NSF and Cyberinfrastructure: Background, Vision, and Activities

Daniel E. Atkins

Director, OCI

datkins@nsf.gov

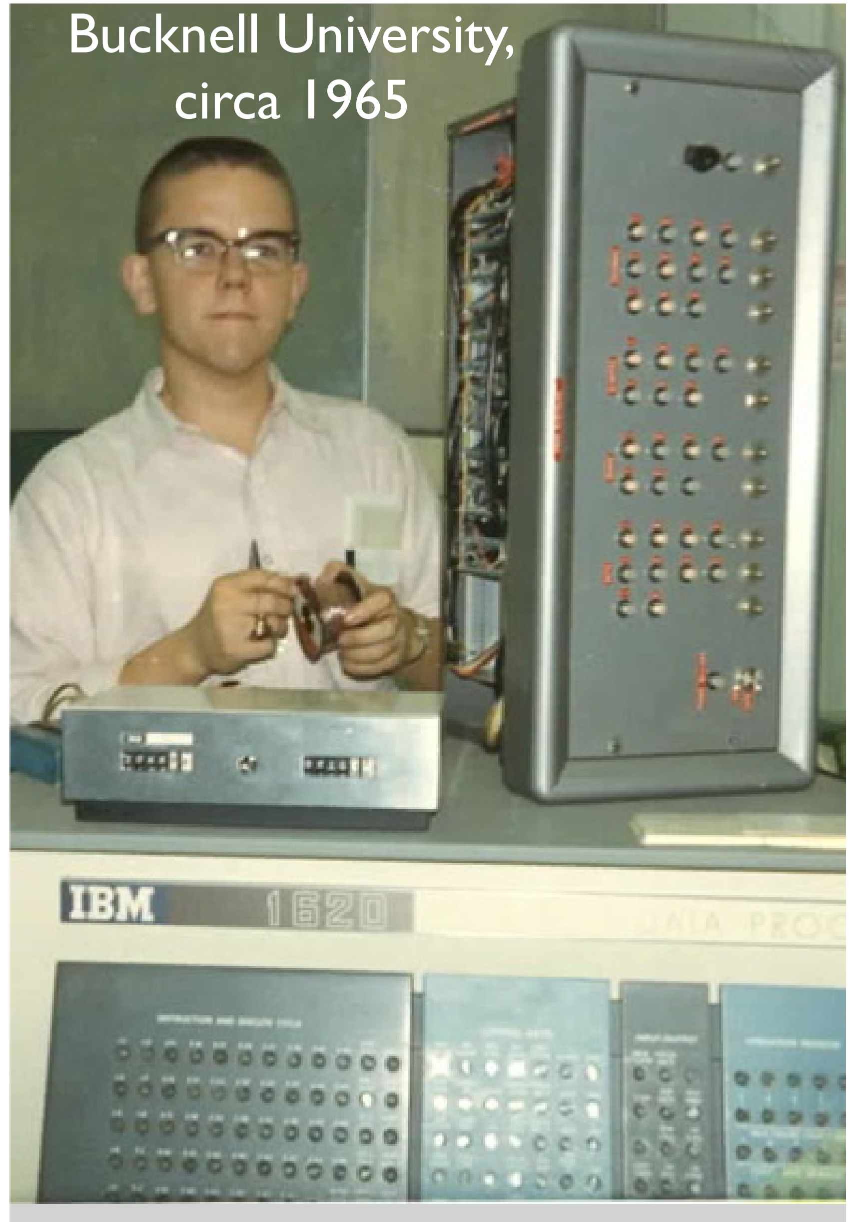
www.nsf.gov/oci

TeraGrid

TeraGrid 2006
June 13, 2006

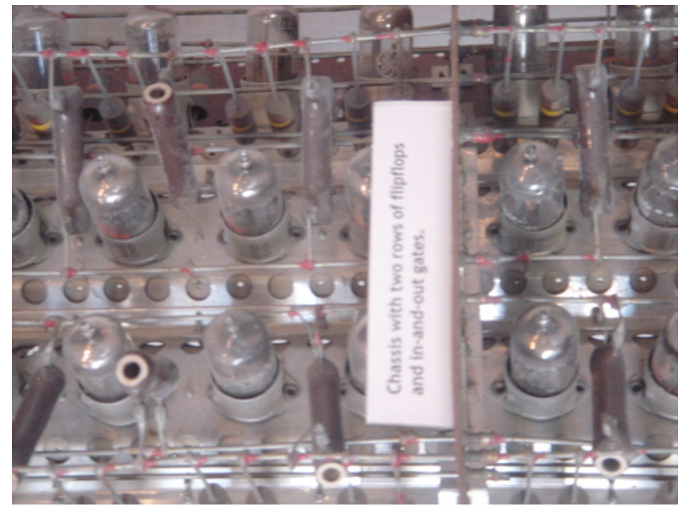
National Science Foundation
www.nsf.gov

Personal Privilege and Amazement: From Digiclock to Petascale Grid Communities



Bucknell University, circa 1965

Illiac I (1949-1962)

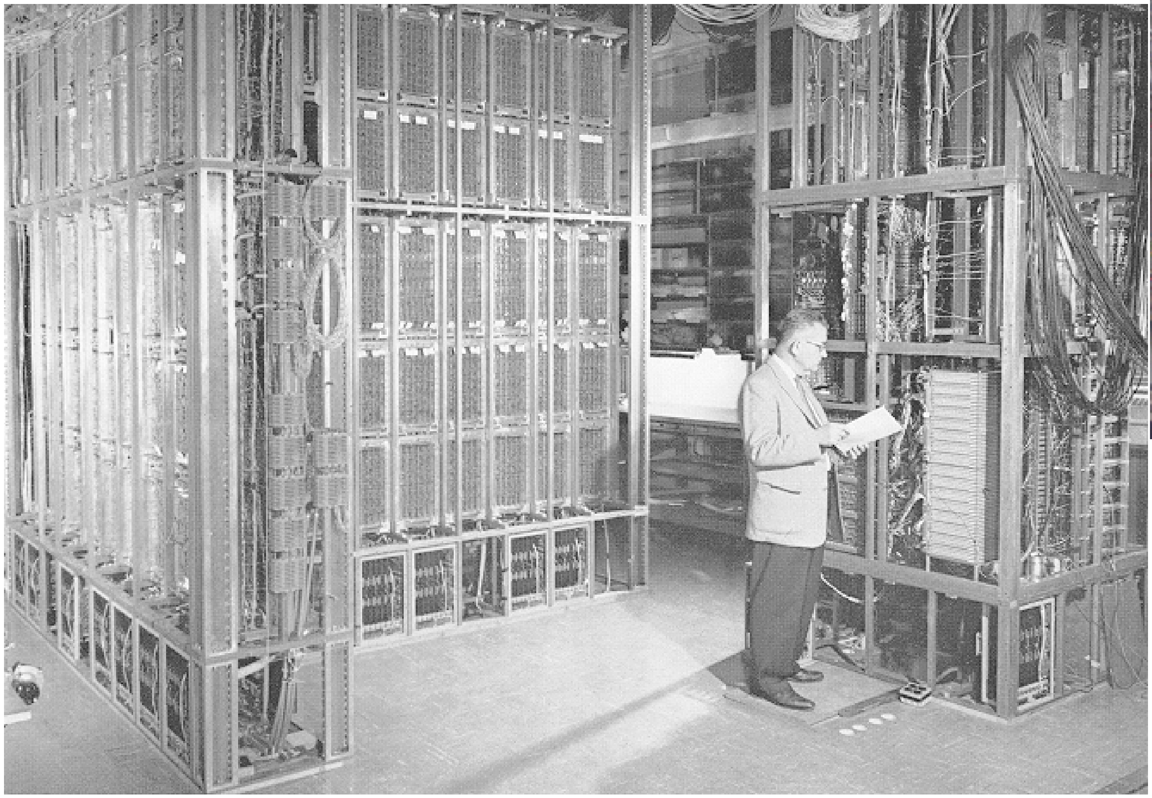


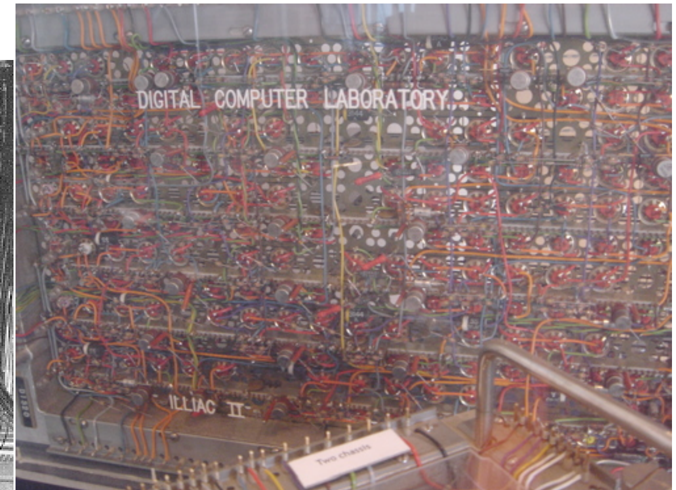
Vacuum tube circuits

Illiac I was the first computer built and owned entirely by a university. It went on line on September 22, 1952. With the computing power of a modern-day handheld calculator, Illiac had 2,800 vacuum tubes and weighed 10,000 pounds. It was more than 10 ft long, 2 ft high, and 8 ft high, with a 5k main memory and 64k Drum memory.

D. E. Atkins
atkins@umich.edu

Illiac 2 (1956-68)





One of the first transistor machines. Each transistor hand tested and cost $80 each

D. E. Atkins
atkins@umich.edu

Illiac III (1960-68)

- The **ILLIAC III** was a fine-grained SIMD pattern recognition computer built by the University of Illinois in 1966.
- This ILLIAC's initial task was image processing of bubble chamber experiments used to detect nuclear particles. Later it was used on biological images.
- The machine was destroyed in a fire, caused by a Variac shorting on one of the wooden-top benches, in 1968. (Including my four floating point AUs.)

From Wikipedia

D. E. Atkins
atkins@umich.edu

Illiac IV (1965-82)

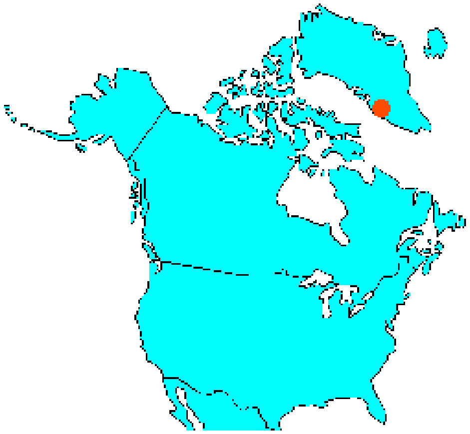
ILLIAC IV at the Institute for Advanced Computation, Moffett Field, California

Vietnam Era, highly parallel GigaFLOPS machine

D. E. Atkins
atkins@umich.edu

The University of Michigan Upper Atmospheric Research Collaboratory (UARC)

The Initial Facility at Sondrestrom, Greenland



D. E. Atkins
atkins@umich.edu

UARC Interface

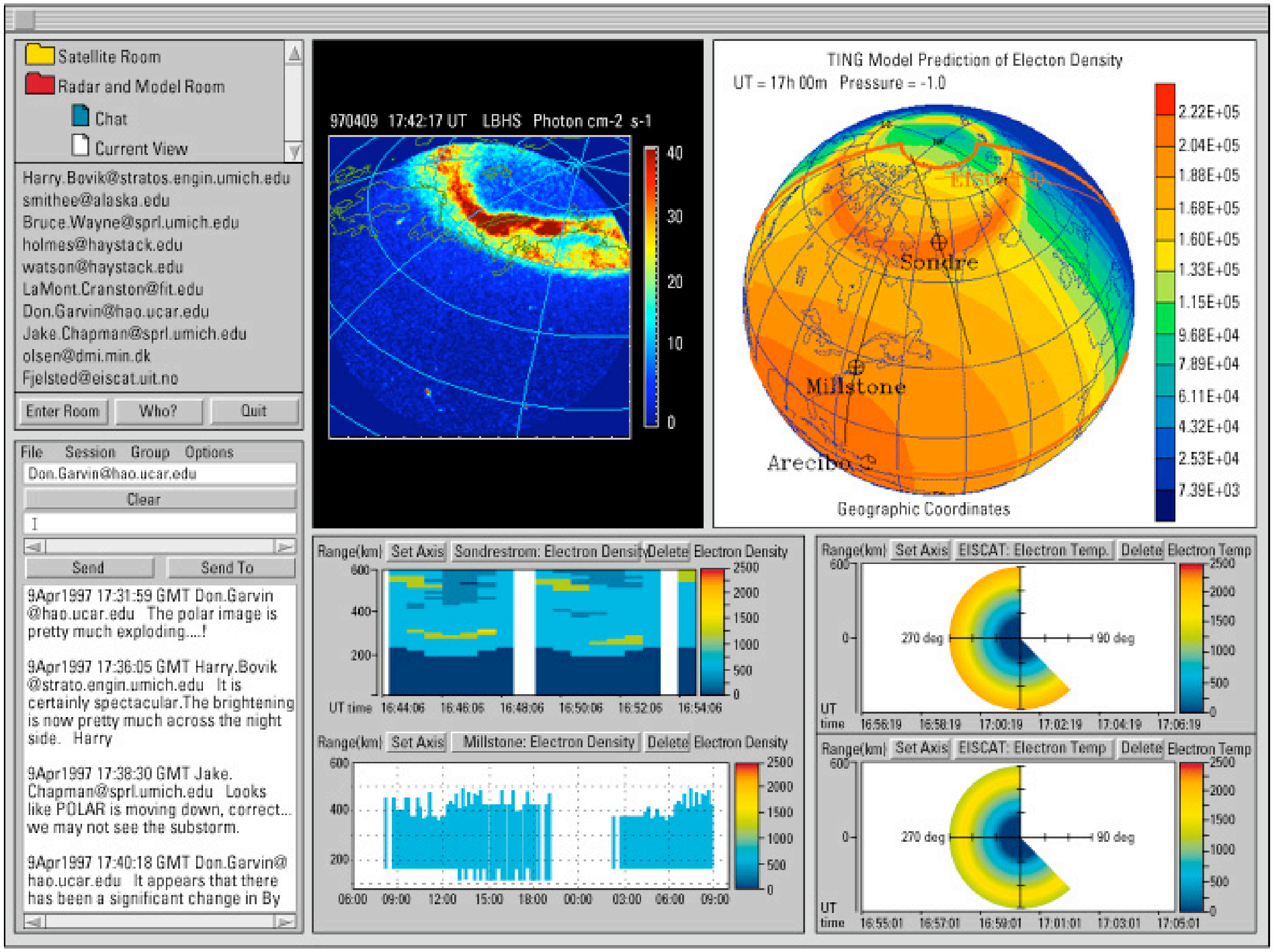
Real-time instruments   *computational models*

dynamic work rooms

team chat

annotation

Session replay



Archival data   *Journals*

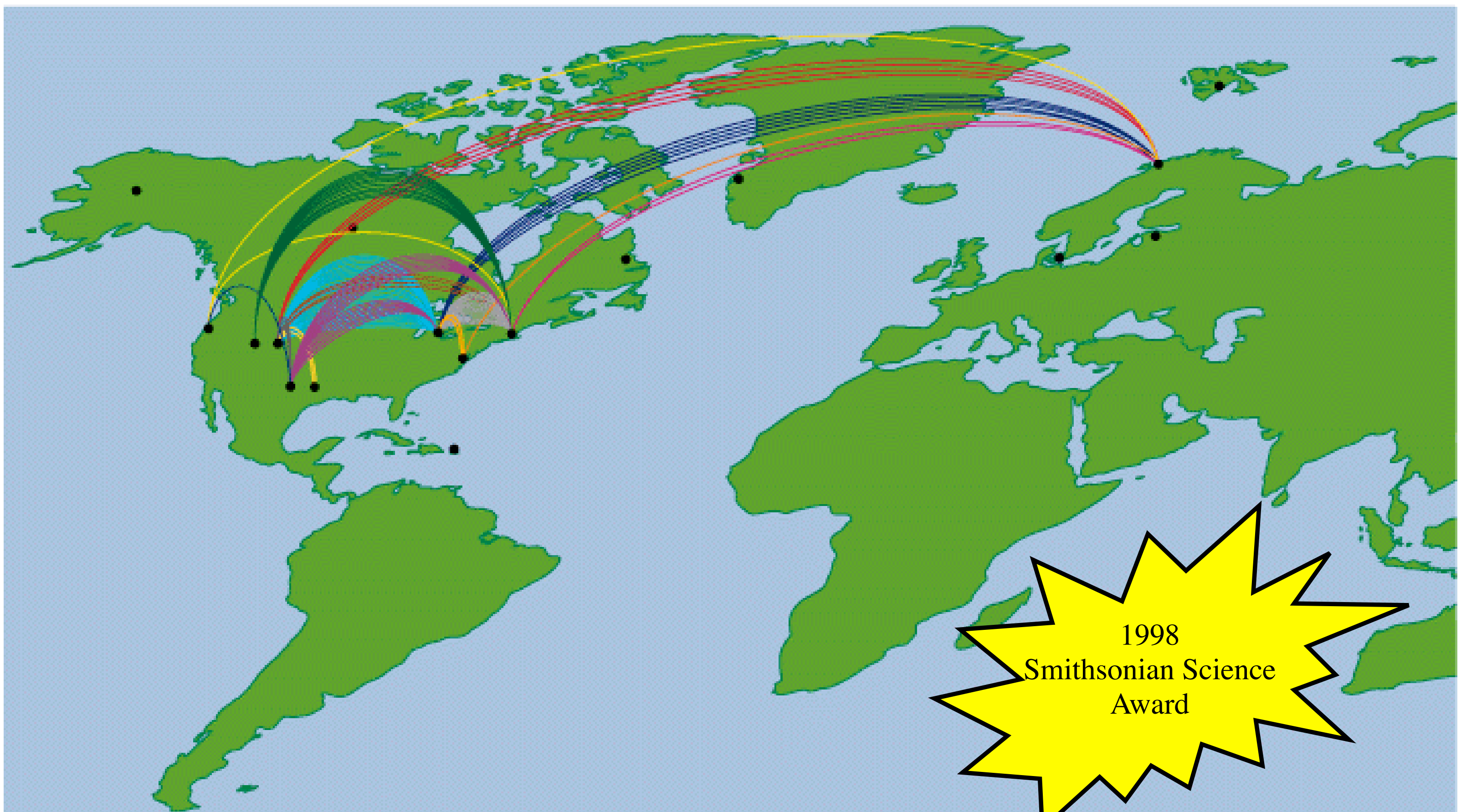
D. E. Atkins
atkins@umich.edu

Evolved into a Network of Instruments (one global instrument)

D. E. Atkins
atkins@umich.edu

UARC Patterns of Communication

Pattern of Communication, UARC Campaign, April 9, 1997



1998
Smithsonian Science
Award

D. E. Atkins
atkins@umich.edu

Vignettes from UARC/SPARC

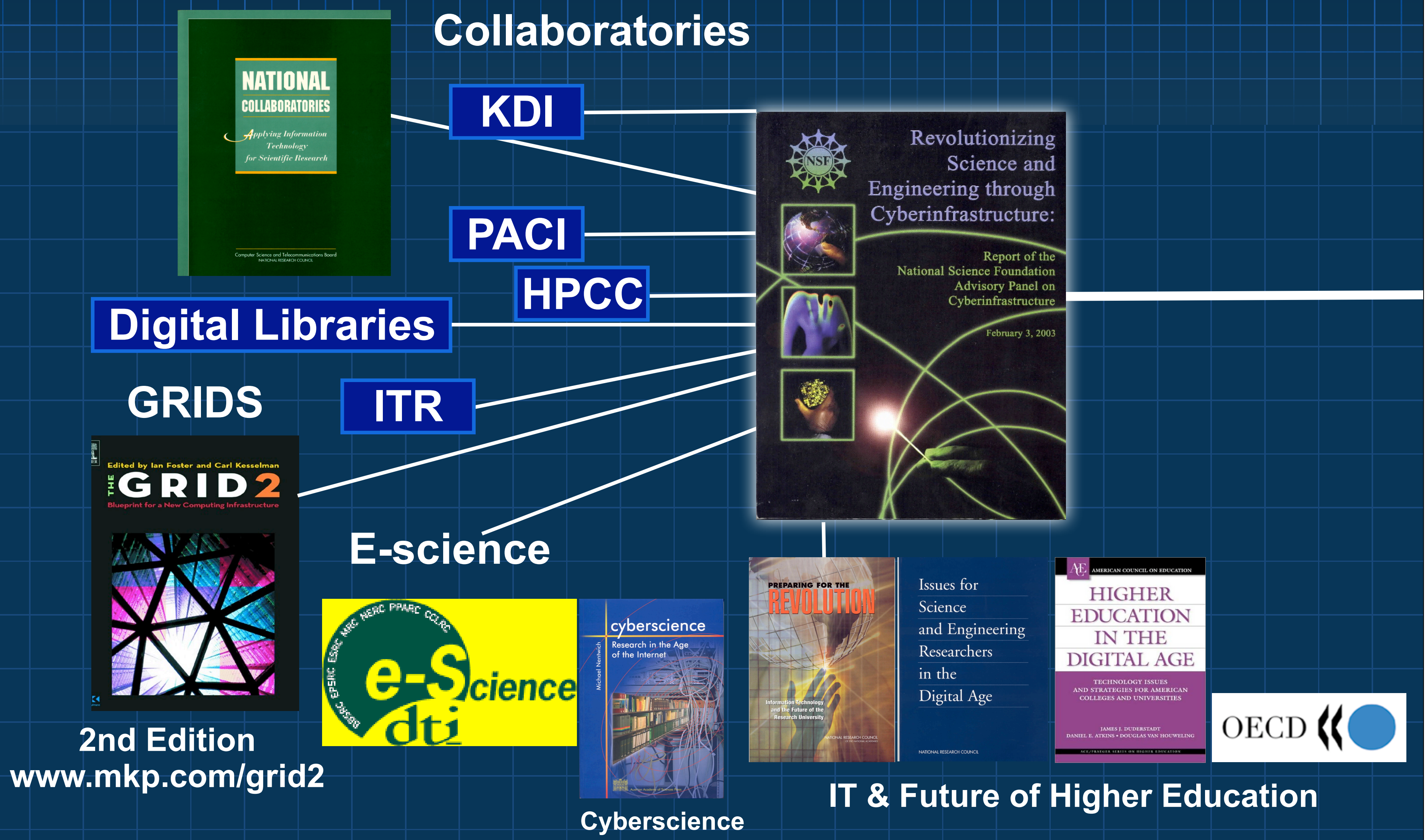
- Shared, tele-instruments & expertise.
- Rapid response, opportunistic campaigns.
- Multi-eyes, complementary expertise.
- Isolated instruments became a global instrument chain.
- Cross-mentoring/training.
- New & earlier opportunities/exposure for grad students.
- Enhanced participation. Legitimate peripheral participation.
- Support for authentic, inquiry-based learning at UG and pre-college level.
- Distributed workshops for post-campaign data analysis.
- Session re-play for delayed participation.
- Data-theory closure.
- A "living specification" to stretch vision of possibilities.

D. E. Atkins
atkins@umich.edu

Background and Call to Action

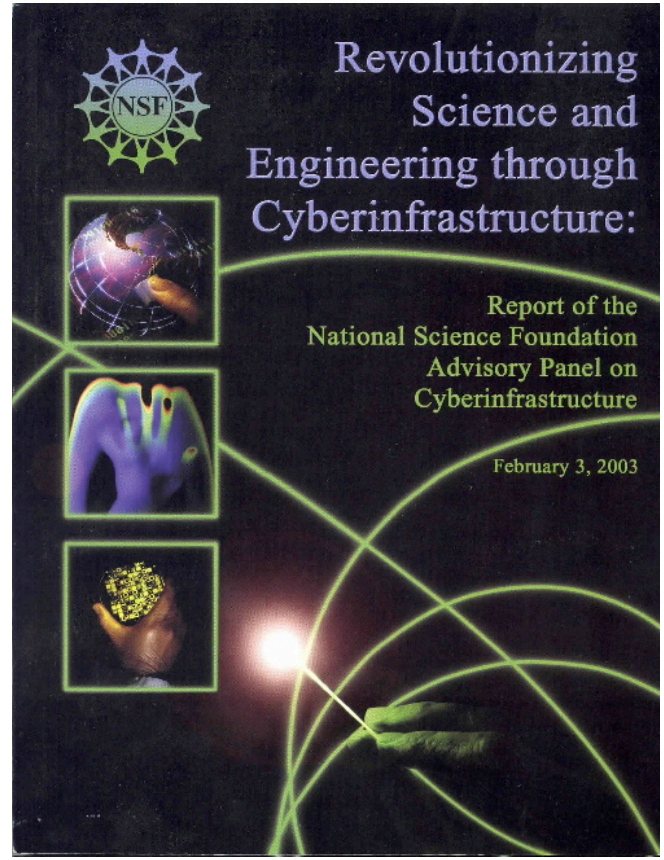
National Science Foundation
www.nsf.gov

CI Genealogy & Movement

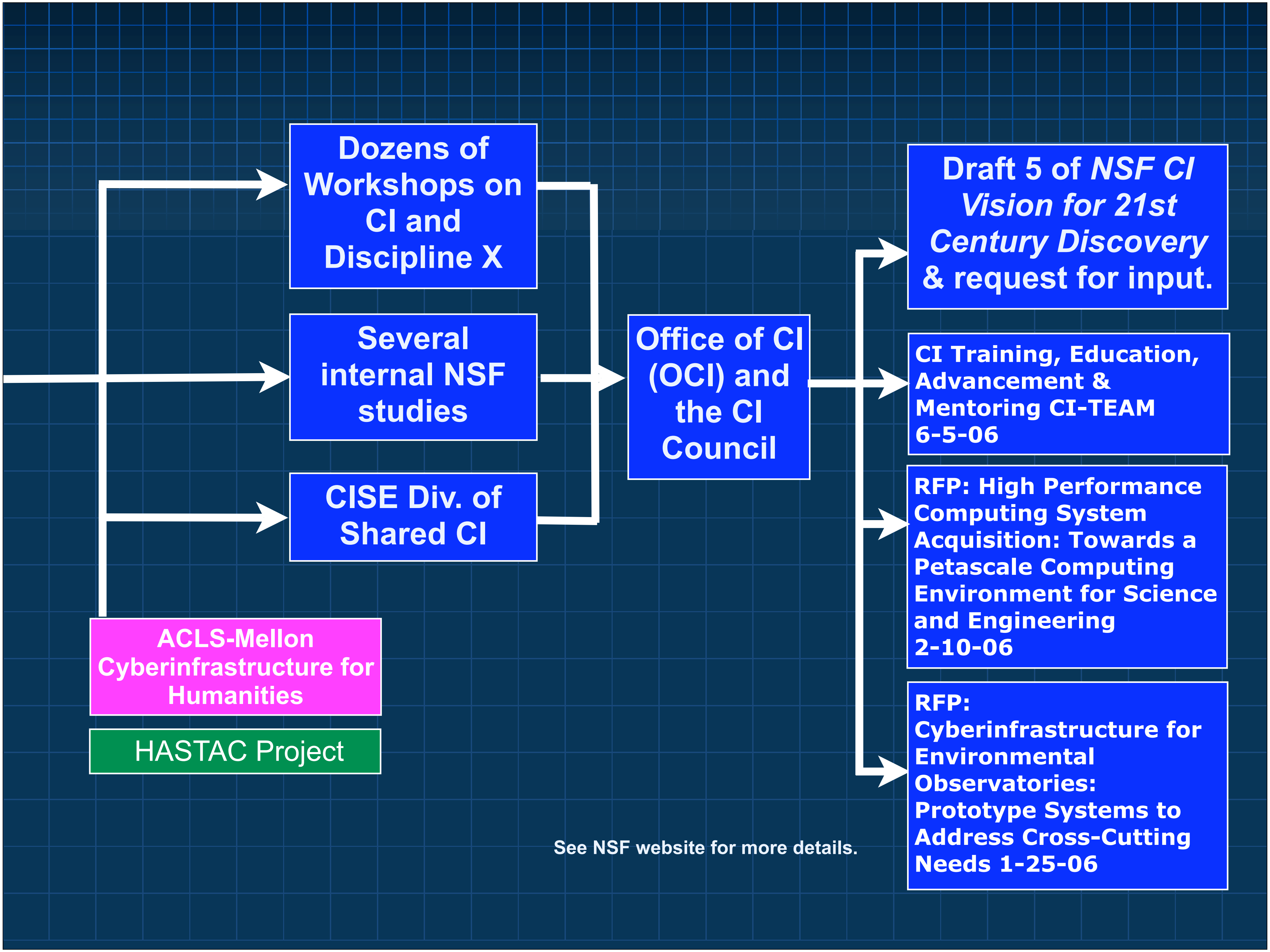


NSF Blue Ribbon Advisory Panel on Cyberinfrastructure

"a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology, and pulled by the expanding complexity, scope, and scale of today's challenges. The capacity of this technology has crossed thresholds that now make possible a comprehensive "cyberinfrastructure" on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy."



http://www.nsf.gov/od/oci/reports/toc.jsp


Dozens of Workshops on CI and Discipline X
Several internal NSF studies
CISE Div. of Shared CI
ACLS-Mellon Cyberinfrastructure for Humanities
HASTAC Project
Office of CI (OCI) and the CI Council
Draft 5 of *NSF CI Vision for 21st Century Discovery* & request for input.
CI Training, Education, Advancement & Mentoring CI-TEAM 6-5-06
RFP: High Performance Computing System Acquisition: Towards a Petascale Computing Environment for Science and Engineering 2-10-06
RFP: Cyberinfrastructure for Environmental Observatories: Prototype Systems to Address Cross-Cutting Needs 1-25-06
See NSF website for more details.

www.nsf.gov/oci/

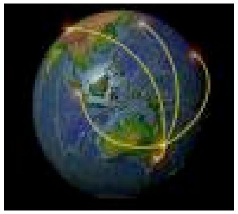
OCI Website - Visit often and provide feedback on the Vision document.

- Scan ACCI Membership List
- See and consider opportunities to serve as a Program Officer

NSF'S CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY

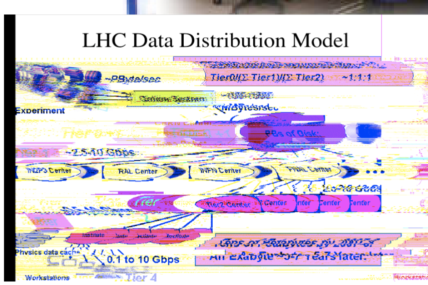
NSF Cyberinfrastructure Council

National Science Foundation
www.nsf.gov

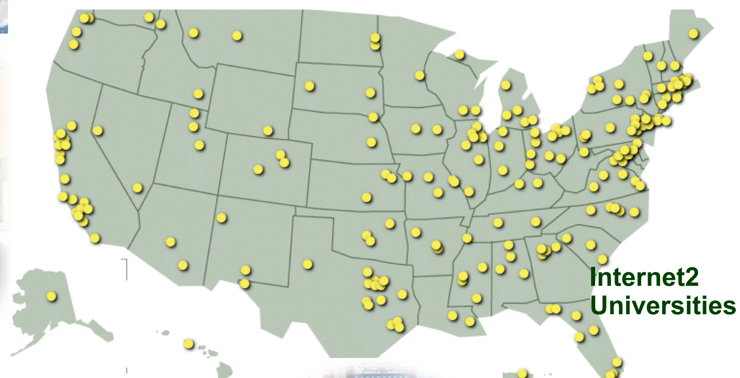


Cyberinfrastructure Vision & Commitment to Leadership

NSF will lead the development and support of a comprehensive cyberinfrastructure essential to 21st century advances in science and engineering.



NSF is the only agency within the U.S. government that funds research and education across all disciplines of science and engineering. ... **Thus, it is strategically placed to leverage, coordinate and transition cyberinfrastructure advances in one field to all fields of research.**



Cyberinfrastructure-enhanced Knowledge Communities (Networks)

Outcomes: New Ideas, New Tools, Education & Career Development, Outreach*

Attributes: Collaborative, Multidisciplinary, Geographically Distributed, Inter-institutional*

Specific Cyber Environments: collaboratories, grids, e-science community, virtual teams, community portal, ...

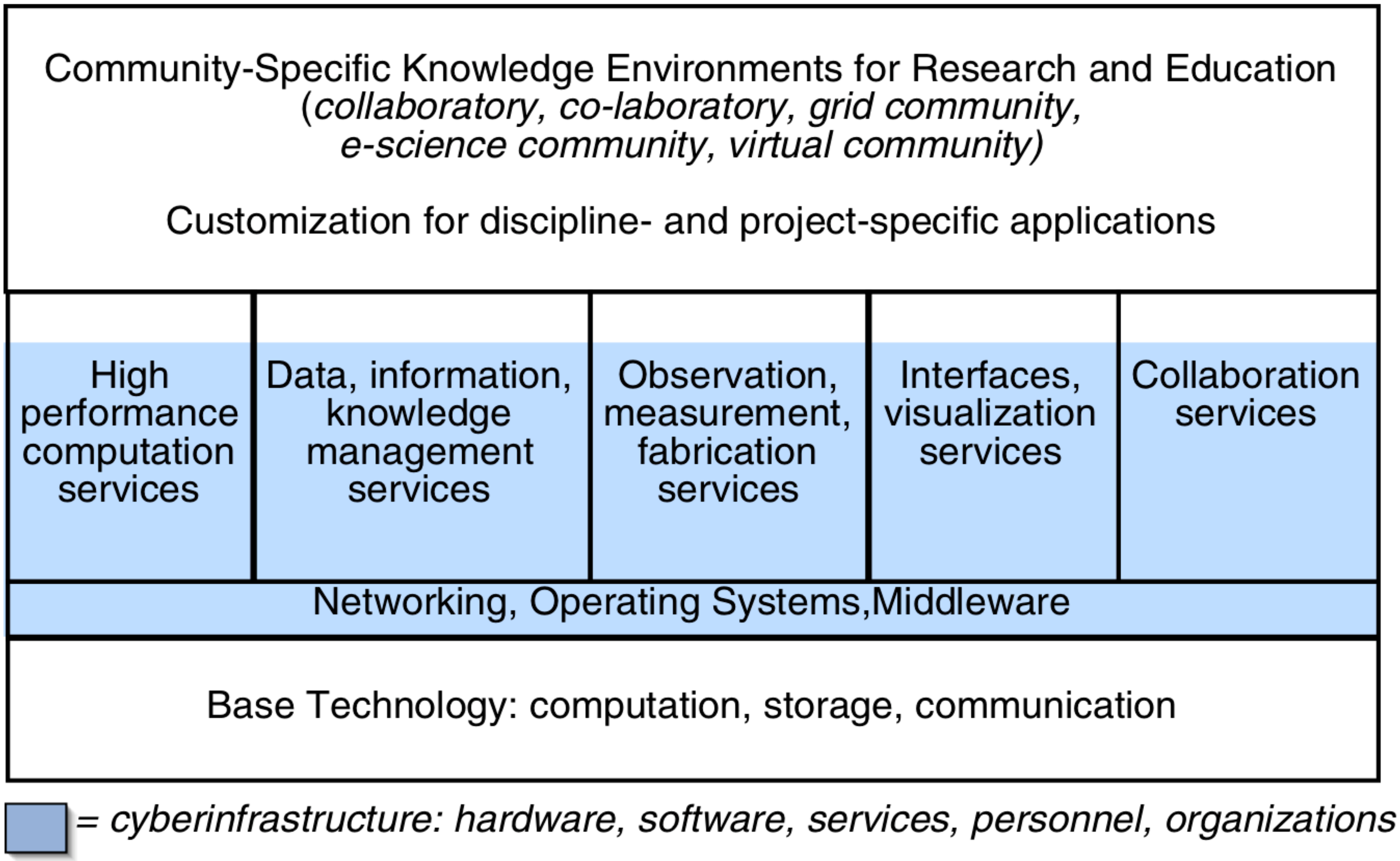
Cyber-infrastructure Services
Equipment, Software, People, Institutions

Computation, Storage, Communication and Interface Technologies

→ Broader Application to other disciplines and types of activity.

* From Cummings & Kiesler (2003) report on KDI Initiative: Multidisciplinary scientific collaborations, see http://www.p2design.com/papers/kdi.pdf

From CI Advisory Panel Report



towards functionally complete CKCs

D. E. Atkins
atkins@umich.edu

From NSF Cyberinfrastructure Vision for 21st Century Discovery

4. Education and Workforce

3. Collaboratories, observatories, virtual organizations

"sophisticated" science application software

1. Distributed, scalable up to petaFLOPS HPC

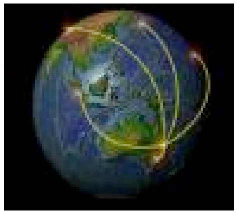
includes networking, middleware, systems software?

2. Data, data analysis, visualization

includes data to and from instruments?

- *provide sustainable and evolving CI that is secure, efficient, reliable, accessible, usable, and interoperable*
- *provide access to world-class tools and services*

Spiral Design*



NSB

Community
Input

Final version to be
released Summer 2006

NSF Directorates & Offices

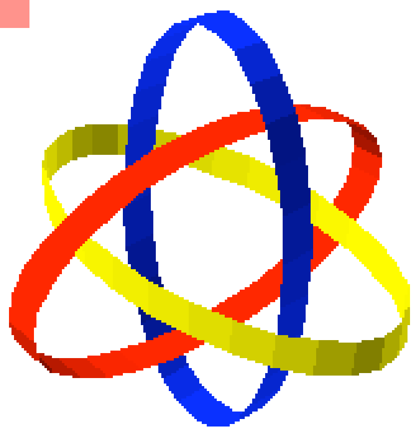
*of both CI activities & OCI role and structure

Borromean Ring Synergy

OCI as a broker of informed mutual self-interest

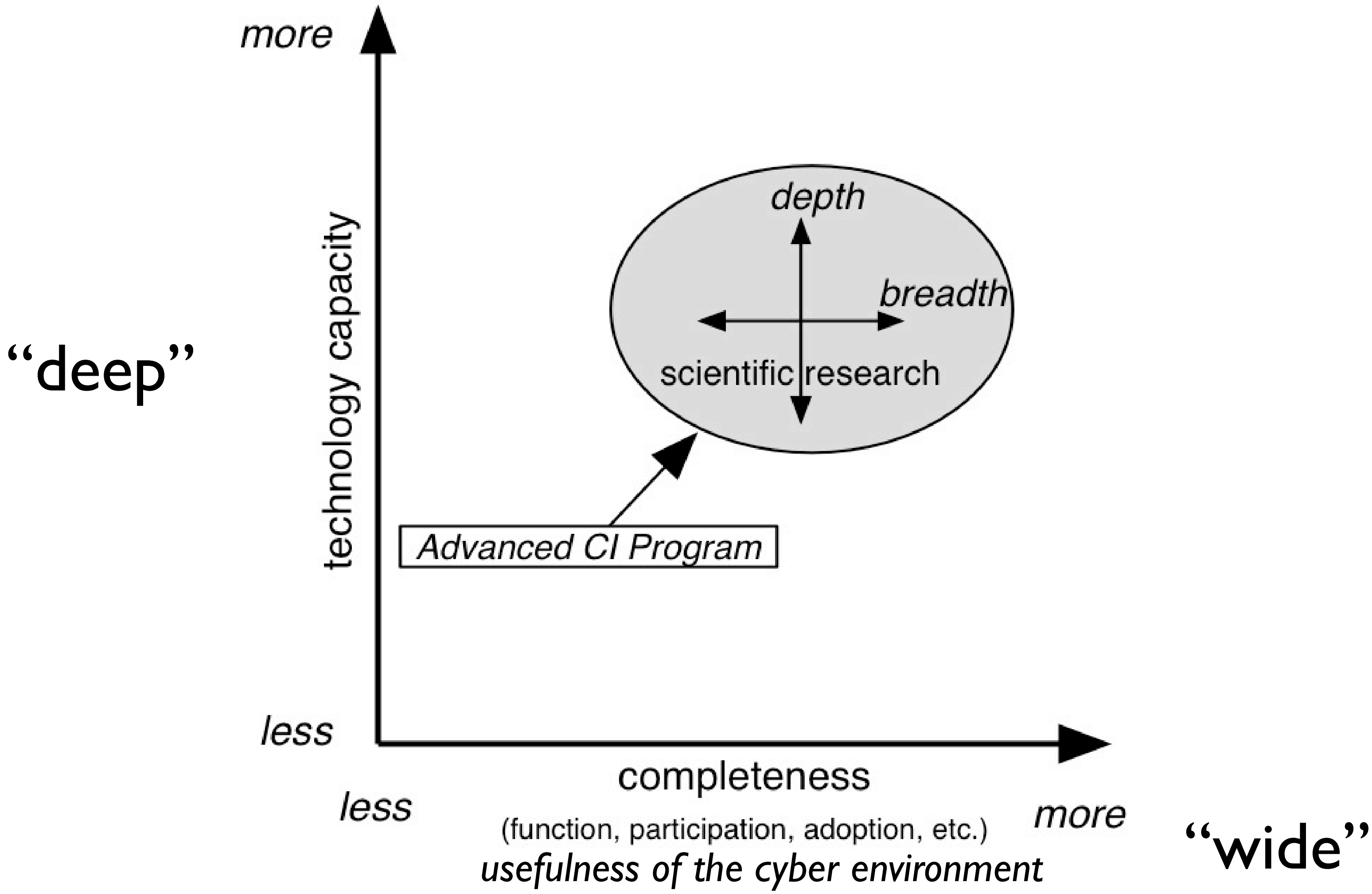
Provisioning -Effective creation, deployment and operation of advanced CI

Application -Innovative use to enhance discovery & learning

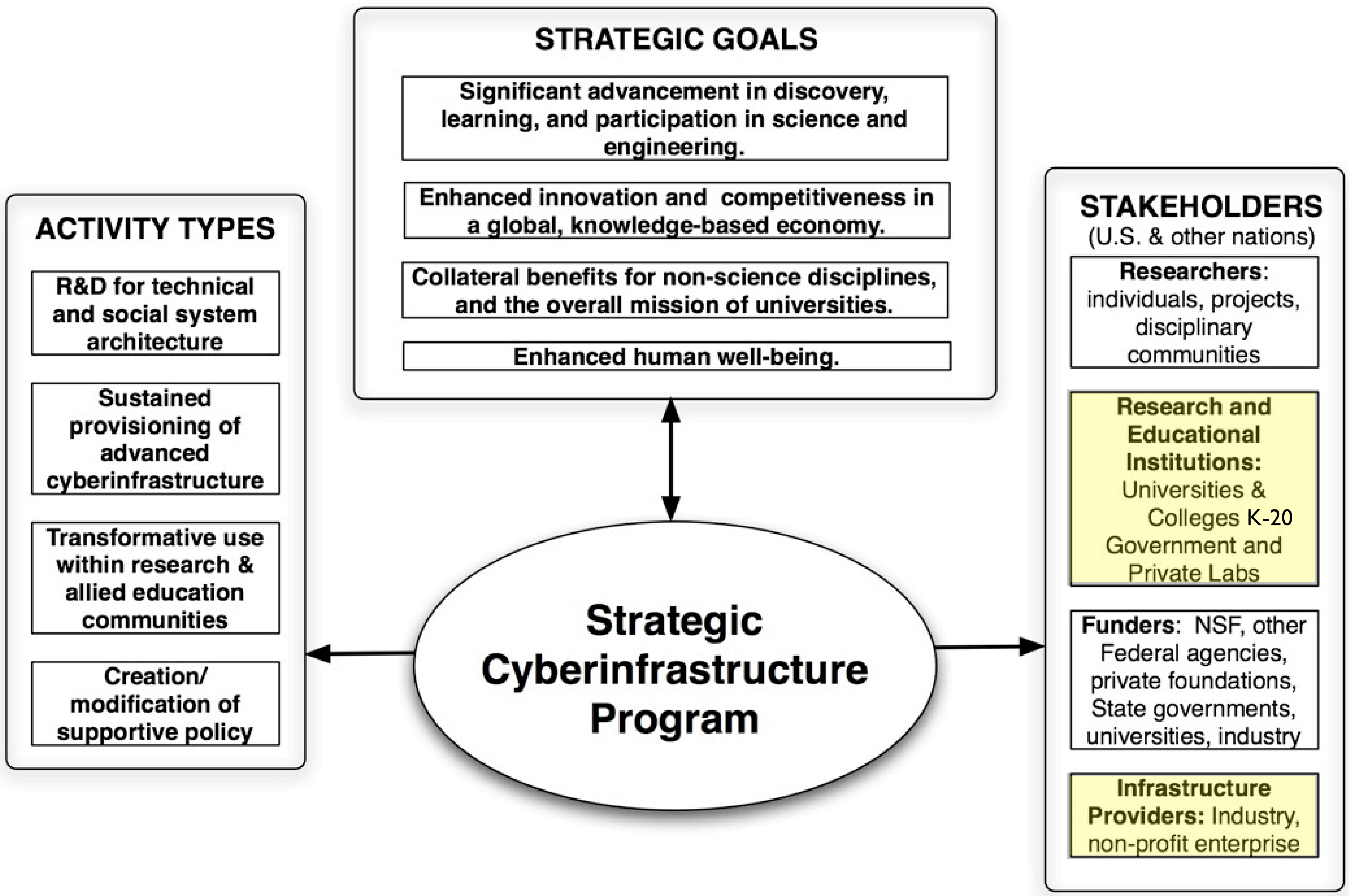


R&D to enhance technology and technical and social dimensions of future CI systems

Balanced investment in enhancement of technology capacity and "completeness" of function and use of CI environments



Alignment of stakeholders towards achieving strategic goals

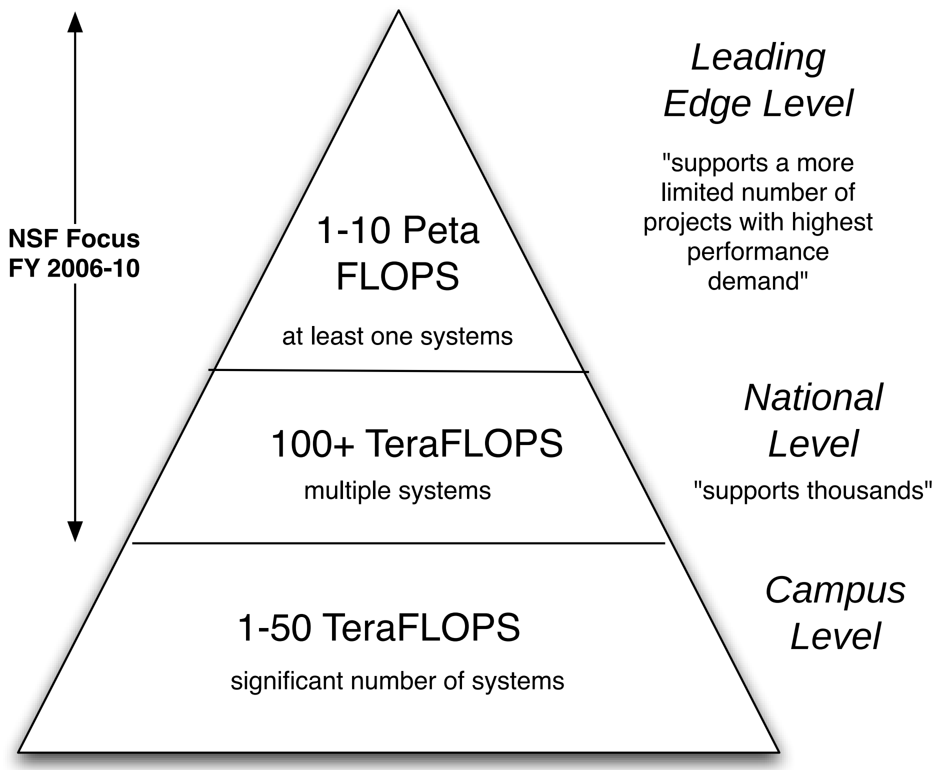


Attributes of NSF HPC Strategy

- Science driven
  - representatives of science users in solicitation and review
  - science based benchmarks
- Three tracks (leading edge, national, campus)
  - shared investment
  - support for capacity and capability
- Connected, grid
- Architectural diversity
- Open competition
- Cooperation with other agencies
- Part of a larger CI vision

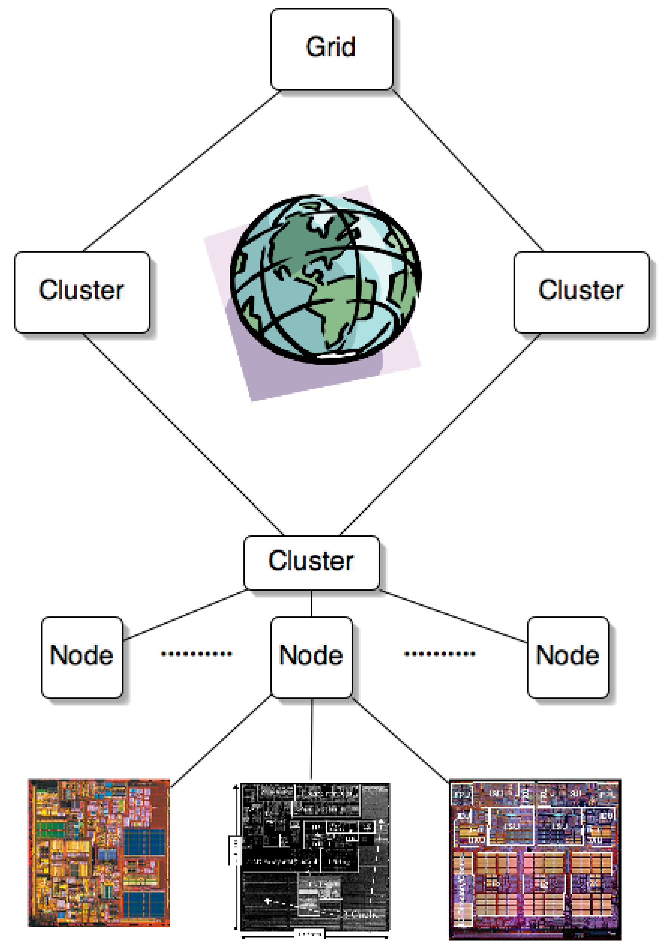
But in an unbundled funding model. i.e. not the PACI core model.

HPC Strategy



On demand & dynamic resource allocation.
HPC in a relevant-time loop.

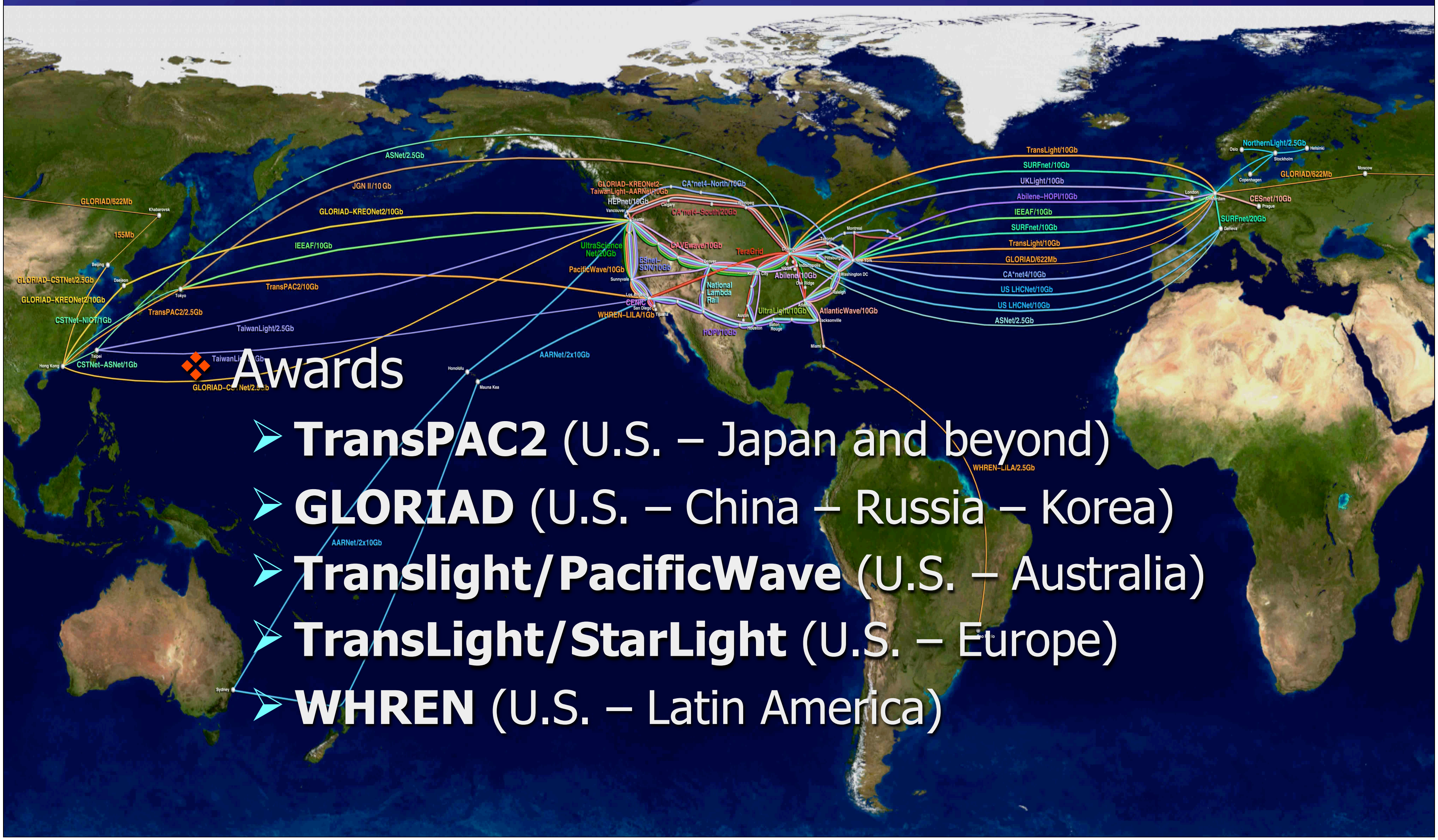
System Diversity



Single chip processors with various internal architectures

D. E. Atkins
atkins@umich.edu

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)

Current International Partners – Internet2 View

Europe

- ARNES (Slovenia)
- BELNET (Belgium)
- CARNET (Croatia)
- CESnet (Czech Republic)
- DANTE (Europe)
- DFN-Verein (Germany)
- FCCN (Portugal)
- GARR (Italy)
- GIP- RENATER (France)
- GRNET (Greece)
- HEAnet (Ireland)
- HUNGARNET (Hungary)
- NORDUnet (Nordic Countries)
- PSNC/PIONER (Poland)
- RedIRIS (Spain)
- RESTENA (Luxembourg)
- RIPN (Russia)
- SANET (Slovakia)
- Stichting SURF (Netherlands)
- SWITCH (Switzerland)
- TERENA (Europe)
- JISC, UKERNA (United Kingdom)

Africa

- MCIT [EUN/ENSTIN] (Egypt)
- TENET (South Africa)

Middle East

- Israel-IUCC (Israel)
- Qatar Foundation (Qatar)

South Asia

- ERNET/CDAC (India)

Asia-Pacific

- AAIREP (Australia)
- APAN (Asia-Pacific)
- ANF (Korea)
- CERNET/CSTNET/ NSFCNET (China)
- ERNET/CDAC (India)
- JAIRC (Japan)
- JUCC (Hong Kong)
- NECTEC/UNINET (Thailand)
- NGI-NZ (New Zealand)
- SingAREN (Singapore)
- TANet2 (Taiwan)

Americas

- CANARIE (Canada)
- CEDIA (Ecuador)
- CLARA (Latin America & Caribbean)
- CNTI (Venezuela)
- CR2NET (Costa Rica)
- CUDI (Mexico)
- REUNA (Chile)
- RETINA (Argentina)
- RNP [FAPESP] (Brazil)
- SENACYT (Panama)

http://international.internet2.edu

Data, Data Analysis and Visualization

- The CI Advisory Panel heard as much about data needs as about computation. Driven by increasing multi-discipline community simulations/modeling and increased power of data mining and visualization.
- Heard "curation" often.
- Heard avoid "data mortuaries".

D. E. Atkins
atkins@umich.edu

Information Services for CKCs

- Online access to complete credentialed, archival literature.
- Stewardship and curation services for enormous collections of scientific data.
- Long-term, federated digital repositories for diverse digital objects as instructional material and works in progress.
- Continuing retrospective digitization, esp. special collections.
- More *continuous* (vs. batch) and *open* forms of scholarly communication.
- Individual and community customized information services.

D. E. Atkins
atkins@umich.edu

Increased Demand for Data Curation Services

- Supercomputer simulations of complex systems require multidisciplinary expertise, computational models and data.
- Increased power of data mining.
- Enormous data streams from smart sensor arrays.
- Data validation and metadata quality enhancement over time.

D. E. Atkins
atkins@umich.edu

Coherent Data CI in Complex Global Context

- Catalyze the development of (a federated, global) system of science and engineering data collections that is open, extensible, evolvable, (and appropriately curated and long-lived.)
  - Challenges of "petascale" and multi tier integration.
  - Substantial policy issues and challenges to find sustainable economic models.
- Support development of a new generation of tools and services to facilitate data mining, integration, analysis, visualization essential to transforming data into knowledge.

D. E. Atkins
atkins@umich.edu

Virtual Organizations

- CI relaxes/reduces constraints of time and distance (geographical, institutional, disciplinary).
- Flattens the world/

D. E. Atkins
atkins@umich.edu

TIME

GEOGRAPHIC PLACE	Same (synchronous)	Different (asynchronous)
Same	P: Physical mtgs. I: Print-on-paper books, journals F: Hands on labs, shops, studios	P: Shared notebook I: Library reserves F: Time-shared labs, shops, studios
Different	P: AV Conference I: Web search F: Online, real time instruments	P: Email I: Knowbots F: Autonomous instruments, session objects

P: people, I: information, F: facilities, instruments

D. E. Atkins
atkins@umich.edu

Potential for

- "functionally complete" VOs
- "one stop shopping"
- entered into via tailorable portals/gateways.

Realising the Potential of e-Science
– the Role of Institutional Infrastructures

- Fulfilling the promise of e-Science will require more than ingenious computer hardware and software engineering, and more than sophisticated system design of new tools which can be readily used by individual scientists and their organisations.
- No less important will be appropriate "institutional" contexts (i.e., informal norms and formal rule structures) to facilitate collaboration within communities of scientific and technical researchers – both on the ground and in cyberspace.
- The institutional and organisational environment of e-Science encompasses a wide and diverse array of interrelated social, economic and legal factors that
  - create both incentives for, and constraints upon individuals' and collective actions;
  - thereby shape the production, utilisation, consumption, and governance of e-Science capabilities and artefacts.

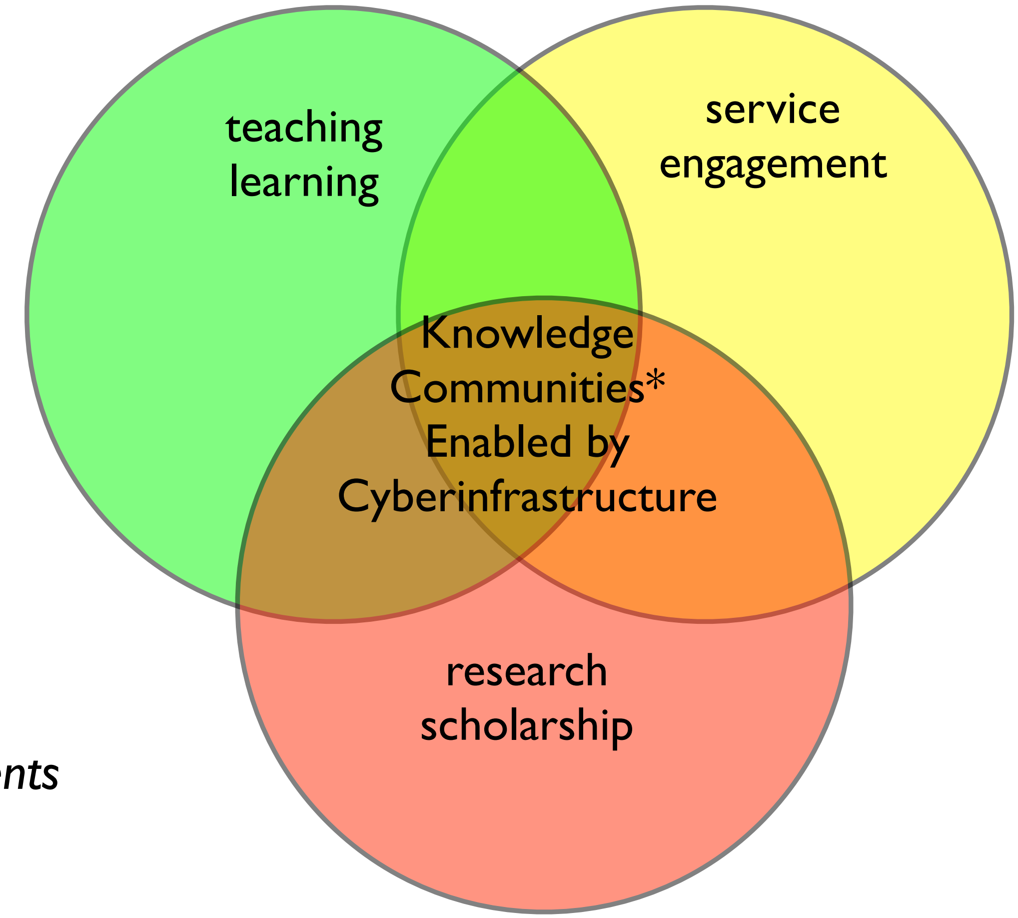
From Prof. Paul David,
Stanford & Oxford

Learning and Workforce Issues

- Learning *supported by* CI and CI-enhanced knowledge communities.
  - learning ecologies: formal, informal, authentic, inquiry-based, life-long.
  - broadened participation; open educational resources
- Learning *requirements* (human capacity building) for the socio-technical design/ evaluation and transformative application *of CI* to science and engineering research and allied education;

And while we are at it...

- can we create CI environments in support of research, learning, and societal engagement in ways that exploit complementarity between them?
  - *Pasteur's Quadrant research*
  - *Ubiquitous learning environments*
  - *Authentic learning*
  - *Professional development*



The Openness Movement → *The CLEAR Agenda*

The OPEN CLEAR Agenda

D. E. Atkins
atkins@umich.edu

Some Questions for the TeraGrid Community

- In what ways is the TeraGrid more than the sum of its parts? Major successes? Major barriers/problems? How do we "instrument" the TeraGrid to better understand its structure, use, and impact.
- Can/should the TeraGrid become the primary fabric for comprehensive virtual communities (collaboratories.)? What are the challenges to extensibility in both scale and resource diversity?
- How do we provide (global) interoperability and/or consolidation across a variety of grid community projects. How do we mitigate wasteful competition and redundancy?
- What are the right models for multi-stakeholder shared sustainability: operation, maintenance, enhancement?
- What else should I be asking?

Topics for ACCI Breakout Groups

- 1. CI and National Competitiveness - J. Duderstadt
- 2. CI Data Issues - S. Graves
- 3. CI Integration Architecture and Software Issues - S. Feldman
- 4. CI and Universities - J. Bottum

Topics for the Future:
CI and Education
International Dimensions
O&M Strategy
What other topics do you suggest?

Questions and Discussion

National Science Foundation
www.nsf.gov