

## METAFOR EC Dissemination Material Month 27 METAFOR Deliverable 7.3 M27

| <b>PROJECT</b>      |   |
|---------------------|---|
| 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                                    |
| <b>DOCUMENT</b>     |   |
| Deliverable         | D7.3 Month 27   |
| Deliverable Title   | METAFOR EC Dissemination Material Month 18  |
| Document Identifier | METAFOR-D7.3_M27  |
| Date                | June 30 <sup>th</sup> , 2010  |
| Work Package        | WP7 Training and Dissemination  |
| Authors             | BADC  |
| Document Status     | Final   |
| Document Link       | <a href="http://metaforclimate.eu/documents">http://metaforclimate.eu/documents</a> |

| <b>Dissemination Level</b> |   |   |
|----------------------------|---|---|
| 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 has also created 2 new leaflets to publicise the work done on the CIM and on the CMIP5 metadata questionnaire.



## Common Metadata for Climate Modelling Digital Repositories Newsletter 4 – March 2010

In this issue:

- *Progress with the CMIP5 metadata questionnaire*
- *Grid controlled vocabulary*
- *CIM updates*
- *Plans for Metafor's final year*
- *New members of the Metafor team*

### *Progress with the CMIP5 metadata questionnaire*

Work on the CMIP5 questionnaire is continuing, and the questionnaire is currently in its beta4 version. Feedback from beta testers from modelling groups in the USA, France and UK was collected and used to improve the functionality and usability of the questionnaire.

The development paradigm for the CMIP5 questionnaire required that a working version be available at all times. The questionnaire has evolved through numerous iterations as bugs have been found and fixed, and the functionality has been extended and improved following feedbacks. Initially the regular Alpha deployments of the questionnaire were released only to the Metafor community, but in late 2009 the first Beta deployment was made available to the wider climate community. The iteration cycle of Beta deployments occurred on a longer timeframe than the Alpha deployments and substantial beta upgrades of the questionnaire were given new URLs. The long iteration cycle and the persistence of old versions of the questionnaire ensured that the beta testers were offered a stable product in which to test their model descriptions.

If you are interested in beta testing and commenting on the questionnaire then please let us know by emailing [sarah.callaghan@stfc.ac.uk](mailto: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.

### *Grid controlled vocabulary*

During the past quarter a special effort was done to reach a higher level of standardization for the vocabulary used to describe the numerical grids of the models. The objective is to make the model grid description independent of the description of model components themselves (atmosphere, ocean, land surface, sea ice, etc.) in order to fully exploit the CIM grid schema class capabilities and allow further standard queries for the search tools.

The work consisted of (i) extracting current information about horizontal and vertical grids from existing standards such as the gridSpec, CMOR and the CF convention and additionally from scientist interviews, (ii) expressing and organizing it in a generic way and (iii) identifying and resolving gaps and redundancies. This new structured and controlled vocabulary for grids forms a separated page of the CMIP5 Questionnaire, allowing a model component using the same (horizontal/vertical) grid as another to simply point to a pre-existing grid description. For example, in the situation where Land Surface uses the same horizontal grid as the Atmosphere, the user won't have to describe the horizontal grid twice.



Screen shot of the beta version of the questionnaire, which can be found at <http://q.cmip5.ceda.ac.uk>



## CIM updates

Significant changes from the last version of the CIM include:

- The use of IDs and versions within the CIM has been rationalised. A GUID type has been introduced which ensures unique instances. This, along with a version (internal and external) and a metadata schema ID and version are associated with a CIM Record. Other IDs are used throughout the CIM, but simple character strings are used rather than GUIDs (to correspond to how modelling centres name their products).
- The use of codelists to capture controlled vocabularies (CVs) has been expanded. In the soon to be released v1.4, a CIM codelist not only includes all of the vocabulary terms but also details on the vocabulary server. However, these details are not currently being used and a future version of the CIM will integrate external CVs with the CIM.
- CIM documents can reference other documents or parts of other documents. In the latest version of the CIM, a reference includes the concept of a modification. This allows the CIM to refer to a “base” document or element and describe how it has been changed in the current context. This is particularly relevant in light of how the CMIP5 questionnaire handles ensembles; ensemble members are treated as a set of changes made to a single simulation rather than a set of separate simulations. This makes the resultant CIM documents much more compact and efficient.
- The concept of conformance – as in a simulation must *conform* to particular experimental requirements – has been fleshed out based on examples coming from CMIP5. It is now possible to associate a set of numerical requirements with a simulation and describe how the simulation has been modified (or not, as the case may be) in order to meet those requirements. It is also possible to record that a simulation is purposefully non-conformant with certain requirements.

The output of the CMIP5 Questionnaire are actual CIM instances. Many of these changes are the result of ensuring that the CIM structure can accommodate the information provided by the questionnaire (while still retained enough generality to be useful for future user groups beyond CMIP5). The current version of the CIM is available from the subversion repository on the Metafor website at: <http://metaforclimate.eu/trac/browser/CIM/tags/version-1.3>

## Plans for Metafor's final year

The Metafor project recently had its year 2 All Partners meeting, which means that the project is now entering its third and final year of work. In year 3 the project will:

- Produce and host an open dissemination workshop (in Autumn 2010). If you would like to attend, please let us know by emailing [sarah.callaghan@stfc.ac.uk](mailto:sarah.callaghan@stfc.ac.uk)
- Finalise the CIM, and produce a plan for its governance and development past the end of the Metafor project.
- Develop and operate the tools and services such as the query tool and presentation tool to allow the CIM to be used in an effective and useful way by climate modellers.
- Release the CMIP5 questionnaire and support the questionnaire users in order to capture the metadata for the CMIP5 archive.

## New members of the Metafor team

Metafor are pleased to welcome Catalan Institute for Climate Sciences (IC3) and Allworlds Geothinking as subcontractors to the project. IC3 will be joining the work done in WP3 for service content creation and testing, and Allworld Geothinking will use model driven architecture techniques to ensure that the CIM is compliant with GML-UML encoding rules

## Metafor at a glance:

Project title: [Common Metadata for Climate Modelling Digital Repositories \(Metafor\)](#)

Web site: <http://metafordclimate.eu>

Project coordinator: Dr Eric Guilyardi (University of Reading, UK and IPSL, France)  
email: [Eric.Guilyardi@locean-ipsl.upmc.fr](mailto:Eric.Guilyardi@locean-ipsl.upmc.fr)

Project manager: Dr Sarah Callaghan (BADC- UK)  
email: [sarah.callaghan@stfc.ac.uk](mailto: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 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](mailto: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](mailto:E.D.A.Guilyardi@reading.ac.uk)

Project manager: [Dr Sarah Callaghan \(BADC- UK\)](#)  
email: [sarah.callaghan@stfc.ac.uk](mailto: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 |



... addressed the fragmentation and gaps in availability of metadata as well as duplication of information collection, and problems of identifying, accessing or using climate data that are currently found in existing repositories.

The main objectives were to:

- develop a Common Information Model (CIM) to describe climate data and the models that produce it in a standard way,
- to ensure the wide adoption of the CIM.
- develop, deploy, and evaluate a prototype infrastructure that allows key data and models to be discovered and compared between distributed digital repositories.



## The METAFOR Project

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](mailto:E.D.A.Guilyardi@reading.ac.uk)

Project manager: Dr Sarah Callaghan  
(BADC- UK)  
email: [sarah.callaghan@stfc.ac.uk](mailto:sarah.callaghan@stfc.ac.uk)  
tel.: +44 1235 445770  
fax.: +44 1235 446140

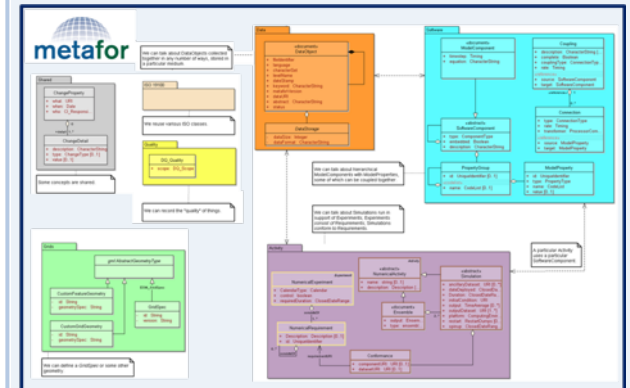
Project participants:

|                |    |
|----------------|----|
| 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

## The Common Information Model (CIM)



<http://metaforclimate.eu>

metafor@metaforclimate.eu



e-infrastructure



# METAFOR: The Common Information Model (CIM)

The CIM has been broken down into several packages:

- **Shared** – contains those elements that are used in many different packages.
- **Quality** – contains elements used to express diverse quality metrics for CIM metadata or the artefacts that metadata describes.
- **Grids** – provides a complete description of the horizontal and vertical discretisation of modelling elements: this may refer to grids that data is mapped onto, software adheres to and/or activities constrain.
- **Activity** – specifies the experimental design including the experimental requirements and descriptions of how simulations conform to these requirements.
- **Software** – specifies all the modelling software components used within the experiment process.
- **Data** – describes the data output from the climate modelling process as well as that for any input data.

... is a formal metadata model of the climate modelling process

... extends the traditional metadata description of climate data with the provenance of that particular data

... *re-uses* rather than *replaces* existing metadata systems and builds on existing metadata standards used in climate research

*An essential aim of Metafor is that the conceptual model is not changed by the manner in which it is used or applied.*

Climate modelling is a complex process with a wide degree of variability between different models and different modelling groups. To accommodate this, the CIM has been designed to be highly generic and flexible.

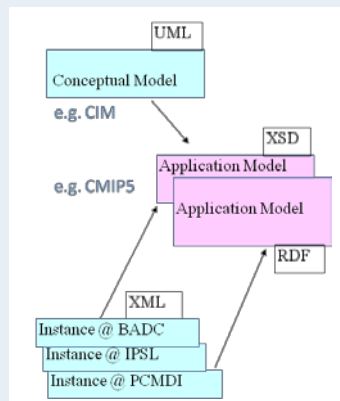
We describe the climate modelling process simply as "an activity undertaken using software on computers to produce data." This has been described as separate UML packages (and, ultimately, XML schemas).

## Conceptual CIM (ConCIM)

- An attempt to standardise our understanding of climate modelling at a very high level.
- A "common" model that all interested parties can agree on
- Written in UML
- Gets modified in response to user needs

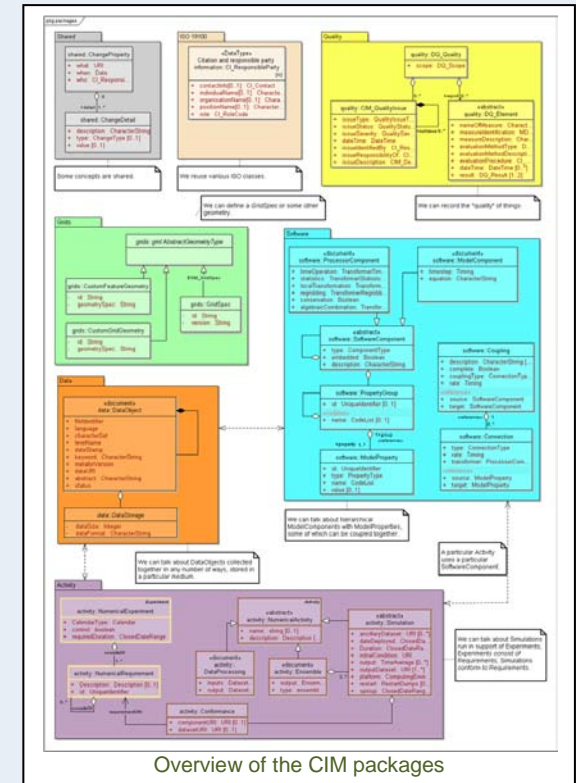
## Application CIM (AppCIM)

- An "application-specific" model (derived from the CONCIM) that is used for a particular community
- Can be written in XSD, RDF or other appropriate language



... includes descriptions of:

- the experiments being undertaken,
- the simulations being run in support of these experiments,
- the software models and tools being used to implement the simulations
- and the data generated by the software.



Overview of the CIM packages



More people than ever now have a need to discover the results of climate models in order to prepare for and mitigate against the potentially severe impacts of global climate change. But climate modeling is a complex process, which requires accurate and complete metadata (data describing data) in order to identify, assess and use the climate data stored in digital repositories.

The main objective of Metafor was to develop a Common Information Model (CIM) to describe climate data and the models that produce it in a standard way, and to ensure the wide adoption of the CIM. Metafor also developed, deployed, and evaluated a prototype infrastructure that allows key data and models to be discovered and compared between distributed digital repositories.



## The METAFOR Project

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](mailto:E.D.A.Guilyardi@reading.ac.uk)

Project manager: Dr Sarah Callaghan  
(BADC- UK)  
email: [sarah.callaghan@stfc.ac.uk](mailto:sarah.callaghan@stfc.ac.uk)  
tel.: +44 1235 445770  
fax.: +44 1235 446140

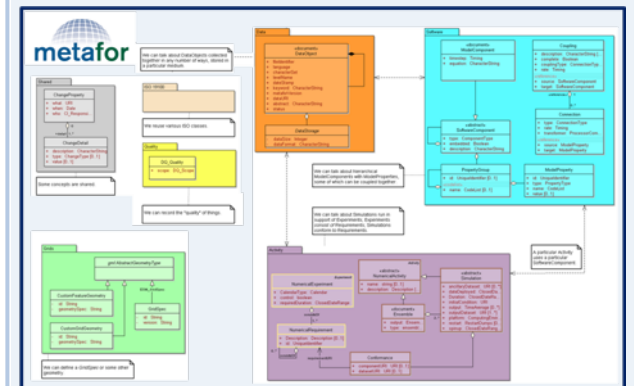
### Project participants:

|                |    |
|----------------|----|
| 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

## The Common Information Model (CIM) and the CMIP5 Metadata questionnaire



<http://metaforclimate.eu>  
metafor@metaforclimate.eu



# METAFOR: The Common Information Model (CIM) and the CMIP5 metadata questionnaire

## METAFOR

... has been charged by the Working Group on Coupled Modelling (WGCM) via the Coupled Model Inter-comparison Project (CMIP) panel to define and collect model and experiment metadata for the CMIP Phase 5 (CMIP5) project.

...has developed a web-based questionnaire to collect information and metadata from the CMIP5 modeling groups on the details of the models used, and how they conform to the CMIP5 experiment requirements.

The questionnaire is primarily a tool to document models in sufficient detail so that the CMIP5 data can be compared in a scientifically meaningful way.



e-infrastructure

## The CMIP5 questionnaire

...is an ambitious metadata collection tool and will help scientists to provide the most comprehensive metadata of any climate model inter-comparison project.

... allows users to enter descriptions of components which are not already specified by the questionnaire controlled vocabulary.

... XML output complies with the Metafor Common Information Model (CIM), allowing tools and services developed using the CIM to be applied to the questionnaire outputs

*The core archive of model data produced by CMIP5 will be used for the next Intergovernmental Panel on Climate Change (IPCC) assessment, due in 2013.*

In METAFOR, we have engaged with the climate modeling community by collecting controlled vocabulary from domain experts, via a series of interviews with climate modelers.

The results of the interviews are interactively summarised in mind maps, allowing us to not only build up the lists of controlled vocabulary, but also build a structure for the way the information is collected.

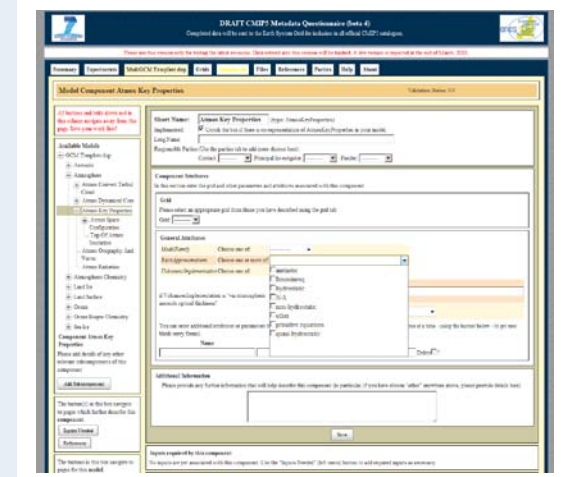
The controlled vocabulary lists and structure resulting from these interviews will eventually be governed independently from the Metafor project, providing a valuable resource for the climate modelling community.

The mind maps then feed directly into the questionnaire and feedback from scientists about the questionnaire content can be integrated quickly without exposing the questionnaire code.

...aims to collect enough detail to allow users to easily...

- browse the archive & find desired datasets
- easily differentiate between the "genealogy" (related models & experiments) of datasets

... will populate the CIM repository with CIM instances complying with the CIM ontology



Screenshot of the questionnaire