Towards a web-based simulation experiment description repository

Michael A. Guravage and Roeland M.H. Merks

NCSB-NISB Biomodeling and BioSystems Analysis - Life Sciences Group, CWI, Amsterdam

The Problem with Models

Modellers are producing more and more complex models. Unless these models are sufficiently characterised and made available to the research community their reuse will be minimal, and reproducing simulation experiments incorporating them will prove problematic. Consensus on the content and form of experiment recipes that combine models and simulations will encourage model sharing and facilitate reuse.

A set of guidelines specifying the Minimum Information About a Simulation Experiment (MIASE)\[^{[1]}\] proposes a common set of information necessary to reproduce simulation experiments that incorporate quantitative models.

We have instantiated these guidelines in a web-based content management system. Our system you can create Simulation and Experiment Descriptions, enrich them with experimental data and annotate them with domain meta-information to facilitate classification, searching and cross referencing - all with the goal of reusing your models and reproducing your experimental results.

Simulation & Experiment Description Markup Language

One instantiation of the MIASE guidelines is SED-ML\[^{[2–4]}\] - an XML schema, instances of which are recipes describing the combination of models and simulations into reproducible experiments. In particular, SED-ML describes five components essential to compose a simulation experiment description, i.e.:

- simulations – a description of the simulations’ method, type and algorithm (KiSAO)\[^{[5]}\]
- models – a description of the models’ location, language and modifications
- tasks – the glue that combine models and simulations into experiments
- data generators – how to present the simulations’ results, e.g. 2D graph
- outputs – how to transform raw simulation output into numerical or graphical results

SEDMLType: extension of SEDBase

We have extended the SED-ML standard to accommodate source-code models, and enriched the output types with descriptive text, images and animations – which makes the repository useful as a laboratory notebook.

Insuring Experiment Results are Reproducible

The Simulation & Experiment Description Meta Language (SED-ML) is a means – like a recipe – to describe the combination of simulations and models in reproducible experiments. We have built a web-based simulation & experiment description repository based on SED-ML. Implemented as an add-on product to the open-source content management system named Plone\[^{TM}\], our repository allows researchers to:

- create simulation & experiment descriptions (SEDs) compliant with SED-ML
- annotate simulations & experiments with domain meta-information from various ontologies
- search for simulations & experiments based on a variety of criteria, e.g. algorithm type
- collaborate with colleagues by allowing them to modify your simulations & experiments
- curate simulations & experiments to build a collection of verified and approved SEDs
- export simulations & experiments in SED-ML compliant XML

In contrast to EBI’s existing BioModels database, which hosts only biochemical models written in SBML, our repository will uniquely record any simulation experiment, including those written in C++, thus making the tool generally applicable to the types of simulation models used within the NCSB.

results

Future Work

- Add the ability to import SEDs from external SED-ML files.
- Develop tools to automatically reproduce experiments expressed in SED-ML.
- Improve the user interface by displaying related information simultaneously.
- Distribute our repository as an open-source Plone add-on product.