

METAFOR EC Dissemination Material Month 27 METAFOR Deliverable 7.3 M33

PROJECT	
Project acronym	METAFOR
Project full title	Common Metadata for Climate Modelling Digital Repositories
Grant agreement no:	211753
Funding Scheme	Combination of Collaborative Projects & Coordination and Support Actions
Call Topic	INFRA-2007-1.2.1 Scientific Digital Repositories
DOCUMENT	
Deliverable	D7.3 Month 33
Deliverable Title	METAFOR EC Dissemination Material Month 33
Document Identifier	METAFOR-D7.3_M33
Date	December 16th, 2010
Work Package	WP7 Training and Dissemination
Authors	BADC
Document Status	Final
Document Link	http://metaforclimate.eu/documents

Dissemination Level		
PU	Public	X
PP	Restricted to other programmes participants	
RE	Restricted to a group specified by the Consortium	
CO	Confidential	

Abstract

In the past 6 months, METAFOR has produced 2 more editions in a series of quarterly newsletters, aimed at publicising the work of the project team and encouraging interaction with other climate scientists and EU projects.

The project also collaborated with WePreserve to produce the sixth Team Digital Preservation adventure, Team Digital Preservation and the Metafor Common Information Model, which is freely available on Youtube.



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 channel is "wepreserve" with 9 videos and a subscribe button. The video duration is 2:29 / 5:08. The video player shows a cartoon with a female character in a yellow and pink outfit with "CIM" on her chest, and a character with a white chicken-like body and "D-M" on its chest. The video player interface includes a search bar, video title, channel name, video duration, resolution, and view count.

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



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 TDS2CIM 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 by 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 be 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

WePreserve and Metafor: Team Digital Preservation and the Metafor Common Information Model

We 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>

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!



Please feel free to make use of these animations as part of your own work to raise awareness and understanding about digital preservation.

Future animations will be released on our YouTube Channel at <http://www.youtube.com/user/wepreserve>



Metafor is funded by the EU 7th Framework Programme as an e-infrastructure (project # 211753) To learn more about the Common Information Model and the Metafor project, please visit <http://www.metaforclimate.eu>



WePreserve is committed to raise awareness about digital preservation issues among the widest possible audience and to breaking down barriers to access.

To learn more about WePreserve please go to

<http://www.wepreserve.eu/>