Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2020

Falk Schreiber[1,2*], Björn Sommer[3], Gary D. Bader[4], Padraig Gleeson[5], Martin Golebiewski[6], Thomas E. Gorochowski[7], Michael Hucka[8], Sarah M. Keating[9], Matthias König[10], Chris Myers[11], David Nickerson[12] and Dagmar Waltemath[13]

[1]Dep. of Computer and Information Science, University of Konstanz, Germany

[2]Faculty of IT, Monash University, Clayton, Australia

[3]Royal College of Arts, London, UK

[4]The Donnelly Centre, University of Toronto, Canada

[5]Dep. of Neuroscience, Physiology and Pharmacology, University College London, UK

[6]Heidelberg Institute for Theoretical Studies (HITS), Germany

[7]School of Biological Sciences, University of Bristol, UK

[8]California Institute of Technology, USA

[9]Heidelberg University, Germany

[10]Humboldt-University Berlin, Germany

[11]University of Utah, USA

[12]Auckland Bioengineering Institute, University of Auckland, New Zealand

[13]University Medicine Greifswald, Germany

[*]To whom correspondence should be addressed. Email:falk.schreiber@uni-konstanz.de

Abstract

This special issue of the *Journal of Integrative Bioinformatics* presents papers related to the 10th COMBINE meeting together with the annual update of COMBINE standards in systems and synthetic biology.

1 Introduction

COMBINE ('COmputational Modeling in BIology' NEtwork) [1, 2], the formal entity which coordinates the development of standards in systems and synthetic biology, celebrated 10 years of activity in 2019. COMBINE not only coordinates standard developments, but also fosters and moderates discussions; designs and implements dissemination strategies; and organises two annual community meetings – COMBINE and HARMONY.

This special issue contains two papers, one new standard, and six updates of standards. Waltemath *et al.* [3] discuss the first 10 years of the international coordination network for standards in systems and synthetic biology, and summarises the COMBINE meeting in Heidelberg in July 2019. Brunak *et al.* [4] present ongoing works and open questions towards standardisation guidelines for *in silico* approaches in personalised medicine.

COMBINE standards and associated initiatives cover a wide range of disciplines, see Fig. 1. This special issue only highlights updates over the last year, namely the CellML 2.0 specification, the SBGNML Milestone 3, the SBML Level 3 Packages Distrib and Multi, the SBOL Version 3.0.0, and the SBOL Visual Version 2.2. Additionally, one new standard has officially been added, OMEX metadata specification 1.0, to harmonise the descriptions of metadata.

Further information on all standards and activities as well as links to the community

websites are available from the COMBINE web site at `https://co.mbine.org/`. Detailed overviews of COMBINE, its history and its organisation have been provided, for example, by Hucka *et al.* [1], Myers *et al.* [2] or Waltemath *et al.* [5] . The annual special issue on COMBINE standards has become a tradition since its launch in 2016. Earlier editions provide summaries of updates for the years 2015-18 [6, 7, 8, 9].

We hope that this editorial is helpful in identifying the relevant specification documents for standards in systems in synthetic biology in the year 2020.

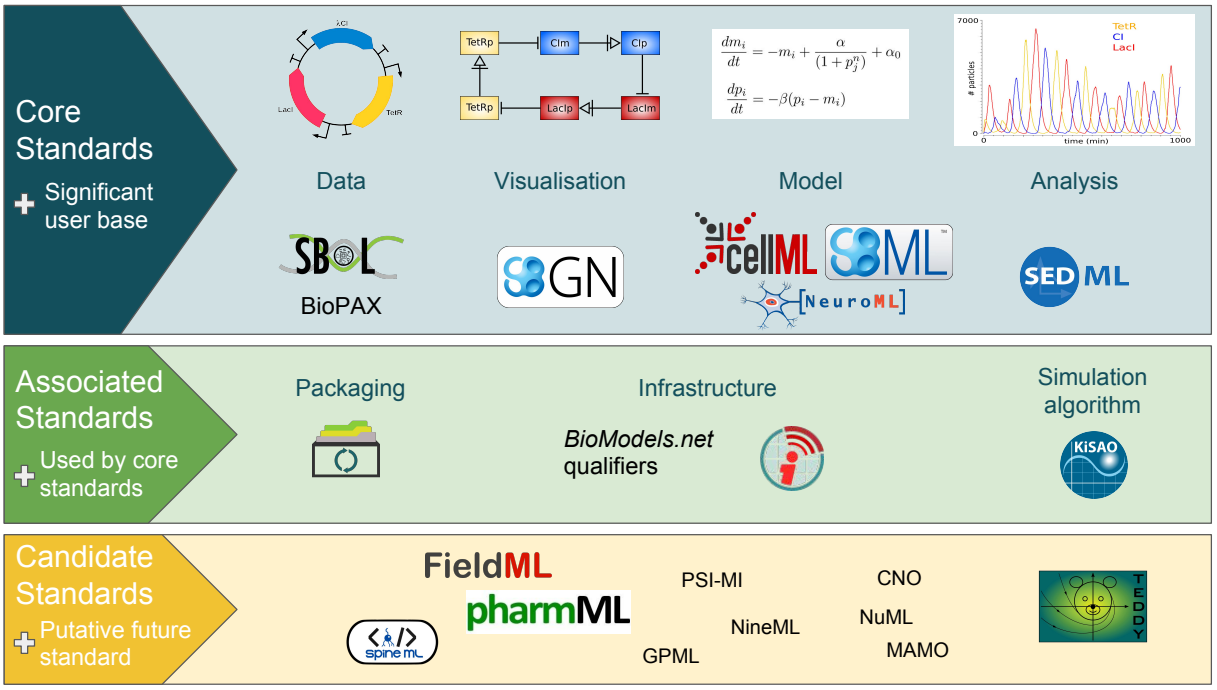

Figure 1: Standards and associated efforts in systems and synthetic biology (from [7])

2 Current versions of COMBINE standards

Please refer to the following specifications when using COMBINE standards. New specifications or updates of existing specifications are highlighted with **NEW**.

2.1 Core Standards

2.1.1 BioPAX (Biological PAthway eXchange)

BioPAX is a standard language for integration, exchange and analysis of biological pathway data. It is expressed in OWL. The current specification is listed in Table 1, no updates in 2019.

Standard	Specification	Reference
BioPAX [10]	BioPAX	[11]

Table 1: Latest specifications for BioPAX

2.1.2 CellML

The CellML language is an XML markup language to store and exchange computer-based mathematical models. The current specifications are listed in Table 2.

NEW CellML 2.0 [12] The development of CellML 2.0 was guided by observing the use of CellML 1.1 in the community for over 10 years. The syntax of CellML 1.1 has been clarified in areas where discrepancies in model interpretation are often seen and simplified to remove features that are never used. These enhancements are primarily aimed at improving the model reuse capabilities of CellML. The single substantial addition to CellML that is introduced in CellML 2.0 is the concept of *resets*, rules that define a change in model state dependent on specified conditions being met during a simulation experiment.

Standard	Specification	Reference
CellML [13]	CellML 2.0	[12]
	CellML Metadata Framework 2.0	[14]
	CellML 1.1	[15]

Table 2: Latest specifications for CellML

2.1.3 NeuroML

The Neural Open Markup Language (NeuroML) is an XML based description language that provides a common data format for defining and exchanging descriptions of neuronal cell and network models. The current specification is listed in Table 3.

Standard	Specification	Reference
NeuroML [16, 17]	NeuroML version 2.0	[16]

Table 3: Latest specifications for BioPAX

2.1.4 SBGN (Systems Biology Graphical Notation)

The Systems Biology Graphical Notation (SBGN), is a set standard graphical languages to describe visually biological knowledge. It is currently made up of three languages describing Process Descriptions, Entity Relationships and Activity Flows. The current specifications are listed in Table 4.

NEW SBGNML Milestone 3 [18] includes new developments such as support for multiple SBGN maps within a single file, complete support for the submap glyph, and the possibility to store colors and annotations through extensions. In addition the *language* attribute has been deprecated to add a more detailed *version* attribute and the SBGN

AF perturbation glyph has been deprecated to align with the SBGN AF specification.

Standard	Specification	Reference
SBGN [19]	SBGN Process Description Level 1 Version 2	[20]
	SBGN Entity Relationship Level 1 Version 2.0	[21]
	SBGN Activity Flow Level 1 Version 1.2	[22]
	SBGN Markup Language Version 0.3	[18]

Table 4: Latest specifications for SBGN

2.1.5 SBML (Systems Biology Markup Language)

The Systems Biology Markup Language (SBML) is a computer-readable XML format for representing models of biological processes. SBML is suitable for, but not limited to, models using a process description approach. SBML development is coordinated by an elected editorial board and central developer team.

The current specifications are listed in Table 5.

NEW SBML Level 3 Package: Distributions, Version 1, Release 1 [25] introduces distributions and uncertainties to SBML. Biological models often contain elements that have inexact numerical values, since they are based on values that are stochastic in nature or data that contains uncertainty. The SBML Level 3 Core specification does not include an explicit mechanism to include inexact or stochastic values in a model, but it does provide a mechanism for SBML packages to extend the Core specification and add additional syntactic constructs. The SBML Distributions package for SBML Level 3 adds the necessary features to allow models to encode information about the distribution and uncertainty of values underlying a quantity.

Standard	Specification	Reference
SBML [23]	SBML Level 3 Core, Version 2 Release 2	[24]
	SBML Level 3 Package: Distributions, Version 1, Release 1	[25]
	SBML Level 3 Package: Flux Balance Constraints Version 2	[26]
	SBML Level 3 Package: Groups, Version 1	[27]
	SBML Level 3 Package: Hierarchical Model Composition, Version 1	[28]
	SBML Level 3 Package: Layout, Version 1	[29]
	SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1 Release 2	[30]
	SBML Level 3 Package: Qualitative Models, Version 1	[31]
	SBML Level 3 Package: Render, Version 1, Release 1	[32]

Table 5: Latest specifications for SBGN

NEW SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1 Release 2 [30]

2.1.6 SBOL (Synthetic Biology Open Language)

The Synthetic Biology Open Language (SBOL) is a language for the description and the exchange of synthetic biological parts, devices and systems. The current specifications are listed in Table 6.

NEW Synthetic Biology Open Language (SBOL) Version 3.0.0 [33] condenses and simplifies previous versions of SBOL based on experiences in deployment across a variety of scientific and industrial settings. In particular, SBOL 3.0.0, (1) separates sequence features from part/sub-part relationships, (2) renames ComponentDefinition/Compo-

nent to Component/SubComponent, (3) merges Component and Module classes, (4) ensuring consistency between data model and ontology terms, (5) extends the means to define and reference SubComponents, (6) refines requirements on object URIs, (7) enables graph-based serialization, (8) moves Systems Biology Ontology (SBO) for Component types, (9) makes all sequence associations explicit, (10) makes interfaces explicit, (11) generalizes SequenceConstraints into a general structural Constraint class, and (12) expands the set of allowed constraints.

NEW Synthetic Biology Open Language Visual (SBOL Visual) Version 2.2 [34] is a refinement to the standard harmonising the ontology used and extending the glyph library to capture new biological parts. Specifically, the changes in SBOL Visual 2.2. include, (1) the grounding of the molecular species glyphs is changed from BioPAX to SBO to better align with the use of SBO terms for interaction glyphs, (2) new glyphs are added for proteins, introns, and polypeptide regions (e.g., protein domains), (3) the prior recommended macro- molecule glyph is deprecated in favor of its alternative, and (4) small polygons are proposed as alternative glyphs for simple chemicals.

Standard	Specification	Reference
SBOL [35]	SBOL Version 3.0.0	[33]
	SBOL Visual Version 2.2	[34]

Table 6: Latest specifications for SBOL

2.1.7 SED-ML (Simulation Experiment Description Markup Language)

The Simulation Experiment Description Markup Language is an XML-based format for encoding simulation experiments. SED-ML allows to define the models to use, the

experimental tasks to run and which results to produce. SED-ML can be used with models encoded in several languages, as long as they are in XML.

Standard	Specification	Reference
SED-ML [36]	SED-ML Level 1 Version 3	[37]

Table 7: Latest specifications for SED-ML

2.2 Associated Standards

Associated standards provide an additional layer of semantics to COMBINE representation formats. A COMBINE archive is a single file bundling the various documents necessary for a modelling and simulation project, and all relevant information. The archive is encoded using the Open Modeling EXchange format (OMEX). COMBINE archive metadata provides a harmonized, community-driven approach for annotating a variety of standardized model and data representation formats within a COMBINE archive. BioModels.net qualifiers are standardized relationships (predicates) that specify the relation between an object represented in a description language and the external resource used to annotate it. The relationship is rarely one-to-one, and the information content of an annotation is greatly increased if one knows what it represents, rather than only know it is ”related to” the model component. MIRIAM Unique Resource Identifiers allow one to uniquely and unambiguously identify an entity in a stable and perennial manner. MIRIAM Registry is a set of services and resources that provide support for generating, interpreting and resolving MIRIAM URIs. Through the Identifiers.org technology, MIRIAM URIs can be dereferenced in a flexible and robust way.

MIRIAM URIs are used by SBML, SED-ML, CellML and BioPAX controlled annotation

schemes. The Systems Biology Ontology (SBO) is a set of controlled, relational vocabularies of terms commonly used in Systems Biology, and in particular in computational modeling.

Each element of an SBML file carries an optional attribute sboTerm which value must be a term from SBO. Each symbol of SBGN is associated with an SBO term.

The Kinetic Simulation Algorithm Ontology (KiSAO) describes existing algorithms and their inter-relationships through their characteristics and parameters.

KiSAO is used in SED-ML, which allows simulation software to automatically choose the best algorithm available to perform a simulation and unambiguously refer to it.

NEW OMEX Metadata Specification Version 1.0 [38] The OMEX Metadata Specification is a technical implementation of the community consensus across COMBINE standards to harmonise the way we describe computational models and other resources with metadata [39]. This specification defines how COMBINE Archives should be annotated with additional knowledge to enable reproducible comprehension of the biological system being modelled in the archive. Including this knowledge in the COMBINE Archive framework helps to ensure that all the required information is shared in a consistent manner, allowing software tools to reason over this metadata, and helping modellers and model users reuse models in COMBINE standards with confidence.

References

[1] M. Hucka, D. P. Nickerson, G. D. Bader et al. Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative. *Frontiers in Bioengineering and Biotechnology*, 3:19, 2015.

Associated standard	Specification	Reference
COMBINE Archive [40]	COMBINE Archive 1.0	[41]
OMEX Metadata	OMEX Metadata Version 1.0	[38]
BioModels.net qualifiers [42]	-	[43]
Identifiers.org URIs [44]	-	[45]
Systems Biology Ontology [46]	[external] Bioportal	[47]
Kinetic Simulation Algorithm Ontology [46]	[external] Bioportal	[48]

Table 8: Latest specifications of associated standards

[2] C. Myers, G. D. Bader, P. Gleeson, M. Golebiewski, M. Hucka, N. Le Novère, D. Nickerson, F. Schreiber and D. Waltemath. A brief history of COMBINE. In *Proceedings of the 2017 Winter Simulation Conference*, pages 884–895. 2017.

[3] D. Waltemath, M. Golebiewski, M. L. Blinov et al. The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE). *Journal of Integrative Bioinformatics*, 17(2):5, 2020.

[4] S. Brunak, C. B. Collin, K. E. O Cathaoir, M. Golebiewski, M. Kirschner, I. Kockum, H. Moser and D. Waltemath. Towards standardization guidelines for in silico approaches in personalized medicine. *Journal of Integrative Bioinformatics*, 17(2):6, 2020.

[5] D. Waltemath, F. T. Bergmann, C. Chaouiya et al. Meeting report from the fourth meeting of the computational modeling in biology network (COMBINE). *Standards in Genomic Sciences*, 9(3):1285–1301, 2014.

[6] F. Schreiber, G. D. Bader, M. Golebiewski et al. Specifications of standards in systems and synthetic biology. *Journal of Integrative Bioinformatics*, 12(2):258,

2015.

[7] F. Schreiber, G. D. Bader, P. Gleeson, M. Golebiewski, M. Hucka, N. Le Novère, C. Myers, D. Nickerson, B. Sommer and D. Waltemath. Specifications of standards in systems and synthetic biology: Status and developments in 2016. *Journal of Integrative Bioinformatics*, 13(3):289, 2016.

[8] F. Schreiber, G. D. Bader, P. Gleeson et al. Specifications of standards in systems and synthetic biology: Status and developments in 2018. *Journal of Integrative Bioinformatics*, 15(1):13, 2018.

[9] F. Schreiber, B. Sommer, G. D. Bader et al. Specifications of standards in systems and synthetic biology: Status and developments in 2019. *Journal of Integrative Bioinformatics*, 16(2):35, 2019.

[10] E. Demir, M. P. Cary, S. Paley et al. The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28:935–942, 2010.

[11] BioPax, 2017. http://www.biopax.org/.

[12] M. Clerx, M. T. Cooling, J. Cooper, A. Garny, K. Moyle, D. P. Nickerson, P. M. F. Nielsen and H. Sorby. CellML 2.0. *Journal of Integrative Bioinformatics*, 17(2):21, 2020.

[13] A. A. Cuellar, C. M. Lloyd, P. F. Nielsen, D. Bullivant, D. Nickerson and P. Hunter. An overview of CellML 1.1, a biological model description language. *Simulation*, 79(12):740–747, 2003.

[14] M. T. Cooling and P. J. Hunter. The CellML metadata framework 2.0 specification. *Journal of Integrative Bioinformatics*, 12(2):260, 2015.

[15] A. A. Cuellar, W. Hedley, M. Nelson, C. M. Lloyd, M. D. B. Halstead, D. P. Bullivant, D. P. Nickerson, P. J. Hunter and P. M. F. Nielsen. The CellML 1.1 specification. *Journal of Integrative Bioinformatics*, 12(2):259, 2015.

[16] R. C. Cannon, P. Gleeson, S. Crook, G. Ganapathy, B. Marin, E. Piasini and R. A. Silver. LEMS: A language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Frontiers in Neuroinformatics*, 8(79), 2014.

[17] P. Gleeson, S. Crook, R. C. Cannon et al. NeuroML: A language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Computational Biology*, 6(6):e1000815, 2010.

[18] F. T. Bergmann, T. Czauderna, U. Dogrusoz, A. Rougny, A. Dräger, V. Toure, A. Mazein, M. L. Blinov and A. Luna. Systems Biology Graphical Notation Markup Language (SBGNML) Version 0.3. *Journal of Integrative Bioinformatics*, 17(2):16, 2020.

[19] N. Le Novère, M. Hucka, H. Mi et al. The Systems Biology Graphical Notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[20] A. Rougny, V. Toure, S. Moodie et al. Systems biology graphical notation: Process description language level 1 version 2. *Journal of Integrative Bioinformatics*, 16(3):22, 2019.

[21] A. A. Sorokin, N. Le Novère, A. Luna, T. Czauderna, E. Demir, R. Haw, H. Mi, S. L. Moodie, F. Schreiber and A. Villéger. Systems biology graphical notation: Entity relationship language level 1 version 2. *Journal of Integrative Bioinformatics*, 12(2):264, 2015.

[22] H. Mi, F. Schreiber, S. L. Moodie, T. Czauderna, E. Demir, R. Haw, A. Luna, N. Le Novère, A. A. Sorokin and A. Villéger. Systems biology graphical notation: Activity flow language level 1 version 1.2. *Journal of Integrative Bioinformatics*, 12(2):265, 2015.

[23] M. Hucka, A. Finney, H. M. Sauro et al. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[24] M. Hucka, F. T. Bergmann, C. Chaouiya et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *Journal of Integrative Bioinformatics*, 16(3):21, 2019.

[25] L. P. Smith, S. L. Moodie, F. T. Bergmann, C. Gillespie, S. M. Keating, M. König, C. J. Myers, M. J. Swat, D. J. Wilkinson and M. Hucka. SBML Level 3 Package: Distributions, Version 1, Release 1. *Journal of Integrative Bioinformatics*, 17(2):18, 2020.

[26] B. G. Olivier and F. T. Bergmann. SBML level 3 package: Flux balance constraints, version 2. *Journal of Integrative Bioinformatics*, 15(1):82, 2018.

[27] M. Hucka and L. P. Smith. The Systems Biology Markup Language (SBML) Level 3 Package: Groups, Version 1 Release 1. *Journal of Integrative Bioinformatics*, 13(3):290, 2016.

[28] L. P. Smith, M. Hucka, S. Hoops, A. Finney, M. Ginkel, C. J. Myers, I. I. Moraru and W. Liebermeister. SBML level 3 package: Hierarchical model composition, version 1 release 3. *Journal of Integrative Bioinformatics*, 12(2):268, 2015.

[29] R. Gauges, U. Rost, S. Sahle, K. Wengler and F. T. Bergmann. The systems biology markup language (SBML) level 3 package: Layout, version 1 core. *Journal of Integrative Bioinformatics*, 12(2):267, 2015.

[30] F. Zhang, L. P. Smith, M. L. Blinov et al. SBML Level 3 package: Multistate, Multicomponent and Multicompartment Species, Version 1, Release 2. *Journal of Integrative Bioinformatics*, 17(2):15, 2020.

[31] C. Chaouiya, S. M. Keating, D. Bérenguier, A. Naldi, D. Thieffry, M. P. van Iersel, N. Le Novère and T. Helikar. The systems biology markup language (SBML) level 3 package: Qualitative models, version 1, release 1. *Journal of Integrative Bioinformatics*, 12(2):270, 2015.

[32] F. T. Bergmann, S. M. Keating, R. Gauges, S. Sahle and K. Wengler. SBML level 3 package: Render, version 1, release 1. *Journal of Integrative Bioinformatics*, 15(1):78, 2018.

[33] H. Baig, P. Fontanarrosa, V. Kulkarni et al. Synthetic Biology Open Language (SBOL) Version 3.0.0. *Journal of Integrative Bioinformatics*, 17(2):17, 2020.

[34] H. Baig, P. Fontanarossa, V. Kulkarni et al. Synthetic Biology Open Language Visual (SBOL Visual) Version 2.2. *Journal of Integrative Bioinformatics*, 17(2):14, 2020.

[35] M. Galdzicki, K. P. Clancy, E. Oberortner et al. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology*, 32(6):545–550, 2014.

[36] D. Waltemath, R. Adams, F. T. Bergmann et al. Reproducible computational biology experiments with SED-ML – the Simulation Experiment Description Markup Language. *BMC Systems Biology*, 5(1):198, 2011.

[37] F. Bergmann, J. Cooper, M. König, I. Moraru, D. Nickerson, N. L. Novère, B. Olivier, S. Sahle, L. Smith and D. Waltemath. Simulation Experiment Description Markup Language (SED-ML) Level 1 Version 3 (L1V3). *Journal of Integrative Bioinformatics*, 15(1):86, 2018.

[38] M. L. Neal, J. H. Gennari, D. Waltemath, D. P. Nickerson and M. König. OMEX Metadata Specification Version 1.0. *Journal of Integrative Bioinformatics*, 17(2):20, 2020.

[39] M. L. Neal, M. König, D. Nickerson et al. Harmonizing semantic annotations for computational models in biology. *Briefings in Bioinformatics*, 20(2):540–550, 2019.

[40] F. T. Bergmann, R. Adams, S. Moodie et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. *BMC Bioinformatics*, 15(1):369, 2014.

[41] F. T. Bergmann, N. Rodriguez and N. Le Novère. COMBINE archive specification version 1. *Journal of Integrative Bioinformatics*, 12(2):261, 2015.

[42] C. Li, M. Courtot, N. Le Novère and C. Laibe. Biomodels. net web services, a free and integrated toolkit for computational modelling software. *Briefings in bioinformatics*, 11(3):270–277, 2010.

[43] BioModels.net Qualifiers, 2020. http://co.mbine.org/standards/qualifiers.

[44] N. Juty, N. Le Novère and C. Laibe. Identifiers. org and miriam registry: community resources to provide persistent identification. *Nucleic Acids Research*, 40(D1):D580–D586, 2012.

[45] Identifiers.org, 2020. http://identifiers.org/.

[46] M. Courtot, N. Juty, C. Knüpfer et al. Controlled vocabularies and semantics in systems biology. *Molecular Systems Biology*, 7(1):543, 2011.

[47] Systems Biology Ontology on BioPortal, 2019. http://bioportal.bioontology.org/ontologies/SBO.

[48] Kinetic Simulation Algorithm Ontology on BioPortal, 2019. http://bioportal.bioontology.org/ontologies/KISAO.