

The Earth System Grid II: Turning Climate Model Datasets into Community Resources

PI's: Ian Foster (ANL), Don Middleton (NCAR),
& Dean Williams (LLNL/PCMDI)

ESG Personnel: Shishir Bharathi (USC), David Bernholdt (ORNL), David Brown (NCAR), Kasidit Chanchio (ORNL), Ann Chervenak (USC/ISI), Luca Cinquini (NCAR), Bob Drach (LLNL), Peter Fox (NCAR), Jose Garcia (NCAR), Carl Kesselman (USC/ISI), Rob Markel (NCAR), Veronika Nefedova (ANL), Line Pouchard (ORNL), Arie Shoshani (LBNL), Alex Sim (LBNL), and Gary Strand (NCAR) Additional Collaborators: William Allcock (ANL), Lawrence Buja (NCAR), Joe Link (ANL), Laura Pearlman (USC/ISI), Von Welch (ANL), John Drake (ORNL), Ilana Stern (NCAR)

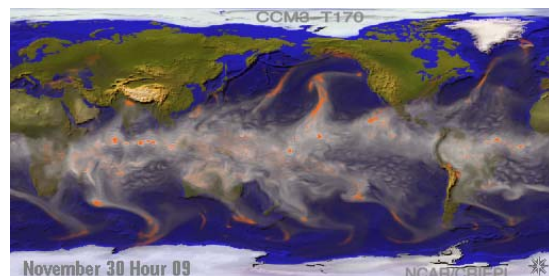
Summary

In pursuit of DOE Climate Change research goals, global climate simulations are being run on supercomputers across several DOE sites and at NCAR. The resulting data archive, distributed over several sites, currently contains upwards of one hundred terabytes of simulation data. Looking towards mid-decade and beyond we must prepare for distributed climate data holdings of many petabytes. The Earth System Grid (ESG) is a collaborative interdisciplinary project aimed at addressing the challenge of enabling management, discovery, access, and analysis of these enormous and extremely important data assets.

By late 2003, DOE-sponsored climate change research had produced roughly 100 terabytes of scientific data that is stored across several of the DOE sites and at NCAR. For the modeling teams, the daily management and tracking of their data is already proving to be a significant problem. The primary customers for the data are climate researchers who reside at various centers and universities across the U.S. and their ability to discover and use the data is extremely limited. That's today, and the problem is rapidly escalating.

In the future, the computers we run the models on will be much faster and the models themselves will become increasingly complex. Furthermore, the geographic resolution that is studied will be much finer. Much of the current modeling activity is focused upon simulations aimed at the upcoming Intergovernmental Panel on

Climate Change (IPCC) assessment and these simulations have twice the horizontal resolution of the models that have been run for the past several years, roughly a 4X increase in data volume. The image below depicts new work representing yet another factor of two resolution increase.



All of this adds up to an enormous increase in the volume and complexity of data that will be produced. Moving it will become increasingly costly and we will often be strongly motivated to leave it where it at its computational point of origin.

ESG's mission is to chart a viable course into the future that allows us manage and *collaboratively* use the climate simulation data. To this end, our goals include researching new strategies, developing new technologies and tools, and building an operational environment for scientists.

The heart of ESG is a simple, elegant, and powerful web portal that allows scientists to register, search, browse, and acquire the data that they need. This portal is ready for early users and we demonstrated it to good effect at the SC'03 conference.

The ESG portal exposes a number of capabilities, many of which represent milestones for the project. In particular, ESG has developed new strategies and technologies for capturing and recording rich, detailed metadata about the climate model simulations. ESG's current web portal exposes a catalog of approximately one hundred datasets (roughly the last half-decade of simulation activity) with a rich body of semantic and usage-level metadata – a decisive step towards capturing elements of the “scientific notebook” for reuse by a broad community. We are employing powerful new technologies based on the OGSA-DAI project results.

ESG has also developed a simple, streamlined Globus-based registration system that allows us to sign up users much faster and more easily than before. This is potentially a technology that could be valuable to many other projects. Efforts are also underway in the security area to incorporate groups and roles and this work is undertaken in cooperation with the SciDAC *Security and Policy for Group Collaboration* project.

We have developed new data management capabilities that provide robust interoperability among DOE (HPSS) and NCAR (MSS) archival systems, and this reflects a close and very productive collaborative effort with the SciDAC

Scientific Data Management ISIC. It also earned us the paraphrased comment from a user: “a hundred times faster – and easier!”

One of ESG's strategies is to dramatically reduce the amount of data that needs to be moved over the network and we are engaged in groundbreaking work in developing generalized remote data access capabilities. This involves heavy collaboration with the SciDAC *High-Performance Datagrid Toolkit* project, as well as joint work with the community *OpenDAP* project. We are in an alpha-testing phase of a new suite of *Virtual Data Services* that will be extremely important to the scientists who use ESG.

ESG is firmly invested in the SciDAC team approach and engages in constant close interaction among climate and computer scientist. We also use the AccessGrid on a weekly basis for project meetings and interactions with scientists and other collaborators. ESG will be an early adopter of new AccessGrid technology releases from the *Middleware to Support Group to Group Collaboration* SciDAC project.

In Summer of 2004, ESG will begin serving IPCC and other model data to a global community in close partnership with the IPCC effort and the World Meteorological Organization (WMO). Efforts throughout 2004 will involve formal user testing and analysis along with battle-hardening and performance enhancement of the underlying systems.

For further information on this subject:

<http://www.earthsystemgrid.org/>

Contact the ESG PI Team at:

esg-manage@earthsystemgrid.org

Or contact Don Middleton

NCAR Scientific Computing Division

don@ucar.edu, 303-497-1250