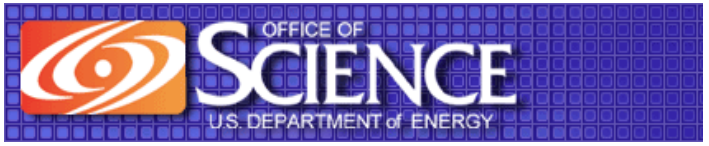


Unified Spaces for Distributed Science

News from National Collaboratories

ESnet Steering Committee Meeting

March 18, 2003

Mary Anne Scott
Dept of Energy
Office of Science

The National Collaboratories Program: Three Strongly Linked Efforts

- **Deployment and operation of a DOE Science Grid**
  - Requires coordinated effort by multiple labs
  - Manages basic software plus imports other R&D work as available
  - Provides basic “infrastructure”
    - ESnet for directory (?) and certificate services
- **Application partnerships linking computer scientists and application groups**
  - Test and validate enabling technologies for discipline-specific applications
  - Explore how to exploit Grid infrastructure to facilitate DOE applications
- **Enabling R&D - middleware**
  - Extending technology base for Grids and packaging Grid software for deployment
  - Extending technology base for collaboration
  - Developing application toolkits
  - Developing web services for science portals

Pilot Collaboratories: **Early implementations of virtual laboratories**

NC Pilots are …

- Focused on a problem of national scientific or engineering significance clearly related to the mission of DOE and have high visibility
- Involve geographically separated groups of personnel and/or facilities that are inherently required to collaborate or be used remotely for success of the project
- Implementations to test and validate that scientific research programs can integrate unique and expensive DOE research facilities and resources for remote collaboration, experimentation, simulation and analysis
  - Utilize middleware technology to enable ubiquitous access to remote resources – computation, information and expertise
  - Demonstrate capabilities that make it easier for distributed teams to work together, over the short and long term

Middleware: *software that connects or mediates between two otherwise separate programs*

NC Middleware Program is focused on…

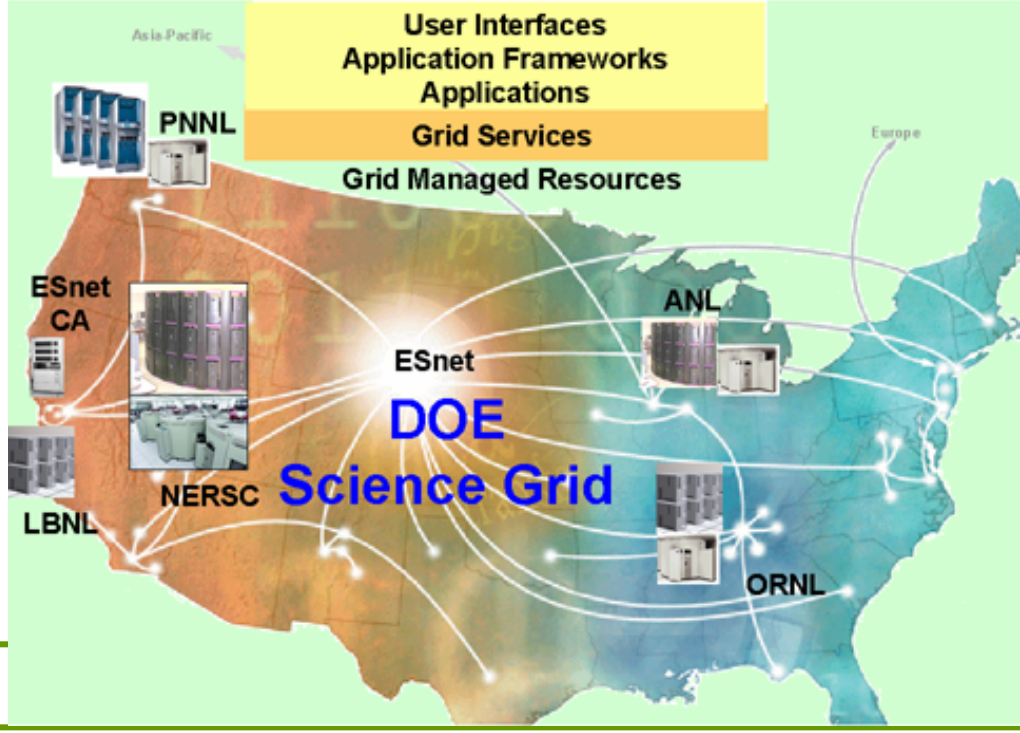
- Technology to enable ubiquitous access to remote resources – computation, information and expertise
- Capabilities that make it easier for distributed teams to work together, over the short and long term
- Standard services and protocols for access to networked resources, that aid software development/interoperability
- Middleware advances that enable scientific computing, e.g., high performance for scientific applications

Goal: It's routine; it's easy.

The DOE Science Grid

Computing and Data Infrastructure For Large-Scale Science

William Johnston, PI, LBNL Ray Bair, co-PI, PNNL Ian Foster, co-PI, ANL Al Geist, co-PI, ORNL William Kramer, co-PI, NERSC	Engineering Working Group Keith Jackson, Chair, LBNL Tony Genovese, ESnet Mike Helm, ESnet Von Welch, ANL Steve Chan, NERSC Kasidit Chanchio, ORNL Scott Studham, PNNL



Objectives

- Grid technology dissemination to DOE science projects
- Middleware for uniform, secure, and highly capable access to large and small scale computing, data, and instrument systems, all of which are distributed across organizations
- Services supporting construction of application frameworks and science portals
- Persistent cyberinfrastructure for building and operating distributed applications (e.g. security services and resource access)
- Persistent resource infrastructure, e.g. NERSC on the Grid

Accomplishments

- **Established effective inter-Lab coordination to build Science Grid and have resources committed to the Science Grid (e.g. PDSF at NERSC, Jazz cluster at ANL, Compact cluster at PNNL)**
- **Development of site firewall policy requirements for Grids**
- **Python wrapper for Globus – used for various monitoring applications**
- **Auditing and fault monitoring system**
- **Advances in Grid authorization - driven by HEP and Fusion**
- **ESnet DOEGrids CA project – a close collaboration w/ Science Grid**
  - PKI authentication infrastructure supporting large-scale DOE science collaborations (HENP, Fusion, Science Grid)
  - Will sign subordinate CA certs (e.g. for Grid integration w/ Kerberos)
  - Provides a directory service for all certificates to support advanced authorization systems
  - Integrated policy with European Data Grid – HEP collaborators in US and Europe can now share resources, NERSC will accept EDG certs

Accomplishments (cont.)

- NERSC
  - Risk mitigation plan for NERSC systems to be on the Grid
  - NERSC funded HPSS integration with Grid was tested and validated on the Science Grid
  - IBM collaboration for Globus on AIX/SP
  - A unified Grid authorization system based on LDAP server across all the production hosts
  - NERSC accounting system integrated with Grid ids
  - Working toward using Grid Security Infrastructure and certificates as a foundation for a single sign-on at NERSC

Available Software and Technology

- DOEGrids CA is available to the DOE-related science community to issue PKI identity certificates for collaborations
- pyGlobus (Python callable Globus modules)
  - NetSaint Grid status monitoring
  - Grid administration tools
- Grid helpdesk software
- Grid auditing and fault monitoring (prototype)

Collaborations and Interactions

- **Grid based Genome Analysis and Databases Update application (ANL)**
- **Authentication and authorization infrastructure for Particle Physics Data Grid (PPDG) and Fusion community**
- **Application "top-to-bottom, end-to-end" Grid performance monitoring for HEP/Atlas**
- **Cross-Grid resource sharing with iVDGL (HEP/Atlas)**
- **Grid testing platform for**
  - ECCE chemistry workbench
  - PNNL subsurface transport model
- **PDSF cluster at NERSC on Grid for Atlas and STAR HENP experiments**

Reviewer's Comments—DSG

- **What the project is doing right…**
  - Dedication to application support
    - ESG, PPDG, Genome,…
    - Vis portal
  - All CA activities, most recently leading the Tokyo accord—PMA approach is good/GridPMA.org
  - Firewall/IDS investigation
  - Positive impact on grid integration into local site security
  - Investigation into performance of grid services (starting with MDS)
  - Python tools (pyGlobus)
  - Authorization, Authentication—lowering barrier to entry (handling private key)

Reviewer Comments Relevant to ESSC

- **"A Persistent DOE Grid"**
  - Resource allocation—grid-wide process
  - Lower barriers to entry
    - Common accounts
    - Common user environment
  - Build a DOE Grid community by working with partners to
    - Establish a Grid Operations Center
    - Explore ways to harmonize site practices
    - Define rules of operation/engagement

Interactions-Collaborations-Relationships

	National Fusion Collab.	Particle Physics Data Grid	DOE Science Grid	Earth System Grid	Collab for Multiscale Chemical Science
Intra-discipline	*Yes*	*Yes*	*Yes (inter-discipline)*	*Yes*	*Yes*
Intra-SciDAC projects	*Goal 1* *Goal 2* *Goal 3*	*Goal 2* *Goal 3*	*Goal 2* *Goal 3*	*Goal 1* *Goal 2* *Goal 3*	*Goal 1* *Goal 3*
Base projects	*Middleware and services*	*Middleware and sevices*	*Middleware and services* *Additional applications*	*Middleware and services*	*Middleware and services*
Interagency		*EU* *NSF (Trillium)*	*EU*	*UK* *NSF* *NASA*	

Other National Collaboratories

- **National Fusion Grid**
  - General Atomics, MIT, Princeton, Univ. Utah, PPPL ANL, LBNL
  - Contact:: David P. Schissel (GA)
- **Particle Physics Data Grid**
  - ANL, BNL, Caltech, FNAL, JLab, LBNL, SDSC, SLAC, U. Wisconsin, UCSD, USC/ISI
  - Contact: Ruth Pordes (FNAL), Doug Olson (LBNL)
- **The Earth System Grid**
  - ANL, LBNL, LLNL, NCAR, ORNL, USC
  - Contact:: Don E. Middleton (NCAR)
- **Collaboratory for Multi-Scale Chemical Science**
  - SNL, PNNL, ANL, LLNL, LANL, NIST, MIT, UC Berkeley
  - Contact: Larry Rahn (SNL)

National Collaboratories

Middleware Projects

- **A High-Performance Data Grid Toolkit**
  - ANL, USC, U. Wisconsin
  - Contact: Bill Allcock (ANL)
- **Middleware Technology to Support Science Portals**
  - Indiana University
  - Contact: Dennis Gannon (IU)
- **Commodity Grid (CoG) Kits**
  - ANL (Gregor von Laszewski), LBNL
- **Storage Resource Management for Data Grid Applications**
  - LBNL (Arie Shoshani), FNAL
- **Pervasive Collaborative Computing Environment**
  - LBNL (Deb Agarwal), U. Wisconsin
- **Scientific Annotation Middleware**
  - PNNL (Jim Myers), ORNL
- **Middleware to Support Group to Group Collaboration**
  - ANL (Rick Stevens)

National Collaboratories

Middleware Projects, Continued

- **Reliable and Secure Group Communication**
  - LBNL (Deb Agarwal)
- **Distributed Security Architectures**
  - LBNL (Mary Thompson)
- **Grid Services**
  - ANL (Kate Keahey), LBNL (Keith Jackson)
- **Distributed Monitoring Framework**
  - LBNL (Brian Tierney)
- **A Scalable and Secure Peer to Peer Information Sharing Tool**
  - LBNL (Karlo Berket)
- **Portal Web Services**
  - UT Austin (Mary Thomas), Indiana U., SDSC, GA