CyberInfrastructure @ the National Science Foundation

EGEE'09

Barcelona, Spain

September 2009

Dr. José Muñoz

Director (Acting)

Office of CyberInfrastructure (OCI)

National Science Foundation

National Science Board

Director
Deputy Director

Biological Sciences($733 M)

Computer & Information Sci & Eng ($633M)

Engineering ($764M)

Geo-Sciences ($909M)

Mathematical & Physical Sciences ($1380M)

Social, Behavioral & Economic Sciences ($257M)

Office of Cyber-infrastructure ($219M)

Office of Polar Programs

Office of Budget, Finance & Award Management

Office of International Sci & Engr

Office of Information Resource Management

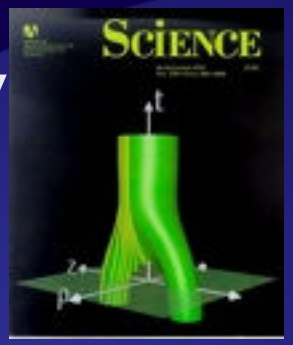
Office of Integrated Activities

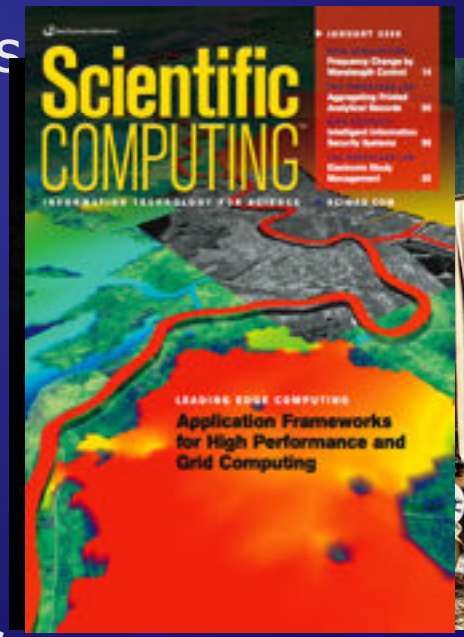
Education & Human Resources

Data-Driven Multiscale Collaborations* for Complexity

Great Challenges of 21st Century

- Multiscale Collaborations
  - General Relativity, Particles, Geosciences, Bio, Social...
  - And all combinations...
- Science and Society being transformed by CI and Data
  - Completely new methodologies
  - "The End of Science" (as we know it)
- CI plays central role
  - No community can attack challenges
  - Technical, CS, social issues to solve
- *Places requirements on computing, software, networks, etc*

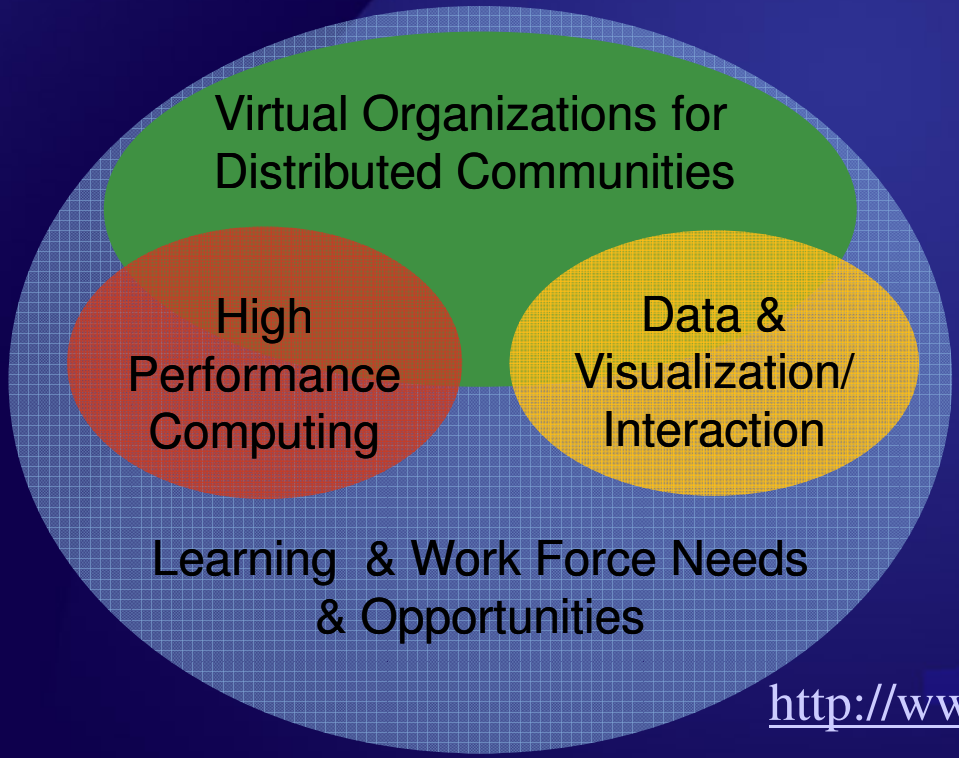


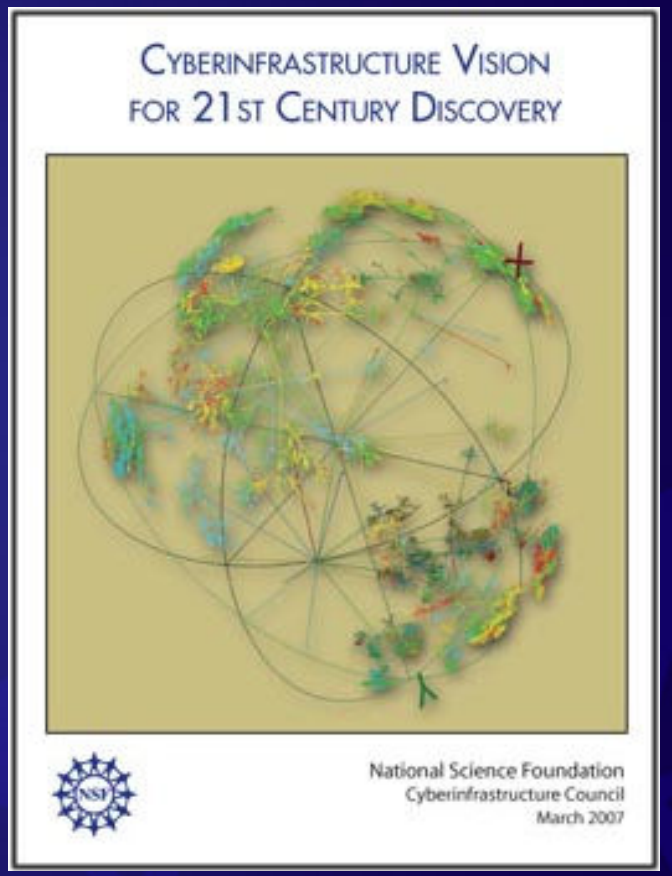


*Small groups still important!

NSF Vision for Cyberinfrastructure

- "National-level, integrated system of hardware, software, data resources & services... to enable new paradigms of science"





http://www.nsf.gov/pubs/2007/nsf0728/index.jsp

Office of Cyberinfrastructure (OCI)

- Development of collaborative computational science
  - Research and development of a comprehensive CI
  - Application of CI to solve complex problems in science, engineering, behavioral science, economics and education
- Provide stewardship for computational science at NSF, in strong collaborations with other offices, directorates, and other agencies
- Supports the preparation and training of current and future generations of researchers and educators to use Cyberinfrastructure to further research and education goals

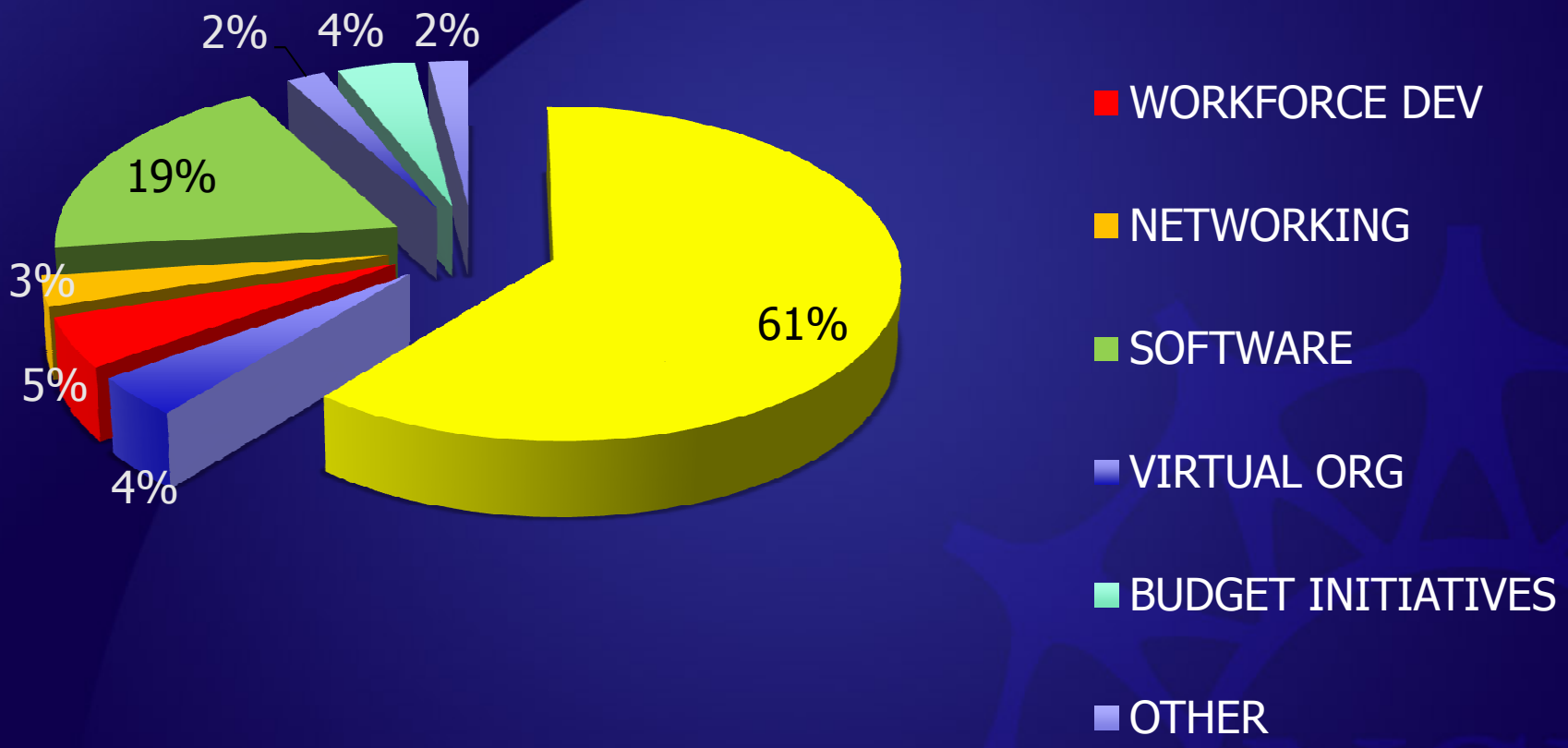
OCI Program Areas

- High Performance Computing
  - Dr. Rob Pennington
  - Dr. Barry Schneider
- Data/Visualization
  - Dr. Phil Bogden
  - Jon Stoffel
- Software
  - Dr. Manish Parashar
  - Dr. AbaniPatra
  - Dr. Jennifer Schopf
- Networking/CyberSec
  - Alan Blatecky
  - Dr. Jennifer Schopf
- Virtual Organizations
  - Dr. Susan Winter
- Learning and WF Dev
  - Dr. Manish Parashar
  - Dr. Rob Pennington
  - Dr. Susan Winter

first initialllastname@nsf.gov

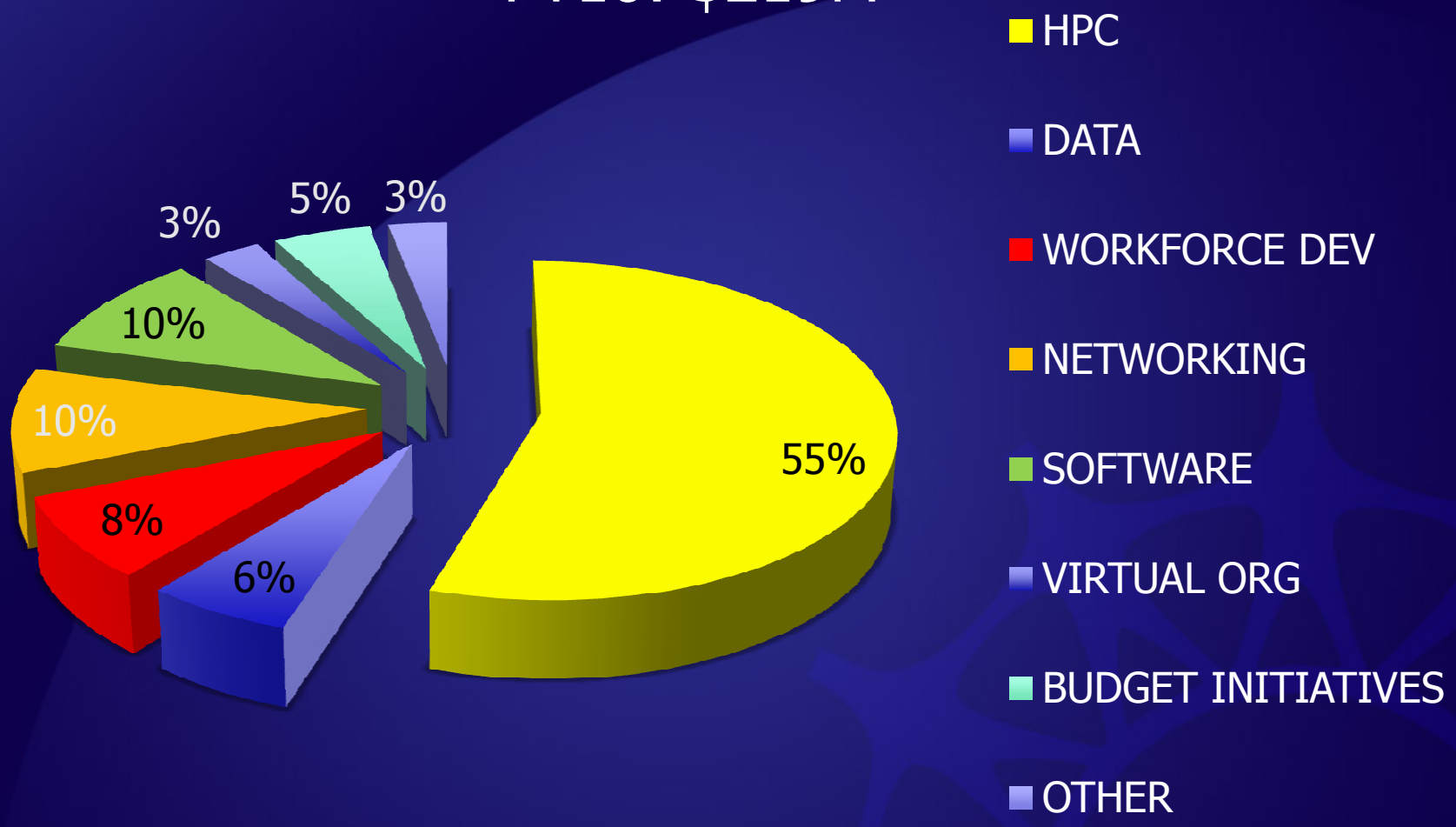
OCI FY09 BUDGET BREAKDOWN

FY09: $279M (includes ARRA)



OCI BUDGET BREAKDOWN

FY10: $219M



Shared Resource Environments

Computers

Data analysis & visualization

People

Modeling and simulation

User support

Training

Common user environments

Tools for educators

Science Gateways

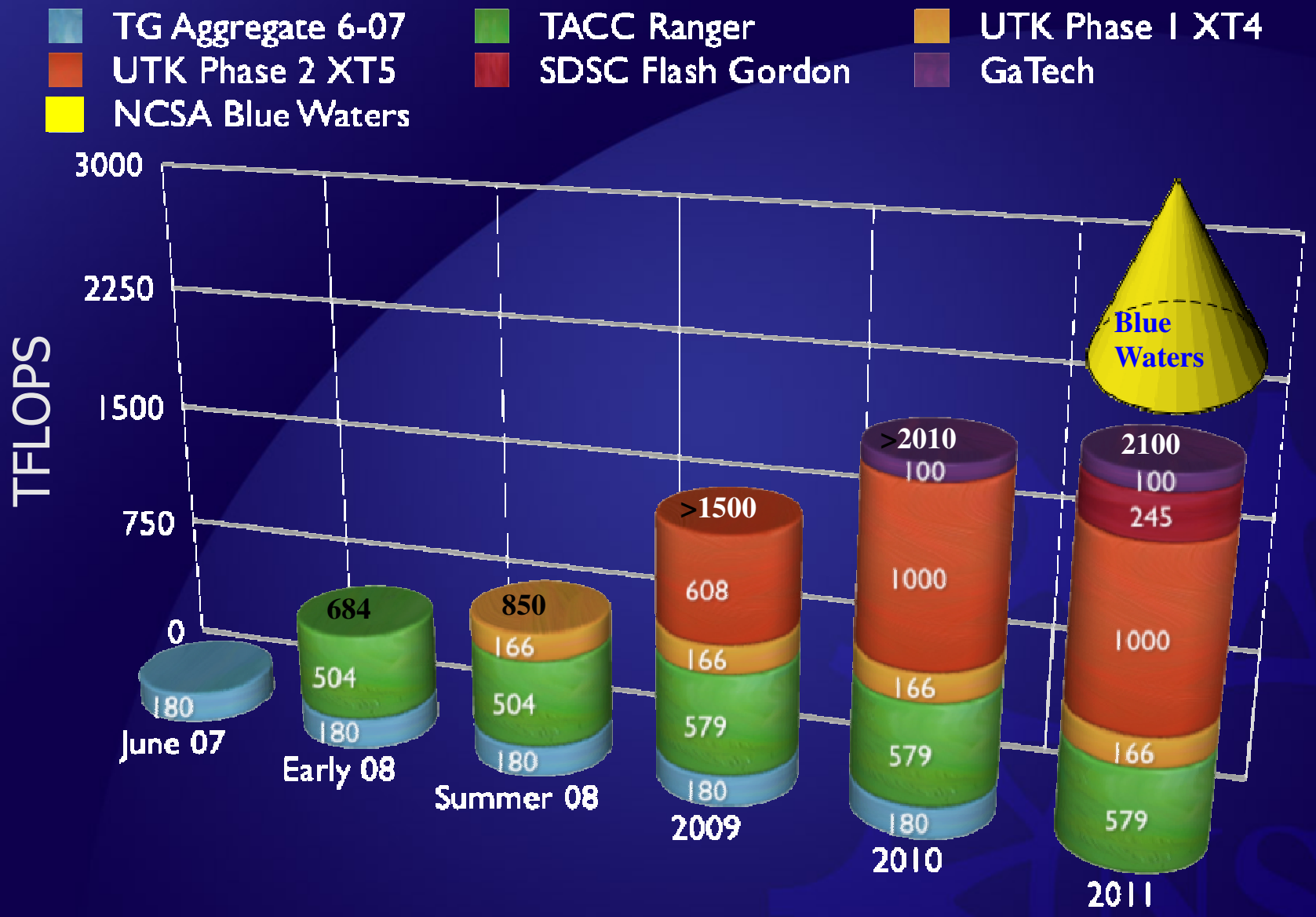
Courtesy of University of Indiana

TeraGrid

- Provides the national community with access to HPC resources and services
- Interlinked network of 14 supercomputers that provide a broad range of HPC platforms including user support
- Has enabled research in areas including:

-Subatomic Particle Physics
-Protein Folding
-Drug Design
-Nanotechnology
-Gas Turbine Design

TeraGrid Growth



Blue Waters Petascale System (2011)

- Blue Waters General Characteristics
  - Site: National Center for Supercomputing Applications (NCSA)
  - 1 petaflops *sustained* performance on *real* applications
- Blue Waters System Characteristics
  - > 200,000 cores using multicore POWER7 processors
  - > 32 gigabytes of main memory per SMP
  - > 10 petabytes of user disk storage
  - > 100 Gbps external connectivity (initial)
  - Fortran, Co-Array Fortran, C/C++, UPC, MPI/MPI2, OpenMP, Cactus, Charm++
- Available for use on "Grand Challenge" projects with users selected via a competitive process
- Blue Waters System Training and Support

Operational June 2011

Cyber-enabled Discovery and Innovation (CDI)

- Transformative Research
- Innovation in, or innovative use of "computational thinking"
- Multidisciplinary
- Themes
  - Virtual Organizations
  - From Data to Knowledge
  - Understanding Complexity
- Solicitations: 07-603, 08-604, 10-XXX

Virtual Organizations as Socio-technical Systems (VOSS)

- What constitutes effective virtual organizations? How do they enhance research and education production and innovation? Supports scientific research directed at advancing the understanding of what constitutes effective Virtual Organizations
- Multi–disciplinary
  - Anthropology, complexity sciences, CS, decision and management sciences, economics, engineering, organization theory, organizational behavior, social and industrial psychology, public administration, sociology
- Broad variety of qualitative and quantitative methods
  - Ethnographies, surveys, simulation studies, experiments, comparative case studies, network analyses.
- Grounded in theory, rooted in empirical methods

http://www.nsf.gov/pu200bs/9/nsf09540/nsf09540.htm

Virtual Organizations as Socio-technical Systems (VOSS)

- Understanding
  - How to develop effective virtual organizations
  - How virtual organizations can enhance scientific, engineering and education research production and innovation
- Grounded in theory
  - advances social, organizational and/or design science
- Employing empirical methods

- Solicitations: 08-550, 09-540 ,10-XXX

DataNet: Sustainable Digital Data Preservation and Access Network Partners Program

- Four primary goals:
  - Provide reliable digital preservation, access, integration, and analysis capabilities for science/engineering data over decades-long timeline
  - Achieve long-term preservation and access capability in an environment of rapid technology advances
  - Create systems and services that are economically and technologically sustainable
  - Empower science-driven information integration capability on the foundation of a reliable data preservation network
- Each project needed to develop a model for shared governance and the standards and protocols to enable interoperability

First Two DataNet Awards

- Data Conservancy: John Hopkins University
  - Initial focus on observational data about astronomy, turbulence, biodiversity and environmental science
  - Especially suited to terabyte-scale data sets but with strong focus on "the long tail of small science."
- DataNetOne: University of New Mexico
  - Designed to enable long-term access to and use of preserved earth observation data
    - Example: spread of diseases, the impact of human behavior on the oceans, relationships among human population density and greenhouse gas production
- [Up to] 3 new awards in 2010

International Research Network Connections (IRNC)

- Goals:
  - Provide network connections linking U.S. research with peer networks in other parts of the world
  - Stimulate the deployment and operational understanding of emerging network technology and standards in an international context
  - Support science and engineering research and education applications
- Proposals due Aug. 21, 2009
  - $40M over 5 years

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503382

Task Forces

Campus Bridging

Data &Viz

- Timelines: 12-18 months
- Led by NSF Advisory Committee on Cyberinfrastructure
- Workshop(s)
  - Recommendations
- We then go back and develop programs

Software

HPC (Clouds, Grids)

Education Workforce

Grand Challenge VOs

Office of International Science and Engineering (OISE)

http://www.nsf.gov/funding/pgm_list.jsp?org=OISE

- International Research Experience for Students (IRES)
  - supports US citizens/permanent residents as they undertake research activities internationally
    - e.g. international REU in Chemistry at the University of Cadiz
- Partnerships for International Research and Education (PIRE) NSF 09-505
  - supporting innovative, international research and education collaborations.
- NSF POC for Spain: Dr. Graham Harrison

New Opportunity

- USA/International HPC & Computational Scientist Exchange
  - USA HPC personnel and Computational Scientists spend an extended period of time at a foreign site
  - International HPC personnel and Computational Scientists spend an extended period of time at a USA site (initially TeraGrid)
  - Opportunity to exchange ideas, approaches, policies, access to each other's resources
  - Summer 2010?
  - Opportunity to extend beyond TeraGrid (e.g. iPlant)

Next Priorities

- Integrating all activities into a much more *comprehensive* cyberinfrastructure
- New programs in software
- Creating deeper partnerships with DoE and other agencies; international partnerships
- Cyber-learning
- Creating a computational science research agenda that crosses NSF and reaches out to other agencies and other countries

Gracias!!

Strategic Technologies for Cyberinfrastructure: STCI

- Support work leading to the development and/or demonstration of innovative Cyberinfrastructure services
- Fill gaps left by more targeted funding opportunities
- Projects outside the scope of targeted solicitations
- Can be used for sustainability funding
- Emphasizes the broad applicability of the work and identifies current and prospective end users

Cyberinfrastructure Reuse

- A venture fund set up by OCI to promote reuse of CI elements including software, data collections, and other computer/data/networking based entities
- Promotes use of CI elements by groups of users in different disciplines and/or the conversion of existing research tools into use for the larger community

SDCI: Software Development for Cyberinfrastructure (FY07)

- HPC, Data, and Middleware target areas defined
- 2 types of proposals
  - New development – show compelling case for new software development
  - Improvement and Support - original software must have a track record of use and impact
- Required characteristics for proposals
  - Multiple application areas and expected usage
  - Awareness/distinction among alternatives
  - Project plan with proof-of-concept and metrics
  - Open source and use of NMI Build and Test facility
  - Demonstration in first 2 years

http://www.nsf.gov/pubs/2007/nsf07503/nsf07503.htm

PetaApps

- Develop the future simulation, optimization and analysis tools that use emerging petascale computing
- Will advance frontiers of research in science and engineering with a high likelihood of enabling transformative research
- Areas examined include:
  - -Climate Change
  - -Earthquake Dynamics
  - -Storm Surge Models
  - -Supernovae simulations

http://nsf.gov/pubs/2008/nsf08592/nsf08592.pdf

Stimulus Update

- Balancing portfolio: New efforts in computational science and education
  - People: CAREER, GRF, Curriculum development
  - CI Research Agenda: Applications, Data, Software, e.g., PetaApps, STCI, Interop
- New programs from ARRA
  - ARI-R2: $200M
    - Can be used to address campus research infrastructure, e.g., network infrastructure
  - MRI-R2: $200M
    - "Supporting advances in supercomputing technology"
    - Awards up to $6M
- Stay tuned for new things, Dear Colleague Letters

NSF Overview

- Mission: "To Promote the progress of science; to advance the national health, prosperity and welfare; to secure the national defense."
  - National Science Foundation Act of 1950
- An independent agency to support basic research and education
- Peer review based on intellectual merit and broader impact
- Supports large facilities nationwide
- Discipline-based structure
- Supports all fields of Science/Engineering
- Cross-Disciplinary Mechanisms

Finding a Foundation for CS&E
The Third Pillar Needs a Place to Stand

- OCI can help create this foundation in NSF
- NSF supports all disciplines, all universities
  - OCI can be neutral catalyst across NSF for Computational Science & Applications
  - Work with CISE, MPS, all directorates
- Create new CI/CSE-research agenda
  - Software, new, computer systems - not just HPC
  - Support those who prototype, use CI
- Education and Workforce development
  - Curriculum, best practices, rewards...

Track 2 - TACC

- Fall 2006
- $59 million award
  - $30 million system
  - 4 years of operating costs
- Ranger
  - Peak performance of 579 teraflops
  - Over 60,000 processing cores
  - 125 TB memory
  - 1.7 PB
- Available since Feb, 2008

High Performance Computing- Track 2

- NSF Solicitation to deploy/support HPC machines to a wide range of researchers nationwide
- Two machines completed with two more planned
- Systems are used in various research simulation & modeling projects
- Machine operating costs/maintenance/user support $7.5M/yr-$9M/yr

Track 2 - UTK

- 2007
  - University of Tennessee at Knoxville
- $65 million, 5-year project
  - Partners at ORNL, TACC, NCAR
- Kraken
  - Cray XT5
  - 8256 compute nodes, 66,048 computational cores
  - More than 100 terabytes of memory
  - 2,300 trillion bytes of disk space
  - Peak performance of more than 607 teraflops
- Full production Feb 2009

TeraGrid Phase III – eXtreme Digital (XD)

- New infrastructure to deliver next generation high-end national digital services
- Goals:
  - Advance science and engineering
  - Providing researchers and educators with the capability to work with extremely large amounts of digitally represented information
  - Make it easy to move between local and national
- Anticipate researchers working with much larger range of digital artifacts, including digital text, digitized physical samples, real-time data streams, …