# Scaling the Earth System Grid to Petascale Data

With prior SciDAC support, we 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 of data in archives distributed around the nation. From the 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 data sets grow dramatically. Building upon the ESG success and experience, we will establish the Earth System Grid Center for Enabling Technologies (ESG-CET) aimed at scaling ESG into the petascale realm and equipping it to serve future scientific requirements. This environment will enable broad community access to, and deep analysis of, simulation and experimental data from a distributed network of sources.

The goals of this 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 next major international climate study, the Intergovernmental Panel on Climate Change assessment in 2010, and (d) support climate model evaluation activities under the proposed SciDAC2 climate application.

To do this, we will broaden ESG to support multiple types of model and observational data, provide federation of ESG data sites, 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.

The ESG-CET project is a multi-institution project that includes USC Information Sciences, the National Center for Atmospheric Research, Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Lawrence Livermore National Laboratory, Argonne National Laboratory, and Los Alamos National Laboratory.

ISI's work on the ESG-CET project will include distributed data management, monitoring of the ESG infrastructure and packaging of ESG software.

In the area of data management, ISI will work with ESG to support the distribution of the ESG infrastructure among data centers from many institutions that provide climate data sets. This work will include continued support for the Globus Replica Location Service and the GridFTP data transport protocol. We will also work with other ESG participants to design a distributed metadata service, identify requirements for consistency and security of metadata accesses and explore design alternatives for exchanging metadata. In this context, we will explore the use of peer-to-peer technologies for ESG catalogs and services.

ISI has also led work on monitoring the existing ESG infrastructure to identify problems with resources in the infrastructure. The monitoring infrastructure allows the ESG administrators to

specify conditions that signify failures, such as when a data transfer service fails to respond to a request. When such conditions are detected, administrators are immediately informed, allowing failures to be addressed quickly, thereby minimizing infrastructure downtime and allowing more consistent access to ESG resources by climate researchers. In the new project, we will expand the richness of the resources being monitored and provide more failure recovery capabilities for ESG. We will also incorporate a monitoring archive that allows ESG to understand patterns of failures over time.

Finally, we will provide technical support for migration of the existing ESG software to the latest version of the Globus Toolkit and for packaging ESG services so that they can be easily deployed on resources in the distributed ESG infrastructure.