*On behalf of the COMBINE Coordinators*

Michael Hucka, Ph.D.
Beckman Institute 139-74
California Institute of Technology
Pasadena, CA 91125

December 5, 2014

National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

**Notice Number: NOT-ES-15-002**

Dear NIH,

We are the coordinators of COMBINE (Computational Modeling in Biology Network), an initiative to coordinate the development of many popular standards and formats for computational modeling in biology. We are grateful for the opportunity that NIH is providing to respond to the *Request for Information (RFI) Making Data Usable – A Framework for Community-Based Data and Metadata Standards Efforts for NIH-relevant Research.* We would like to take this opportunity to describe the experiences of the standardization efforts under the COMBINE umbrella.

### *The history and goals of COMBINE*

COMBINE was formed in 2009 by the groups involved in developing file formats and other standards in systems biology, including SBML [11, 12], SBGN [14], BioPAX [6], CellML [4, 9], SED-ML [17], SBOL [7], NeuroML [8], and others. The impetus was the realization that many individuals were involved in multiple standardization efforts, traveling to separate international workshops year after year and performing many of the same organizational tasks multiple times for each standards community. Eventually, two "super meetings" were held involving many of the groups, and slowly we realized that not only could there be cost savings in co-locating meetings: the various efforts could also benefit from common infrastructure, operating procedures, and potentially a common voice to seek support. The Le Novère group (then at the EBML European Bioinformatics Institute near Cambridge, UK) undertook the creation and maintenance of a home website for COMBINE at

<center>http://co.mbine.org</center>

The primary goal of COMBINE is to coordinate the development and other activities of the various community standards used in the area of computational modeling. By doing so, we hope that the federated projects will develop standards that are more interoperable and less overlapping than if the efforts proceeded separately. COMBINE offers a format specification infrastructure, announcement lists, and more, as discussed below.

An important point about COMBINE is that it *does not dictate what individual standardization efforts should do.* Actions are entirely up to the leaders and members of the communities involved in the individual efforts. COMBINE does offer examples of what has worked in terms of community organization approaches, as well as some common infrastructure for such things as cataloguing standards specifications, but the degree of participation is up to the groups behind the efforts.

### *Groups involved in COMBINE today*

COMBINE today includes the efforts listed in the following table. All are open community efforts, with freely available specifications, open community participation, etc. They cover a range of topics: raw data standards, model format standards, graphical notation standards, ontologies, and minimum information guidelines.

| Category | Name | COMBINE page or other reference |
|---|---|---|
| COMBINE representation standards | BioPAX (*Biological Pathways Exchange*) | http://co.mbine.org/standards/biopax |
| | CellML | http://co.mbine.org/standards/cellml |
| | SBGN (*Systems Biology Graphical Notation*) | http://co.mbine.org/standards/sbgn |
| | SBML (*Systems Biology Markup Language*) | http://co.mbine.org/standards/sbml |
| | SBOL (*Synthetic Biology Open Language*) | http://co.mbine.org/standards/sbol |
| | SED-ML (*Simulation Experiment Description Markup Language*) | http://co.mbine.org/standards/sed-ml |
| Associated standardization efforts | BioModels.net Qualifiers | http://co.mbine.org/standards/qualifiers |
| | COMBINE Archive | http://co.mbine.org/standards/omex |
| | MIASE (*Minimum Information About a Simulation Experiment*) | http://co.mbine.org/standards/miase |
| | MIRIAM (*Minimal Information Required In the Annotation of Models*) | http://co.mbine.org/standards/miriam |
| | KiSAO (*Kinetic Simulation Algorithm Ontology*) | http://co.mbine.org/standards/kisao |
| Related standardization efforts | BioSharing | [16] |
| | CNO (*Computational Neuroscience Ontology*) | [13] |
| | FieldML (*Field Markup Language*) | [2] |
| | GPML (*GenMAPP Pathway Markup Language*) | [1] |
| | MAMO (*Mathematical Modeling Ontology*) | [18] |
| | NeuroML | [8] |
| | NuML (*Numerical Markup Language*) | [5] |
| | PSI-MI (*Proteomics Standards Initiative*) | [10] |
| | SpineML (*Spiking Neural Markup Language*) | [15] |
| | TEDDY (*TErminology for the Description of DYnamics*) | [3] |

The differences in the categories are as follows:

- The *COMBINE representation standards* meet a number of basic criteria which include the following: (i) represent information in biology, (ii) possess democratically-elected editorial boards, (iii) possess full specifications of version 1.0 or higher, (iv) have API library implementations supporting the standard, and (v) have continued development supported by a unified group of identifiable people.

- The *Associated standardization efforts* are either in a more fledgling state of development, or are not standard formats per se but rather tools or services that facilitate the use or interoperability of the COMBINE representation standards.

- The *Related standardization efforts* are other efforts that are either candidate COMBINE standards in early stages of development, or else are mature efforts in their own right that have their own substantial communities and, while not part of COMBINE, are efforts that the COMBINE community is involved in.

A standardization community that wants to become part of COMBINE begins as a *Related standardization effort*. The developers of the standard should join the COMBINE announcement mailing lists and, especially, participate in the annual COMBINE meetings discussed below. If the effort meets the basic criteria to be considered a full-fledged standard, the *COMBINE Coordinators* will accept the effort as a COMBINE standard.

# References

[1] GPML pathway format. http://developers.pathvisio.org/wiki/EverythingGpml, 2014.

[2] G. Richard Christie, Poul M. F. Nielsen, Shane A. Blackett, Chris P. Bradley, and Peter J. Hunter. FieldML: concepts and implementation. *Philosophical Transactions of the Royal Society of London A*, 367(1895):1869–1884, 2009.

[3] Mélanie Courtot, Nick Juty, Christian Knüpfer, Dagmar Waltemath, Anna Zhukova, Andreas Dräger, Michel Dumontier, Andrew Finney, Martin Golebiewski, Janna Hastings, Stefan Hoops, Sarah Keating, Douglas B. Kell, Samuel Kerrien, James Lawson, Allyson Lister, James Lu, Rainer Machne, Pedro Mendes, Matthew Pocock, Nicolas Rodriguez, Alice Villéger, Darren J. Wilkinson, Sarala Wimalaratne, Camille Laibe, Michael Hucka, and Nicolas Le Novère. Controlled vocabularies and semantics in systems biology. *Molecular Systems Biology*, 7(1):543, September 2011.

[4] Autumn A Cuellar, Catherine M Lloyd, Poul F Nielsen, David P Bullivant, David P Nickerson, and Peter J Hunter. An overview of CellML 1.1, a biological model description language. *Simulation*, 79(12):740–747, 2003.

[5] Joseph O. Dada, Irena. Spasic, Norman W. Paton, and Pedro Mendes. SBRML: a markup language for associating systems biology data with models. *Bioinformatics*, 26(7):932–938, 2010.

[6] Emek Demir, Michael P. Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming N. Wu, Peter D'Eustachio, Carl Schaefer, Joanne Luciano, Frank Schacherer, Irma Martinez-Flores, Zhenzjun J. Hu, Veronica Jimenez-Jacinto, Geeta Joshi-Tope, Kumaran Kandasamy, Alejandra C. Lopez-Fuentes, Huaiyu Y. Mi, Elgar Pichler, Igor Rodchenkov, Andrea Splendiani, Sasha Tkachev, Jeremy Zucker, Gopal Gopinath, Harsha Rajasimha, Ranjani Ramakrishnan, Imran Shah, Mustafa Syed, Nadia Anwar, Özgün O. Babur, Michael Blinov, Erik Brauner, Dan Corwin, Sylva Donaldson, Frank Gibbons, Robert Goldberg, Peter Hornbeck, Augustin Luna, Peter Murray-Rust, Eric Neumann, Oliver Reubenacker, Matthias Samwald, Martijn van Iersel, Sarala Wimalaratne, Keith Allen, Burk Braun, Michelle Whirl-Carrillo, Kei-Hoi H. Cheung, Kam Dahlquist, Andrew Finney, Marc Gillespie, Elizabeth Glass, Li Gong, Robin Haw, Michael Honig, Olivier Hubaut, David Kane, Shiva Krupa, Martina Kutmon, Julie Leonard, Debbie Marks, David Merberg, Victoria Petri, Alex Pico, Dean Ravenscroft, Liya Y. Ren, Nigam Shah, Margot Sunshine, Rebecca Tang, Ryan Whaley, Stan Letovksy, Kenneth H. Buetow, Andrey Rzhetsky, Vincent Schachter, Bruno S. Sobral, Ugur Dogrusoz, Shannon McWeeney, Mirit Aladjem, Ewan Birney, Julio Collado-Vides, Susumu Goto, Michael Hucka, Nicolas Le Novère, Natalia Maltsev, Akhilesh Pandey, Paul Thomas, Edgar Wingender, Peter D. Karp, Chris Sander, and Gary D. Bader. The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28(9):935–942, 2010.

[7] Michal Galdzicki, Kevin P. Clancy, Ernst Oberortner, Matthew Pocock, Jacqueline Y. Quinn, Cesar A. Rodriguez, Nicholas Roehner, Mandy L. Wilson, Laura Adam, J. Christopher Anderson, Bryan A. Bartley, Jacob Beal, Deepak Chandran, Joanna Chen, Douglas Densmore, Drew Endy, Raik Grünberg, Jennifer Hallinan, Nathan J. Hillson, Jeffrey D. Johnson, Allan Kuchinsky, Matthew Lux, Goksel Misirli, Jean Peccoud, Hector A. Plahar, Evren Sirin, B. Stan, Guy-Bart, Alan Villalobos, Anil Wipat, John H. Gennari, Chris J. Myers, and Herbert M. Sauro. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology*, 32(6):545, 2014.

[8] Padraig Gleeson, Sharon Crook, Robert C. Cannon, Michael L. Hines, Guy O. Billings, Matteo Farinella, Thomas M. Morse, Andrew P. Davison, Subhasis Ray, Upinder S. Bhalla, Simon R. Barnes, Yoana D. Dimitrova, and R. Angus Silver. 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.

[9] Warren J. Hedley, Melanie R. Nelson, D. P. Bullivant, and Poul F. Nielson. A short introduction to CellML. *Philosophical Transactions of the Royal Society of London A*, 359:1073–1089, 2001.

[10] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, R. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, YX X. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, SGN G. N. Grant, C. Sander, P. Bork, WM M. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, L. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, and R. Apweiler. The HUPO PSI's Molecular Interaction format—a community standard for the representation of protein interaction data. *Nature Biotechnology*, 22(2):177–183, 2004.

[11] Michael Hucka, Frank T. Bergmann, Stephan Hoops, Sarah M. Keating, Sven Sahle, James C. Schaff, and Lucian P. Smith. The Systems Biology Markup Language (SBML): Language specification for Level 3 Version 1 Core. Available via the World Wide Web at http://sbml.org/Documents/Specifications, 2010.

[12] Michael Hucka, Andrew Finney, Herbert M. Sauro, Hamid Bolouri, John C. Doyle, Hiroaki Kitano, Adam P. Arkin, Benjamin J. Bornstein, Dennis Bray, Athel Cornish-Bowden, Autumn A. Cuellar, Sergey Dronov, Ernst-Dieter Gilles, Martin Ginkel, Victoria Gor, Igor I. Goryanin, Warren J. Hedley, T. Charles Hodgman, Jan-Hendrik. Hofmeyr, Peter J. Hunter, Nick S. Juty, J. L. Kasberger, Andreas Kremling, Ursula Kummer, Nicolas Le Novère, Leslie M. Loew, Daniel Lucio, Pedro Mendes, Eric D. Mjolsness, Y. Nakayama, Melanie R. Nelson, Poul F. Nielsen, Takeshi Sakurada, James C. Schaff, Bruce E. Shapiro, Thomas S. Shimizu, Hugh D. Spence, Jörg Stelling, Koichi Takahashi, Masaru Tomita, John Wagner, and J. Wang. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[13] Yann Le Franc, Andrew P. Davison, Padraig Gleeson, Fahim T. Imam, Birgit Kriener, Stephen D. Larson, Subhasis Ray, Lars Schwabe, Sean Hill, and Erik De Schutter. Computational Neuroscience Ontology: a new tool to provide semantic meaning to your models. *BMC Neuroscience*, 13(Suppl 1):P149, 2012.

[14] Nicolas Le Novère, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, Sarala M Wimalaratne, et al. The Systems Biology Graphical Notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[15] Paul Richmond, Alex Cope, Kevin Gurney, and David J Allerton. From model specification to simulation of biologically constrained networks of spiking neurons. *Neuroinformatics*, 12(2):307–323, 2014.

[16] Susanna-Assunta Sansone, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, Steffen Neumann, Weida Tong, Linda Amaral-Zettler, et al. Toward interoperable bioscience data. *Nature Genetics*, 44(2):121–126, 2012.

[17] Dagmar Waltemath, Richard Adams, Frank T Bergmann, Michael Hucka, Fedor Kolpakov, Andrew K Miller, Ion I Moraru, David Nickerson, Sven Sahle, Jacky L Snoep, et al. Reproducible computational biology experiments with SED-ML—the Simulation Experiment Description Markup Language. *BMC Systems Biology*, 5(1):198, 2011.

[18] Anna Zhukova, Dagmar Waltemath, Maciej J. Swat, Jon Olav Vik, Nicolas Rodriguez, and Nicolas Le Novère. Mathematical Modelling Ontology (MAMO). https://sourceforge.net/p/mamo-ontology/wiki/Home/, 2014.