

## METAFOR Governance Plans Report

### METAFOR Deliverable 2.6 M42

<b>PROJECT</b>	
Project acronym	METAFOR
Project full title	Common <u>Meta</u> data for Climate Modelling Digital Repositories
Grant agreement no:	211753
Funding Scheme	Combination of Collaborative Projects & Coordination and Support Actions
Call Topic	INFRA-2007-1.2.1 Scientific Digital Repositories
<b>DOCUMENT</b>	
Deliverable	D2.6 Month 42
Deliverable Title	Governance Plans Report
Document Identifier	METAFOR-D2.6_M42
Date	6 <sup>th</sup> October 2011
Work Package	WP2 Coordinating the CIM definition
Authors	BADC
Document Status	Final
Document Link	<a href="http://metaforclimate.eu/documents">http://metaforclimate.eu/documents</a>

<b>Dissemination Level</b>		
PU	Public	x
PP	Restricted to other programmes participants	
RE	Restricted to a group specified by the Consortium	
CO	Confidential	

<b>Document History</b>			
Version	Date	Comment	Author/Partner
3.1	October 6 <sup>th</sup> , 2011	Final version for the EU	Bryan Lawrence/BADC

## Metafor CIM Governance

### Introduction

The “Metafor” Common Information Model (CIM) provides a formal structure for describing earth system simulations, and key elements of the accompanying workflow. It has been developed by an international consortium (Metafor) funded primarily by the European Commission – but with significant contributions from other partners – in particular those of the U.S. “Curator” project.

The CIM is encoded in the Unified Modelling Language (UML), and as such, is primarily a conceptual artefact. In practice the CIM is currently (2011) only instantiated as extensible markup language (XML) “documents” which conform to an XML application schema (the so called “ApCIM”, to be compared with the conceptual UML schema, the “ConCIM”). It is entirely possible that in the future there might be more than one serialisation of the ConCIM into application schemae (e.g. perhaps a variant of RDF-S<sup>1</sup>, or OWL<sup>2</sup>, or even some sort of set of rules as to how a JSON<sup>3</sup> document might be circumscribed and validated<sup>4</sup>)

The CIM (ConCIM and ApCIM) is accompanied by a set of controlled vocabularies which provide much of the key scientific content. One can think of the CIM as providing the structure, and the vocabularies, the dominant components of the language, needed to describe simulation.

Thus far the evolution of the CIM and the accompanying vocabularies have been governed by the Metafor project team. Now that the Metafor project has come to an end, it is necessary for the CIM and the accompanying vocabularies to find a new governance. Given the aspiration (and current reality) that the CIM be an international activity, the governance needs to be international. This document describes the proposed governance.

### What needs governing

In practice there are three artefacts (or groups of artefacts) that need governance:

1. The formal UML definition of the CIM (the “ConCIM” - including the metamodel, which encodes the rules as to how UML is used to construct the ConCIM). The only stable version in use at the moment is 1.5, stored in the metafor trac repository at

<http://metafortrac.badc.rl.ac.uk/trac/browser/CIM/tags/version-1.5>

A future version is under discussion, and thus there might be more than one stable version in circulation by the time the governance structure is fully developed. One would expect the governance activity gatekeep the release of stable versions.

1 RDF-S: Resource Description Framework Schema.

2 OWL: Web Ontology Language.

3 JSON: Javascript Object Notation, a format for encoding information, primarily to and from web browsers.

4 There is currently no widely accepted JSON schema, but a draft is available at <http://tools.ietf.org/html/draft-zyp-json-schema-03>

2. The formal XML schema version of the CIM (the “ApCIM”). There is only one version in widespread use at the moment, an XML schema document (XSD) corresponding to ConCIM 1.5. However, several variants of it potentially exist. It would be better were the provenance of these schema better controlled. This would involve controlling two more collections of information:
  - a) The “rules” for getting from the ConCIM to ApCIM (in any given serialisation), and
  - b) The official versions of the ApCIM (be they in XSD or any other document definition and constraint formalism).
3. The various vocabularies that together constrain the language that can be used in some CIM documents. These are currently encoded as mind-maps in [http://metafortrac.badc.rl.ac.uk/trac/browser/controlled\\_vocabularies](http://metafortrac.badc.rl.ac.uk/trac/browser/controlled_vocabularies) and as term lists within the Metafor CMIP5 questionnaire <http://metafortrac.badc.rl.ac.uk/trac/browser/cmip5q/tags/version-1.1/trunk/cmip5q/cmip5q/vocabs>

---

## For whom is the CIM governed?

---

The target audience of CIM governance is the scientific community who will use CIM descriptions of earth simulation, and the software developers who build the tools to create and consume CIM documents. Both constituencies will need to approve CIM changes, since the CIM needs to be scientifically relevant and usable in real software systems. Key scientific stakeholders include:

1. The European Network for Earth Simulation (as a representative body of the European climate modelling community),
  - There are good links to this community via the existing Metafor and is-enes projects.
2. The World Meteorological Organisation (WMO), via the Working Group on Coupled Modelling (WGCM)
  - It will be seen that the proposed governance will be embedded in this community.
3. The Numerical Weather Prediction Community
  - The current version of the CIM does not cover NWP, but future versions will. The NWP community already has aspirations to describe simulations, so bringing these activities together is desirable.
  - The WMO provides another route into this community, via the Expert Team on Metadata.
4. The Impact Community (being a catch all phrase to cover all users of simulation descriptions who are not themselves in the simulation community that produce and evaluate such).
  - As there is no obvious representative group, we can only anticipate approaching individuals from specific sub-communities.

The primary software specialists who need to be involved in the CIM governance are those in the climate community who build workflow systems, and those who build tools to manipulate and display the CIM itself.

---

## Governance Structure

---

The CIM shall be managed by a sub-committee of the Climate Forecast conventions committee (<http://cf-pcmdi.llnl.gov/governance>). Any change to the terms of reference and membership of the CIM-committee would need to be approved by the CF-conventions committee.

### Terms of Reference

The “Common information Model” committee will be responsible for the maintenance and development of standard descriptions of numerical experiments, earth system models, and simulations, comprising:

- A) A formal CIM description (currently encoded in UML) which may be serialised in a number of ways (currently, at least GML). To do this, the CIM committee will govern a “metamodel” which provides rules for how to construct the model (currently that would describe how to use UML), as well as the UML model itself.
- B) The set of approved “application schema” descriptions of the UML, suitable for use in software, along with the rulesets which govern their construction from the UML.
- C) A set of terms and scientific options which provide controlled vocabularies for possible and actual earth system model configurations.

### Membership

At least two representatives of

1. Major earth system modelling groups,
2. Those responsible for the technical development of CIM tools.

At least one representative of

3. The Numerical weather prediction community
4. Model simulation user communities unconnected to those who build and run models.
5. Model data archive centres.

ex officio:-

6. The chair of the CF governance panel
7. A representative of the WMO Inter-Programme Expert Team on Metadata & Data Interoperability
8. A representative of WCRP WGCM
9. A representative of the IPCC DDC
10. A representative of any organisation providing 0.25 (or more) FTE towards supporting the work of the committee.

A committee member may fulfil more than one of these roles (and in order to keep membership numbers down, individuals with dual roles will be desirable members.)

Where discussions require it, the committee may also co-opt members with specific scientific backgrounds and/or standards expertise. An ex-officio member may nominate someone else to take their role at any or all committee deliberations.

#### Normal Working Method

The committee shall elect a chair, from time to time, as necessary. The chair shall be expected to sit on the CF conventions committee.

The committee would normally expect to receive electronic proposals for change, and either vote on them directly, or where the matter is expected to be relatively non-contentious, delegate authority for decision making to a named individual or sub-committee.

The minimum number of members voting in quorum would be six.

More particularly:

- The day-to-day vocabulary maintenance and extension is expected to be one of the relatively non-contentious activities, which could be managed without recourse to the committee except where there is community disagreement (by analogy to standard\_names management by the CF standard names committee).
- The committee would normally have a standing sub-committee to technically vet proposals for the full committee, and provide a recommendation to the full committee. All matters which appear before the full committee should be decided by consensus, but where that is impossible, all members will have one vote. A tied vote shall be considered to be a vote for the status quo.
- Where possible meetings should be by teleconference or videoconference.

(The particulars may be revisited at any time without recourse to the CF-conventions committee.)