The CellML Metadata 1.0 Specification
Working Draft - 16 January 2002

1   Introduction

1.1   Need for Metadata in CellML

Metadata is usually defined as "data about data". It is the supporting information that provides context to a resource. In CellML, the model (i.e. the structure and mathematics of the model) is the resource. Information that puts the model into a wider context is metadata. Metadata in CellML includes information such as the literature reference that supports the model, the identity of the creator(s) of the model, and the species for which the model is relevant.

The CellML project needs metadata for two primary reasons:

- It will be difficult to reuse other people's models and components without metadata to provide the context for these objects. A modeller considering reusing someone else's model component will need to know things such as: what biological entity the component represents, for which species the component is relevant, and when the component was created and last modified (to help determine whether it is likely to incorporate the most recent experimental results).

- As the number of models and components grows, metadata will provide the only scalable method for locating particular models and components. Experience in other biological fields shows that as a field grows, powerful search techniques are needed to enable researchers to find relevant resources. These search techniques require structured metadata.

Metadata in CellML can be used in many different ways, such as:

- To support searches of a model repository (or at least to make it possible to automate loading of a database that will support such searches).

- To enable automatic discovery of models published on remote websites, such as laboratory websites.

- To allow the documentation for a model to be kept in the same document as the model itself, which will keep the documentation from becoming obsolete as work continues on the model.

The metadata structure should be flexible and extensible because it is almost certain that we have not thought of all possible uses of CellML Metadata.

1.2   The Larger Metadata Picture

Currently it is not particularly easy to find a specific piece of information on the web, and, once you have found the information, it is not easy to determine whether or not you should trust it. Metadata can address both of these problems. Therefore, there is a push to begin to incorporate metadata into web resources to allow users to get the maximum use out of the information on the web. Tim Berners-Lee has been particularly active in advocating a "semantic web" in which resources would include the semantic information necessary to allow machines to *understand* (not just read) them. The W3C has set up a semantic web activity[1] to promote what they view as phase two of the Internet. Some software projects, such as Mozilla[2], have

---

[1]http://www.w3.org/2001/sw/
[2]http://www.mozilla.org/

begun trying to take advantage of the metadata that is currently available about web resources. The CellML development team is working to make CellML compatible with the semantic web activity.

E-mail questions, criticism, submissions or info to info@cellml.org
Input document last modified : Tue Feb 10 16:18:55 NZDT 2004