Scaling the Earth System Grid to Petascale Data

PRINCIPAL INVESTIGATOR: Galen M. Shipman

PARTICIPATING STAFF: Meili Chen, Ross G. Miller, Feiyi Wang

PROJECT START DATE: October 1, 2010

PROJECT END DATE: September 30, 2011

SPONSOR: US DOE, Office of Science, Office of Biological and Environmental Research (BER)

PARTNERS: Los Alamos National Laboratory (LANL), Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), National Center for Atmospheric Research (NCAR)

PROJECT DESCRIPTION

With prior Scientific Discovery through Advanced Computing (SciDAC) support, the research team has developed and deployed the Earth System Grid (ESG) to make climate simulation data easily accessible to the climate modeling community. ESG currently has 2500 registered users, and manages 160 terabytes (TB) of data in archives distributed around the nation. In this past year alone, more than 200 scientific journal articles have been published from analyses of data delivered by the ESG.

Despite these successes, ESG faces significant challenges in coming years as the size, complexity and number of climate datasets grows dramatically. The goals of this proposed five-year project are to (a) sustain the successful existing ESG system, (b) address projected scientific needs for data management and analysis, (c) extend ESG to support the major Intergovernmental Panel on Climate Change (IPCC) assessment in 2010, (d) support the Climate Science Computational End Station at the DOE Leadership Computing Facilty at ORNL, and (e) support climate model evaluation activities under the proposed SciDAC2 climate application. To do this, the team will broaden ESG to support multiple types of model and observational data, provide more powerful (client-side) ESG access and analysis services, enhance interoperability between common climate analysis tools and ESG, and enable end-to-end simulation and analysis workflow.

SIGNIFICANCE

DOE's investment in climate change research is broad, spanning model development, climate change simulation, model intercomparisons, observational programs, and supporting infrastructure for the Intergovernmental Panel on Climate Change (IPCC). Climate change research is increasingly data intensive, involving the analysis and intercomparison of simulation and observation data from many sources. Continued scientific progress depends upon powerful, effective enabling technologies that allow the core climate science community to coherently manage and publish a diverse collection of what in a few years will be petascale data, such that a broad, global community can access and analyze it.