The Earth System Grid: Turning Climate Datasets into Community Resources
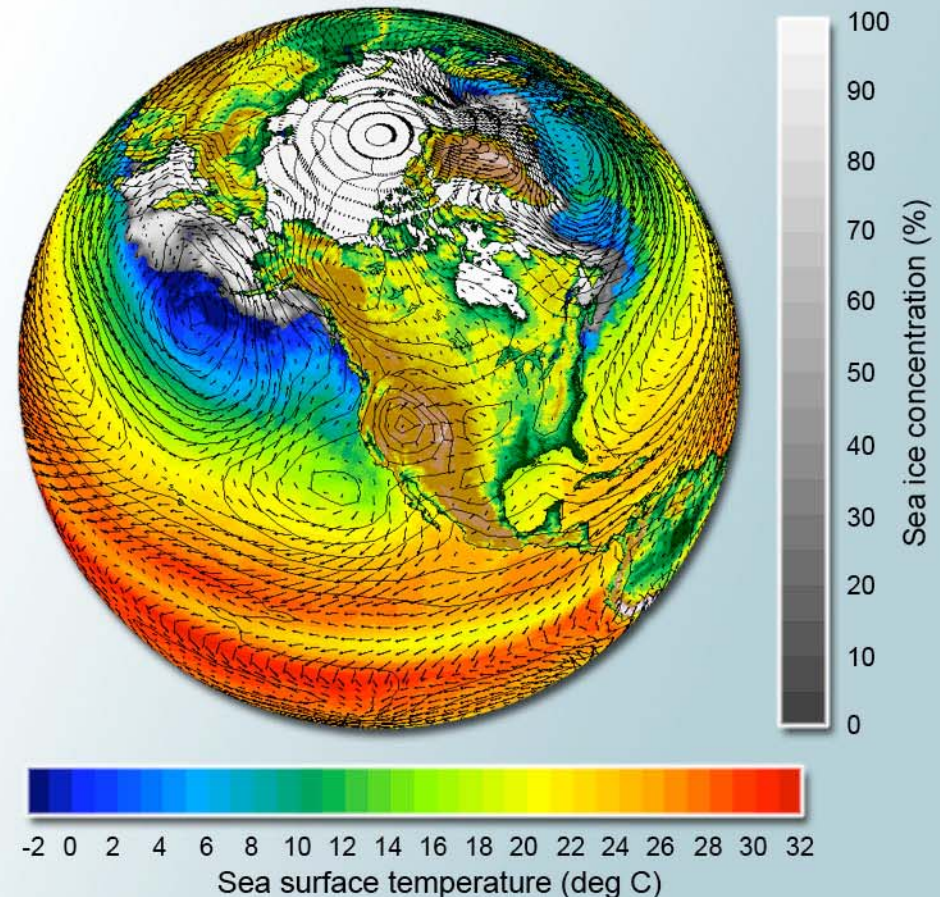
www.earthsystemgrid.org

SC07

Office of Science
U.S. DEPARTMENT OF ENERGY

SciDAC
Scientific Discovery through Advanced Computing

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.

Sea ice concentration (%)

Sea surface temperature (deg C)
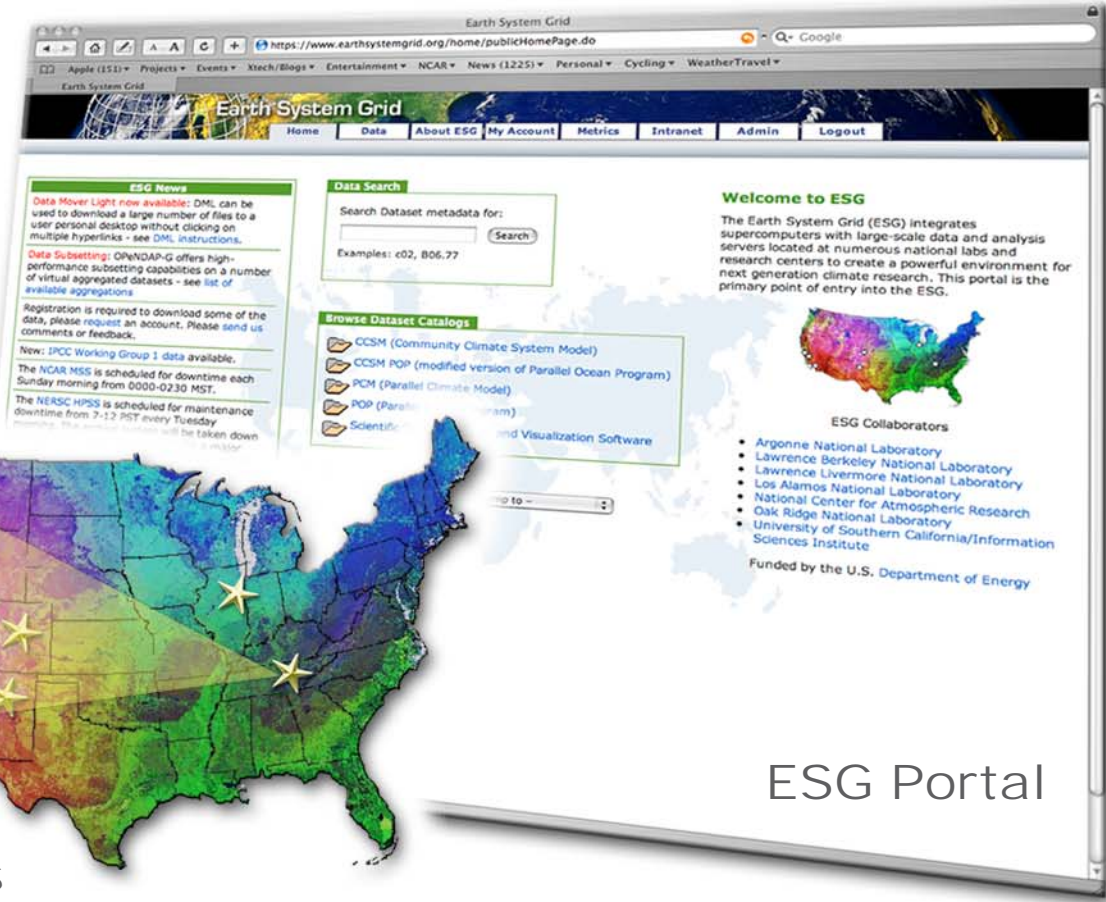
Earth System Grid objectives

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

ESG Portal

ESG facts and figures

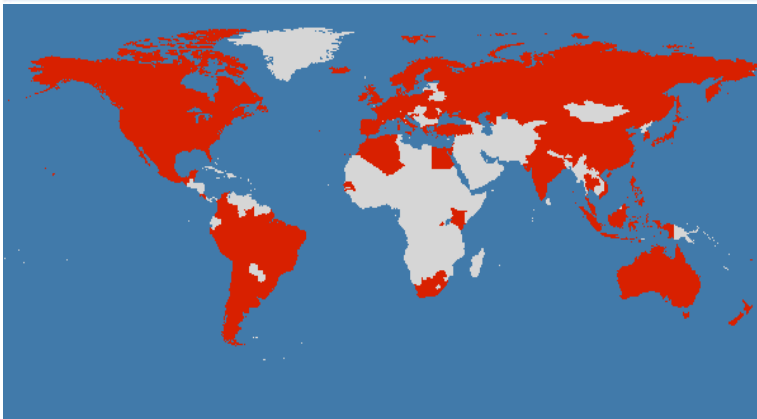| Main ESG Portal	CMIP3 (IPCC AR4) ESG Portal
146 TB of data at four locations • 1,059 datasets • 958,072 files • Includes the past 6 years of joint DOE/NSF climate modeling experiments	**35 TB of data at one location** • 77,400 files • Generated by a modeling campaign coordinated by the Intergovernmental Panel on Climate Change • Model data from 13 countries
4,910 registered users	**1,314 registered users**
Downloads to date • 30 TB • 106,572 files	**Downloads to date** • 245 TB • 914,400 files • 500 GB/day (average)

CMIP3 (IPCC AR4) Daily Downloads (through 7/2/07)
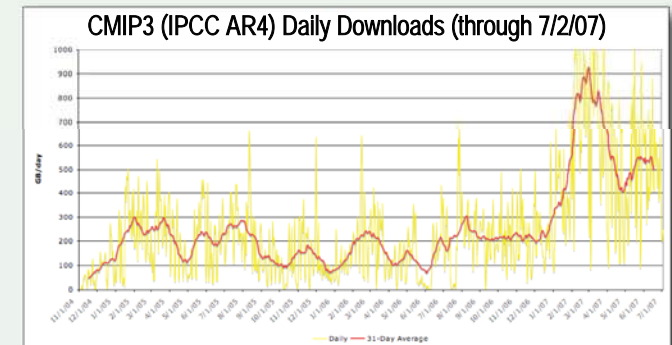
> **> 300 scientific papers published to date based on analysis of CMIP3 (IPCC AR4) data**

Worldwide ESG user base

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
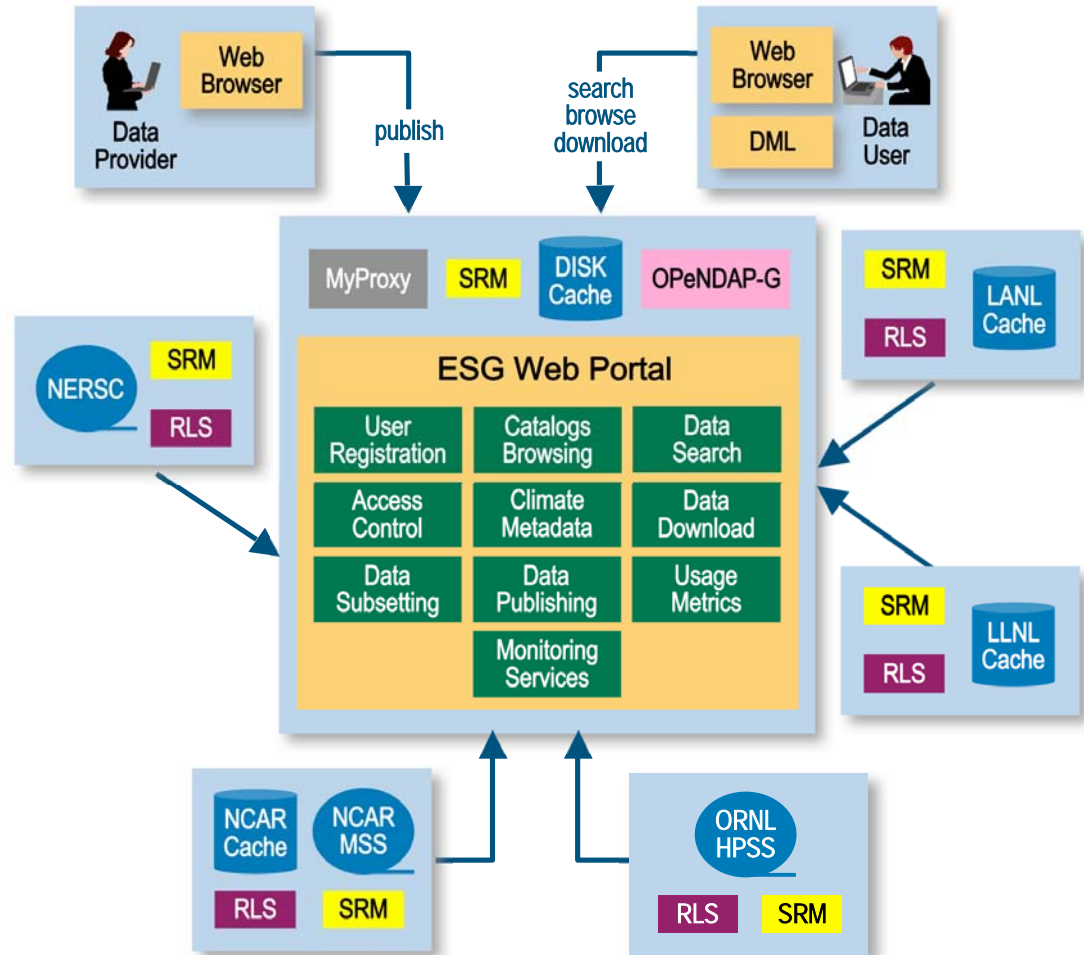
MSS, HPSS: Tertiary data storage systems

Evolving ESG to petascale

Earth System Grid

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

ESG Data Archive

CSSM, IPCC, satellite, In situ biogeochemistry, ecosystems
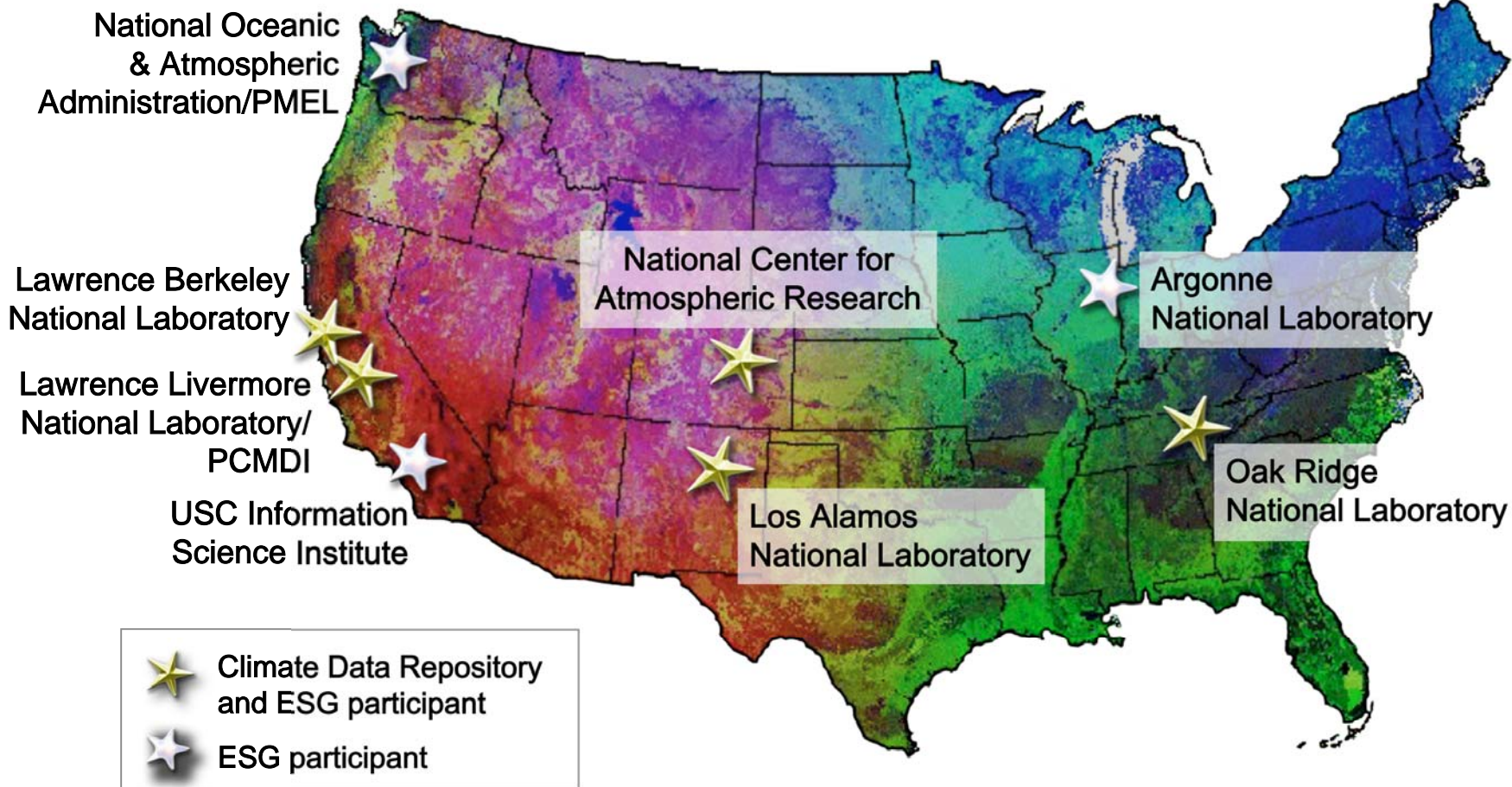
Terabytes **Petabytes**

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**

Second Generation ESG Architecture

Federated ESG Deployment
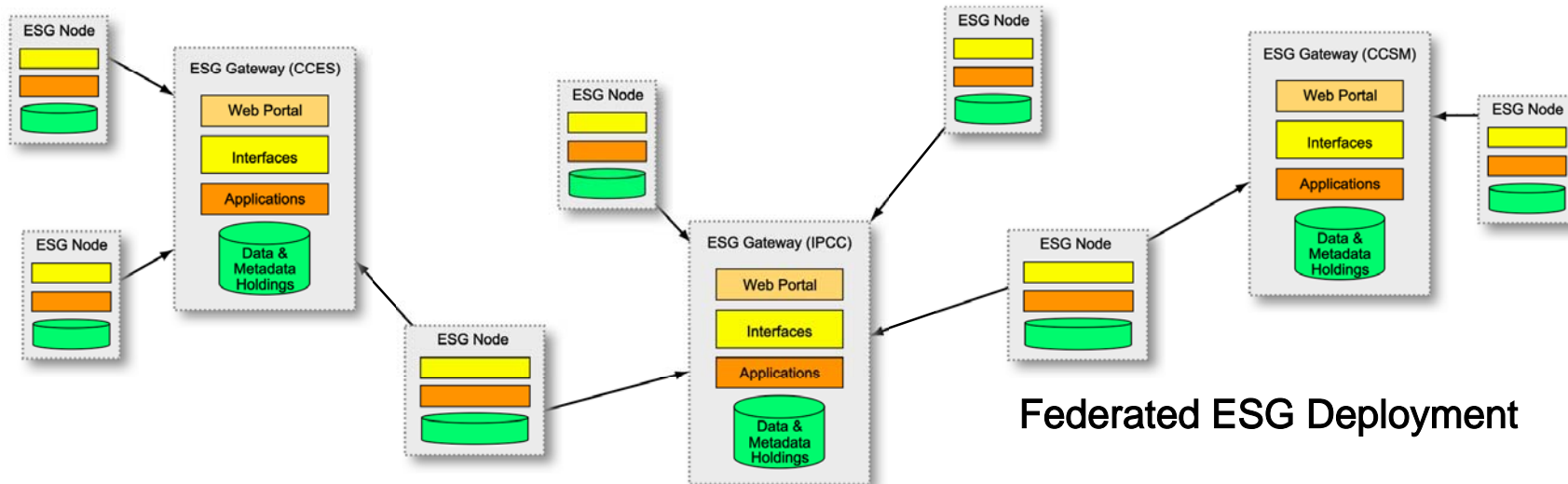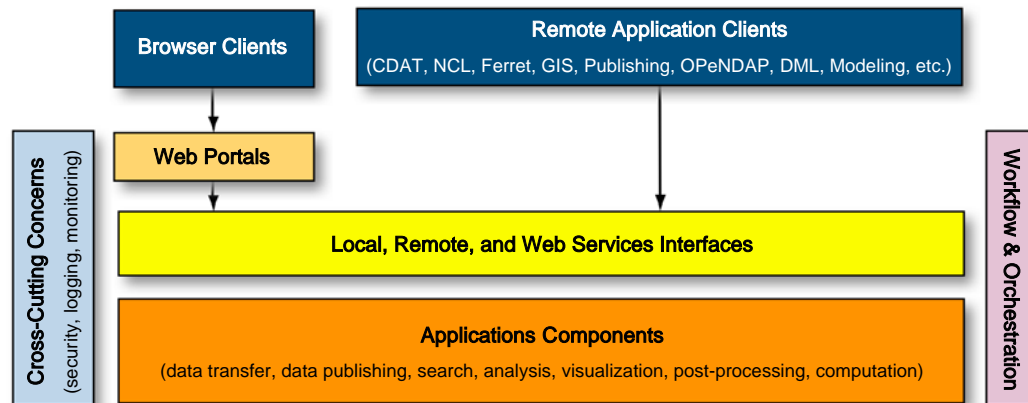
The team and sponsors

For more information...

ORNL booth at SC2007

- David Bernholdt

Other booths at SC2007

- ANL/Global Grid Forum (Booth 551)   Ann Chervenak

- LBNL (351)   Arie Shoshani, Alex Sim

- NCAR (361)   Don Middleton

Internet

- http://www.earthsystemgrid.org

- Esg-manage@earthsystemgrid.org