Collaborative Model Metadata Development with ESG/Curator/METAFOR

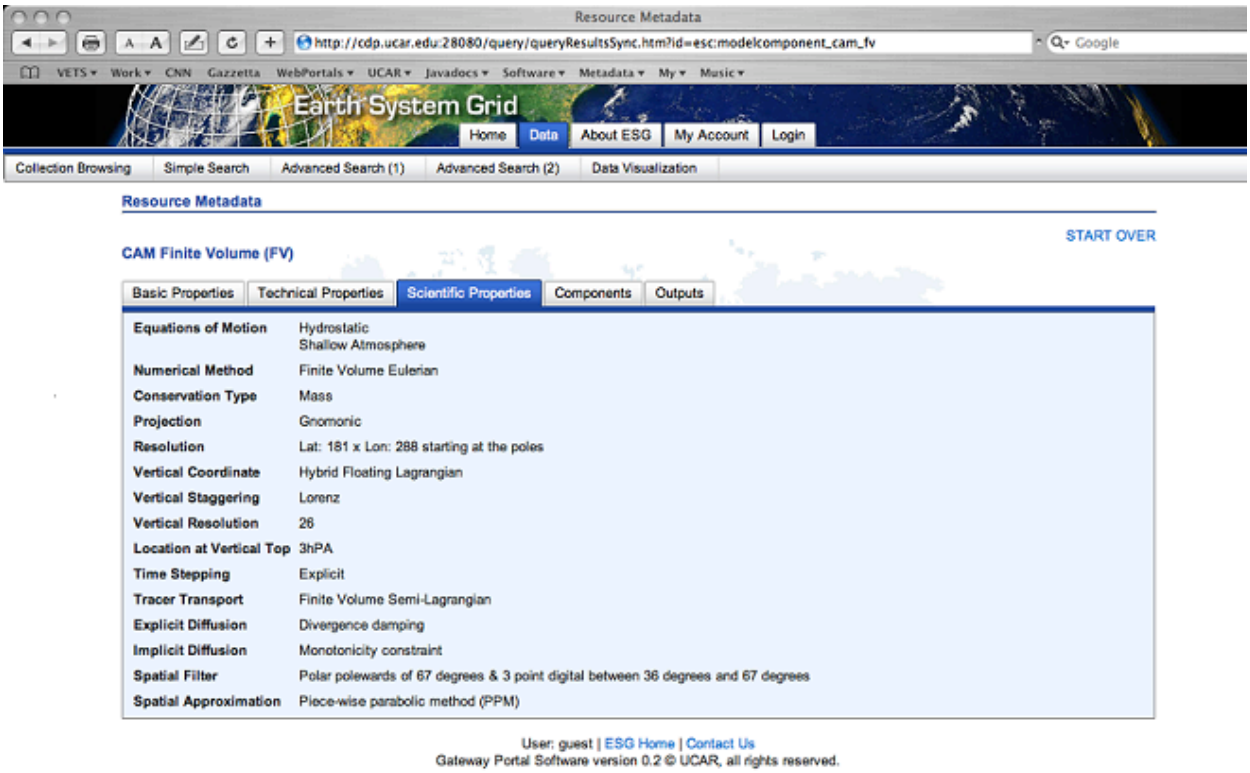
Sylvia Murphy[1], Luca Cinquini[1], Julien Chastang[1], Cecelia DeLuca[1], Don Middleton[1], and V. Balaji[2]

[1]National Center for Atmospheric Research; [2]Geophysical Fluid Dynamics Laboratory

Summary

The metadata associated with climate science artifacts, such as models and datasets, enables them to be discovered, understood, compared, and used with analysis tools and automated workflows. The structured vocabulary, or ontology, employed by the Earth System Grid (ESG) project is increasingly influenced by that being developed by the European Common Metadata for Climate Modelling Digital Repositories (METAFOR) project. The Earth System Curator project is working with the METAFOR team on the development of their ontology and is serving as a liaison between METAFOR and ESG. Members of the Curator team at NCAR are implementing portions of the METAFOR ontology in the ESG Gateway, and working with ESG and METAFOR to reconcile differences. The Curator team has focused initially on implementation of a detailed, hierarchical grid specification developed at GFDL. Curator has also added detailed metadata describing atmospheric dynamical cores, developed for a summer 2008 workshop at NCAR. New ontology additions are being integrated into search and browse features of the Gateway. A "trackback" interface has been developed, which enables users to move easily between datasets and descriptions of the simulations that generated them.

Figure 1: Scientific properties for a finite volume dynamical core, displayed in a new interface in the ESG Gateway. The goal is to enable scientists to find detailed information about models, model components, and datasets through highly accessible web interfaces.



Researchers who used the ESG IPCC archive for AR4 noted ways that model and experiment metadata could be enhanced, including identification of input datasets for model simulations, identification of specific subcomponents and versions used in simulations, and descriptions of key algorithms and processes.[1] The Earth System Curator effort, in collaboration with ESG and other community projects, has been working to supplement the model metadata in the ESG Gateway and make detailed information about the scientific and technical properties of climate models readily available on-line.

The Curator project began in 2005 with funding from the National Science Foundation, and new support has come from NASA. Partners include NCAR, the Georgia Institute of Technology, NOAA GFDL, and MIT. Curator members at NCAR have focused on work on the ESG Gateway, including:

1. Extending the ESG ontology to include information describing the scientific and technical properties of models and their components. The recent focus has been on the incorporation of detailed grid metadata developed at GFDL.
2. Developing new interfaces in the ESG Gateway to display this metadata, and through "trackback" pages, linking datasets with descriptions of the model configurations that generated them.
3. Generating metadata from climate models in a form that is readily digestible by ESG and other science gateways.

The Curator strategy has been to build upon national and international community efforts in order to create infrastructure that leverages the efforts of many individual groups, will persist over time, and offers functionality beyond what a single project could achieve. ESG, Curator, and the European METAFOR project have had complementary roles.

ESG provided the base ontology, originally focused on data collections. It is defined in the Web Ontology Language (OWL), which offers great flexibility in expressing relationships; it can easily represent the nested structure of a model, or enable a scientist to direct their search through ESG collections by selecting which aspects of a simulation (e.g. has volcanic forcing, runs on my institution's computer) are most important to them. In addition to the initial ontology, members of the ESG staff at NCAR have provided guidance and support for implementation.

The METAFOR project was funded in 2008 by the European Union to define climate data and the models that produce it in a standard way. METAFOR has been endorsed by the World Climate Research Programme as the source of CMIP5 metadata definition. Curator is serving as a liaison between METAFOR and ESG, by implementing portions of the METAFOR ontology in the ESG Gateway, and working to reconcile differences.

A prototype Gateway based on ESG that utilized metadata-enhanced capabilities was deployed for the 2008 "Numerical Techniques for Global Atmospheric Models" Colloquium at NCAR. About 40 students and their mentors ran, analyzed, and compared nine different atmospheric dynamical cores. Curator staff assisted the workshop organizers in defining metadata that described the different cores, and collaborated with Gateway developers to create new interfaces like the one shown in Figure 1, which displays the scientific properties of one of the cores. This early deployment was a useful test and demonstration of the ESG prototype, and proved instrumental to the colloquium's success.

The metadata enhancements to the Gateway are being migrated to ESG production code as they become increasingly robust. They represent a productive collaboration between ESG and the Curator and METAFOR projects, and will help scientists to understand and compare the data produced for future experiments such as CMIP5.

For further information on this subject contact:
Name: Cecelia DeLuca
Organization: National Center for Atmospheric Research
Email: cdeluca@ucar.edu
Phone: (303) 497-1828

[1] E.g. http://www.realclimate.org/index.php/archives/2008/02/ipcc-archive/