



# Common **Metadata for** Climate Modelling Digital Repositories

Gerard Devine (NCAS, University of Reading)  
and the Metafor Team

GO-ESSP, Hamburg, Oct' 2009

## Current issues in climate simulations



- Simulations have a key role in climate science in constructing understanding, and in producing predictions.
- Discriminating between two simulations is not easy, even when you were responsible for them!
- Documentation currently revolves around (at best) the runtime, but not the scientific detail and relevance of the model components.
- Little or no documentation of the “simulation context” (the whys and wherefores and issues associated with any particular simulation).

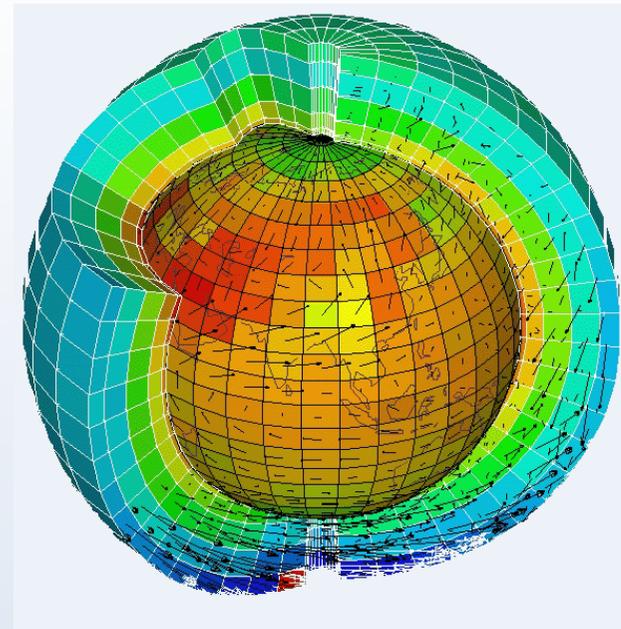
# Goals of Metafor



“The main objective of METAFOR is 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”

## Requirements for success:

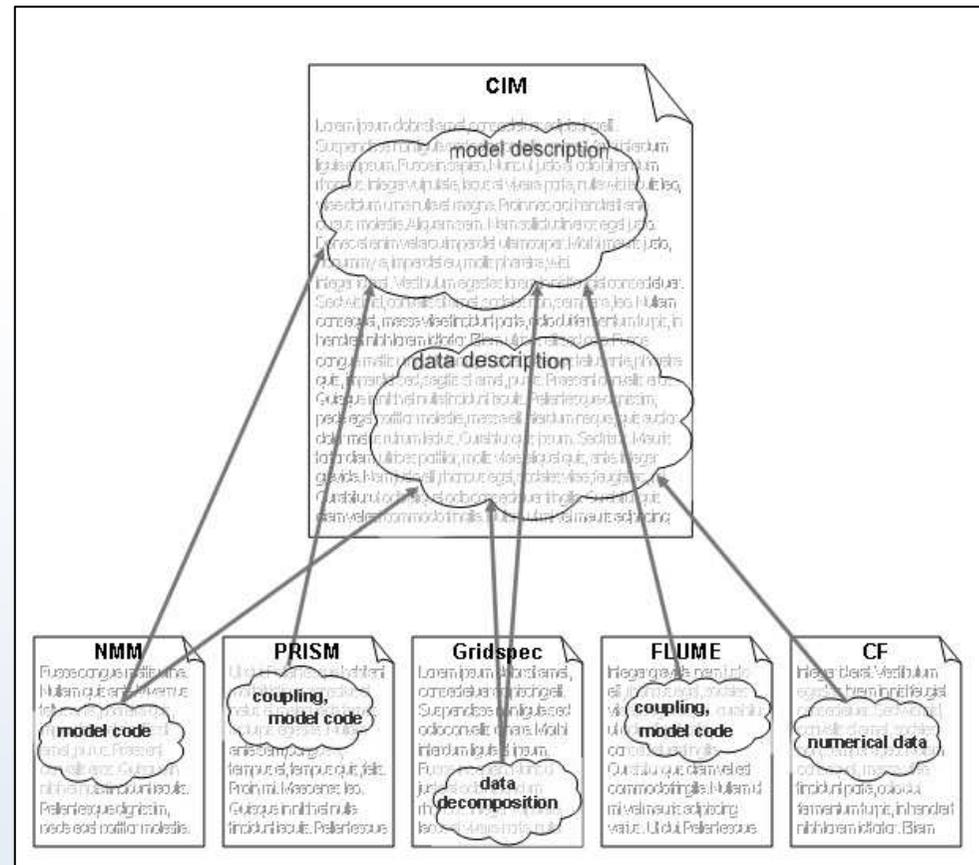
- Gather top field experts
- Engage with similar existing activities
- Work towards community adoption
- Capture wider community needs
- Produce Metafor EU deliverables



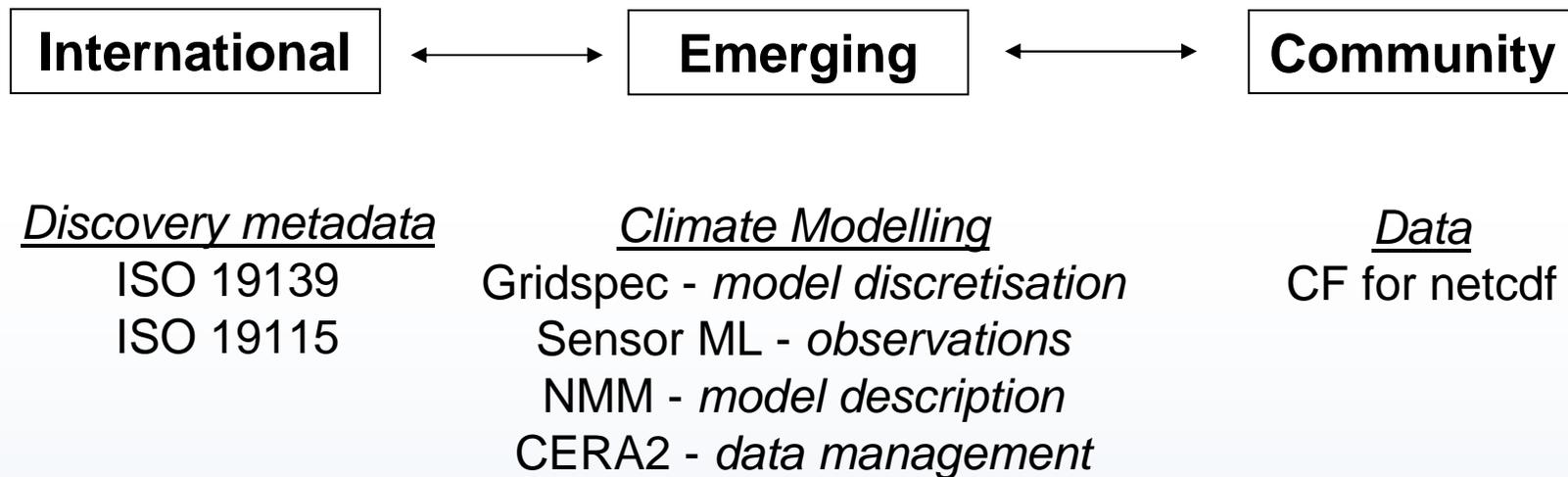
# Metafor objectives



- To allow essential data, model and experiment distinctions to be understood
- To build on existing metadata standards used internationally in climate (CF, CDML, CSML, Curator, NMM, FLUME, etc.)
- Use existing formats and frameworks (XML, RDF, etc.)



# Metafor Metadata Standards



**METAFOR** will coordinate the filling of the metadata gaps, mapping to different standards, aggregating the metadata and, if necessary, creating new standards.

# Metafor Metadata Standards



## Discovery metadata

ISO 19139

ISO 19115

## Climate Modelling

Gridspec - *model discretisation*

Sensor ML - *observations*

NMM - *model description*

CERA2 - *data management*

## Data

CF for netcdf

## Guiding Principles for metadata

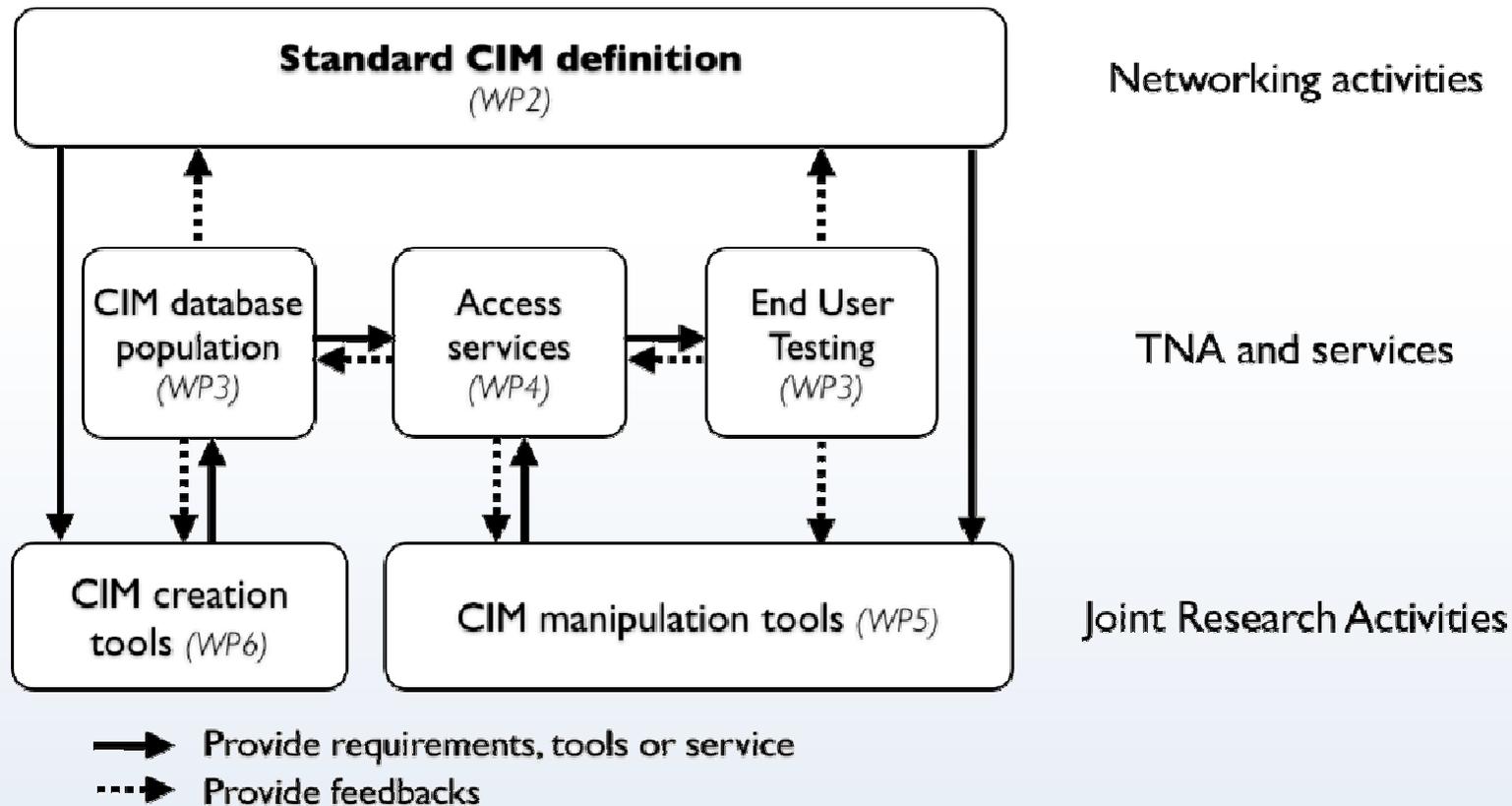
- integration of existing standards (ISO, climate modeling community,...)
- flexibility to support emerging standards both from within Metafor as well as from the broad community
- maintaining the “separation of concerns” (modularity)
- providing clear governance policies

# Metafor partners

- 11 partners
- EU contribution of 2.2M€
- Started March 2008, duration 3 years
- BADC, Science and Technology Facilities Council, UK
- CERFACS, France
- Models and Data, Max Planck Institute for Meteorology, Germany
- NCAS, University of Reading, UK (Coordinator)
- Institute Pierre-Simon Laplace, CNRS, France
- University of Manchester, UK
- Met Office, UK
- Administratia Nationala de Meteorologie, Romania
- Météo France, CNRM, France
- CLIMPACT, France
- CICS, Princeton University, USA



# Metafor Work Plan



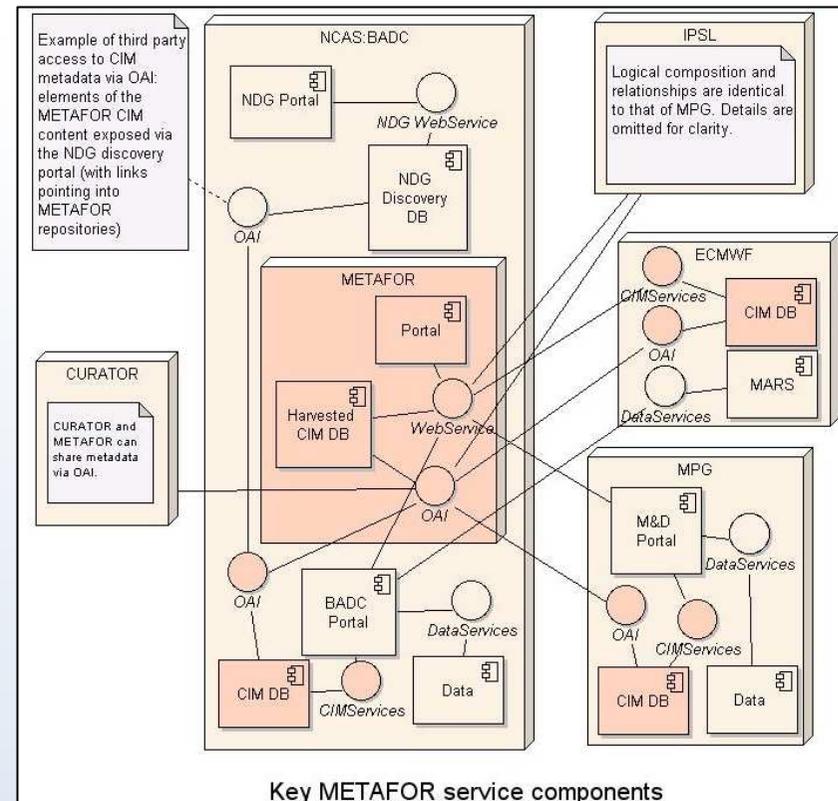
*Project management, training and dissemination are organised in WP1 and WP7.*

# Metafor Infrastructure



Develop, deploy, and evaluate a **prototype infrastructure** that will allow key **data** and **models** to be **discovered** and **compared** between **distributed digital repositories**

- single sign-on services to populate and manipulate, the CIM metadata
- services exploit NDG CSML to provide a common Geographic Markup Language interface to climate data
- centralized CIM content harvested from individual repositories using OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).



# CIM development



“The main objective of METAFOR is 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”

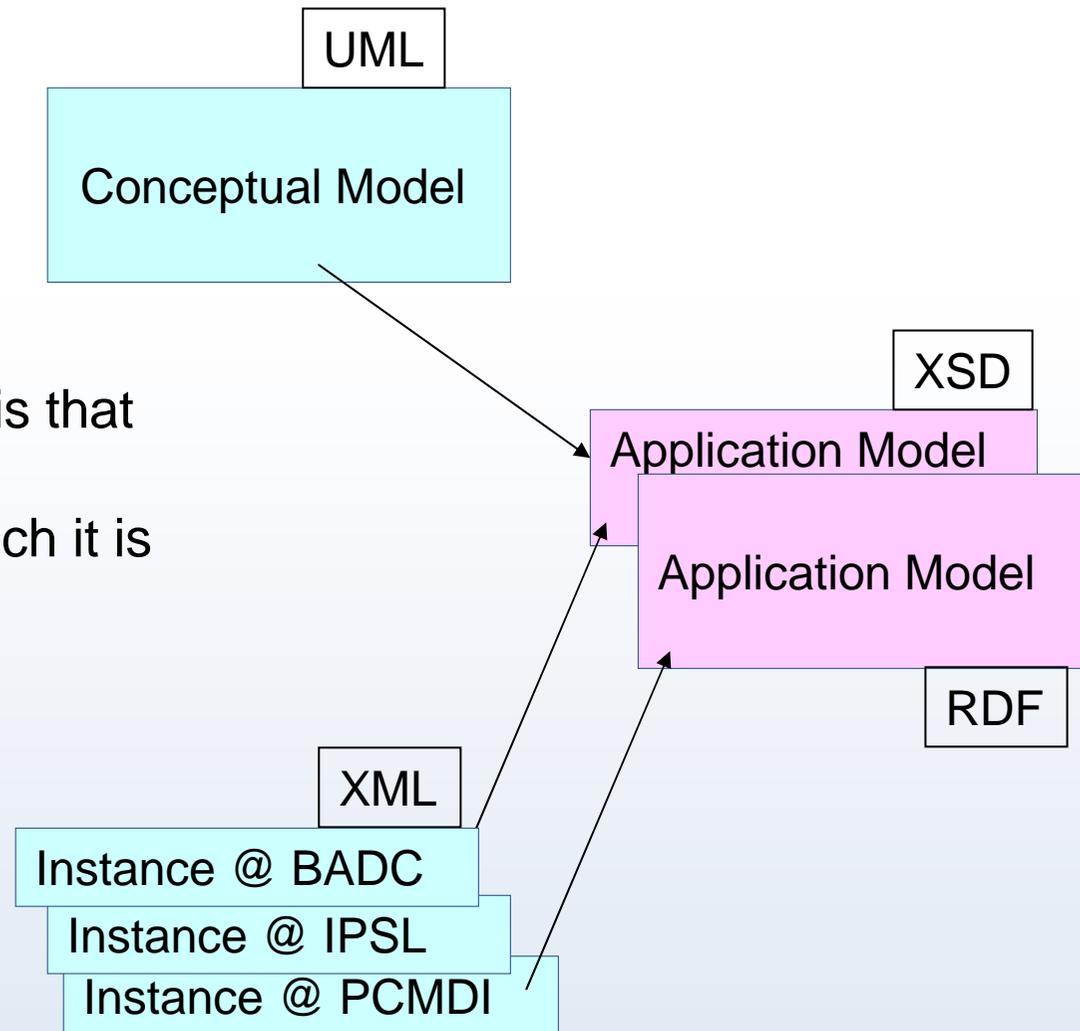
## **Goal :**

- One normative artefact – the UML model
- Derived XSD generated automatically

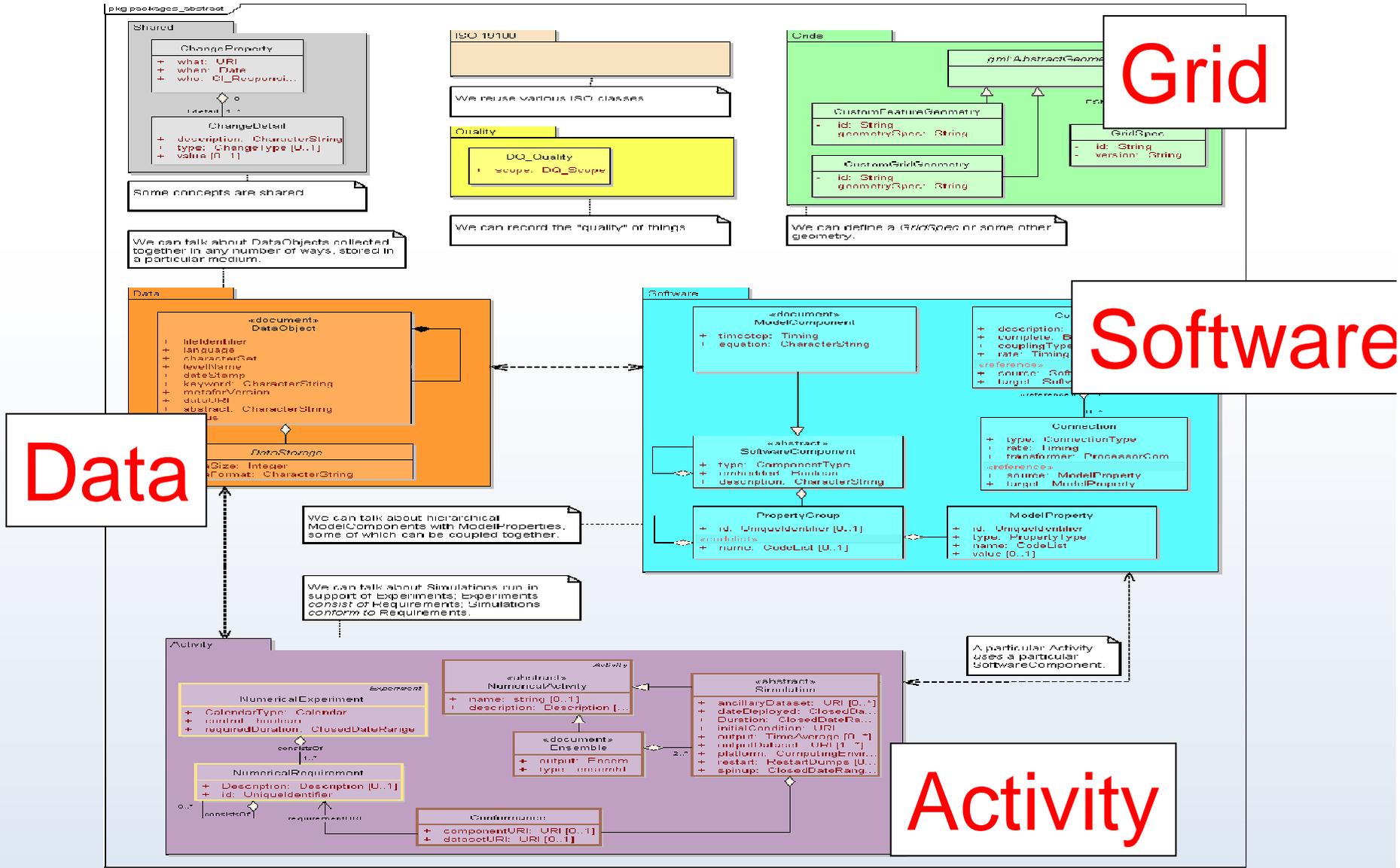
# CIM development



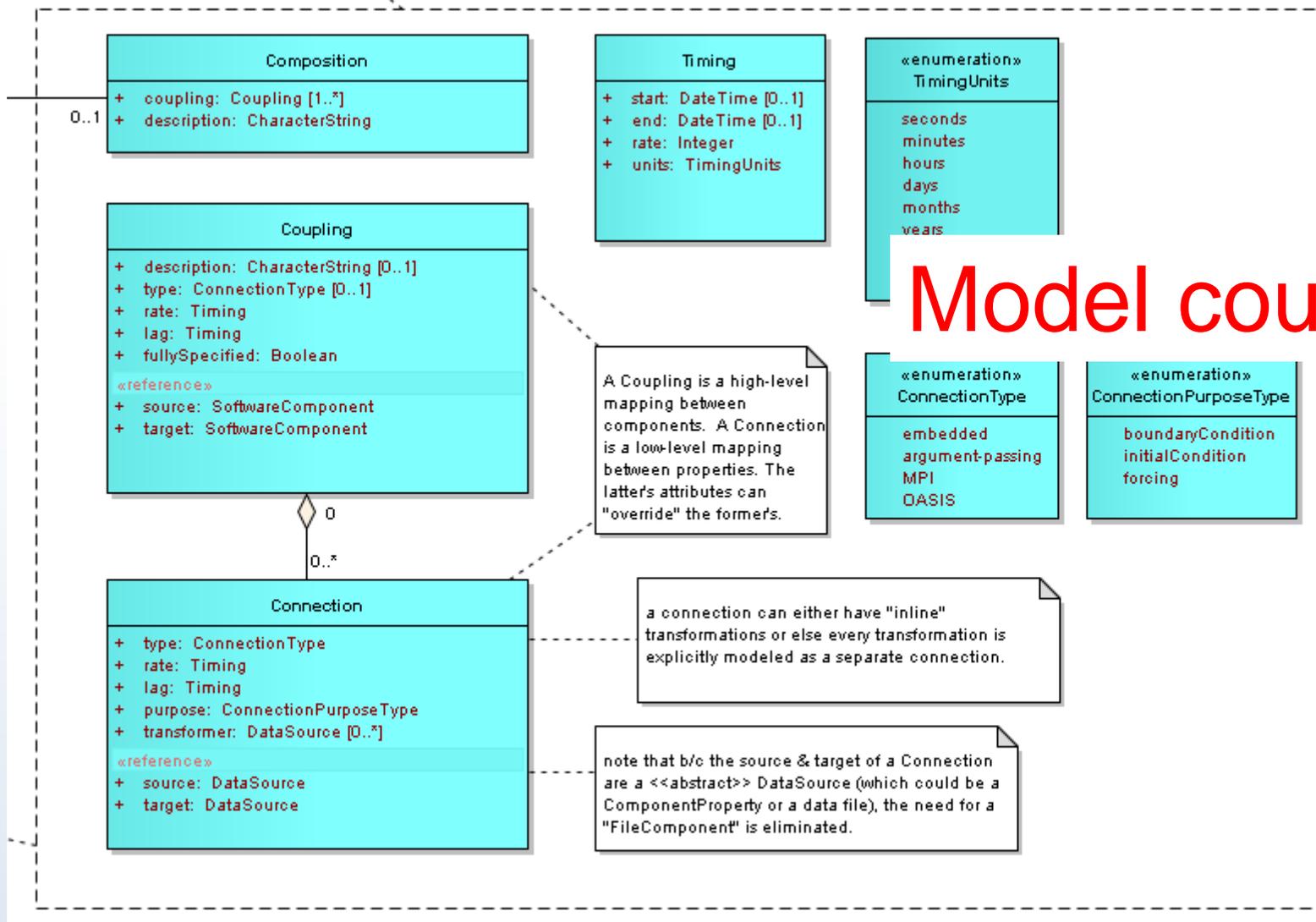
An essential aim of **Metafor** is that the conceptual model is not changed by the manor in which it is used or applied.



# CIM structure



# CIM software package



Model coupling

# Metafor highlights so far

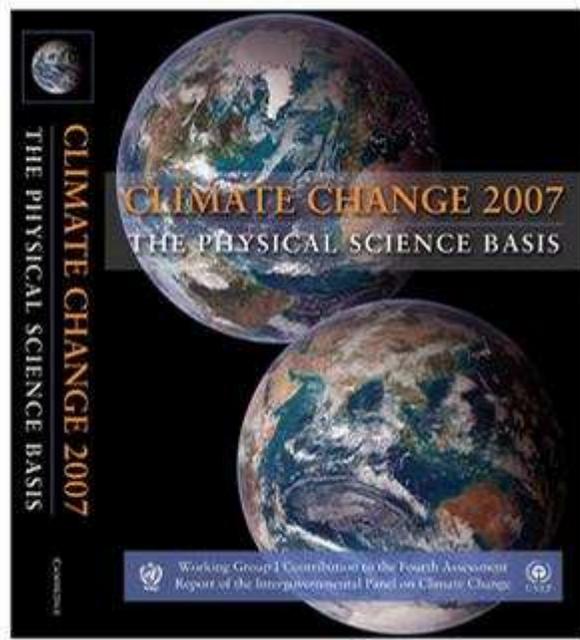


- A dedicated and tightly organised group of experts
- A methodology: CIM development strategy proposed, including conceptual level and meta-model
- A first CIM: v1.0 delivered, freely available at:

**<http://metaforclimate.eu/trac/browser/CIM>**

- Strong international collaboration and links established with USA colleagues in Curator/ESG/PCMDI
- A prototype portal deployed
- Strong community buy-in:
  - Leading the CMIP5 metadata collection
  - An inclusive mail list (~100/month)
  - Future wide-range dissemination planned to tie in with CMIP5 questionnaire and AR5 (the “stick” and the “carrot” !)

# CMIP5/IPCC



- The CMIP5 experimental archives will be ~500TB of model run data
- We need to be able to capture all the details of these experiments (and the component models and platforms used) to allow users of the archive to differentiate between the experiments and the models.
- To do this, Metafor has been tasked by WGCM/CMIP to produce a questionnaire to capture the model metadata.

# CMIP5 questionnaire



A screenshot of a web browser window showing the 'CMIP5 Model Metadata Questionnaire' for the 'UK National Centre for Atmospheric Science'. The browser title is 'CMIP5 Model Metadata Questionnaire - Mozilla Firefox'. The address bar shows 'http://localhost:8000/c mip5/1/'. The page header includes the 'metafor' logo and the tagline 'finding and understanding simulations of past, present and future climate'. Below the header is a navigation menu with links for 'Home:NCAS', 'GCM Template', 'Simulations', 'Files', 'References', 'Help', and 'About'. The main content area is titled 'Summary: UK National Centre for Atmospheric Science' and contains an 'Introduction' section. The introduction text states: 'Each CMIP5 modelling centre is running Simulations which run Models on Platforms. The Models are made up of Components. The Simulations conform to the NumericalRequirements of Experiments via what we call Conformances which consist of either specific code modifications or the use of specific boundary or initial condition Files. The purpose of this questionnaire is to glean information about the entities denoted thus and/or their relationships. We expect to see each centre enter at least one model, one platform, and then multiple simulations, each of which will involve entering descriptions of how they conform to the numerical requirements via conformances. It is not possible to start entering simulation information until at least one model and one platform have been created.' Below the introduction are three sections: 'Models associated with NCAS' with a table showing a 'GCM Template' with an 'Edit' button and a 'placeholder' status; 'Simulations associated with NCAS' with a 'Simulation Master Page' button; and 'Computing platforms associated with NCAS' with an 'Add a new Platform' button. A note states: 'Note that it can take some time to create a new model from the CMIP5 template ... be patient!'. Another note says: 'The status column provides an indicator of how much of the model description has been completed.' At the bottom right of the simulation section, it says 'Number of dataobjects and conformances listed here'. The browser status bar at the bottom left shows 'Done'.

.....to be discussed in a separate talk