U.S. Grid Projects and Involvement in EGEE

Ian Foster

Argonne National Laboratory
University of Chicago

http://www.mcs.anl.gov/~foster

EGEE-LHC Town Meeting, CERN, February 22, 2003

Overview

- U.S. Grid projects
  - Overview of current state in infrastructure, applications, and middleware
- Next steps
  - NSF "Cyberinfrastructure" report
  - U.S. MAGIC Committee
  - Planned "LHC" ITR proposal & GRIDS-2
  - Building a Grid middleware community
- U.S. involvement in EGEE
  - Integration of infrastructure
  - Collaboration on middleware

Current State of U.S. Grid Projects (1): Infrastructure + Applications

- Infrastructure deployment & operation
  - NSF TeraGrid, iVDGL, DOE Science Grid, NASA IPG, BIRN, various regional Grids
  - Good progress, but still far from critical mass as a national "cyberinfrastructure"
- Applications R&D and deployment
  - (HEP) GriPhyN, iVDGL, PPDG [next slide]
  - (other) Earth System Grid, NEESgrid, NEON, GEON, etc.
  - Substantial engagement of large application communities; much more needed

"Trillium": US Physics Grid Projects

- Particle Physics Data Grid
  - Data Grid tools for HENP expts
  - DOE funded, $9.5M
- GriPhyN
  - Data Grid research
  - NSF funded, $11.9M
- iVDGL
  - Development of global Grid lab
  - NSF funded, $14.1M

- Data intensive expts.
- Collaborations of physicists & computer scientists
- Infrastructure development & deployment
- Globus + VDT based

=

iVDgL

PPDG

GriPhyN
Data Intensive Science

foster@mcs.anl.gov

Current State of U.S. Grid Projects (2): Middleware

- Diverse mix of projects & funding sources
  - No-one wants to fund middleware!
- Unified by two long-lived projects: Globus (since 1995), Condor (since 1987)
  - Persistence, strategic direction, skilled staff
  - Foundation for essentially all Grid projects
  - Strong & close coordination between the two
- Much recent progress towards creation of professional, distributed support structures
  - With support from NSF Middleware Initiative/GRIDS Center & GriPhyN (VDT)

Impact of NSF NMI/GRIDS Center: Evolution of GT Processes

- Before 2000
  - Email-based problem tracking, aka "req"
- 2000
  - Detailed documentation, release notes (Q1)
  - Legal framework for external contributions (Q1)
- 2001
  - Packaging; module & binary releases (Q4)
  - Substantial regression tests (Q4)
- 2002
  - Bugzilla problem reporting & tracking (Q2)
  - Processes for external contrib (Q2)
  - Distributed testing infrastructure (Q3)
- 2003 (in progress)
  - Distributed support infrastructure: GT "support centers"
  - Standardized Grid testing framework(s)
  - GT "contrib" components
  - Grid Technology Repository

Overview

- U.S. Grid projects
  - Overview of current state in infrastructure, applications, and middleware
- **Next steps**
  - **NSF "Cyberinfrastructure" report**
  - **U.S. MAGIC Committee**
  - **Planned "LHC" ITR proposal & GRIDS-2**
  - **Building a Grid middleware community**
- U.S. involvement in EGEE
  - Integration of infrastructure
  - Collaboration on middleware

Report of the NSF Blue Ribbon Panel on Cyberinfrastructure

(www.communitytechnology.org/nsf_ci_report)

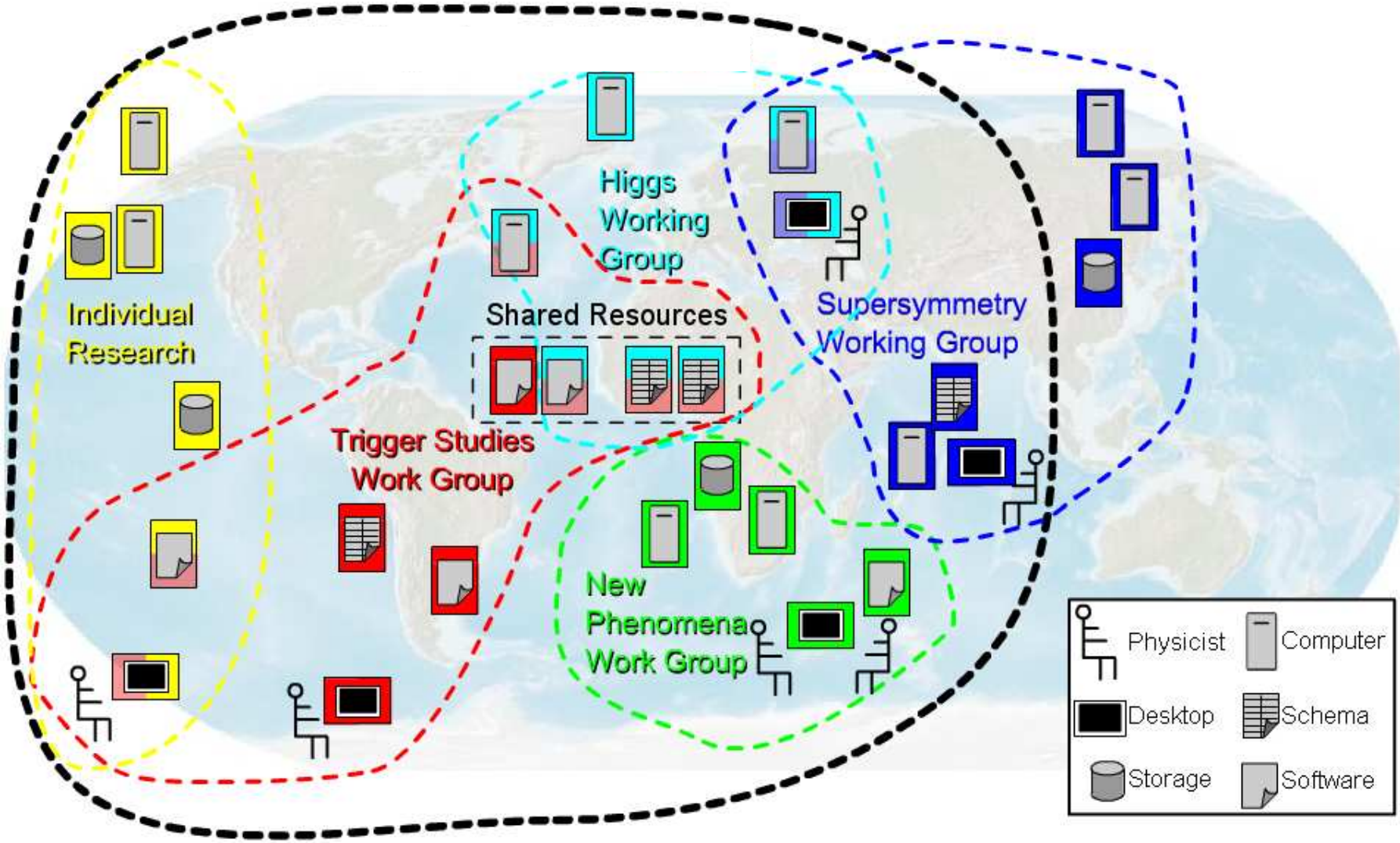
- "A new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology, and pulled by the expanding complexity, scope, and scale of today's challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive *cyberinfrastructure* on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy."
- Recommends $1B/yr new funding for

Report of the NSF Blue Ribbon Panel on Cyberinfrastructure

(www.communitytechnology.org/nsf_ci_report)

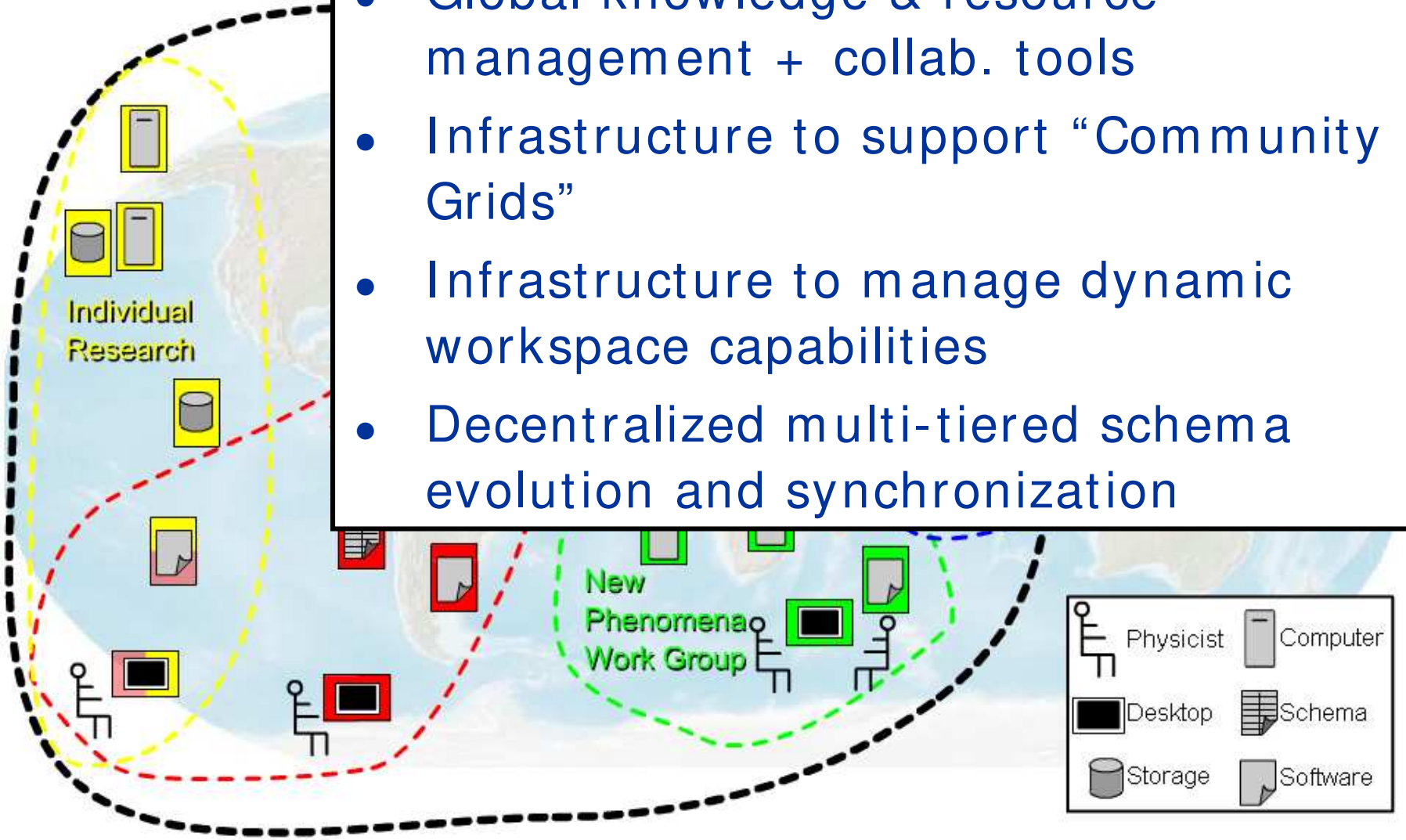
"The National Science Foundation should establish and lead a large-scale, interagency, and internationally coordinated Advanced Cyberinfrastructure Program (ACP) to create, deploy, and apply cyberinfrastructure in ways that radically empower all scientific and engineering research and allied education. We estimate that sustained new NSF funding of $1 billion per year is needed."

NSF ITR: Global Analysis Communities



NSF ITR: Global Analysis Communities

- Global knowledge & resource management + collab. tools
- Infrastructure to support "Community Grids"
- Infrastructure to manage dynamic workspace capabilities
- Decentralized multi-tiered schema evolution and synchronization



GRIDS Center 2 (Proposals to NSF Due March 7th)

- Transition to OGSA standards
- Expand range of functionality supported
- Put in place a distributed integration, testing, and support structure
- Facilitate exporting the NMI toolset to other middleware activities
- Expand the set of communities supported
- Establish international collaborations

GRIDS Center 2: An Open Grid Technology Community

- Success of Grid concept demands effective community mechanisms for coordinated
  - R&D for core technologies
  - Testing, packaging, documentation
  - Support and training of user communities
- All three must become collaborative activities
  - Based on open standards (GGF)
  - Centered on a common code base
  - Supported by appropriate tools
  - United by appropriate processes & governance

Overview

- U.S. Grid projects
  - Overview of current state in infrastructure, applications, and middleware
- Next steps
  - NSF "Cyberinfrastructure" report
  - U.S. MAGIC Committee
  - Planned "LHC" ITR proposal & GRIDS-2
  - Building a Grid middleware community
- **U.S. involvement in EGEE**
  - **Integration of infrastructure**
  - **Collaboration on middleware**

U.S. Involvement in EGEE (1): Integration of Infrastructure

- EGEE (& U.S. equivalents) to serve science communities with international scope
  - Physics, astronomy, environment, bio, …, …
- We must design for international collaboration & coordination from the start
  - Need to learn more about application needs & implications on Grid design and operation
  - But some things clear, e.g., standards; open and productive coordination; applications
- Strong application community interest in establishing these connections

U.S. Involvement in EGEE (2): Collaboration on Middleware

- With OGSA, EGEE, evolving distributed support structures, etc., stars are aligned for true international cooperation
  - Software: transform GT & Condor into international collaboration (think Linux) on a common base for Grid/distributed computing
  - Testing & support: link EGEE staff & systems into international testing framework: procedures, tools, infrastructure, bi-literal agreements
- Not easy: many opportunities for not-invented-here, start-from-scratch perspectives, and small local changes that lead to diversion!

U.S. Involvement in EGEE (3): Specific Activities

- Explicit collaborative efforts aimed at
  - Integrating U.S. and EGEE resources in support of international science
  - Creation & operation of coordinated international testing and support structure
- Direct engagement of U.S. groups in EGEE work packages
  - Joint development in some cases, e.g., monitoring/operations
  - Establishment and operation of the structures above, to ensure common evaluation, testing, and support