

# The CellML Metadata 1.0 Specification

## Working Draft - 2 November 2001

### 4 General Metadata

#### 4.1 Model Builder

Model builder metadata stores information about the person or persons who coded the model into CellML. A given element can have multiple model builders who may need to be considered as individuals or as members of a group. If they are members of a group, the group may or may not need to be ordered.

Model builder metadata is defined using the Dublin Core creator element, `<dc:creator>`. Repeating this element for a given CellML element indicates that the people listed worked independently on the model. This definition is shown in Figure 10. Listing multiple people in the `<dc:creator>` element using an `<rdf:Bag>` container indicates that the group of people worked together on the model, and that they are all considered equal contributors. This definition is shown in Figure 11. Listing multiple people in the `<dc:creator>` element using an `<rdf:Seq>` container indicates that the group of people worked together on the model and that their contributions are ordered (the first member of the list is first author, the second member is second author, and so on). Metadata authors are free to use the `<rdf:Alt>` container (as long as they produce valid RDF). However, CellML Metadata compliant software is not required to be able to consistently interpret the meaning of an `<rdf:Alt>` container in this context. Note that in the examples shown here, the basic vCard “name” construct is used to store the name of the model builder.

#### 4.2 Contributor

Contributor metadata indicates that a person contributed to a resource but did not actually create it (such as an editor).

Contributor metadata is defined using the Dublin Core contributor element, `<dc:contributor>`, as shown in Figure 12. Multivalued contributor metadata is handled exactly as multivalued model builder metadata. Simple repetition of the element indicates that the people contributed to the resource independently. The use of RDF bag (`<rdf:Bag>`) or sequence (`<rdf:Seq>`) containers indicates that the people contributed to the resource as an unordered or ordered group, respectively.

#### 4.3 Publisher

The publisher is the person or organization responsible for providing the model, model component, or other CellML element. A given CellML element can have multiple publishers.

Publisher metadata is defined with the Dublin Core publisher element (`<dc:publisher>`), as shown in Figure 13. Multivalued publisher metadata is handled exactly as multivalued model builder metadata. Simple repetition of the element indicates that the people or organisations publish the resource independently. The use of RDF bag (`<rdf:Bag>`) or sequence (`<rdf:Seq>`) containers indicates that the people or organisation published the resource as an unordered or ordered group, respectively.

#### 4.4 Copyright

The copyright metadata refers to the copyright that protects the CellML document, model, model component, or other CellML element. It is defined using the Dublin Core rights element (`<dc:rights>`), and,

---

```
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dc:creator>
      <vCard:N rdf:parseType="Resource">
        <vCard:Family>Flintstone</vCard:Family>
        <vCard:Given>Fred</vCard:Given>
      </vCard:N>
    </dc:creator>
    <dc:creator>
      <vCard:N rdf:parseType="Resource">
        <vCard:Family>Brown</vCard:Family>
        <vCard:Given>Charlie</vCard:Given>
      </vCard:N>
    </dc:creator>
    <dc:creator>
      <vCard:N rdf:parseType="Resource">
        <vCard:Family>Doo</vCard:Family>
        <vCard:Given>Scooby</vCard:Given>
      </vCard:N>
    </dc:creator>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 10: Recommended definition of model builder metadata in which multiple people worked independently on the model.

---

---

```

<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dc:creator>
      <rdf:Bag>
        <rdf:li>
          <vCard:N rdf:parseType="Resource">
            <vCard:Family>Flintstone</vCard:Family>
            <vCard:Given>Fred</vCard:Given>
          </vCard:N>
        </rdf:li>
        <rdf:li>
          <vCard:N rdf:parseType="Resource">
            <vCard:Family>Brown</vCard:Family>
            <vCard:Given>Charlie</vCard:Given>
          </vCard:N>
        </rdf:li>
        <rdf:li>
          <vCard:N rdf:parseType="Resource">
            <vCard:Family>Doo</vCard:Family>
            <vCard:Given>Scooby</vCard:Given>
          </vCard:N>
        </rdf:li>
      </rdf:Bag>
    </dc:creator>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 11: Recommended definition of model builder metadata in which multiple people worked together on the model and all are considered equal contributors.

---

```

<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dc:contributor rdf:parseType="Resource">
      <vCard:N rdf:parseType="Resource">
        <vCard:Family>Flinstone</vCard:Family>
        <vCard:Given>Fred</vCard:Given>
      </vCard:N>
    </dc:contributor>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 12: Recommended definition of contributor metadata.

---

---

```

<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="">
    <dc:publisher>
      University of Auckland, Bioengineering Research Group
    </dc:publisher>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 13: Recommended definition of publisher metadata. Note that the empty **about** attribute indicates that this metadata refers to the CellML *document* (as opposed to the model or a specific element in the model).

---

therefore, a given CellML element can technically have multiple copyrights. However, the recommended practice is to include only one copyright for any given element.

Figure 14 demonstrates the definition of the copyright metadata.

---

```

<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dc:rights>Physiome Sciences, 2000</dc:rights>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 14: Recommended definition of copyright metadata.

---

## 4.5 Creation Date

The creation date is the date upon which the model or model part was coded into CellML. A given CellML element can have only one creation date.

Creation date metadata is defined using the Dublin Core date qualifier element, **<dcq:creation>**. The creation date is further qualified by using the Dublin Core date encoding scheme qualifier element, **<dcq:W3CDTF>**, which indicates a YYYY-MM-DD format. The definition of creation date metadata is demonstrated in Figure 15.

## 4.6 Modified Date

The modified date is the date upon which the content of a CellML element was changed. The modified date metadata is defined with the Dublin Core date qualifier element, **<dcq:modified>**. Otherwise, its definition is exactly the same as that of the creation date metadata. The definition of modified date metadata is demonstrated in Figure 16.

---

```
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dcq:created>
      <dcq:W3CDTF>2000-10-05</dcq:W3CDTF>
    </dcq:created>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 15: Recommended definition of the creation date metadata.

---

---

```
<rdf:RDF
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dcq:modified>
      <dcq:W3CDTF>2000-10-05</dcq:W3CDTF>
    </dcq:modified>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 16: Recommended definition of the last modified date metadata.

---

## 4.7 Alternative Names

Alternative name metadata provides human-readable names for CellML elements. This preferred name could be used by software whenever it needs to display a human-readable name. The use of this metadata allows us to limit the values of **name** attributes on CellML elements to enable efficient code generation, without worrying about whether or not the name will be sufficiently meaningful to human readers.

Alternative name metadata is defined with the Dublin Core title qualifier element, **<dcq:alternative>**.

One element may have multiple alternative names. Only one should be considered the preferred human-readable name. The preferred name should be stored in a **<dc:title>** element. Additional names should be stored in the **<dcq:alternative>** element.

Figure 17 shows the definition of alternative name metadata. The element referenced by "#cellml\_element\_id" is given two human-readable names. The preferred one is "EGF-EGFR complex".

---

```
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dc:title>EGF-EGFR complex</dc:title>
    <dcq:alternative>epidermal growth factor-epidermal growth factor receptor complex</dcq:alternative>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 17: Recommended definition of alternative name metadata.

---

## 4.8 Species

Species metadata refers to the biological species (such as human, dog, pig, *Palaemon affinis*, etc.) for which an element is relevant. A given CellML element may be relevant for multiple species. It may also be relevant for an entire class of species, such as all mammals.

Species metadata is defined with a CellML-specific metadata element, **<cmeta:species>**, as shown in Figure 18. The content of this metadata must be a valid scientific name for a species or group of species. Notwithstanding recent arguments among taxonomists about the impact of genomic data on species classifications, scientific names are considered to be sufficiently standard to obviate the need to use a formal controlled vocabulary. The CellML Metadata specification recommends using [NCBI's Taxonomy Browser](http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html)<sup>1</sup> as a resource for scientific names. If a modeller needs to refer to a discontinuous group of species (i.e., one that cannot be specified by a single scientific name) he/she can include multiple **<cmeta:species>** elements. Multiple values for the species metadata will always mean that the CellML element is relevant for any one of the species listed. Relevance to all of the species as a group would imply some sort of population dynamics model, which is outside of the scope of CellML.

## 4.9 Sex

Sex metadata refers to the sex for which a CellML element is relevant. A given element may be relevant for more than one sex.

<sup>1</sup><http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html>

---

```

<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 rdf:about="#cellml_element_id">
    <cmeta:species>Mammalia</cmeta:species>
    <cmeta:species>Xenopus laevis</cmeta:species>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 18: Recommended definition of species metadata. The element referenced by "#cellml\_element\_id" is relevant for all mammals and the African clawed frog, *Xenopus laevis*.

---

Sex metadata is defined with the CellML-specific element, `<cmeta:sex>`, as shown in Figure 19. The valid content of this element must be chosen from the following controlled vocabulary:

- male
  - female
  - hermaphrodite
  - other
  - all
  - undefined (the element is explicitly specified not to have a defined relevance to any particular sex).
- 

```

<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 rdf:about="#cellml_element_id">
    <cmeta:sex>male</cmeta:sex>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 19: An example of the use of sex metadata

---

## 4.10 Biological Entity

This area of the metadata will almost certainly be expanded in future versions of CellML. For now it is simply a name or database unique identifier for a biological entity, such as an ion channel, signalling pathway, or specific cell type, that is represented by the model or model component. A given CellML element can represent multiple biological entities either as a complete group or as a list of alternatives. A CellML element that represents a list of alternative biological entities would probably be a “superclass”

component, that will be re-used multiple times in a model, each time to represent a different entity on the list of alternatives. For instance, a modeller might define a general “calcium-binding protein” component and then re-use this component three times in his/her model: once to represent calmodulin, once to represent troponin C, and once to represent parvalbumin. [Note that the component re-use capabilities are not yet defined in CellML. They will be a part of a future version of CellML. The list of alternative biological entities metadata construct is provided now for use in the future.]

Biological entity metadata is defined using a CellML-specific element, **<cmeta:bio\_entity>**. A biological entity may be identified by name, database identifier, or both. Multiple database identifiers may be provided, but all except one must be marked “alternative”. The name of the biological entity is defined exactly as alternative names for CellML elements are defined (with the **<dcq:alternative>** element).

Each database identifier is stored in a **<cmeta:identifier\_scheme>** element that identifies the database. The CellML metadata specification will control names for certain encoding schemes (see below). The **<cmeta:identifier>** element may also be qualified by a **<cmeta:identifier\_type>** element. This element should have a value of “alternative” for all **<cmeta:identifier>** elements except for one, which is considered the primary identifier. This addresses a concern about allowing multiple database identifiers that might actually refer to different biological entities. Such an error may still occur, but marking all identifiers except one as “alternative” provides software a method by which to determine which identifier should be given precedence.

The CellML metadata specification will define the following encoding schemes:

- SWISS-PROT (SWISS-PROT protein database)
- GenBank (GenBank nucleic acid database)
- GO Consortium (Gene Ontology controlled vocabulary)
- OMIM (Online Mendelian Inheritance in Man catalog of human genes and genetic disorders)
- LocusLink (LocusLink genetic loci database)
- Unigene (GenBank non-redundant gene clusters database)
- URI (URI for a web resource providing info about the biological entity)

Model authors and authors of processing software are free to define additional encoding schemes, by putting the **<entity\_scheme>** element in an application-specific namespace. However, software claiming to be “CellML metadata compliant” is not required to recognize these schemes.

RDF containers can be used to indicate that a given CellML element is relevant for more than one biological entity. An **<rdf:Bag>** element can be used to indicate that the CellML element is relevant for an entire group of biological entities. An **<rdf:Alt>** element can be used to indicate that the CellML element can be relevant for one member of a group of entities. Note that the first member listed in the **<rdf:Alt>** element will be considered the preferred value. The use of the **<rdf:Bag>** element is shown in Figure 20. The use of the **<rdf:Alt>** element would be identical. “CellML metadata compliant” software will be required to recognize RDF containers in biological entity metadata. The use of RDF containers is preferred to simply repeating the **<cmeta:bio\_entity>** element because it removes all ambiguity about how the group of biological entities relates to the referenced CellML element.

Figure 20 demonstrates the definition of biological entity metadata. Using the **<rdfs:label>** element is an alternative to using the **<dc:title>** element: it is a human-readable title of the database value.



---

```

<rdf:RDF
  xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
  xmlns:dc="http://purl.org/dc/elements/1.0/"
  xmlns:dcq="http://purl.org/dc/qualifiers/1.0/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdf:Description rdf:about="#cellml_element_id">
    <cmeta:bio_entity>
      <rdf:Bag>
        <rdf:li rdf:parseType="Resource">
          <dc:title>calmodulin</dc:title>
          <dcq:alternative>CaM</dcq:alternative>
          <cmeta:identifier rdf:parseType="Resource">
            <cmeta:identifier_scheme>SWISS-PROT</cmeta:identifier_scheme>
            <rdf:value>CALM_HUMAN</rdf:value>
          </cmeta:identifier>
        </rdf:li>
        <rdf:li rdf:parseType="Resource">
          <dc:title>troponin C</dc:title>
        </rdf:li>
        <rdf:li rdf:parseType="Resource">
          <cmeta:identifier rdf:parseType="Resource">
            <cmeta:identifier_scheme>SWISS-PROT</cmeta:identifier_scheme>
            <rdf:value>PRVA_HUMAN</rdf:value>
            <rdfs:label>parvalbumin</rdfs:label>
          </cmeta:identifier>
        </rdf:li>
      </rdf:Bag>
    </cmeta:bio_entity>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 20: Recommended definition of biological entity metadata. The referenced CellML element represents the following group of proteins: calmodulin, troponin C, and parvalbumin (the protein identified by SWISS-PROT entry PRVA\_HUMAN). The calmodulin biological entity has an alternative name and a database entry. The troponin C biological entity is only identified by name. The PRVA\_HUMAN protein is identified by database reference and a human-readable name.

---

## 4.11 Mathematical Problem Type

The mathematical problem type is a classification of the type of problem encoded in the math associated with the model or model component. It is specified using NIST's [GAMS classification tree](#)<sup>2</sup>.

Mathematical problem type is defined using a CellML-specific element, `<cmeta:GAMS>`. Modellers are free to use a different controlled vocabulary for the math problem classifications by ... However, CellML Metadata compliant software is not required to recognize any classification scheme other than the GAMS tree.

Figure 21 shows the recommended definition of mathematical problem type metadata.

---

```
<rdf:RDF
  xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdf:Description rdf:about="#cellml_element_id">
    <cmeta:GAMS>
      <rdfs:label>1st order ODE- Initial Value Problem</rdfs:label>
      <rdf:value>I1a</rdf:value>
    </cmeta:GAMS>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 21: Recommended definition of the mathematical problem type metadata.

---

## 4.12 Description

Description metadata is a short description of the referenced resource.

Description metadata is defined with either of the Dublin Core description qualifier elements, `<dcq:abstract>` or `<dcq:tableOfContents>`. Use of the `<dcq:abstract>` element will probably be most common in CellML Metadata.

Figure 22 shows how to define description metadata.

## 4.13 Annotations

There are four types of annotations that are recognized by the CellML Metadata specification. Model authors are free to create additional types. However, CellML Metadata compliant software will not be required to recognize any annotation types except for the following four:

- **comment:** free-form comment of the person who coded the model into CellML.
- **limitation:** brief description of the limitations/scope of the content of the CellML element.
- **modification:** description of a change made to the content of the CellML element.
- **validation:** description of the level of validation of the content of the CellML element. This may be a code. Note that validation codes are unlikely to be interoperable.

---

<sup>2</sup><http://gams.nist.gov/>

---

```
<rdf:RDF
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="#cellml_element_id">
    <dcq:abstract>
      This element uses simple mass-action kinetics to describe the
      A + B <-> C + D reaction.
    </dcq:abstract>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 22: Recommended definition of the Dublin Core description metadata.

---

Each annotation also has creator and creation date metadata that refers to it.

Annotation metadata is defined using CellML-specific elements: **<cmeta:comment>**, **<cmeta:limitation>**, **<cmeta:modification>**, and **<cmeta:validation>**.

The author metadata associated with an annotation is defined exactly as the model builder metadata (Section 4.1), and creation date metadata associated with an annotation is defined exactly as the general creation date metadata (Section 4.5).

Figure 23 demonstrates the definition of comment and limitation annotations. Figure 24 demonstrates the definition of modification and validation annotations.

---

```

<rdf:RDF
  xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#">

  <rdf:Description rdf:about="#cellml_element_id">
    <cmeta:comment rdf:parseType="Resource">
      <rdf:value>This model does not include the data of Jones, et al.
      about the corresponding pathway in canine.</rdf:value>
      <dc:creator>
        <vCard:N rdf:parseType="Resource">
          <vCard:Family>PowerPuff</vCard:Family>
          <vCard:Given>Bubbles</vCard:Given>
        </vCard:N>
      </dc:creator>
      <dcq:created>
        <dcq:W3CDTF>2001-04-01</dcq:W3CDTF>
      </dcq:created>
    </cmeta:comment>
    <cmeta:limitation rdf:parseType="Resource">
      <rdf:value>
        This component is only valid for temperatures above 20 degrees C.
      </rdf:value>
      <dc:creator>
        <vCard:N rdf:parseType="Resource">
          <vCard:Family>Doo</vCard:Family>
          <vCard:Given>Scooby</vCard:Given>
        </vCard:N>
      </dc:creator>
      <dcq:created>
        <dcq:W3CDTF>2001-03-28</dcq:W3CDTF>
      </dcq:created>
    </cmeta:limitation>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 23: Recommended definition of comment and limitation annotation metadata.

---

---

```

<rdf:RDF
  xmlns:cmeta="http://www.cellml.org/metadata/1.0#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#">

  <rdf:Description rdf:about="#cellml_element_id">
    <cmeta:modification rdf:parseType="Resource">
      <rdf:value>Changed the equation for the sodium current.</rdf:value>
      <dc:creator>
        <vCard:N rdf:parseType="Resource">
          <vCard:Family>PowerPuff</vCard:Family>
          <vCard:Given>Bubbles</vCard:Given>
        </vCard:N>
      </dc:creator>
      <dcq:created>
        <dcq:W3CDTF>2001-04-01</dcq:W3CDTF>
      </dcq:created>
    </cmeta:modification>
    <cmeta:validation rdf:parseType="Resource">
      <rdf:value>Physiome level 2</rdf:value>
      <dc:creator>
        <vCard:N rdf:parseType="Resource">
          <vCard:Family>Too</vCard:Family>
          <vCard:Given>Shaggy</vCard:Given>
        </vCard:N>
      </dc:creator>
      <dcq:created>
        <dcq:W3CDTF>2001-03-28</dcq:W3CDTF>
      </dcq:created>
    </cmeta:validation>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 24: Recommended definition of modification and validation annotation metadata.

---