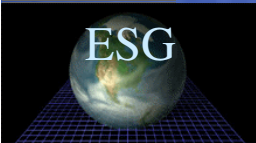


# The Earth System Grid (ESG)

Presented by Don Middleton & Luca Cinquini  
NCAR Scientific Computing Division  
**On Behalf of the ESG Team**

SCD Executive Committee  
February 25, 2003



# The Earth System Grid

<http://www.earthsystemgrid.org>

- U.S. DOE Scidac funded R&D effort
- Build an “Earth System Grid” that enables management, discovery, distributed access, processing, & analysis of distributed terascale climate research data
- A “Collaboratory Pilot Project”
- Build upon ESG-I, Globus Toolkit™, DataGrid technologies, and *deploy*
- Potential broad application to other areas



The Earth System Grid



# ESG Team

- ANL
  - Ian Foster (PI)
  - Veronika Nefedova
  - (John Bresenhan)
  - (Bill Allcock)
- LBNL
  - Arie Shoshani
  - Alex Sim
- ORNL
  - David Bernholdt
  - Kasidit Chanchio
  - Line Pouchard
- LLNL/PCMDI
  - Bob Drach
  - Dean Williams (PI)
- USC/ISI
  - Anne Chervenak
  - Carl Kesselman
- NCAR
  - David Brown
  - Luca Cinquini
  - Peter Fox
  - Jose Garcia
  - Don Middleton (PI)
  - Gary Strand

NCAR





# Basic Numbers

- T42 CCSM (current, 280km)
  - 7.5GB/yr, 100 years -> .75TB
- T85 CCSM (140km)
  - 29GB/yr, 100 years -> 2.9TB
- T170 CCSM (70km)
  - 110GB/yr, 100 years -> 11TB



The Earth System Grid



# Capacity-related Improvements

Increased turnaround, model development, ensemble of runs

***Increase by a factor of 10, linear data***

- Current T42 CCSM

- 7.5GB/yr, 100 years  $\rightarrow .75\text{TB} * 10 = 7.5\text{TB}$

The ESG logo features a stylized globe with a grid pattern, set against a dark background. The letters "ESG" are prominently displayed in a white, serif font above the globe.

ESG

The Earth System Grid

The logo for "Scientific Discovery through Advanced Computing" features a circular, multi-colored pattern resembling a stylized flower or a complex network, with the text "Scientific Discovery through Advanced Computing" written in a white, sans-serif font to its right.

Scientific Discovery  
through Advanced  
Computing

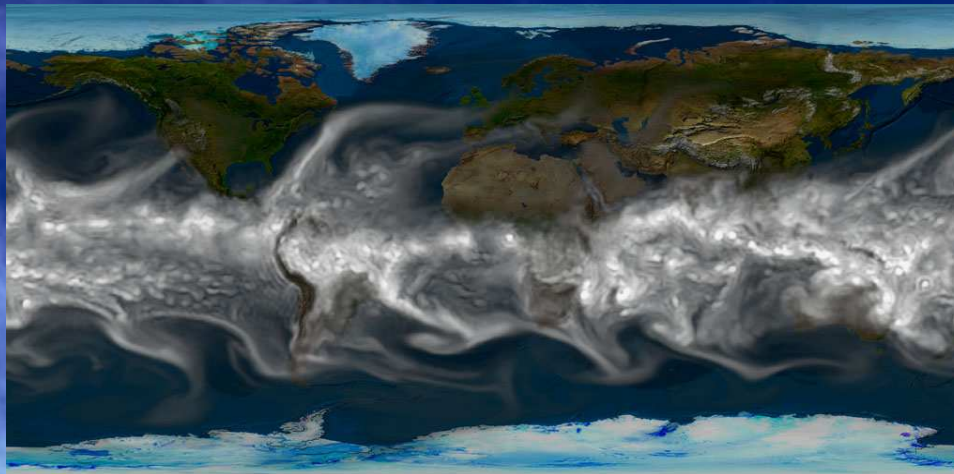
# Capability-related Improvements

Spatial Resolution: T42 -> T85 -> T170

***Increase by factor of ~ 10-20, linear data***

Temporal Resolution: Study diurnal cycle, 3 hour data

***Increase by factor of ~ 4, linear data***



CCM at T170 (70km)

ESG

The Earth System Grid





# Capability-related Improvements

Quality: Improved boundary layer, clouds, convection, ocean physics, Improved land model, river runoff, new sea ice

***Increase by another factor of 2-3, data flat***

Scope: Atmospheric chemistry (sulfates, ozone...)

Biogeochemistry (carbon cycle, ecosystem dynamics)

Middle Atmosphere Model

***Increase by another factor of 10+, linear data***

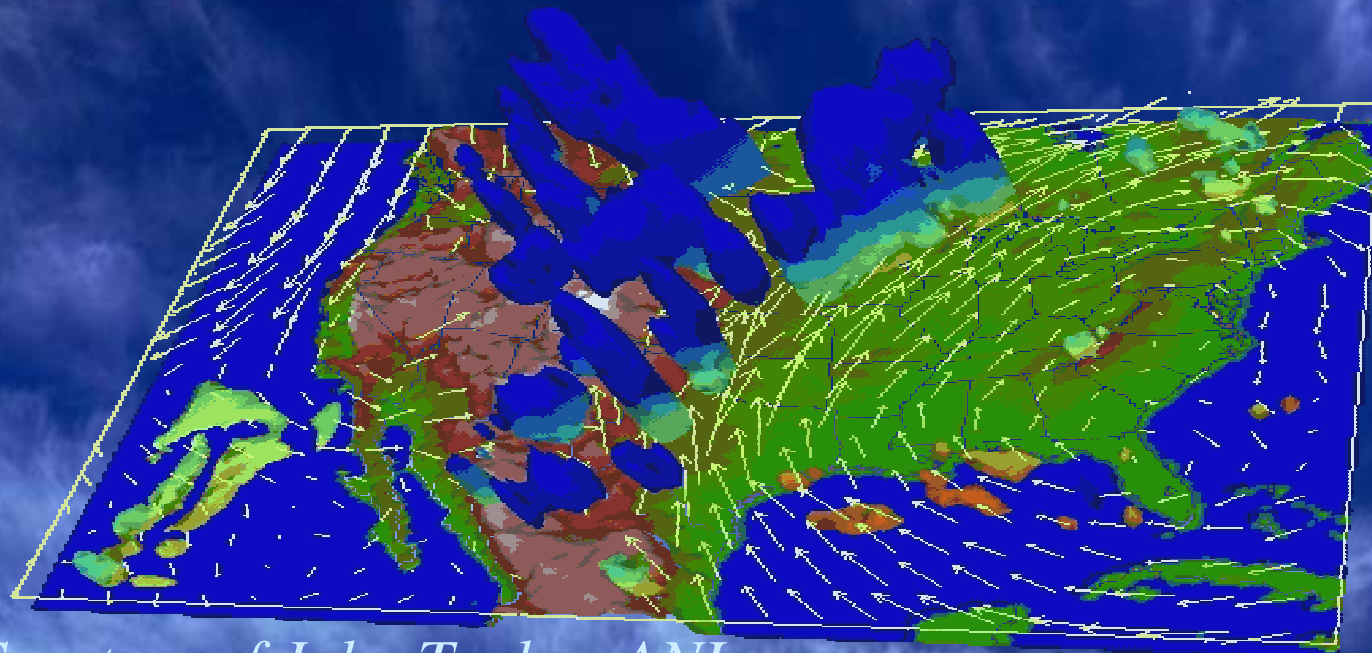


The Earth System Grid



# Approaching Mesoscale (i.e. “weather”) Resolution

05:00:00  
06 Jul 93  
5 of 24  
Tuesday



*Courtesy of John Taylor, ANL*

ESG

The Earth System Grid

Vis5D

Scientific Discovery  
through Advanced  
Computing



# Model Improvements cont.

Grand Total:

*Increase compute by a Factor  $O(1000-10000)$*

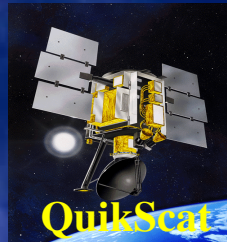
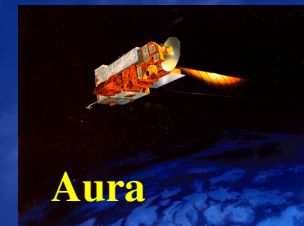
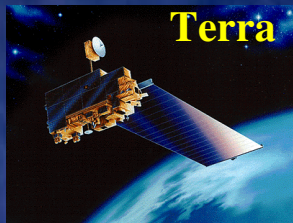


The Earth System Grid

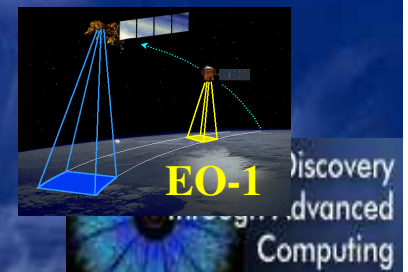
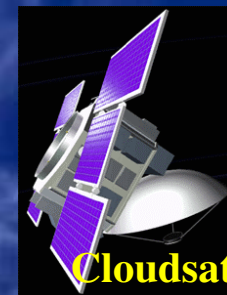
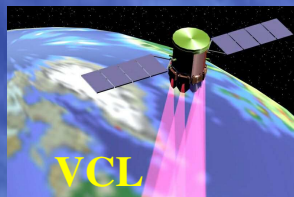


# We Will Examine Practically Every Aspect of the Earth System from Space in This Decade

## Longer-term Missions - Observation of Key Earth System Interactions



## Exploratory - Explore Specific Earth System Processes and Parameters and Demonstrate Technologies



The Earth System Grid



## Primary ESG Servers

Mass storage,  
disk cache,  
and computation



Web and applications-  
based access to  
management, discovery,  
analysis, and  
visualization

**NCAR:** Climate  
change  
prediction and  
data archive

**LBNL/NERSC:**  
Climate  
data archive

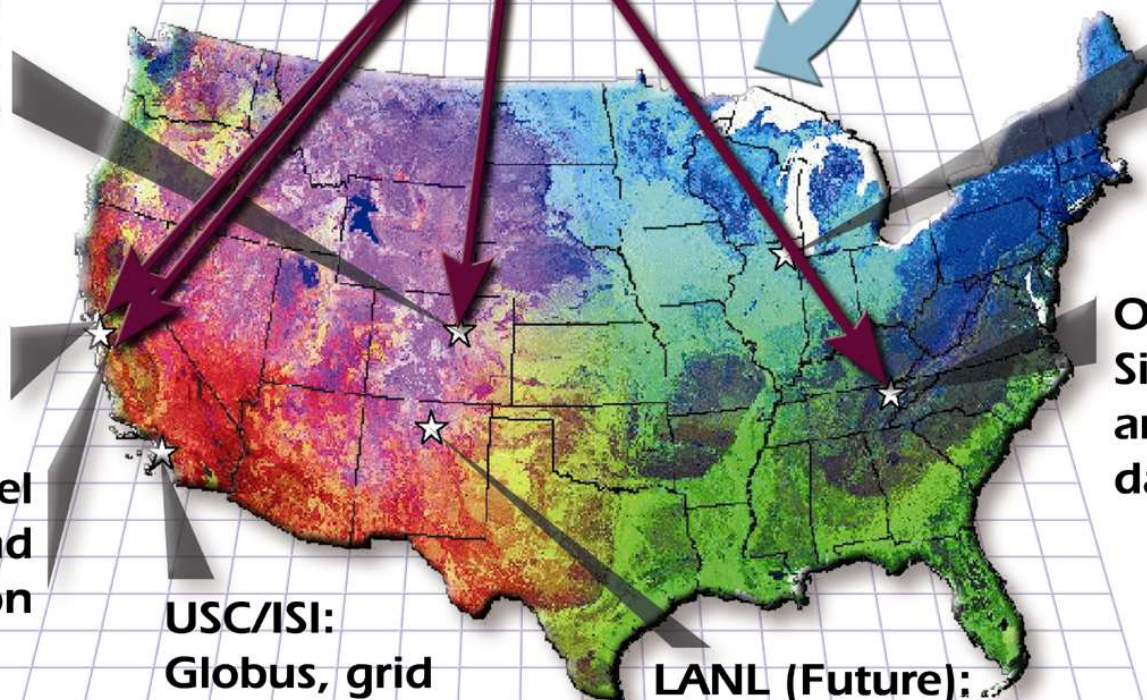
**LLNL:** Model  
diagnostics and  
inter-comparison

**USC/ISI:**  
Globus, grid  
applications, and  
metadatabases

**LANL (Future):**  
Climate and ocean  
data archive

**ANL:**  
Globus  
and grid  
applications

**ORNL:**  
Simulation  
and climate  
data archive





# What Is “The Grid”?

*Central Concept: “Coordinated resource sharing and problem-solving in dynamic multi-institutional virtual organizations”*

- Analogous to the “power grid”
- A “megatrend”...
- Foundations for a meta-OS?



The Earth System Grid



# The Globus Toolkit™



## *An Open Source Project*

- Security (!)
- Directory, Metadata, and Replica Services
- Resource Management
- Data Access and Management
- Distributed Computation
- Coming Soon – Open Grid Services Architecture (OGSA)
  - Reliable, persistent web services

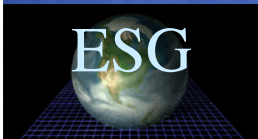


The Earth System Grid



# Corporate Commitments...

- Compaq
- Cray
- Sun
- SGI
- Veridian
- Entropia
- Microsoft
- IBM
- NEC
- Fujitsu
- Hitachi
- Platform Computing
- Cisco



The Earth System Grid





# ESG: Challenges

- Enabling the simulation and data management team
- Enabling the research community in analyzing and visualizing results
- Enabling broad multidisciplinary communities to access simulation results

*We need integrated “cyberinfrastructure” to enable smooth WORKFLOW for knowledge development: compute platforms, collaboration & collaboratories, data management, access, distribution, and analysis.*



The Earth System Grid



# ESG: Strategies

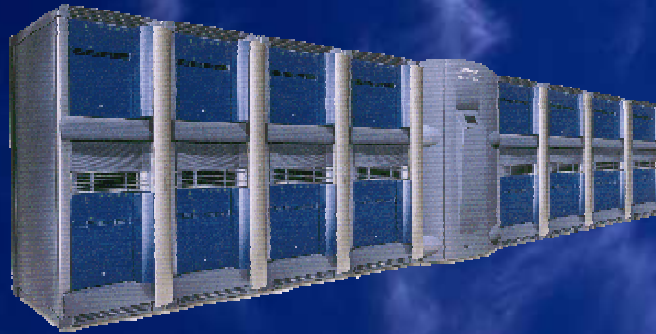
- Move data a minimal amount, keep it close to computational point of origin when possible
  - *Data access protocols, distributed analysis*
- When we must move data, do it fast and with a minimum amount of human intervention
  - *Storage Resource Management, fast networks*
- Keep track of what we have, particularly what's on deep storage
  - *Metadata and Replica Catalogs*
- Harness a federation of sites
  - *Globus Toolkit -> The Earth System Grid -> The UltraDataGrid*



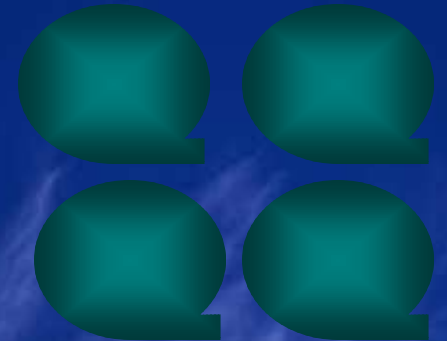
The Earth System Grid



Tera/Peta-scale  
Archive



HRM



Server

Storage Resource  
Management tools  
for reliable staging,  
replication,  
transport

HRM

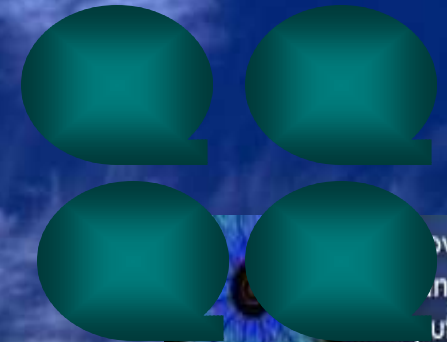
Server



Client  
Selection  
Control  
Monitoring

HRM

Tera/Peta-scale  
Archive



The Global System Grid

very  
anced  
uting

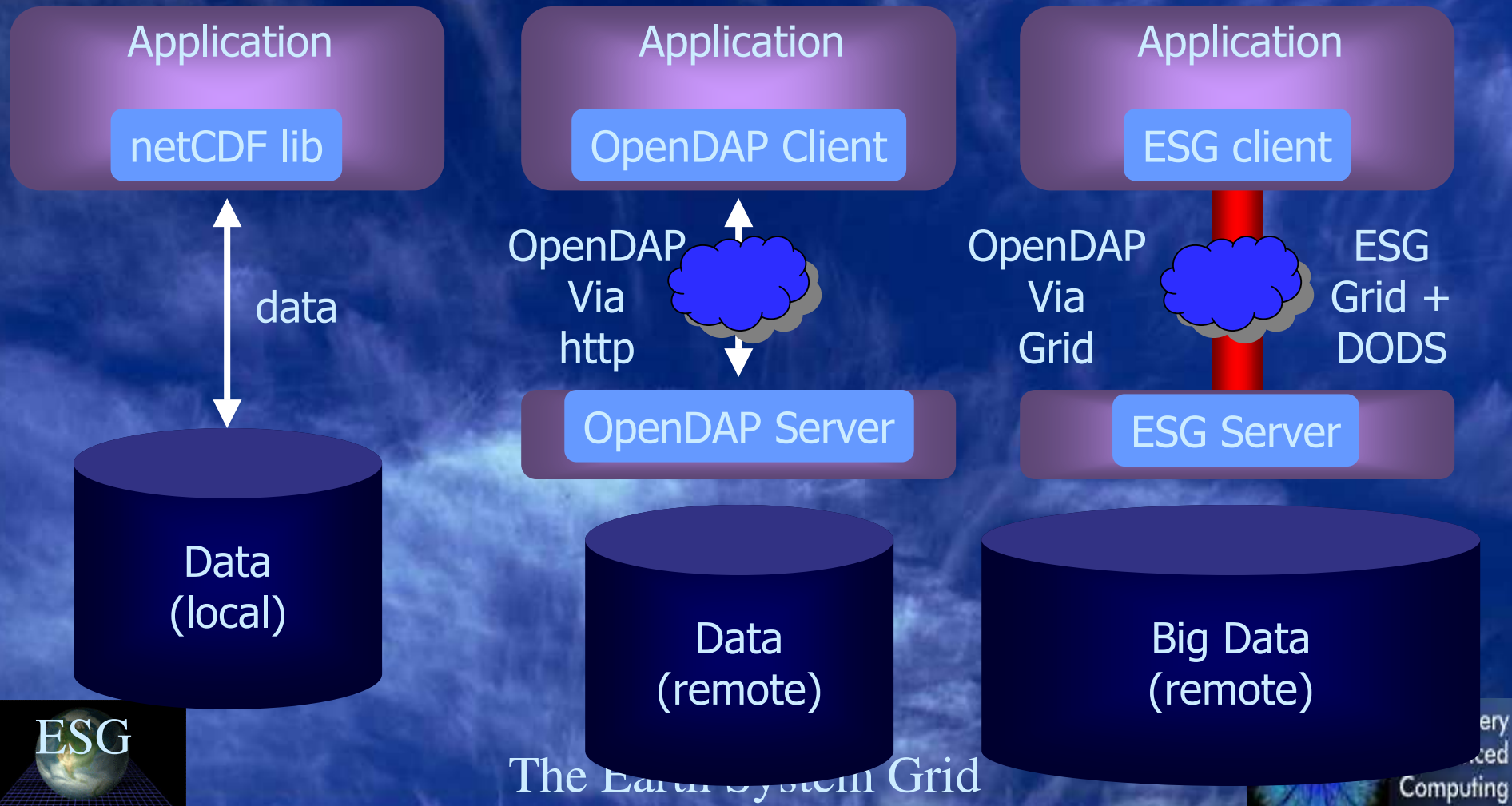


# Distributed Data Access Protocols

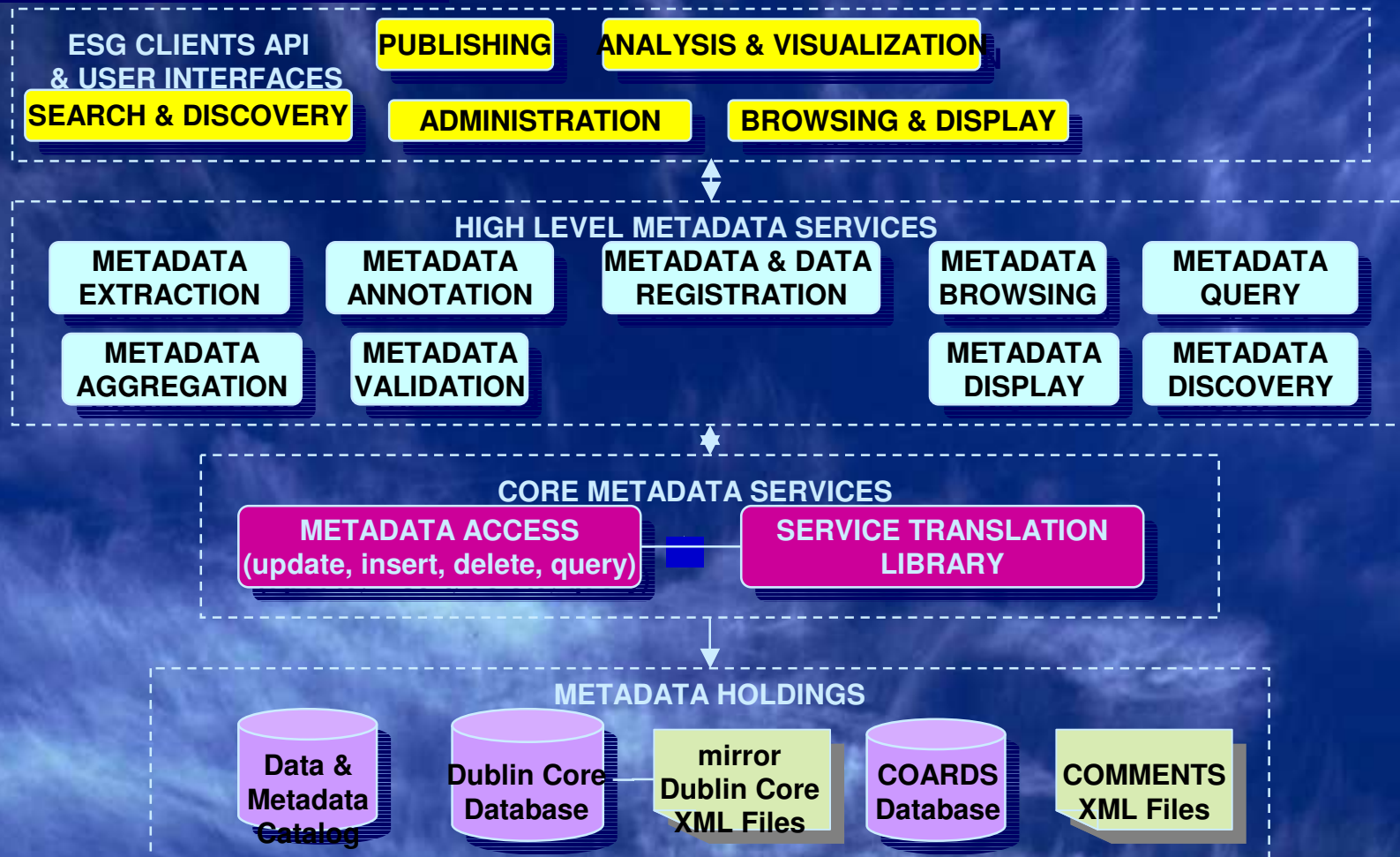
Grid + OpenDAP  
-Transparency  
-Performance  
-Security  
-Resource Mgmt  
-Analysis functions

Typical Application

Distributed Application



# ESG: Metadata Services



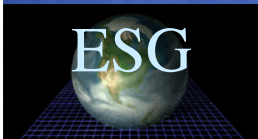
ESG

The Earth System Grid

Scientific Discovery  
through Advanced  
Computing

# Metadata

- Co-developed NcML with Unidata
- Finalizing a specific schema for PCM/CCSM
- Addressing interoperability via the generation of DIF/FGDC
- Addressing interoperability with digital libraries via the creation of Dublin Core
- Experimenting with relational and native XML databases
- Exploratory work for first-generation ontology
- Catalog population begins in the next 30 days



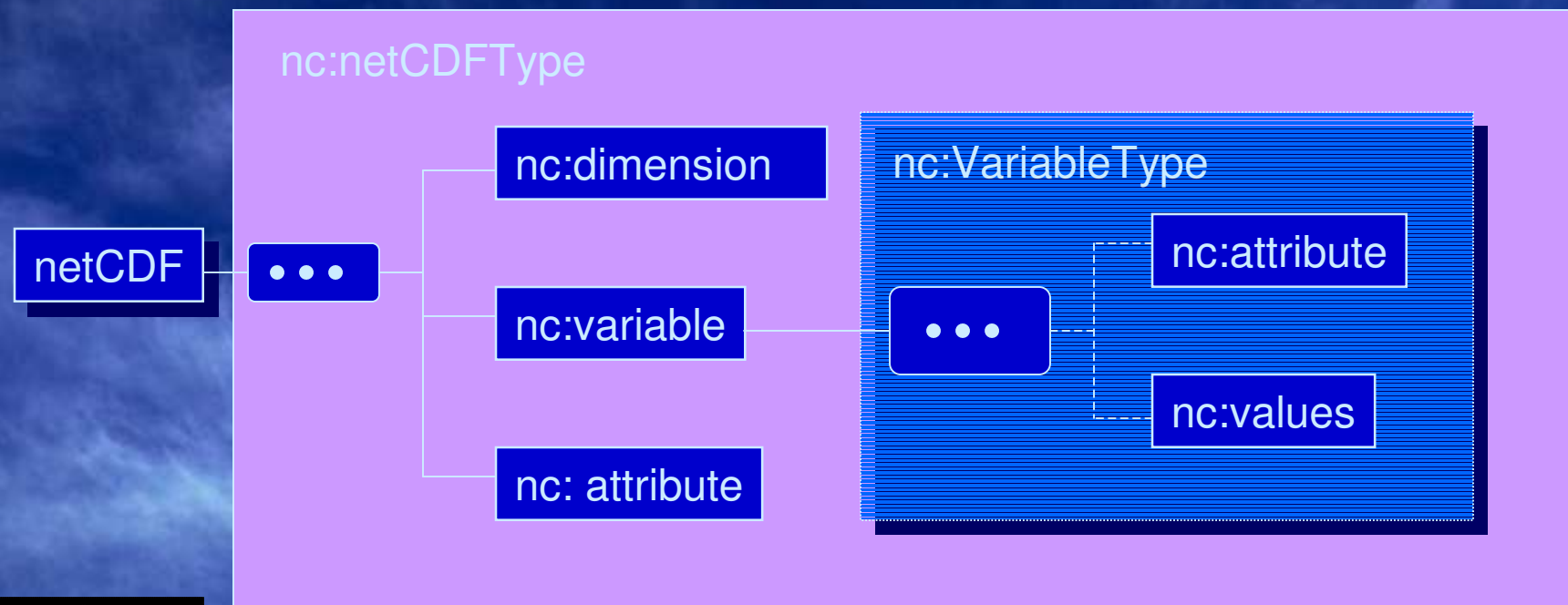
The Earth System Grid



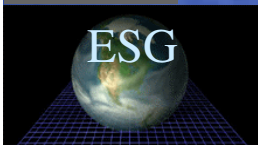


# ESG: NcML Core Schema

- For XML encoding of metadata (and data) of any generic netCDF file
- Objects: netCDF, dimension, variable, attribute
- Beta version reference implementation as Java Library  
([http://www.scd.ucar.edu/vets/luca/netcdf/extract\\_metadata.htm](http://www.scd.ucar.edu/vets/luca/netcdf/extract_metadata.htm))

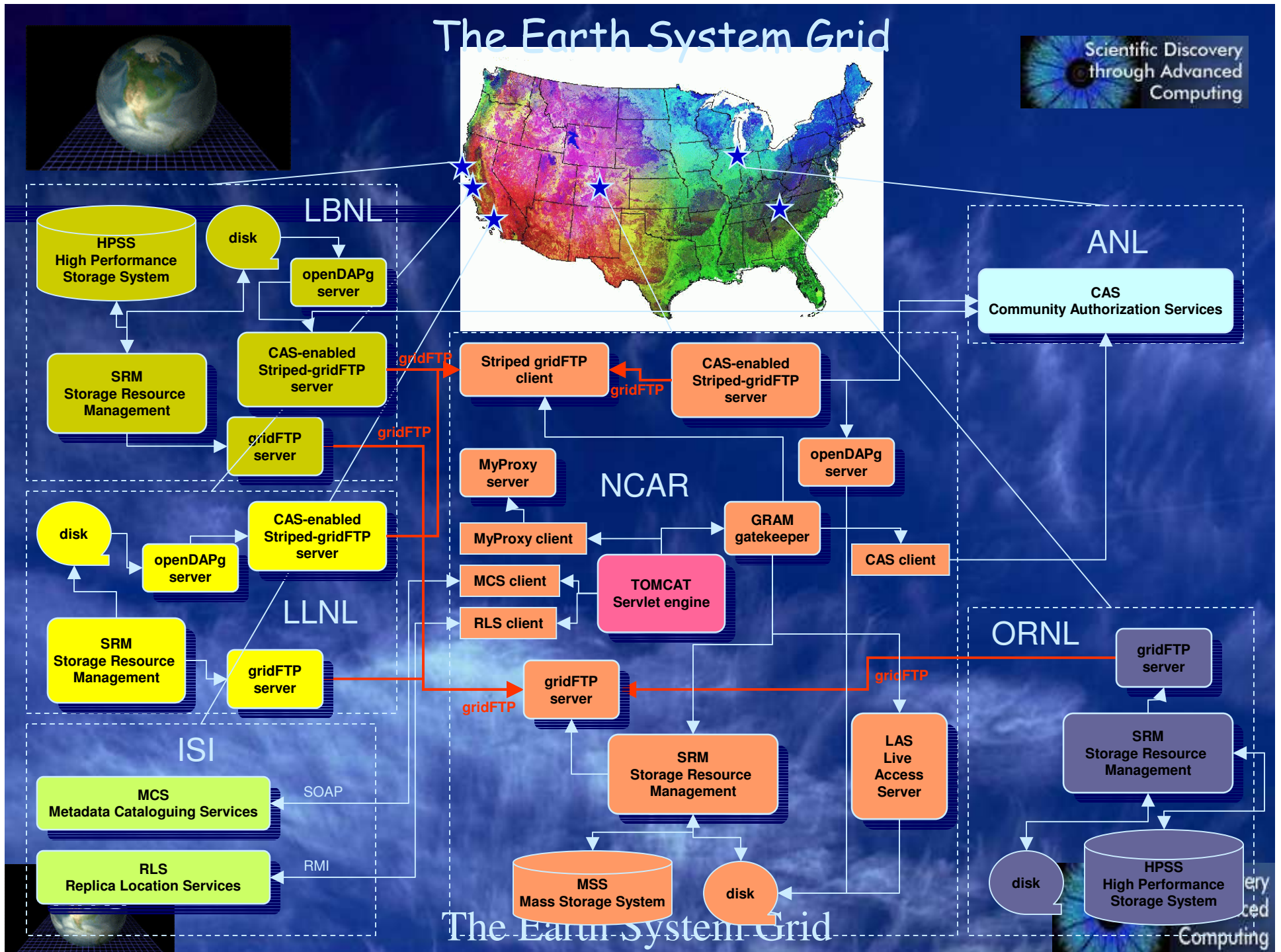


# Technology Demonstration



# The Earth System Grid

Scientific Discovery  
through Advanced  
Computing





# Collaborations & Relationships

- CCSM Data Management Group
- The Globus Project
- Other SciDAC Projects: Climate, Security & Policy for Group Collaboration, Scientific Data Management ISIC, & High-performance DataGrid Toolkit
- OPeNDAP/DODS (multi-agency)
- NSF National Science Digital Libraries Program (UCAR & Unidata THREDDS Project)
- U.K. e-Science and British Atmospheric Data Center
- NOAA NOMADS and CEOS-grid
- Earth Science Portal group (multi-agency, intl.)

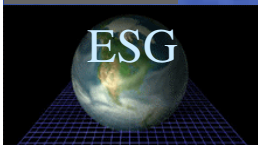
ESG

The Earth System Grid

Scientific Discovery  
through Advanced  
Computing

*<http://www.earthsystemgrid.org>*

Questions?



# END

