

CellML Specification

Draft — 2 March 2001

8 Metadata Framework

8.1 Introduction

Metadata is “*data about data*”. In a CellML document, the principal data defined 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.

The CellML metadata structure is defined in a parallel document. This section of the CellML specification presents a framework for the use of metadata in a CellML document.

8.2 Basic Structure

Metadata is defined in a CellML document using the [Resource Description Framework](#)¹ (RDF), which is a W3C recommendation. Two CellML RDF Schema are being developed for the convenience of model authors and developers of CellML processing software. The first schema will define a data model for storing elements from the Dublin Core element set, modification history information, inline documentation and specific biological metadata. The second schema will define how information about literature references should be stored in a CellML document. This schema will be an RDF serialization of the [Object Management Group’s Bibliographic Query Service](#)² (BQS) data model. The CellML RDF Schema will be defined and discussed in a companion metadata specification.

The table in [Section 2.2.2](#)³ defines five 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`. Dublin Core elements and Dublin Core qualifier attributes are placed in the appropriate namespaces, which should be mapped to the prefixes `dc` and `dcq`, respectively. CellML metadata elements and BQS citation elements each have their own namespace, mapped to prefixes of `meta` and `bqs`, respectively.

CellML processing software is free to ignore any and all metadata. However, it is hoped that software will 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](#)⁴. However, doing so decreases the likelihood that CellML processing software will be able to do anything useful with the metadata in the model.

Metadata is defined within an `<RDF>` element in the RDF namespace, as shown in Figure 18. The recommended practice is to define the RDF namespace and any namespaces used by the enclosed metadata on the RDF element, even if these namespaces are already defined on the `<model>` element. This increases the re-usability of the RDF block. Furthermore, RDF processing software that does not recognise the CellML namespace can still parse a CellML document, extract the RDF blocks, and perhaps provide useful functionality with the information described in the RDF.

¹<http://www.w3.org/RDF/>

²<http://www.omg.org/lsr/>

³http://www.cellml.org/public/specification/20010302/cellml_specification.html#sec_fundamentals_namespaces

⁴http://www.cellml.org/public/specification/20010302/cellml_specification.html#sec_fundamentals_extending_cellml

The `<rdf:RDF>` element contains an `<rdf:Description>` element, which defines an `about` attribute. The value of the `about` attribute must be a valid [Uniform Resource Identifier](#)⁵ (URI). A URI that points to a resource in the current document consists of a hash (#) followed by the value of that resource's `id` attribute.

Metadata is associated with a CellML *document* by assigning the `about` attribute an empty value (" "). Any CellML *element* that has associated metadata must define an `id` attribute in the CellML metadata namespace (defined in [Section 2.2.2](#)⁶). This attribute is of type ID, as defined in the [XML specification](#)⁷. Its value must be unique across the CellML document, but need not have any meaning. Metadata is associated with a CellML element by assigning the `about` attribute on the `<rdf:Description>` element a value equal to the value of the `cmeta:id` attribute on the CellML element.

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 `<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 `<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 the 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 about which resource the RDF block contains metadata.

8.3 Examples

Figure 18 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 companion CellML metadata specification 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" for the `about` attribute on the `<rdf:Description>` element. This indicates that this metadata provides information about the model that is delimited by the `<model>` element with an `cmeta:id` attribute with a value of "model01". The final RDF block provides metadata about the `membrane` component. This is indicated by assigning a value of "#comp01" to the `about` attribute on the `<rdf:Description>` element.

Note that 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 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 about a particular element in a

⁵<http://www.ietf.org/rfc/rfc2396.txt>

⁶http://www.cellml.org/public/specification/20010302/cellml_specification.html#sec_fundamentals_namespaces

⁷<http://www.w3.org/TR/2000/REC-xml-20001006>

```
<model
  name="example_metadata_model"
  cmeta:id="model01"
  xmlns="http://www.cellml.org/2001/03/cellml"
  xmlns:cellml="http://www.cellml.org/2001/03/cellml"
  xmlns:cmeta="http://www.cellml.org/2001/03/metadata">

  <!-- This metadata block is about the CellML document -->
  <rdf:RDF
    xmlns:cmeta="http://www.cellml.org/2001/03/metadata"
    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/2001/03/metadata"
    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/2001/03/metadata"
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
      <rdf:Description about="#comp01">
        <!--
          Some metadata content, such as an annotation describing
          limitations of this representation of the membrane
        -->
      </rdf:Description>
    </rdf:RDF>

  </component>

</model>
```

FIGURE 18: An example of the use of metadata in a CellML document.

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](#)⁸ recommendation from the W3C.
[Avoid the abbreviated syntax defined in the recommendation to ensure maximum portability of the metadata.]

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

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

- The content of an **<rdf:Description>** element must conform to the [Resource Description Framework \(RDF\) Model and Syntax Specification](#)⁹ recommendation from the W3C.
[The recommended practice is for contained elements to adhere to an RDF schema, and to use the CellML metadata schema wherever possible.]

2. Allowed values of the **about** attribute

- The **about** attribute on an **<rdf:Description>** element must either be empty or have a value equal to a valid URI that points to an element in the current document (i.e., is equal to the value of a **cmeta:id** attribute on an element in the current document preceded by a hash (#)).
[An **<rdf:Description>** element with an empty **about** attribute contains information about the CellML document. An **<rdf:Description>** element with an **about** attribute that references a **cmeta:id** attribute value contains information about the element in the current document identified by the **cmeta:id** attribute.]

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

- The **cmeta:id** attribute may appear on any element in a CellML document.
[In this and subsequent rules, the `cmeta` prefix places elements and attributes in the CellML metadata namespace.]
- The value of the **cmeta:id** attribute must be unique across the CellML document.
- A **cmeta:id** attribute must be defined on any element in the CellML or MathML namespaces for which RDF metadata is defined.

8.5 Rules for Processor Behaviour

8.5.1 Metadata is optional

All metadata is optional. A model without any metadata is a valid CellML model. However, we strongly recommend that the modeller 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<http://www.w3.org/TR/1999/REC-rdf-syntax-1999022>

⁹<http://www.w3.org/TR/1999/REC-rdf-syntax-1999022>

8.5.2 Associating metadata with resources

Software must associate the metadata contained within an RDF block with a CellML document, a CellML model, or a specific element within the CellML model according to the following rules:

- If the **about** attribute on an **<rdf:Description>** is empty, then the metadata contained within the **<rdf:Description>** element refers to the entire CellML document.
- If the **about** attribute on an **<rdf:Description>** points to a **<model>** element, then the metadata contained within the **<rdf:Description>** element is associated with the referenced model.
- If the **about** attribute on an **<rdf:Description>** points to any other element within the current document, then the metadata contained within the **<rdf:Description>** element is associated with the referenced element.

8.5.3 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 companion CellML metadata specification.

- 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.