



Portable Infrastructure for the Metafor Metadata System











PIMMS provides institutions with tools to capture information about the workflow of running simulations

PIMMS provides institutions with tools to capture information about the workflow of running simulations from the design of experiments to the implementation of experiments via simulations running models. PIMMS uses the Metafor methodology for simulation documentation which consists of a common information model



The initial PIMMS deployment will be based on the controlled vocabularies collected by the Metafor project in support of CMIP5 (5th Climate Model Intercomparison Project). The CMIP5 controlled vocabularies will be extended to cover paleoclimate simulations by PIMMS partners at the University of Bristol.

(CIM), a set of controlled vocabularies (CV) and software tools. PIMMS software tools provide for the creation and consumption of CIM content via a web infrastructure and portal.

PIMMS will:

- Refactor the Metafor questionnaire for CMIP5 for use in university departments
- Create a new tool for defining experiments
- Engage with university community on usage (and CIM content)
- Connect data holdings at Bristol University and Reading

University to Metafor metadata.

• Promulgate the Metafor approach elsewhere in UK

Python Parser

A python parser processes the XML files generated by the mind maps

<component name="Radiation">

- <definition status="missing">Definition of component type Radiation required</definition> <parameter name="RadiativeTimeStep" choice="keyboard">
- <definition status="missing">Definition of property name RadiativeTimeStep required</definition>
- </parameter>
- <parametergroup name="Longwave"> <value name="Wide-band model"/> <value name="Wide-band (Morcrette)"/>
- <value name="K-correlated"/>

- </parameter>
- <value name="Two stream"/>





Mind maps are used to capture

information requirements from domain

experts and build a controlled vocabulary.

http://www.djangoproject.com

Django is used to generate the dynamic structure and content of the web based questionnaire from processed xml and the METAFOR Common Information Model (CIM)

PIMMS controlled vocabularies are recorded using mind maps. Mind maps not only collate lists of controlled vocabulary, but also provide a structure for the way the information is collected in the PIMMS web interface.

> PIMMS will provide a local portal so that research groups can view and search their own content, as well as publish their metadata content to institutional, national and international services. PIMMS will also include data node software so that data documented with PIMMS can be published to the web, both locally, and to national and international services.

CMIP5 Questionnaire

http://q.cmip5.ceda.ac.uk/

Summary Experiments Model: G	M Template dup Grid:Grid Template dup Simulat	Files References Parties He	Pap About Validation Status: 0.0
Noter Component Annos An		and and a second se	
All buttons and links above and in this column navigate away from this page. Save your work first! Available Models GCM Template dup Atmosphere Atmosphere Atmos Convect Turbul Cloud Atmos Dynamical Core Atmos Advection Atmos Advection Atmos Corography And Waves Atmos Radiation Atmos Radiation	Short Name: HiGEM_AtmosAdvec (type: A Implemented: ✓ Untick the box if there is no represent Long Name: HiGEM Atmospheric Advection compo Responsible Parties (Use the parties tab to add more choices I Contact: NCAS Principal Invest Grid Please select an appropriate grid from those you have descu Grid: TestGrid Vou can enter attributes or parameters here if you wish (to forms). Name	AtmosAdvection) tation of AtmosAdvection in your model. unent here): tigator: NCAS Funder: NCAS C tes associated with this component . ribed using the grid tab ts Value Value	Copy Parties to sub-components t a time - using the button below - to get new blank entry
Land Ice	Name	Value	Delete 2
Ocean Biogeo Chemistry Sea Ice Component Atmos Advection Please add details of any other relevant subcomponents of this component Add Subcomponent	Tracers SchemeName Choose one of: other SchemeType Choose one of: semi-Lagra ConservedQuantities Enter string value: ConservationMethod ConservationMethod Choose one or more of: conservation You can enter additional attributes or parameters here if yo blank entry forms). conservation	 angian on fixer ✓ u wish (to enter more than one, you need to save to sav	them one at a time - using the button below - to get new
	Name	Value	
The button(s) in this box navigate			Delete ?
component.	Momentum		
Inputs Needed References	SchemeName Enter string value: SchemeType Choose one of: ConservedQuantities Enter string value: O		
The buttons in this box navigate to pages for this model. Resolve Inputs Export Text View	ConservationMethod Choose one or more of: You can enter additional attributes or parametions blank entry forms). Name OMass-co Other	ian grangian onserving onserving / Finite volume (Lin-Rood)	hem one at a time - using the button below - to get new Delete?
Validate Export XML Send to CMIP5 (You can fix it later if ncessary)	Additional Information Please provide any further information that v		sen "other" anywhere above, please provide details here).
		Save	
	i nputs required by this component To inputs are yet associated with this component. Use the "Inp	outs Needed" (left menu) button to add required in	iputs as necessary.

UML overview of the Metafor Common Information Model (CIM). CIM XML documents are used to generate experiment descriptions and conformance pages in the web interface.

Screen shot from the CMIP5 model documentation questionnaire showing a metadata entry page for describing the Atmosphere Radiation component of a model.

The controlled vocabulary captured in the mind map interviews is used to populate drop-down lists in the web interface and the structure of the mind maps generates a navigation tree. The navigation tree shown on the left hand frame of the CMIP5 questionnaire screen shot, allows users to get to any component of the model description. The web interface also allows users to enter descriptions of components which are not covered by the mind maps. The mind map driven sections of the web interface ensure that we collect a standardised set of metadata about each model but if users wish to describe their models in more detail then the web interface is flexible enough to allow them to do it.

PIMMS will deploy a system for documenting the provenance of computer simulations of real world earth system processes into two university departments: at the University of Bristol and the University of Reading. This system will be a key part of the data management ecosystem for simulation data – without "PIMMS class" information simulation data can become effectively useless very soon after it is created. By extending and repackaging a prototype infrastructure developed by METAFOR and originally deployed to support CMIP5, PIMMS will be able to achieve significant leverage on previous R&D funding to provide a relatively mature system targeted at the university community.

Beyond PIMMS

Beyond the PIMMS project itself, the eventual goal of the work will be to see the PIMMS system implemented across multiple institutions and extended to support other scientific domains. With PIMMS in place, community based federated data infrastructures which support simulation data reuse, yet are based on institutional data holdings, can become a reality.

If you have any comments or questions about PIMMS or the CIM we'd love to hear from you! Further information about the PIMMS project can be found at http://proj.badc.rl.ac.uk/pimms/wiki Information about Metafor and the CIM can be found at http://metaforclimate.eu

The key to the customisation of PIMMS is in the modularity of its tools and the clear separation of structure (CIM) from content (CV). The PIMMS project will extend the CMIP5 controlled vocabulary to encompass descriptions of paleoclimate models and will also demonstrate how the CIM can be used to document an Integrated Assessment Model (IAM). This proof of concept prototype will create a new controlled vocabulary in collaboration with the Ermitage project and use it to reconfigure PIMMS to collect metadata in a different discipline. PIMMS will further explore how the CV that is used to configure PIMMS may be of further use to our stake holders and the wider JISC community through the development of the University of Cambridge chemicaltagger tool.

PIMMS will provide a local portal so that research groups can view and search their own content, as well as publish their metadata content to institutional, national and international services. In addition PIMMS will also include data node software so that data documented with PIMMS can also be published to the web, both locally, and to national and international services.