

LibSBGN: electronic exchange of SBGN maps

M.P. van Iersel^{1, 2, 9}, S.E. Boyd³, F. Bergmann⁴, S. Moodie⁵, F. Schreiber^{6, 7}, T. Czauderna⁶, E. Demir⁸, N. Le Novère⁹, A. Sorokin⁵, H. Mi¹⁰, A. Luna^{11, 12}, U. Dogrusoz¹³, Y. Matsuoka¹⁴, A. Funahashi¹⁵, H. Kitano^{14, 16, 17}, M.I. Aladjem¹¹, M.L. Blinov¹⁸, A.C. Villéger^{19, 20}

1) Department of Bioinformatics BiGCaT, Maastricht University, Maastricht, The Netherlands. 2) Netherlands Consortium for Systems Biology (NCSB), The Netherlands. 3) AgriBio, La Trobe University, Melbourne, Australia. 4) Department of Bioengineering, University of Washington, Seattle, Washington, USA. 5) School of Informatics, University of Edinburgh, UK. 6) Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. 7) Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany 8) Memorial Sloan Kettering Cancer Center - Computational Biology Center, New York, NY, USA. 9) EMBL European Bioinformatics Institute, Hinxton, UK. 10) SRI International, Menlo Park, California, USA. 11) National Cancer Institute, Bethesda, Maryland, USA. 12) Department of Bioinformatics Boston, MA, USA 13) Bilkent Center for Bioinformatics, Bilkent University, Ankara, Turkey. 14) The Systems Biology Institute, Tokyo, Japan. 15) Department of Biosciences and Informatics, Keio University, Yokohama, Japan. 16) Okinawa Institute of Science and Technology, Okinawa, Japan. 17) Sony Computer Science Laboratories, Tokyo, Japan. 18) Department of Computational Biology, University of Medicine, USA. 19) School of Computer Science, University of Manchester, UK. 20) Manchester Interdisciplinary Biocentre, Manchester, UK.

What is SBGN?

The Systems Biology Graphical Notation (SBGN, see also Le Novère et al., 2009) is a standard notation for the visual representation of biochemical and cellular processes. SBGN maps can be used to exchange complex biological knowledge in a clear, concise and unambiguous manner. See figure 1 for an example of the SBGN Process Description (PD) language.

What is LibSBGN?

As SBGN is becoming adopted more widely, there is a need for exchanging maps electronically between the various systems biology tools. Exchange using graphicsonly file formats (such as SVG) is often insufficient, because biological meaning of elements is lost. There is a need for a toolset to exchange maps while preserving biological meaning and relations.

Figure 1: Example SBGN map Figure 2: SBGN-ML example Catalysis arcs end Both arcs and glyphs There are two main with an open circle Circles represent Rounded rectangles have a class attribute, elements at the top and connect an pools of **simple** represent pools of which determines both the level: glyph and arc. <?xml version="1.0" encoding="UTF-8"?> enzyme to a reaction. chemicals. sbgn xmlns="http://sbgn.org/libsbgn/@ macromolecules. biological meaning and <map language="process description"> <glyph class="simple chemical" id="id1"> the appearance of that <label text="DHAP"/> glyph. <bbox x="604" y="90" w="40" h="40"/> glucose-6P isomerase phospho fructokinase <glyph class="simple chemical" id="id2"> <label text="GA-3P" />

 triose-P isomerase <glyph class="macromolecule" id="id3"> <label text="ADH1" /> <bbox x="675" y="150" w="60" h="30"/> <glyph class="process" orientation="vertical"</pre>

 Glyphs have an <port x="624" y="160" id="id4.1"/> <port x="624" y="170" id="id4.2"/> optional label. In the current version, only the ADP <arc class="production" id="id5" source="id</pre> bounding box of glyphs is stored. <start x="624" y="160" /> Further details, such as e.g. the <end x="624" y="130" /> roundness of corners, are left for <arc class="catalysis" id="id6" source="id3</pre> <start x="630" y="165" /> future versions. <end x="674" y="165" /> pyruvate kinase **Clone markers** enolase PGK1 <arc class="production" id="id7" source="id4.2" target="id2"> indicate duplication, <start x="624" y="170" /> and are used for <end x="624" y="200" /> A reaction consists of Square **process** common chemicals. The connection between consumption and **nodes** represent </sbgn> arcs and glyphs is stored **production** arcs. processes such as using a system of biochemical reactions. identifiers and references.

Implementation

To meet this need, we defined a new markup language named SBGN-ML. This file format is based on an XML Schema definition (XSD). See figure 2 for an example.

In addition we developed a supporting software library called LibSBGN. Besides reading and writing Reference maps have SBGN-ML files, this library will also be used to been made for all three languages of SBGN. validate SBGN maps against the specifications, and convert to- and from related systems biology standards, such as SBML (with layout extension) and BioPAX. LibSBGN is currently implemented in Java and C++. LibSBGN is still under development, but it is already being adopted by several tools (see the project wiki for an up-to-date list).

A test suite has been created; it consists of dozens of reference maps covering every aspect of SBGN. Corresponding reference documents have been encoded in XML. To test the adherence of supporting tools to the SBGN-ML specification, a rendering comparison pipeline was set up (see figure 3); it automatically renders SBGN maps through different programs, thus enabling comparison against the reference map.

Community & Infrastructure

LibSBGN is a community driven effort, involving people from institutes all over the world, representing a wide selection of tools. To organize this community, we have set up the following infrastructure:

- A wiki for documentation, bug tracker and code repository available at SourceForge: http://libsbgn.sourceforge.net
- Mailing list: sbgn-libsbgn@lists.sourceforge.
- Monthly online meetings, open to any interested party.

Figure 3: Renderer comparison Currently, Three pathway tools support conversion to the LibSBGN file format, namely

By comparing three implementations, we can root out ambiguities and tool-(T) (P) (P) ① F P specific artifacts. P@T286 T306 mt:prot

PathVisio, Render extension

and SBGN-ED.

STAT1α P@/701 P@/727 STAT1α P@Y701 P@Y727 IRF1-GAS

Some features are optional. For

example, not all tools can do

curved edges. The biological

regardless.

meaning of the map is the same

Roadmap

Milestone 1: (completed Feb 2011)

- Implement semantics of Process Description language Level 1 Version 1.1.
- Only high-level graphics specification.
- Low-level validation with XML Schema
- Beta support for C++

Milestone 2:

- Implement all three languages: PD, ER and AF
- Support for third party extensions
- High-level validation with Schematron

Milestone 3:

• Pixel-perfect graphics specification

And beyond...

- Conversion to and from SBML (with layout extension)
- Conversion to and from BioPAX

Acknowledgements

The authors acknowledge the support of a BBSRC United States Partnering Award (BB/ G530441/1) for travel and other collaboration costs.

References

Converting in each tool

automatically helps

detect errors.

N. Le Novère et al., The Systems Biology Graphical Notation, Nature Biotechnology 2009, 27(8):735-41