Saint: a lightweight integration environment for model annotation
Allyson L. Lister, Matthew Pocock, Morgan Taschuk and Anil Wipat
Centre for Integrated Systems Biology of Ageing and Nutrition, Newcastle University and
School of Computing Science, Newcastle University, Newcastle Upon Tyne, UK

Received on June 8, 2009; revised on August 4, 2009; accepted on August 26, 2009
Advance Access publication September 7, 2009
Associate Editor: Alex Bateman

ABSTRACT
Summary: Saint is a web application which provides a lightweight
annotation integration environment for quantitative biological
models. The system enables modellers to rapidly mark up models
with biological information derived from a range of data sources.
Availability and Implementation: Saint is freely available for use
on the web at http://www.cisban.ac.uk/saint. The web application
is implemented in Google Web Toolkit and Tomcat, with all major
browsers supported. The Java source code is freely available for
download at http://saint-annotate.sourceforge.net. The Saint web
server requires an installation of libSBML and has been tested on
Linux (32-bit Ubuntu 8.10 and 9.04).
Contact: helpdesk@cisban.ac.uk; a.l.lister@ncl.ac.uk
Supplementary information: Supplementary data are available at
Bioinformatics online.

1 INTRODUCTION
Quantitative modelling is at the heart of systems biology. Model
description languages such as the Systems Biology Markup
Language (SBML; Hucka et al., 2003) or CellML (Lloyd et al.,
2008) allow the relationships between biological entities to be
captured and the dynamics of these interactions to be described
mathematically. Currently, however, many dynamic models include
only the mathematical information required to run simulations, and
do not explicitly contain the full biological context.
The efficient exchange, reuse and integration of models is aided
by the presence of biological information in the model. Model
annotations are necessary to describe how the model has been
generated and to define the meaning of the components that make
up the model in a computationally accessible fashion. If biological
information about a model is added consistently and thoroughly, the
model becomes useful not just for simulation, but also as an input
to other computational tasks and as a reference for researchers.

Without a biological context, models are only easily understandable
by their creators. Interested third parties must visit many web sites, applications and interfaces in order to identify
relevant information, and may not be aware of all potentially useful
databases. Because modellers add information manually, it is very
difficult to annotate exhaustively.

The addition of biological annotations to a model is usually a
manual, time-consuming process; there is no single resource that
encompasses all suitable data sources. A modeller usually has to
visit many web sites, applications and interfaces in order to identify
relevant information, and may not be aware of all potentially useful
models. Currently, however, many dynamic models include
difficult to annotate exhaustively.

Whilst annotations are vital for model sharing and reuse, they
do not contribute to the mathematical content of a model and are
not critical to its successful functioning. The addition of biological
knowledge must be performed quickly and easily in order to make
annotation worthwhile to a modeller. A large number of tools
are available for the construction, manipulation and simulation of
models, but there is currently a lack of tools to facilitate rapid
and systematic model annotation. While web sites and applications
specializing in integrating disparate data sources exist, such as
BioMart (Smedley et al., 2009) and Pathway Commons, none are
designed to put information directly into a model.

In this article, we describe a lightweight SBML model annotation
tool called Saint, specifically designed to identify and integrate
biological information relevant to computational models. Saint is
an application which supports the addition of basic annotation to
SBML entities and identifies new reactions which may be valuable
for extending a model. Whilst the addition of biological annotation
does not modify the behaviour of the model, the incorporation of
new reactions or species adds new features that can later be built
upon to potentially change the model’s output.

2 IMPLEMENTATION
On the client side, Saint is a web application implemented in
Google Web Toolkit and hosted on a Tomcat server, with a query
translation service connecting to a number of external web services
running on the server side. New annotation is presented to the user
in a single integrated view after retrieval by the server-side queries.
Reactions and associated species are added directly to the
SBML model, whereas the majority of the remaining biological
annotation is added to Annotation elements according to the
Minimal Information Required in the Annotation of Biochemical
Models (MIRIAM) specification (Le Novère et al., 2005).

*To whom correspondence should be addressed.

© The Author(s) 2009. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
MIRIAM annotations are resource annotations that are added to SBML in a standardized way which link external resources such as ontologies and data sources to a model. MIRIAM, among other things, defines an annotation scheme accessible via web services which specifies the format and set of standard data types which should be used for these URIs (Laibe and Le Novère, 2007). The use of the MIRIAM format provides a standard structure for explicit links between the mathematical and biological aspects of an SBML model.

Saint facilitates the biological annotation of SBML models by using query translation to present an integrated view of data sources and suggested ontological terms. Data sources include UniProtKB (The UniProt Consortium, 2008), STRING (Jensen et al., 2008) and Pathway Commons. Supported ontologies and standards include MIRIAM, the Systems Biology Ontology (SBO; Le Novère, 2006) and Gene Ontology (GO; Ashburner et al., 2000). Query translation within Saint occurs when the query for each species is translated into a set of queries over these resources’ web services. Data are matched to a species through syntactic equivalence between the query term and the external data source. The combined query results are then displayed in the web browser.

If a model is valid, Saint displays the parts of the model available for annotation. The display is organized around species, which are the main target of annotation. Saint makes use of the Google Web Toolkit to provide both asynchronous calls to external resources and cross-browser compatibility. New annotation can be viewed by the modeller, even if the other species are not annotated yet. The modeller can select or delete annotations as it suits their model, or hide entire species from consideration. When the modeller is satisfied with the new state of the model, it can be converted back to SBML and saved. Parsing and validation of the models are handled with libSBML (Bornstein et al., 2008).

As an example, a Saccaromyces cerevisiae model containing a species with a single, simple identifier of ‘cdc13’ is loaded into Saint. Saint suggests the SBO term ‘macromolecule’ (SBO:0000245), which is added as an attribute to that species element, as the best SBO match to a protein. This term was suggested both because ‘cdc13’ was found within UniProtKB and because the Pathway Commons interaction set identified the species as a protein. Saint also suggests the UniProtKB accession P32797, and GO terms including ‘nuclear telomere cap complex’ (GO:0000783) and ‘single-stranded telomeric DNA binding’ (GO:0034047) as retrieved from UniProtKB. This information is stored within the model via MIRIAM annotations. Extensions to the model are also suggested. For each species, new reactions and their associated species and reactions are retrieved from both Pathway Commons and STRING. More examples and comparisons are available in the Supplementary Material.

3 DISCUSSION

To date, there are few tools available for automating the retrieval and integration of data for the annotation of SBML models. The Saint application was developed as an interactive web tool to annotate models with new MIRIAM resources and reactions, keeping track of data provenance so that the modeller can make an informed decision about the quality of the suggested annotation. The system makes it easy for modellers to add explicit biological knowledge to their models, increasing a model’s usefulness both as a reference for other researchers and as an input for further computational analysis.

A small number of similar tools are available. SemanticSBML provides MIRIAM annotations via a combination of data warehousing and query translation via web services as part of a larger application. The Java library libAnnotationSBML (Swainston and Mendes, 2009) uses query translation to provide annotation functionality with a minimal user interface. Unlike libAnnotationSBML, Saint is accessible through an easy-to-use web interface and unlike both tools is unique in its ability to add new reactions and associated species.

Saint is under active development. Future enhancements will include the addition of new data sources and ontologies, annotation of elements other than species and species and reactions and support for other modelling formalisms such as CellML.

ACKNOWLEDGEMENTS

We acknowledge the support of Newcastle University Systems Biology Resource Centre (SBRC) and Newcastle University Bioinformatics Support Unit.

Funding: BBSRC/EPSCR funding for CSIBAN (BB/C008200/1 to A.L.L., M.T. and A.W.); BBSC ONDEX project (BB/F006063/1 to M.P.).

Conflict of Interest: none declared.

REFERENCES


http://sysbio.molgen.mpg.de/semanticsbml/