DE GRUYTER Journal of Integrative Bioinformatics. 2018; 20180013

Falk Schreiber[1,2] / Gary D. Bader[3] / Padraig Gleeson[4] / Martin Golebiewski[5] / Michael Hucka[6] / Sarah M. Keating[7] / Nicolas Le Novère[8] / Chris Myers[9] / David Nickerson[10] / Björn Sommer[1,2] / Dagmar Waltemath[11]

Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2017

[1] Department of Computer and Information Science, University of Konstanz, Konstanz, Germany, E-mail: falk.schreiber@uni-konstanz.de
[2] Faculty of IT, Monash University, Clayton, Australia, E-mail: falk.schreiber@uni-konstanz.de
[3] The Donnelly Centre, University of Toronto, Toronto, Canada
[4] Department of Neuroscience, Physiology and Pharmacology, University College London, London, UK
[5] Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany
[6] California Institute of Technology, Pasadena, CA, USA
[7] European Bioinformatics Institute, EMBL-EBI, Cambridge, UK
[8] Babraham Institute, Cambridge, UK
[9] University of Utah, Salt Lake City, UT, USA
[10] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand
[11] University of Rostock, Rostock, Germany

Abstract:
Standards are essential to the advancement of Systems and Synthetic Biology. COMBINE provides a formal body and a centralised platform to help develop and disseminate relevant standards and related resources. The regular special issue of the *Journal of Integrative Bioinformatics* aims to support the exchange, distribution and archiving of these standards by providing unified, easily citable access. This paper provides an overview of existing COMBINE standards and presents developments of the last year.

Keywords: COMBINE, Systems Biology, Synthetic Biology, Standards

DOI: 10.1515/jib-2018-0013

Received: February 8, 2018; **Revised:** February 8, 2018; **Accepted:** February 8, 2018

1 Standards in Systems and Synthetic Biology

Standardisation is a driver of innovation and development. Standards support organisations, people and companies with a basis for mutual understanding and information exchange. Standardisation facilitates communication, measurement, manufacturing and more. In short, standards shape our everyday life and are often driven by urgent needs of science, technology and society. An example is the Baltimore fire in 1904, when more than 1500 buildings were destroyed even though fire departments from Washington, Philadelphia, New York and other towns provided aid – the fire departments which were helping found their hose couplings did not match the city's fire hydrants, which reduced that help significantly. This led afterwards to a national standard for diameter and number of threads per inch for hose couplings and fire hydrants [1].

Standards are essential in Systems and Synthetic Biology. The 'COmputational Modeling in BIology' NEtwork (COMBINE) [2], [3] provides a centralised platform and formal body to help develop and disseminate standards for these fields of research. COMBINE "*…is a network formed by the communities developing standards and formats to share computational models. Working together, it is expected that the federated projects will develop a set of interoperable standards covering all the aspects of computational modelling. Building on the experience of mature projects, which already have stable specifications, software support, user-base and community governance, COMBINE helps foster or support fledging efforts aimed at filling gaps or new needs*" [4]. The benefits and synergy effects of standardisation bodies working together under the umbrella of COMBINE were, for example, shown in our work on the encoding of whole-cell models in SBML and SBGN [5]. This coordination of the development of standards with

---

Falk Schreiber is the corresponding author.

all its benefits and synergies certainly helped for the COMBINE standards to have become some of the most widely used standard formats for modelling in the field as assessed by surveys in the community [6].

The COMBINE initiative published the first collection of Systems and Synthetic Biology standards as a special issue of the *Journal of Integrative Bioinformatics* in 2015 [7]. Since then a regular special issue of JIB serves as both an overview of existing standards as well as an update of the current state of standards in the domain. The latest update was published in 2016 [8]. This special issue contains novel developments of standards and related initiatives in 2017.

2 COMBINE Standards

Standards and associated standardisation initiatives within COMBINE cover a broad range of topics in Systems and Synthetic Biology, see also Figure 1. In the following, the COMBINE standards are presented. Wherever possible we refer to the most current specification published in the series of JIB special issues about COMBINE standards. Some specifications are not yet published; they are currently only available from web sites provided.

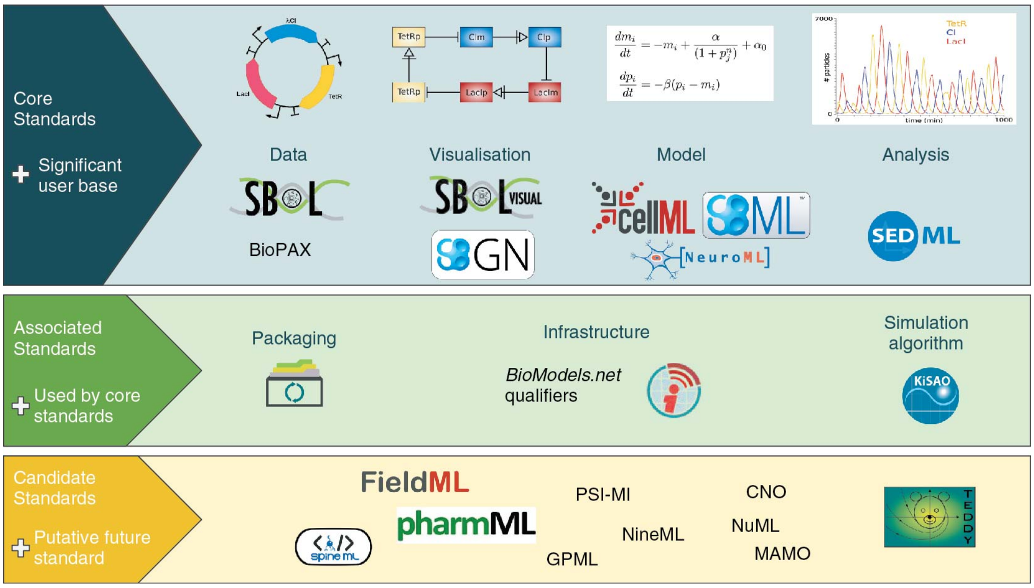

Figure 1: Overview of standards and associated standardisation efforts in Systems and Synthetic Biology covered by COMBINE (from [8]).

For convenience new standard (and standard extension) versions published this year are highlighted with the term *NEW*.

BioPAX (Biological Pathway Exchange) [9] is a format in the Web Ontology Language, to store biological pathways and support the exchange, integration and analysis of biological pathway data. The current specification can be found on the BioPAX webpage [10].

CellML is an XML format to store and exchange reusable, modular computer-based mathematical models [11]. Current specifications are

- CellML 1.1 [12]
- CellML Metadata Framework 2.0 [13]

NeuroML (Neural Open Markup Language) [14], [15] is an XML based language to allow standardisation of model descriptions in computational neuroscience. The latest version of the language (version 2) enables greater extensibility and interoperability with other COMBINE standards, due to the use of the LEMS (Low Entropy Model Specification) language for machine readable model definition. The current specification is

- NeuroML version 2.0 [14]

SBGN (Systems Biology Graphical Notation) [16] is a set of graphical languages used to represent visually processes and networks studied in systems biology. It consists of three orthogonal SBGN languages to represent different aspects of biological systems at different levels of detail. SBGN maps can be exchanged using the format SBGN-ML [17]. Current specifications are

- SBGN Process Description Level 1 Version 1.3 [18]
- SBGN Entity Relationship Level 1 Version 2.0 [19]
- SBGN Activity Flow Level 1 Version 1.2 [20]

SBML (Systems Biology Markup Language) [21] is a format to represent and exchange computational models in systems biology such as models of metabolism, signal transduction and gene regulation. The most recent definition is SBML Level 3, which consists of a base format (Core) and packages that extend the core. Current specifications are

- SBML Level 3 Core, Version 1 Release 2 [22] *NEW*
- SBML Level 3 Core, Version 2 [23] *NEW*
- SBML Level 3 Package: Flux Balance Constraints, Version 2 [24]
- SBML Level 3 Package: Hierarchical Model Composition, Version 1 [25]
- SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1 Release 1 [26] *NEW*
- SBML Level 3 Package: Render, Version 1, Release 1 [27] *NEW*
- SBML Level 3 Package: Qualitative Models, Version 1 [28]
- SBML Level 3 Package: Layout, Version 1 [29]
- SBML Level 3 Package: Groups, Version 1 [30]

SBOL (Synthetic Biology Open Language) [31] is an XML format to exchange data about synthetic biology designs including both structural information, such as hierarchically annotated DNA, RNA, and protein sequences for design components, and behavioural information, such as the interactions between these components. A companion standard is SBOL visual, a set of standard graphical symbols used to represent synthetic biology designs. Current specifications are

- SBOL Version 2.2.0 [32] *NEW*
- SBOL Visual Version 2.0 [33] *NEW*

SED-ML (Simulation Experiment Description Markup Language) [34] is an XML format to describe the procedures to analyse and simulate models, including model identification, pre-processing, simulation setup, post-processing of simulation results and presentation thereof. The current specification is

- SED-ML Level 1 Version 3 [35] *NEW*

A set of associated standards provide an additional layer of semantics to COMBINE representation formats and facilitate their use and interoperability.

COMBINE Archive [36] is a zip-compressed container to support the exchange of all information necessary to reuse and reproduce a simulation study. It includes a manifest file, describing the content of the archive; an optional metadata file, adding a layer of annotation; and the original files, encoded in standard formats. Any other type of file can also be added. An example of a fully featured COMBINE Archive to reproduce a simulation study on syncytial mitotic cycles in Drosophila embryos is provided in [37]. The current specification is

- COMBINE Archive 1.0 [38]

Identifiers.org URIs [39] is a standard set of URIs to reference data and other resources used in the life science community.

Systems Biology Ontology [40] provides controlled vocabularies of terms commonly used in systems biology and computational modelling.

Kinetic Simulation Algorithm Ontology [40] provides support for referring to simulation algorithms and specifying their setups in simulation studies.

BioModels.net qualifiers [41] supports the precise qualification of an annotation used to define the relationship between a model component and an entity or resource used in the annotation of that component.

An overview of the COMBINE core standards can be found in a dedicated collection at the FAIRsharing.org resource [42]. More details on some of the most commonly used COMBINE standards and other formats for modelling can be found in the NormSys registry [43], [44], a freely available online resource that not only lists the standards, but also compares their major features, their possible fields of biological application and use cases (including model examples), as well as their relationships, commonalities and differences.

COMBINE is an open initiative and everybody is invited to join. Two meetings take place each year: the HARMONY (Hackathons on Resources for Modeling in Biology) workshops focus on the development and interoperability of standards and the software that support them; and the COMBINE forums offer presentations, discussions, posters and breakout sessions that focus on standards development and their scientific applications. More information can be found on the COMBINE web site http://co.mbine.org/ or in COMBINE-related publications [2], [3], [45].

Acknowledgement

MG was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) (NormSys, FKZ 01FS14019), and is supported by the German Federal Ministry of Education and Research (BMBF) as part of the Liver Systems Medicine Network (LiSyM, FKZ 031L0056), and by the Klaus Tschira Foundation (KTS). CM is supported by the National Science Foundation under grant nos. 1522074, CCF-1218095, and DBI-1356041. DN is supported by the Aotearoa Foundation. MH and NL are supported by the US National Institute of General Medical Sciences (NIGMS) under grant R01GM070923. NL is also supported by the British Biotechnology and Biological Sciences Research Council under grant nos BB/P013384/1, BB/P013406/1 and BB/P013414/1. DW was supported by the German Federal Ministry of Education and Research (BMBF) as part of the e:Bio program (SEMS, FKZ 0316194). FS was supported by the German Federal Ministry of Education and Research (BMBF) as part of the e:Bio program (SBGN-ED+, FKZ 0316181).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Author Contributions: GB, MG, PG, MH, SK, NL, CM, DN, FS and DW are current or past COMBINE coordinators; BS and FS compiled the special issue. Contact the COMBINE coordinators at combine-coord@googlegroups.com.

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal's publication ethics and publication malpractice statement available at the journal's website and hereby confirm that they comply with all its parts applicable to the present scientific work.

References

[1] Seck MD, Evans DD. Major U.S. cities using national standard fire hydrants, one century after the great Baltimore fire. Gaithersburg: National Institute of Standards and Technology, NISTIR, 7158; 2004. p. 7–9.
[2] Hucka M, Nickerson DP, Bader GD, Bergmann FT, Cooper J, Demir E, et al. Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative. Front Bioeng Biotechnol. 2015;3:19.
[3] Myers C, Bader GD, Gleeson P, et al. A brief history of COMBINE. In: Proceedings of the 2017 Winter Simulation Conference; 2017. p. 884–95.
[4] Combine. 2017. http://co.mbine.org/. Accessed 15 December, 2017.
[5] Waltemath D, Karr JR, Bergmann FT, Chelliah V, Hucka M, Krantz M, et al. Toward community standards and software for whole-cell modeling. IEEE Trans Biomed Eng. 2016;63:2007–14.
[6] Stanford NJ, Wolstencroft K, Golebiewski M, Kania R, Juty N, Tomlinson C. The evolution of standards and data management practices in systems biology. Mol Syst Biol. 2015;11:851.
[7] Schreiber F, Bader GD, M. Golebiewski et al. Specifications of standards in systems and synthetic biology. J Integr Bioinform. 2015;12:258.
[8] Schreiber F, Bader GD, Gleeson P, Golebiewski M, Hucka M, Le Novère N, et al. Specifications of standards in systems and synthetic biology: status and developments in 2016. J Integr Bioinform. 2016;13:289.
[9] Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, et al. The BioPAX community standard for pathway data sharing. Nat Biotechnol. 2010;28:935–42.
[10] BioPax. 2018. http://www.biopax.org/. Accessed 15 December, 2017.
[11] Cuellar AA, Lloyd CM, Nielsen PF, Bullivant D, Nickerson D, Hunter P. An overview of CellML 1.1, a biological model description language. Simulation 2003;79:740–7.
[12] Cuellar AA, Hedley W, Nelson M, Lloyd CM, Halstead MDB, Bullivant DP, et al. The CellML 1.1 specification. J Integr Bioinform. 2015;12:259.
[13] Cooling MT, Hunter PJ. The CellML metadata framework 2.0 specification. J Integr Bioinform. 2015;12:260.

[14] Cannon RC, Gleeson P, Crook S, Ganapathy G, Marin B, Piasini E, et al. LEMS: a language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. Front Neuroinform. 2014;8:79.

[15] Gleeson P, Crook S, Cannon RC, Hines ML, Billings GO, Farinella M, et al. NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. PLoS Comput Biol. 2010;6:e1000815.

[16] Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The systems biology graphical notation. Nat Biotechnol. 2009;27:735–41.

[17] Van Iersel MP, Villéger AC, Czauderna T, Boyd SE, Bergmann FT, Luna A, et al. Software support for SBGN maps: SBGN-ML and LibSBGN. Bioinformatics 2012;28:2016–21.

[18] Moodie SL, Le Novère N, Demir E, Mi H, Villéger A. Systems biology graphical notation: process description language level 1 version 1.3. J Integr Bioinform. 2015;12:263.

[19] Sorokin AA, Le Novère N, Luna A, Czauderna T, Demir E, Haw R, et al. Systems biology graphical notation: entity relationship language level 1 version 2. J Integr Bioinform. 2015;12:264.

[20] Mi H, Schreiber F, Moodie SL, Czauderna T, Demir E, Haw R, et al. Systems biology graphical notation: activity flow language level 1 version 1.2. J Integr Bioinform. 2015;12:265.

[21] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19(4):524–531.

[22] Hucka M, Bergmann FT, Dräger A, Hoops S, Keating SM, Le Novère N, et al. The systems biology markup language (SBML): language specification for level 3 version 1 release 2 core. J Integr Bioinform. 2018;15:20170080.

[23] Hucka M, Bergmann FT, Dräger A, Hoops S, Keating SM, Le Novère N, et al. The systems biology markup language (SBML): language specification for level 3 version 2 core. J Integr Bioinform. 2018;15:20170081.

[24] Olivier BG, Bergmann FT. SBML level 3 package: flux balance constraints, version 2. J Integr Bioinform. 2018;15:20170082.

[25] Smith LP, Hucka M, Hoops S, Finney A, Ginkel M, Myers CJ, et al. SBML level 3 package: hierarchical model composition, version 1 release 3. J Integr Bioinform. 2015;12:268.

[26] Zhang F, Meier-Schellersheim M. SBML level 3 package: multistate, multicomponent and multicompartment species, version 1, release 1. J Integr Bioinform. 2018;15:20170077.

[27] Bergmann FT, Keating SM, Gauges R, Sahle S, Wengler K. SBML level 3 package: render, version 1, release 1. J Integr Bioinform. 2018;15:20170078.

[28] Chaouiya C, Keating SM, Bérenguier D, Naldi A, Thieffry D, van Iersel MP, et al. The systems biology markup language (SBML) level 3 package: qualitative models, version 1, release 1. J Integr Bioinform. 2015;12:270.

[29] Gauges R, Rost U, Sahle S, Wengler K, Bergmann FT. The systems biology markup language (SBML) level 3 package: layout, version 1 core. J Integr Bioinform. 2015;12:267.

[30] Hucka M, Smith LP. The systems biology markup language (SBML) level 3 package: groups, version 1 release 1. J Integr Bioinform. 2016;13:290.

[31] Galdzicki M, Clancy KP, Oberortner E, Pocock M, Quinn JY, Rodriguez CA, et al. The synthetic biology open language (SBOL) provides a community standard for communicating designs in synthetic biology. Nat Biotechnol. 2014;32:545–50.

[32] Cox RS, Madsen C, McLaughlin JA, Nguyen T, Roehner N, Bartley B, et al. Synthetic biology open language (SBOL) version 2.2.0. J Integr Bioinform. 2018;15:20170074.

[33] Cox RS, Madsen C, McLaughlin J, Nguyen T, Roehner N, Bartley B, et al. Synthetic biology open language visual (SBOL visual) version 2.0. J Integr Bioinform. 2018;15:20170075.

[34] Waltemath D, Adams R, Bergmann FT, Hucka M, Kolpakov F, Miller AK, et al. Reproducible computational biology experiments with SED-ML – the simulation experiment description markup language. BMC Syst Biol. 2011;5:198.

[35] Bergmann F, Cooper J, König M, Moraru I, Nickerson D, Novère NL, et al. Simulation experiment description markup language (SED-ML) level 1 version 3 (L1V3). J Integr Bioinform. 2018;15:20170086.

[36] Bergmann FT, Adams R, Moodie S, Cooper J, Glont M, Golebiewski M, et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. BMC Bioinform. 2014;15:369.

[37] Scharm M, Waltemath D. A fully featured COMBINE archive of a simulation study on syncytial mitotic cycles in drosophila embryos. F1000Research. 2016;5; 2421. doi: 10.12688/f1000research.9379.1.

[38] Bergmann FT, Rodriguez N, Le Novère N. COMBINE archive specification version 1. J Integr Bioinform. 2015;12:261.

[39] Juty N, Le Novère N, Laibe C. Identifiers. org and miriam registry: community resources to provide persistent identification. Nucleic Acids Res. 2012;40:D580–6.

[40] Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, et al. Controlled vocabularies and semantics in systems biology. Mol Syst Biol. 2011;7:543.

[41] BioModels.net, 2018. http://co.mbine.org/standards/qualifiers.

[42] COMBINE FAIRsharing, 2018. https://fairsharing.org/collection/ComputationalModelingCOMBINE.

[43] Nickerson D, Atalag K, de Bono B, Geiger J, Goble C, Hollmann S, et al. The human physiome: how standards, software and innovative service infrastructures are providing the building blocks to make it achievable. Interface Focus. 2016;6:20150103.

[44] NormSys registry. 2018. http://normsys.h-its.org. Accessed 15 December, 2017.

[45] Waltemath D, Bergmann FT, Chaouiya C, Czauderna T, Gleeson P, Goble C, et al. Meeting report from the fourth meeting of the computational modeling in biology network (COMBINE). Stand Genomic Sci. 2014;9:1285–301.