



Harvesting metadata in the IMDI Earth System modelling environment

S. Legutke

Model and Data Group at the Max-Planck Institute for Meteorology, Hamburg, Germany
(stephanie.legutke@zmaw.de / Phone: +49-40-41173 104)

Earth System (ES) science projects undertaking experiments with numerical models increasingly involve large international scientific communities and in most cases many different ES models (ESM) often including components developed at different institutions. Metadata describing ESMs and experiments therefore become more and more important to allow scientists to evaluate model output with regard to differences in model formulation and experiments. Also, detailed descriptions of ESM experiments are needed in order to meet 'rules to ensure good scientific practices' as they are e.g. formulated by the German Max Planck Society or by the German Science Foundation. These rules say that sufficient and complete documentation that allows reproduction of data has to be accessible to people with vested interest for at least 10 years. 10 years is also the time primary data should be archived for later checks and control.

It is obvious that a huge amount of information is needed to meet these requirements and that the information should be human and machine readable. A popular format to store such information is xml formatted schemas which can easily be transformed into human readable html files that can be displayed by browsers. On the other hand, such xml formatted files, since they are machine readable, can easily be compared or processed. They are often used in data archives to make information in the archives browsable and searchable.

Filling the appropriate values into xml files is not a trivial task in particular if no graphical user interface exists through which the data values can be entered. This is not only due to the large amount of information required, but also since this information

is not always known to the scientists.

To help scientists with this task, the Model and Data group at the Max Planck Institute for Meteorology has decided to enable its modelling environment to write out all information known during the execution of ESM experiments. The modelling environment named 'Integrated Model and Data Infrastructure' (IMDI) defines standardized procedures to store model source code, configure and compile components, as well as to define and perform numerical experiments. The latter optionally include preprocessing of input data, model integration and graphical experiment monitoring, postprocessing of output data, as well as data archiving. The environment has e.g. been used to produce the contribution of the Max Planck Institute for Meteorology to the IPCC AR4 data archive with the ECHAM5/MPIOM model. It has also been used in the German ES modelling consortium effort to perform regional dynamical downscaling of global IPCC AR4 scenarios to the European region with the CLM regional climate model.

It is described in the presentation what kind of data are available for experiments within IMDI, how data from other metadata models can be included, and how they are harvested during the experiments.