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

## www.earthsystemgrid.org





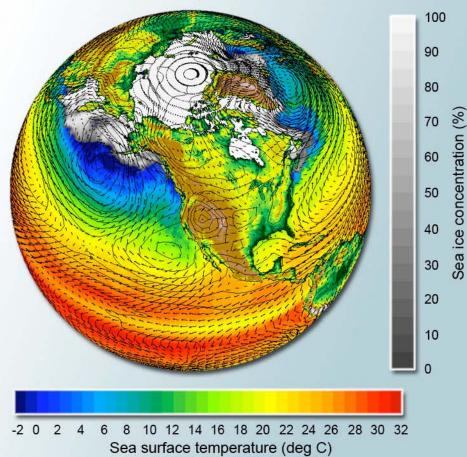


## The growing importance of climate simulation data



- 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.

#### Earth System Grid HPC Q . Q. Coogle A A C + Ohttps://www.earthsystemgrid.org/home/publicHomePage.do Apple (151) \* Projects \* Events \* Xtech/Blogs \* Entertainment \* NCAR \* News (1225) \* Personal \* Cycling \* WeatherTravel \* hardware running Earth System Grid Home Data About ESG My Account Metrics Intranet climate models **Data Search** Welcome to ESG Ir Light now available: DML can be used to download a large number of files to a Search Dataset metadata for: The Earth System Grid (ESG) integrates ser personal desktop without clicking on supercomputers with large-scale data and analysis servers located at numerous national labs and Search nultiple hyperlinks - see DML instructions Examples: c02, 806.77 Oata Subsetting: OPeNDAP-G offers high-performance subsetting capabilities on a number research centers to create a powerful environment for next generation climate research. This portal is the of virtual apprepated datasets - see list of primary point of entry into the ESG. rvailable apprepations legistration is required to download some of the data, please request an account. Please send us Browse Dataset Catalog comments or feedback. CCSM (Community Climate System Model) New: IPCC Working Group 1 data available. CCSM POP (modified version of Parallel Ocean Program) The NCAR MSS is scheduled for downtime each Sunday morning from 0000-0230 MST. PCM (Parallel Climate Model) The NERSC HPSS is scheduled for maintenance POP (Paral me from 7-12 PST every Tuesday ESG Collaborators Scientif. nd Visualization Software Die taken de 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 Funded by the U.S. Department of Energy **ESG** Portal **ESG** Sites

# **ESG** facts and figures



Main ESG Portal	CMIP3 (IPCC AR4) ESG Portal
<ul> <li>146 TB of data at four locations</li> <li>1,059 datasets</li> <li>958,072 files</li> <li>Includes the past 6 years of joint DOE/NSF climate modeling experiments</li> </ul>	<ul> <li>35 TB of data at one location</li> <li>77,400 files</li> <li>Generated by a modeling campaign coordinated by the Intergovernmental Panel on Climate Change</li> <li>Model data from 13 countries</li> </ul>
4,910 registered users	1,314 registered users
<ul> <li>Downloads to date</li> <li>30 TB</li> <li>106,572 files</li> </ul>	<text></text>
	> 300 scientific papers published to date based on analysis of CMIP3 (IPCC AR4) data

Worldwide ESG user base

# ESG architecture and underlying technologies

- Earth System Grid
- Web Web Browsei Browser search browse publish Data Data download DML User Provider DISK Cache SRM **OPeNDAP-G** SRM **MyProxy** LANL Cache RLS SRM ESG Web Portal NERSC RLS Catalogs Data User Registration Browsing Search Climate Data Access Download Control Metadata Usage Data Data Subsetting Publishing Metrics SRM LLNL Monitorina Cache RLS Services NCAR MSS NCAR ORNL HPSS Cache SRM RLS SRM RLS

MSS, HPSS: Tertiary data storage systems

### Climate data tools

- Metadata catalog
- NcML (metadata schema)
- OPenDAP-G
   (aggregation and subsetting)
- Data management
  - <u>Data Mover Lite</u>
  - <u>Storage Resource Manager</u>
- Globus toolkit
  - <u>G</u>lobus <u>Security</u> <u>Infrastructure</u>
  - GridFTP
  - Monitoring and Discovery Services
  - <u>Replica</u> Location Service
- Security
  - Access control
  - MyProxy
  - User registration

#### First Generation ESG Architecture

# **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

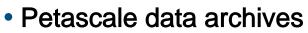
Full data sharing (add to testbed ... )

- 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



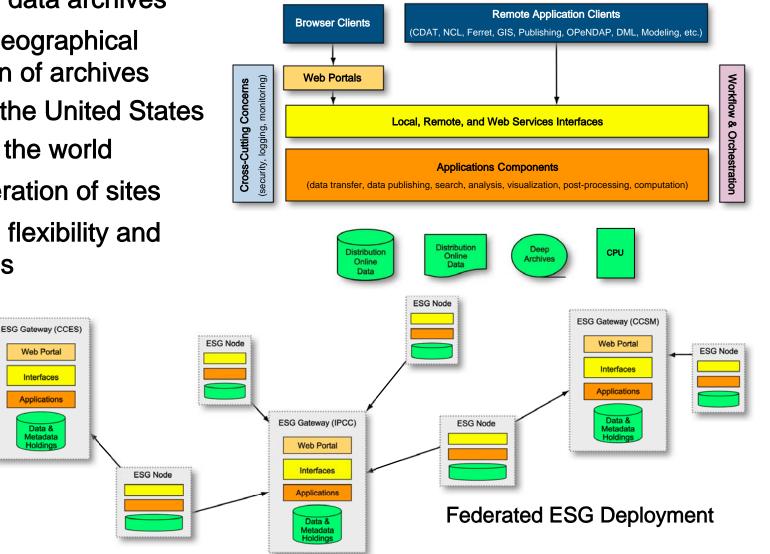
## Architecture of the next-generation ESG





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

#### Second Generation ESG Architecture

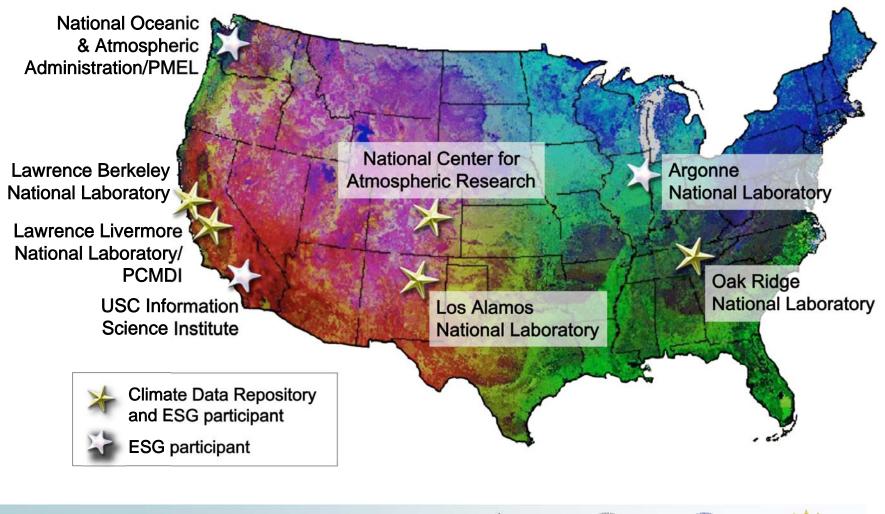


ESG Node

ESG Node

# The team and sponsors







# For more information...



## **ORNL booth at SC2007**

David Bernholdt

## **Other booths at SC2007**

- ANL/Global Grid Forum (Booth 551)
- LBNL (351)
- NCAR (361)

## Internet

- http://www.earthsystemgrid.org
- Esg-manage@earthsystemgrid.org

Ann Chervenak Arie Shoshani, Alex Sim Don Middleton