



Common Metadata for Climate Modelling Digital Repositories Newsletter 5 – June 2010

In this issue:

- *CMIP5 metadata questionnaire update*
- *Statistical downscaling vocabulary*
- *CIM updates*
- *TDS2CIM tool*
- *METAFOR and Team Digital Preservation – a new cartoon*

CMIP5 metadata questionnaire update

The next few weeks see final preparations for releasing the CMIP5 questionnaire to our beta testers. Many thanks to people at the UK Met Office, CSIRO, JAMSTEC and CMCC for agreeing to beta test the questionnaire. New features in the latest version of the questionnaire include better support for the description of the model grids, and updates to the handling of simulations and ensembles. The information entered by our beta testers will be migrated to the final version of the questionnaire. Filling the model part of the questionnaire has a direct link with IS-ENES for the European ESMs groups as the result will be used to fulfill one IS-ENES deliverable (D4.2).

If you are interested in beta testing and commenting on the questionnaire then please email sarah.callaghan@stfc.ac.uk. The questionnaire can be found at <http://q.cmip5.ceda.ac.uk>. Selecting the Example modelling centre will allow you to explore the questionnaire without previously registering for a modelling centre account.

Statistical downscaling vocabulary

The controlled vocabulary METAFOR has established so far is devoted to the description of climate models, suitable for the GCMs that will run in the next IPCC assessment. METAFOR has now started to pay attention to regional climate data and the methods used to generate them. There are typically two different ways to produce climate projections at regional-scale. The first method consists of running a regional climate model (RCM) forced by large-scale reanalysis or global climate model (GCM) outputs at its boundaries, the so-called "dynamical downscaling" method. The alternative is to use "statistical downscaling" techniques for which a statistical relationship is established between some relevant large-scale variables (the predictors, usually reanalysis fields) and some small-scale variables of interest (the predictands, usually the observed fields) over a learning period (present time). This relation is then applied to climate projections performed by a GCM in order to know the evolution of the given small-scale variables over a given region in the future (e.g.: precipitations, wind, etc...).

Dynamical downscaling can easily be described through the metadata designed for the climate models, so we focus on statistical downscaling for which no high-level metadata exist yet. The work has been initiated within an initial set of experts whose groups are involved in METAFOR (CLIMPACT, CERFACS, Univ. of Cantabria). A first controlled vocabulary mind map draft has been produced and will be reviewed by these "internal" experts in order to identify the gaps in its ability to describe their own statistical method. Because of the wide range of method families in statistical downscaling ("transfer function", "weather typing", "weather generator", etc.), we will ask for external scientists to review this vocabulary for methods where no expertise exists within the METAFOR groups (for example in France: CNRM, CLIMATOR project, EDF).



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

METAFOR intend to address some of the needs expressed by the impact/adaptation communities with the metadata devoted to downscaling data. This work is consistent with the objectives of the WCRP/CORDEX project and collaboration with CORDEX will be explored.

CIM Updates

CIM v1.4 is currently available from the subversion repository on the METAFOR website at: <http://metaforclimate.eu/trac/browser/CIM/tags/version-1.4> . We anticipate one more version of the CIM being released before the end of the project. Many of the changes being made to the CIM are to ensure a clear match between the information provided by the Questionnaire and the structure provided by the CIM (while still allowing the CIM to remain general enough for user groups beyond CMIP5). The current outputs of the CMIP5 Questionnaire are actual CIM instances that conform to version 1.4.

Significant changes being made include:

- Added the concept of a "standardName" to identify certain CIM concepts using terminology from groups outside of METAFOR. A single element can have multiple standardNames.
- Several controlled vocabularies have been expanded as a result of the Questionnaire.
- Several concepts within the CIM have been made optional, as the Questionnaire did not use them, but it was thought that other groups might.
- Dates and calendars in the CIM have been reviewed. In this case changes to the questionnaire were made to match the CIM, rather than the other way around.
- Several new terms were added to DataHierarchyType. These correspond to DRS terminology and reflect how THREDDS categorises its data. Within a hierarchical set of DataObjects, the DataHierarchyLevelName and DataHierarchyLevelValue define the logic for grouping together DataObjects at that particular level.

TDS2CIM tool

All the CMIP5 data nodes will publish their model output data with the ESG publisher software. As a side effect of this publishing process Thredds records are generated and published in a web service container, the Thredds data Server, TDS. (For more information about Thredds please refer to: <http://www.unidata.ucar.edu/projects/THREDDS/>.) CIM metadata must be created for all the CMIP5 models, experiments, simulations and of course the data itself, and this metadata should feed into the CIM system and the CIM repository. The number of CMIP5 data output files will be very large - at least several 100,000 up to millions. It is impossible to do this metadata generation by hand.

A solution for this big issue is a useful way to access the TDS and capture as much as possible of the stored information (e.g. geospatial and temporal coverage, variables content, location) about the data files and create a CIM document describing the data. This is what the TDS2CIM tool developed by METAFOR does. The TDS2CIM tool is simply a python program. It uses several other python packages especially the lxml package which allows all kinds of XML access and manipulation.

METAFOR and Team Digital Preservation – a new cartoon

METAFOR are pleased to announce that they will be collaborating with the animation team who created Team Digital Preservation (<http://www.digitalpreservationeurope.eu/>) to produce a youtube cartoon to publicise the work of METAFOR and the CIM. Team Digital Preservation provides key messages about digital preservation as they battle the evil forces of Team Chaos in a series of thrilling adventures available from <http://www.youtube.com/user/wepreserve>

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: E.D.A.Guilyardi@reading.ac.uk

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
Univ.Cantabria	ES