Metadata activities: ESMF and ESC

V. Balaji
Princeton University and NOAA/GFDL

PRISM/ESMF Metadata Meeting
Exeter UK
9 May 2006

ESMF's metadata-laden data structures

Earth system models can broadly be described as composed of components in which physical quantities are integrated on a physical grid. In a framework like ESMF, these are described in terms of 5 layers of abstractions consisting of ***metadata-laden data structures***. These layers are:

`grid` describes the physical grid in a standard way, so that component-neutral regridding software can be used to transform quantities from one grid component to another, with no knowledge of those components themselves. We seek to inscribe the grid metadata within community standards and conventions, so that analysis tools cognizant of these conventions may take advantage of grid information.

`field` consists of the physical variable discretized on a **`grid`**, along with metadata describing the physical quantity itself. The field metadata in ESMF have been designed to resemble the CF convention, so that CF-compliant model output may be produced if desired.

`attribute` configuration attributes of a component: these are very generic, but are intended to contain all the physical input parameters used to configure a model.

ESMF's metadata-laden data structures

`state` is the instantaneous state of some set of **`field`**s within a model component. Typically these are used as part of "import" and "export" states that are exchanged between components; but they are often used to contain the entire model state as well.

`component` the top level entity of this design. Components are hierarchical: that is, they may be composed of other components. The top-level **`component`** is the application or model itself.

These software layers exist in the ESMF, and ESMF-compliant models in the near future will be using these abstractions, rich in metadata, to describe a wide range of models across the weather and climate community. Simply by using these abstractions and encoding them in model output, we are creating a layer of ***formal, structured, hierarchical metadata***. We call this the ***model metadata layer***, and it is the core of the Curator. The model metadata layer is what makes possible for either a fully-configured model configuration or a model dataset to be the result of a database query.

The MAPS project

MAPS is an ESMF follow-on project between NASA/GMAO, GFDL, and MIT.

- ESMF has a very general notion of a “component”: MAPS seeks to develop a best-practice description of coupled climate models.
- The MAPS “standard architecture” will develop generic ESMF couplers for typical coupling functions.
- The test configuration will couple the GEOS-5 atmospheric model to a choice of three oceans: Poseidon, GFDL-MOM4, MITgcm.
- Runtime configuration will be done using the same XML descriptors as PRISM/OASIS4!

Initial study has shown us some basic incompatibilities... the ORAD component is below the OCORE in MOM4, parallel to OCORE in Poseidon. Requires MOM4 import state to undergo basic retooling to accept 3D heating rate; to turn off ORAD.

Comparative study of model metadata

There is a controversy within the community about the feasibility and advisability of treating components as interchangeable bits of code that can be slotted together at will. An understanding of component diversity will mark the limits of such an approach. We seek to answer two questions:

- Are components ostensibly labelled “atmosphere”, say, sufficiently similar that a single physical interface may be defined? Or, to put it another way, to what extent to two such components share a `state`?

- Do different models see component granularity the same way? What incompatibilities are introduced if one model treats atmospheric chemistry say, as an indivisible entity within an atmosphere component, whereas another treats them as independent `component`s?

It appears, based on the MAPS experience, that the more granular framework forces its structure upon the less granular one.

The Earth System Curator

ESC begins with a crucial insight: that the descriptors used for comprehensively specifying a model configuration are needed for a scientifically useful description of the model output data as well. Thus ***the same attributes may be used to specify a model as well as the model output dataset***: thus leading to a ***convergence of models and data***.

ESC is best considered a pilot project building prototype elements of a future ESME. The current project (2005-07) is funded by NSF and brings together NCAR, Princeton, MIT and GA Tech.

ESC metadata activities

Survey of model metadata:

- GFDL Curator DB: operational at GFDL for IPCC AR4 data holdings;
- ESMF component DB: based upon components listed as ESMF-compliant;
- NMM: we have begun dialogue between Curator and NMM, would have been great if Rocky Dunlap could have been here!

Grid metadata is also a Curator milestone: draft will be presented tomorrow.