



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

In this issue:

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

METAFOR and Team Digital Preservation – cartoon now launched!

METAFOR are delighted to announce the sixth Team Digital Preservation adventure, Team Digital Preservation and the Metafor Common Information Model, is now ready for viewing at

<http://www.youtube.com/watch?v=76MCRXK4Itc>

The screenshot shows a YouTube video player. The video title is "WePreserve and Metafor: Team Digital Preservation and the Metafor Common Information Model". The video is from the channel "wepreserve" and has 387 views. The video player shows a cartoon scene with a character wearing a yellow mask and a purple top with "CIM" on it, and another character wearing a white shirt with "D-M" on it. The video player interface includes a search bar, a video title, a duration of 2:29 / 5:08, and a view count of 387. There are also buttons for Like, Save to, Share, and Embed.

Blizzard and his team of chaos causing henchmen have escaped yet again and this time they have stolen all of the Climate Model Data being used to plan and build the Really Big Dam. The beautiful and flexible CIM comes to the rescue and makes a big impression on Digiman!

CMIP5 metadata questionnaire update

We are coming to the end of CMIP5 questionnaire beta testing – many thanks to our beta testers for all their feedback. In response to the feedback from our beta testers we have simplified the information we ask from users to describe their model forcing data. We now use descriptive free text to describe how simulations conform to specific requirements of CMIP5 experiments. It is no longer necessary to associate each requirement with input files and model modifications, instead this forcing information is gathered at the level of the experiment and it is up to the user to decide what forcing information to share. These changes place responsibility in the hands of users to choose the level of detail they wish to offer when describing the set up of their simulations for CMIP5.

The slim-line version of the questionnaire is due to be released to the CMIP5 community very soon. The questionnaire can be found at <http://q.cmip5.ceda.ac.uk> . Selecting the Example



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

modelling centre will allow you to explore the questionnaire without previously registering for a modelling centre account.

CIM Updates

The CIM v1.5 continues to evolve in response to 1) fine-tuning the CMIP5 Questionnaire, 2) exercising the Quality Package, and 3) separating out the Controlled Vocabularies from the CIM Schemas. Despite this, the overall structure of the CIM remains largely unchanged. This is on purpose. Once the Questionnaire goes live, significant changes to the CIM structure will require significant effort to re-map the Questionnaire output to a valid CIM XML document.

Changes to the CIM include refactoring the way that calendars and durations are recorded in both simulations and experiments. A simulation now "inherits" the calendar of the experiment for which it is being run. Its duration can either be recorded as a startPoint and a length of time to add to that point or an absolute endPoint. These must conform to any SpatioTemporal constraints imposed by the experiment.

Additionally, the Quality class has been given a "document" stereotype, which means that CIM Quality reports can exist as stand-alone XML documents (which reference other CIM documents). A tool for generating CIM quality reports is being developed.

Up until now the terms belonging to Controlled Vocabularies had been embedded within the CIM Schemas. A long-term goal has been to separate out these terms so that they can be governed independently from the CIM. To that end, the "value" attribute of a CIM CodeList is no longer bound to an enumerated type within the CIM. Instead, it can take any value and the CodeList must refer to a CV and a CV server location (those structures had been in the CIM since v1.2, but they hadn't been used until now). A separate piece of Python code is used to scan a CIM document and determine which CVs should be used and where in the document they apply. This information is used to generate a set of Schematron rules defining the constraints that the CV imposes upon the CIM document. This process will form part of the CIM Validation Service, which can already incorporate Schematron files into the set of constraints it checks.

TDS2CIM tool

The TDS2CIM is now in a 'pre-production' state. CMIP5 data is published by the ESG data node software. The publication process produces Thredds records (for more information about THREDDS please refer to: <http://www.unidata.ucar.edu/projects/THREDDS/>) and puts them onto a thredds data server (TDS). This thredds data server is parsed with a parser module (written in python script language). The python module for harvesting THREDDS catalogs, implemented in TDS2CIM, has been used for populating a database, which supports the query tool deployed in the IS-ENES portal. This parsed information is serialized and written to key/value pairs into a local database. From this database different output formats are possible. As output format the CIM dataObject is used for data ingest into the CIM search repository database. Also the CERA dataset format is implemented as output format. RDF or other could be easily implemented. All output formats are strictly separated.

The way to parse the TDS is to write it into a temporary storage and afterwards produce output. XML decouples the parsing process from the output generation process. So the output could be generated at any time and incremental output generation is available.

An implementation in a Pylons framework has been done in on a test server. It could be used to provide the tool abilities in the Metafor portal.

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