Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2021

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

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

[2]Faculty of Information Technology, Monash University, Clayton, Australia

[3]University College London, UK

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

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

[6]California Institute of Technology, USA

[7]Institute for Theoretical Biology, Humboldt-University Berlin, Germany

[8]Dept. of Electrical, Computer, and Energy Eng., University of Colorado Boulder, USA

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

[10]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* contains updated specifications of COMBINE standards in systems and synthetic biology. The 2021 special issue presents four updates of standards: Synthetic Biology Open Language Visual Version 2.3, Synthetic Biology Open Language Visual Version 3.0, Simulation Experiment Description Markup Language Level 1 Version 4, and OMEX Metadata specification Version 1.2. This document can also be consulted to identify the latest specifications of all COMBINE standards.

1   Introduction

COMBINE ('COmputational Modeling in BIology' NEtwork) [1, 2] is the formal entity coordinating the development of standards in systems and synthetic biology. It was founded in 2009, and since then supports and coordinates standard developments across the globe and thereby fosters and moderates discussions; designs and implements dissemination strategies; and offers a central access point to specifications and library support. The COMBINE coordination board organises two annual community meetings – the COMBINE forum and the HARMONY hackathon.

Figure 1 shows an overview of COMBINE standards and associated initiatives. This editorial will present the latest specifications of all COMBINE standards, and this special issue highlights updates over the last year, namely the releases of the following specifications: OMEX Metadata Version 1.2, SED-ML Level 1 Version 4, SBOL Visual 2.3 and 3.0. Special issues on COMBINE standards have been published since 2016, and earlier editions [4, 5, 3, 6, 7] provide updates for the years 2015 to 2020.
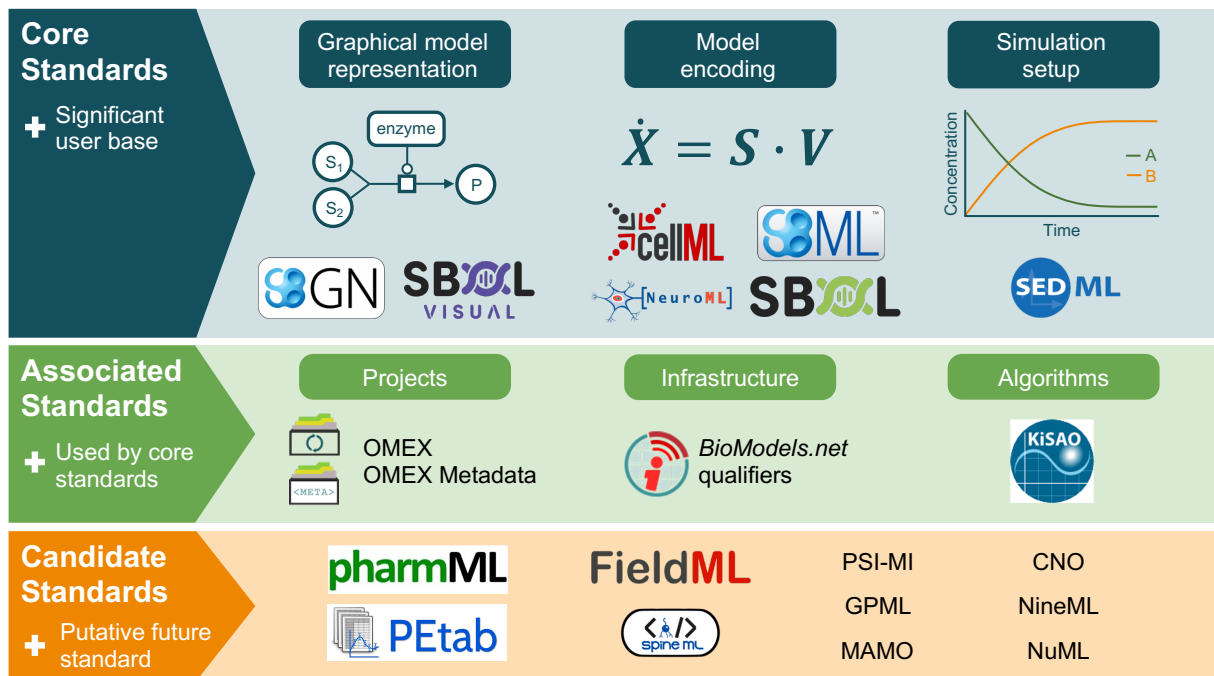
Figure 1: COMBINE standards and associated efforts (updated from [3])

The community websites of the different standards and initiatives are available from the COMBINE web site at `https://co.mbine.org/`, this web site also contains links to COMBINE events such as the COMBINE forum and the HARMONY hackathon. Additional information, the history of COMBINE as well as its organisation can be found in publications, for example, by Hucka *et al.* [2], Myers *et al.* [8], Waltemath *et al.* [9] or Dräger and Waltemath [10].

As in the last editorial [7] this editorial will in the remaining part also provide a brief updated overview of the latest specifications of all COMBINE standards. Please refer to those versions of specifications for the most current versions. Again, we hope that this editorial helps in identifying relevant specification documents for the different standards in systems in synthetic biology in the year 2021.

2 Current versions of COMBINE standards

Please refer to the following (most up-to-date) 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:

Standard	Specification	Reference
BioPAX [11]	BioPAX	[12]

2.1.2 CellML

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

Standard	Specification	Reference
CellML [13]	CellML 2.0	[14]
	CellML Metadata Framework 2.0	[15]
	CellML 1.1	[16]

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:

Standard	Specification	Reference
NeuroML [17, 18]	NeuroML version 2.1	[17]

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. In addition, SBGN-ML is a XML-based file format describing the geometry of SBGN maps, while preserving their underlying biological meaning. The current specifications are:

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	[23]

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:

Standard	Specification	Reference
SBML [24]	SBML Level 3 Core, Version 2 Release 2	[25]
	SBML Level 3 Package: Distributions, Version 1, Release 1	[26]
	SBML Level 3 Package: Flux Balance Constraints Version 2	[27]
	SBML Level 3 Package: Groups, Version 1	[28]
	SBML Level 3 Package: Hierarchical Model Composition, Version 1	[29]
	SBML Level 3 Package: Layout, Version 1	[30]
	SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1, Release 2	[31]
	SBML Level 3 Package: Qualitative Models, Version 1	[32]
	SBML Level 3 Package: Render, Version 1, Release 1	[33]

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. SBOL Visual (SBOLv) is a complementary standard that provides a standard set of glyphs and rules for drawing genetic circuit diagrams. The current specifications are listed in the following table.

NEW Synthetic Biology Open Language Visual (SBOL Visual) Version 2.3 [34] updates SBOL Visual Version 2.2 with several new features. These include higher-level "interactions with interactions," the representation of the binding with a nucleic acid backbone using overlapping glyphs, and a new "unspecified interaction" glyph. In this version, the "insulator" glyph is deprecated and replaced with a new "inert DNA spacer" glyph, and the polypeptide region glyph is now recommended for showing 2A sequences.

NEW SBOL Visual Version 3.0 [35] is major revision of the SBOL Visual standard in which the diagrams and glyphs are now defined with respect to the SBOL 3 data model rather than the SBOL 2 data model. In addition, the use of dashed undirected lines for subsystem mappings has been removed, and deprecated material has been removed from the collection of glyphs. Finally, the deprecated BioPAX alternatives to SBO terms has been removed.

Standard	Specification	Reference
SBOL [36]	SBOL Version 3.0.0	[37]
	SBOL Visual Version 2.3	[34]
	SBOL Visual Version 3.0	[35]

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. The current specification is listed in the following table.

NEW The Simulation Experiment Description Markup Language (SED-ML) Level 1 Version 4 [38] clarifies previous versions of SED-ML and extends the language with multiple new features. Specifically, enhancements introduced in SED-ML Level 1 Version 4 include (1) enriched plotting capabilities, (2) dimension reductions and math on multidimensional data, (3) support for simple parameter fitting experiments, and (4) a generic Analysis task. Further refinements in this version of the specification aim to clarify the use of SED-ML with non-XML model description formats.

Standard	Specification	Reference
SED-ML [39]	SED-ML Level 1 Version 4	[38]

2.2   Associated Standards

Associated standards provide an additional layer of semantics to COMBINE representation formats. The current specifications are listed in the following table:

Standard	Specification	Reference
COMBINE Archive [40]	COMBINE Archive 1.0	[41]
OMEX Metadata	OMEX Metadata Version 1.2	[?]
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]

A COMBINE archive is a single file bundling the various documents and all relevant information necessary for a modelling and simulation project. The archive is encoded using the Open Modeling EXchange format (OMEX). COMBINE archive metadata provides a harmonised, community-driven approach for annotating a variety of standardised model and data representation formats within a COMBINE archive. BioModels.net qualifiers are standardised relationships (predicates) that specify the relation between an object represented in a description language and the external resource used to annotate it. 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 modelling. 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.

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 [49].

NEW The OMEX Metadata Specification Version 1.2 [50] clarifies the OMEX Metadata Specification 1.0 [51]. Specifically, the main changes introduced in OMEX Metadata Specification Version 1.2 include (1) clarification of authorship and provenance information, (2) clarification of distinction between document- and model-level annotations, and (3) the introduction of an archive-level namespacing convention to ensure reproducible sharing and interpretation of the knowledge graph encoded in the annotations.

Acknowledgements

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] F. Schreiber, G. D. Bader, P. Gleeson, M. Golebiewski, M. Hucka, N. L. 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.

[6] 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.

[7] F. Schreiber, B. Sommer, T. Czauderna et al. Specifications of standards in systems and synthetic biology: status and developments in 2020. *Journal of Integrative Bioinformatics*, 17(2-3), 2020.

[8] 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.

[9] 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.

[10] A. Dräger and D. Waltemath. Overview: Standards for modeling in systems medicine. 2020.

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

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

[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. 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.

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

[16] 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.

[17] 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.

[18] 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.

[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] 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] H. Baig, P. Fontanarossa, V. Kulkarni et al. Synthetic Biology Open Language Visual (SBOL Visual) Version 2.3. *Journal of Integrative Bioinformatics*, 18(3):2020.0045, 2021.

[35] J. Beal, H. Baig, P. Fontanarrosa, J. A. McLaughlin, J. Scott-Brown, P. Vaidyanathan, T. Gorochowski, G. Misirli and C. Myers. Synthetic Biology Open Language Visual (SBOL Visual) Version 3.0. *Journal of Integrative Bioinformatics*, 18(3):2021.0013, 2021.

[36] 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.

[37] 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.

[38] L. Smith, F. Bergmann, A. Garny, T. Helikar, J. Karr, D. Nickerson, H. Sauro and M. König. Simulation experiment description markup language (SED-ML): Level 1 version 4. *Journal of Integrative Bioinformatics*, 18(3):2021.0021, 2021.

[39] 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.

[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.

[49] 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.

[50] J. H. Gennari, M. König, G. Misirli, M. L. Neal and D. P. Nickerson. OMEX Metadata Specification Version 1.2. *Journal of Integrative Bioinformatics*, 18(3):2021.0020, 2021.

[51] 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.