



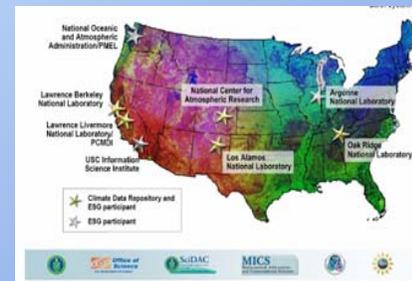
Earth System Grid: A Federated Climate Model Data Management Infrastructure

R. B. McCoy, C. Doutriaux, R. Drach, K. Halliday, V. Mlaker, D. N. Williams
Lawrence Livermore National Laboratory, PCMDI



The Earth System Grid (ESG) is a **virtual collaborative environment** that links distributed centers, users, models, and data in a Grid computing environment. The primary goal of ESG is to support the infrastructural needs of the national and international climate community by providing crucial technology to securely access, monitor, catalog, transport, and distribute data. The **next generation ESG Center for Enabling Technologies (ESG-CET)** will support **petabyte dataset volume** in a distributed environment through the **federation of data centers**.

ESG Team and Sponsors



Current ESG-II

- The ESG was enhanced to support the **IPCC Fourth Assessment Report (AR4)**.
- PCMDI assembled an unprecedented set of model data from **12 experiments, 13 countries and 23 models**, so called **CMIP3 database**.



CMIP3 AR4 (2004-2007) Worldwide User Base

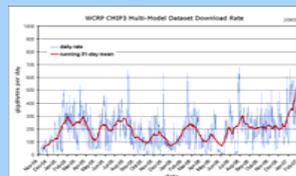
PCMDI - Program for Climate Model Diagnosis and Intercomparison at LLNL
IPCC - Intergovernmental Panel on Climate Change
CMIP3 - Coupled Model Intercomparison Project phase 3

CMIP3 Facts

- 32 TB** of data at the PCMDI (73,000 files)
- 1000 registered users**
- FTP, web portal, and analysis tool access
- "Virtual Datasets"** with subsetting, aggregation
- 247 scientific papers** published

Downloads to date:

- 171 TB**
- 703,000 files**
- 300 GB/day (average)**



ESG Portal



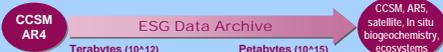
<http://www-pcmdi.llnl.gov>
<https://esg.llnl.gov:8443/index.jsp>

Next Generation ESG-CET

Evolving ESG for the future

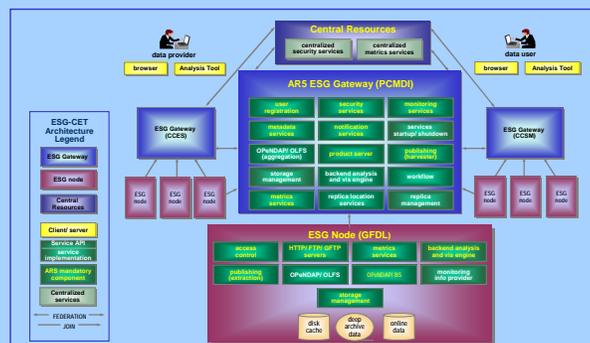
ESG Data System Evolution

2006	Early 2009	2011
Central database <ul style="list-style-type: none"> Centralized curated data archive Time aggregation Distribution by file transport No ESG responsibility for analysis Shopping-cart-oriented web portal ESG connection to desktop analysis tools (i.e., CDAT and CDAT-LAS) 	Testbed data sharing <ul style="list-style-type: none"> Federated metadata Federated portals Unified user interface Quick look server-side analysis with CDAT Location independence Distributed aggregation Manual data sharing Manual publishing 	Full data sharing (leads to testbed...) <ul style="list-style-type: none"> Synchronized federation > metadata, data Full suite of server-side analysis with CDAT Model/observation integration ESG embedded into desktop productivity tools with CDAT GIS integration Model intercomparison metrics User support, life cycle maintenance



CDAT - Climate Data Analysis Tools developed mainly at PCMDI; LAS - Live Access Server
CCSM - Community Climate System Model (NCAR); GIS - Geographic Information System

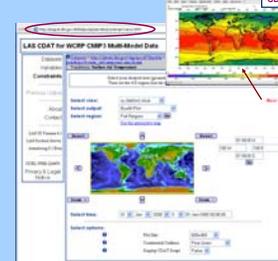
Future ESG-CET architecture



AR5 - IPCC Fifth Assessment Report (-year 2012)

Current Scenario

- Browse PCMDI's centralized database
- Select data
- Perform time aggregation
- Quick Look server-side (LAS - CDAT)
- Download data



<http://esgcat.llnl.gov:8090/pcdas/>

Future Scenario

- Search, browse and discover distributed data
- Remote site:
 - Request data
 - Regrid
 - Diagnostics
- ESG return results

