

Towards an Ontology of Biomodelling

Larisa Soldatova, Qian Gao, and David Gilbert

Department of Information Systems and Computing, Brunel University, Uxbridge,
UB8 3PH, London, UK.

{Larisa.Soldatova,Qian.Gao,David.Gilbert}@brunel.ac.uk

<http://brunel.ac.uk/siscm/disc/>

Abstract. We present a core Ontology of Biomodelling (OBM), which formally defines principle entities of modelling of biological systems, and follows a structural approach for the engineering of biochemical network models. OBM is fully interoperable with relevant resources, e.g. GO, SBML, ChEBI, and the recording of biomodelling knowledge with Ontology of Biomedical investigations (OBI) ensures efficient sharing and re-use of information, reproducibility of developed biomodels, retrieval of information regarding tools, methods, tasks, bio-models and their parts. An initial version of OBM is available at disc.brunel.ac.uk/obm.

Keywords: ontology; knowledge representation; systems biology, modelling.

1 Introduction

We propose an Ontology of Biomodelling (OBM) that enables formally defined description of the key information about the design and analysis of biological models, motivated by the need for interoperability and re-usability of scientific knowledge. OBM is an important element in BioModel Engineering, a structured approach for the engineering of biochemical network models [2], which facilitates the design, construction and analysis of computational models of biological systems.

Ontology engineering is a popular solution for integration, interoperability and re-usability of scientific knowledge and is related to several biomodelling ontological resources. Systems Biology Ontology (SBO)¹ is a set of controlled, relational vocabularies of terms commonly used in systems biology, in particular in computational modelling, and informs the development of SBML². The Ontology of Data Mining (OntoDM) enables recording of most essential information about predictive modelling as a type of data mining [8]. The Ontology of Biomedical investigations (OBI) provides semantic descriptors to report the most essential information about scientific investigations carried out in biomedical domains [4].

OBM employs an OBI approach to the reporting of investigations [4], incorporating all the relevant representations from other resources such as OntoDM

¹ <http://www.ebi.ac.uk/sbo>

² <http://sbml.org>

and SBO, and is fully interoperable with GO (Gene Ontology), ChEBI (Chemical Entities of Biological Interest), SBML and other biomedical resources. It is designed as a foundation for an environment to support the key steps of the construction and analysis of models of biological systems (see section 2). Such an environment would assist in the selection of methods and tools for the construction and development of a model, searching over available models and their parts, advising appropriate validation methods, and reporting output models in standard formats. An ontology-driven environment would serve as an integration platform for most existing tools.

2 A workflow of biomodelling

The following steps are most commonly presented in a typical scenario of biomodelling:

1. Identification of tasks and requirements for a model construction. Construction of a model of a bio-system is a purpose-led process. Such purposes or tasks along with specified requirements could and should be recorded and collected for the benefit of the research community. Some scientific domains already have such dedicated task ontologies [6]. A constructed model can be checked for how well it satisfies the specified requirements. Such analysis can be used for meta-learning to find patterns of what models and design methods are suitable for particular tasks and requirements [1].

2. Modelling a domain of interest. Currently developers of bio-models rely on manual literature searches, interviews with biologists and chemists, analysis of experimental results in order to provide an unambiguous representation of the knowledge about a target biochemical system. Such domain background knowledge relevant for modelling should be represented in a form of domain modelling ontologies and preserved for future re-use. Domain modelling ontologies could be populated by automatic text mining searches that would extract required information from scientific literature, e.g. parameters of a system and recommended values, lists of genes, proteins, and chemical reactions that are associated with a target system; and by facts from already existing resources such as data bases, knowledge bases, and other domain ontologies.

3. Selection of model type and associated construction method. There are different types of models, e.g. static, dynamic – qualitative, quantitative (continuous, stochastic, hybrid), and it is important to select an appropriate for the purpose. OBM (empowered by a task ontology and a methods ontology) enables queries over available methods and models such as “Is there already an existing model A that can be modified to satisfy requirements R?”, “Find all methods that are applicable for a task X with requirements R”, “What modifications of method A would lead to the satisfaction of requirement R?”, “Why is method A not suitable for task B?”. Additionally, ready-made ‘building blocks’ of models could be re-used to construct a new model. Breitling et al. proposed typical building blocks for modelling of cellular signaling models (see [3] for more

detail). In a similar way ready building blocks could be collected for other areas of interest, and offered to developers for specified tasks.

4. Selection of tools for model construction, analysis, and validation. There is a great variety of software available for modelling tasks. The Software Ontology (SWO)³ is a resource for describing software tools, their types, tasks, versions, provenance and data associated. OBM (empowered by a task ontology and SWO) enables such queries over available methods and tools as “List all available tools for a task X and a method M”, “Is a tool T well supported and has a large user community?”, “Is software S freely available for academic purposes?”.

5. Verification of a model. Once a model for a biological system has been created, it needs to be validated in a principled way. Does it produce reasonable predictions of system behavior? What datasets were used to test the model, and what properties do they have? Is the model safe against deadlock and other system failures? Has it been tested in wet laboratory experiments? OntoDM provides formalized description of various model verification methods, and also of datasets that are used to test models [8], [7]. OBM extends OntoDM descriptors by the description of model checking by wet experimentations.

6. Exploration of a model. Once a model for a biological system has been created and verified, it can be used for simulating system behavior under various conditions. A domain modelling ontology could supply parameters and their values as an input for model simulation. All produced versions of a model should be recorded for further analysis, meta-learning, and re-use. Some tools, e.g. BioNessie⁴, allow recording of model versions and simulations runs, but it is still not a common practice in systems biology to record and report this information.

7. Reporting. Currently SBML is widely used for recording and reporting of biomodels. However, it is important to do this not only the final model, but also for all steps in its development, versions, and verification so that scientist could make informative decisions about how to use models. Many parts of models are re-usable and should also be recorded as separate entities. OBM enables the recording and reporting of the key information about the process of the development, analysis, and verification in a machine processable form.

The basic methodology of biomodelling can be extended to meet the challenges of multiscale modelling of complex biological systems (see [5] for more detail).

3 OBM: a core ontology of biomodelling

OBM follows OBI in the representation of a typical scientific workflow [4]. OBM imports from OBI classes that are relevant for the area of modelling of biological systems, e.g. investigator, planned-process, objective, conclusion-textual-entity,

³ <http://theswo.sourceforge.net/>

⁴ disc.brunel.ac.uk/bionessie/

and relations between these classes, e.g. has-specified-input, precedes. Additionally, OBM defines biomodelling - specific classes for the representation of the area of modelling of biological systems, e.g. model, model-component, task-identification, model-verification, model-representation, and relations between them, e.g. is-model-of.

We have followed the best practices in ontology engineering in the development of OBM. OBM employs standard upper level classes and relations where possible to ensure full interoperability with key biomedical ontologies and other resources, i.e. ChEBI, GO, OBI. OBM is designed in such a way that it complements other ontological resources, e.g. SWO for the description of software, OntoDM for the description of predictive modelling, SBO for the description of a model, that are necessary for the efficient recording of the most essential information about biomodelling.

Future work. The development and application of OBM will have the following next stages: (1) an extension of the coverage of OBM in order to include various biomodelling scenarios, and not only most typical ones; (2) an instantiation of OBM in order to enable search over workflows and their steps (currently OBM provides the conceptual description of biomodelling workflows, i.e. at the class level); (3) support of the development of an ontology-driven environment for biomodelling.

References

1. P. Brazdil, G. C. Carrier, C. Soares, and R. Vilalta. *Metalearning. Applications to Data Mining*. Springer, 2009.
2. R. Breitling, R. Donaldson, D. Gilbert, and M. Heiner. Biomodel engineering - from structure to behavior. *Trans. on Comput. Syst. Biol.*, 12:1–12, 2010.
3. R. Breitling, D. Gilbert, M. Heiner, and R. Orton. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. *Briefings in Bioinformatics*, 12, 2008.
4. R. R. Brinkman, M. Courtot, D. Derom, J. M. Fostel, Y. He, P. Lord, J. Malone, H. Parkinson, B. Peters, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, L. N. Soldatova, C. J. Stoeckert, J. A. Turner, J. Zheng, and T. OBI Consortium. Modeling biomedical experimental processes with OBI. *J. of Biomedical Semantics*, 1:1–12, 2010.
5. Q. Gao, D. Gilbert, M. Heiner, F. Liu, D. Maccagnola, and D. Tree. Multiscale modelling and analysis of planar cell polarity in the drosophila wing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press, 2012.
6. R. Mizoguchi, J. Vanwelkenhuysen, and M. Ikeda. *Towards very large knowledge bases. Task ontology for reuse of problem solving knowledge*. IOS Press, 1995.
7. P. Panov, S. Džeroski, and L. N. Soldatova. OntoDM: An ontology of data mining. In *ICDMW '08: Proceedings of the 2008 IEEE International Conference on Data Mining Workshops*, pages 752–760. IEEE Computer Society, 2008.
8. P. Panov, L. N. Soldatova, and S. Džeroski. Towards an ontology of data mining investigations. In *Discovery Science, 12th International Conference, DS 2009, Porto, Portugal, October 3-5, 2009*, volume 5808 of *LNCS*, pages 257–271. Springer, 2009.