

Metadata in CellML

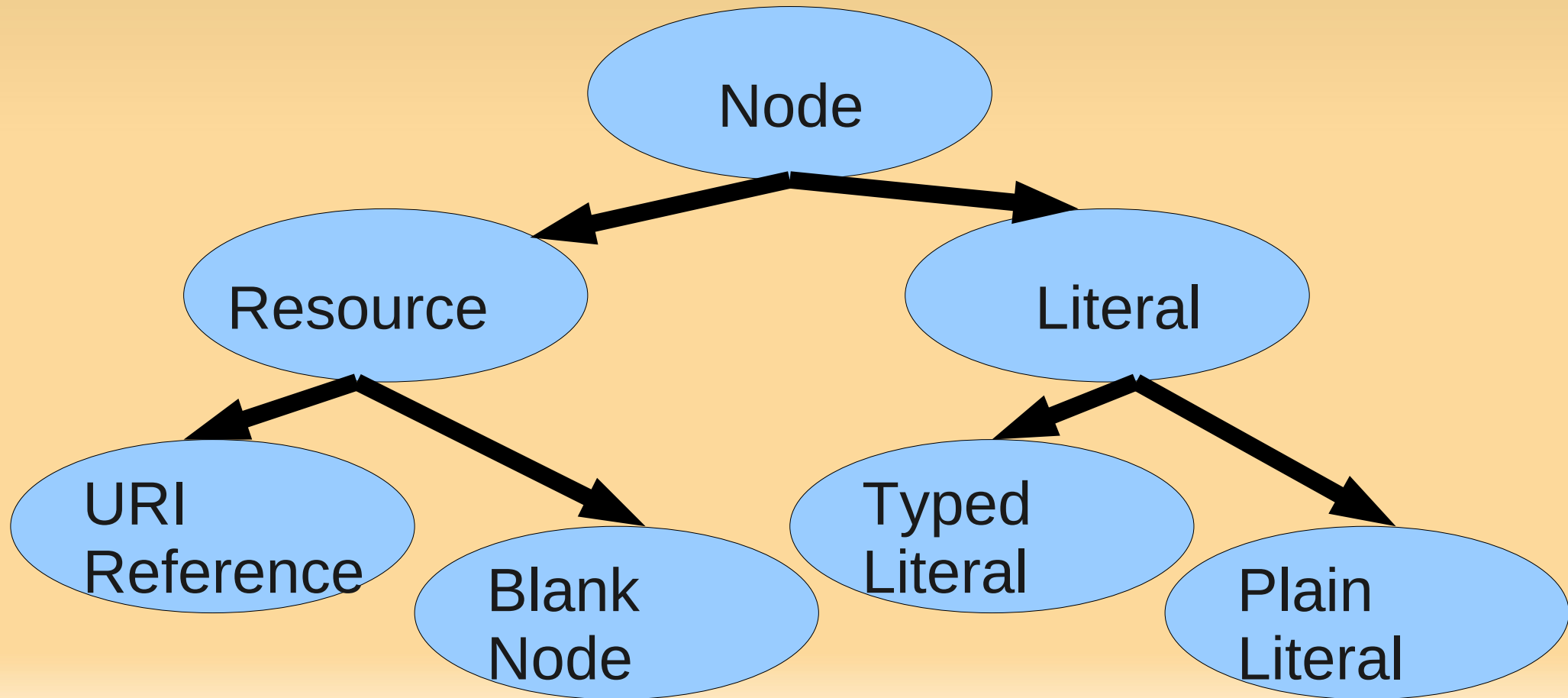
Andrew Miller <ak.miller@auckland.ac.nz> &
James Lawson <j.lawson@auckland.ac.nz>
Auckland Bioengineering Institute, University of Auckland

The current situation

- CellML Metadata 1.0 draft was written around 2001.
- There has not been much work on it since then, except for a recent restart of work.

Use of RDF

- CellML Metadata is encoded in RDF.
- In RDF, there are the following types of node:



RDF

- RDF describes everything as a triple. A triple is a statement of the form: Subject <Resource> Predicate <Resource> Object <Node>.
- Predicate is usually a URI reference from a controlled vocabulary. This is interpreted as saying 'the property described by the predicate takes the value in the object, for the subject'.
- Because the URI reference acts like a namespace, different specifications are unlikely to interpret the same predicate differently.

RDF & RDF/XML

- RDF itself is nothing more than abstract triples and nodes, from which other structures (like sequences) can be built. It is not a format.
- RDF/XML describes how RDF can be represented as XML. It provides a certain level of syntactic sugar to create complex structures involving blank nodes and containers, rather than directly listing off the triples.
- RDF/XML describing arbitrary RDF graphs be embedded in CellML and SBML models.

The Cmeta specification

- The CellML Metadata specification 1.0 came out before there was a clean separation between RDF and RDF/XML, and so is a bit antiquated.
- It describes how the `cmeta:id` is used on CellML elements as an identifier for URI References from the RDF. For example `<component cmeta:id="mycomponent" ...`, followed by a reference in metadata to `#mycomponent`

RDF triples in cmeta

- Cmeta 1.0 references a number of other specifications like Dublin Core to describe publications, and a draft vCard in RDF specification to describe people.
- It defines predicates for modification history, species, sex, bio-entity (which allows references to a number of databases), mathematical problem type, as well as free-form comment, limitation, and validation information.

Implementations

- The CellML repository makes use of the CellML metadata for publications and author descriptions.
- Generally speaking, support for the metadata has been limited to date. In particular, automatic 'semantic web' typed processing applications have yet to materialise.

Generality of metadata

- The current cmeta 1.0 specification allows for the same information to be represented in a number of different ways, potentially using several different specifications (e.g. Dublin Core vs vCard).
- This complicates its use.
- This is probably a bigger issue for representing biological entities like proteins.
- Some of these issues will be partially addressed as best practices for model annotation emerge.

Cmeta 1.1

- This is primarily a cleanup of cmeta 1.0.
- The document is being split up into a core specification, with additional specifications for things like bioentities and citations.
- This modularity means that we can more easily change or add to the metadata without changing the 'core' specification which covers the fundamentals.
- It is being developed in a public git repository, and everyone is welcome to contribute.

Questions and discussion

- Questions about cmeta 1.0 and 1.1.
- Discussions about metadata for models.