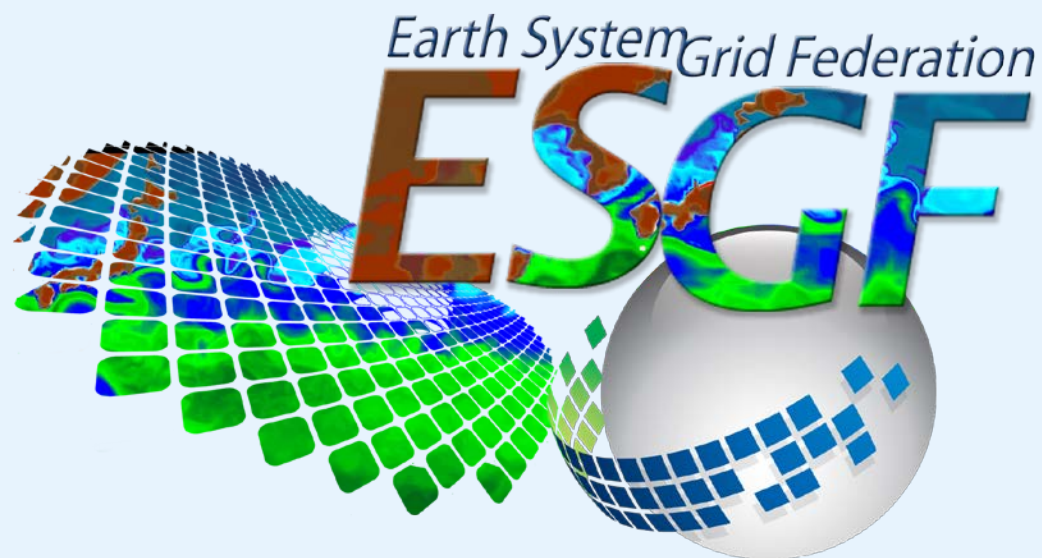


Earth System Science Portals: Earth System Grid Federation (ESGF)

Dean N. Williams on behalf of the ESGF Community

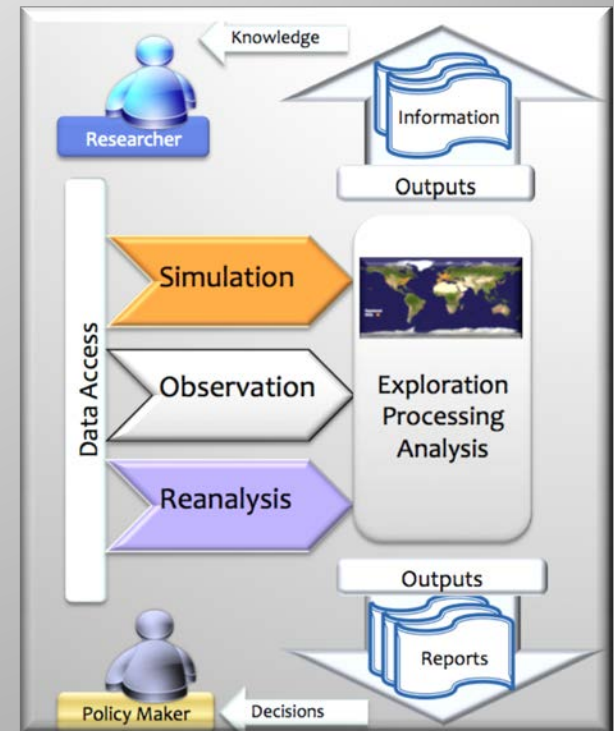
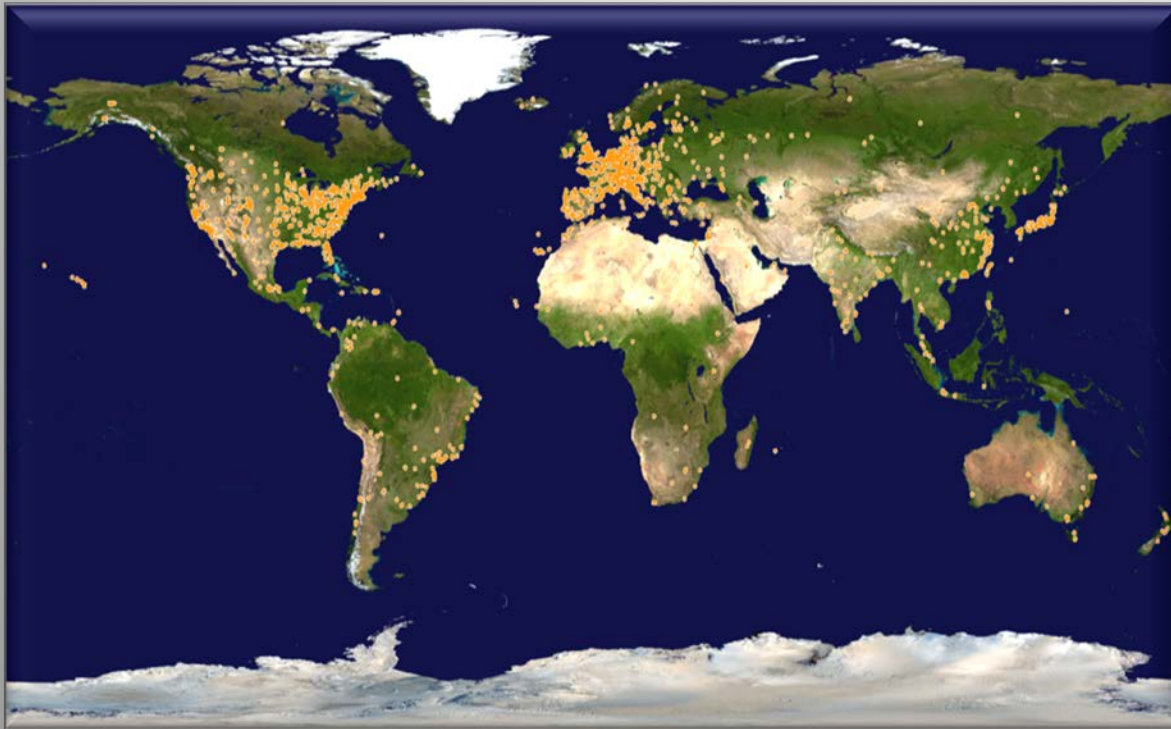
Federation and Integrations of Data from Multiple Sources

Climate and Earth System Modeling (CESM) Principal Investigators' Meeting ♦ September 22, 2011



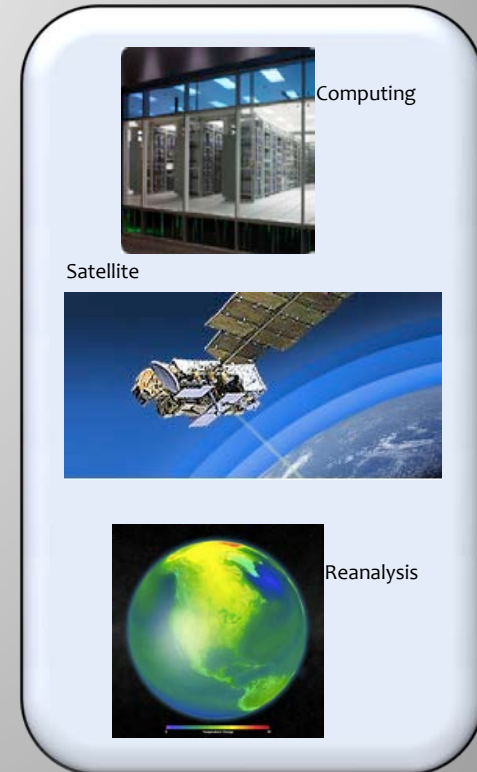
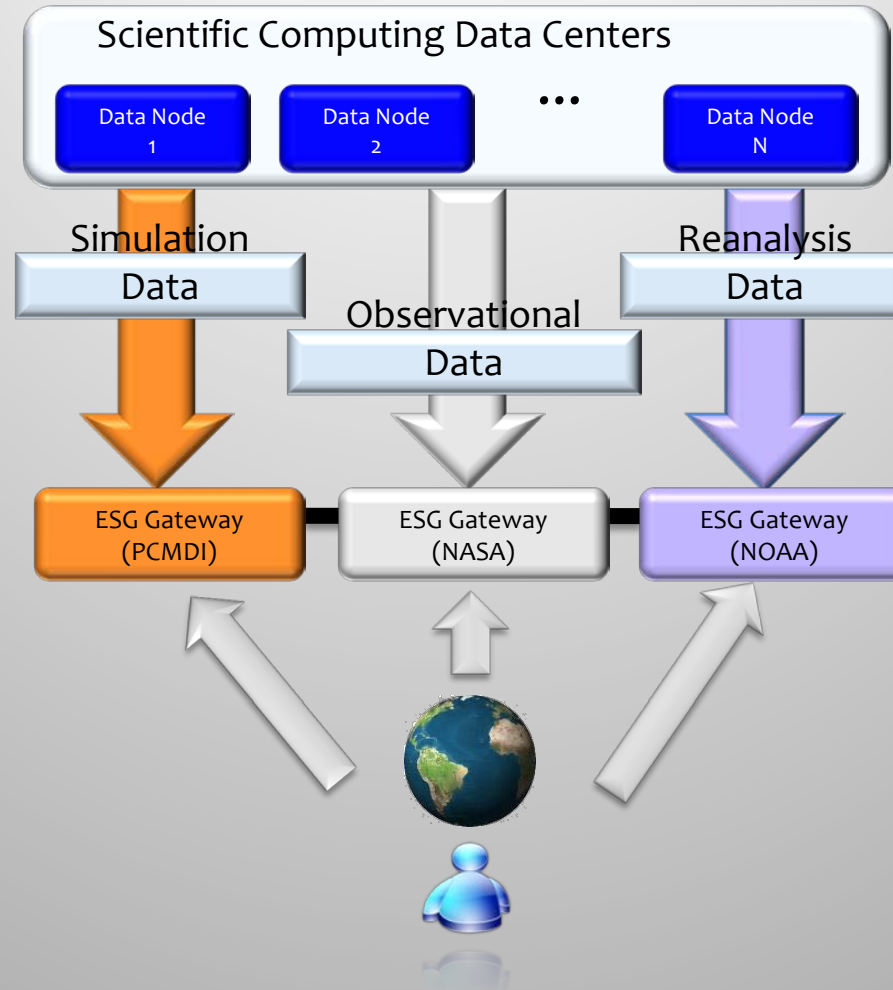
Earth System Grid Federation (ESGF): Coupled Model Intercomparison Project

- ESGF is a free, open consortium of institutions, laboratories and centers around the world that are dedicated to **supporting research of Climate Change**, and its environmental and societal impact
- Historically originated from **Earth System Grid (ESG)** project, expanded beyond its constituency and mission to include many other **partners in the U.S., Europe, Asia, and Australia**
- **Groups** working at many projects: ESG, Earth System Curator, Metafor, Global Interoperability Program, Infrastructure for the European Network for Earth System Modeling, and many more
- **U.S. funding** from DOE, NASA, NOAA, NSF



ESGF development of climate data management, access and analysis

- Too much data to move, **must leave in place**
- Promote **sharing** of knowledge, software and tools among partners
- Define **APIs and protocols** for interoperability among data centers
- **Collaborative development** of some software components
- Deployment of **common software infrastructure**

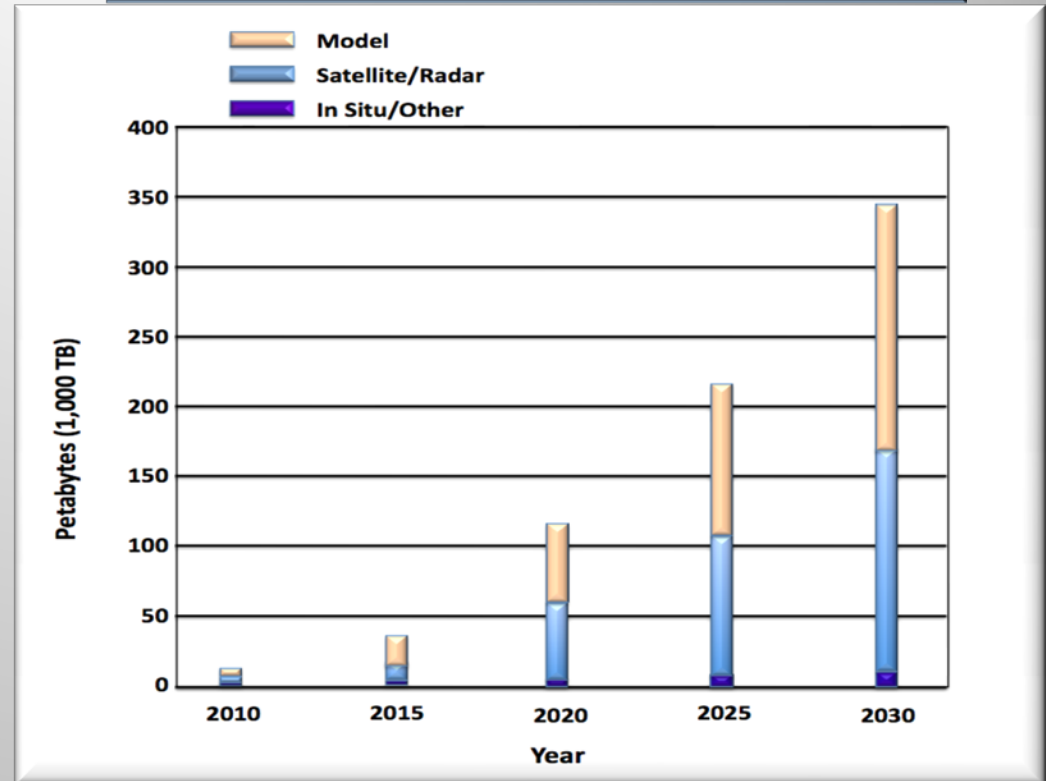


Community climate data storage requirements expands rapidly every five years as computer capabilities increase and climate simulations become more complex

- Rapidly expanding worldwide global **climate data holdings**
- Local and remote servers with **web-based and analysis tool access**.
- **>1 Tbps networking** must fit in with the timelines shown in the graph
 - End-to-end monitoring for centers and individual desktop transfers and **remote access and server-side analysis**
- **Climate network mailing list:** climatedata@es.net

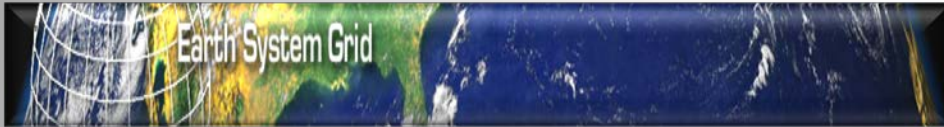
J.T. Overpeck, G.A. Meehl, S. Bony, D.R. Easterling

"Climate Data Challenges in the 21st Century", *Science*, 11 February 2011, Vol 331, 700 (2011); DOI: 10.1126/science.1197869"



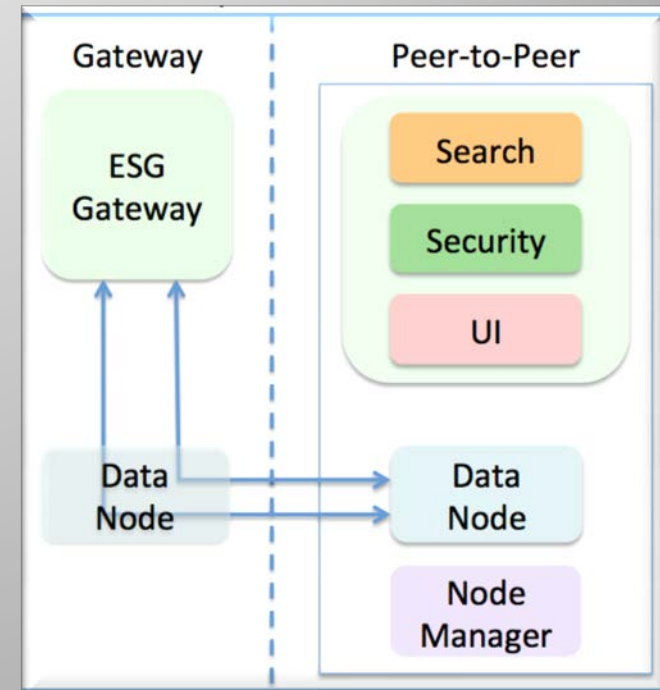
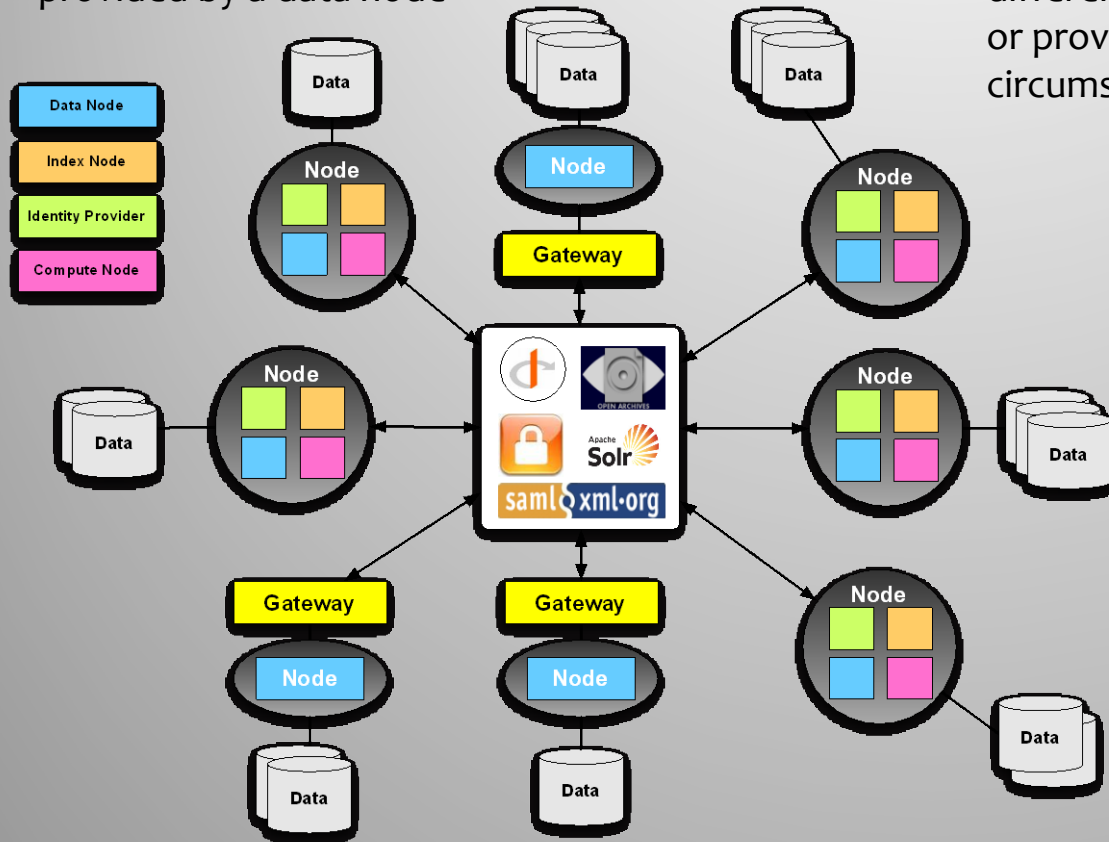
Transfer Rate	Time to Transport 1 TB of Data	Time to Transport 33.3 TB of Data	Time to Transport 1 PB of Data	Time to Transport 3 PB of Data
10-Gbps	13.65 minutes	7.6 hours	9.7 days	29.1 days
100-Gbps	81.9 seconds	45.5 minutes	23.3 hours	2.9 days
1-Tbps	8.19 seconds	4.55 minutes	2.33 hours	6.7 hours

Balanced Ecosystem: Overall Architecture Design



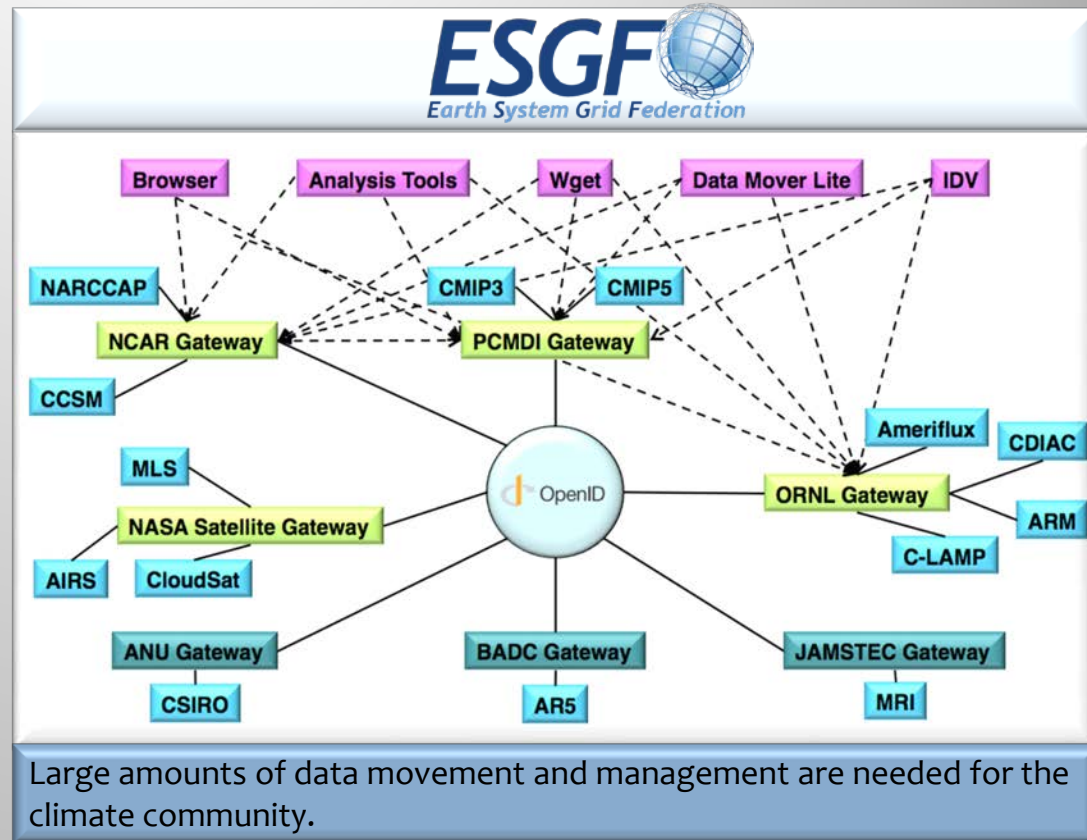
Gateway-data node (client-server): A more traditional model, based on a specialized gateway application that acts as a broker towards services provided by a data node

Peer-to-Peer (P2P): An innovative paradigm in which all participating sites interact as equal partners, can be flexibly configured to expose different sets of services, and can act as consumers or providers of services depending on circumstances



The ESGF is a distributed data archival and retrieval system

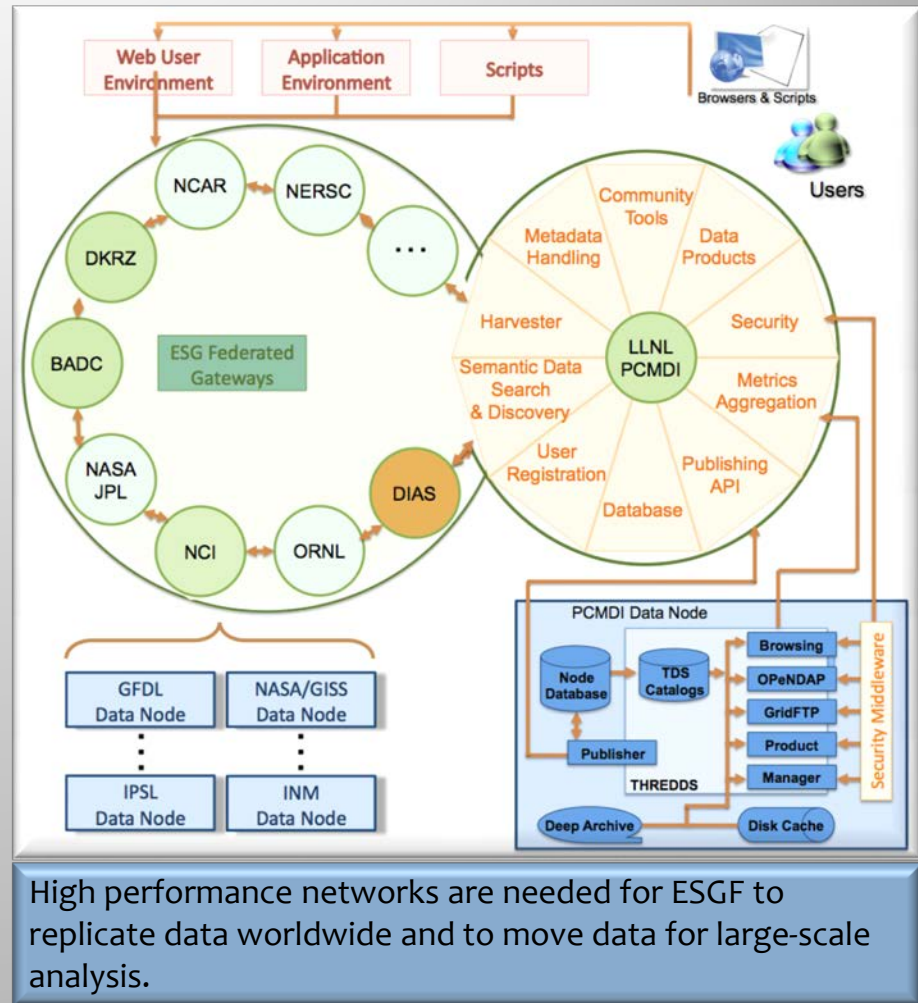
- Distributed and federated architecture
- Support discipline specific Gateways
- Support browser-based and direct client access
- Single Sign-on
- Automated GUI-based publication tools
- Full support for data aggregations
 - A collection of files, usually ordered by simulation time, that can be treated as a single file for purposes of data access, computation, and visualization
- User notification service
 - Users can choose to be notified when a data set has been modified



Networking the climate science community for large-scale data and science discovery

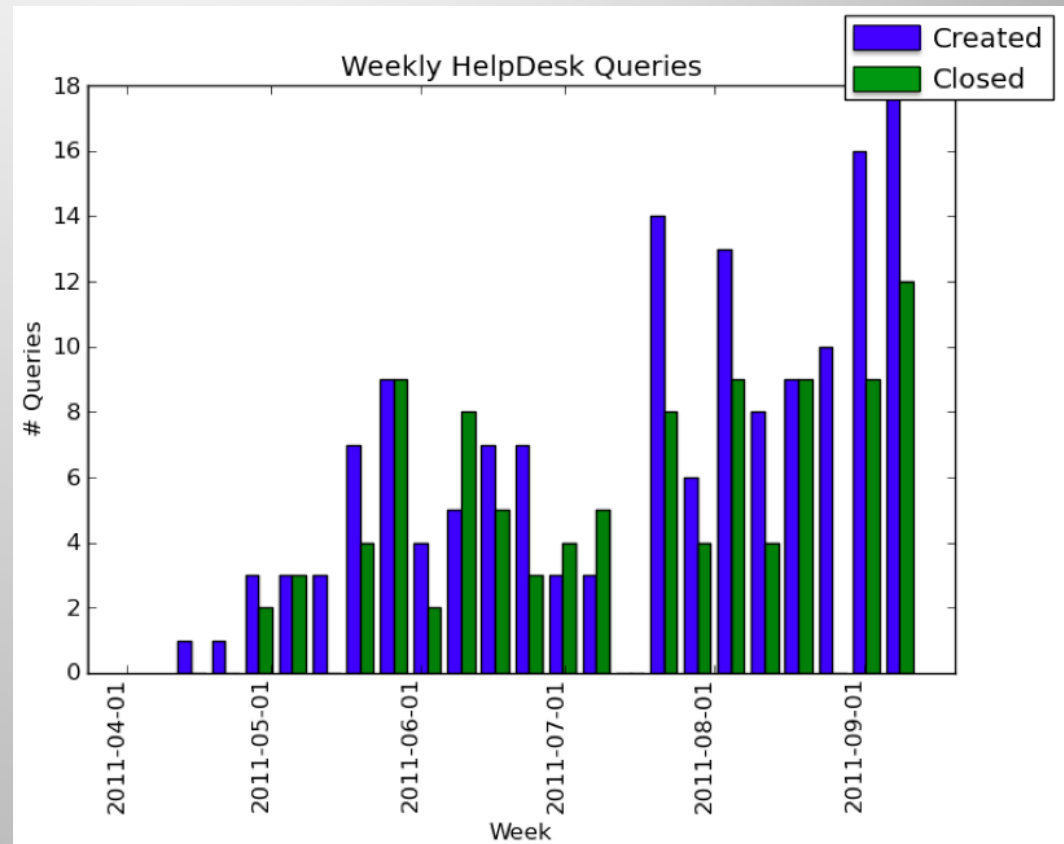


- CMIP = **Coupled Model Intercomparison Project**
 - Phase 1: Idealized simulations of present-day climate
 - Phase 2: Idealized simulations of future climate changes
 - Phase 3: More realistic simulations
- **CMIP 5** multi-model archive expected to include
 - 3 suites of experiments
 - 25 modeling centers in 19 countries
 - 58 models
 - Total data, ~10 PB
 - Replica 2 – 3 PB
- **Global distribution**
- **Timeline fixed** by IPCC (2011 – 2013)
- **Wide adoption** by the Climate community: CMIP, CCSM, CSSEF, ARMBE, Obs4MIPs, TAMIP, NCPP, NARCCAP, NCA, etc.



ESGF Help Desk Weekly Traffic

- About half of these queries come directly to cmip5-helpdesk@stfc.ac.uk and the other half to the esg-support mailing list esg-support@earthsystemgrid.org
- Questions that are resolved are placed on the [ESGF FAQ](#) list
- **Scientists** at BADC, PCMDI, and DKRZ are charged with addressing CMIP5 data questions
- **Technical staff** at BADC, PCMDI, DKRZ, NCAR, and JPL are charged with addressing ESGF system questions



Gateway and Peer-to-Peer Web Front End

ESG-NCAR Gateway - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.earthsystemgrid.org/home.htm

ESG-NCAR Gateway

Earth System Grid

Home Data Account About Contact Us Login

ESG Gateway at the National Center for Atmospheric Research

Search: Datasets for: Search Start Over

To conduct a search, select a category from the pull down menu and/or enter free text into the text box.

Search Categories

- Project
 - CCSM
 - CMIP5
 - NARCCAP
 - PCM
 - PyNGL
 - PyNIO
 - obs4MIPs
- Institute
- Model
- Experiment
- Frequency
- Product
- Realm
 - Aerosol
 - Atmosphere
 - Land
 - Land ice
 - Ocean
 - Sea ice
- Variable
- Ensemble

Global Climate Models

Community Earth System Model (CESM)

- CCSM 4.0 Model Output
- CCSM 3.0 Model Output
- Parallel Climate Model (PCM)

Regional Climate Models

NARCCAP: North American Regional Climate Assessment Program

Analysis & Visualization Software

- NCL: NCAR Command Language
- PyNGL: Python Interface to the NCL Graphic Libraries
- PyNIO: Python Interface for Multi-format Geoscientific Data I/O

Quick Links

- Getting Started Guide
- Create Account
- Browse Catalogs
- Search for Data

ESG Data Gateways

- ESG-BADC Gateway
- ESG-NCL Gateway
- ESG-NERSC Gateway
- ESG-DRML Gateway
- ESG-PCMDI Gateway
- ESG-WDCC Gateway
- NASA JPL Gateway

Other Gateways

- CADIS (Arctic)

http://www.earthsystemgrid.org/home.htm#

ESGF Portal

http://esgf-node2.llnl.gov/esgf-web-fe/

ESGF Earth System Grid Federation

P2P

Home Search Login

Welcome to pcmdi-node2

Quick Search

Keyword: Search

About pcmdi-node2

Resources

Quick Links

- Create Account
- Search for Data
- MyEzra Login

ESGF Earth System Grid Federation

Geospatial Search

Home Search Login

Search Type:

Envelopes Overlaps

Enter address:

Clear Markers

[1] lat 43.38, lon -87.98

[2] lat 34.59, lon -104.78

Define Area:

Square Circle

Submit Geospatial Constraints

Data Preview (server-side and client-side analysis)

The image displays a complex interface for data analysis, split into a web browser window and a desktop application window.

Web Browser (ESGF LAS):

- URL: `esg-datanode.jpl.nasa.gov/las/localGetUI.do?dsid=893EB2D5C79AD40EE2436A3F118649CE_ns_obs4MIPs.NASA-JPL.AIRS.mon.ta.1.aggregation.1`
- Navigation: Choose dataset, Update Plot, Set plot options, Animate, Compare, Google Earth, Show Values, Export to Desktop Application, Save As..., Link To..., Print.
- Dataset: obs4MIPs NASA-JPL AIRS L3 Monthly Data/obs4MIPs.NASA-JPL.AIRS.mon.ta.1.aggregation/obs4MIPs.NASA-JPL.AIRS.mon.ta.1.aggregation - Subset 1
- Map: Latitude-Longitude view with coordinates 90 N, 0.5 W, 0.5 W, 90 S.
- Plots: Includes a QFLX plot (Longitude-Depth) and a Cloudness Plot (Latitude-Longitude).
- Navigation: Choose dataset, Update Plot, Set plot, QFLX.
- Plots: Includes a QFLX plot (Longitude-Depth) and a Cloudness Plot (Latitude-Longitude).
- Navigation: Choose dataset, Update Plot, Set plot, QFLX.
- Plots: Includes a QFLX plot (Longitude-Depth) and a Cloudness Plot (Latitude-Longitude).

Desktop Application (UV-CDAT):

- Variable: `ct`
- Plot: `Isotiff`
- Variable Properties: Name: `ct`, Load variable from: `ESGF`, Search: `project=obs4cmip5`
- Search Results: `Search 1 Variable=hus`, `Unmapped Files`, `hus`, `Obs-AIRS`, `obs4MIPs.NASA-JPL.AIRS.mon.hus.AIRS.L3.RetStd-v5_20...`, `HTTPServer @ http://esg-datanode.jpl.nasa.gov/thredds...`, `OPENDAP @ http://esg-datanode.jpl.nasa.gov/thredds...`, `Obs-MLS`, `obs4MIPs.NASA-JPL.MLS.mon.hus.MLS.L3_v02-2x_20040...`
- Authentication: Please enter your credentials for Openid on ESGF, `openid3.jpl.nasa.gov`, `2119`, `donovan1`, Password: `*****`

US Department of Commerce NOAA | OAR | PMEL | Contacts | Privacy Policy | Disclaimer | mailto: brownrig@u



Examining future developments for the ESGF has potential impact on other scientific domains

- **Other scientific domains:** biology, chemistry, nuclear energy, high-energy physics, etc.
- **Peer-to-Peer Architecture** – building an ecosystem
- **Analysis** services for:
 - extremely large data sets
 - multiple large data sets are not co-located
 - cloud computing
- **Data integration** and advanced metadata capabilities
- **Advanced product services** via multiple scripting languages
- Integration of security assertion Markup Languages (SAML) identity providers
- **Measuring replication and data access patterns** in extreme scale ESG
- **Workflow and provenance**
- **Virtual Organization** management as **Software as a Service**
- **Advanced networks** as easy-to-use community resources
- Management of **open source, community-driven** software development

