Meeting Minutes 8 October

Metadata in CellML

Author:

Melanie Nelson (Physiome Sciences Inc.)

Contributors:

Warren Hedley (Bioengineering Institute, University of Auckland)

Poul Nielsen (Bioengineering Institute, University of Auckland)

David Bullivant (Bioengineering Institute, University of Auckland)

1   Introduction

This document outlines a proposed metadata structure for CellML. It uses the Dublin Core elements, extended as recommended by Data Model Working Group of the Dublin Core Initiative in their document offering guidance on expressing the Dublin Core with RDF[1], as well as some CellML specific metadata.

2   Basic Features

The metadata requested by the modelers at Physiome Sciences is listed in Section 16 of the CellML Requirements document[2]. The crack CellML development team has decided that all metadata should be allowed on any CellML element. The official claim is that this was done to avoid cluttering the data model, but it is also possible that Melanie just couldn't face going through and making rational decisions about which metadata made sense for which elements. It was also decided that none of the metadata should be required. However, some metadata will be strongly recommended by the forthcoming CellML best practices document. Furthermore, a given CellML element can only have one "data model's worth" of metadata, i.e., it can only have one block of metadata.

3   Handling of Metadata

The following table matches each piece of metadata with a proposed method for handling it in CellML, and a multiplicity rule. The following items from the list in the requirements document are not included:

1. **Name** is now considered an identifier, not metadata.

2. **References** are discussed in a separate document.

3. **Classification** will be part of the ontology included in later versions of CellML.

---

[1]http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/

[2]http://www.cellml.org/private/fundamentals/requirements.html#sec metadata

Metadata	Handling	Multiplicity
Alias	DC:Title	0..*
Model Builder	DC:Creator (extended)	0..*
Species	CellML-specific RDF	0..*
Sex	CellML-specific RDF	0..1
Description	DC:Description	0..1
CellML-coding date	DC:Date (extended)	0..1
Model-builder comments	CellML-specific RDF	0..*
Limitations	CellML-specific RDF	0..*
Biological entity	CellML-specific RDF	0..*
Copyright	DC:Rights	0..* (best practice: exactly one)
Publisher	DC:Publisher	0..*
Modification	CellML-specific RDF	0..*
Derivation	CellML-specific RDF	0..*
Mathematical problem type	CellML-specific RDF	0..*

4   Dublin Core Extensions

The Dublin Core Creator and Date elements need to be extended to meet the CellML metadata requirements. The extensions proposed here are in accordance with the guidelines found in the Dublin Core Data Model Working Group's document offering guidance on expressing the Dublin Core with RDF[3].

The **Creator** element needs to be extended to include the following information about the model builder:

- surname

- first_name

- mid_initials

- email

- affiliation

- mailing address

  - type
  - street
  - city
  - territory
  - postcode
  - country

- phone number

  - type
  - number

This additional information can be added to the RDF serialization of the Dublin Core data model using value components (see section 3.2.7 of the Dublin Core Data Model Working Group document). There is currently no existing metadata standard that provides the info we need to store about people. Therefore, we

```
<rdf:RDF
    xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
    xmlns:crdf="http://www.cellml.org/2000/cellml/RDF"
    xmlns:dc="http://purl.org/dc/elements/1.0">

  <rdf:Description rdf:about="my_favorite_model.xml">
    <dc:creator>
      <rdf:Description>
        <crdf:surname>Nelson</crdf:surname>
        <crdf:first_name>Melanie</crdf:first_name>
        <crdf:mid_initials>R</crdf:mid_initials>
        <crdf:email>mnelson@physiome.com</crdf:email>
        <crdf:affiliation>
          Physiome Sciences, Inc.
        </crdf:affiliation>
        <crdf:mail_address
              crdf:city="Princeton"
              crdf:country="USA"
              crdf:postcode="08540"
              crdf:street="307 College Road East"
              crdf:territory="NJ"
              crdf:type="work" />
        <crdf:phone_number
              crdf:number="609-987-1199 x245"
              crdf:type="work" />
        <crdf:phone_number
              crdf:number="609-987-9393"
              crdf:type="fax" />
      </rdf:Description>
    </dc:creator>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 1: An example extension of the Dublin Core Creator element.

can create elements in our CellML-specific RDF schema to represent this information, and mix this into the Dublin Core Creator element, as shown in Figure 1.

Note that it is possible for a CellML element to have more than one Creator. In this case, it is up to the model builder to decide whether to simply repeat the **<dc:creator>** metadata element, or to group the creators into a bag (unordered collection of objects) or sequence (ordered collection of objects). If the user wishes to indicate that multiple people worked together to create a model, but does not want to order the authors, he should group the creators into a bag. If the user wishes to make a statement about the order of the authors, he should use a sequence. If the user does not wish to indicate that the creators worked together as a group (for instance, if three people worked on a model at three different points in time), he should simple repeat the **<dc:creator>** element.

The **Date** element needs to be extended to include the meaning of the date. The original requirements only included an original coding date. However, the crack CellML development team sees no harm in adding a last modified date. The Date element can be extended to include this type information, as shown in Figure 2.

```
<rdf:RDF
    xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
    xmlns:crdf="http://www.cellml.org/2000/cellml/RDF"
    xmlns:dc="http://purl.org/dc/elements/1.0"
    xmlns:dcq="http://purl.org/dc/qualifiers/1.0">

  <rdf:Description rdf:about="my_favorite_model.xml">
    <dc:date
        dcq:dateScheme="W3CDTF"
        dcq:dateType="created">2000-10-05</dc:date>
    <dc:date
        dcq:dateScheme="W3CDTF"
        dcq:dateType="modified">2000-10-08</dc:date>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 2: An example extension of the Dublin Core Date element.

5   Extra Dublin Core Elements

The Dublin Core element set also contains the following elements, which we cannot prevent users from including:

- Subject
- Contributor
- Type
- Format
- Identifier
- Language

---

[3]http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/

- Relation
- Coverage

6   CellML-specific RDF

All metadata that does not map to the Dublin Core elements will be defined using a CellML-specific RDF schema. The following metadata can be implemented as simple RDF elements in the CellML-specific RDF schema's namespace:

- **Species**: The recommended best practice is to use the scientific name for the species, as this is unambiguous.

- **Sex**: The recommended best practice is to choose from the following values: male, female, either, unknown.

- **Biological entity**: This area of the metadata will almost certainly be expanded in future versions of CellML, but for now it is simply a name of a biological entity, such as a cell or protein, that a model or model component represents.

- **Mathematical problem type**: the recommended best practice is to choose from the GAMS classification tree.[4]

The following metadata can all be considered annotations on the model or model element. The RDF element structure for these pieces of metadata will be handled with one type of element: annotation. The three types of annotations that are currently defined are listed below. A given CellML element can have multiple RDF annotation elements of the same type. Each annotation element has a Creator and a CreateDate associated with it.

- **Model builder comments**
- **Limitations**
- **Modifications**

An example RDF expression of the CellML-specific metadata is shown in Figure 3.

The final type of CellML-specific metadata is listed below. It also has internal structure, as described, and an example is given in Figure 4.

- **Derivation**: Each derivation has the name of the parent element, a boolean flag indicating whether or not any modification has been made to the parent element, and an optional comment associated with it. There can be multiple derivations associated with a given CellML element (usually a model or model component). Multiple derivations will in general be independent of each other, and can therefore be represented in the RDF by repeating the derivation element structure.

This document covers the important aspects of the metadata data model, but does not formally specify that model. A full RDF Schema will follow shortly.

---

[4]http://gams.nist.gov/

```
<rdf:RDF
    xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
    xmlns:crdf="http://www.cellml.org/2000/cellml/RDF"
    xmlns:dc="http://purl.org/dc/elements/1.0"
    xmlns:dcq="http://purl.org/dc/qualifiers/1.0">

  <rdf:Description rdf:about="my_favorite_model.xml">
    <crdf:species>Cavia porcellus</crdf:species>
    <crdf:sex>male</crdf:sex>
    <crdf:bio_entity>HERG</crdf:bio_entity>
    <crdf:math_problem_type
         problemScheme="GAMS">I1a</crdf:math_problem_type>
    <crdf:annotation annotationType="limitation">
      <rdf:Description>
        <crdf:annotText>
            This model is only valid on alternative Sundays
        </crdf:annotText>
        <dc:creator>
          <rdf:Description>
            <crdf:surname>Nelson</crdf:surname>
            <crdf:first_name>Melanie</crdf:first_name>
          </rdf:Description>
        </dc:creator>
        <dc:date
            dcq:dateScheme="W3CDTF"
            dcq:dateType="created">2000-10-05</dc:date>
      </rdf:Description>
    </crdf:annotation>
    <crdf:annotation annotationType="comment">
      <rdf:Description>
        <crdf:annotText>
            I really enjoyed writing this.
        </crdf:annotText>
        <dc:creator>
          <rdf:Description>
            <crdf:surname>Nelson</crdf:surname>
            <crdf:first_name>Melanie</crdf:first_name>
          </rdf:Description>
        </dc:creator>
        <dc:date
            dcq:dateScheme="W3CDTF"
            dcq:dateType="created">2000-10-05</dc:date>
      </rdf:Description>
    </crdf:annotation>
    <crdf:annotation annotationType="modification">
      <rdf:Description>
        <crdf:annotText>changed conductance</crdf:annotText>
        <dc:creator>
          <rdf:Description>
            <crdf:surname>Nelson</crdf:surname>
            <crdf:first_name>Melanie</crdf:first_name>
          </rdf:Description>
        </dc:creator>
        <dc:date
            dcq:dateScheme="W3CDTF"
            dcq:dateType="created">2000-10-08</dc:date>
      </rdf:Description>
    </crdf:annotation>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 3: An example of an RDF structure for the CellML-specific metadata.

```xml
<rdf:RDF
    xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
    xmlns:crdf="http://www.cellml.org/2000/cellml/RDF"
    xmlns:dc="http://purl.org/dc/elements/1.0"
    xmlns:dcq="http://purl.org/dc/qualifiers/1.0">

  <rdf:Description rdf:about="my_favorite_model.xml">
    <crdf:derivation modified="yes">
      <rdf:Description>
        <crdf:parent>
            my_least_favorite_model.xml
        </crdf:parent>
        <crdf:derivComment>
            I have no comment.
        </crdf:derivComment>
      </rdf:Description>
    </crdf:derivation>
  </rdf:Description>
</rdf:RDF>
```

FIGURE 4: The CellML-specific RDF extensions for describing model derivations.