Federal Networking & Information Technology Research Opportunities at NSF

Deborah L. Crawford

Directorate for Computer and Information Sciences and Engineering

National Science Foundation

Federal Support for Research

- National Science Foundation (NSF)
  - Opportunity driven; longer term
- National Institutes of Health (NIH)
  - Use inspired basic research; longer term
- Mission Agencies: DARPA, DOE, NSA, NASA, etc.
  - Applied research; shorter term

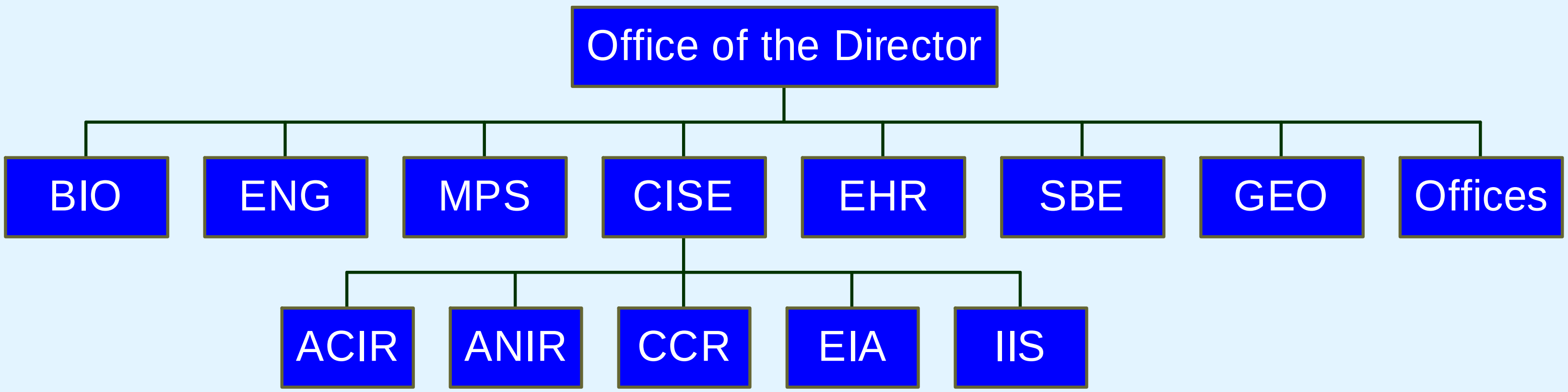
National Science Foundation (NSF)

- Created in 1950

  "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes"

- Organization
  - Directorates and Offices generally organized by discipline
  - Directorates have divisions

Organization: NSF's Alphabet Soup



NSF/CISE Strategic Priorities

- IDEAS
  - Strengthen the computer/information sciences/engineering knowledge base
- PEOPLE
  - Prepare a world-class IT workforce
- TOOLS
  - Provide computing resources and tools that empower the research and education community
- Improve organizational effectiveness

CISE Funding Profile (FY 2003)

CORE

- Advanced Computational Infrastructure and Research (ACIR) – $85M
- Advanced Networking Infrastructure and Research (ANIR) – $68M
- Computer-Communications Research (C-CR) – $70M
- Experimental and Integrative Activities (EIA) – $62M
- Information and Intelligent Systems (IIS) – $51M

CROSSCUTTING

- NSF-wide Information Technology Research (ITR) Priority Area
  - $213M (CISE) + $76M (Other Directorates/Offices) + $10M (Terascale)

CISE CORE

CCR: Computer – Commmunications Research

- Intellectual thrusts
  - – Trusted Computing
  - – Embedded & Hybrid Systems
  - – Distributed Systems and Compilers
  - – Computer Systems Architecture
  - – Design Automation
  - – Signal Processing Systems
  - – Graphic, Symbolic and Geometric Computation
  - – Theory of Computing
  - – Software Engineering & Languages
  - – Communications Research

IIS: Information and Intelligent Systems

- Intellectual Thrusts
  - Digital Society and Technologies
  - Human Computer Interaction
    - Includes Universal Access
  - Information and Data Management
  - Knowledge and Cognitive Systems
  - Robotics and Human Augmentation
  - Digital Libraries

ACIR: Advanced Computational Infrastructure and Research

- Advanced Computational Infrastructure
  - supports two advanced computational infrastructure partnerships: National Computational Science Alliance & National Partnership for Advanced Computational Infrastructure
  - coordinates Terascale Computing investments, construction of the Extensible Terascale Facility
- Advanced Computational Research
  - data handling and visualization
  - software tools
  - scalable algorithms

ANIR: Advanced Networking Infrastructure and Research

- Advanced Network Research
  - high speed networking, security and trustworthiness
  - sensor networks
  - networking research testbeds
- Advanced Network Infrastructure
  - network middleware infrastructure
  - high performance network connections
  - strategic technologies for the internet
  - experimental infrastructure testbeds

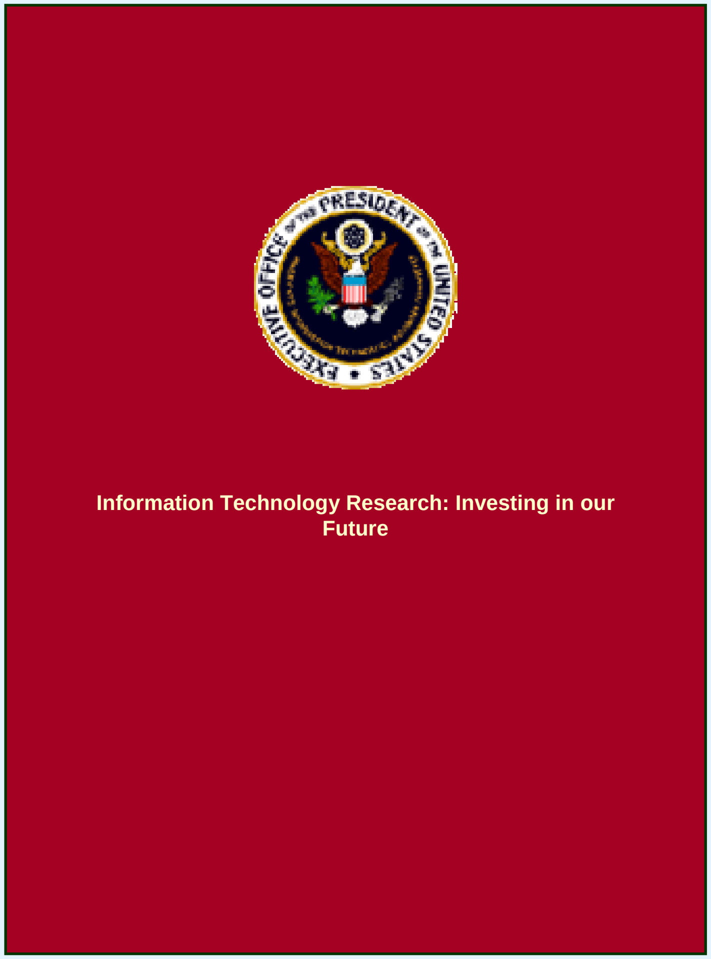
EIA: Experimental and Integrative Activities

- Instrumentation and infrastructure for core CISE research and education
- Education and workforce programs
- Multidisciplinary research and education – mainly at the frontiers of Bio and IT
- Special projects: symposia, travel, international activities, studies

ITR

PITAC Report (February 1999)

http://www.ccic.gov/ac/report/



"Information technology will be one of the key factors driving progress in the 21st century - it will transform the way we live, learn, work, and play. Advances in computing and communications technology will create a new infrastructure for business, scientific research, and social interaction."

ITR: Information Technology Research

- 1999 Report recommended doubling support for Federal IT R&D with an addition to annual investment of $1.37 billion
- NSF as lead agency
- Diversify modes of support – larger, more interdisciplinary, longer duration
- Focus Areas evolved over the years

ITR Priority Area (FY 2000 – 2004)

- NSF-wide Priority Area
  - 4 Program Solicitations to date
  - Crosscutting Activity
    - All areas of science and engineering participating
  - Range of award sizes
    - “Small” grants ≤ $500K total for 3 years
    - “Medium” grants added later (between $500K & $1M/year for up to 5 years)
    - “Large” awards ≤ $3M/year for up to 5 years

ITR Objectives

- ITR addresses the fundamental relationship between the creation of knowledge and the information tools that support the quest for new knowledge.
  - ITR theoretical research addresses the nature of information and the limits of computation.
  - ITR awards advance knowledge in every field of science, medicine, industry, education, and government.

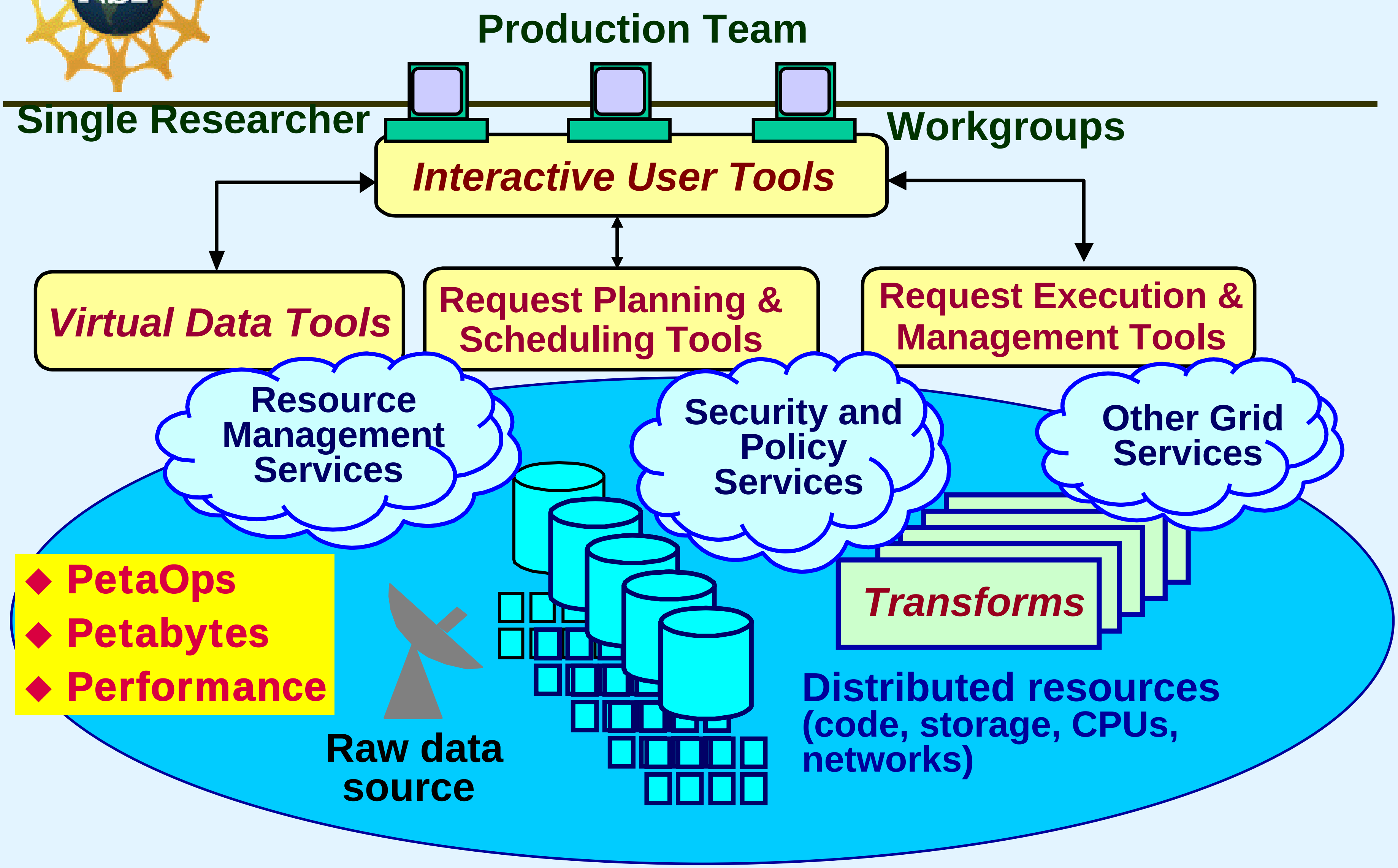
ITR Successes

- GriPhyN/iVDGL - An IT Infrastructure for Collaborative Science
- Networked InfoMechanical Systems - Adding New Dimensions to Embedded Networked Sensing
- Health Monitoring of Highway Bridges and Civil Infrastructure

GriPhyN and iVDGL Grid

- Basic composition -- (~120 people)
  - GriPhyN: 12 funded univ., SDSC,3 labs (~80people)
  - iVDGL: 16 funded insts., SDSC, 3 labs (~80 people)
  - Experts: US-CMS, US-ATLAS, LIGO, SDSS/NVO
  - Large overlap of people, institutions, management
- Grid research/infrastructure and Grid deployment
  - GriPhyN: CS research, Virtual Data Toolkit (VDT) development
  - iVDGL: Grid laboratory deployment
  - 4 physics experiments provide frontier challenges
  - VDT in common

PetaScale Virtual-Data Grids



Networked Infomechanical Systems
Adding New Dimensions to
Embedded Networked Sensing

W. Kaiser and D. Estrin
UCLA

ITR, Sensors and the Environment

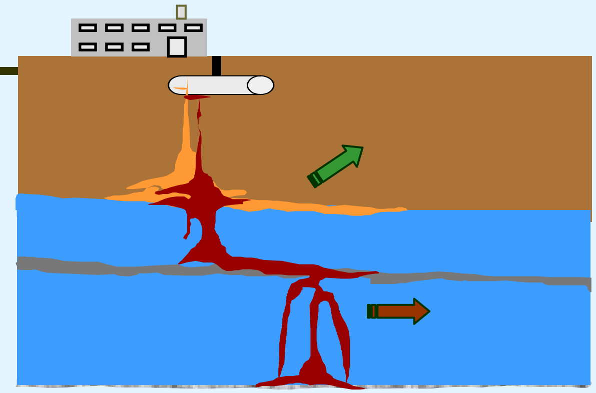
Seismic Structure response

Marine Microorganisms

Micro-sensors, on-board processing, wireless interfaces feasible at very small scale--can monitor phenomena "up close"

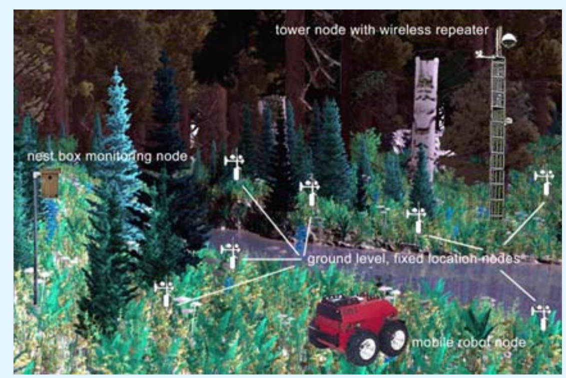
Enables spatially and temporally dense environmental monitoring

Embedded Networked Sensing will reveal previously unobservable phenomena



Contaminant Transport

Ecosystems, Biocomplexity

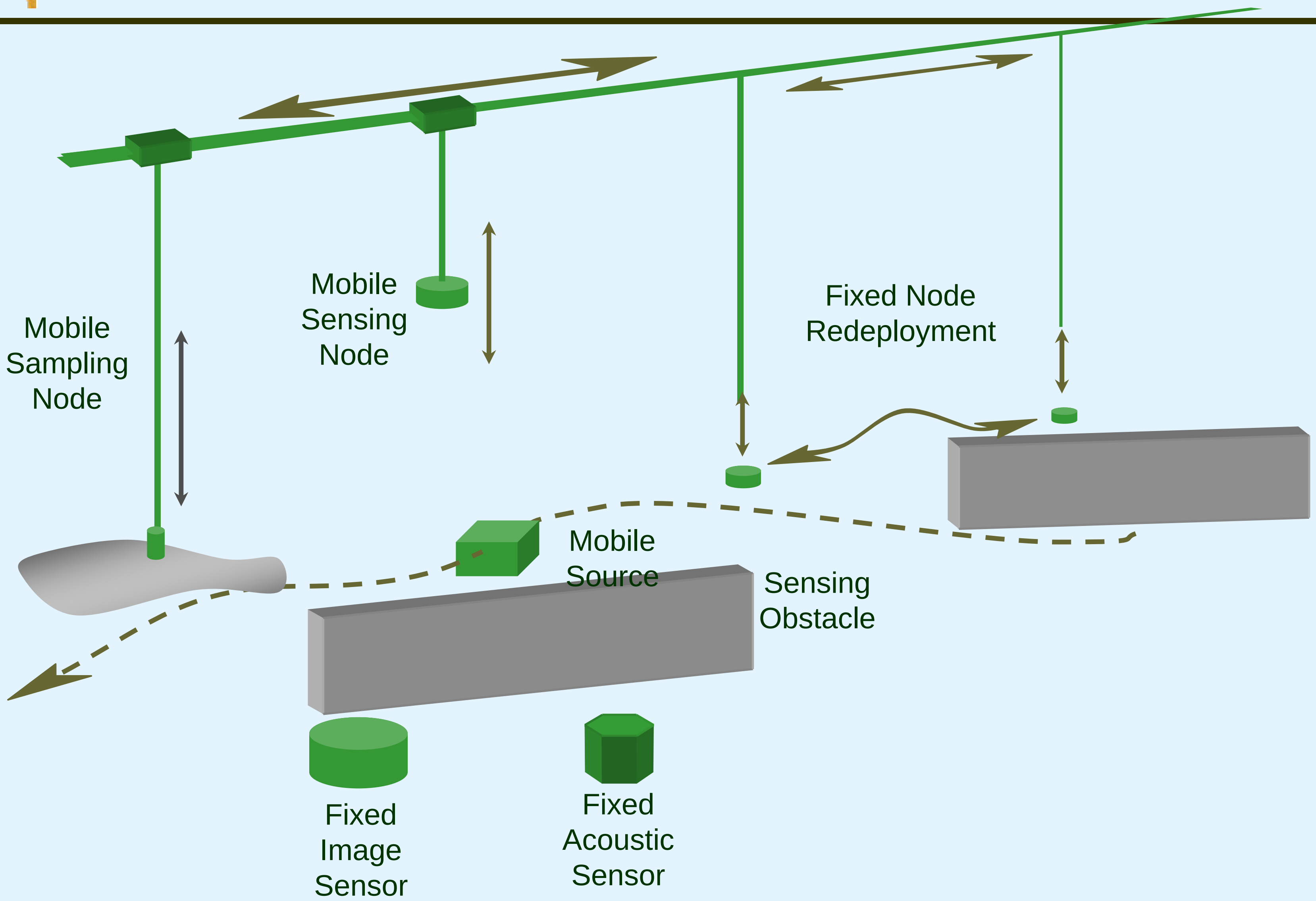


NIMS Architecture

Distributed, autonomous embedded systems

- Self-aware of their system-wide *sensing uncertainty*
- Exploit sensing *diversity*
- Exploit physical world reconfigurability to optimize system performance
- Guided by information theoretic framework

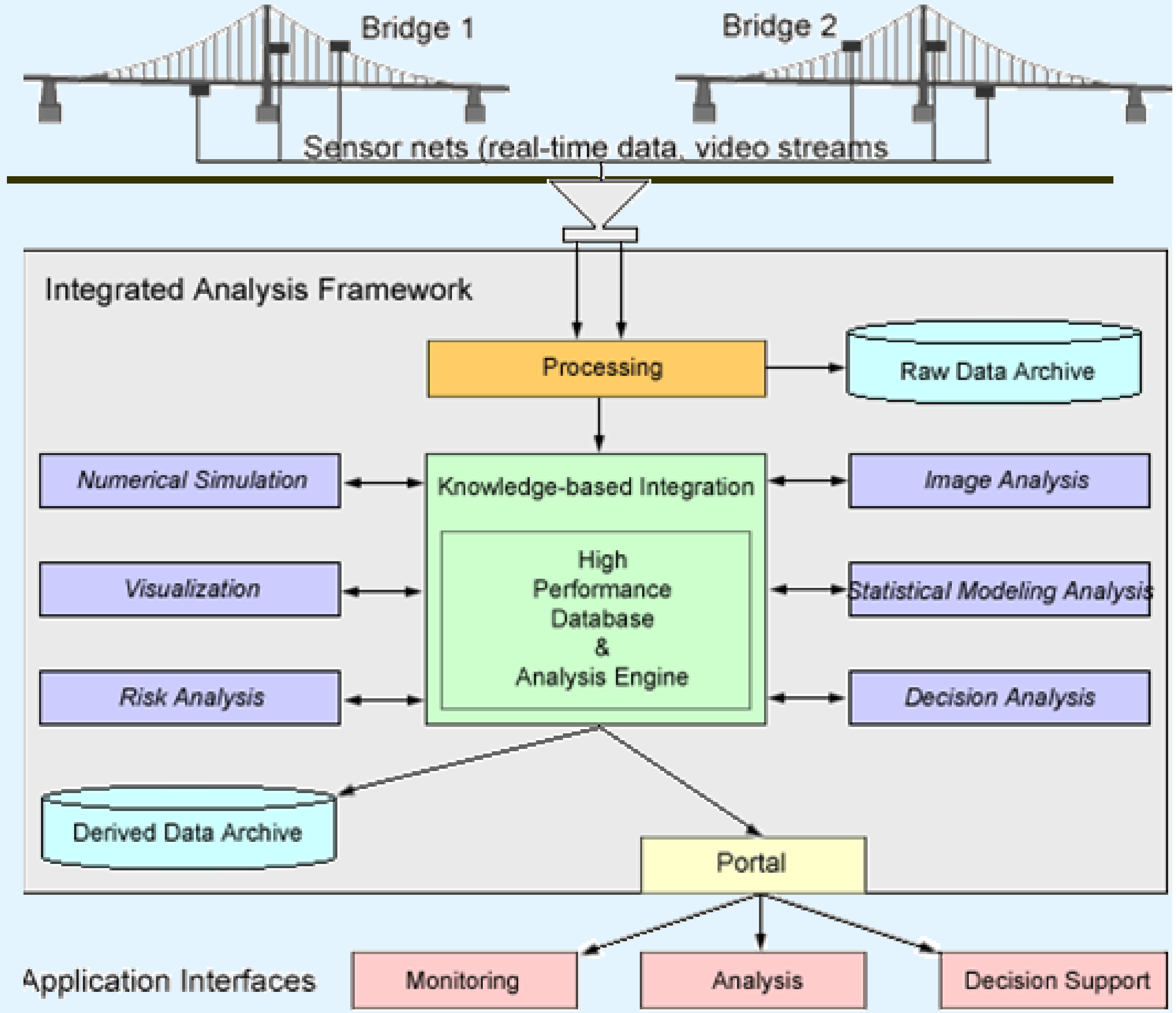
Sensing Diversity



Health Monitoring of Highway Bridges and Civil Infrastructure

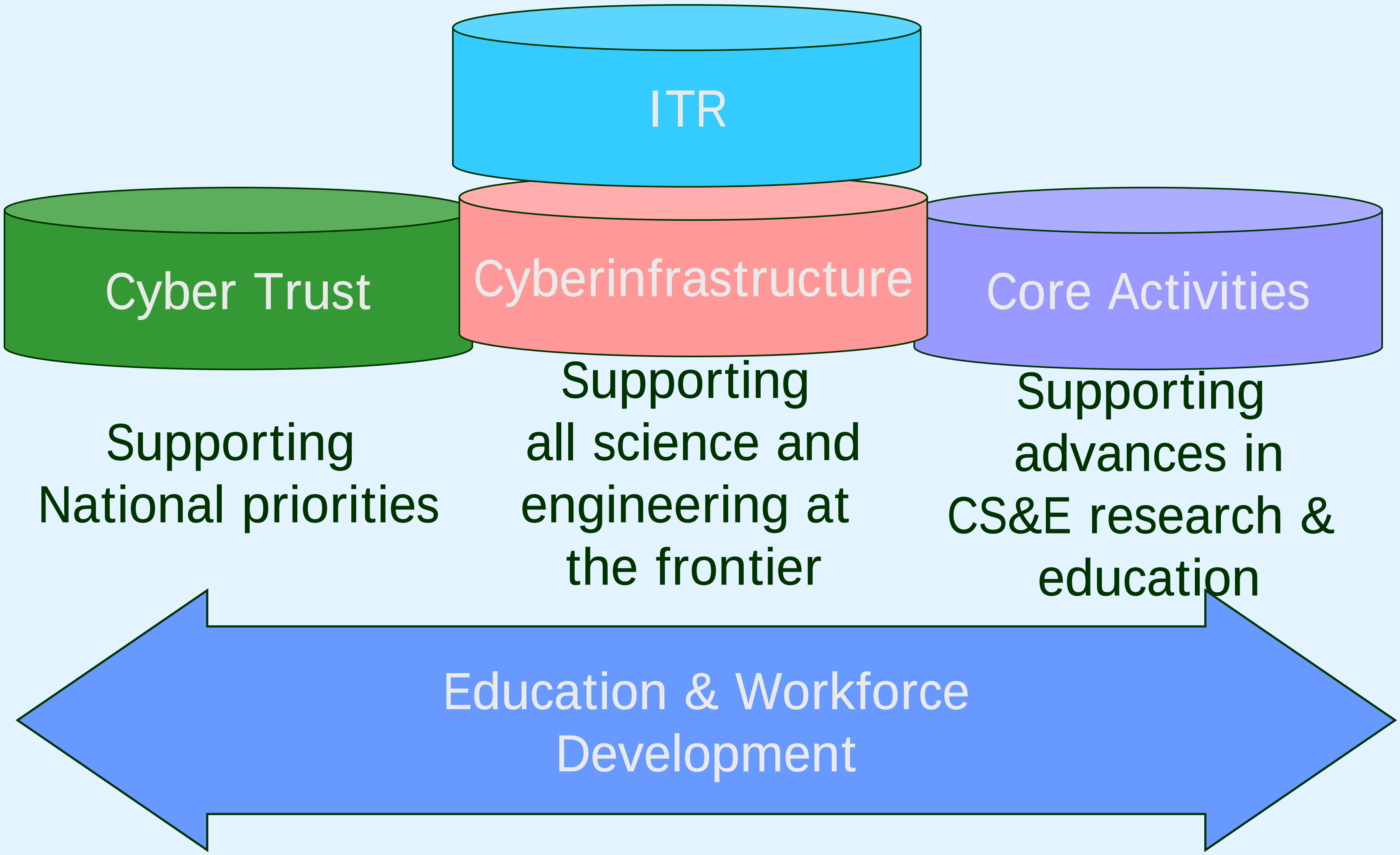
A. Elgamel
University of California, San Diego

- Novel health monitoring strategies for Highway Bridges and Constructed Facilities
- To develop a next-generation, versatile, efficient, and practical health monitoring strategy
- To detect and assess the level of damage to the civil infrastructure due to severe loading, caused by natural loads or man-made events, and/or progressive environmental deterioration.



FY 2004 Priorities

CISE Priorities in FY 2004



Cyber Trust

- Vision: A society in which
  - Computing systems operate securely and reliably
  - Computing systems protect sensitive information
  - Systems are developed and operated by a well-trained and diverse workforce
- Congressional mandate to build national capacity in cybersecurity research and development
  - Cybersecurity Research & Development Act (Public Law 107-305)
- Research on foundations, network security, systems software, and information systems
- Integrated education and workforce activities
- Program announcement to be released in the fall

Core CISE Priorities (under development)

- Science of Design
  - Software is component of computing technology that is enabling… or not. Software engineering tools, techniques, processes insufficient to task ahead
  - Develop better software “building blocks” based on body of scientific knowledge and coherent, systemized approach – science of design
- Information Integration
  - Exploit available science and engineering data to create new knowledge
  - Tools and techniques to efficiently access disparate information systems in real-time in response to unanticipated information needs…. unifying data models, data fusion algorithms, data federation strategies ….

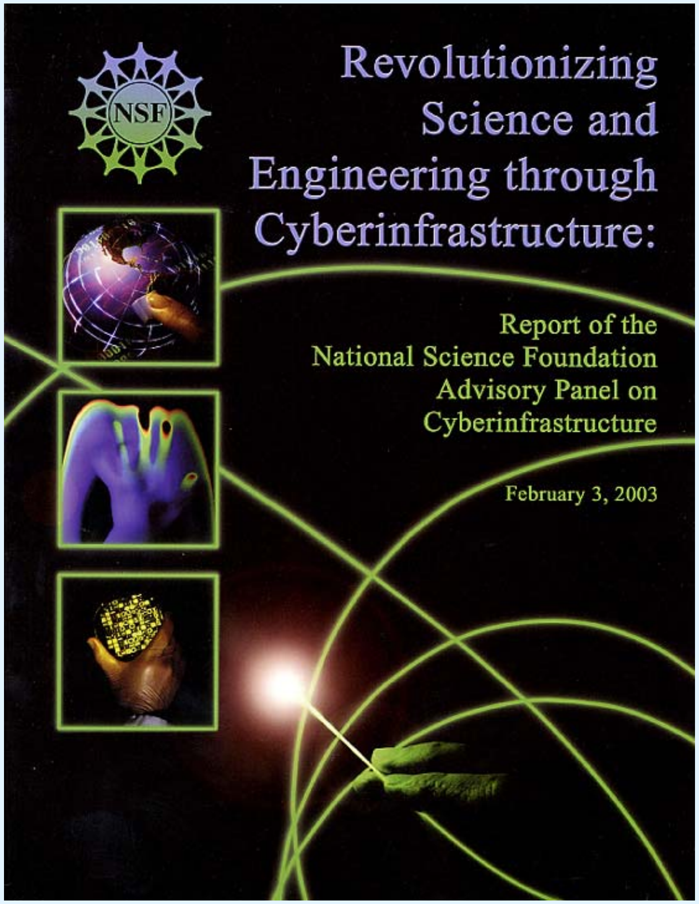
Education and Workforce

- Goal
  - Focused integration of education and workforce development with research
- New Announcements:
  - CRCD/EI (late fall)
  - Workforce program: ITWF (late fall)
  - Increasing synergy with other CISE activities
- Fiscal Year 05
  - One program with education, workforce, and integration

ITR in 2004

- Final year as a Priority Area
  - Continue innovative, high-risk, interdisciplinary IT research
  - Build on success and transition to the future
- Overarching vision
  - Explosion of connectivity
    - Enables activities never before possible
    - Increases complexity and interdependency
  - Increased importance when deployed for national priority
- Focus on National Priorities
  - Advances in science and engineering
  - Economic prosperity and vibrant civil society
  - National and homeland security

Cyberinfrastructure: The Atkins Report



Daniel E. Atkins, Chair
University of Michigan
Kelvin K. Droegemeier
University of Oklahoma
Stuart I. Feldman
IBM
Hector Garcia-Molina
Stanford University
Michael L. Klein
University of Pennsylvania
David G. Messerschmitt
University of California at Berkeley
Paul Messina
California Institute of Technology
Jeremiah P. Ostriker
Princeton University
Margaret H. Wright
New York University

http://www.cise.nsf.gov/evnt/reports/toc.htm

"[Science is] a series of peaceful interludes punctuated by intellectually violent revolutions . . .[in which] . . . one conceptual world view is replaced by another."

--Thomas Kuhn

From *The Structure of Scientific Revolutions*

NSF’s Cyberinfrastructure Objective

- To lead the country in providing an integrated, system of computing, data facilities, connectivity, software, services, and instruments that ...

- enables all scientists and engineers to work in new ways in the advancement of discovery, learning and innovation…

- at scale, across disciplinary, organizational and national boundaries, at higher levels of inclusion, investigating new methods, …

Is There a Definition of Cyberinfrastructure (CI)?

- Not really - means different things to different groups - but there are commonalities
- Literally, infrastructure composed of “cyber” elements
- Includes High-End Computing (HEC, or supercomputing), grid computing, distributed computing, etc. etc.

A working definition:

- *an integrated system of interconnected computation/communication/information elements that supports a range of applications*

Note: There is an **extant CI** today. What we are really talking about is an **emergent CI**.

Cyberinfrastructure is the means;
"e-Science" is the result

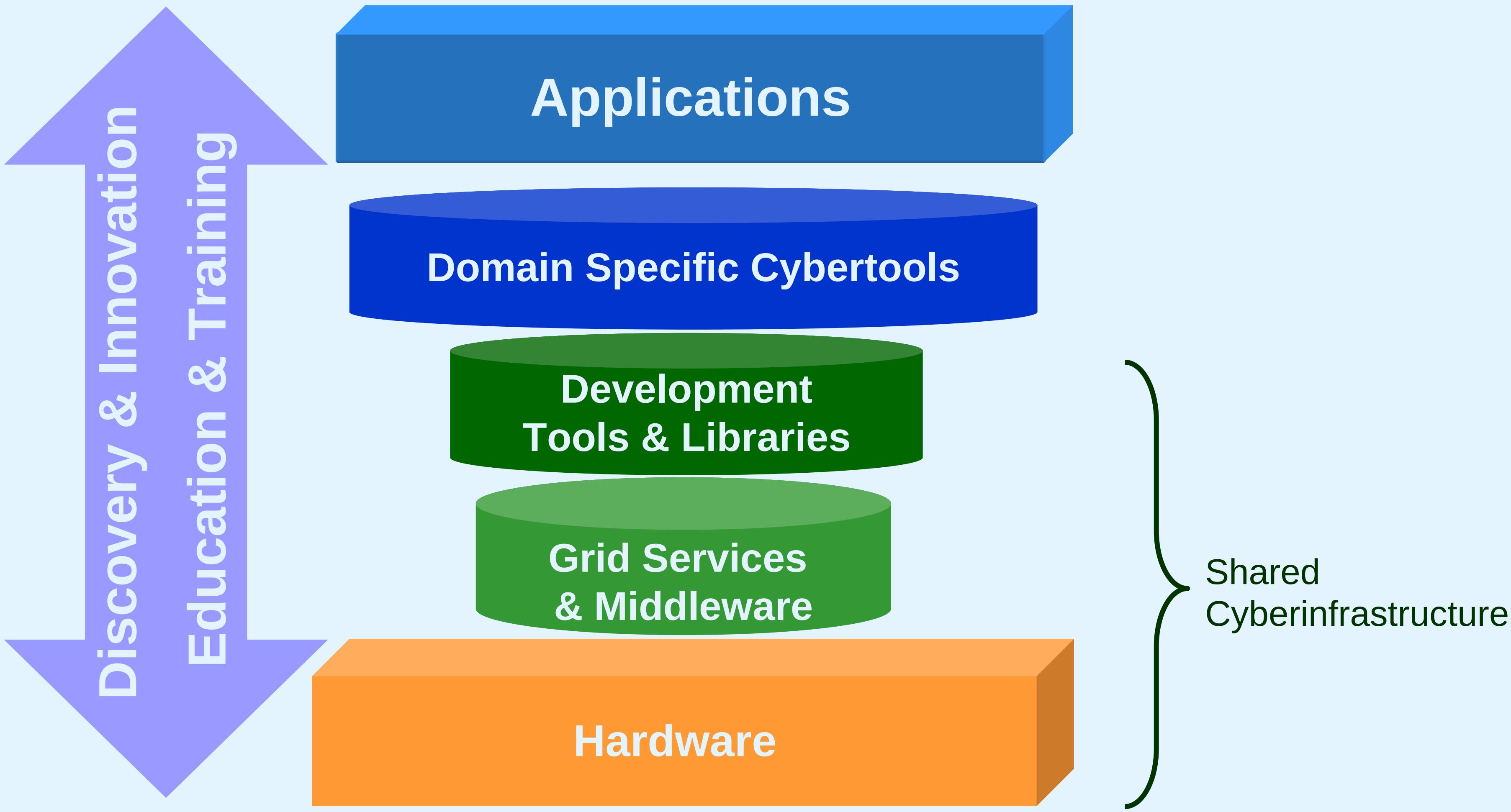
Cyberinfrastructure consists of …

- Computational engines (supercomputers, clusters, workstations, small processors, …)
- Mass storage (disk drives, tapes, …)
- Networking (including wireless, distributed, ubiquitous)
- Digital libraries/data bases
- Sensors/effectors
- Software (operating systems, middleware, domain specific tools/platforms for building applications)
- Services (education, training, consulting, user assistance)

All working together in an integrated fashion.

Integrated Cyberinfrastructure…

meeting the needs of a community of communities



In Ten Years, a CI That Is…

- rich in resources, comprehensive in functionality, and ubiquitous;
- easily usable by all scientists and engineers, from students to emertii;
- accessible anywhere, anytime needed by authenticated users;
- interoperable, extendable, flexible, tailorable, and robust;
- funded by multiple agencies, states, campuses, and organizations;
- supported and utilized by educational programs at all levels.

Some Characteristics of a Future Cyberinfrastructure

- Built on broadly accessible, highly capable network: 10’s of terabits backbones down to intermittent, wireless connectivity at very low speeds;
- Contains significant and varied computing resources: 100’s of petaflops at high end, with capacity to support most scientific work;
- Contains significant storage capacity: exabyte collections common; high-degree of DB confederation possible;
- Allows wide range of sensors/effectors to be connected: sensor nets of millions of elements attached;
- Contains a broad variety of intelligent visualization, search, database, programming and other services that are fitted to specific disciplines

Technical Challenges

- How to build the components?
  - Networks, processors, storage devices, sensors, software
- How to shape the technical architecture?
  - Pervasive, many cyberinfrastructures, constantly evolving/changing capabilities
- How to customize CI to particular S&E domains

Operational Challenges

- Data standards
- General interoperability
- Resource allocation
- Security and privacy
- Training
- Continuous evolution

Funding/Ownership Challenges

- Cooperation among agencies
- Cooperation between federal and state/private levels
- Role of campuses
- Interaction with private industry
- $$$$$!

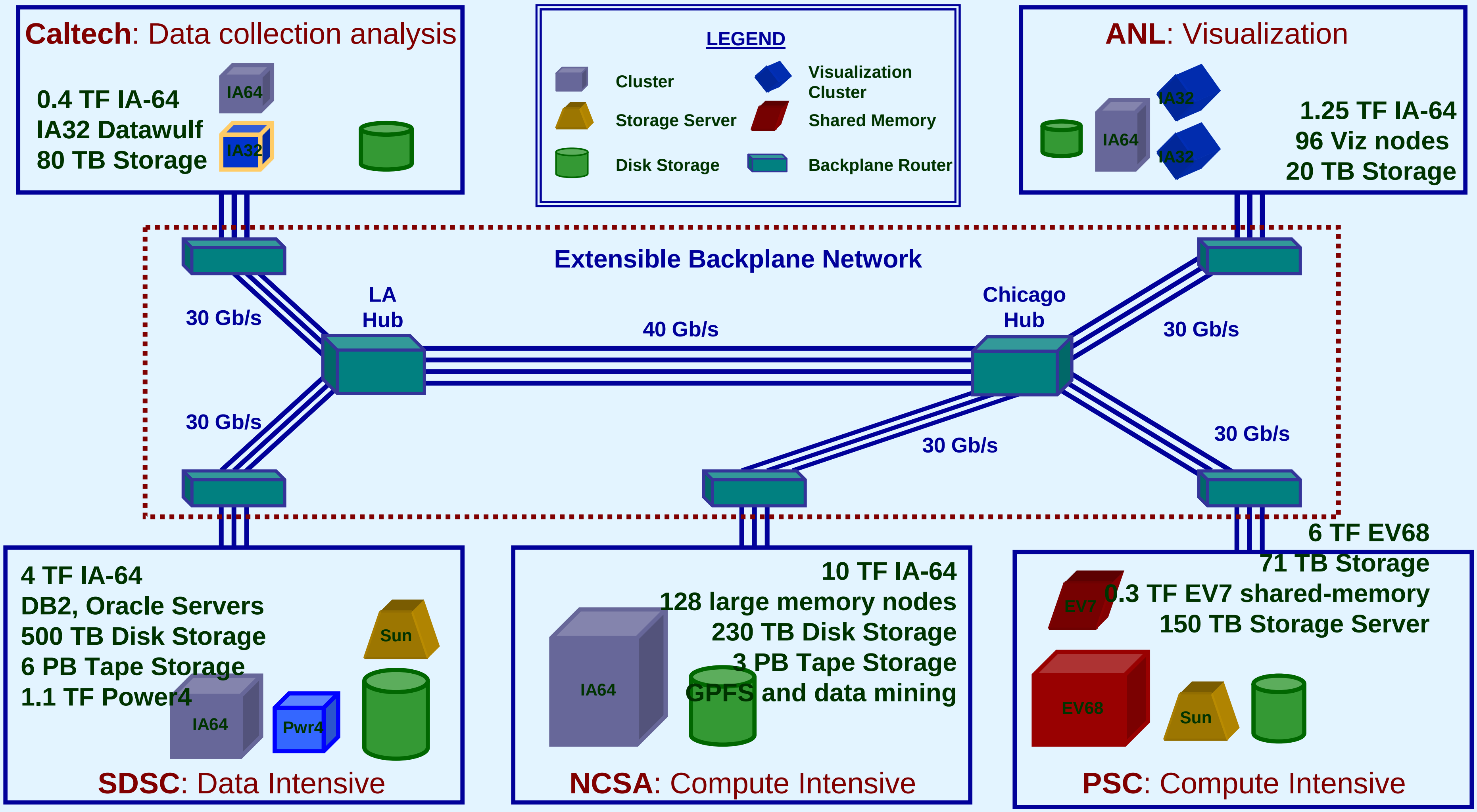
Educational Challenges

- How to make sure that future generations of scientists and engineers can fully utilize CI
  - New paradigms, methods, objectives
- How to retrain current scientists and engineers
- How to make sure that new ideas for extending CI continue to come from those that are using it

Extant Projects

- Partnerships for Advanced Computational Infrastructure (PACI)
- Grid Physics Network (GriPhyN)/ iVGDL
- Network for Earthquake Engineering Simulation (NEES)
- National Ecological Observatory Network (NEON)
- National Science Digital Library (NSDL)
- Bioinformatics Research Network (BIRN)
- Extensible Terascale Facility

Extensible TeraGrid Facility (ETF)



Has the Grid been Grid oversold?

The promise of the Grid has been not been oversold but the difficulty of developing the requisite Grid infrastructure has been underestimated.

- Fran Berman, Director, SDSC

Measures of ETF Success

- Use a single node on TeraGrid
  - Portals, SW, scheduling allows access to designated individual resources
- Use as a wide-area cluster computer
  - Use multiple designated resources of the same type for a single computation
- Use as a simple grid
  - Use multiple resources of different types in a staged or concurrent computation
- Use as a full grid
  - Use multiple nodes as an ensemble via advanced SW environment

Focus on Policy and Social Dynamics

- Policy issues must be considered up front
- Social engineering will be at least as important as software engineering
- Well-defined interfaces will be critical for successful software development
- Application communities will need to participate from the beginning

Required Long-term Investments for Cyberinfrastructure

- Fundamental CS&E research
- Operational cyberinfrastructure
- Shared cybertools (software)
- Domain-specific cybertools
- Education and outreach (workforce)

To support research and education in all science and engineering fields.

Cyberinfrastructure in Transition

Principles:

- Build on what we’ve learned to date
- Provide new funding opportunities for extant and emerging providers and users
- Encourage partnerships between CI users and computing specialists
- Promote flexibility, interoperability and competition for best ideas

Conclusion

- The NSF mission is broad
  - Serve the National well-being
  - by enriching the science and engineering research and education community
- For CISE…
  - Rapid changes in computing and national priorities
  - leads to changes for CISE directorate

I am reminded…

- Blessing:
  May you live in interesting times
- Curse:
  May you live in interesting times

Enabling the nation's future through discovery, learning and innovation

Evolution of the Computational Infrastructure