

Portable Infrastructure for the Metafor Metadata System



Charlotte Pascoe¹, Gerry Devine²
¹ NCAS-BADC, ² NCAS-CMS University of Reading



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 (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:

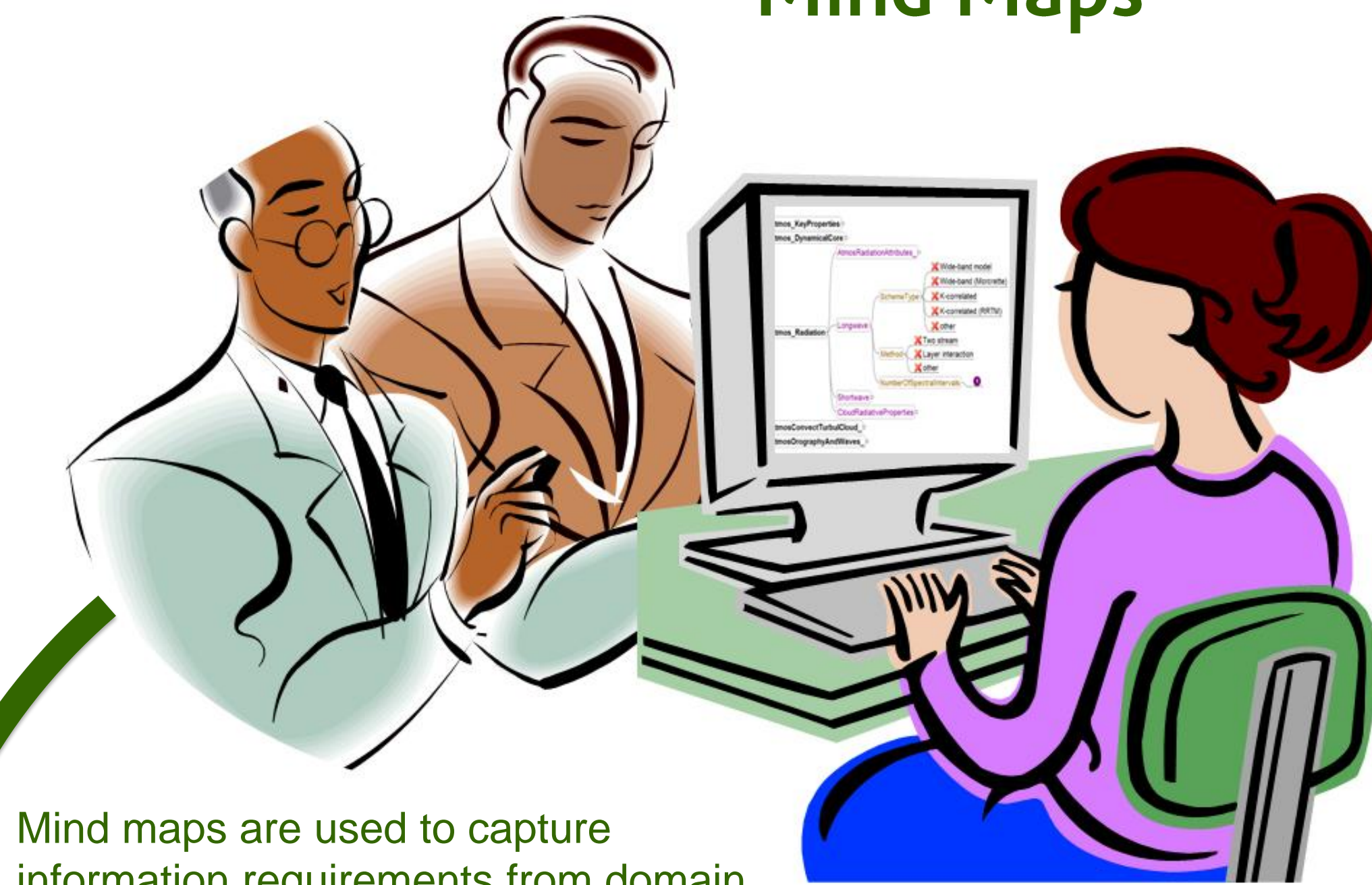
- Refactor the Metafor questionnaire for CMIP5 for use in networks
- Create a new tool for defining experiments
- Engage with university community on usage (and CIM content)
- Connect data holdings at Bristol University and Reading University to Metafor metadata.
- Promulgate the Metafor approach elsewhere in UK

Python Parser

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

```
<component name="Radiation">
  <definition status="missing">Definition of property Radiation required</definition>
  <parameter name="RadiativeTimeStep" choice="keyboard">
    <definition status="missing">Definition of property name RadiativeTimeStep required</definition>
    <value format="numerical" name="time step" units="time units"/>
  </parameter>
  <parametergroup name="Longwave">
    <parameter name="SchemeType" choice="XOR">
      <definition status="missing">Definition of property
      <value name="Wide-band model"/>
      <value name="Wide-band (Morcrette)"/>
      <value name="K-correlated"/>
      <value name="K-correlated (RRTM)"/>
      <value name="other"/>
    </parameter>
    <parameter name="Method" choice="XOR">
      <definition status="missing">Definition of property
      <value name="Two stream"/>
      <value name="Layer interaction"/>
      <value name="other"/>
    </parameter>
    <parameter name="NumberOfSpectralIntervals" choice="keyboard">
      <definition status="missing">Definition of property name NumberOfSpectralIntervals required</definition>
      <value format="numerical" name=""/>
    </parameter>
  </parametergroup>
</component>
```

Mind Maps



Mind maps are used to capture information requirements from domain experts and build a controlled vocabulary.

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.

PIMMS controlled vocabularies are recorded using mind maps. Mind maps not only collect lists of recorded 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 also be published to the web, both locally, and to national and international services.

CMIP5 Questionnaire

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



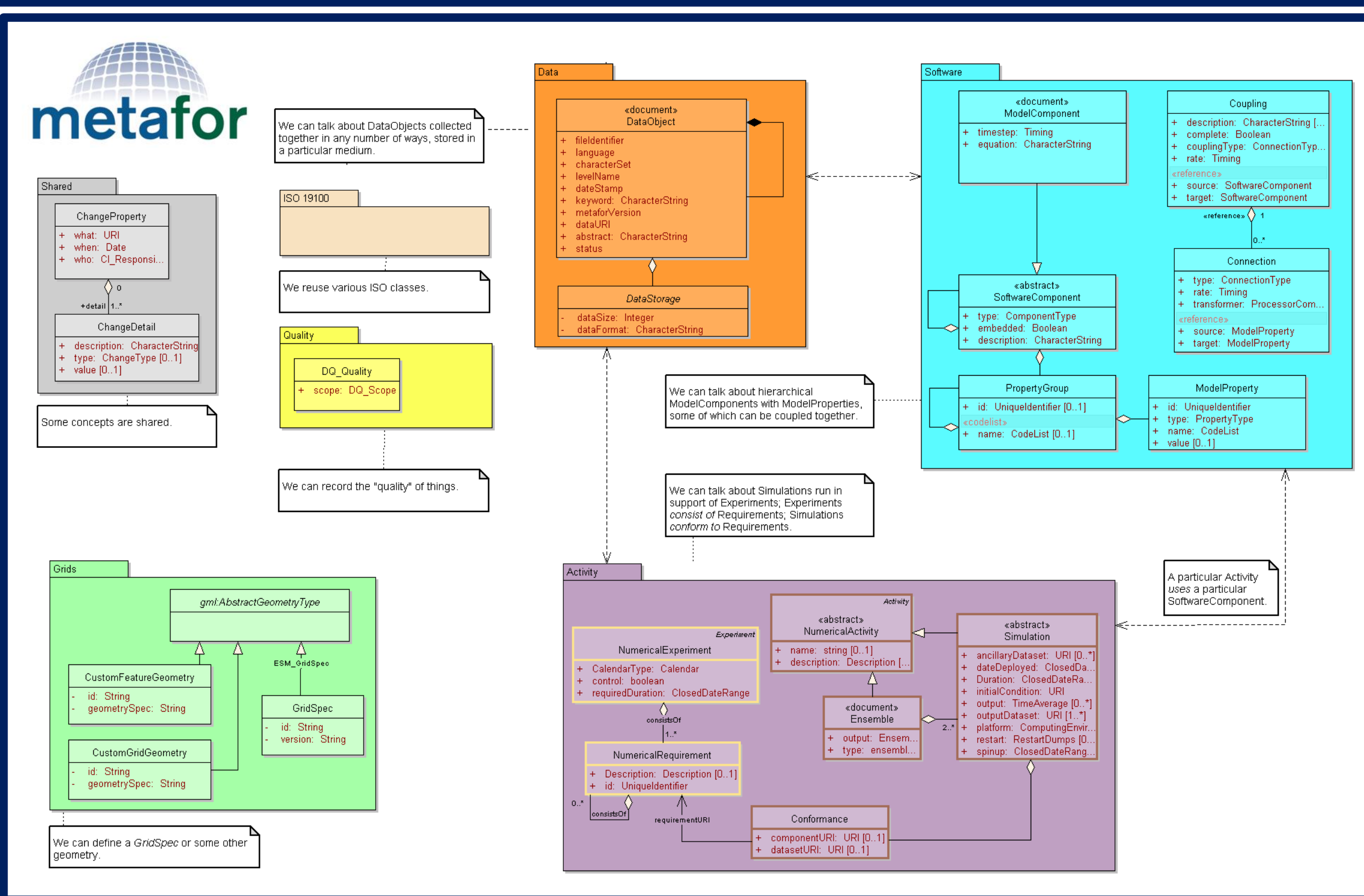
<http://www.djangoproject.com>

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)

Django is a high-level Python Web framework which makes it possible to build web applications with minimal code.

The CMIP5 questionnaire, on which PIMMS is based, makes much use of Django templates. The templates allow us to separate design, content and Python code.

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



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

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 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.

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 also demonstrate the CMIP5 controlled vocabulary to encompass descriptions of paleoclimate simulations and will extend the CMIP5 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 stakeholders and the wider PIMMS 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.