

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



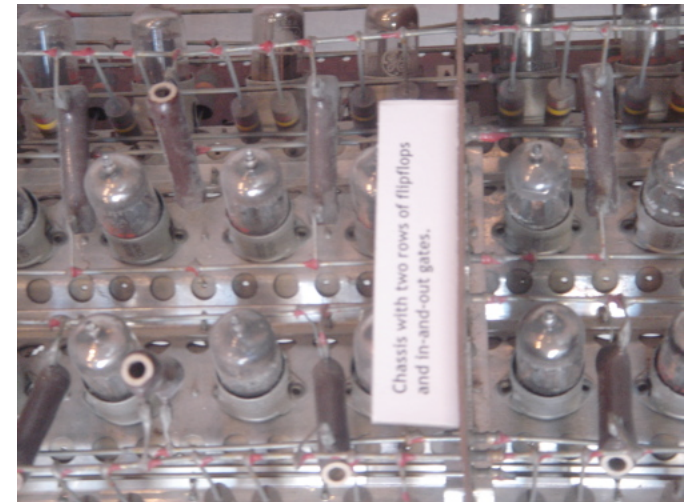
National Science Foundation
www.nsf.gov

Personal
Privilege and
Amazement:
From Digiclock
to Petascale
Grid
Communities

Bucknell University,
circa 1965



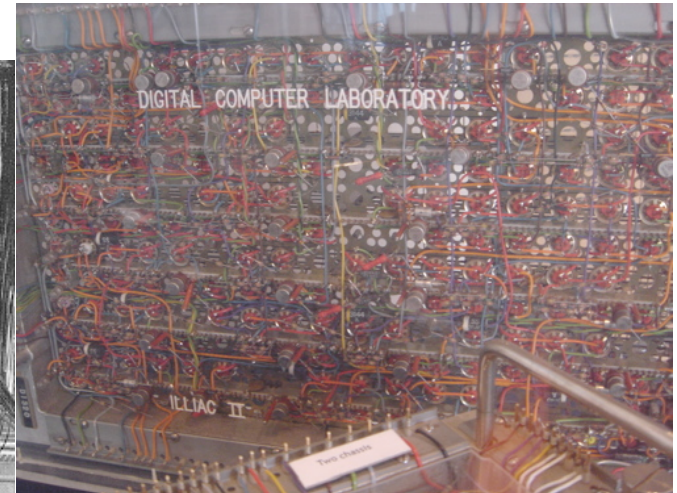
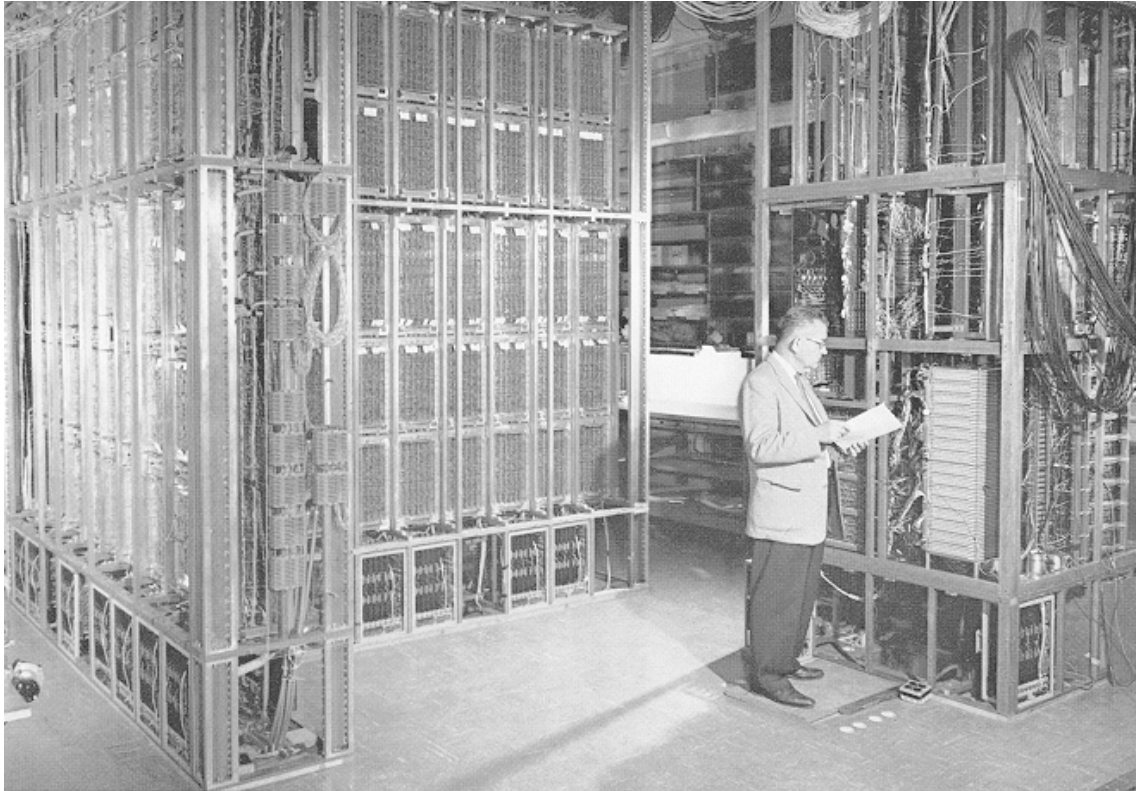
Illiac I (1949-1962)



Vacuum tube circuits

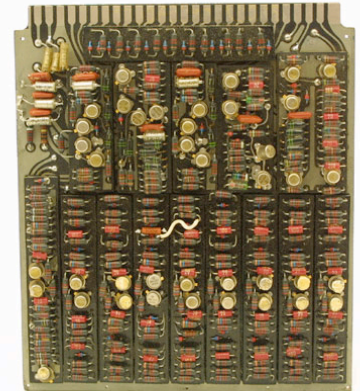
Illiac I was the first computer built and owned entirely by a university. It went on line on September 22, 1952. With **the computing power of a modern-day handheld calculator**, Illiac had 2,800 vacuum tubes and weighed 10,000 pounds. It was more than 10 ft long, 2 ft high, and 8 ft high, with a 5k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

Illiac 2 (1956-68)



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

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

Illiac IV (1965-82)



Illiac IV to perform defense tasks

Spent

(Continued from Page 1)

Area Limitation Talks these going on in Helsinki, Finland.

Sources had speculated that much of the testing of the new IBM workbooks will now be done by computer simulation.

Shatzk, however, said that the Illiac IV is not a classified facility and that it will not have the security safeguards required to process classified data.

Common carrier

"The Defense Department will not be able to bring the bulk of their data here. Illiac IV will have to be involved in what Shatzk termed "scientific subproblems" which would in turn be used as input to classified military problems. Shatzk said data from Illiac IV will also be accumulated and available to the public.

Shatzk admitted that Illiac IV could be tied up to other computer facilities, "all of them on the campus at MIT, Stanford and UCLA, as well as the Rand Corp. But Shatzk said all such connections were on common carrier (like telephone) lines. National security regulations would not allow transmission of classified data on a common carrier line.

Shatzk also indicated the computer would be in use by the Space Defense System, which attempts to track every man-made object in space, as well as for satellite monitoring.

The cost capabilities of the machine, Shatzk said, will also allow "a vastly upgraded processing system for radar, sensor and science data."

Shatzk said he was aware of the new law and felt "there would be enough technological fallout from Illiac IV of interest to the Defense Department," so expected continued handling of the program.

The computer itself is being fabricated in Pennsylvania by the Burroughs Corp. and will draw anything new in existence. Howard Clumant, member and former president of the University Board of Trustees, announced last fall the facility was so vast it would increase the U. S. computer capacity by 25 per cent "the moment was plugged in."

A million instructions

Shatzk said Illiac IV is a highly parallel computing system which allows a single instruction system to act on a great number of variables at once.

The original Illiac IV program as outlined in the first Air Force contract, envisioned a computer with 200 units able to handle a billion instructions per second.

The computer that will be attached here, however, will handle only 84 units and handle only 10 million instructions per second (4 computer built to date).

Shatzk said he felt the Bureau is also a result of subcontractors on the part of scientists in questionable business practices to the holders of the computer.

"We're ending up spending far more money on it than we had planned to spend on IBM," he said.

Shatzk said he was aware of the new law and felt "there would be enough technological fallout from Illiac IV of interest to the Defense Department," so expected continued handling of the program.

The computer itself is being fabricated in Pennsylvania by the Burroughs Corp. and will draw anything new in existence. Howard Clumant, member and former president of the University Board of Trustees, announced last fall the facility was so vast it would increase the U. S. computer capacity by 25 per cent "the moment was plugged in."

Signed into law

Manuscript when Defense Secretary Melvin Laird on Nov. 25, 1968, to check compliance with the military procurement authorization of 1970, signed into law by President Nixon on Nov. 25.

The procurement act compels the Defense Department to award funding for non-military research. The act, however, has not been applied to the Illiac project, which received its latest funding increase in Nov. 28.

Parkard replied to Mansfield on Nov. 2 that the department was complying fully with the new law in a sense that same day he declassified program. Parkard reiterated the department's intent not to support any "work which does not have a direct, apparent, and clearly documented relationship to one or more specifically identified military functions or operations."

Aware of law

Parkard announced that this policy would also apply to the "non-classified" research, the

direct of what the nation does with the DOD versus what you're continuously working to achieve.

"These military things are not my interests but Department of Defense interests. They are interested in ways of meeting their objectives while at the same time they are offering me a way to meet my objectives."

The Daily Illini confirmed the primarily military intent of the Illiac IV facility through correspondence between Illinois Democratic leader Mike Mansfield and David Parkard, deputy secretary of defense.

Shatzk said the site of the computer may make possible "a far more hostile and powerful weapons system by creating a body of scientific knowledge for such weapons systems."

But Shatzk said he felt the project was justified in spite of these factors. "I know the military side of the Department of Defense (DOD), some of them are dangerous facts. But their power base is not dependent on me. We 'inventors' "

"If I could have gotten \$50 million from the Red Cross I would not have minded with the DOD. But, you, legislators, measure, the

Is Illiac Killiac?

Or: Is defense money better than none at all?

Senate committee okays Illiac computation center

Proposal goes to vote N

JH

cut/1

II

THE

DI

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



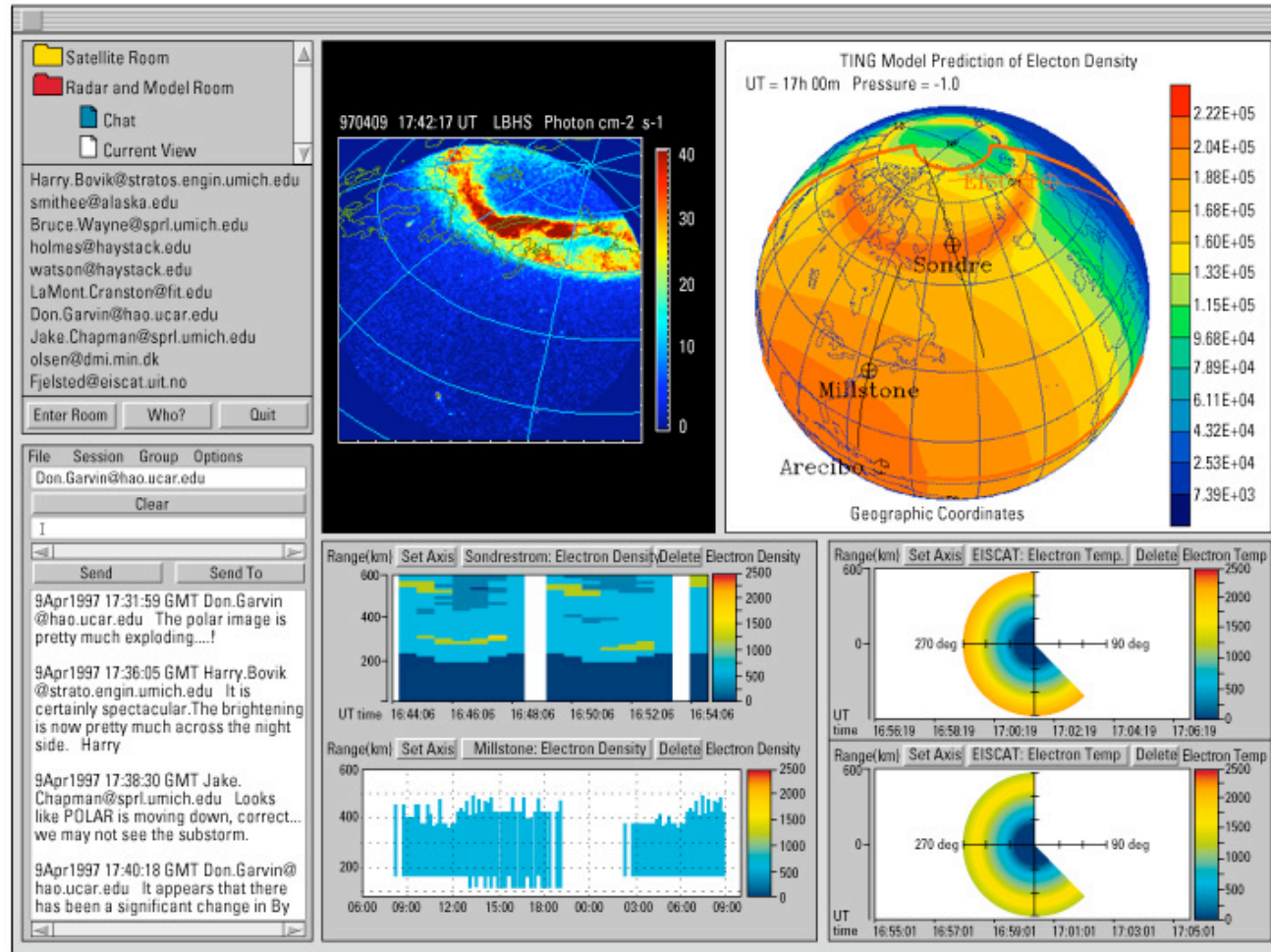
D. E. Atkins
atkins@umich.edu

UARC Interface

Real-time instruments *computational models*

*dynamic
work
rooms*

*team
chat*



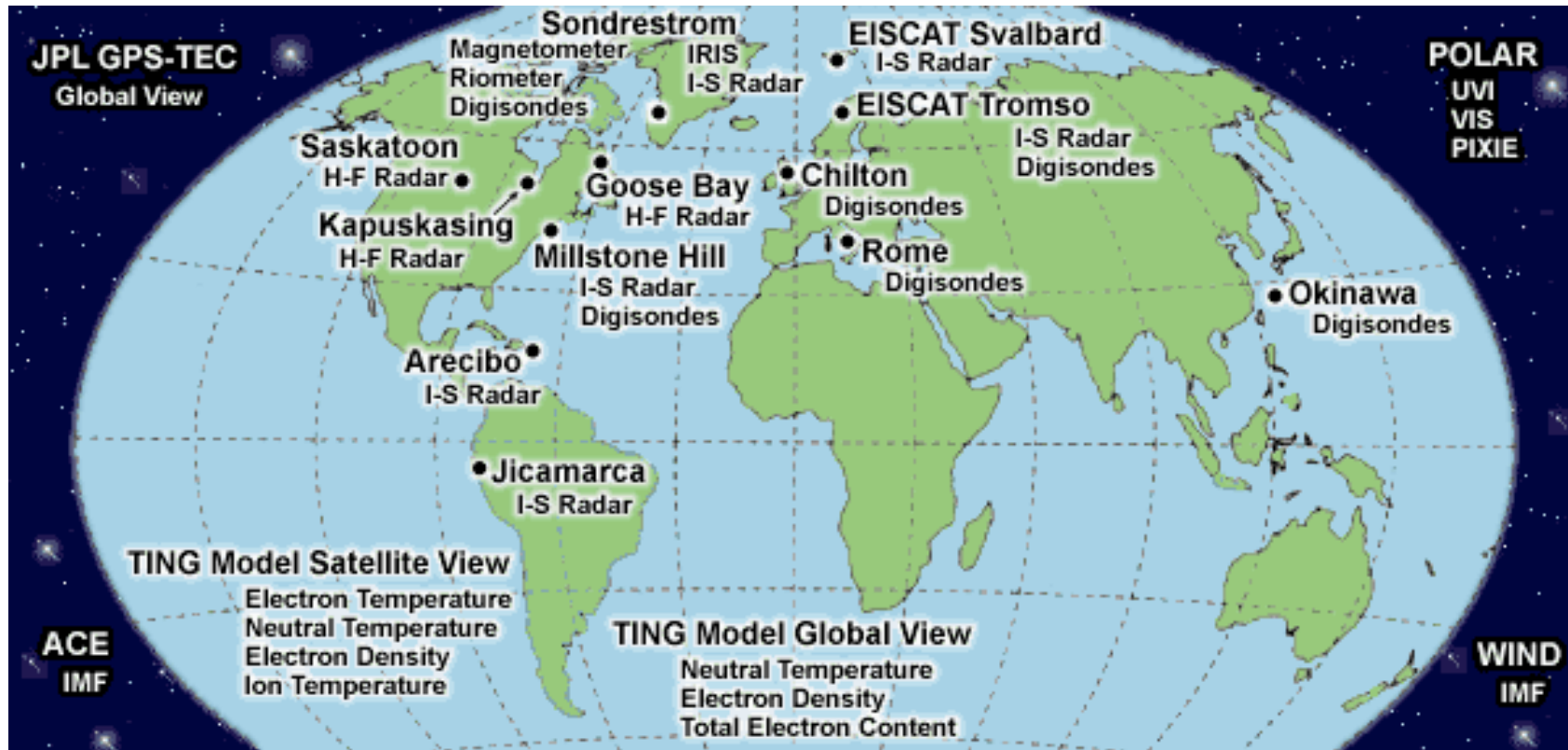
annotation

Session replay

Archival data

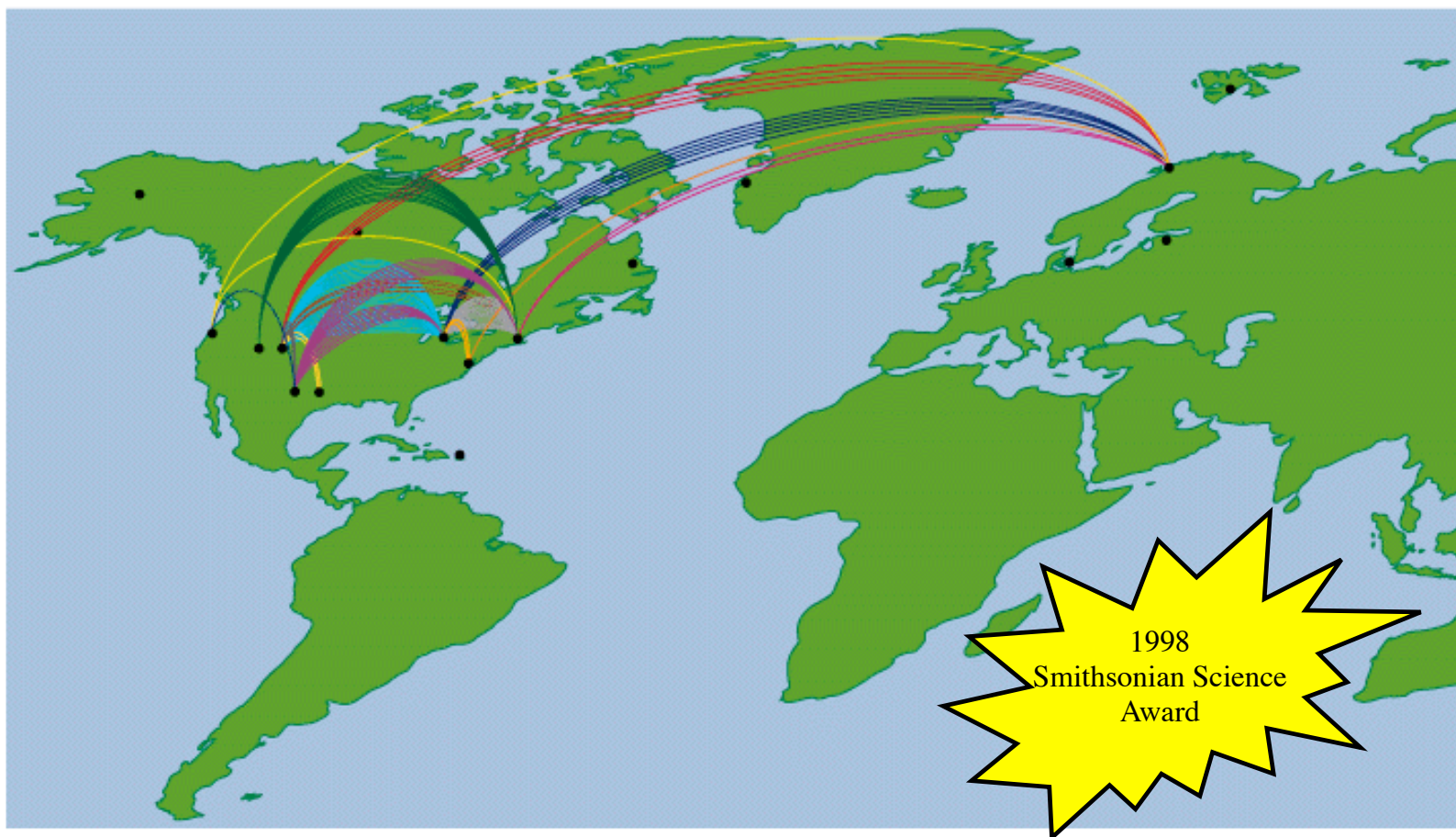
Journals

Evolved into a Network of Instruments (one global instrument)



UARC Patterns of Communication

Pattern of Communication, UARC Campaign, April 9, 1997



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.



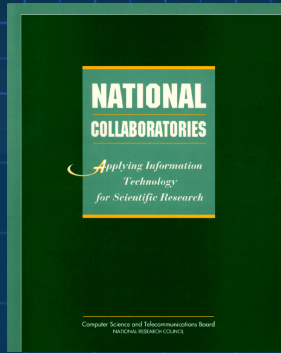
Background and Call to Action



National Science Foundation
www.nsf.gov

CI Genealogy & Movement

Collaboratories



KDI

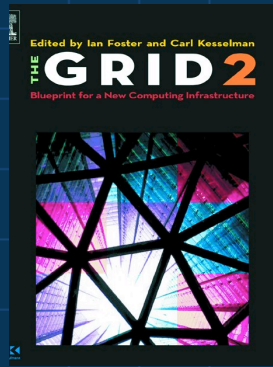
PACI

HPCC

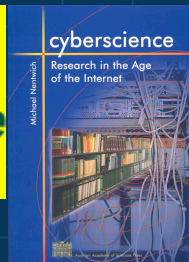
Digital Libraries

GRIDS

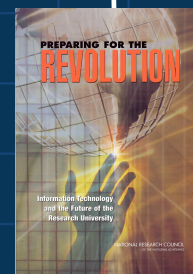
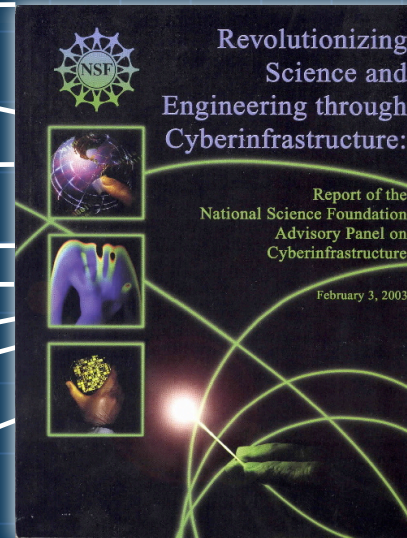
ITR



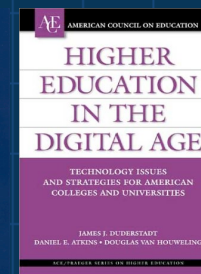
E-science



Cyberscience



Issues for Science and Engineering Researchers in the Digital Age

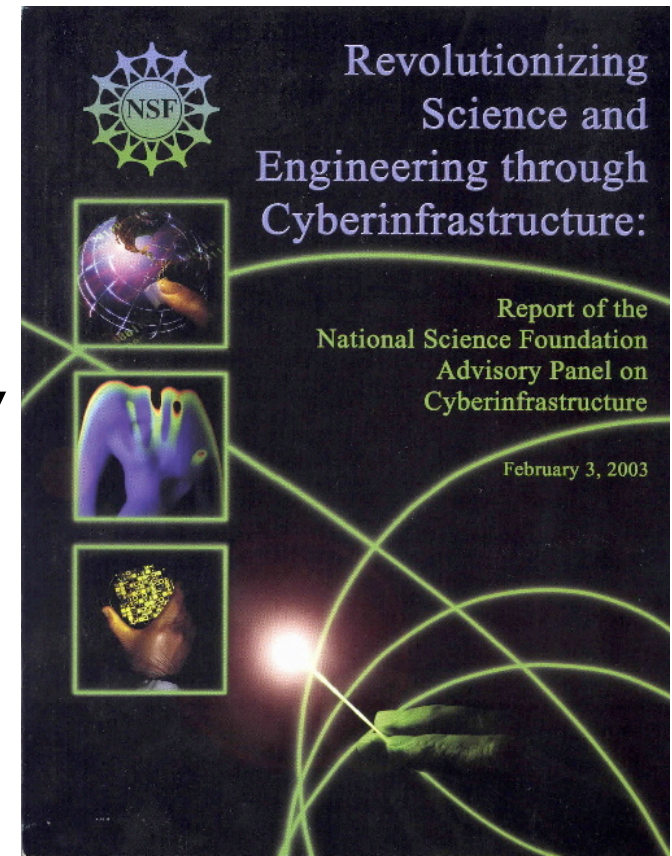


2nd Edition
www.mkp.com/grid2

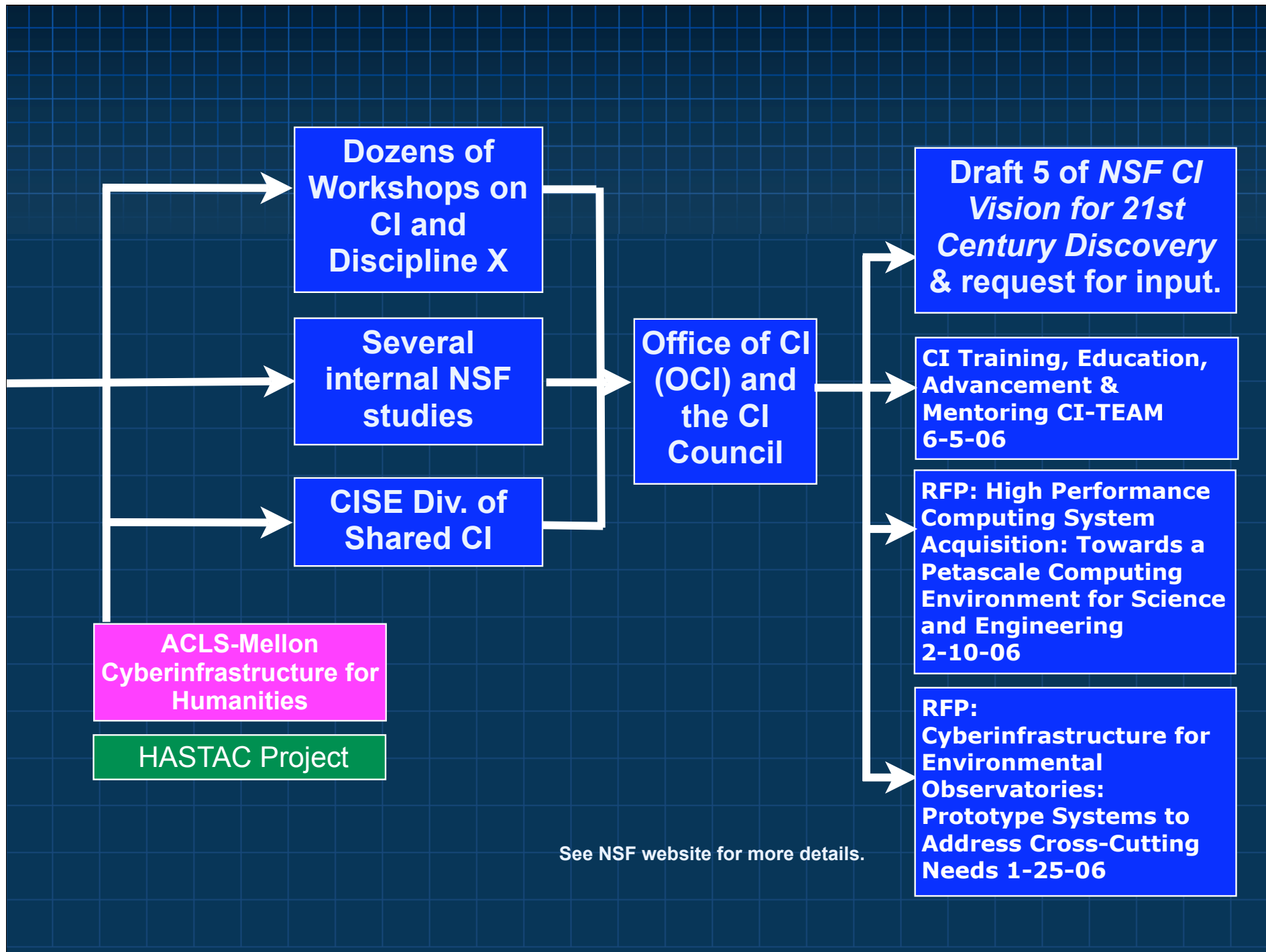
IT & Future of Higher Education

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>



www.nsf.gov/oci/

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

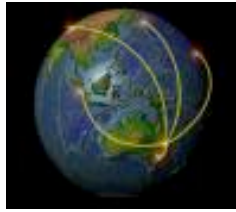
- Scan ACCI Membership List
- See and consider opportunities to serve as a Program Officer

NSF'S CYBERINFRASTRUCTURE VISION FOR
21ST CENTURY DISCOVERY

NSF Cyberinfrastructure Council



National Science Foundation
www.nsf.gov

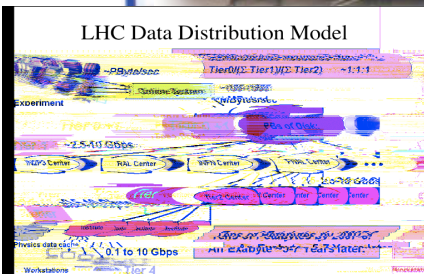
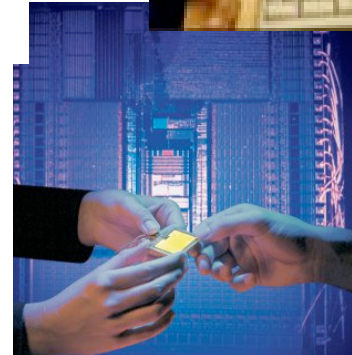


Cyberinfrastructure Vision & Commitment to Leadership



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

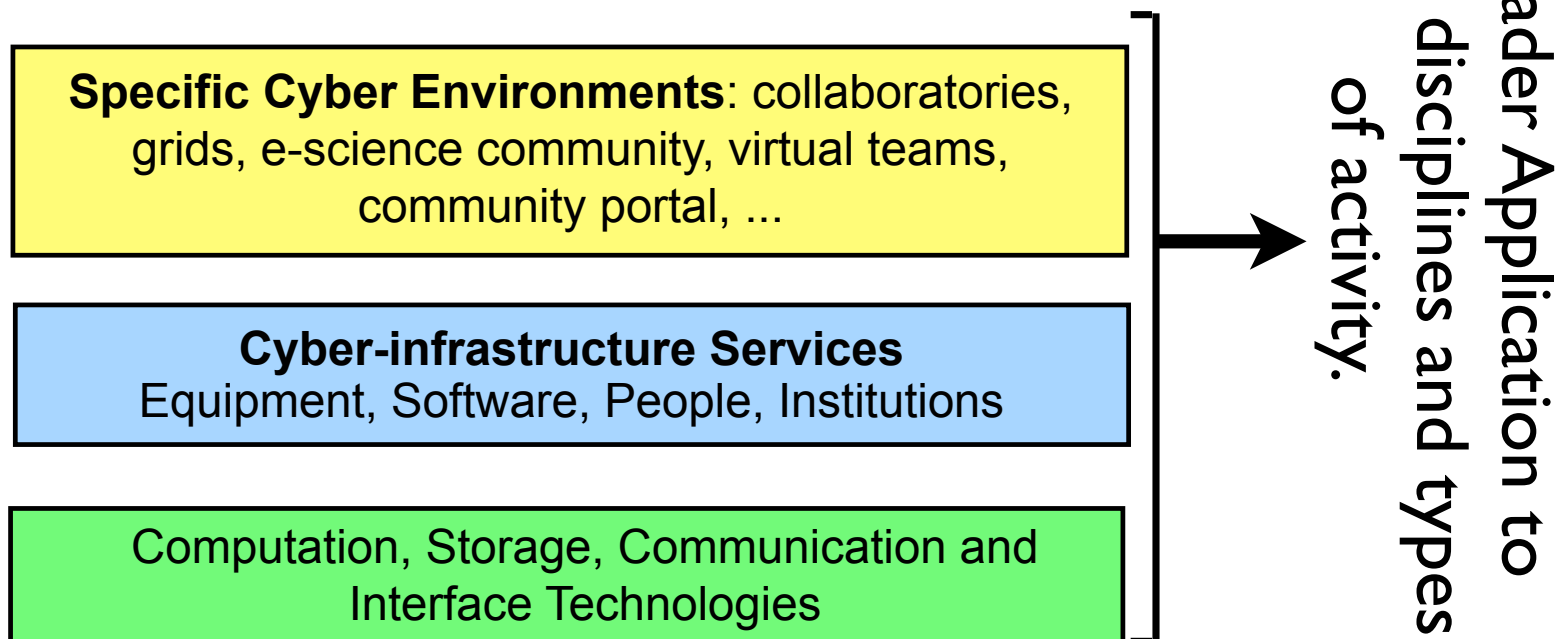
NSF is the only agency within the U.S. government that funds research and education across all disciplines of science and engineering. ... **Thus, it is strategically placed to leverage, coordinate and transition cyberinfrastructure advances in one field to all fields of research.**



Cyberinfrastructure-enhanced Knowledge Communities (Networks)

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

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

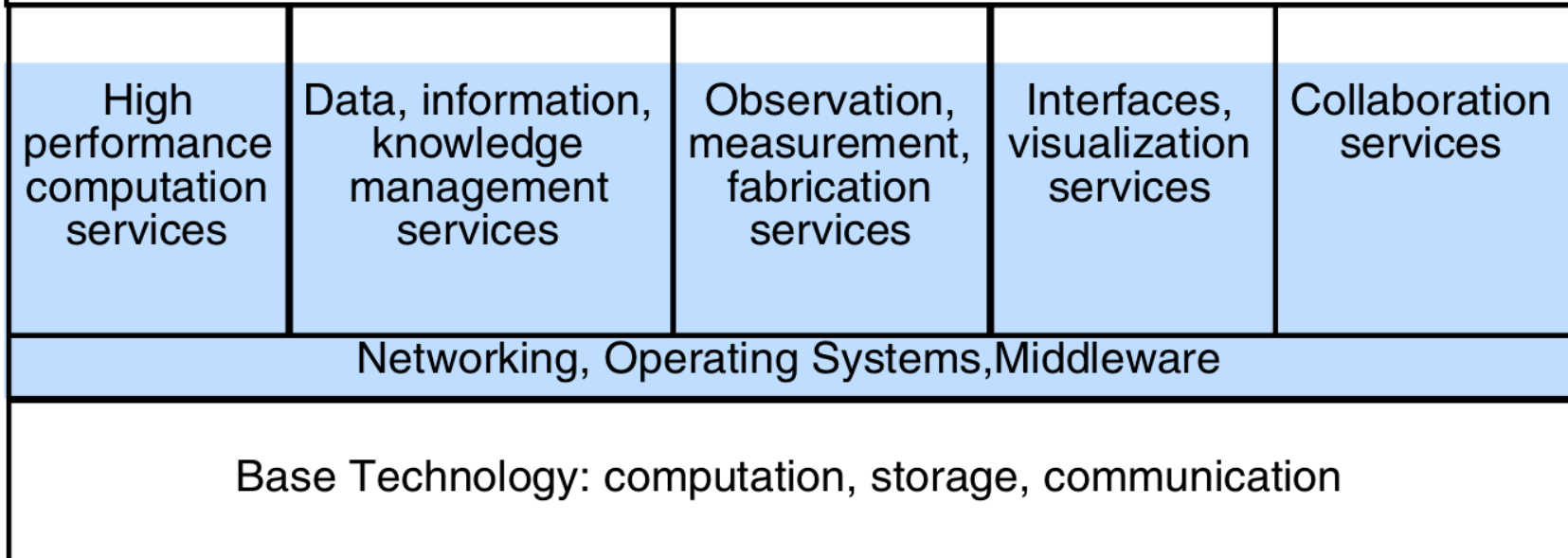



* From Cummings & Kiesler (2003) report on KDI Initiative: Multidisciplinary scientific collaborations, see <http://www.p2design.com/papers/kdi.pdf>

From CI Advisory Panel Report

Community-Specific Knowledge Environments for Research and Education
(*collaboratory, co-laboratory, grid community, e-science community, virtual community*)

Customization for discipline- and project-specific applications



 = *cyberinfrastructure: hardware, software, services, personnel, organizations*

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

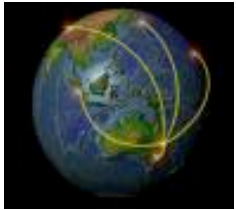
includes networking,
middleware, systems
software?

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*



We are here



X

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

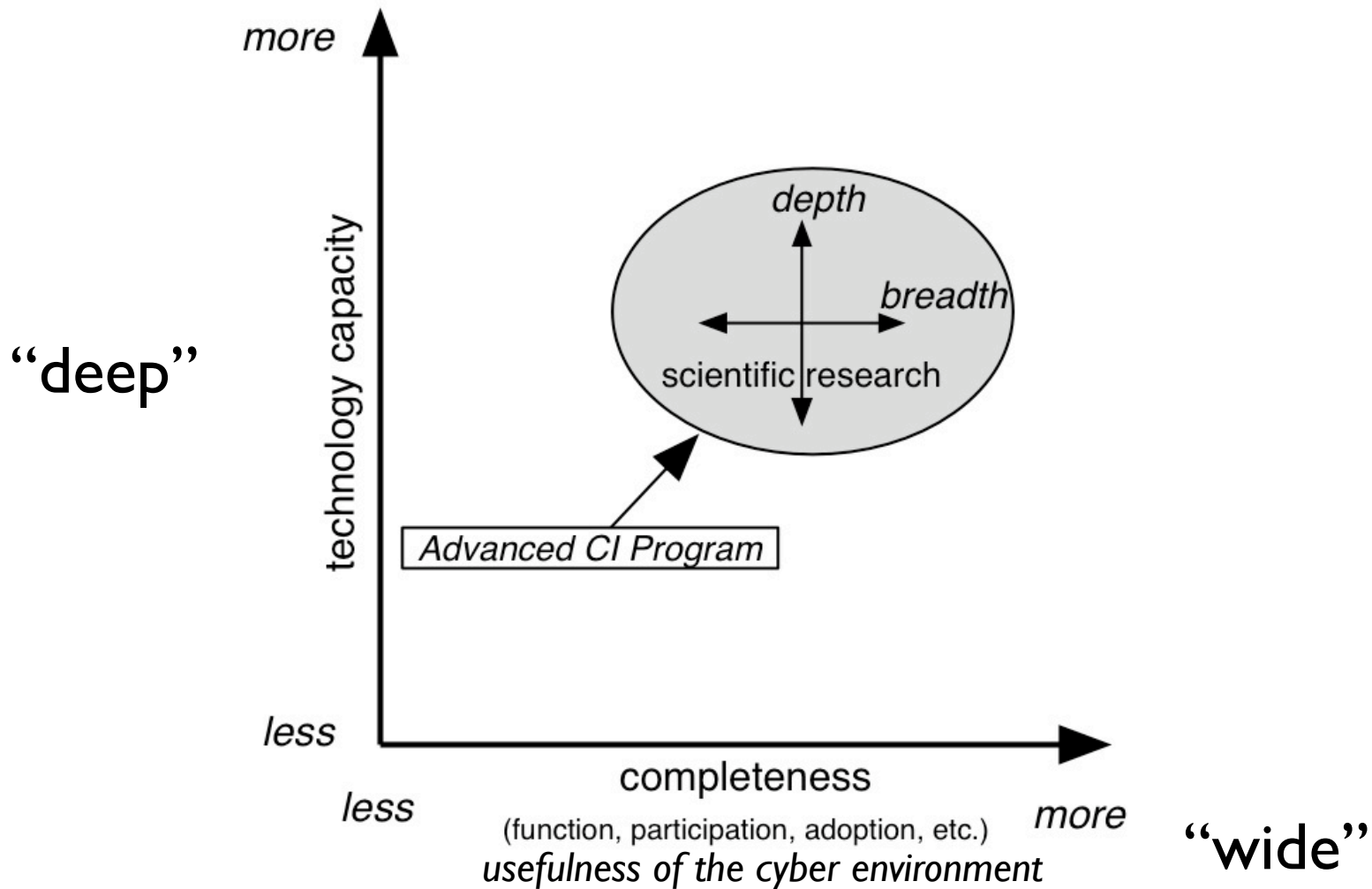
Provisioning -Effective creation, deployment and operation of advanced CI

Application -Innovative use to enhance discovery & learning

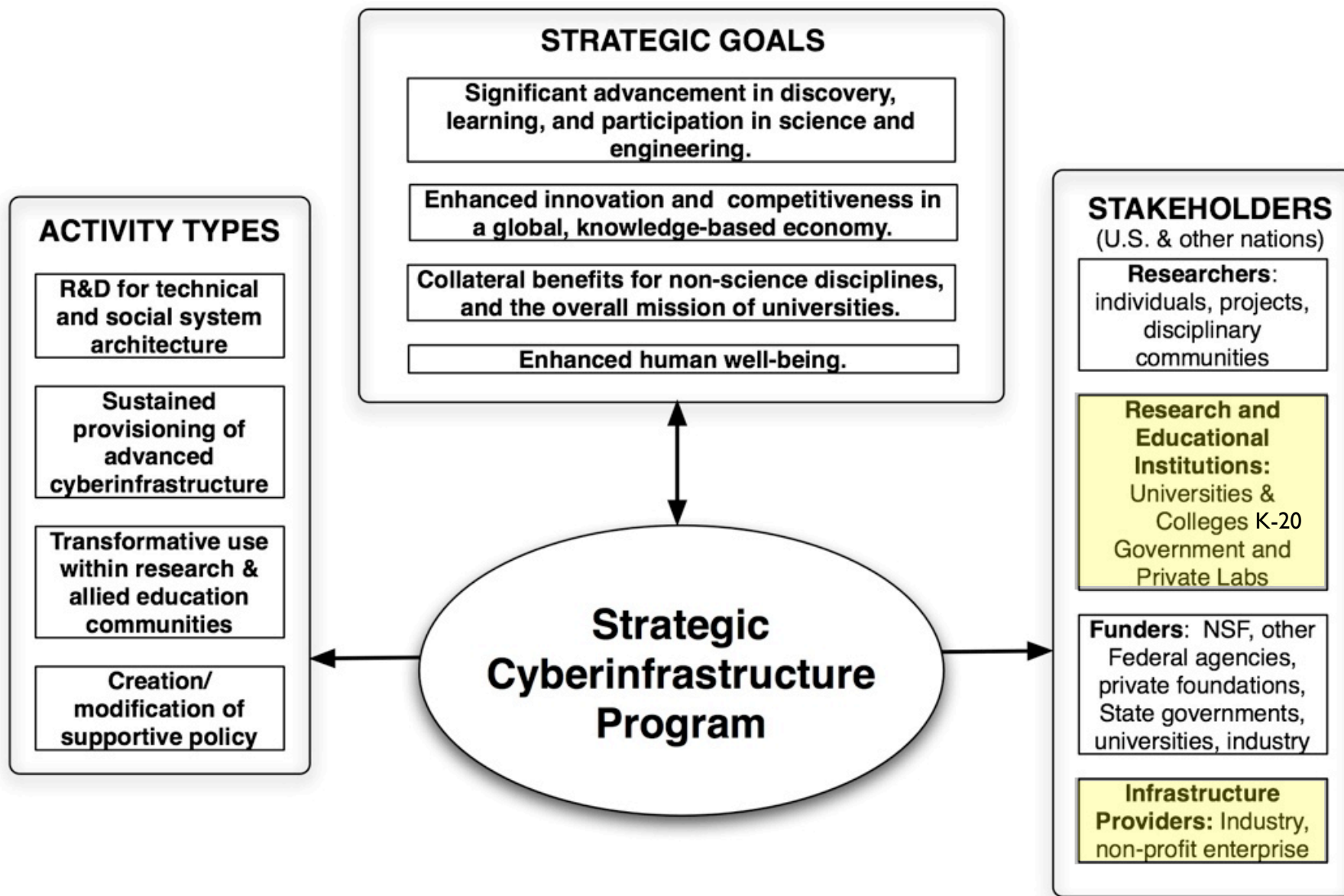


R&D to enhance technology and technical and social dimensions of future CI systems

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



Alignment of stakeholders towards achieving strategic goals



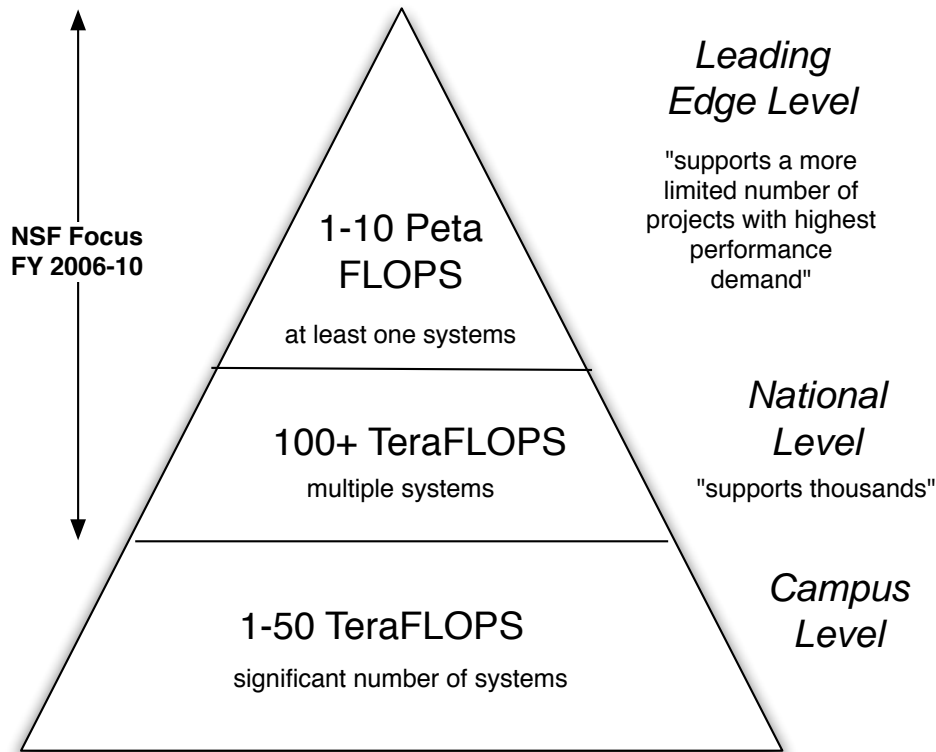
Attributes of NSF HPC Strategy



- ▶ 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

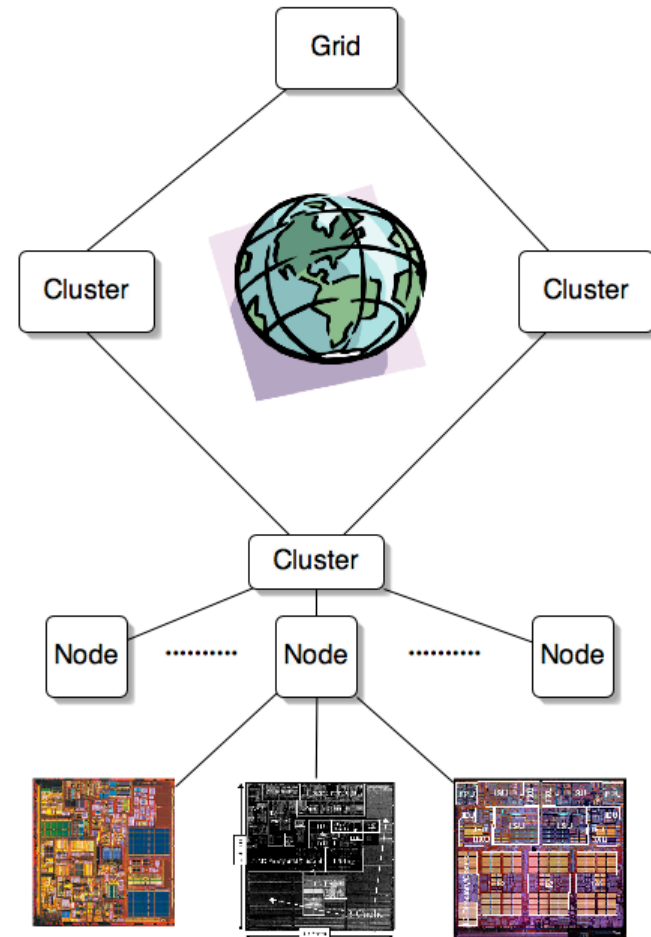
But in an unbundled funding model. i.e. not the PACI core model.

HPC Strategy



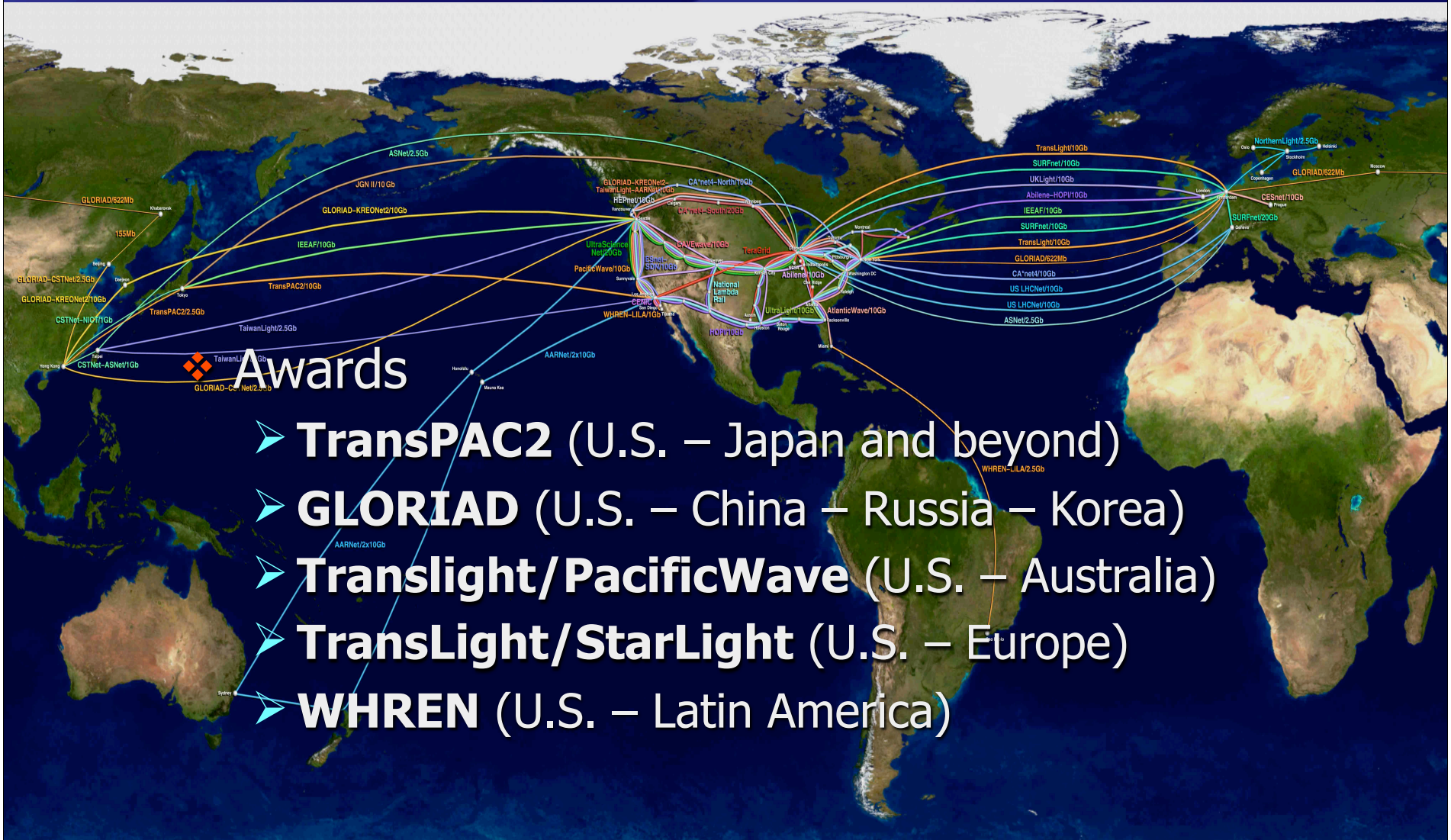
On demand & dynamic resource allocation.
HPC in a relevant-time loop.

System Diversity





International Research Network Connections (IRNC)





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)

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”.

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.



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.

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.



Virtual Organizations

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

**GEOGRAPHIC
PLACE**

TIME

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

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.

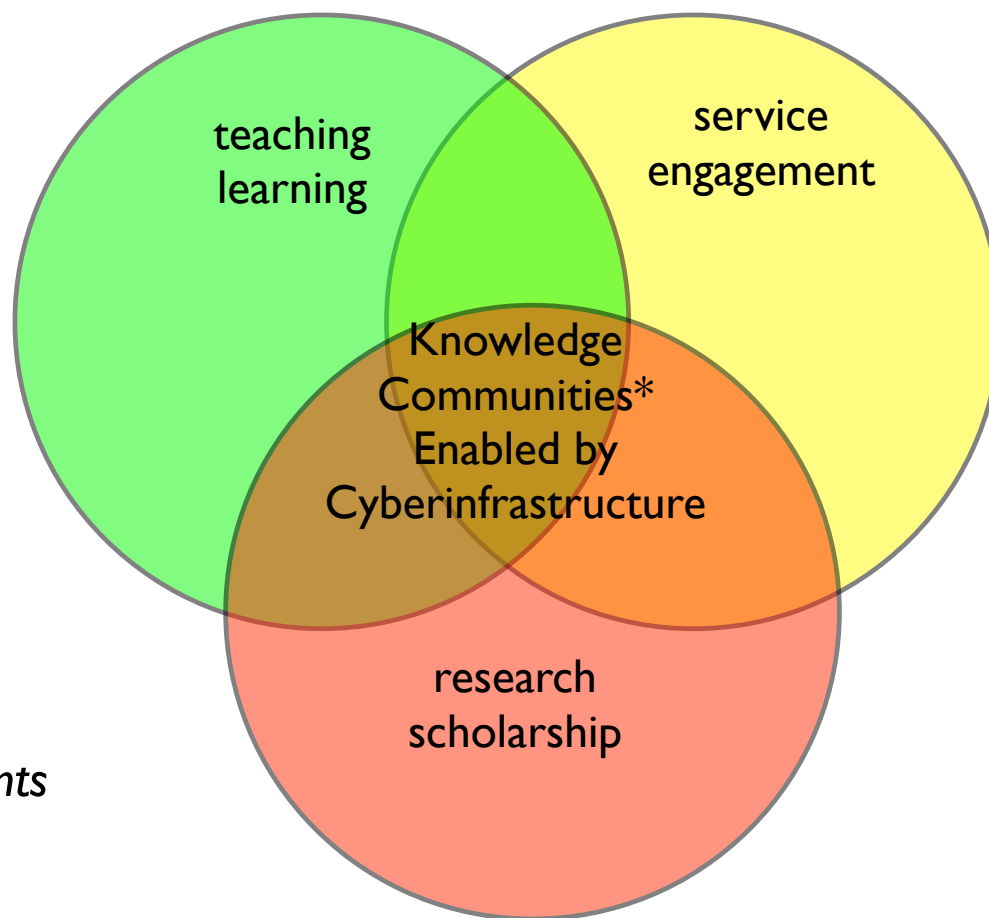
Learning and Workforce Issues

- 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