

National Science  
Foundation  
Advisory Committee  
for  
CyberInfrastructure (ACCI)

Jose L Munoz

(Acting) Director, OCI

# ACCI Represents All of NSF

Engineering

Computer and Information  
Systems Eng

Math and Physical Sciences

Education and Human  
Resource

Biology

Social, Behavioral &  
Economics

- Geosciences

- Office of Polar Program

- Office of Integrative  
Activities

- Office of Cyber-  
infrastructure

- Office of International  
Science & Eng

# Task Forces

Campus Bridging

Data and Visualization

Grand Challenges and Virtual  
Organizations

“High Performance Computing”

Learning and Workforce Development

Software

# Task Force Structure

Chair/Co-Chair by ACCI member

OCI and NSF personnel provide support

Task forces populated by members of the science and engineering communities..

**YOU!**

12-24 month time-frame

# NSF Strategic Plan for a Comprehensive National CyberInfrastructure

J. Bottum, Chair ACCI

Taskforce Leadership: D. Keyes - Software, J. T. Oden GCC,  
T. Zachariah HPC, S. Baker Data, C. Stewart CB, A.  
Ramirez LWD, V. Taylor Software, O. Ghattas GCC, T.

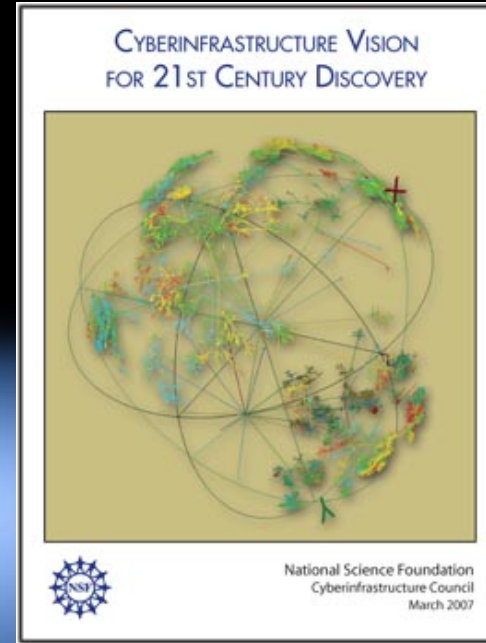
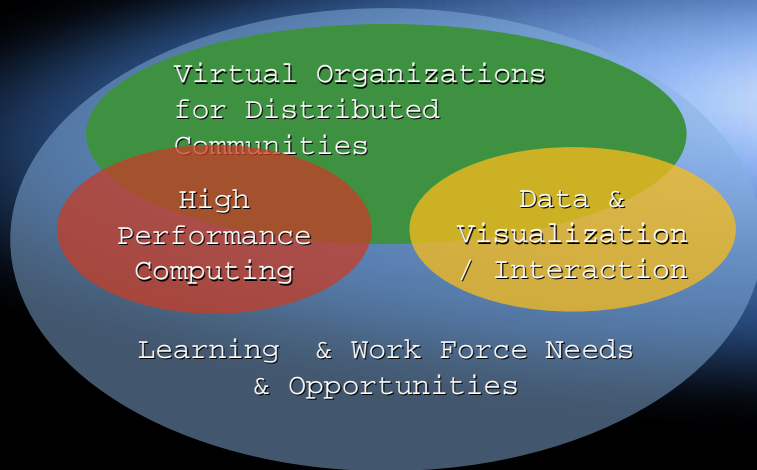
Hey Data, ...

# Agenda

- Outline of Taskforce Activity -- J. Bottum, Chair, ACCI -- 35 minutes
- Questions -- 20 minutes
- Closing Remarks and Offline Feedback Mechanisms -- 5 minutes

# 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>

# Realizing the Vision -- Office of Cyberinfrastructure (OCI)

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

# Advisory Committee on Cyberinfrastructure

- Committee of distinguished scientists and engineers from academia, government and industry with expertise in different disciplines and a computational focus
- Advise *all of* NSF on cyberinfrastructure activities
- Oversee vision and strategic direction -- operational oversight left to Committee of Visitors
- Membership and other details at <http://www.nsf.gov/od/oci/advisory.jsp>

# A Strategic Planning Exercise

**Campus  
Bridging**

**Data & Viz**

**Software**

**HPC  
(Clouds, Grids)**

**Education  
Workforce**

**Grand  
Challenge VOs**

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

# GCC & VO Task Force: Membership

*J. Tinsley Oden (chair)*

*Omar Ghattas (co-chair)*

*John King (co-chair)*

*Barry Schneider (NSF contact)*

## **HPC**

*James Kinter*

*Robert Moser*

*Abani Patra (lead)*

*Tamar Schlick*

*Klaus Schulten*

*Mary Wheeler*

## **Data & Visualization**

*Chris Johnson (co-lead)*

*Cathy Wu (lead)*

## **Software**

*Thom Dunning (lead)*

*Michael Gurnis*

*Bill McCurdy*

*Christian Ott*

*Kathy Yelick (co-lead)*

## **Education**

*Klaus Bartschat*

*Sharon Glotzer (lead)*

*J. Tinsley Oden*

*Linda Petzold (co-lead)*

*Jeroen Tromp*

## **Computational Methodologies**

*Don Estep (co-lead)*

*Omar Ghattas (lead)*

*David Keyes*

*J. Tinsley Oden*

## **VO's and GCC's**

*John Drake*

*Sara Kiesler*

*Sangtae Kim*

*John King (lead)*

*Gerhard Klimeck*

*Victoria Stodden (co-lead)*



# GCC & VO Task Force: Charge

- Develop a thorough understanding of the requirements of science and engineering applications on the cyberinfrastructure that must be developed to make significant progress toward resolving Grand Challenge (GC) Problems.
- Identify methods for enabling different communities to work together to solve complex problems. This will involve the study of Virtual Organizations and tools to support them.
- Interact with other task forces to come forth with a set of well conceived recommendations on ideas for new programs that might be developed within OCI that will more tightly couple advanced problem solving in science and engineering with continuing investments.

# GCC & VO Task Force: Timeline

07/02	GCC's and VO's Taskforce Convened (telcon)
July/Aug	Bi-Weekly Telecons
08/25	Taskforce Workshop at NSF
09/18	Taskforce Leads Telecon
09/30	Sections Submitted for Draft #1 of the TF Report
10/20	TF Report Draft #1 Forwarded to the Committee
11/04	Report of TF Primary Findings to NSF MPS Board
11/10	Input Due for Draft #2 of the TF Report
11/13	Draft # 2 of the TF Report Circulated to Committee
11/17	BoF Meeting at SC09
12/02	Presentation to NSF ACCI
Feb/Mar '10	Final Workshop with all TFs
Mar '10	Final Report to NSF

# GCC & VO Task Force: Progress

- A ~70 page 2<sup>nd</sup> draft of the task force report is circulating among task force members.
- The report is organized by sections that describe the implications and needs of next generation grand challenge science and engineering problems for hardware, software, computational methods, data and visualization, education, and virtual organizations.
- A third draft will be presented to other task forces and the broader CS&E community at a workshop in early 2010; the input from that community will inform the final report of the task force.
- While not yet in final form, the report contains a number of key recommendations, including the overarching recommendation that NSF create a permanent program in **interdisciplinary computational science & engineering**, and that OCI be responsible for its stewardship, in coordination with other programs and directorates.

# HPC Membership

- Thomas Zacharia - Chair
- Jim Kinter – Co-Chair
- Dick McCombie
- Theresa Windus
- Dave Randall
- Thomas Sterling
- Horst Simon
- Bill Harrod
- Rubin Landau
- Ron Cohen
- Tiziana Di Matteo
- George Karniadakis
- Ravi Arimilli
- Steve Scott
- Nancy Wilkins-Diehr
- Michael Macy
- Larry Davis
- Rich Loft
- Rob Pennington, NSF Liaison

# HPC Drivers

- Access to advanced computing resources needed in 2011-2015 time frame
- Applications development and support is complex, long lead time
- Computer science and engineering is presenting new opportunities
  - Innovations to advance development and use
- Integration of research and education needed for new researchers
  - Pre-college through post-graduate
- Policies and programmatic activities are needed to bring this about

# HPC Charge & Timeline

- Goals:
  - Build/sustain computing environment that stimulates advances in using HPC
  - Build/sustain savvy and capable community that fully exploits HPC resources and drives development
- Provide specific advice on the broad portfolio of HPC investments that NSF should consider
- Five to ten year span of time
- Cyberinfrastructure to support research
- Research on cyberinfrastructure
- Training
- Hardware and software

# HPC TF Progress

- Community Workshop to be held Dec 4 in Washington DC
- Invitations have been issued
- Major Topics
  - Requirements for the next generations of CI and access to advanced scientific and engineering computing/data/viz/networking resources
  - Sustainability models for CI development and support

# Software TF Membership

## ACCI

D. Keyes, KAUST  
T. Hey, Microsoft  
S. Feldman, Google  
V. Taylor, TAMU

## NSF:

A. Patra, M. Parashar, OCI  
B. Schneider, J. Schopf, OCI  
T. Russell, OIA  
P. McCartney, BIO  
C. Cooper, ENG  
P. Westmoreland, C. Cooper, GEO  
N. Sharp, MPS  
G. Solomon, F. Choudhury, SBE

## Community

P. Collela, LBNL  
J. Dongarra, UT  
T. Dunning, UIUC  
M. Ellisman, UCSD  
I. Foster, ANL  
W. Gropp, UIUC  
C. Johnson, Utah  
C. Kamath, LLNL  
C. Moler, MathWorks  
S. Parker, NVIDIA  
P. Raghavan, PennState  
A. Trefethen, Oxford

# Drivers

- Software infrastructure is an important component of Cyberinfrastructure
  - Perhaps least coordinated and systematically funded
  - Recognized in the PITAC reports as the “Grand Challenge” of Cyberinfrastructure
- Software critical for simulation and data synthesis/analysis
  - Used in science discovery, engineering design, decision support, etc.
  - It is not the computer science community that bemoans lack of investment in software, so much as the application community at large
  - Scientists expect new hardware capabilities and observational facilities to yield new scientific results and software is often the bottleneck

# Recognized stresses on SW

- Hardware responds to forces that are market-driven more than science-driven; software must continually bridge the gulf between what's "out there" (many core, memory wall, power wall, faults, etc.) and what's needed
- Community to be supported becomes exponentially larger and on average less sophisticated than the technologically oriented pioneers who invented the infrastructure we have today
- Desired functionality for software continually expands
- Reliability requirements for software continually become stricter

# Motivation for TF

- “Future development and maintenance of this critical infrastructure will be difficult if our support mechanisms continue in the current *ad hoc* mode. The confluence of these stresses, especially with respect to rapid hardware change and ripeness of scientific opportunity, suggests that we are at a most opportune time for NSF to rethink the research, development, and maintenance of the nation’s software infrastructure.” (SW TF Charge to Committee, 2009)

# Official charge

- Identify specific needs and opportunities across the spectrum of scientific software infrastructure.
- *Characterize such specific needs, and analyze technical gaps and opportunities for NSF to meet those needs through both individual and systemic approaches.*
- Design responsive approaches.
- *Develop initiatives and programs led (or co-led) by NSF to grow, develop, and sustain the software infrastructure needed to support NSF's mission of transformative research and innovation leading to scientific leadership and technological competitiveness.*
- Address issues of institutional barriers.
- *Anticipate, analyze and address both institutional and exogenous barriers to NSF's promotion of such an infrastructure.*

# Scope of SW TF (1)

- OCI's interest in software is different from that of CISE
  - CISE community does research in software (programming languages, compilers, operating systems, software engineering, etc.)
  - Scientific community at large looks at software as a good
    - To be developed initially under research contracts
    - To be maintained, ported, supported, and disseminated to users under infrastructure contracts
- Of course, software requirements of the scientific community can and do spin-off software research for CISE

# Scope of SW TF (2)

- OCI's interest in software is concentrated in scientific software
- This includes the entire scientific toolchain:
  - Low level: operating systems, compilers, debuggers, profilers, schedulers
  - Middle level: I/O management, communication protocols, libraries, frameworks, development environments, source-to-source translators
  - High level: applications, mesh generators, partitioners, object-to-object translators, visualizers, data bases, data miners, data sharing environments, collaboration environments
- Of course, scientific software can be and is leveraged for many other sectors, e.g., business, entertainment, communications, embedded systems, etc.

# Balance needed

- Standardization is efficient for investment – if we pick the “right” standards
- Good software reliably and efficiently encodes expertise in processing data and delivers it across idiomatic interfaces to users and other developers and integrators of that expertise
- Ad hoc, loose, or no coordination may lead to redundant efforts, but also gives creative, unorthodox ideas chance to demonstrate their potential and influence the entire cyber-ecosystem
- Principles for software design and metrics for software evaluation are controversial and should not be prescribed too narrowly or rigidly

# NSF's role

- Currently, NSF sponsors a lot of software creation and very little software maintenance, porting, support, training
- The mission of NSF, “*To promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense...*,” would seem to allow scope for expanding the role of NSF in software, post-creation
- This precedent has been set in NSF's support for user facilities for science
- No single agency (or country full of agencies) can pretend to lead the global cyber-ecosystem, but can aspire at best to set forth compelling principles and examples and to offer compelling incentives for compatibility of others

# Principles for investment

- Academia, industry, and national laboratories are all stakeholders in the scientific software infrastructure and NSF's efforts must fit well into the larger cyber-ecosystem context
- Scope of new programs should involve *all* software critical to NSF's mission, not only the high-end software that poses the most exciting challenges and rewards to scientists
- There is much tie-in to the other five task forces
- Must be more comprehensive than the improvement of each software tool individually
- An agency initiative can enforce end-to-end coordination

# Contemplated programs

- For software of community importance, in which the authors are still engaged, multi-year grants to authors to do code curation
- Not being constrained by NSF's standard review criteria of originality and publication in the small, but in the context of the larger community use of the software
- For software of community importance, which has been “orphaned” by the developers or their obvious successors, long-term NSF-sponsored centers to do curation for multiple codes
- Such centers would also teach “best practices” to active developers and would train users

# Progress

- Hold one-hour telecons every two weeks
- Leverage from existing reports and activities
  - International Exascale Software Project
  - NSF Workshop on Data-Centric Workflows (2009)
  - Contacts with other agencies: NSF, DOE
- Developed detailed outline for report
- Working on full draft report by Dec. ACCI meeting
- Seek more formal community feedback in Spring 2010

# Campus Bridging Membership

- Craig Stewart, Chair
- Jim Bottum, Co-chair
- Guy Almes\* (Texas A & M)
- Gary Crane (SURA)
- Andrew Grimshaw\* (University of Virginia)
- Sandra Harpole (Mississippi State)
- Dave Jent\* (Indiana University)
- Ken Klingenstein\* (Internet 2)
- Miron Livny\* (University of Wisconsin)
- Lyle Long (Penn State University)
- Clifford Lynch (CNI)
- Gerald Giraud (Oglala Lakota College)
- Brian Voss (Louisiana State University)
- John McGee\* (Renaissance Computing Institute)
- Michael R Mundrane\* (University of California, Berkeley)
- Jan Odegard (Rice University)
- Jim Pepin (Clemson University)
- Larry Smarr\* (Cal-IT2)
- Von Welch\* (National Supercomputing Applications)
- NSF: Alan Blatecky, Jennifer Schopf
- Operational support: Dale Lantrip (Indiana University) [dlantrip@iupui.edu](mailto:dlantrip@iupui.edu)

\* Indicates member of executive committee

# Drivers

- The nation's existing cyberinfrastructure –broadly considered - is not adequate to meet the needs of the current U.S. science and engineering community.
- We are not using the existing cyberinfrastructure effectively or efficiently enough. This is at least in part due to the barriers of migration between campus and national CI
- The proliferation of distributed devices capable of producing large amounts of data is making the data deluge problematic in new ways (e.g. 454 sequencers)

# Charge & Timeline

- This task force is meant to address the broad issues involving improving campus interactions with cyberInfrastructure, broadly construed. It will include a number of different types of bridging:
- Campus grids to national infrastructure (both compute and data oriented approaches) and international CI
- Campus networks to state, regional and national
- Departmental cluster to campus HPC infrastructure
- Campus-to-campus and campus-to-state/regional approaches

**Plan to produce reports on particular topics rather than one grand document at end. Themes:**

- **Networking (Dave Jent, lead)**
- **Identity Management (Ken Klingenstein & Von Welch co-leads)**
- **Data-centric Campus Bridging Issues (Guy Almes, lead)**
- **Middleware (Andrew Grimshaw and John McGee, co-leads)**
- **Campus Leadership Engagement (Michael Mundrane, lead)**

# Progress

- Monthly telcons, F2F meeting in October
- Proposal for workshop funded by NSF; to be held in Feb. Position papers to be solicited from community as part of workshop preparation
- Making progress in several areas; first reports 1H 2010 likely

# Learning and Workforce Development Membership

- Alex Ramirez, Hispanic Association of Colleges and Universities - Chair
- Michael Stubblefield, Southern University
- Nora Sabelli, SRI International & Learning in Informal & Formal Environments (LIFE) Science of Learning Center
- Henry Neeman, University of Oklahoma
- Geoffrey FOX, Indiana University
- Susan Winter, NSF, OCI
- Rob Pennington, NSF, OCI
- Full Task Force Membership still to be completed

# Drivers

- Need to educate and train the current and next generation of scientists and engineers in computational science, computational thinking and use of cyberinfrastructure and the social dynamics of virtual organizations and remote collaborative research and education
- CI provides interesting learning resources to be incorporated into the curriculum enabling more active and authentic pedagogy
- Deluge of science data to incorporate into education at all levels, and a possible deluge of learning data to incorporate into teaching, administration and learning science
- Need for informal education, professional development and public awareness
- Emerging computational science areas
- Emerging cyber-learning science, pedagogy and curriculum resources

# Charge & Timeline

- Foster the broad deployment and utilization of CI-enabled learning and research environments
- Support the development of new skills and professions needed for full realization of CI-enabled opportunities
- Promote broad participation of underserved groups, communities and institutions, both as creators and users of CI;
- Stimulate new developments and continual improvements of CI-enabled learning and research environments;
- Facilitate CI-enabled lifelong learning opportunities ranging from enhancement of public understanding of science to meeting the needs of the workforce seeking continuing professional development;
- Support programs that encourage faculty who exemplify the role of teacher-scholars through outstanding research, excellent education the integration of education and research in computational science and computational science curriculum development;
- Support the development of programs that connect K-12 students and educations with the types of computational thinking and computational tools that are being facilitated by cyberinfrastructure

# Progress

Dates (tentative)	Activity
	Difficulty in getting started. Community input solicited – Broad spectrum of areas many with their own communities with varying perspectives
December 2009	Task Force Membership Finalized
January 2010	Plans for workshops completed
March 2010	First workshop
May 2010	Second workshop
July 2010	Third workshop
September 2010	Workshop reports completed
October 2010	Reports disseminated broadly for community review and comment
November 2010	Draft of Final Report completed and disseminated
December 2010	Report completed

# Data TF Membership

- Shenda Baker (Chair) Professor of Chemistry, Global Clinic Director, Harvey Mudd College
- Elizabeth Lyon, Director, UKOLN, University of Bath
- Tom Dietterich, Prof. and Associate Dean of Intelligent Systems Res.,  Oregon State University
- Stuart Feldman, Vice President, Engineering, Google, Inc.
- Brewster Kahle, Founder, Digital Librarian, Internet Archive
- John King, Vice Provost for Academic Information, Univ. of Michigan
- Advisory/Adjunct Bodies: NSF Data Working Group, Cyberinfrastructure Coordinating Committee (CICC),
- Philip Bogden, Jon Stoffel, NSF Office of Cyberinfrastructure

# Data/Viz TF Drivers

- Charge: Examine the increasing importance of data and their visualization in driving grand challenge science, engineering and education
- Emphasis: Value of the data, its complexity, and organic aspects. The role of data in research, the value of metadata and ontologies for integration, etc.
- Goal: Catalyze a network of science and engineering data collections that is open, extensible and sustainable; Enable multiple fields of science and engineering research and education - including new types of data-driven computational science, interdisciplinary research and cross-disciplinary education

# Data/Viz TF Progress

**Phase I: 1 May 2009 => 1 May 2010**

- **Build from Grand Challenge TF findings & National Academy workshops – query the community**
- **Virtual workshop in Fall 2009 with follow-on workshops in spring/summer 2010**

**Phase II: June 2010 => December 2010**

- **Document findings, clarify recommendations and inform necessary changes in policy**

# DISCUSSION

- Community input requested
- Please email short “white papers” to [accl-task-forces@nsf.gov](mailto:accl-task-forces@nsf.gov)