A Common Information Model paired with scientific Controlled Vocabularies for Climate Models and Statistical Downscaling

Charlotte Pascoe (1), Marie-Pierre Moine (2), Allyn Treshansky (3), Gerard Devine (4), Sebastien Denvil (5), Michel Kolasinski (6), and Rupert Ford (7)

(1) RAL Space, STFC, STFC, United Kingdom (charlotte.pascoe@stfc.ac.uk), (2) CERFACS, France (marie-pierre.moine@cerfacs.fr), (3) Coelancath Consulting, Ltd., United Kingdom (allynt@coelacanthconsulting.com), (4) University of Reading, United Kingdom (g.m.devine@reading.ac.uk), (5) IPSL, France (sebastien.denvil@ipsl.jussieu.fr), (6) CLIMPACT, France (ko@climpact.com), (7) University of Manchester, United Kingdom (rupert@manchester.ac.uk)

The results of climate models are now of more than purely academic interest: governments and the private sector also have a need to discover the results in order to prepare for and mitigate against the potentially severe impacts of global climate change. However, climate modeling is a complex process, which requires accurate and complete metadata (data describing data) in order to identify, assess and use the climate data stored in digital repositories.

The EU METAFOR Project (http://metaforclimate.eu) has developed a Common Information Model (CIM) to describe, in a standard way, climate data and the models and modeling environments that produced it. The CIM is constructed at a very high conceptual level using UML. It is then converted to an application-specific format using XSD. The structure of the CIM is highly generic so it can accommodate the wide degree of variability between different climate models. The generic structure of the CIM is paired with specific controlled vocabularies to create a metadata framework of valid CIM instances. The separation of CIM structure and controlled vocabulary gives the CIM the flexibility to support different communities (climate forecast, impact, etc.) through the development of domain specific controlled vocabularies.

This paper will present the CIM and the controlled vocabulary that has been collected by METAFOR to describe the climate models contributing to the fifth climate model inter-comparison project, CMIP5. It will show how the CMIP5 controlled vocabulary has been carefully decided upon through detailed consultation and feedback from a number of earth system domain specialists. It will also be demonstrated that the CIM is well designed (1) for application to a specific scientific domain once paired to an appropriate set of controlled vocabularies; (2) for hosting metadata collected through a web-based questionnaire; and (3) for the management of metadata instances and the interoperability with other ontologies (eg. Earth System Grid).

This paper will also demonstrate how the CIM and controlled vocabulary partnership is being exploited to extend the METAFOR methodology to statistical downscaling methods and thus address the metadata needs of the climate impacts community.