XSEDE: Reflections on Establishing a Foundation for a Cyberinfrastructure Ecosystem

John Towns
PI and Project Director, XSEDE
Director, Collaborative Cyberinfrastructure Programs, NCSA
jtowns@ncsa.illinois.edu

Extreme Science and Engineering Discovery Environment

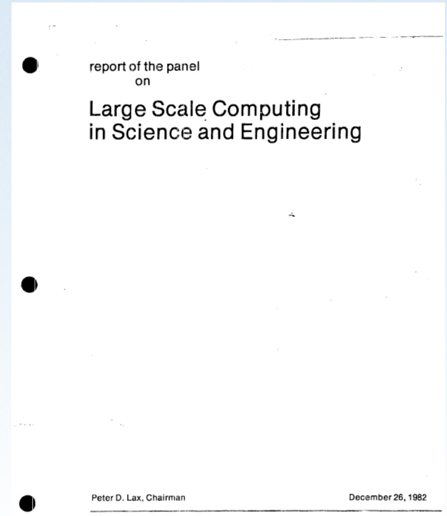
In the beginning…

…there was nothing.

But let’s not start there….

In the 1980s…

- Monolithic supercomputers were needed for open science research
  - the Lax Report (1982)


report of the panel
on
Large Scale Computing
in Science and Engineering

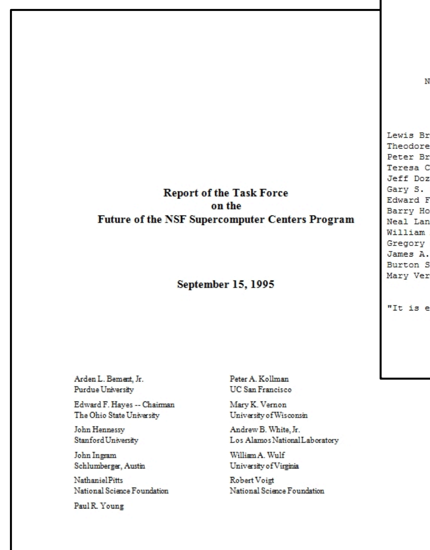
Peter D. Lax, Chairman
December 26, 1982

… and thus was born the NSF Supercomputer Centers Program

- Focus was on large scale simulation-based science and engineering
  - astrophysics, CFD, chemistry, high energy physics…
- … and a nod to the need to move data
  - NSFnet created

In the 1990s…

- Multiple supercomputers were needed to support leading edge science
  - NSF centers start to work together at some level
  - joint allocations (limited) was the extent of most of that
  - coining of the “metacenter” term
- Some visionary work on resource and data sharing
  - iGrid
  - vBNS networking project
- But we needed something more
  - Branscomb report (1993)
  - Hayes report (1995)
  - APEX meetings (1995-1996)



Report of the Task Force
on the
Future of the NSF Supercomputer Centers Program

September 15, 1995

Arden L. Bement, Jr.
Purdue University

Edward F. Hayes -- Chairman
The Ohio State University

John Hennessy
Stanford University

John Ingram
Schlumberger, Austin

Nathaniel Pitts
National Science Foundation

Paul R. Young

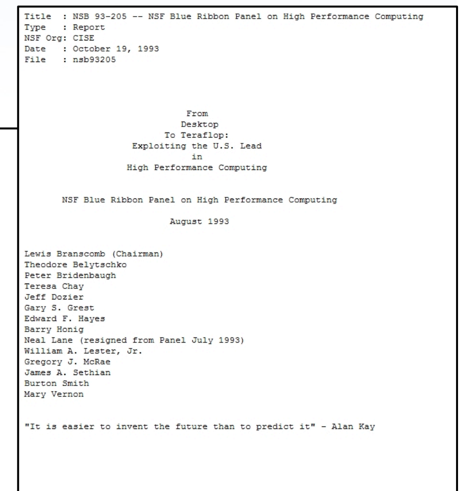
Peter A. Kollman
UC San Francisco

Mary K. Vernon
University of Wisconsin

Andrew B. White, Jr.
Los Alamos National Laboratory

William A. Wulf
University of Virginia

Robert Voigt
National Science Foundation



Title : NSB 93-205 -- NSF Blue Ribbon Panel on High Performance Computing
Type : Report
NSF Org: CISE
Date : October 19, 1993
File : nsb93205

From
Desktop
To Teraflop:
Exploiting the U.S. Lead
in
High Performance Computing

NSF Blue Ribbon Panel on High Performance Computing

August 1993

Lewis Branscomb (Chairman)
Theodore Belytschko
Peter Bridenbaugh
Teresa Chay
Jeff Dozier
Gary S. Grest
Edward F. Hayes
Barry Honig
Neal Lane (resigned from Panel July 1993)
William A. Lester, Jr.
Gregory J. McRae
James A. Sethian
Burton Smith
Mary Vernon

"It is easier to invent the future than to predict it" - Alan Kay

Mid to Late 1990s…

- NSF initiates PACI Program
  (Partnerships for Advanced Computational Infrastructure)
  - large partnerships each sporting a range of computational resources
  - resource allocations process unified
- Two awards made
  - the Alliance: National Computational Science Alliance
  - NPACI: National Partnership for Advanced Computational Infrastructure

ALLIANCE
Led by NCSA/UIUC

NPACI

NSF

XSEDE

In the 2000s…

- Discovered PACI was not working
  – NSF shuts down the program early
- We figured out what we thought we needed
  – not just HPC resources and networks
  – an integrated set of resources and support organization
- Distributed Terascale Facility program initiated
  – TeraGrid is born
    - *"supercomputing, storage, visualization systems, data collections, and science gateways, integrated by software services and high bandwidth networks"*

  – fundamentally an HPC-centric, computer science and technology play
    - delivered many Good Things!!

Tremendous Impact – looking just at last year:

- More than 1,200 projects supported
  - 54 examples highlighted in most recent TG Annual Report
    - atmospheric sciences, biochemistry and molecular structure/function, biology, biophysics, chemistry, computational epidemiology, environmental biology, earth sciences, materials research, advanced scientific computing, astronomical sciences, computational mathematics, computer and computation research, global atmospheric research, molecular and cellular biosciences, nanoelectronics, neurosciences and pathology, oceanography, physical chemistry
- 2010 TeraGrid Science and Engineering Highlights
  - 16 focused stories
  - http://tinyurl.com/TeraGridSciHi2010-pdf
- 2010 EOT Highlights
  - 17 focused stories
  - http://tinyurl.com/TeraGridEOT2010-pdf

In the mid-2000s…

- "cyberinfrastructure" term becomes popular
  - even though most had (and still have) differing definitions
- TeraGrid was a Good Thing, but not satisfying the needs of enough researchers
- Significant community soul searching on cyberinfrastructure needs
  - many reports sponsored by NSF and other agencies


The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences (for public comment)
Computation As a Tool for Discovery in Physics
Establishing a Petascale Collaboratory for the Geosciences
Building a
NSF'S CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY
NSF Cyberinfrastructure Council
National Science Foundation
January 20, 2006
Version 5.0
Trends in
Geoscience Education
National Science Board
Visualization
Management
Sensors
November 7, 2004
Second Printing—MARCH 2004
September 2005

XSEDE

By the end of the 2000s…

- Discovered the TeraGrid was too narrow in scope
  - focus was on HPC and technology
  - community was primarily high-end users of HPC
    - resource sharing and data were secondary
  - struggled to reach to broader community for researchers
    - It didn't really offer what they needed!
  - was difficult to defend across the NSF since it did not support a broad enough range of research
- NSF recognizes need for a different plan
  - supports development of ideas for what comes next
  - TeraGrid Futures committee produces recommendation report

Home People Events

TeraGrid Future

the online home for the teragrid planning process

TeraGrid Planning Process

Background

The National Science Foundation (NSF) is providing support for a community-driven, participatory planning process whose goal is to provide information that will help guide the future evolution of TeraGrid. TeraGrid is a collaboration among nine partner sites that uses high-performance network connections to integrate high-performance computers, data resources and tools, and experimental facilities around the country. As of early 2007, the integrated resources include more than 102 teraflops of computing capability and more than 15 petabytes (quadrillions of bytes) of online and archival data storage. Through the TeraGrid, researchers can access over 100 discipline-specific databases.

Current awards for the operation, user support, and enhancement of the TeraGrid facility will expire in 2010. By this date, a petascale computing resource will be on the horizon, the user community will have grown and diversified, and new policies and services are likely to be needed to meet the needs of users and the expanding pool of high-performance computing resources. In anticipation of these changes, the planning process is focusing on the needs of current and emerging user communities as a critical aspect in the development of a path forward for TeraGrid in 2010 and beyond.

Planning Process Leadership

The process will be led by a steering committee drawn from key stakeholder communities; this group is currently being formed. The Chair of the committee is Dr. Timothy Killeen, Director of the National Center for Atmospheric Research (NCAR), and the Associate Chair is Dr. Roberta Balstad, Senior Fellow and former Director of Columbia University's Center for International Earth Science Information Network (CIESIN). The University of Michigan's School of Information is facilitating the planning process and providing support to the steering committee.

XD Solicitation/XD Program

- eXtreme Digital Resources for Science and Engineering (NSF 08-571) -- Extremely Complicated
  - High-Performance Computing and Storage Services
    - aka Track 2 awardees
  - High-Performance Remote Visualization and Data Analysis Services
    - 2 awards; 5 years; $3M/year
    - proposals due November 4, 2008
  - Integrating Services (5 years, $26M/year)
    - Coordination and Management Service (CMS)
      - 5 years; $12M/year
    - Technology Audit and Insertion Service (TAIS)
      - 5 years; $3M/year
    - Advanced User Support Service (AUSS)
      - 5 years; $8M/year
    - Training, Education and Outreach Service (TEOS)
      - 5 years, $3M/year
  - two phase proposal process for IS
    - pre-proposals November 4, 2008
    - final proposals due ~~June 15, 2009~~ July 16, 2010

XSEDE

Setting the Context of the eXtreme Digital (XD) Program

- XD represents the third-generation of the TeraGrid program
  - 2002-2005: Distributed/Extended Terascale Facility (DTF/ETF)
  - 2005-2011: Grid Infrastructure Group (GIG) + Resource Providers
  - 2010-2016: eXtreme Digital (XD) + Service Providers
- XD represents the integration services across resource providers, not the resources themselves
  - with exception of the two Remote Visualization and Data Analysis systems, hardware resources are not included in the XD solicitation
- Resource Provider awards are separate
  - some resources are being retired at the end of TeraGrid
  - many resources will bridge from TeraGrid to XD era

TeraGrid: a project in transition

- Currently in a one year (plus three months) extension of project
  - start of XD for CMS/AUSS/TEOS deferred for one year (July 2011)
    - TeraGrid Extension funded to bridge to XD program
      - 12-month funding to support most GIG functions and some non-Track 2 RP resources
    - still some uncertainty in sequence/timing of events
      - additional 3 months of support being provided
- All activity areas have effort reserved for TeraGrid → XD transition as appropriate
  - planned period for transition: April-July 2011
  - transition issues exist for nearly all areas
- Continued support of users during transition is our highest priority

So where did that leave us?

- We continued to refine what we believed the research community needs
  - XD appropriately called for a much broader engagement with the research community
- Interpret broadly that which needs to be integrated
  - act as integrator and not operator…

Boundary Conditions and Principles

- XSEDE inherited TeraGrid environment
- XSEDE inherited TG community and their expectations

- Point of view has changed
  - not an HPC/CS/tech play
  - about productivity and creating the environment necessary to be productive
- Finally figured out that the project must define a solution that is designed to evolve!
- Identify the greatest needs and start there
  - Don't forget what you have learned – both good and bad!

Convenience requirements will always increase.

Each generation of users requires more convenience than the former: thus we must always be adding new layers of software while maintaining and extending existing reliability and capability.

Change is the only Constant – Heraclitis 535BC-475BC

Source: Phil Andrews, UTK

XSEDE Vision

The eXtreme Science and Engineering Discovery Environment (XSEDE):

enhances the productivity of scientists and engineers by providing them with new and innovative capabilities

and thus

facilitates scientific discovery while enabling transformational science/engineering and innovative educational programs

Science requires diverse digital capabilities

- XSEDE is a comprehensive, expertly managed and evolving set of advanced heterogeneous high-end digital services, integrated into a general-purpose infrastructure.
- XSEDE is about increased user productivity
  - increased productivity leads to more science
  - increased productivity is sometimes the difference between a feasible project and an impractical one

XSEDE supports a breadth of research

From direct contact with user community as part of requirements collections

- Earthquake Science and Civil Engineering
- Molecular Dynamics
- Nanotechnology
- Plant Science
- Storm modeling
- Epidemiology
- Particle Physics
- Economic analysis of phone network patterns
- Brain science
- Analysis of large cosmological simulations
- DNA sequencing
- Computational Molecular Sciences
- Neutron Science
- International Collaboration in Cosmology and Plasma Physics

Just a <u>sampling</u> of much larger set. Many examples are new to XSEDE. Range from petascale to disjoint HTC, many are data driven. XSEDE will support thousands of projects.

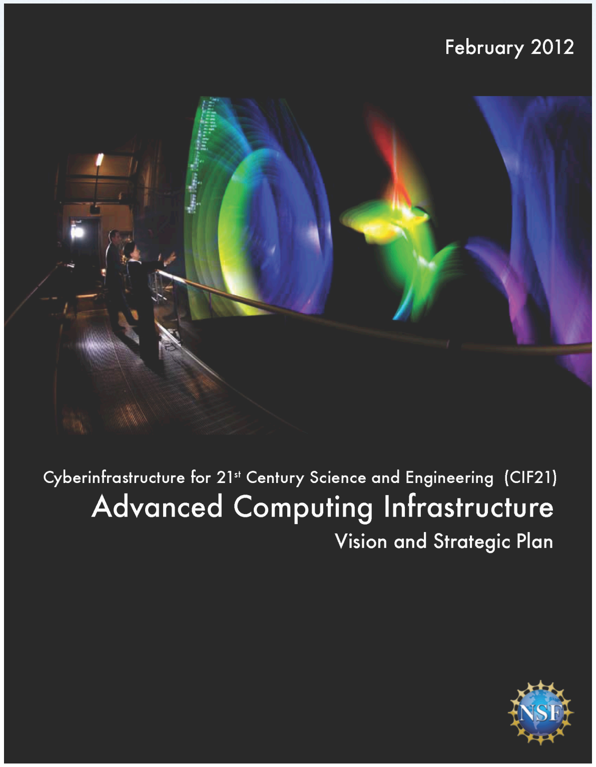
What do you mean by "Advanced Digital Services?"

- Often use the terms "resources" and "services"
  - these should be interpreted very broadly
  - most are likely ***not*** operated by XSEDE
- Examples of resources
  - compute engines: HPC, HTC (high throughput computing), campus, departmental, research group, project, …
  - data: simulation output, input files, instrument data, repositories, public databases, private databases, …
  - instruments: telescopes, beam lines, sensor nets, shake tables, microscopes, …
  - infrastructure: local networks, wide-area networks, …
- Examples of services
  - collaboration: wikis, forums, telepresence, …
  - data: data transport, data management, sharing, curation, provenance, …
  - access/used: authentication, authorization, accounting, …
  - coordination: meta-queuing, …
  - support: helpdesk, consulting, ECSS, training, …
  - And many more: education, outreach, community building, …

XSEDE

XSEDE's Distinguishing Characteristics

- Foundation for a national cyberinfrastructure (CI) ecosystem
  - comprehensive suite of advanced digital services will federate with other high-end facilities and campus-based resources



- *NSF's Advanced Computing Infrastructure: Visions and Strategic Plan* statement
  - http://www.nsf.gov/pubs/2012/nsf12051/nsf12051.pdf

A little about the project and its organization and governance…

- High order bits
  - 5 year, $121M project
    - option for additional 5 years of funding upon major review after PY3
  - no funding for major hardware
    - coordination, support and creating a national/ international cyberinfrastructure
  - ~120 FTE funded across 17 partner institutions

XSEDE

XSEDE's Distinguishing Characteristics - Governance

- World-class leadership from CI centers with deep experience: partnership led by NCSA, NICS, PSC, TACC and SDSC
  - PI: John Towns, NCSA/Univ of Illinois
  - Co-PIs: Jay Boisseau, TACC/Univ of Texas Austin
    Robert Harrison (interim), NICS/Univ of Tenn-Knoxville
    Ralph Roskies, PSC/CMU
    Nancy Wilkins-Diehr, SDSC/UC-San Diego
- Partners who strongly complement these CI centers with expertise in science, engineering, technology and education

Univ of Virginia	Ohio Supercomputer Center
SURA	Cornell
Indiana Univ	Purdue
Univ of Chicago	Rice
Berkeley	NCAR
Shodor	Jülich Supercomputing Centre

Infrastructure Designed for Innovation & Evolution

- An environment in which all resources, data and services relevant to a researcher can be embedded and shared
  - campus bridging creating a single virtual system with interactive data transfer and resource sharing capabilities
    - "make my data accessible everywhere I want to be"
  - coordinated archival approach to ensure persistence of important datasets beyond the lifetime of particular service providers
- An underlying infrastructure to support this
  - open architecture with judicious use of standards designed to evolve in a non-disruptive way
  - interoperability of XSEDE with other CIs

XSEDE's Distinguishing Characteristics - Architecture

- There *is* an architecture (being) defined
  - based on set of design principles
  - rooted in the judicious use of standards and best practices
  - clearly defined transition plan from TeraGrid to XSEDE
- Professional systems engineering approach
  - responds to evolving needs of existing, emerging, and new communities
    - incremental development/deployment model
  - new requirements gathering processes
    - ticket mining, focus groups, usability panels, shoulder surfing
  - ensure robustness and security while incorporating new and improved technologies and services
  - process control, quality assurance, baseline management, stakeholder involvement

System and Software Engineering (SSE)

"You‘ve got to start with the customer experience and work back toward the technology - not the other way around."

Steve Jobs, May 1997

- XSEDE is driven by user needs and requirements
  - baseline requirements set was a mergers of those identified by the XSEDE and XROADS teams
    - best estimate of the early capabilities required
- SSE is responsible for ongoing requirement collection, analysis and processing
  - DOORS XSEDE Requirements database (XRDB).
  - requirements used to determine the functionality needed
- SSE organizes and supports the User Requirements, Evaluation and Prioritization Working Group (the “UREP”)
  - inclusive membership: reps from major project areas, appropriate subject experts, UAC members
  - reviews, analyzes and prioritizes requirements
- Prioritized requirements drive activities of the project

XSEDE

XSEDE Software Development and Integration (SD&I)

- Delivers high quality software to
  - Operations, Service Providers, campuses, developers and researchers
- Applies technical expertise in
  - data and execution software and services
  - security software and services
  - discovery software and services
  - engineering process and support services
  - testing process and support services
- Specifically contributes to
  - production software support
  - user requirements management
  - XSEDE architecture definition and design

XSEDE SD&I recent activities …

- Delivering 3 new capabilities to Operations:
  - Production Globus Data Transfer Service/GO-Data
  - Beta Execution Management Service/EMS (Genesis/Unicore)
  - Beta Global Federated File System/GFFS (Genesis)
- Planning and launching ~14 PY1 Q4 projects
- Planning ~70 candidate PY2 projects

XSEDE

Working with other CI Providers

- Realizing a national/international CI ecosystem requires confederation of CI providers
  - XSEDE is working closely with OSG and PRACE today
    - initial work with others under way
  - still much to address though!
- OSG is a significant CI in the US
  - ties to CI (eScience infrastructure) providers internationally
  - is a Level 1 Service Provider in XSEDE
- PRACE is a significant CI in Europe
  - subsumed DEISA last year
  - working toward joint allocations call later this calendar year
- Partnerships allow expanding the scope of XSEDE
  - OSG has a focus on HTC opportunistic resources complement traditional HPC resources inherited from TeraGrid
  - PRACE represents both large scale HPC and distributed resources

Integration Requires Solid Resources and Services to Integrate!

- XSEDE takes an opportunistic view
  - architecture designed in an implementation agnostic manner
  - SD&I team looks to identify existing implementations that satisfy design (or can be readily adapted)
    - original XD solicitation discourages (forbids?) software development
    - encourages adoption of other development products (particularly NSF-funded efforts)
- Leverage efforts of related XSEDE TIS (Technology Insertion Service) award
  - separately funded award from NSF
  - focus on technology tracking and evaluation for use in XSEDE
  - https://tis.xsede.org/

What does this mean to me though?

- Architectural design drives processes in XSEDE to produce useful capabilities for XSEDE-supported researchers

- Some examples currently in process:
  - Globus Online
    - reliable, high-performance file transfer …as a service
  - Global Federated File System (GFFS)
    - data sharing
    - resource sharing

Wide Area Data Transfer

So how do you move that data?

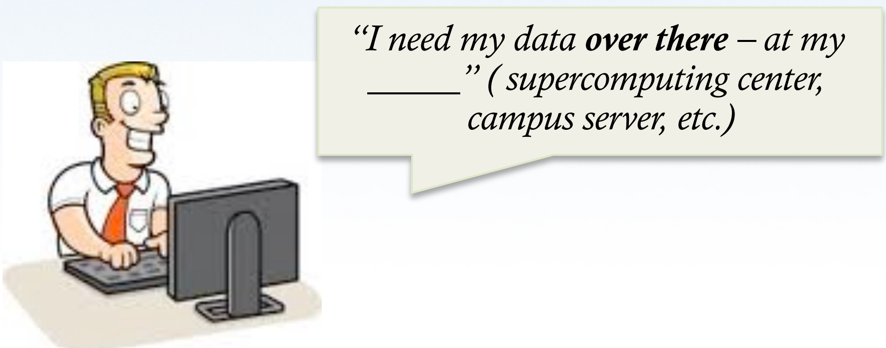
- **Bulk data transfer is *still* one of the most important distributed applications**
- **"Traditional" data transfer modes**
  - FTP and variants
    - UberFTP, GridFTP
    - scp
- **"Modern" data transfer modes**
  - parallel FTP and variants
    - GridFTP
      - striped transfers
      - third party transfers
    - tgcp, globus-url-copy, UberFTP
      - user interfaces built on top of GridFTP
  - wide-area filesystems
    - GPFS-WAN, Luster-WAN

From July 2006 presentation

NCSA

Globus Online: Moving Big Data *Easily*

- The Challenge:
  - what should be trivial…



  - can be painfully tedious and time-consuming

Adapted from slides from Steve Tuecke, UChicago

Excellent Example of Synergy

- A capability satisfying identified need
  - documented in our use cases and requirements
- An implementation satisfying architectural design
  - Borg-like attitude toward architectural designs
    - identify architecture behind known useful implementations
    - adapt/integrate to XSEDE's environment
  - GlobusOnline was ready to go and easily integrated into the XSEDE environment
- Integration process provides feedback for improvements to original developers
  - GlobusOnline team is implementing several improvements for full integration/acceptance as a production service in XSEDE

Transparency: publicly available project information

- Beginning to roll out documents now:
  - www.xsede.org/web/guest/project-documents
- Project Definition documents
  - Original Project Summary
- Policy documents
  - Service Provider Definition
  - SPs as members of the XSEDE Federation *(coming very soon)*
- Project Reports
  - *quarterly and annual reports to appear shortly*
- Architecture and Design documents
  - Level 1 and Level 2 Decomposition
- Engineering documents
  - System Requirements Specification
- Production Environment and Operations documents
  - Software and Service Baseline
  - Technical Security Baseline

Current XSEDE Compute Resources

- Kraken @ NICS
  - 1.2 PF Cray XT5
- Ranger @ TACC
  - 580 TF Sun Cluster
- Gordon @ SDSC
  - 341 TF Appro Distributed SMP cluster
- Lonestar (4) @ TACC
  - 302 TF Dell Cluster
- Forge @ NCSA
  - 150 TF Dell/NVIDIA GPU Cluster
- Trestles @ SDSC
  - 100TF Appro Cluster
- Steele @ Purdue
  - 67 TF Dell Cluster
- Blacklight @ PSC
  - 36 TF SGI UV (2 x 16TB shared memory SMP)

https://www.xsede.org/web/xup/resource-monitor

NSF

XSEDE

Current XSEDE Visualization and Data Resources

- Visualization
  - Nautilus @ UTK
    - 8.2 TF SGI/NVIDIA SMP
    - 960 TB disk
  - Longhorn @ TACC
    - 20.7 TF Dell/NVIDIA cluster
    - 18.7 TB disk
  - Spur @ TACC
    - 1.1 TF Sun cluster
    - 1.7 PB disk

https://www.xsede.org/web/xup/resource-monitor#advanced_vis_systems

- Storage
  - Albedo
    - 1 PB Lustre distributed WAN filesystem
  - Data Capacitor @ Indiana
    - 535 TB Lustre WAN filesystem
  - Data Replication Service
    - 1PB iRODS distributed storage
  - HPSS @ NICS
    - 6.2 PB tape
  - MSS @ NCSA
    - 10 PB tape
  - Golem @ PSC
    - 4 PB tape
  - Ranch @ TACC
    - 70 PB tape
  - HPSS @ SDSC
    - 25 PB tape

https://www.xsede.org/web/xup/resource-monitor#storage_systems

Current XSEDE Special Purpose Resources

- Condor Pool @ Purdue
  - 150 TF, 27k cores
- Keeneland @ GaTech/NICS
  - developmental GPU cluster platform
  - production GPU cluster expected in July 2012
- FutureGrid
  - Experimental/development distributed grid environment

https://www.xsede.org/web/xup/resource-monitor#special_purpose_systems

How to be successful at IT in a research environment…

- Recognize the ugly truths…
  - what you do is not important
    - compared to the science
  - your customers typically don't know what they want, much less what they need
    - though they will tell you anyhow
  - IT folks are still the "whipping boy"
  - what makes sense to an IT person does not necessarily make sense to the researcher
  - centralization/de-centralization will forever be an issue
  - standards are a Good Thing

    … but balance with best practices!!
    - there tends to be a lot of them
    - many are "community standards" that don't encompass the community

Some strategies…

- Always remember that what you do is about enabling science
  - it is NOT about cool tech
- Have a beer with a scientist
  - understand their requirements, not their proposed solutions
  - make sure they understand the difference
- Be the whipping boy
  - embrace the role, do a good job and it will all be OK
- The "better" is the enemy of the "good enough"
  - the latest and greatest technology is not always the most effective
  - research teams are not willing to jettison significant investments for a "better" answer

Identify the greatest needs and start there

Don't forget what you have learned – both good and bad!

- Balance diversity
  - common infrastructure is a Good Thing
    - it is often at odds with the inertia of a research project
  - be willing to adopt multiple technologies if they provide paths to success
- The customer is always right
  - usually…
- Protect your greatest asset
  - expertise of your staff

What are the grand challenge scientific goals in cyberinfrastructure?

- Wrong question, or at least a presumed context
  - implies cyberinfrastructure is still an experimental activity
  - implies shortcomings in deploying production cyberinfrastructure
  - to play devil's advocate, I will take the position of a production cyberinfrastructure advocate and not the role of cyberinfrastructure developer
- What is the right question then?
  - first and foremost, recognize the this is a muti-faceted engineering problem and not a science or computer science problem
    - systems engineering
    - software engineering
    - social engineering
  - most challenges well understood; problem is that there are many incompatible/non-interoperable solutions
- What are the challenges in designing, deploying and supporting a general purpose (inter)national cyberinfrastructure?
  - understanding requirements
  - distributed systems architecture, implementation and support (running trains)
  - community support, workforce development, support for development of novel services and capabilities

From US-India CI Workshop in December 2012


Our reach will forever
exceed our grasp, but,
in stretching our horizon,
we forever improve our world.
XSEDE
Extreme Science and Engineering
Discovery Environment