

A modular semantic annotation framework: CellML Metadata Specifications 2.0

Cooling, M. T. (m.cooling@auckland.ac.nz)

Auckland Bioengineering Institute, The University of Auckland, New Zealand.

Aim: Transform mathematical models from the esoteric to the informative

The last ten years has seen the development of model encoding schemes for documenting and sharing mathematical models. Now hundreds of published computational biology models are available for free, online, encoded in languages such as the XML-based CellML [1] and SBML [2].

Availability of models has been achieved, but what about **Utility**?

Currently it is often difficult and time-consuming to:

- Understand exactly what the model elements represent
- Find the authors to ask them questions, or credit them
- See what assumptions are embodied
- Determine how best to add new model elements

From Esoteric to Informative

Semantic annotation allows humans and computers to better determine the meaning, source and provenance of model constructs (supporting minimum information standards [3]) – crucial information as model complexity grows.

CellML is facilitating successful research in both Systems and Synthetic Biology [4,5]. Its inherent support for modular models makes it an excellent language for testing annotation schemes with respect to their support for complex models and smart aggregation.

We have developed a modular metadata specification framework to better enable the annotation of CellML models. The framework consists of a Core that prescribes how to link RDF statements to CellML model constructs, with multiple consistent yet loosely-coupled Satellite specifications.

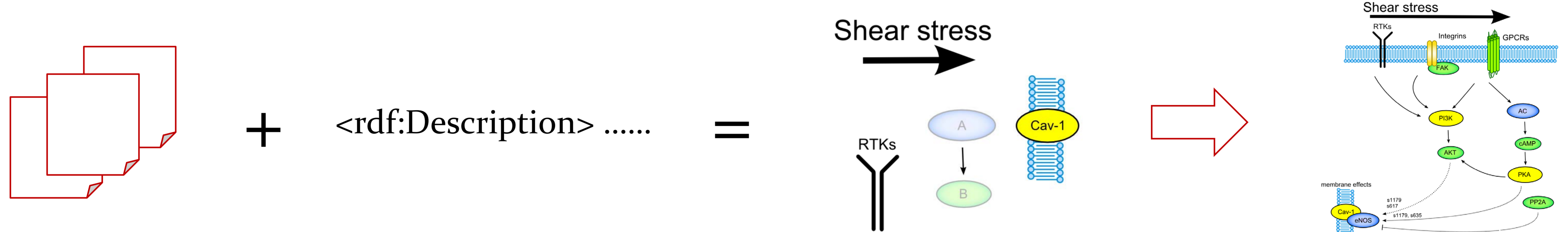
Specification Framework

Annotations are written in RDF - either in-file as RDF/XML, or decoupled from model files via XPATH [6].

Satellites follow consistent use of existing standards such as Dublin Core [7], Friend-of-a-Friend (FOAF) [8], RDF Schema (rdfs:comment) [9], Biomedical Model Qualifiers [10] and the Bibliographical Ontology (BIBO) [11] in order to support the following:

- Basic information - provides model elements with people, timestamps and supporting comments
- Licensing - linking models with legal information
- Citations - linking model elements to academic documents and personal communications
- Biological meaning - linking model elements to the biological entities and processes that they represent

Application – Vascular Endothelial Cell Signalling Research



Model Files (CellML)

Metadata (RDF/XML)

Components with Semantics

Smart Aggregation

Annotated mathematical models unambiguously represent biological entities and processes to both humans and machines. Imbued model components can then be more readily combined in larger models, as shown here for the key players in endogenous nitric oxide synthase (eNOS) signaling in endothelial cells. A resulting repository of model components for investigating questions in the onset of cardiovascular disease is to be made available online [12].

Future Work

- Proposing enhancements to biological ontologies and databases to facilitate precise annotation
- Extending the satellite specification suite to include modelling construct and mathematical formalism information
- Designing and developing smart model aggregation services

[1] www.cellml.org [2] sbml.org [3] Laibe, C. *et al.* (2007) BMC Sys. Bio. [4] Hunter, P. & Borg, T. K. (2003) Nat. Rev. Mol. Cell Bio. [5] Cooling, M. T. *et al.* (2010) Bioinf. [6] www.w3.org/TR/xpath/ [7] dublicore.org [8] www.foaf-project.org [9] www.w3.org/TR/rdf-schema/ [10] www.ebi.ac.uk/miriam/main/qualifiers/ [11] bibliontology.org [12] models.cellml.org