



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

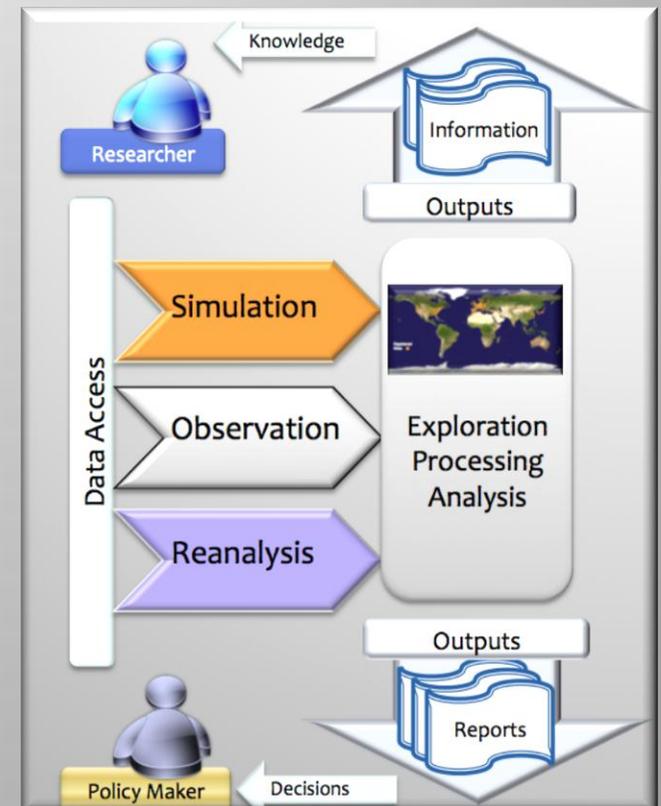
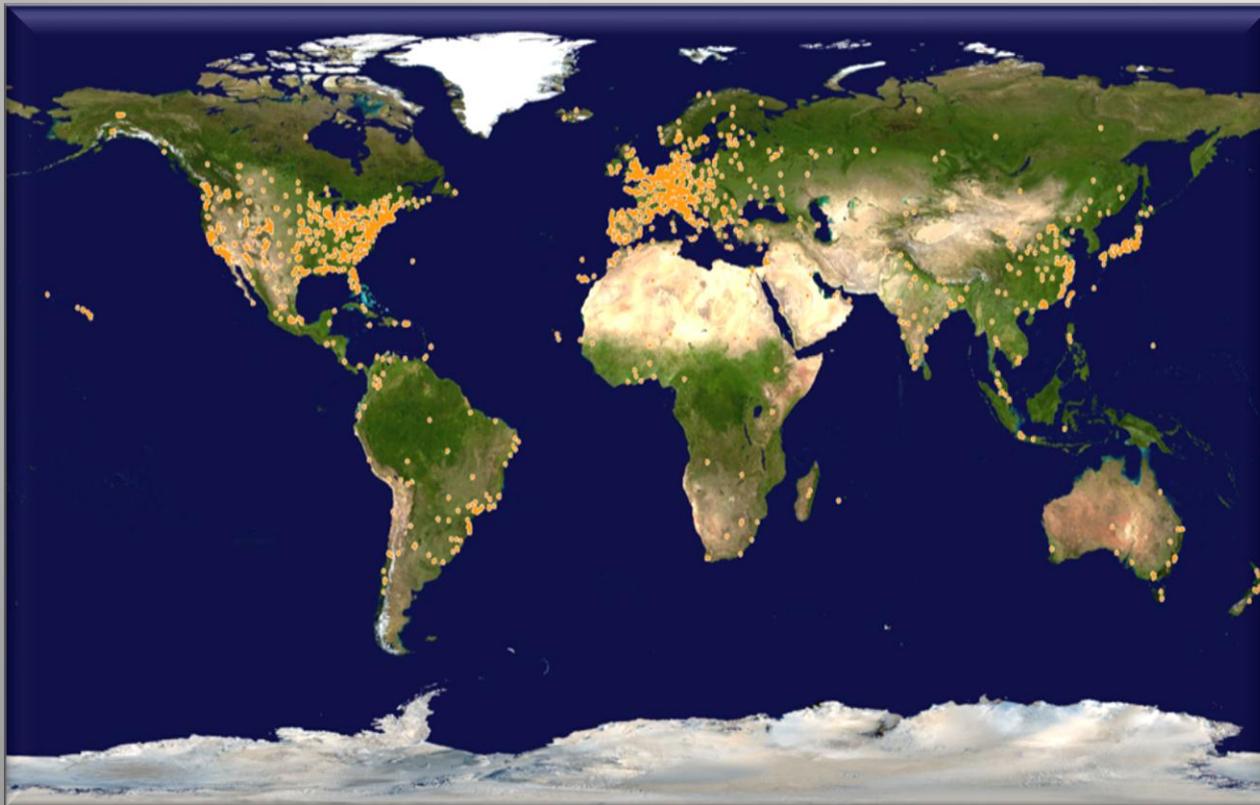
Dean N. Williams on behalf of the ESGF P2P Community  
Federation and Integrations of Data from Multiple Sources

ESPC Workshop ♦ March 21-23, 2012



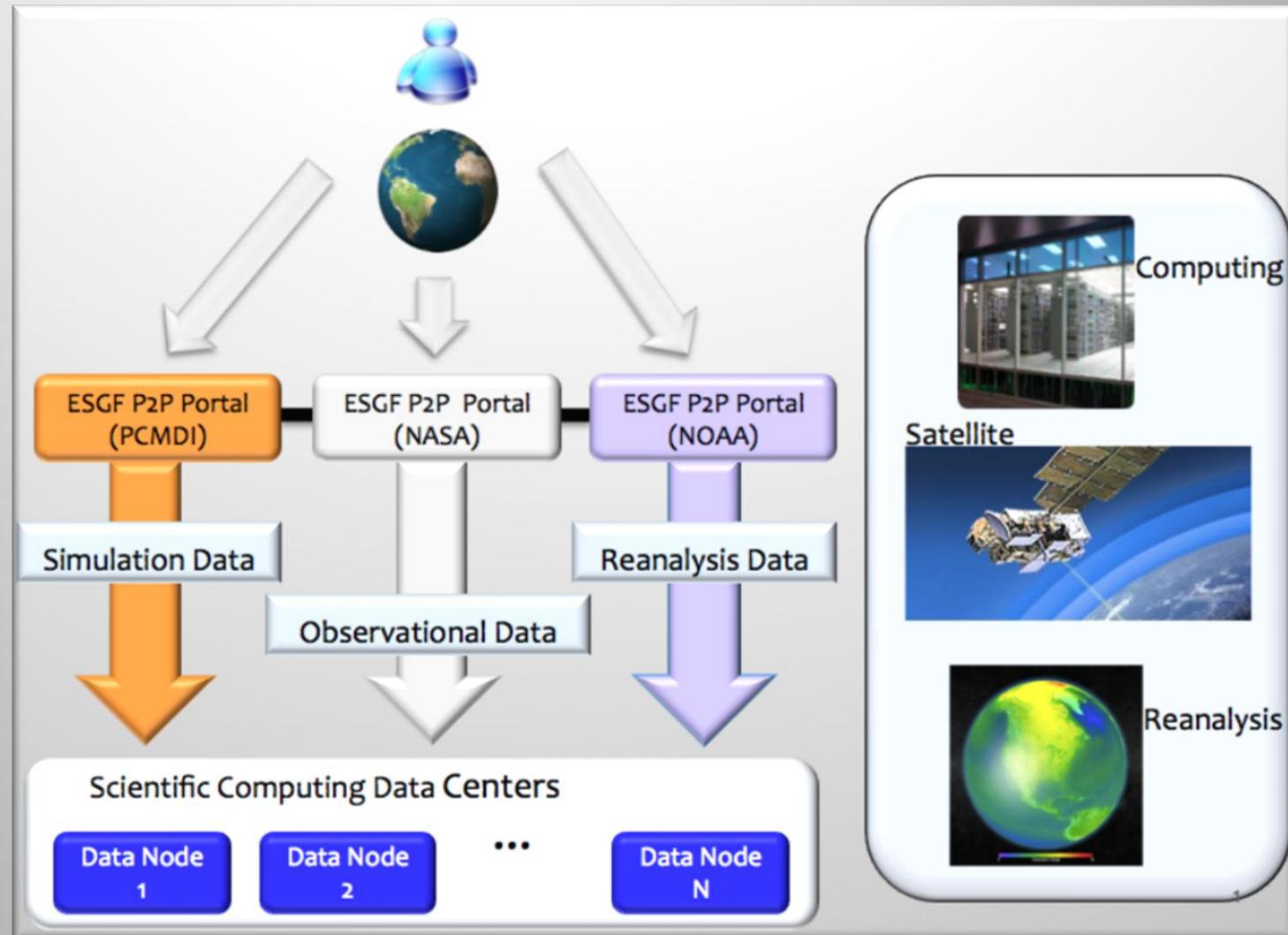
# Earth System Grid Federation (ESGF): Coupled Model Intercomparison Project

- 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



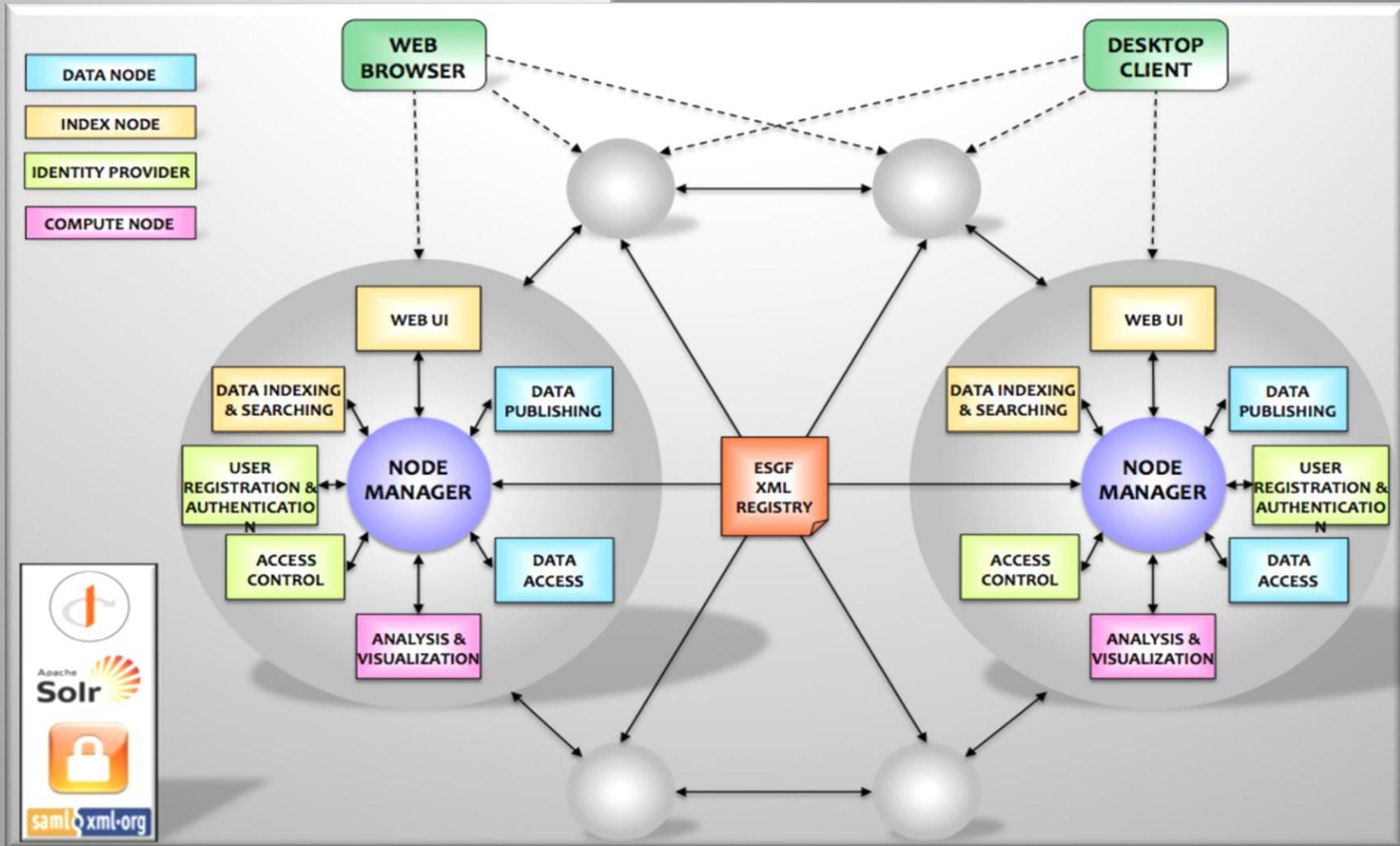
# ESGF development of climate data management, access and analysis

- 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**



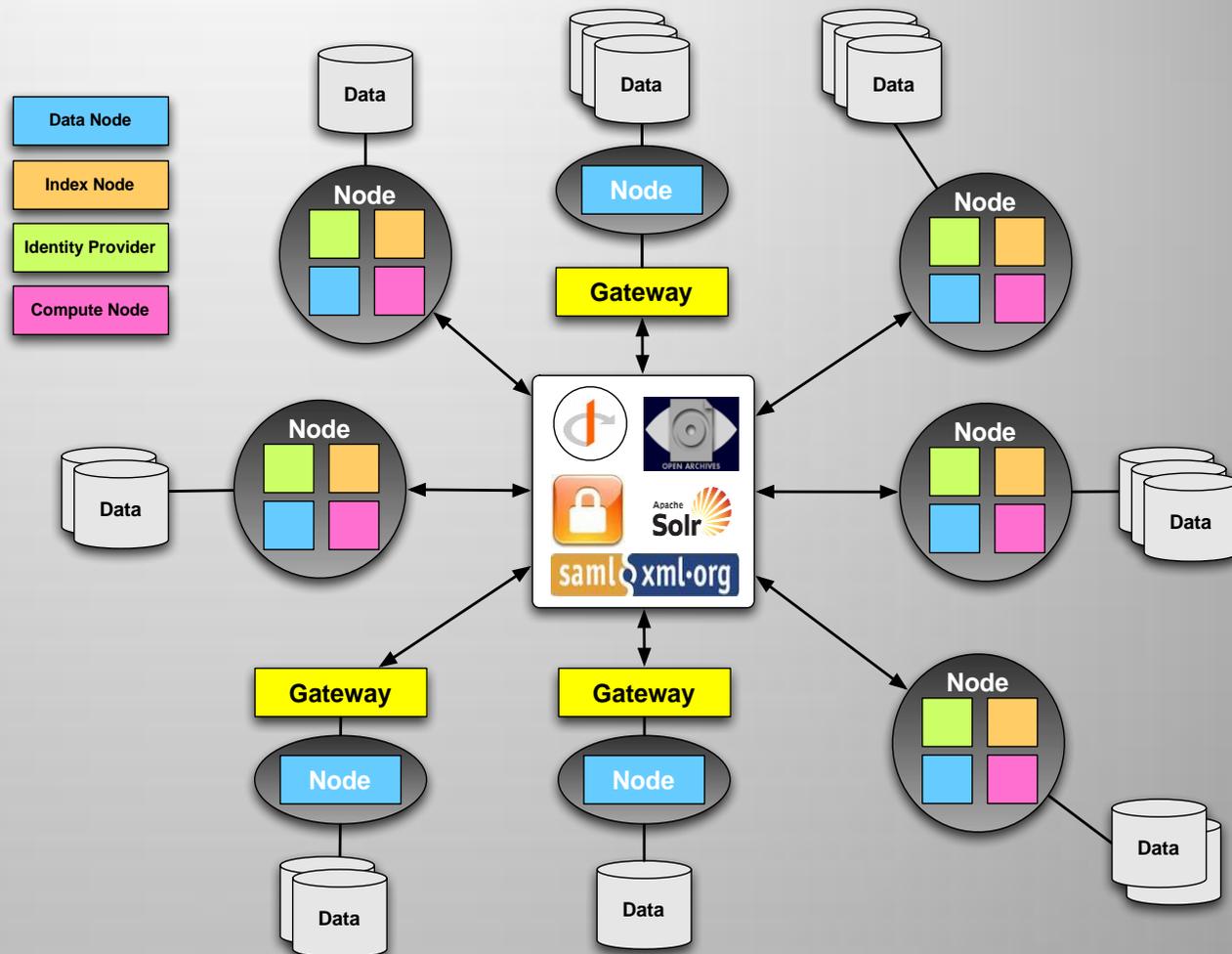
# Balanced Ecosystem: Overall Architecture Design

**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

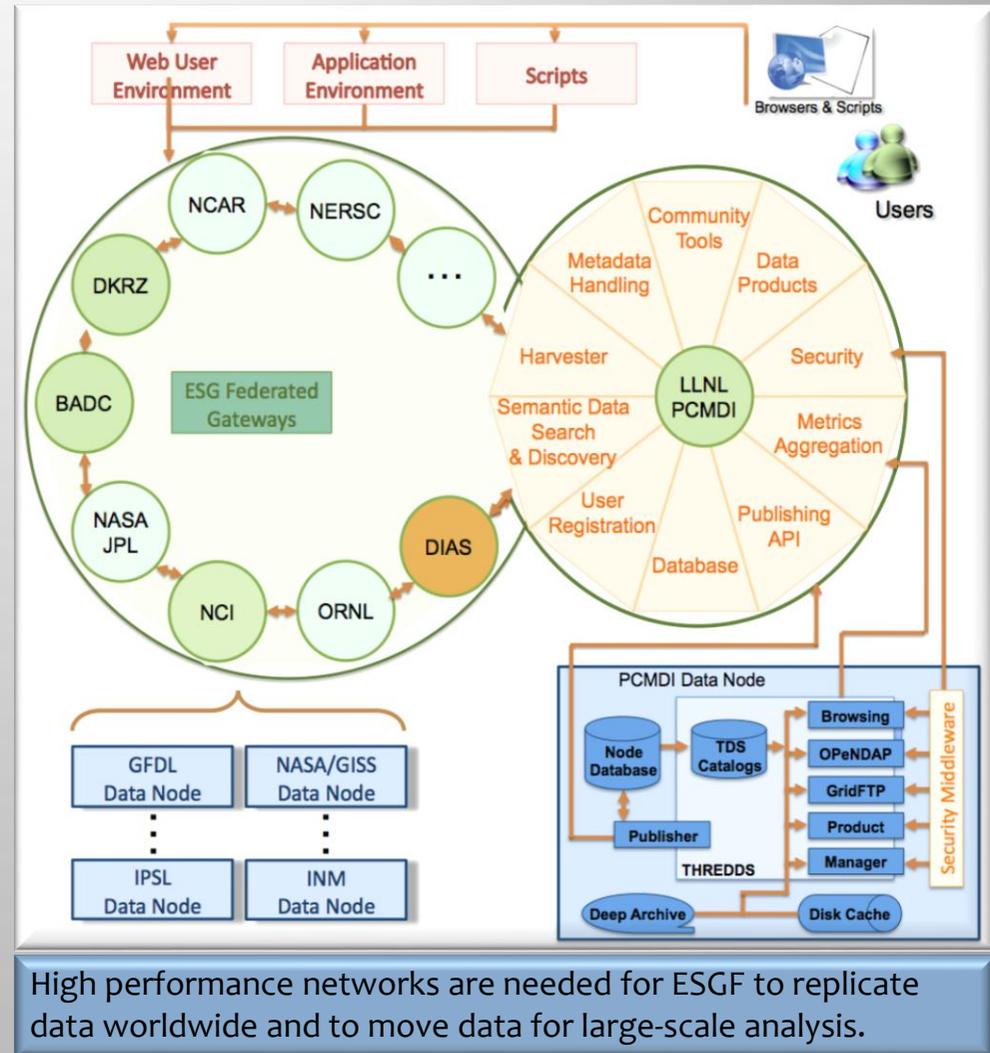
- **Peer-to-Peer Architecture** – building an ecosystem
- **Analysis** services for:
  - extremely large data sets
  - multiple large data sets are not co-located
  - cloud computing
- **Data integration** and advanced metadata capabilities
- **Advanced product services** via multiple scripting languages
- Integration of security assertion Markup Languages (SAML) identity providers
- **Measuring replication and data access patterns** in extreme scale ESGF
- **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



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

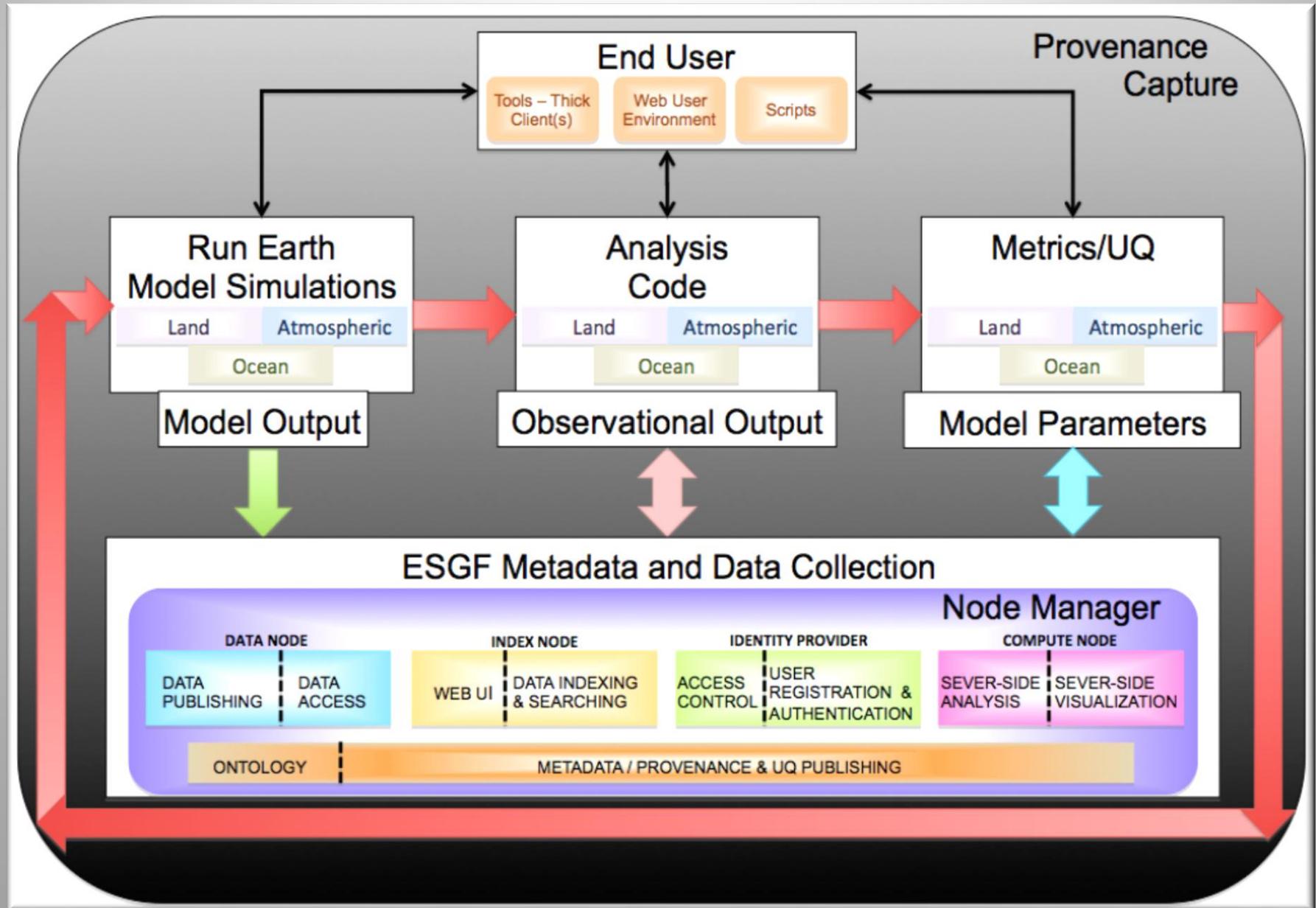


- CMIP = **Coupled Model Intercomparison Project**
  - Phase 1: Idealized simulations of present-day climate
  - Phase 2: Idealized simulations of future climate changes
  - Phase 3: More realistic simulations
- **CMIP 5** multi-model archive expected to include
  - 3 suites of experiments
  - 25 modeling 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.



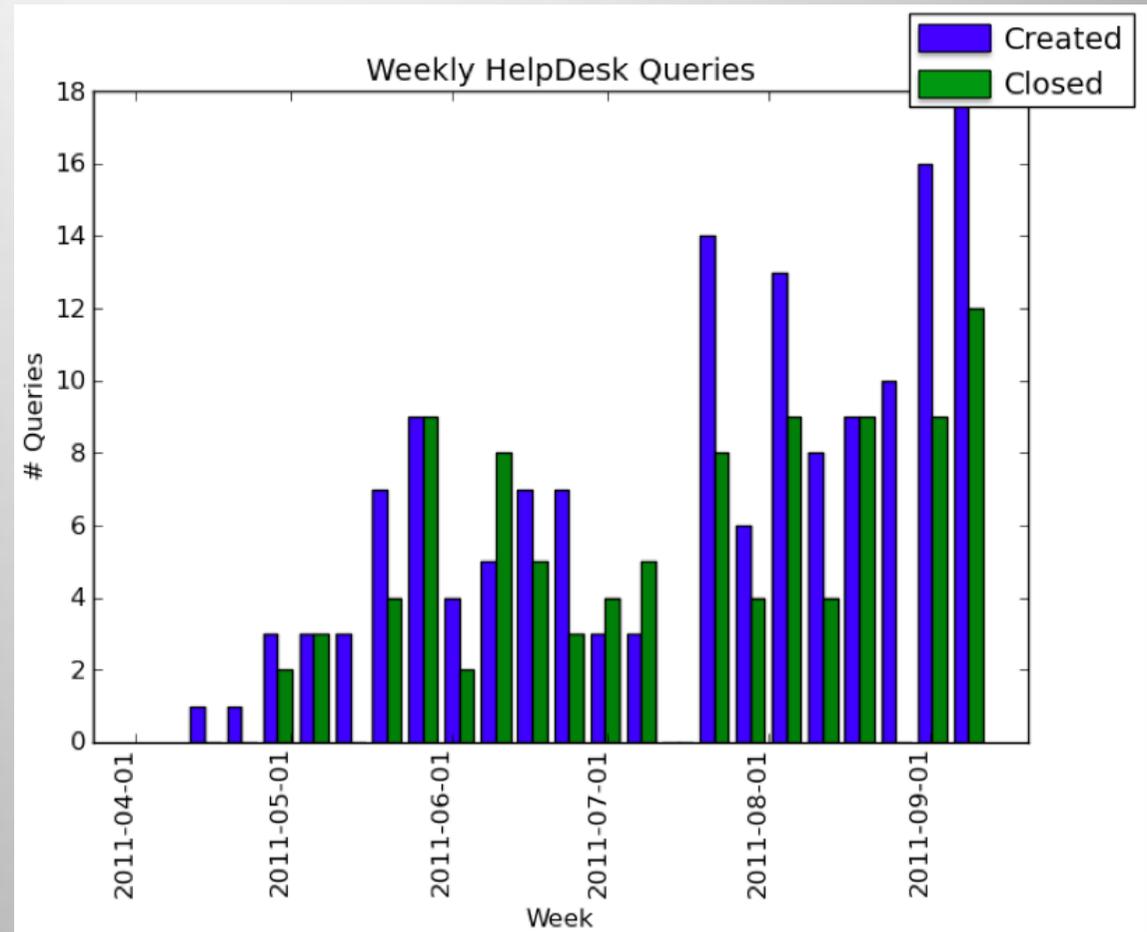
High performance networks are needed for ESGF to replicate data worldwide and to move data for large-scale analysis.

# Climate Science for a Sustainable Energy Future (CSSEF)



# ESGF Help Desk Weekly Traffic

- About half of these queries come directly to [cmip5-helpdesk@stfc.ac.uk](mailto:cmip5-helpdesk@stfc.ac.uk) and the other half to the esg-support mailing list [esg-support@earthsystemgrid.org](mailto: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



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

ESGF Portal

http://pcmdi9.lln.gov/esgf-web-fe/

ESGF Earth System Grid Federation

PCMDI

Welcome to this ESGF P2P Node

Quick Search

Keyword:  Search

Advanced Search (Category, Geospatial, Temporal, and more)...

Peer Nodes

- ANL Node
- CMCC Node
- DKRZ Node
- IPSL Node
- NASA-JPL Node
- ORNL Node
- PCMDI Node
- PNNL Node

About esgf-pcmdi-9

The PCMDI mission is to develop improved methods and tools for the diagnosis and intercomparison of general circulation models (GCMs) that simulate the global climate. The need for innovative analysis of GCM climate simulations is apparent, as increasingly more complex models are developed, while the disagreements among these simulations and relative to climate observations remain significant and poorly understood. The nature and causes of these disagreements must be accounted for in a systematic fashion in order to confidently use GCMs for simulation of putative global climate change.

Resources

Quick Links

- Create Account
- MyProxy/Logon
- Expert Search (XML)
- Wget Script Generator
- ESGF aggregated RSS feed
- Contact ESGF

Instructions

- ESGF Full User Guide
- Search Help
- Wget Scripts FAQ
- Wget Scripting
- Tutorial: Download Strategies
- Using Globus Online
- Subscribing to RSS Notification

ESGF LAS

Live Access Server About LAS

Choose dataset Update Plot Set plot options Animate Co

obs4MIPs NASA-JPL AIRS L3 Monthly Data/obs4MIPs NASA-JPL AIRS m

ta

Search Categories

- project
- model
- experiment
- frequency
- realm
- instrument
- variable
- cf\_variable

Z (Pa)

TIME : 1

80°N

40°N

0°

40°S

90 N

0.5 W

0.5 W

90 S

MAPS

- Latitude-Longitude

DEPTH PROFILES

- Longitude-Depth
- Latitude-Depth

Geospatial Search

Search Type:  Encloses  Overlaps

Enter address:

Clear Markers

[1] lat 43.38, lon -87.98  
[2] lat 34.59, lon -104.76

Define Area:  Square  Circle

Submit Geospatial Constraints

SUBSURFACE METEOROLOGY DATA - NCDC (E5F)



UV-CDAT

Projects

Cloudiness Plot Plot 2

Variables

New Variable

var 1  
var 2  
var 3  
var 4

Workflow

```

    graph TD
      A[CMWS_FilterWorkflow] --> B[Difference CMWS_FilterWorkflow]
      B --> C[CMWS_InstrumentWorkflow]
      C --> D[VolumeRender]
      D --> E[SV3DGui]
      E --> F[LevelSurface]
      E --> G[SV3DGui]
      F --> H[LevelSurface]
      F --> I[SV3DGui]
  
```

```

vslicer = load_workflow_as_function('vtdv3d.vt', 'slicer')
vslicer(variable='temperature')
vrrender = load_workflow_as_function('vtdv3d.vt', 'vr')
vrrender(variable='temperature')
    
```

Script

Provenance

Hovmöller Plots

- Latitude-Longitude
- Longitude-Time
- Latitude-Time

LINE PLOTS

- Time Series
- Longitude
- Latitude

Date: Feb 100

Apply analysis

0 30E 60E 90E 120E 150E 180 150W 120W

0 4e-05 8e-05 0.000

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