



Common Metadata for Climate Modelling Digital Repositories

Newsletter 2 – September 2009

This newsletter is the second in the series of quarterly publications of the Metafor project team, and aims to keep interested parties up-to-date with the recent developments in the project.

In this issue:

- *Interactions with climate scientists*
 - *the controlled vocabulary*
 - *the CMIP5 metadata questionnaire*
- *Changes in the Common Information Model (CIM)*

Interactions with climate scientists

A key part of the Metafor project is our interactions with climate scientists and climate modeling groups. This is a continuing process which ensures that the CIM is developed in such a way so that it meets the needs of those who will eventually be using it as a tool for comparing and discovering model metadata. Getting scientists involved with the project and the development of the CIM also means that the users of the CIM will have ownership of it when the Metafor project ends, and will be interested in governing the CIM development further.

In the past quarter, the interactions between the Metafor project team and climate scientists have focused on two key areas – controlled vocabulary and the CMIP5 metadata questionnaire.

The controlled vocabulary

The Metafor team have conducted a series of interviews with climate modellers in order to capture the controlled vocabulary that should be used both in the CMIP5 questionnaire and the CIM. The results of these interviews have been summarised in mind maps, which can be downloaded from the Metafor project site at <http://metaforclimate.eu/trac/wiki/ticket/198>.

The Controlled Vocabulary describes the terms that can be used to fill in the content of the CIM and the relationships among those terms. For example, while the CIM might describe a metadata structure that supports components with child components each of which has a componentType, the CV would constrain the metadata content by describing the terms that can be used for componentTypes (ie: "atmosphere," "ocean," "radiation," etc.) and which types can be child components (i.e.: an atmosphere can have a child radiation component, but a radiation component cannot have a child atmosphere



METAFOR is funded by the EU 7th Framework Programme as an e-infrastructure (project # 211753)

component). These constraints are handled separately from the CIM itself because it is expected that they may change according to the needs of the CIM users.

The use of mind maps, created interactively during the conversations with scientists, allows us to not only build up the lists of controlled vocabulary, but also build a structure for the way we collect information. These mind maps then feed directly into the questionnaire, which means that feedback from scientists about the content of the questionnaire can be done without them ever having to see a piece of xml/html or any other code – a far more user-friendly way of interacting.

All of the controlled vocabulary mind maps are available on-line at http://metaforclimate.eu/trac/browser/controlled_vocabularies/trunk/Software , where it is possible to interact with them, for example by opening and closing nodes to view the lower levels of the maps.

The CMIP5 questionnaire

The CMIP5 questionnaire is being developed to define and collect model and experiment metadata for CMIP5. It will be a web-based questionnaire developed using Django, which is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. As mentioned earlier, the structure, content and layout of the questionnaire is generated directly from the controlled vocabulary mind maps.

The questionnaire is currently in alpha3 stage of release. It is anticipated that the first beta release will happen at the end of September. If you are interested in beta testing and commenting on the draft questionnaire the project team would be very happy to hear from you!

Changes in the CIM:

The CIM is currently in version 1.1. Most of the work on the CIM has concentrated on updating the supporting documentation, as well as general housekeeping. A more intuitive way of describing model parameters has been incorporated into the software package; users can now categorise parameters as they want to instead of being required to conform to artificially imposed categories of "scientific," "numerical," or "technical." The data package has been integrated more closely with the activity package. Classes to track the "genealogy" of a document (ie: how it is related to prior documents) have been added. And users now have a choice of whether references to other CIM documents are described by pointing to a physically separate document or by embedding the referenced document's content within another document. A lot of time has been spent ensuring that the concepts used in the CIM are compatible with those being used by ESG as part of the Curator project.

A new release of the CIM is under way which will form Deliverable 2.3 of the EU project. The current version is available from the subversion repository on the Metafor website at: <http://metaforclimate.eu/trac/browser/CIM/tags/version-1.1>

Metafor at a glance:

Project title: **Common Metadata for Climate Modelling Digital Repositories (Metafor)**

Web site: <http://metaforclimate.eu>

Project coordinator: Dr Eric Guilyardi (University of Reading, UK and IPSL, France)
email: Eric.Guilyardi@locean-ipsl.upmc.fr

Project manager: Dr Sarah Callaghan (BADC- UK)
email: sarah.callaghan@stfc.ac.uk
tel.: +44 1235 445770
fax.: +44 1235 446140

Project participants:

| | |
|-----------|----|
| UREAD | UK |
| BADC | UK |
| CERFACS | FR |
| MPG | DE |
| CNRS/IPSL | FR |
| UNIMAN | UK |
| UKMO | UK |
| NMA | RO |
| MeteoF | FR |
| CLIMPACT | FR |
| PrinceU | US |