integration and can accommodate unforeseen evolution of concepts and methodologies. We aim at defining a metamodeling architecture for integration of biological data based on domain ontologies and on standards that have been developed for various technological spaces [4]. In this paper we present a case study based on the Gene Ontology and some of its extensions that are proposed in the domain litterature.

1.2 Overview of our approach

Within integration architectures, metamodels define languages that are then used to describe different applications. Analogous to the linguistic similarity of Romance languages which makes translating from French to Spanish easier than translating from French to German, semantical similarities of metamodels facilitate accurate information exchanges between applications. We have proposed in [28] to use metamodeling integration architectures to precisely determine which parts of the knowledge structure are shared by cooperating information systems.

For the biological domain, we define such integration architectures as follows. Upper-level metamodels, which include major types of OBO relationships, offer a general relational perpective on the biomedical domain. Metamodels derived from such upper-level metamodels are subjected to various constraints that diminish the number of usable relationships. Knowing precisely which relationships are explicitly stated in biological applications may be of a major interest, particularly in the case when applications compliant with the *close world assumption* have to cooperate with non-compliant applications.

Our proposal is based on the three types of relationships that were introduced by OBO and BFO [10]:

Intra-ontological relationships are limited to a single ontology. For example, the fundamental relationships (*is_a* and *part_of*), spatial relationships (*located_at*, *contained_in*, and *adjacent_to*), temporal relationships (*transformation_of*, *derives_from*, and *preceded_by*), participation relationships (*has_participant*, and *has_agent*) are all intra-ontological relationships.

Trans-ontological relationships relate two terms from different ontologies that have the same definition of relevance (i.e., they describe the same domain). For example, SNAP-SNAP relationships express either qualitative changes (e.g., change in qualitative creation, or qualitative destruction), spatial changes, or substantial changes. They are trans-ontological relationships. Similarly, SNAP-SPAN relationships (e.g., participation relationships) also belong to trans-ontological relationships.

Meta-ontological relationships relate two terms from ontologies with different definitions of relevance.

In order to construct the core of our integration architecture, we first define an upper-level metamodel that must allow description of any application in the biomedical domain. We base this upper-level metamodel on OBO's set of relations. Such a metamodel thus includes all intra-, trans-, and meta- OBO's relationships (denoted by suffixes *m*, *t*, and *i*, respectively, in metamodel names). This metamodel is denoted in Figure 1 by *OBOmti*, and it constitutes the root of an inheritance hierarchy of metamodels. We also define two intermediate metamodels. The metamodel *OBOti*, derived from *OBOmti* by *specialization link 1* in Figure 1, is defined by forbiding use of all meta-relationships. The metamodel *OBOi*, derived from *OBOti* by *specialization link 2* is defined by forbiding use of meta- and trans- relationships. All derived metamodels are thus defined by forbiding relationships that are not used in specific families of descriptions.

2 Description, analysis, and metamodeling of GO and its extensions

We provide a short presentation of GO, as well as of some of the GO extensions that will be used in our metamodeling architecture. From the above core OBO-based architecture we derive a specific metamodel for GO and each of its extensions. This way we obtain the GO case study architecture, as described in Figure 1.

2.1 Gene Ontology

The Gene Ontology [18] was launched in 1998 to elicit a consensual language for describing gene products and functions. GO thus begun the current standardization trends in systems biology [4]. Up to now, over 170.000 genes have been annotated by using the GO terms in the consortium databases [5]. Nevertheless, as the biological domains keep getting more complex, increasingly numerous terms are needed for specifying details of gene function, as well as for future developments of GO towards immunology, transport, signaling, and neurobiology [6].

GO was constructed as a directed acyclic graph (DAG) of three independent ontologies corresponding to *cellular component*, *molecular function* and *biological process*. Each of these three independent ontologies uses *is_a* and *part_of* links. Though the three GO roots can be linked between each other, there is no link between concepts belonging to different sub-ontologies. Each GO concept is defined by its identifier (a unique label, e.g., *GO:0303335*), one or more terms (e.g., *positive regulation of cell migration*, *upregulation of cell migration*), and its location in the DAG (single path or multiple paths to the root of the concept sub-ontology). The rule of single inheritance does not apply to GO descriptions. The "*True Path Rule*" requires that for each GO concept, every path

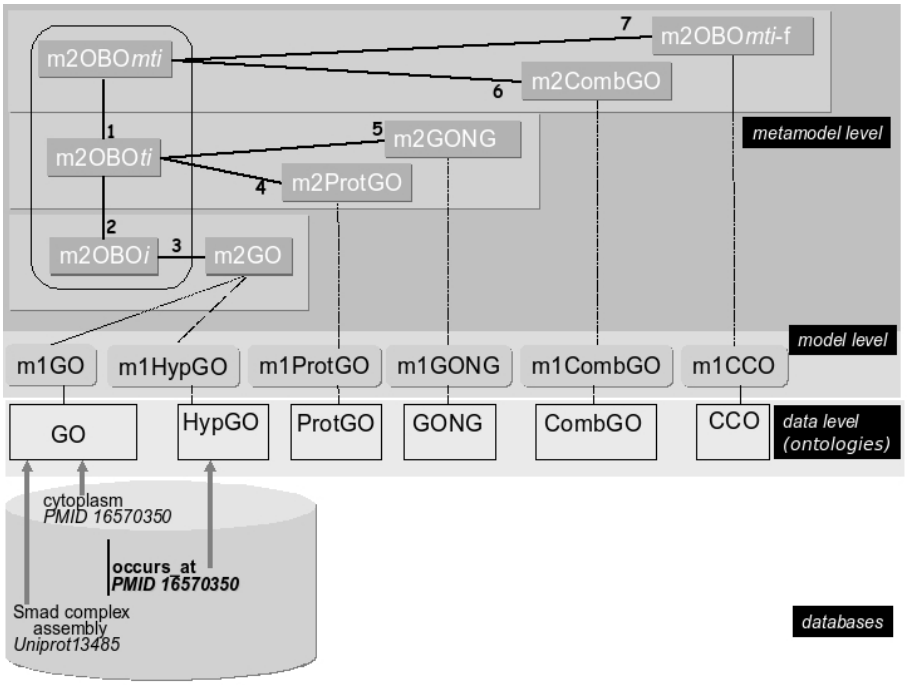

Figure 1: Architecture of metamodels and models. In the upper half, metamodels to ontologies: thick black lines indicate derivation links between metamodels, thin lines indicate instantiation links (between metamodels and models, as well as between models and ontologies). In the lower half, an annotated database: two terms and a relationship between them (denoted by a thick black line), gray arrows indicate annotation references.

from that concept to the corresponding root is *meaningful* (e.g., the example provided on GO's web pages: concept *nuclear chromosome*, identifier *GO:0000228*, path "nuclear chromosome *is_a* chromosome which is *part_of* cytoplasm", and path "nuclear chromosome is *part_of* nucleus which is *part_of* cytoplasm").

Since GO uses only intra-relationships, namely *is_a* and *part_of*. The GO metamodel, $m2GO$, is thus built as a specialization of the $OBOi$ metamodel (*specialization link 3* in Figure 1). The GO model, denoted by $m1GO$ in Figure 1, is derived from the $m2GO$ metamodel.

2.2 GO extensions

The meaning of a GO concept may be expressed in various ways: 1) names associated with the concept, 2) the concept's location in GO DAGs, 3) explicitly specified relationships with other concepts, 4) implicit relationships with other concepts that are revealed by linguistic analysis of names of concepts. Most GO extensions aim at achieving better correspondances between the above approaches to expression of concept semantics. Below, we outline three families of extensions:

Firstly, Lee [16] and Mungall [21] apply linguistics analysis methods to ontological descriptions in order to add new terms and relationships to the biological knowledge. New terms, as generated by Lee, are positioned as sub-terms of the existing GO terms. Mungall's OBOL language uses the *genus* & *differentiae* paradigm for term definitions: the *genus* part of a term achieves broad categorization while the *differentiae* part introduces variations in meaning. OBOL defines thirteen domains (e.g., general, anatomy, enzyme, function) and includes the *has-part* relation while GO uses a single relation *part_of*. We denote by *LA* this family of extensions.

Secondly, many authors [22, 29, 30] point out that implicit information is contained in GO concept names due to the compositional structure of terms. Yeh [30] and Wroe [29] introduce additional relationships between existing concepts. These two authors pre-process the GO ontology in order to improve the description quality (Yeh uses Protégé and Wroe uses DAM+OIL). We denote by *CO* this family of extensions.

Thirdly, Hill [11] and Bada & al. [19] generate new terms by "*creating cross products between orthogonal DAGs*". Concepts corresponding to these new terms are related to concepts of the initial DAGs. We denote by *CP* this family of extensions.

We selected some of the above extensions, representative of the above three families, to form our GO extensions' case study. They are presented in more detail below.

HypGO (Lee's extension [16]) This GO extension belongs to the *LA* family. It uses hyponymy links between names of existing concepts to generate new concepts in sub-ontologies. In principle, any term could be inserted either at the parent concept or at the child concept of the hyponym link. In order to avoid overload of irrelevant terms, rules were introduced in order to define a context-sensitive language of combined words. The benefit of such an extension was to automatically create new terms that could then be validated by consulting the domain literature.

While Lee's is a modification of the ontology since it adds terms, it is not a modification of the structure of the ontology as no new relationships are added. The model of Lee's extension which is denoted by $m1HypGO$ is thus derived from the $m2GO$ metamodel.

ProtGO (Yeh's extension [30]) This GO extension belongs to the *CO* family. Yeh transferred GO to Protégé in order to verify the ontology consistency while adding new concepts and relations (in order to cope with the problem of ever growing knowledge). Protégé is a knowledge base management system (KBMS) that allows ontol-

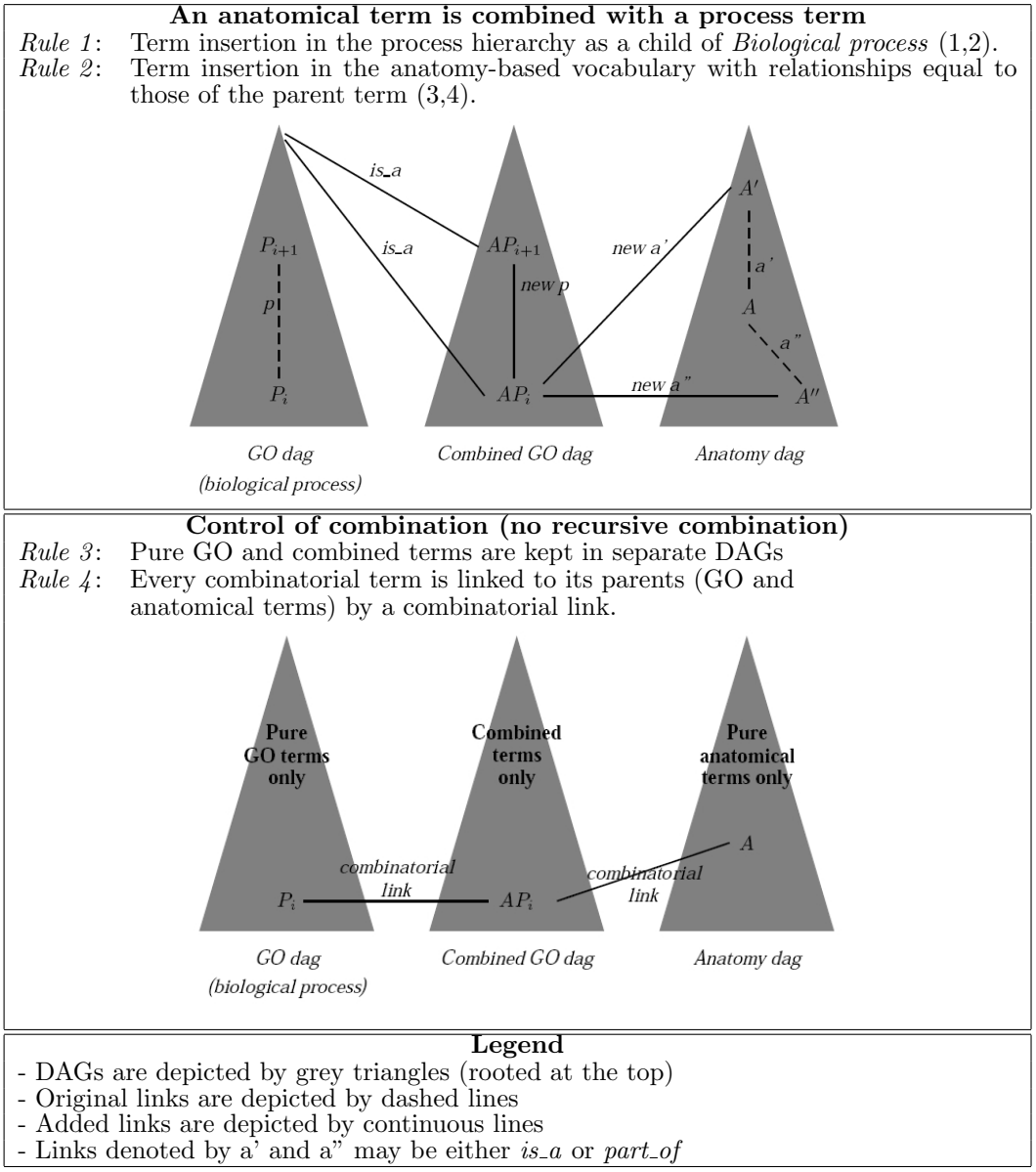

Figure 2: Definition of rules for combination of process and anatomy terms in Hill's extension

ogy development and its maintenance[1]. Various relations were created between sub-ontologies, e.g., *Part_of_Process* as an attribute of *molecular function*, in order to identify the process in which a particular function is involved, *Occurs_at_Component* as an attribute of *molecular function* and *biological process*, in order to specify a location where a molecular function or a biological process occurs. In addition, new concepts were added to GO, e.g., *Cellular space* as a subclass of *cellular component*, *Macromolecule* and *Complex* as subclasses of *cellular component*.

The two relationships (*Part_of_process* and *Occurs_at_component*) link concepts across distinct sub-ontologies, which remain separate in the Gene Ontology, and thus constitute a non-trivial extension of the ontology and of its structure.

The Yeh's metamodel, denoted by $m2ProtGO$, is derived from the *OBOti* metamodel (*specialization link 4*), as Yeh's extension uses intra-ontological (*is_a* and *part_of*) and trans-ontological relationships (between the *biological process* and the *molecular function* or between the *biological process* and *cellular component* sub-ontologies).

Yeh's model, denoted by $m1ProtGO$, is derived from the $m2ProtGO$ metamodel.

GONG (Wroe's extension [29]) This GO extension belongs to the *CO* family. Wroe uses a description logics paradigm (semantics of each concept is precisely defined by its position in the description hierarchy) implemented with DAML+OIL. The DAML+OIL language is designed for ontology development in the context of the semantic web. It allows to describe the application's domain in an object-oriented fashion by using classes and properties as constructs. The language supports relations such as *subClassOf*, *intersectionOf*, etc. Although it does not require that the subclass-relation between classes is acyclic, multiple inheritance is replaced by restrictions on objects and properties in order to avoid inconsistencies. Wroe used DAML+OIL to reorganize GO's DAG into a formal ontology[2].

[1] At the time of Yeh's work (2002), the *is_a* hierarchy of GO was not fully elaborated (due to the presence of "orphan terms"). Tool-specific relations were added between concepts to solve this problem. Such relations are not included in our work.

[2] The authors also reorganized the *part_of* hierarchy and intro-

The new organization is meant to facilitate the formal ontology analysis while metadata indicate that the new hierarchy is not to be used for annotation [29]. New concepts were created by linking GO concepts to concepts of external ontologies (e.g., Kegg enzyme database, Enzyme Classification, MESH ontology, UMLS).

GONG extension uses intra-ontological and trans-ontological relationships. Yet, the GONG metamodel differs from the $m2ProtGO$ metamodel since GONG uses trans-ontological relationships that are different from those used by Yeh. The GONG metamodel, denoted by $m2GONG$, is derived from the $OBOti$ metamodel (*specialization link 5*).

The GONG model, denoted by $m1GONG$, is derived from the $m2GONG$ metamodel.

CombGO (Hill's extension [11]) This GO extension belongs to the *CR* family. The purpose of Hill's extension, denoted by CombGO, was to automatically construct a vocabulary for extending GO by integrating orthogonal ontologies. Hill gives an example cross-product of a GO DAG with a DAG derived from an anatomical vocabulary[3] in which he "*condensed the anatomical vocabulary and defined* is_a *and* part_of *relationships between anatomical terms*. Hill's method combines the most general term of the first ontology with all the terms of the second ontology. Hill requires that "*no combinatorial term containing an anatomical component can be combined with another anatomical component term*". Three possible solutions are discussed. The author chooses to keep the pure GO DAG separated from the combined GO DAG. The author thus links (by a *combinatorial_link*) each concept corresponding to a GO combinatorial term with its anatomical concept, as well as with its GO parent concept. Figure 2 presents extension rules that allow producing combinatorial terms and inserting them into the GO DAG.

The Hill's metamodel is derived from the $OBOmti$ metamodel (*specialization link 6*) since Hill's combinatorial process uses all three types of relationships (meta-, trans-, and intra- relationships).

Hill's model, denoted by $m1CombGO$, is derived from the $m2CombGO$ metamodel.

CCO (Cell cycle ontology) [7] This GO extension belongs to the *CR* family. The Cell Cycle Ontology (CCO) was recently developed to expand GO in order to improve the Cell Cycle branch (by organizing concepts, introducing new relations and terms, and modifying the ontology structure to support logic-based reasoning [7, 13, 25]). The main design principle consists of the single *is-a* inheritance and the use of a set of relations described in Relation Ontology (RO) [24]. CCO aims to integrate temporal and dynamic aspects of the cell cycle process. CCO introduced 304 terms originating from GO, and 15 relations originating from RO (that are intra- and meta-ontological relations). Meta-ontological relations can be used to handle multi-granularity in descriptions.

The CCO extension also uses all three types of relationships (meta-, trans-, and intra-relationships). Yet, CCO is built as a formal ontology that induces additional constraints expressed at the metamodel level[4]. The CCO's metamodel, denoted by $m2OBOmti\text{-}f$, is derived from $OBOmti$ metamodel (*specialization link 7*) since CCO uses all three types of relationships.

The CCO model, denoted by $m1CCO$, is derived from the $m2OBOmti\text{-}f$ metamodel.

3 Conclusion

Various authors have emphasized the semantical fragility of ontology-based annotations due to precise meanings of relationships between concepts [20], possible semantical variations between paths and name compositions of concepts [22], instability of domain knowledge [12], and difficulties in maintaining the quality of ontology structures. For application domains that are in constant evolution, the amount of knowledge being introduced to or modified in ontologies is large. Such ontologies are often extended, as well as translated from one formalism to another. In databases that have been annotated with different versions of a particular ontology, the same concept can have slightly different meaning from one version to another, thus introducing semantical variations between data that were annotated with variants of this concept. Such semantical variations must be therefore taken into account.

The most common use and evolution of the Gene Ontology occur in the context of gene annotation that is based on new data published in the scientific literature. Databases are either manually or automatically annotated by consulting published articles in order to describe gene activity in GO terms. At this time, over 25.000 terms are used in GO. Nevertheless, due to evolution of concepts, GO terms might not be sufficiently expressive to account for the exponentially growing knowledge. Several hundred thousands of gene entries in databases have already been annotated using GO, with gene domains ranging from bacteria to plants and humans. Various GO extensions strive to enrich GO expressive capabilities. Yet, such improvements may introduce unforeseen terms and relationships that will be difficult to trace if there is no indication on the GO extension used. For example (see Figure 1, lower half), in RGD Database[5] the rat's *Smad4* gene (RGD:3033) is annotated with GO terms including a cytoplasm (PMID 16570350), a SMAD protein complex assembly (UniProt 13485), etc. The use of *ProtGO* allows specifying that the SMAD protein complex assembly is found in (i.e., *occurs_at*) the cytoplasm (PMID 16570350).

In this paper, we introduced a metamodeling architecture for integration of OBO ontologies. OBO classifies biological concepts as well as biological relationships. We based our work on OBO relationships since they appear to be more relevant for GO extensions than OBO concepts. Yet, similar results could be achieved by using OBO's biological concepts.

In our example (Figure 1), databases that use information from GO and Lee's extension (corresponding to $m1GO$ and $m1HypGO$ models) interoperate by using a basis of agreement that is described by their common metamodel $m2GO$ (which is semantically close to both of them). On the contrary, databases exploiting information from GO and Hill's extension (corresponding to $m1GO$ and $m1CombGO$ models) use a basis of agreement that is described

duced three additional concepts (e.g., *part_of_cellular_component*) in order to suppress orphan terms (see Footnote 1).

[3]Anatomical Dictionary used in the Mouse Gene Expression Database (MGED).

[4]Expression of such constraints at the metamodel-level is beyond the scope of this paper.

[5]RGD: Rat Genome Database (`http://rgd.mcw.edu`).

by the *OBOmti* metamodel (which is semantically somewhat different from *m2GO*).

Metamodeling appoaches that allow automatic tracing of semantic "distances" between models and metamodels may be of great help for fine tuning of database cooperation. An example application that illustrates our approach is part of the Life Explorer project (a visualization tool for molecular and synthetic biology [14]). In the future, we will extend our approach to various data sets and technologies. Omics sciences participate in the standardization movement, particularly in efforts pertaining to the storage, the exchange, and the analysis of data. We plan to use standards' specifications to derive models elements for metamodeling.

References

[1] The molecular biology database collection. *Nucl. Acid. Research*, (2007 Update):D3–D4, 2007.

[2] J. Bao, Z.L. Hu, D. Caragea, J. Reecy, and V. Honavar. Collaborative construction of large biological ontologies. In *Proceedings of the DEXA Workshop on Biological Data Management (BIDM'06)*, 2006.

[3] O. Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*, 32(Database issue), 2004.

[4] A. Brazma, M. Krestyaninova, and U. Sarkans. Standards for systems biology. *Nature Rev. Genetics*, 7(8):593–605, 2006.

[5] The Gene Ontology Consortium. The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(Database issue):D258–D261, 2004.

[6] The Gene Ontology Consortium. Gene ontology project in 2006. *Nucleic Acids Research*, 34:D322–D326, 2006.

[7] Antezana E., Tsiporkova E., Mironov V., and Kuiper M. A cell-cycle knowledge integration framework. In *Proc. of the 3rd Int. Workshop on Data Integration in the Life Sciences (DILS), LNCS 4075*, 2006.

[8] SP. Gardner. Ontologies and semantic data integration. *Drug Discovery Today*, 10(14):1001–1007, 2005.

[9] PV. Gopalacharyulu, E. Lindfors, C. Bounsaythip, T. Kivioja, L. Yetukuri, J. Hollmn, and M. Oresič. Data integration and visualization system for enabling conceptual biology. *Bioinformatics*, 21(Suppl. 1):i177–i185, 2005.

[10] P. Grenon. Spatio-temporality in basic formal ontology part i: SNAP and SPAN: Upper level ontology and framework for formalisation. Technical Report 5/03, Iformis, 2003.

[11] D.P. Hill, J.A. Blake, J.E. Richardson, and M. Ringwald. Extension and integration of the gene ontology (GO): Combining GO vocabularies with external vocabularies. *Genome Research*, 12(12):1982–1991, 2002.

[12] A.R. Jones and N.W. Paton. An analysis of extensible modelling for functional genomics data. *BMC Bioinformatics*, 6, 2005.

[13] T. Kushida, T. Takagi, and KI. Fukuda. Event ontology: a pathway-centric ontology for biological processes. pages 152–163, 2006. URL `http://helix-web.stanford.edu/psb06/`.

[14] D. Larivière, E. Fourmentin, O. Mavré, JF. Bonnet, MN. Terrasse, and M. Roux-Rouquié. Lifexplorer : a vizualisation tool for molecular and synthetic biology. Poster session, CNRS Thematic School, Evry, France, 2006.

[15] Stein LD and Thierry-Mieg J. Scriptable access to the caenorhabditis elegans genome sequence and other ACeDB databases. *Genome Research*, 8(12):1308–1315, 1998.

[16] J.B. Lee, J.J. Kim, and J.C. Park. Automatic extension of gene ontology with flexible identification of candidate terms. *Bioinformatics*, 22(6):665–670, 2006.

[17] B. Louie, P. Mork, F. Martin-Sanchez, A. Halevy, and P. Tarczy-Hornoch. Data integration and genomic medicine. *Journal of Biomedical Informatics*, 40(1):5–16, 2006.

[18] Ashburner M & al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature Genetics*, 25(1):25–29, 2000.

[19] Bada M. and Hunter L. Enrichment of OBO ontologies. *Journal of Biomedical Informatics*, 2006. (accepted paper).

[20] Donnelly M., Bittner T., and Rosse C. A formal theory for spatial representation and reasoning in biomedical ontologies. *Artificial Intelligence in Medicine*, 36(1):1–27, 2006.

[21] CJ. Mungall. Obol: integrating language and meaning in bio-ontologies. *Comparative and Functional Genomics*, 5:509–520, 2004.

[22] PV. Ogren, KB. Cohen, GK. Acquaah-Mensah, J. Eberlein, and L. Hunter. The compositional structure of gene ontology terms. In *Proceedings of the 9th Pacific Symposium on Biocomputing*, 2004.

[23] DC. Schmidt. Model-driven engineering. *IEEE*, pages 25–31, 2006.

[24] B. Smith, W. Ceusters, B. Klagges, J. Kohler J, A. Kumar, J. Lomax, C.J. Mungall, F. Neuhaus, A. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(R46), 2005.

[25] B. Smith, J. Williams, and S. Schulze-Kremer. The ontology of gene ontology. In *AMIA Annual Symposium Proceedings*, 2003.

[26] AD. Spear. Ontology for the twenty first century: An introduction with recommandations. 2006.

[27] Kirsten T. and Rahm E. Biofuice: Mapping-based data integration in bioinformatics. In *Proc. of the 3rd Int. Workshop on Data Integration in the Life Sciences (DILS), LNCS 4075*, 2006.

[28] M.N. Terrasse, M. Savonnet, E. Leclercq, and G. Becker. Building platforms for information system interoperability. In *Proceedings of the International Workshop on Engineering Federated Information Systems (EFIS), UK*, 2003.

[29] C. J. Wroe, R. Stevens, C. Goble, and M. Ashburner. A methodology to migrate the gene ontology to a description logic environment using DAML+OIL. In *Proceedings of the 8th Pacific Symposium on Biocomputing*, 2003.

[30] I. Yeh, P.D. Karp, N.F. Noy, and R.B. Altman. Knowledge acquisition, consistency checking and concurrency control for gene ontology (GO). *Bioinformatics*, 19(2):241–248, 2006.

[31] Bukhman YV and Skolnick J. Biomolquest: integrated database-based retrieval of protein structural and functional information. *Bioinformatics*, 17(5):468–478, 2001.