Common Metadata for Climate Modelling Digital Repositories

Newsletter 7 – December 2010

In this issue:

- *CMIP5 metadata questionnaire launch*
- *Open Workshop about the CIM*
- *Statistical downscaling vocabulary*
- *CIM updates*
- *TDS2CIM tool*

CMIP5 metadata questionnaire launch

The CMIP5 metadata questionnaire was launched in Nov 2010 (http://q.cmip5.ceda.ac.uk), and is now in use by several of the CMIP5 modelling centres. Instructions for gaining access to the questionnaire can be found at that same address. The questionnaire has a "Test centre" area where users can experiment before filling out information in their own respective centre pages, and a read-only "Example centre" which gives examples of the sorts of information that is expected. As always, help is at hand by emailing the questionnaire support team at cmip5qhelp@stfc.ac.uk.

Open Workshop about the CIM

The Metafor open dissemination workshop **"Using the Metafor Common Information Model (CIM) to store, discover and locate climate modelling data"**, will be held in Abingdon, UK, on Monday 14th March 2011. The workshop is aimed at climate modellers and end users of climate data but also welcomes metadata experts.

If you have any questions, or would like to express an interest in coming to this workshop then please email sarah.callaghan@stfc.ac.uk . There is no charge for the workshop.

Statistical downscaling vocabulary

Even though METAFOR has been primarily focusing on global climate modelling, it aims to improve data documentation and understandability within all the research fields based on climate data analysis. Regional Climate Models (dynamical downscaling), can easily be described through the metadata framework developed for global climate modelling. The CIM is sufficiently generic to support metadata for statistical techniques and a metadata collecting tool is planned to be derived from the existing CMIP5 questionnaire.

The objective is to apply the metadata building procedure METAFOR used for large-scale Climate Modelling data to Statistical Downscaling (SD) data. The end-user communities targeted are the Impact and Adaptation communities. A standard metadata convention for SD will increase data interoperability and avoid misleading usage of the data.

This task will benefit other projects like the European project IS-ENES and the WCRP-sponsored CORDEX program. IS-ENES will rely on METAFOR expertise and developments for everything related to global and regional climate metadata. The CORDEX coordination program has among its objectives to provide a framework for the evaluation and intercomparison of regional downscaling models and methods and will define standards for production and dissemination of downscaled climate data. METAFOR will contribute to these standards.

A structured and hierarchical controlled vocabulary (CV) able to describe each Statistical Downscaling method needs to be built. The granularity of this CV (i.e. the level of details to be

Season's Greetings from the Metafor team!

The Metafor team would like to wish all our friends and colleagues

and a happy and prosperous 2011!

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

reached) is something to be agreed on relying on expert knowledge. The questions to be answered include:

- "How much do I need to know about the method that produced this downscaled data set to allow its unambiguous distinction from the others?"
- "What are the key parameters useful to scientifically decipher and analyse these downscaled data?"

The spread in the SD methods is large, and some combine several classical methods including: weather typing, weather generator, and transfer functions. That's why we are currently gathering inputs and opinion from a large number of SD experts. A first draft is available here:
http://metaforclimate.eu/trac/browser/controlled_vocabularies/trunk/Downscaling/StatisticalDownscaling.mm

CIM Updates

A final version of CIM v1.5 will be released shortly. This version will be used by all the METAFOR tools and services. Although work on the CIM will continue (in an effort to finally make the CIM fully GML-compatible), it will no longer impact the tools and services. The current version of the CIM, especially the Quality Package, has had some changes to make it more GML-compatible. These include incorporating the appropriate ISO datatypes into the CIM domain model.

The CIM, and the CIM Software Package in particular, has been extended to ensure that it supports the OASIS coupler. It is also being extended to ensure that it supports the Bespoke Framework Generator (BFG).

De-coupling the Controlled Vocabularies (CVs) from the CIM Schemas has continued. Python code has been written which takes a CIM document, finds all the XML nodes that are bound to CVs, extracts relevant information from those CVs, and generates a set of Schematron rules to be passed to the CIM validator. This code assumes that the format of the CVs will be SKOS; however, SKOS has some limitations in the complexity of relationships that it can encode and so the final format of CVs is still being decided. The "CV-checker" code will change as needed to support any new formats.

Grid information is finally being incorporated into the CMIP5 Questionnaire. Therefore, the CIM Grids Package is changing slightly to reflect the structure of the information being asked for by the Questionnaire.

TDS2CIM tool

The first publication of the Pylons-based Metafor portal was done in a demonstration to the Metafor and Curator community at the end of November. Inside the portal the TDS2CIM tool is used to generate CIM XML dataObject documents. Any TDS (Thredds Data Server) can be used to provide input to the TDC2CIM tool, but it must be registered in the portal. The registering process takes the url of the TDS, information about the institute, the responsible person and the harvesting time range. After the url is checked it will be used for capturing of CIM dataObjects. In the portal this is done by a background process, the so called 'process launcher'. In the latest version of the TDS2CIM tool all the functionality is controlled by an external configuration file which will be used be this portal 'process launcher'.

At the moment all generated XML will be inserted directly into the eXist database which is the storage place for the CIM repository. They are available immediately for search, validation and view and all further portal functionality.

A different way to provide these CIM XML documents to other end users is to publish all TDS2CIM output XML by the Atom feed protocol on an internal server. From that server XML documents could also inserted into the CIM repository by a background process.

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