

# The Earth System Grid (ESG): Turning Climate Datasets into Community Resources



Presented by

## The ESG-CET Team

*including*

Lawrence Berkeley National Laboratory

Lawrence Livermore National Laboratory

Los Alamos National Laboratory

National Center for Atmospheric Research

National Oceanic and Atmospheric Administration

Oak Ridge National Laboratory

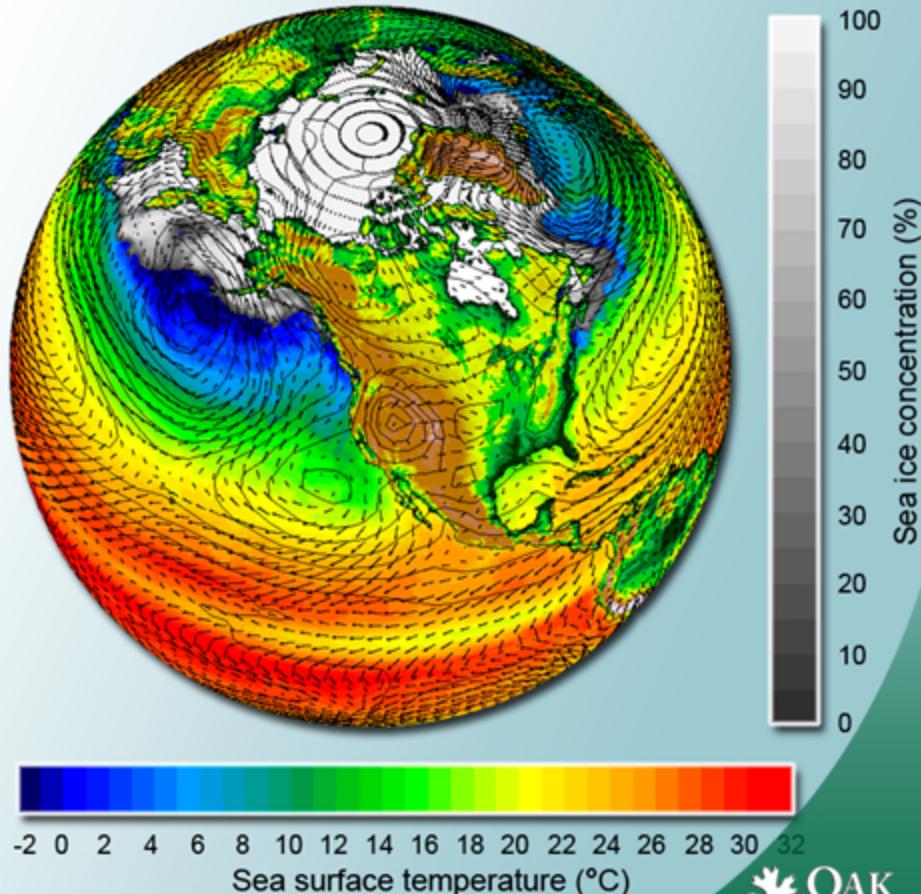
University of Southern California

[www.earthsystemgrid.org](http://www.earthsystemgrid.org)

# The growing importance of climate simulation data

- DOE invests broadly in climate change research:
  - Development of climate models
  - Climate change simulation
  - Model intercomparisons
  - Observational programs
- Climate change research is increasingly data-intensive:
  - Analysis and intercomparison of simulation and observations from many sources
  - Data used by model developers, impacts analysts, policymakers

Results from the Parallel Climate Model (PCM) depicting wind vectors, surface pressure, sea surface temperature, and sea ice concentration. Prepared from data published in the ESG using the FERRET analysis tool by Gary Strand, NCAR.



# Earth System Grid objectives



To support the infrastructural needs of the national and international climate community, ESG is providing crucial technology to securely access, monitor, catalog, transport, and distribute data in today's grid computing environment

## HPC hardware running climate models



## ESG Sites

Earth System Grid

Earth System Grid

ESG News

Data Mover Light now available: DML can be used to download a large number of files to a user's personal computer without clicking on multiple hyperlinks - see [DML instructions](#).

Data Subsetting: OPeNDAP-G offers high-performance subsetting capabilities on a number of virtual aggregated datasets - see [list of available aggregations](#).

Registration is required to download some of the data, please [request an account](#). Please send us comments or feedback.

New: [IPCC Working Group 1 data](#) available.

The NCAR M3S is scheduled for downtime each Sunday morning from 0000-0230 MST.

The NERSC HPSS is scheduled for maintenance downtime from 7-12 PST every Tuesday morning. This maintenance will be taken down in major

Data Search

Search Dataset metadata for:

Examples: c02, B06.77

Search

Welcome to ESG

The Earth System Grid (ESG) integrates supercomputers with large-scale data and analysis servers located at numerous national labs and research centers to create a powerful environment for next generation climate research. This portal is the primary point of entry into the ESG.

ESG Collaborators

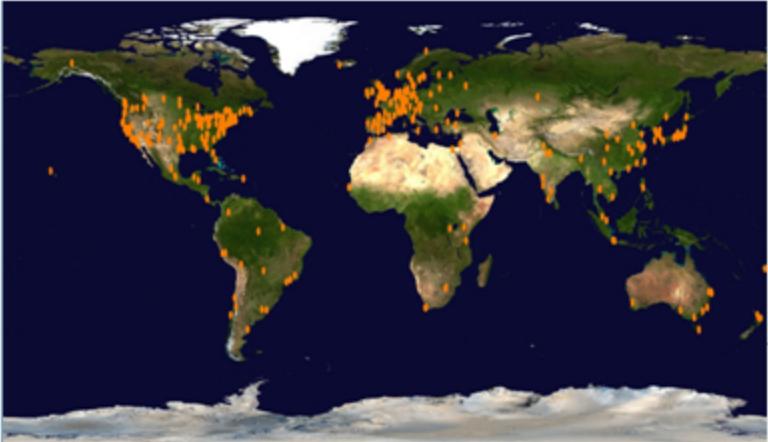
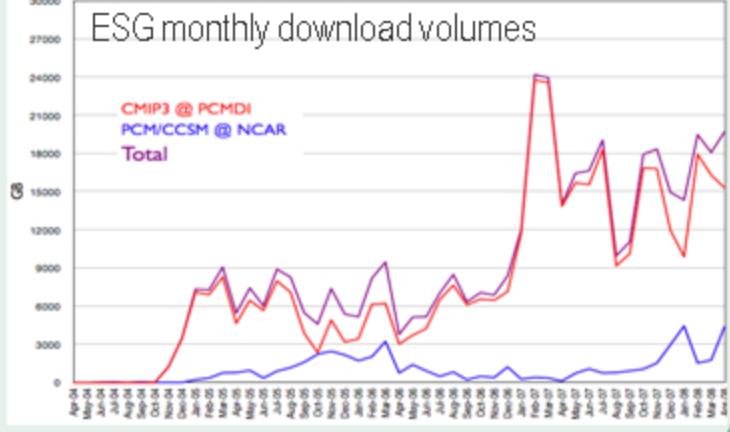
- Argonne National Laboratory
- Lawrence Berkeley National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- National Center for Atmospheric Research
- Oak Ridge National Laboratory
- University of Southern California/Information Sciences Institute

Funded by the U.S. Department of Energy

## ESG Portal

# ESG facts and figures



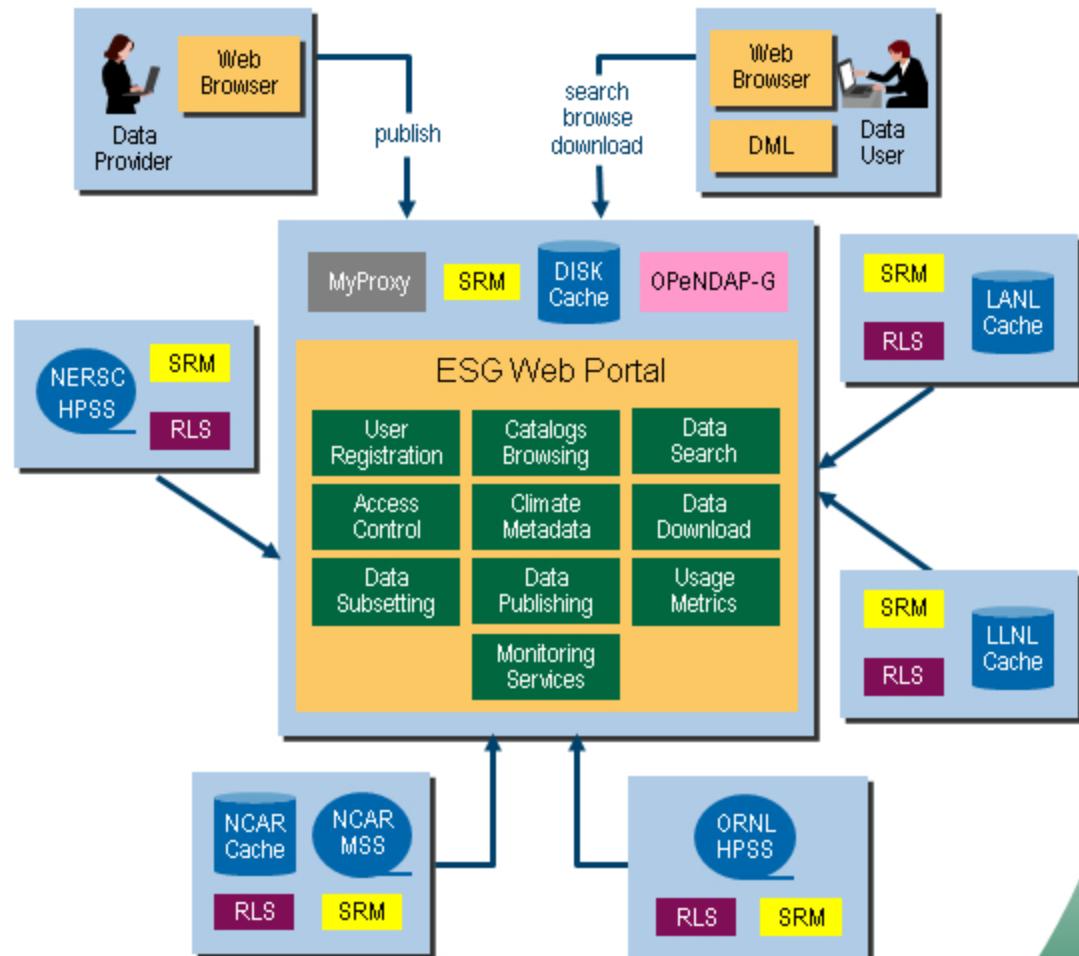
Main ESG Portal	CMIP3 (IPCC AR4) ESG Portal
198 TB of data at four locations <ul style="list-style-type: none"><li>• 1,150 datasets</li><li>• 1,032,000 files</li><li>• Includes the past 6 years of joint Department of Energy/National Science Foundation climate modeling experiments</li></ul>	35 TB of data at one location <ul style="list-style-type: none"><li>• 74,700 files</li><li>• Generated by a modeling campaign coordinated by the Intergovernmental Panel on Climate Change</li><li>• Data from 13 countries, representing 25 models</li></ul>
8,000 registered users	2,000 registered projects
Downloads to date <ul style="list-style-type: none"><li>• 49 TB</li><li>• 176,000 files</li></ul> 	Downloads to date <ul style="list-style-type: none"><li>• 387 TB</li><li>• 1,300,000 files</li><li>• 500 GB/day (average)</li></ul> 
ESG usage: over 500 sites worldwide	<b>400 scientific papers published to date based on analysis of CMIP3 (IPCC AR4) data</b>

# ESG architecture and underlying technologies



- Climate data tools
  - Metadata catalog
  - NcML (metadata schema)
  - OPeNDAP-G (aggregation and subsetting)
- Data management
  - Data Mover Lite
  - Storage Resource Manager
- Globus toolkit
  - Globus Security Infrastructure
  - GridFTP
  - Monitoring and Discovery Services
  - Replica Location Service
- Security
  - Access control
  - MyProxy
  - User registration

First Generation ESG Architecture



**MSS, HPSS:** Tertiary data storage systems

# Evolving ESG to petascale



## ESG Data System Evolution

**2006**

Central database

- Centralized curated data archive
- Time aggregation
- Distribution by file transport
- No ESG responsibility for analysis
- Shopping-cart-oriented web portal

**Early 2009**

Testbed data sharing

- Federated metadata
- Federated portals
- Unified user interface
- Selected server-side analysis
- Location independence
- Distributed aggregation
- Manual data sharing
- Manual publishing

**2011**

Full data sharing (add to testbed...)

- Synchronized federation –metadata, data
- Full suite of server-side analysis
- Model/observation integration
- ESG embedded into desktop productivity tools
- GIS integration
- Model intercomparison metrics
- User support, life cycle maintenance

CCSM  
IPCC

**Terabytes**

## ESG Data Archive

Bonelli\_ESG\_XCOS

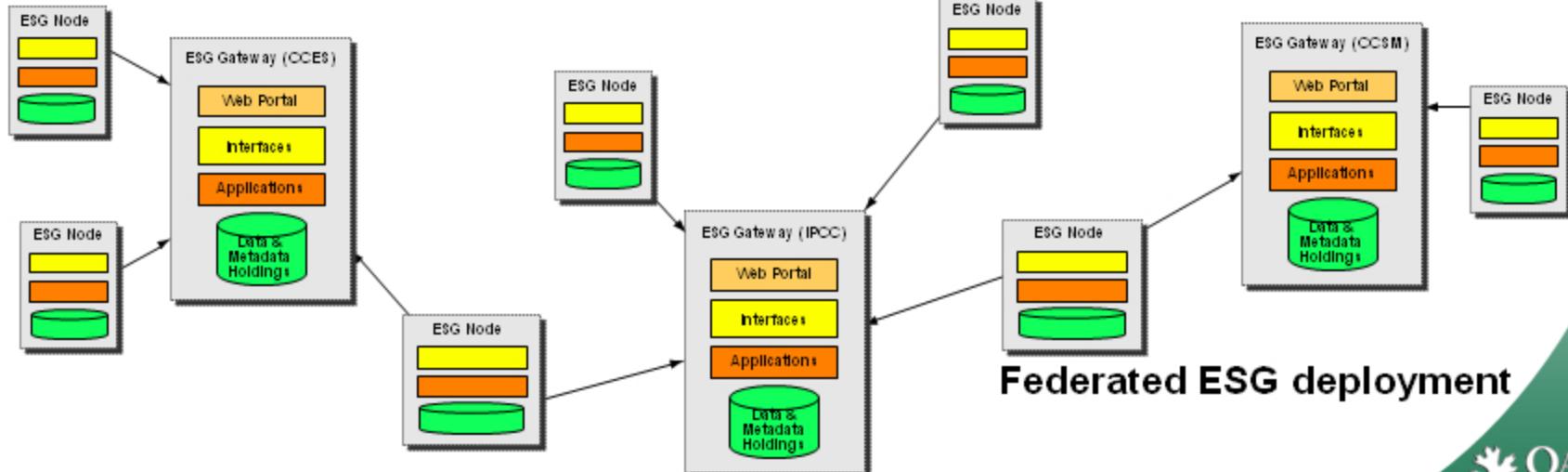
CSSM, IPCC,  
satellite, In situ  
biogeochemistry,  
ecosystems

**Petabytes**

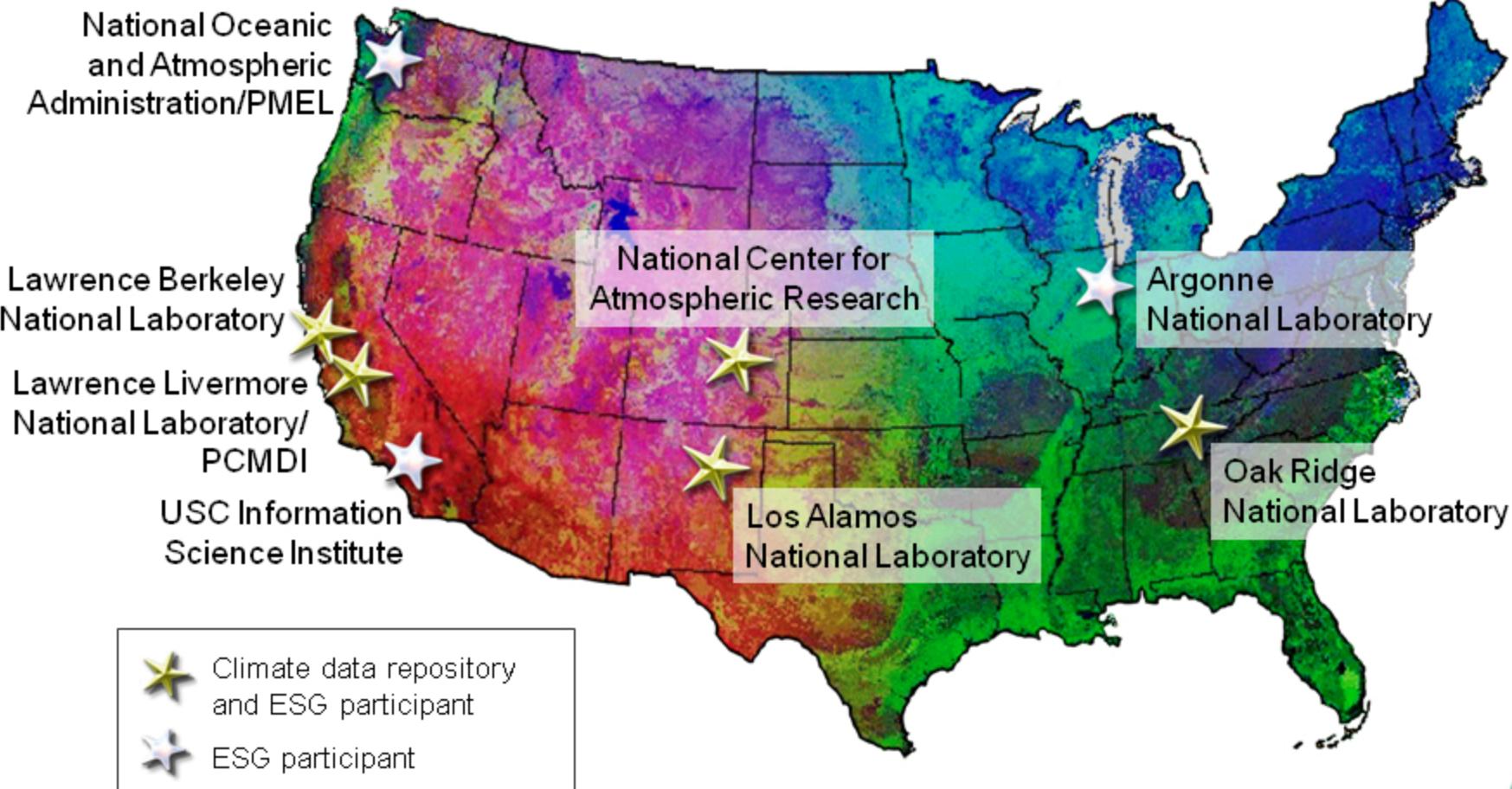


# Architecture of the next-generation ESG

- Petascale data archives
- Broader geographical distribution of archives
  - across the United States
  - around the world
- Easy federation of sites
- Increased flexibility and robustness



# The team and sponsors



# Contact



## ORNL booth at SC2008

- David Bernholdt

## Other booths at SC2008

- ANL (Booth 558)
- LBNL (540)
- NCAR (365)

Ian Foster

Arie Shoshani, Alex Sim

Don Middleton

## Internet

- <http://www.earthsystemgrid.org>
- [esg-xc@earthsystemgrid.org](mailto:esg-xc@earthsystemgrid.org)