CellML Specification
Final Draft — 18 May 2001

8   Metadata Framework

8.1   Introduction

Metadata is "*data about data*". In a CellML document, the principal data is the structure and mathematics of a biological model. Information that provides context for this data is metadata. Metadata can be included in a CellML document to facilitate searches of collections of models and model components. It provides a means for a modeller to include structured descriptive information about the model, which can help other modellers determine whether they can incorporate the model into their own work.

This section of the CellML specification presents a framework for the use of metadata in a CellML document. Methods for identifying types of metadata within that framework are recommended in the CellML Metadata Specification[1]. The use of these methods ensures reliable extraction of metadata from CellML documents across all processors aware of CellML Metadata. The CellML Metadata specification is being developed independently of the CellML specification.

All metadata is optional. A model without any metadata is a valid CellML model. However, it is recommended that a CellML document author provide as much metadata as possible, particularly his/her name and contact information and a reference for a paper that describes the development of the model.

8.2   Basic Structure

Metadata should be embedded in a CellML document using the Resource Description Framework (RDF)[2], the syntax of which is defined in the RDF Model and Syntax Recommendation[3]. For interoperability, CellML processing software should make use of the methods for identifying types of metadata outlined in the CellML Metadata Specification[4].

Section 2.2.2[5] defines two metadata namespaces that CellML processing software is expected to recognise, and recommended prefixes to which these namespaces should be mapped. RDF elements are placed in the RDF namespace, which should be mapped to the prefix **rdf**. CellML Metadata elements and attributes have their own namespace which should be mapped to the prefix **cmeta**.

CellML processing software is free to ignore any and all metadata. However, it is recommended that software at least display metadata. Model authors are free to develop their own RDF schema for metadata, or to store metadata in another format by using the CellML extension mechanism described in Section 2.2.3[6]. However, doing so decreases the likelihood that CellML processing software will be able to do anything useful with the metadata in a CellML document.

Metadata is defined within an **<rdf:RDF>** element as shown in Figure 16. The **rdf**, **cellml**, and **cmeta** prefixes are used throughout this section to indicate that elements and attributes are in the RDF, CellML and CellML Metadata namespaces, respectively. The recommended practice is to define the RDF namespace and any namespaces used by the enclosed metadata on the **<rdf:RDF>** element, even if these namespaces are already defined on the ancestor elements of the **<rdf:RDF>** element. This increases the re-usability of the RDF block. Furthermore, RDF processing software that does not recognise the

---

[1] http://www.cellml.org/public/metadata/index.html

[2] http://www.w3.org/RDF/

[3] http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[4] http://www.cellml.org/public/metadata/index.html

[5] http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_fundamentals_namespaces

[6] http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_fundamentals_extending_cellml

CellML namespace can still parse a CellML document, extract the RDF blocks, and perhaps provide useful functionality with the information described in the RDF.

An **<rdf:RDF>** element typically contains one or more **<rdf:Description>** elements, each of which defines an **about** attribute. The value of the **about** attribute must be a valid Uniform Resource Identifier[7] (URI). Metadata may be associated with the *document* it is defined in by assigning the **about** attribute an empty value (**" "**). Metadata may be associated with an element in the current document by defining an attribute of type ID on that element and assigning the **about** attribute on the **<rdf:Description>** element a value equal to the value of that attribute preceded by a hash (#). An attribute must be given a type of ID in the document type declaration (DTD) or schema associated with an XML document, and its value must be unique across all attributes of type ID in a given document. The correct way to do this in a DTD is described in Section 3.3.1 of the XML 1.0 Recommendation[8].

As was discussed in Section 2.2.1[9], the **name** attribute that occurs on many CellML elements is not of type ID because it is not necessary that CellML identifiers be unique across a document. To facilitate the association of metadata with CellML elements, a **cmeta:id** attribute in the CellML Metadata namespace may be added to any CellML element. The CellML 1.0 DTD (given in Appendix A.6[10]) declares this attribute to be of type ID for all CellML elements. This declaration prevents CellML elements from having any other attributes of type ID, including attributes in extension namespaces. The MathML 2.0 DTD defines an attribute **id** of type ID for all MathML elements. Extension elements may use their own attributes of type ID, or make use of **cmeta:id** attributes, which CellML processing software is required to treat as if it had type ID. The value of an attribute of type ID must conform to the requirements specified in the XML specification.

For interoperability, an RDF block should be stored in the element about which it contains metadata. This makes the element more re-useable. Elements in the MathML namespace are an exception to this recommendation. The MathML content of a **<cellml:component>** element might be extracted for use in a general MathML processor, which might not be able to handle RDF content. Therefore, metadata on MathML elements should be placed in the containing **<cellml:component>** element. If the RDF block contains metadata about the CellML document, it should be included in the root element of the document. Note that simply putting an RDF block inside an element is not sufficient to indicate that the metadata in the block refers to that element. The **about** attribute on the **<rdf:Description>** element must be used to indicate the resource about which the RDF block contains metadata.

8.3   Examples

Figure 16 demonstrates the use of metadata in CellML. Three RDF blocks are shown: one that provides metadata about the CellML document, one that provides metadata about the model, and one that provides metadata about a component contained in the model. Only the RDF framework elements are shown. The actual metadata is not shown here. Examples in the CellML Metadata Specification[11] will demonstrate how to use the recommended metadata elements.

The first RDF block provides metadata about the CellML document. This is indicated by the empty value of the **about** attribute on the **<rdf:Description>** element. The second RDF block has a value of **"#model01"** in the **about** attribute on the **<rdf:Description>** element. This indicates that this metadata provides information about the model that is delimited by the **<cellml:model>** element with a **cmeta:id** attribute value of **"model01"**. The final RDF block provides metadata about the **membrane** component. This is indicated by the **about** attribute with a value of **"#comp01"** on the

---

[7] http://www.ietf.org/rfc/rfc2396.txt

[8] http://www.w3.org/TR/2000/REC-xml-20001006#sec-attribute-types

[9] http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_fundamentals_identifiers

[10] http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_cellml_1_0_dtd

[11] http://www.cellml.org/public/metadata/index.html

```xml
<model
    name="example_metadata_model"
    cmeta:id="model01"
    xmlns="http://www.cellml.org/cellml/1.0#"
    xmlns:cellml="http://www.cellml.org/cellml/1.0#"
    xmlns:cmeta="http://www.cellml.org/metadata/1.0#">

  <!-- This metadata block is about the CellML document -->
  <rdf:RDF
      xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description about="">
      <!-- Some metadata content, such as a last-modified date -->
    </rdf:Description>
  </rdf:RDF>

  <!-- This metadata block is about the CellML model -->
  <rdf:RDF
      xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description about="#model01">
      <!--
       Some metadata content, such as a species for which
       the model is relevant.
      -->
    </rdf:Description>
  </rdf:RDF>

  <component name="membrane" cmeta:id="comp01">

    <!-- This metadata block is about the membrane component -->
    <rdf:RDF
        xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
      <rdf:Description about="#comp01">
        <!--
         Some metadata content, such as a annotation describing
         limitations of this representation of the membrane
        -->
      </rdf:Description>
    </rdf:RDF>

  </component>

</model>
```

FIGURE 16: An example demonstrating how metadata can be embedded in a CellML document using the Resource Description Framework (RDF).

<rdf:Description> element.

All three RDF blocks declare the RDF and CellML Metadata namespaces. This makes the RDF blocks portable: the information needed to interpret the RDF will be preserved even if the blocks are extracted from the CellML document.

8.4 Rules for CellML Documents

8.4.1 Proper use of the **cmeta:id** attribute

- A **cmeta:id** attribute (where the **cmeta** prefix is mapped to the CellML Metadata namespace URI defined in Section 2.2.2[12]) may be defined on any CellML element. A **cmeta:id** attribute may also be defined on extension elements for which no attribute of type ID is declared in the DTD, schema or language specification.

  [On MathML elements, the **mathml:id** attribute must be used. A **cmeta:id** attribute must specifically not be added to MathML elements, because a given element may only contain one attribute of type ID.]

8.4.2 The **<rdf:RDF>** element

1. **Allowed use of the <rdf:RDF> element**

   - Any CellML element may contain any number of **<rdf:RDF>** elements.

     [Metadata may appear on any CellML element, and may be split across multiple **<rdf:RDF>** elements. The recommended practice is to enclose all metadata relevant to a particular resource in a single **<rdf:RDF>** element. In this and subsequent rules, the use of the **rdf** prefix indicates that elements and attributes are in the RDF namespace.]

   - The content of an **<rdf:RDF>** element must conform to the Resource Description Framework (RDF) Model and Syntax Specification[13] recommendation from the W3C.

     [For interoperability, the abbreviated syntax defined in the RDF recommendation should be avoided. However a **rdf:parseType** attribute with a value of `"Resource"` can be added to non-RDF elements to create anonymous resources within an **<rdf:RDF>** element.]

8.5 Rules for Processor Behaviour

8.5.1 Treatment of **cmeta:id** attributes

CellML processing software must treat any **cmeta:id** attributes in a CellML document (where the **cmeta** prefix is mapped to the CellML Metadata namespace URI defined in Section 2.2.2[14]) as if they're of type ID. This has the following consequences for CellML documents:

- A **cmeta:id** attribute must not be defined on a non-CellML element for which the DTD, schema or language specification has already declared an attribute of type ID.

- The values of all **cmeta:id** attributes and any other attributes of type ID in a given CellML document must be unique.

- The values of all **cmeta:id** attributes in a CellML document are potential targets for the values of **about** attributes on **<rdf:Description>** elements.

---

[12]http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_fundamentals_namespaces
[13]http://www.w3.org/TR/1999/REC-rdf-syntax-1999022
[14]http://www.cellml.org/public/specification/20010518/cellml_specification.html#sec_fundamentals_namespaces

8.5.2 General meaning of metadata

Metadata may refer to the CellML document, the CellML model, or a specific element within the CellML model. The following list documents the intended meaning of metadata on each of these resources. More detailed information can be found in the CellML Metadata Specification[15].

- Metadata that refers to the CellML document provides information relevant to the document as a whole, independent from the use of the document to specify a model. Examples of metadata that might appear on a CellML document are last modified date (date on which the document was last edited) and publisher (person or organization distributing the document).

- Metadata that refers to the CellML model provides information relevant to the model as a whole. For instance, the model author is the person who created the complete model, even if some of the components were taken from a shared database and have different authors.

- Metadata that refers to a specific CellML element provides information about that element only. It does not provide information about elements that are contained in the referenced element.

E-mail questions, criticism, submissions or info to info@cellml.org
Input document last modified : Sat May 19 00:36:14 NZST 2001

---

[15]http://www.cellml.org/public/metadata/index.html