Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2016

Falk Schreiber[1*], Gary D. Bader[2], Padraig Gleeson[3], Martin Golebiewski[4], Michael Hucka[5], Nicolas Le Novère[6], Chris Myers[7], David Nickerson[8], Björn Sommer[1] and Dagmar Walthemath[9]

[1]Dept. of Computer and Information Science, University of Konstanz, Germany & Faculty of IT, Monash University, Clayton, Australia

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

[3]University College London, UK

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

[5]California Institute of Technology, USA

[6]Babraham Institute, UK

[7]Dept. of Electrical and Computer Engineering, University of Utah, USA

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

[9]Dept. of Systems Biology and Bioinformatics, University of Rostock, Germany

Summary

Standards are essential to the advancement of science and technology. In systems and synthetic biology, numerous standards and associated tools have been developed over the last 16 years. This special issue of the *Journal of Integrative Bioinformatics* aims to support the exchange, distribution and archiving of these standards, as well as to provide centralised and easily citable access to them.

Norms and requirements for objects, methods, processes and practices shape our everyday lives. Often they are represented in formal documents each describing a standard for a particular domain. The use of standards offers many benefits, such as improving the precision and efficiency of information exchange, reducing the cost of production, allowing parts from different producers to be combined or interchanged, and much more. Standards are not static, but develop and evolve with the progress of science and technology.

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

When new, innovative areas (such as systems and synthetic biology) first emerge, they often either lack proper standards altogether, or the documents describing the standards in use are not available from a centralised location. To help develop and disseminate standards for systems and synthetic biology, the COMBINE (the 'COmputational Modeling in BIology' NEtwork) initiative was formed in 2010 [1]. This initiative "...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" [2].

An important component of COMBINE's mission is outreach, particularly two yearly meetings: (1) HARMONY (Hackathons on Resources for Modeling in Biology) workshops, which are focused on the development and interoperability of standards and the software that support them; and (2) COMBINE forums, which feature presentation, discussion, poster and breakout sessions focused on standards development and their scientific applications. For more information, please see `http://co.mbine.org/`.

When the first systems and synthetic biology standards began to be developed in the late 1990's, the definitions of the standards were published and made accessible in diverse ways, such as the standards' web pages or preprint services. However, there was no centralised, properly citable platform available. Last year, the COMBINE initiative introduced the first collection of up-to-date versions of systems and synthetic biology standards as a special issue of the *Journal of Integrative Bioinformatics* (JIB) to provide a single, easily accessible and citable platform for the publication of these standards [3]. The current special issue is intended to serve as both an update as well as an overview of the current state of standards in the domain.

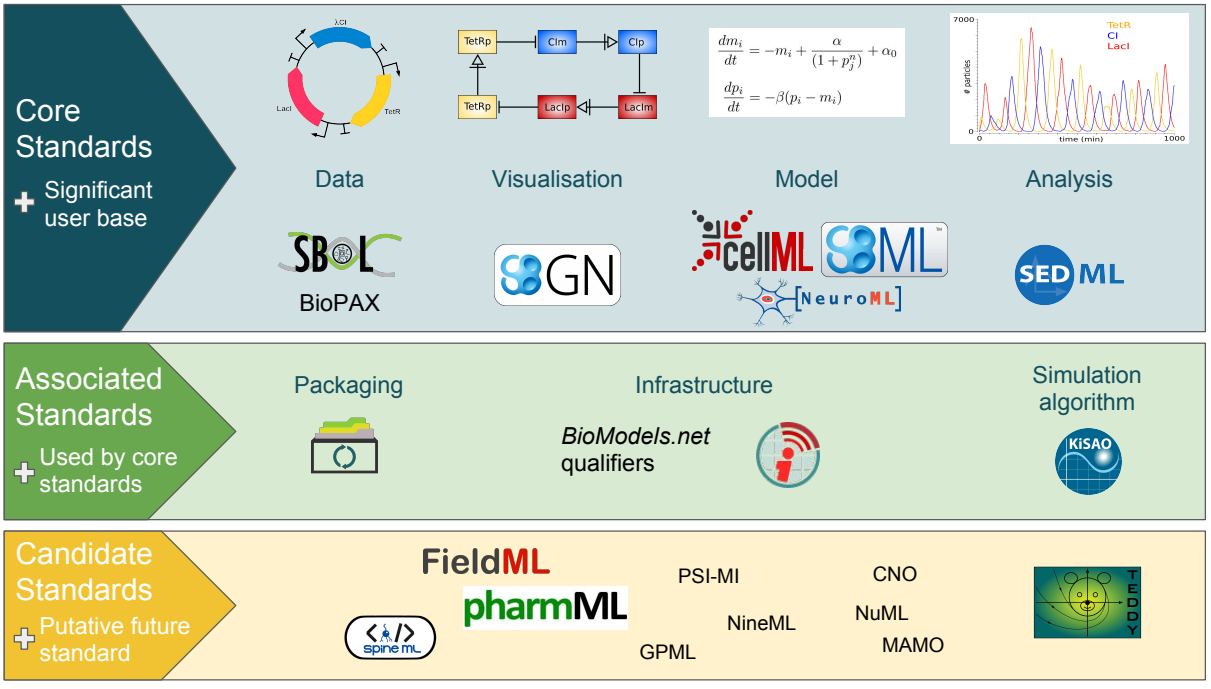

Figure 1: COMBINE standards and associated standardisation efforts.

COMBINE standards and their associated standardisation efforts cover a broad range of topics in systems and synthetic biology. We distinguish core standards (i. e., those with a specification implemented in several software tools, and with a broad user base); associated standards (i. e., those employed by the core standards or serving support roles); and candidate standards (i. e., efforts that have indicated their interest in becoming COMBINE standards in the future). All that are connected to COMBINE in 2016 are shown in Figure 1.

In the following list, we summarize the core COMBINE standards. Where possible, we refer to the most recent specification published in the series of JIB special issues (2015 and 2016). A few standards' specifications are currently only available from their websites, in which case, we refer to that resource instead.

BioPAX (Biological Pathway Exchange), to store biological pathways and support the exchange, integration and analysis of biological pathway data [4]. The current specification can be found on the BioPAX web page [5].

SBGN (Systems Biology Graphical Notation), for the graphical depiction of processes and networks studied in systems biology [6]. Three SBGN sublanguages can represent different aspects of phenomena at different levels of detail:

- SBGN Process Description 1.1.3 [7]
- SBGN Entity Relationship 1.2.0 [8]
- SBGN Activity Flow 1.1.2 [9]

SBML (Systems Biology Markup Language), to represent computational models in systems biology such as models of metabolism, signal transduction and gene regulation [10]. The most recent definition is SBML Level 3, which consists of a base format (*Core*) and *packages* that extend the core:

- SBML Level 3 Core, Version 1 [11]
- SBML Level 3 Package: Flux Balance Constraints, Version 2 [12]
- SBML Level 3 Package: Hierarchical Model Composition, Version 1 [13]
- SBML Level 3 Package: Qualitative Models, Version 1 [14]
- SBML Level 3 Package: Layout, Version 1 [15]
- SBML Level 3 Package: Groups, Version 1 [16]

SED-ML (Simulation Experiment Description Markup Language), to describe the procedures to analyse and simulate models, including model identification, pre-processing, simulation setup, post-processing of simulation result and presentation thereof [17]. The current specification is SED-ML Level 1 Version 2 [18].

CellML, to store and exchange reusable, modular computer-based mathematical models [19]. In addition to the CellML standard, CellML also defines a metadata framework:

- CellML 1.1 [20]
- CellML Metadata Framework 2.0 [21]

SBOL (Synthetic Biology Open Language), to exchange knowledge about synthetic biology designs. SBOL covers 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 [22]. The current specification is SBOL 2.1.0 [23].

NeuroML, a model description language for computational neuroscience [24]. The current specification is NeuroML version 2, beta 4, and can be found on the NeuroML website [25].

In addition to the main COMBINE standards, COMBINE coordinates the development of associated efforts, that are not model or data representation formats *per se*. In some cases, they involve additional software formats that help software tools work with models and data; in other cases, the resources provide an additional layer of semantics that facilitate the use and interoperability of COMBINE standards or otherwise enhance their usefulness. Wherever possible, we cite the most current specification published in the series of JIB special issues (2015 and 2016) on COMBINE standards. A few associated efforts are currently only available from their website, in which case, we refer to that resource.

COMBINE Archive, to support the exchange of information necessary for a modelling and simulation experiment in biology [26]. It is a zip-compressed container that includes a manifest file, an optional metadata file, and all the files necessary to describe the models and their processing. The current specification is COMBINE Archive 1.0 [27].

Identifiers.org URIs, to provide unique and permanent URIs for reference data and other resources used in the scientific community [28].

Systems Biology Ontology, to provide controlled vocabularies of terms commonly used in systems biology, and in particular in computational modelling [29].

Kinetic Simulation Algorithm Ontology, to provide support for referring to specific simulation algorithms in descriptions of simulation experiments [29].

BioModels.net qualifiers, to support 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. See BioModels.net [30] for details.

In conclusion, standards are increasingly important in systems and synthetic biology. Many software tools and databases support them, and many are endorsed by journals. We hope that providing a centralised and easily citable access to them will help to increase the adoption and use of standards in systems and synthetic biology, and thereby support the exchange, distribution and archiving of models.

GB, MG, MH, PG, NLN, CM, DN, FS and DW are COMBINE coordinators; BS and FS compiled the special issue.
Contact the COMBINE coordinators via email at combine-coord@googlegroups.com.

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.

[2] http://co.mbine.org/, 2016.

[3] 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.

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

[5] http://www.biopax.org/, 2016.

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

[7] S. L. Moodie, N. L. Novère, E. Demir, H. Mi and A. Villéger. Systems Biology Graphical Notation: Process Description language Level 1 Version 1.3. *Journal of Integrative Bioinformatics*, 12(2):263, 2015.

[8] A. A. Sorokin, N. L. 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.

[9] H. Mi, F. Schreiber, S. L. Moodie, T. Czauderna, E. Demir, R. Haw, A. Luna, N. L. 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.

[10] 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.

[11] M. Hucka, F. T. Bergmann, S. Hoops, S. M. Keating, S. Sahle, J. C. Schaff, L. P. Smith and D. J. Wilkinson. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *Journal of Integrative Bioinformatics*, 12(2):266, 2015.

[12] B. G. Olivier and F. T. Bergmann. The Systems Biology Markup Language (SBML) Level 3 Package: Flux Balance Constraints. *Journal of Integrative Bioinformatics*, 12(2):269, 2015.

[13] L. P. Smith, M. Hucka, S. Hoops, A. Finney, M. Ginkel, C. J. Myers, I. I. Moraru and W. Liebermeister. The Systems Biology Markup Language (SBML) Level 3 Package: Hierarchical Model Composition, Version 1 Release 3. *Journal of Integrative Bioinformatics*, 12(2):268, 2015.

[14] C. Chaouiya, S. M. Keating, D. Bérenguier, A. Naldi, D. Thieffry, M. P. van Iersel, N. L. 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.

[15] 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.

[16] 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.

[17] 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.

[18] F. T. Bergmann, J. Cooper, N. L. Novère, D. P. Nickerson and D. Waltemath. Simulation Experiment Description Markup Language (SED-ML) Level 1 Version 2. *Journal of Integrative Bioinformatics*, 12(2):262, 2015.

[19] A. A. Cuellar, C. M. Lloyd, P. F. Nielsen et al. An overview of CellML 1.1, a biological model description language. *Simulation*, 79(12):740–747, 2003.

[20] 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.

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

[22] N. Roehner, J. Beal, K. Clancy et al. Sharing structure and function in biological design with SBOL 2.0. *ACS Synthetic Biology*, 5(6):498–506, 2016.

[23] J. Beal, R. S. C. III, R. Grunberg et al. Synthetic Biology Open Language (SBOL) Version 2.1.0. *Journal of Integrative Bioinformatics*, 13(3):291, 2016.

[24] 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.

[25] http://www.neuroml.org/, 2016.

[26] 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.

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

[28] 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.

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

[30] http://co.mbine.org/standards/qualifiers, 2016.