The Importance of Cyberinfrastructure for Research and Higher Education

Edward Seidel
Director, Office of Cyberinfrastructure
National Science Foundation
hseidel@nsf.gov
Center for Computation & Technology
Louisiana State University
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!
How did I get here?
Hints at where I will try to take this...

• High energy physics
• Black hole perturbation theory
  • Very hard, try supercomputers
• Black hole collisions
  • Very hard, try working with computer scientists
• Neutron star collisions
  • Very hard, try larger scale collaborations, add in high speed nets, grids
• Build interdisciplinary structure in LA
  • Very hard, try national scale!

Developing a working code was at least as hard as the theory!
How did I get here?

Hints at where I hope to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Grand Challenge teams, CS experts, new tech, critical to success
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Critical importance of international cooperation and networking
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbations
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Software engineering, tools, reusable components...
How did I get here?  
*Hints at where I will try to take this...*

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Grids: collab., data, computation for complex problems
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Visualization!
Local, remote, large data, collaboration, PR, science!
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

Optical networks, data deluge
How did I get here?
Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

All disciplines can use common components!
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!

How much this all costs!
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!
How did I get here?

Hints at where I will try to take this...

• High energy physics
• Black hole perturbation theory
  • Very hard, try supercomputers
• Black hole collisions
  • Very hard, try working with computer scientists
• Neutron star collisions
  • Very hard, try larger scale collaborations, add in high speed nets, grids
• Build interdisciplinary structure in LA
  • Very hard, try national scale!
How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
  - Very hard, try supercomputers
- Black hole collisions
  - Very hard, try working with computer scientists
- Neutron star collisions
  - Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
  - Very hard, try national scale!
Complex Problems

Require advanced CI

Mission 2009
An Important Complex Problem

Where is it going to go?
OptIPlanet Collaboratory Persistent Infrastructure
Supporting Microbial Research

Photo Credit: Alan Decker

iHDTV: 1500 Mbits/sec Calit2 to UW Research Channel Over NLR/CENIC/PW

Feb. 29, 2008

Ginger Armbrust's Diatoms: Micrographs, Chromosomes, Genetic Assembly

UW's Research Channel
Michael Wellings
We will need dynamically provisioned global lambdas to desktops, searchable distributed data archives, supercomputers, visualization facilities, coupled instruments for next generation science...
LHC, Gamma-ray bursts!

**Great Mysteries of Universe**

- **Gamma-ray bursts!**
  - All energy emitted in lifetime of sun bursts out in a few seconds: what are they?! Colliding BH-NS? SN?
  - GR, hydrodynamics, nuclear physics, radiation transport, neutrinos, magnetic fields: globally distributed collab!
  - Scalable algorithms, complex AMR codes, viz, PFlops*week, PB output!
- **LHC: What is the nature of mass? Higgs particle?**
  - ~10K scientists, 33+ countries, 25PB data, distributed!
  - Planetary lab for scientific discovery!
LHC, Gamma-ray bursts!

Great Mysteries of Universe

- Gamma-ray bursts!
  - All energy emitted in lifetime of sun bursts out in a few seconds: what are they? Colliding BH-NS? SN?
  - GR, hydrodynamics, nuclear physics, radiation transport, neutrinos, magnetic fields: globally distributed collaboration!
  - Scalable algorithms, complex AMR codes, viz. PFlops*week, PB output!

- LHC: What is the nature of mass? Higgs particle?
  - ~10K scientists, 33+ countries, 25PB data, distributed!
  - Planetary lab for scientific discovery!

5PB takes 1 week to move at 100Gb/s!!
LHC, Gamma-ray bursts!
Great Mysteries of Universe

- Gamma-ray bursts!
  - All energy emitted in lifetime of sun bursts out in a few seconds: what are they?!
  - Colliding BH-NS? SN?
  - GR, hydrodynamics, nuclear physics, neutrino transport, neutrino radiation, magnetic fields: globally distributed collaboration!
  - Scalable algorithms, complex AMR codes, viz, 1 Petaflops/week, PB output!

- LHC: What is the nature of mass? Higgs particle?
  - ~10K scientists, 33+ countries, 25PB data, distributed!
  - Planetary lab for scientific discovery!

5PB takes 1 week to move at 100Gb/s!!

Now, compare with observation...LIGO!
Major computation just for data analysis itself!

5PB takes 1 week to move at 100Gb/s!!

Edward Seidel
hseidel@nsf.gov

National Science Foundation
Where Discoveries Begin
Office of Cyberinfrastructure
Science, Engineering, Society are being Transformed

- Complex problems require totally new methodologies
  - Scale computing beyond current experience...
  - Large scale data, beyond...
  - Large scale collaborations, beyond...
- Good news!
  - Cyberinfrastructure!
- Not so good news!
  - We are not moving fast enough, spotty CI coverage!

The quest for knowledge used to begin with grand theories. Now it begins with massive amounts of data. Welcome to the Petabyte Age.
Analysis: What is needed

- **Comprehensive, balanced, integrated national high performance cyberinfrastructure!**
- Integrates campuses, multiple CIs, multiple agencies, instruments, data repositories, grids, HPC environments, grids, optical nets
- With advanced software environments, reusable components, usable algorithms
- Supports large scale distributed collaborations
- **Support for cyber-workforce development**
- Computational Science as a career path
  - Researchers who develop, prototype the next generation CI
  - Advanced scientists who use this for their scientific research
- New scales of interdisciplinarity, proper reward system
Cyberinfrastructure Vision

We actually already have this!

- “Atkins report” - Blue-ribbon panel, chaired by Daniel E. Atkins
- Called for a national-level, integrated system of hardware, software, & data resources and services
- New infrastructure to enable new paradigms of science & engineering research and education with increased efficiency
NSF Vision

1. Virtual Organizations for Distributed Communities

2. High Performance Computing

3. Data & Visualization/Interaction

4. Learning & Work Force Needs & Opportunities
What is being done to address this?
National CI Blueprint
National CI Blueprint
National CI Blueprint

- Track 1
- Track 2
- Track 2
- Track 2

Viz

XD
National CI Blueprint

- Exascale (multi-agency?)
- Nets
- Track 1
- Track 2
- OSG
- LC
- Campus
- DataNet
- Algorithms
- Software
- Viz
- XD
- Nets (campus)
National CI Blueprint

- Track 1
  - Exascale (multi-agency?)
  - Nets
- Track 2
  - DataNet
  - Software
  - Algorithms
- Apps
- DataNet (campus)
- NsD
- Apps
- OSG
- LC
- Campus
- Apps
- Apps
- Apps
- Apps

Edward Seidel
hseidel@nsf.gov

National Science Foundation
Where Discoveries Begin

Office of Cyberinfrastructure
NSF Shared Resource Environments

Courtesy of University of Indiana
NSF Shared Resource Environments

Computers

Data services

Visualization services

People

Courtesy of University of Indiana
NSF Shared Resource Environments

Computers
Data services
Visualization services
People

Modeling and simulation
Data analysis & visualization
User support
Training
Common user environments
Tools for educators
Science Gateways

2007-8 Track 2

Courtesy of University of Indiana

Edward Seidel
hseidel@nsf.gov

National Science Foundation
Where Discoveries Begin
Office of Cyberinfrastructure
NSF Shared Resource Environments

Computers
Data services
Visualization services
People

Modeling and simulation
Data analysis & visualization
User support
Training
Common user environments
Tools for educators
Science Gateways

2007-8 Track 2
2008-9 Track 2

Courtesy of University of Indiana

Edward Seidel
hseidel@nsf.gov

Office of Cyberinfrastructure
National Science Foundation
Where Discoveries Begin
NSF Shared Resource Environments

Computers
Data services
Visualization services
People

Modeling and simulation
Data analysis & visualization
User support
Training
Common user environments
Tools for educators
Science Gateways

2007-8
Track 2

2008-9
Track 2

2009-10
Track 2

2010-11
Track 2

Courtesy of University of Indiana

National Science Foundation
Where Discoveries Begin
Edward Seidel
hseidel@nsf.gov
Office of Cyberinfrastructure
NSF Shared Resource Environments

- Computers
- Data services
- Visualization services
- People

Modeling and simulation
Data analysis & visualization
User support
Training
Common user environments
Tools for educators
Science Gateways

2007-8 Track 2
2008-9 Track 2
2009-10 Track 2
2010-11 Track 2

Courtesy of University of Indiana
- Access to 45,000 Cores, 6 Petabytes Disk, 15 Petabytes Tape
- >15,000 CPU Days/Day
  - ~85% Physics: LHC, Tevatron Run II, LIGO;
  - ~15% non-physics: biology, climate, text mining,
  - Including ~20% Opportunistic use.
- Virtual Data Toolkit: Common software developed between Computer Science & applications used by OSG and others.

Partnering with:
- **US LHC**: Tier-1s, Tier-2s, Tier-3s
- **Campus Grids**: Clemson, Fermilab, Purdue, Wisconsin,
- **Regional & National Grids**: TeraGrid, New York State Grid, EGEE, UK NGS
- **International Collaboration**: South America, Central America, Taiwan, Korea, UK.

Source: Miron Livny
The Worldwide LHC Grid: Open Science Grid (US), EGEE (EU), and beyond...33 countries!

Tier 1
10 - 40 Gbps

Tier 1

Tier 0 +1

Tier 2

Tier 2 Center

Tier 2 Center

Tier 2 Center

Tier 3

Tier 3

Tier 4

Several Petabytes in 2008
An Exabyte ~7 Years later
100 Gbps+ Data Networks

Source: Harvey Newman

Emerging Vision: A Richly Structured, Global Dynamic System
$100M DataNet Program
(Sustainable Digital Data Preservation & Access Network Partners)

• Goals:

• Catalyze development of multi-disciplinary science & engineering data collections: open, extensible & evolvable, sustainable over 50+ years.

• Support development of a new generation of tools & services facilitating data acquisition, mining, integration, analysis, visualization.

STCI: Strategic Technologies for CI

- Core OCI program
  - Support development and/or demonstration of innovative cyberinfrastructure services for science and engineering research and education
  - Want highly innovative cyberinfrastructure education, outreach and training proposals that lie outside the scope of targeted solicitations

- Two dates each year: August and February
  - Cross Foundational program with participation from OCI, MPS, CISE, Engineering, SBE and GEO
  - $26M in 18 awards each of less than $2M over 3-5 years. Wide range of science/engineering Apps
  - $18M investment planned for FY09-10. Solicitation currently open -- due date Oct 30, 31 2008
International Research Network Connections (IRNC)

Awards

- TransPAC2 (U.S. – Japan and beyond)
- GLORIAD (U.S. – China – Russia – Korea)
- Translight/PacificWave (U.S. – Australia)
- TransLight/StarLight (U.S. – Europe)
- WHREN (U.S. – Latin America)
What next?

- We have an excellent high level vision document!
  - We have started to address some parts of it very well...
  - We have started to address other parts of it in pieces...
  - Some parts are missing

- All parts need to be aggressively built out, integrated, coupled to/driven by applications
  - Need *end-to-end* integration; networks critical

- Need to address support of a computational science community
  - Computational Science is 3rd pillar of science & engineering?
  - Who creates, prototypes, uses next generation CI?
Translation to Programs
Themes for future development

- Need overarching CI architecture
  - How do things plug in?
  - Build on, reuse what is out there!
- Much stronger sense of integration
  - Advanced campus infrastructure
  - Loosely coupled grids
  - High end supercomputer grids
  - Instruments, MREs, sensors
  - Data archives, repositories
Translation to NSF Programs

Themes for future development

- High performance optical nets, bandwidth (lambdas?) to desktop
  - GLIF, NLR, I2, RONS, Campuses: end-to-end!
  - Bandwidth doubling every 9 months? Researchers are choked at their labs!
    - A. Bement: “Those massive conduits are reduced to two-lane roads at most college and university campuses”
- National CI must be well integrated with campus, local lab, desktop CI
  - To be developed: campus integration!
  - This is where all the researchers are!
  - Compute, manage, visualize, analyze the data...from lab!
Translation to NSF Programs

Themes for future development

- Advanced application development
- HPC, grids, data, software, tools, nets
- VOs, collaboratories
- Integrated with CI development teams
Campus Integration

- Campus CI as a 1st class element of national CI
  - Example: OptIPuter and CalIT2/UCSD Quartzite
    - Dozens of optical paths connecting campuses and labs within campus, prototyping campus of future
    - NSF ITR and MRI awards

- Integration between the campus and the national and international fabrics
  - Regional grids, Open Science Grid, Teragrid, network connectivity (EPSCoR role), IRNC, Internet-2 and NLR, remote instruments, archives, more
  - File systems: too many places files can exist!
Federated Identity Management

- Researchers live with bewildering array of accounts, logins, passwords!
- Please: a system to adopt local credentials!
  - Single UserID and password for everything!
  - Transparent access to all systems: campus, world
  - Improve security, reduce support

- InCommon
  - Over 70 universities, numerous agencies, 30 corporations (Apple, Microsoft, Internet-2)
    - NSF Pilot program starting
    - FastLane, half-dozen universities, TeraGrid
- How fast can we get there?
OCI Networking

- Under development – input welcome
- Model – the 2003 Experimental Infrastructure Networking Program (EIN)
- EIN Key Characteristics
  - Enabling E-Science apps that cannot be supported by today’s production network capabilities
  - Encourage vertical teaming network-app layer (scientist/user) and industry engagement
- Possible launch of EIN-like program in future
This is All Computational Science!

- This third pillar of science needs real support!
  - Who prototypes, develops next generation CI?
  - Where do they find support?
  - Need to address this!

- Must be done in partnership with
  - Other NSF directorates (e.g., CISE, MPS, BIO, GEO, etc)...
  - Other agencies (DOE, NIH, NASA, NOAA, etc...) 
  - Other countries (EU, Asia, etc...)
Summary

- Excellent Vision already in place
  - Atkins report, creation of OCI, initial steps all good
- Comprehensive, balanced, integrated, national high performance cyberinfrastructure needed!
  - HPC, Grids moving towards Extreme Digital (XD)
    - Vast set of challenges remain
- Many researchers concerned with other types of CI
- Underdeveloped or missing pieces include
  - Crisis in software, tools, applications, VOs
  - Data centric CI, optical networks to the desktop
  - Computational science needs support as a discipline
- Need help! Task forces, workshops coming...