

Institution: University of Hertfordshire
Unit of Assessment: Panel B (11): Computer Science and Informatics
Title of case study: SBML, the Systems Biology Markup Language
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>Research into the operational characteristics and applicability of biological reaction networks, carried out at the university in collaboration with groups at Caltech and Sony Systems, revealed the pressing need for a standard format that could be used for storage and exchange of mathematical models of such systems. Hertfordshire researchers played a crucial role in the initial design, dissemination and early exploitation of the Systems Biology Markup Language, SBML, now recognised as the de facto standard format for this purpose. Several major scientific publishers operating across academic boundaries require their authors to use SBML, and 254 software tools, including MATLAB and Mathematica, are now SBML-compliant. Online forums testify to a sizeable, international user–developer community that encompasses engineers, biologists, mathematicians and software developers.</p>
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>The university's Biocomputation Group led by Hamid Bolouri, Professor of Neural Systems (employed at the university 1998–2002), played a crucial role in the initial design, dissemination, and early exploitation of the Systems Biology Markup Language, or SBML. SBML is now considered the standard format for storage and exchange of biological reaction network models, and has been adopted across the Systems Biology community in academia and the pharmaceutical industry.</p> <p>During the 1990s, work in Bolouri's group had concentrated on the implementation of artificial neural networks in hardware and software. However, after studying the biology of the brain and embryonic development, Bolouri began to investigate whether the control principles that underlie biological development could be applied to electronic and other man-made computational systems. It was clear from the outset that there was a plethora of software packages in existence that would be extremely helpful in this research but, unfortunately, most of these packages were stand-alone and lacked any kind of interoperability or standardisation. Initially, therefore, it seemed more efficient to develop the required modelling and simulation software in house from scratch, and in 2000 the group embarked on an ambitious project to do so.</p> <p>At the same time, Bolouri continued to investigate how existing resources could be made to work together. With John Doyle, Professor of Control and Dynamical Systems at Caltech, Bolouri approached Hiroaki Kitano of the Sony Computer Science Laboratories in Tokyo. Kitano, who had led the development of AIBO, the robot dog, had also become fascinated by the 'ingenuity' of biological systems, and was actively looking for partners in his newly established Kitano Symbiotic Systems Project. As the plans put forward by Bolouri and Doyle appeared to fit remarkably well in his research strategy, the threesome began lobbying potential stakeholders, to set up a project to promote interoperability of computational software for Systems Biology.</p> <p>A core team of software engineers and computational biologists based at Hertfordshire and at Caltech set out under Bolouri's guidance to create a 'Systems Biology Workbench' or SBW, a software framework that would allow individual tools to 'plug in' via a common interface and exchange models and data using a standard messaging protocol. It soon became clear that the workbench would need a common format for model representation to enable the intended</p>

interoperability of software tools, and the growing community of stakeholders in the project decided that such a format should be XML-based. Thus, the development of the Systems Biology Markup Language, or SBML, was initiated, and the first specification of SBML was issued by the core team early in 2001, rapidly followed by improved and more wide-ranging versions. At the University of Hertfordshire, a repository of biochemical network models expressed in SBML was set up, and the first converter that facilitated model exchange with another repository of biological process models was developed. The repository evolved into the BioModels Database, an altogether much larger operation, now hosted at EMBL-EBI.

3. References to the research (indicative maximum of six references)

Publications

- Authors affiliated with the university at the time of publication are indicated by bold type
- The top three publications are indicated by **

1. Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., **Rust, A.G.**, Pan, Z., **Schilstra, M.J.**, **Clarke, P.J.C.**, Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., **Bolouri, H.** (2002). 'A genomic regulatory network for development', *Science* 295, 1669–1678. doi: 10.1126/science.1069883
2. Hucka M, Finney A, Sauro H, **Bolouri H**, Doyle J and Kitano H (2002). 'The ERATO Systems Biology Workbench: Enabling interaction and exchange between software tools for computational biology', *Proc. Pacific Symposium on Biocomputing*, 7, 450–461. <<http://psb.stanford.edu/psb-online/proceedings/psb02/hucka.pdf>>
3. ** Hucka M, **Finney A**, Sauro HM, **Bolouri H** et al. (2003). 'The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models', *Bioinformatics* 19 (4), 524–531. doi: 10.1093/bioinformatics/btg015
4. ** Hucka M, **Finney A**, Bornstein B, **Keating SM**, Shapiro BE, **Matthews J**, Kovitz B, **Schilstra MJ**, Funahashi A, Doyle JC and Kitano H (2004). 'Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) project', *Systems Biology* 1 (1), 41–53. doi: 10.1049/sb:20045008
5. **Schilstra MJ**, Li L, **Matthews J**, **Finney A**, Hucka M and Le Novère N (2006). 'CellML2SBML: Conversion of CellML into SBML', *Bioinformatics* 22 (8), 1018–1020. doi: 10.1093/bioinformatics/btl047
6. ** Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, **Schilstra MJ**, Shapiro B, Snoep JL and Hucka M (2006). 'BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems', *Nucleic Acids Research* 34 (Supp 1), D689–691. doi: 10.1093/nar/gkj092

Selected Funding

2002–4 (2 years), Biotechnology and Biological Sciences Research Council, awarded to Hamid Bolouri (PI), Maria Schilstra and Roderick Adams. Project Title: Model Sharing & Co-Simulation Standards for System. **Total award to University of Hertfordshire: £171,032.**

2002 (1 year), California Institute of Technology, awarded to Hamid Bolouri (PI), Roderick Adams. Project Title: Development of Systems Biology Markup Language. **Total award to University of Hertfordshire: £77,768.**

4. Details of the impact (indicative maximum 750 words)

Since the 1990s, it has become clear that quantitative systems analysis is a crucial step in predicting the effect of drugs and other interventions on the human body. 'Systems Biology' departments exist in most biomedical research institutions; funding agencies have directed significant resources towards the development of computational tools; and old-fashioned quantitative disciplines such as enzyme kinetics, sidelined during the molecular genetics revolution of the 1970s and 80s, have been re-energised.

Quantitative analysis of the responses of complex biological systems requires compound mathematical models and integrated computational approaches. Experimental data and small-scale computational models may be available for small parts of such systems, but the responses of the whole will undoubtedly be different from the sum of the responses of its parts. Recognising the need for integration of computational approaches and mathematical modelling, Hamid Bolouri (University of Hertfordshire), Hiroaki Kitano (Sony), and John Doyle (Caltech) were the prime movers from 2000 onwards in creating, disseminating and promoting SBML, which has become the de facto standard format for storing and exchanging biological reaction network models. Its success can be measured by the fact that SBML has outgrown its original base and continues to be used and developed into the 2008–13 period.

In 2000, Bolouri instigated a series of workshops on Software Platforms for Systems Biology. Around half a dozen modelling and simulation tools developers attended the first meeting, with participant numbers growing rapidly and a computational modelling 'community' becoming established. The consortium's original mission was integrating simulation analysis tools through a 'workbench', a software interoperability framework, but it transpired that the most viable part of the project was the Systems Biology Markup Language or SBML, the XML-based common format for model representation designed to enable model exchange. The consortium therefore concentrated on shaping, expanding and promoting SBML.

A number of application developers were recruited, mainly to the University of Hertfordshire and to Caltech, their task to supply utilities such as editors, converters, libraries and APIs, example simulators, and a model repository which encouraged developers of client software to implement compliance with the language and help end users appreciate SBML's possibilities.

The workshops, renamed SBML Forums in 2002, were initially held twice yearly, but in 2004 they differentiated into one SBML Forum and one SBML 'Hackathon' annually. After becoming satellites of the annual International Conference of Systems Biology, attendance increased dramatically. The Hackathon's web pages (see section 5, ref. 5) state that the event's focus is the 'development of the standards, interoperability and infrastructure', offering a space where 'time is devoted to allowing hands-on hacking and interaction between people focused on practical development of software and standards'. Since 2008, attendee numbers, representing user institutions worldwide, has varied but generally shown an increasing trend, from around 26 in 2008 to 60 in 2011 and 44 in 2012.

The first SBML model repository was designed, populated, and hosted at the University of Hertfordshire; when it was converted into a relational database, BioModels, and moved to a more appropriate host (the EMBL-EBI in Hinxton, Cambridge, UK). The publishers of several major journals whose readership extends well beyond academia, among them the Nature Publishing Group, *Public Library of Science* and BioMedCentral, opted to require or recommend deposition of published models in SBML format in the BioModels Database.

As of July 2013, 254 software tools have built-in or add-on SBML compliance. MATLAB, in the MathWorks SimBiology® package, and Mathematica, in the Wolfram SystemModeler™ module,

support import and export of SBML models, and numerous specialised SBML-compliant tools expose MATLAB or Mathematica APIs. Users and developers of SBML-compliant software, from industry and academia, form a lively international community: the discussion lists on the SBML.org website have seen over 7,500 posts in total since September 2002. The initial project and its later incarnations have brought experimental and theoretical biologists together with software developers, electronic engineers, control system specialists, mathematicians, physicists, and others with an interest in emergent properties of complex biological systems. At least one company, Integrative Bioinformatics (IBI), has, according to its CEO, 'invested heavily in SBML'. IBI bridges the increasing chasm that experimental life scientists need to cross in their interpretation and quantification of their data, by providing consultancy services to R&D labs – those situated in academia as well as in the pharmaceutical industry (see section 5, item 7).

SBML's mission continues: providing support for multi-component, multi-scale models and model composition has proven more problematic than initially envisaged, but the SBML community guarantees that discussions will be held in the open, and that everyone with an idea or an interest can contribute.

5. Sources to corroborate the impact (indicative maximum of 10 references)

1. SBML portal: <http://sbml.org/Main_Page>. See specifically:
 - a) <<http://sbml.org/SBML.org>About>> for a short history of SBML and evidence of the role that members of the Biocomputation Group at Hertfordshire (Bolouri, Finney, Schilstra, Keating and Matthews) played in its establishment; and
 - b) <<http://sbml.org/Community>>, for an overview of the current SBML community
2. Endorsement of SBML in the Nature Publishing Group's *Molecular Systems Biology's* policy: <<http://www.nature.com/msb/about/oa.html>>
3. a) BioModels Database: <<http://www.ebi.ac.uk/biomodels-main/>>
 b) Editorial in *Nature* about BioModels Database, Systems Biology and SBML: <<http://www.nature.com/nature/journal/v435/n7038/full/435001a.html>>
4. *Molecular Systems Biology* and *Public Library of Science (PLoS)* journals both specify in their author guidelines use of the BioModels Database as standard:
 - a) *Molecular Systems Biology* guidelines, <http://mts-msb.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions#deposition>
 - b) *PLoS* guidelines, <<http://www.ploscompbiol.org/static/guidelines#accessionnumbers>>
5. Lists and details of the annual SBML Hackathons: <<http://sbml.org/Events/Hackathons>>
6. Wikipedia entries on BioModels Database and on SBML: <http://en.wikipedia.org/wiki/BioModels_Database>; and <<http://en.wikipedia.org/wiki/SBML>>
7. Comment posted on an SBML online forum by the CEO of IBI, detailing his use of SBML and knowledge of its use outside academia: <<http://sbml.org/Forums/index.php?t=tree&goto=8121&rid=0>>