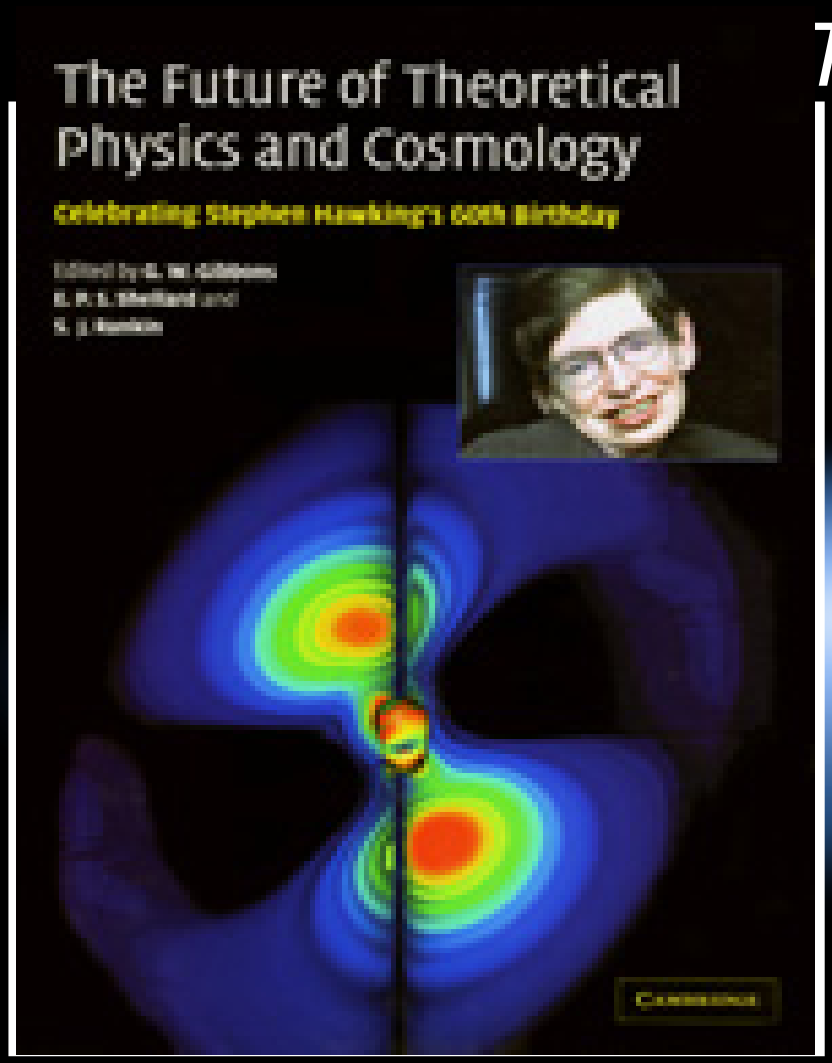
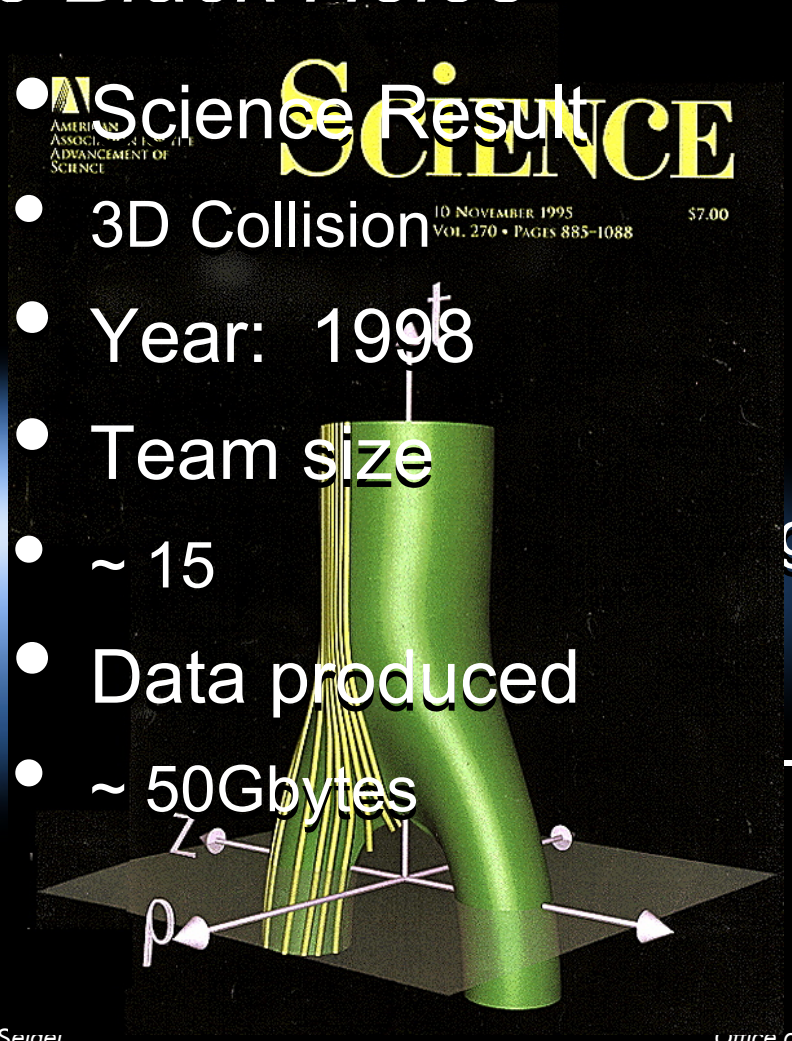
Importance of Cyberinfrastructure for Research and Education

Edward Seidel
Director, Office of Cyberinfrastructure
National Science Foundation
hseidel@nsf.gov

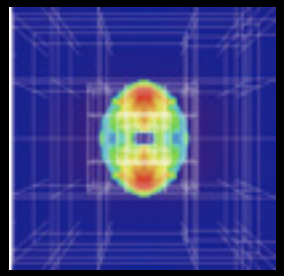
Transformation of Science



Two Black Holes



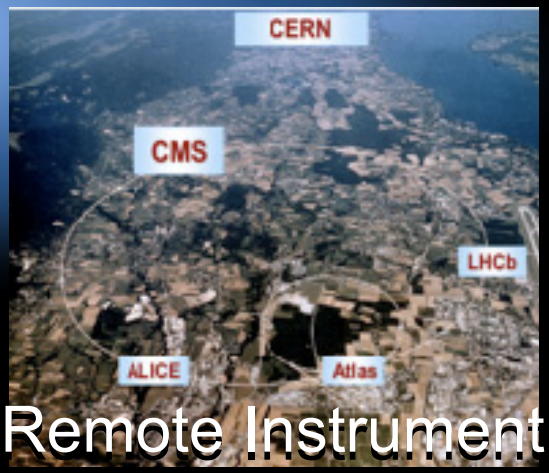
- Science Result
- 3D Collision
- Year: 1998
- Team size
- ~ 15
- Data produced
- ~ 50Gbytes



Complex Universe
LHC, rsts!

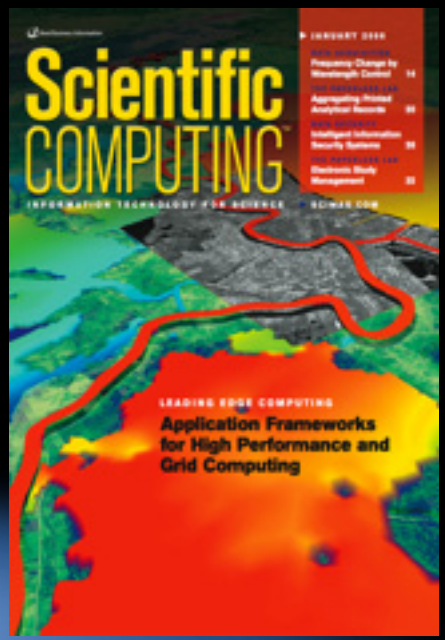
Now, compare with observation...LIGO! GR becomes a data-driven science

- Gamma-ray bursts!
  - All energy emitted in lin out in a few seconds: what 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!

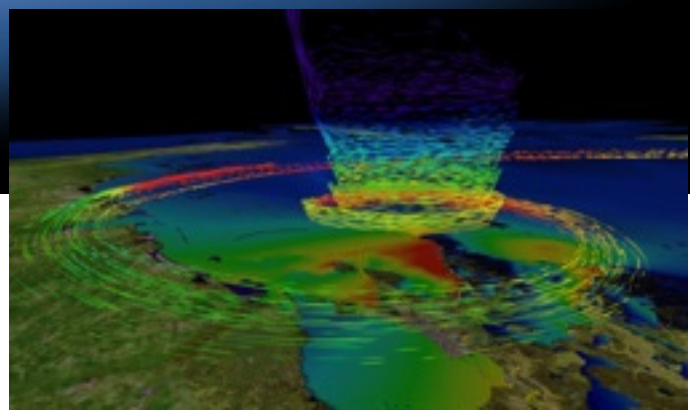


Grand Challenge Communities Combine it All...

Where is it going to go?



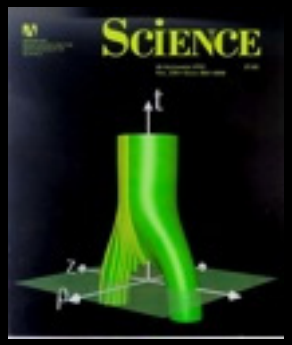
QuickTime™ and a
Video decompressor



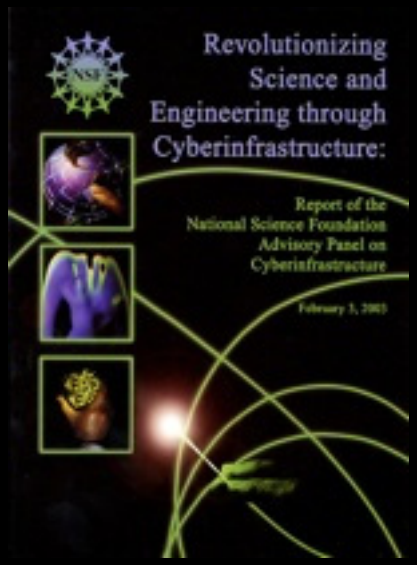
Data-Driven Collaborations for Complex Problems

Great Challenges of 21st Century

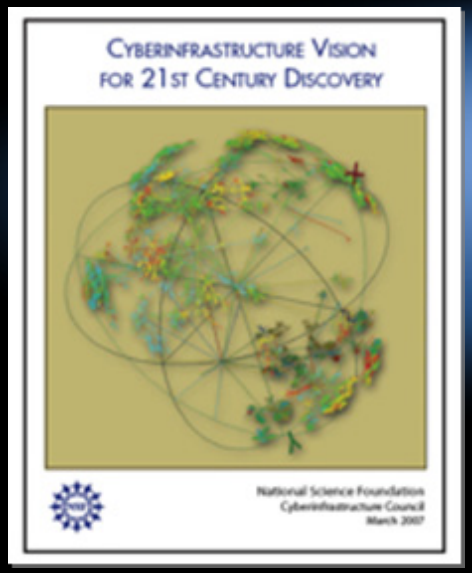
- Every field of science
  - General Relativity, HEP, Geosciences, Bio, SBE...
  - And all combinations...
- Science and Society being transformed by CI and Data
  - Completely new methodologies
  - "The End of Science" (as we know it)
- CI plays central role
  - No community can attack challenges
 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

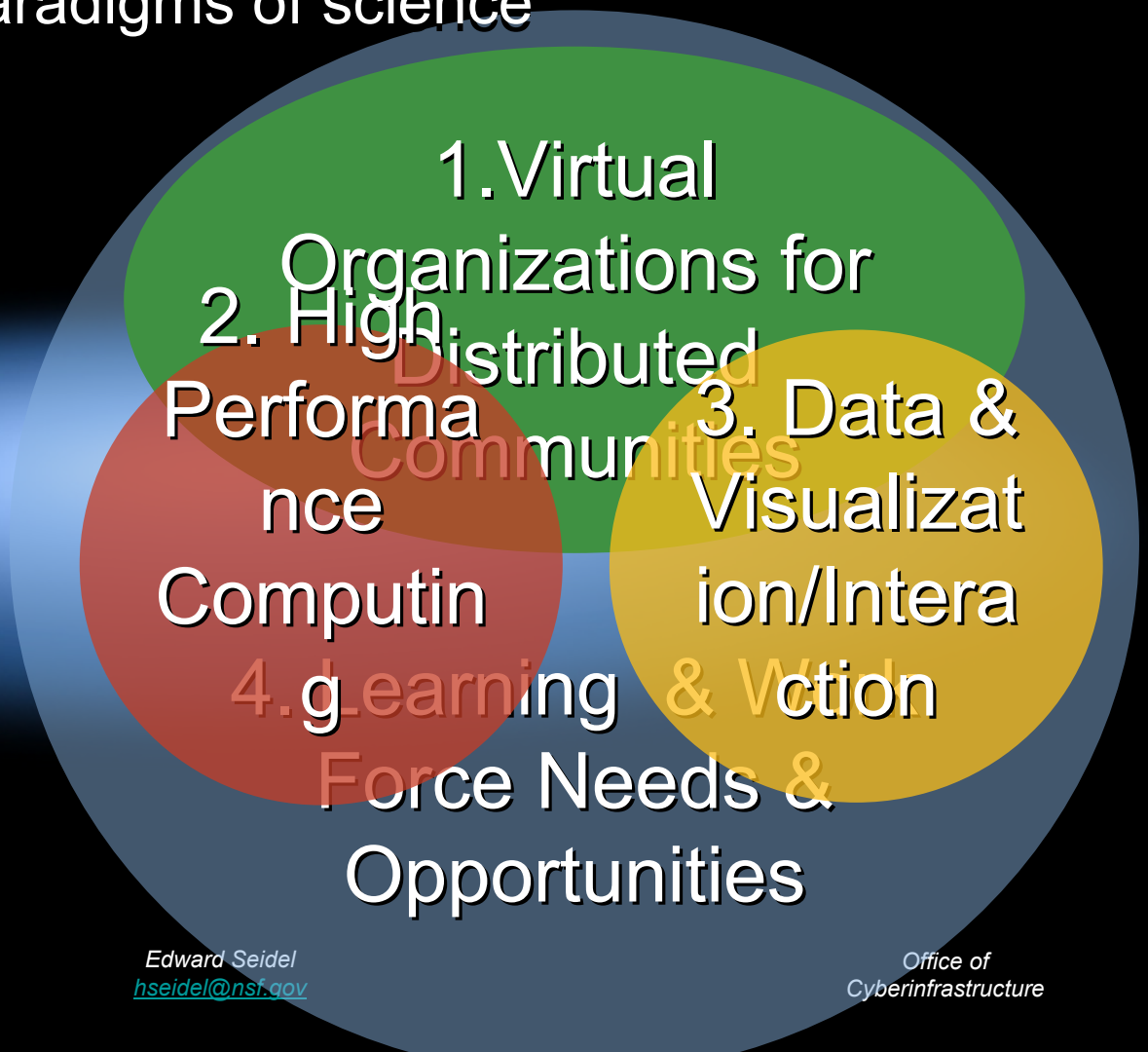


NSF Vision



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



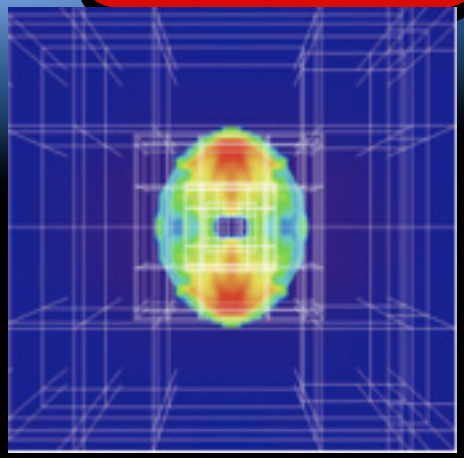


High End Computing

- "Modeling, simulation, and knowledge from data collections, [which] is increasingly essential to scientific and engineering
- Sustained petascale capable systems; going beyond HPC!
- HPC software and tools
- Necessary scalable applications
- Sharing among academic institutions to optimize the accessibility and use of HPC assets deployed and supported at the campus level"

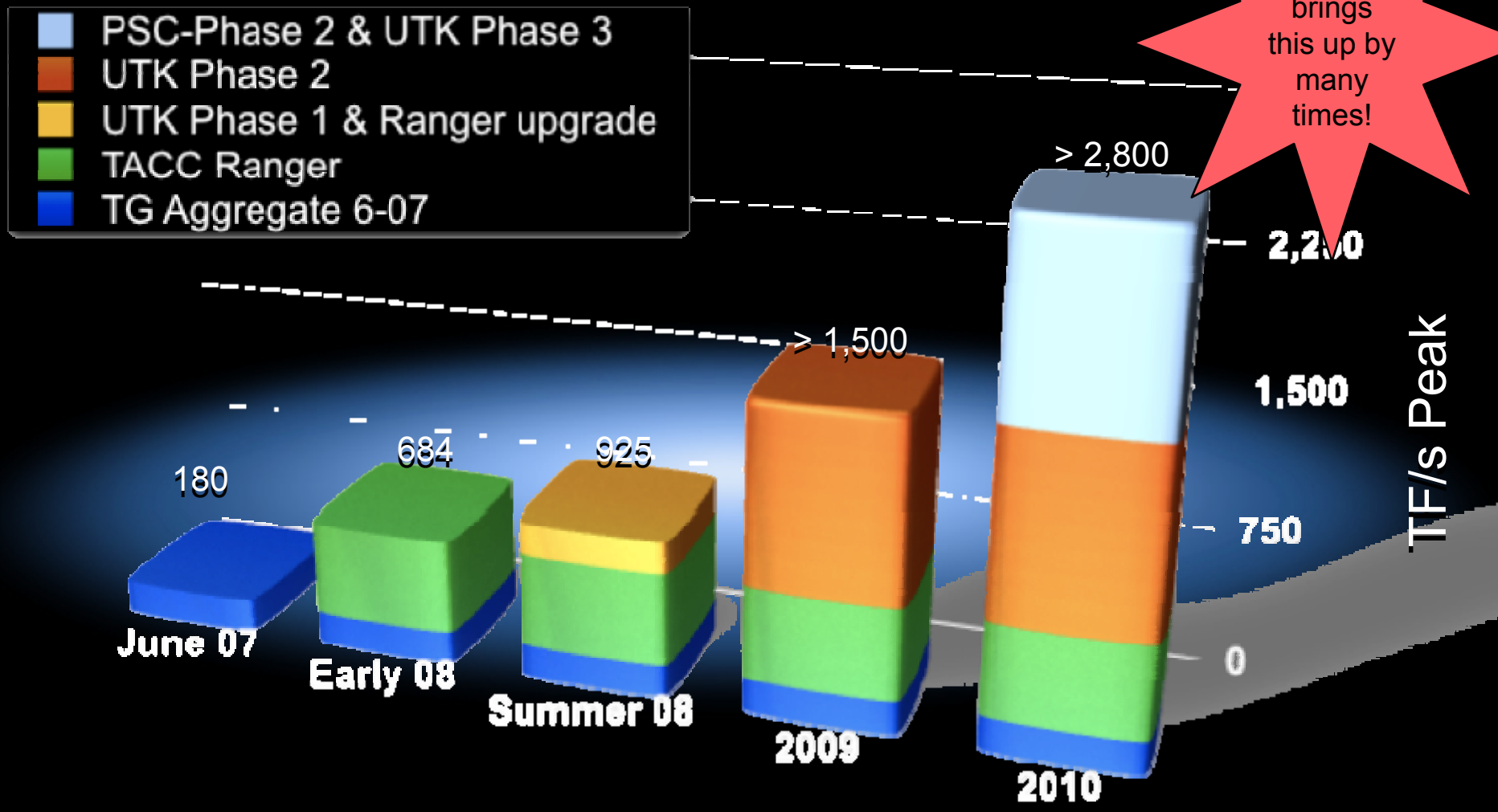
Toolkit for complex CI?

AMR on a million processors?



Impact

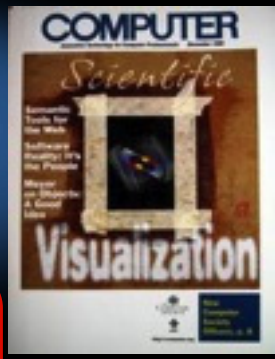
Greatly expanding capacity of the TeraGrid for digital exploration with reduced oversubscription and queue wait times.



Data, Data Analysis, and Visualization

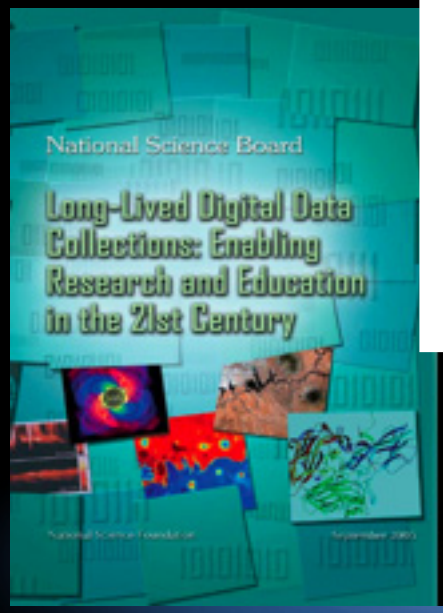
- "Any cogent plan must addres
  phenomenal growth of data ir
- Goals are to
  - Catalyze the development of a system of science and engineering data collections that is open, extensible, and evolvable
  - Support development of a new generation of tools and services for data discovery, integration, visualization, analysis and preservation
- The resulting national digital data framework will be an integral component in nationa

Where is my simulation data? Compare with experiment?

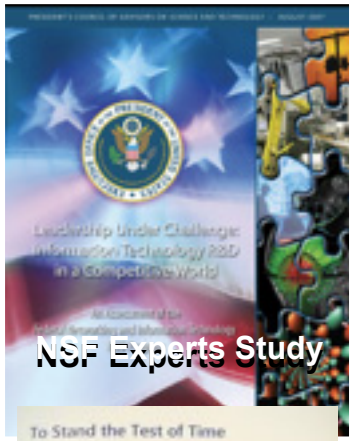


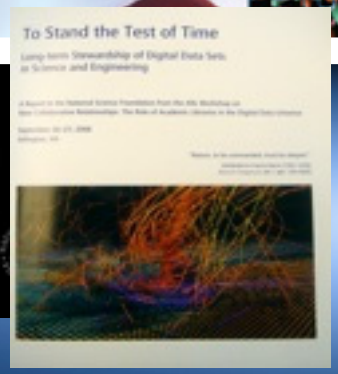
I can't visualize my results!

NSB Report: Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century



PCAST Digital Data





Industry

Storage Networking Industry Association (SNIA) *100 Year Archive Requirements Survey Report*

"there is a pending crisis in archiving… we have to create long-term methods for preserving information, for making it available for analysis in the future." 80% respondents: >50 yrs; 68% > 100 yrs

Wired, Nature

"The data deluge represents an opportunity to advance U.S. leadership in science and technology, and harnessing it has become a national priority"

$100M DataNet Program (5 Years)

(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.
- Status: NSB approved two awards! Round 2 pre-proposals Nov. 13, w/invited full proposals due May 15, 2009.

- User-centric
- Multi-Sector
- Sustainable
- Extensible
- Evolvable
- Nimble
- Reliable

Virtual Organizations for Distributed Communities

- "A VO functions as a coherent unit...through the use of end-to-end CI systems, providing shared access to resources and services, often in realtime
  - Technological framework...experimental facilities, instruments and sensors, applications, tools, middleware, remote access
  - Operational framework from campus level to international scale..."
- Specific interpretation: Next generation Grand Challenge communities for science, engineering, humanities...

Edward Seidel
hseidel@nsf.gov

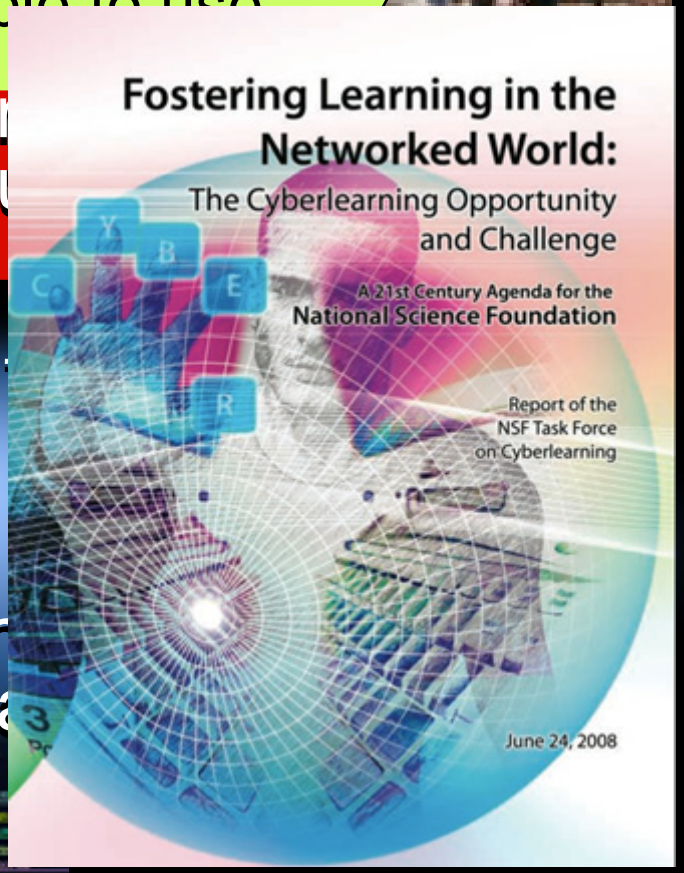
VOSS, VOs

- VOSS
- What constitutes effective virtual organizations? How do they enhance research and education production and innovation?
- CDI
- VOs
- Science Gateways: NanoHub, NEES, LEAD...

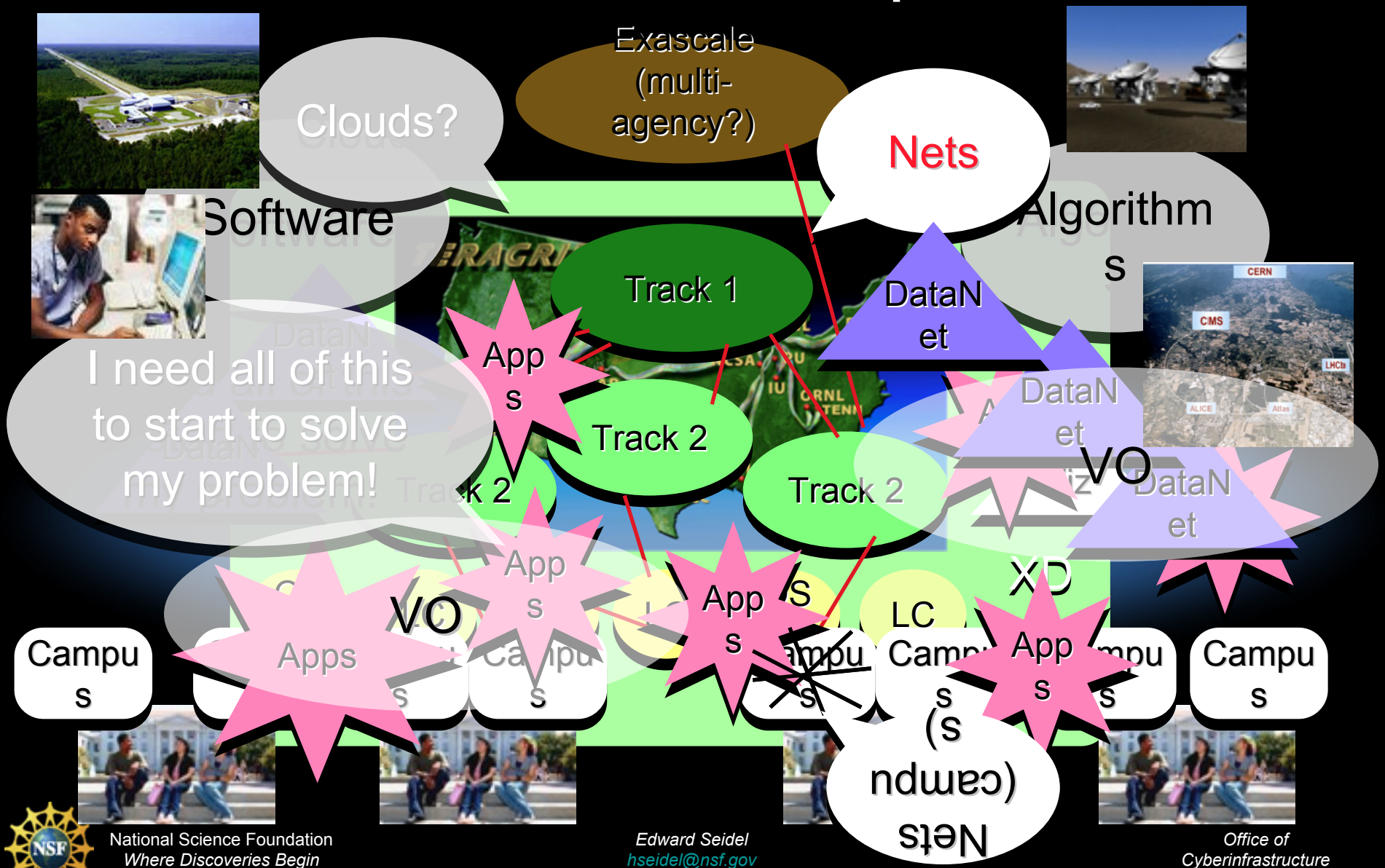
Learning and Develo

We must be much more aggressive about training people to use this!

- "NSF will:
  - Identify and address the barr utilization of cyberinfrastructu services, and resources
  - Promote the training of facul students, researchers
  - Encourage programs that wi exploit cyberinfrastructure, ir advantage of the internationa provides..."



National CI Blueprint



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 entirely
- 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? Note: my def^{n} of comp. science is broad!
  - Who creates, prototypes, uses next generation CI?

What about Campuses?

- The next generation science will require **huge** amounts of data to be manipulated, retrieved, stored, analyzed
- Collaborative environments will be needed at unprecedented levels of sophistication
  - We can barely do low-end video conferencing today! HD, Optiportal-level environments needed
- It is very hard to use modern CI!
- *We need to seriously rethink our campus environments and how they can support new modalities of research, collaboration, education*

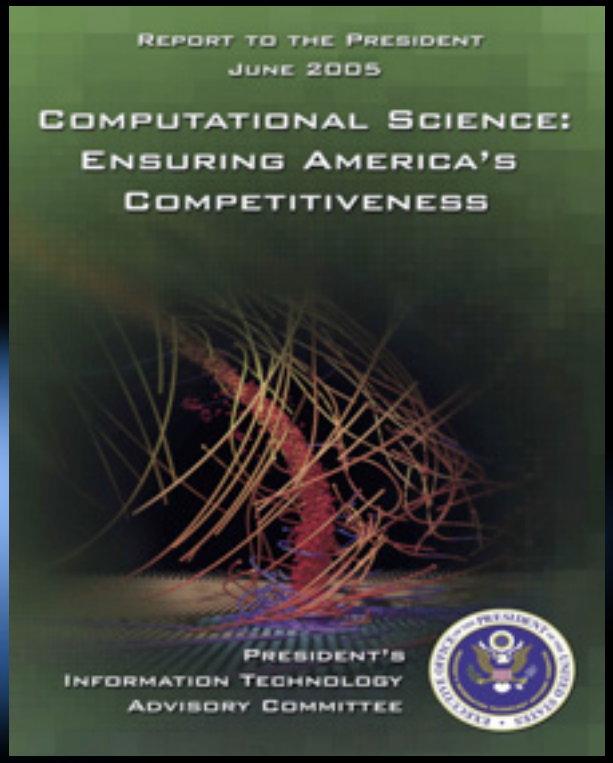
iHDTV: 1500 Mbits/sec Calit2 to UW Research Channel Over NLR/CENIC/PW

Computational Science
PITAC Report Summary

"Universities must significantly change organizational structures; multidisciplinary & collaborative research are needed for US to remain competitive in global science"

"Itself a discipline, computational science serves to advance all of science, and adequate and competitive, adequate and competitive, adequate and competitive, adequate and competitive, now constitutes the third pillar of scientific inquiry… multidisciplinary field"

"Technology is the third pillar of scientific inquiry…"



Edward Seidel & Fahmida Chowdhury
hseidel@nsf.gov & fchowdhu@nsf.gov

Office of Cyberinfrastructure
Social, Behavioral & Economic
Sciences Directorate

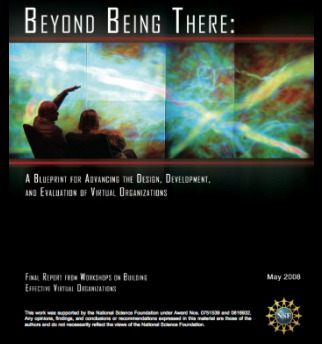
Finding a Foundation for CS&E

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, Bio, Geo, SBE, etc...
  - Create new CI/CSE-research agenda
    - Software, networks, compute systems (not just HPC!)
    - Support those who prototype, use CI
    - These people are central to our f
  - Education and Workforce dev
    - CAREER awards, curriculum de
  - Help universities align
    - Curriculum, best practices, rev ... Nee organize community!

How do we catalyze creation of 21st century universities?

Rough Plan for OCI

- Attempt to fill out holes in current CI coverage
  - Data, software, algorithms, networks...
  - Alternate computing modalities beyond HPC: Clouds, data, campus integration, etc...
  - Task Forces: Campus Bridging, Education, Computing, Software, Data, GC Communities

- Workforce development!
- OCI as base for CI development, deployment & *computational science* research and education
- New programs: VOs attack complex problems at frontiers of science and engineering
  - CI-driven, with above pieces in place first
  - This is the heart of computational science!