The CMIP5 Model Documentation Questionnaire: Development of a Metadata Retrieval System for the METAFOR Common Information Model

Charlotte Pascoe (1), Bryan Lawrence (1), Marie-Pierre Moine (2), Rupert Ford (3), and Gerry Devine (4)

(1) British Atmospheric Data Centre, Rutherford Appleton Laboratory, UK (Charlotte.Pascoe@stfc.ac.uk), (2) CERFACS, Climate Modelling and Global Change, Toulouse, France, (3) University of Manchester, UK, (4) University of Reading, UK

The EU METAFOR Project (http://metaforclimate.eu) has created a web-based model documentation questionnaire to collect metadata from the modelling groups that are running simulations in support of the Coupled Model Intercomparison Project - 5 (CMIP5). The CMIP5 model documentation questionnaire will retrieve information about the details of the models used, how the simulations were carried out, how the simulations conformed to the CMIP5 experiment requirements and details of the hardware used to perform the simulations. The metadata collected by the CMIP5 questionnaire will allow CMIP5 data to be compared in a scientifically meaningful way.

This paper describes the life-cycle of the CMIP5 questionnaire development which starts with relatively unstructured input from domain specialists and ends with formal XML documents that comply with the METAFOR Common Information Model (CIM).

Each development step is associated with a specific tool. (1) Mind maps are used to capture information requirements from domain experts and build a controlled vocabulary, (2) a python parser processes the XML files generated by the mind maps, (3) Django (python) is used to generate the dynamic structure and content of the web based questionnaire from processed xml and the METAFOR CIM, (4) Python parsers ensure that information entered into the CMIP5 questionnaire is output as CIM compliant xml, (5) CIM compliant output allows automatic information capture tools to harvest questionnaire content into databases such as the Earth System Grid (ESG) metadata catalogue.

This paper will focus on how Django (python) and XML input files are used to generate the structure and content of the CMIP5 questionnaire. It will also address how the choice of development tools listed above provided a framework that enabled working scientists (who we would never ordinarily get to interact with UML and XML) to be part the iterative development process and ensure that the CMIP5 model documentation questionnaire reflects what scientists want to know about the models.

Keywords: metadata, CMIP5, automatic information capture, tool development