information technology
UNIVERSITY OF OKLAHOMA

OKLAHOMA SUPERCOMPUTING SYMPOSIUM 2006

The NSF Cyberinfrastructure Vision: Act Locally and Engage Globally

Daniel E. Atkins
Director,
U.S. National Science Foundation
Office of Cyberinfrastructure
datkins@nsf.gov

What I know about the...

Oklahoma Supercomputing Context

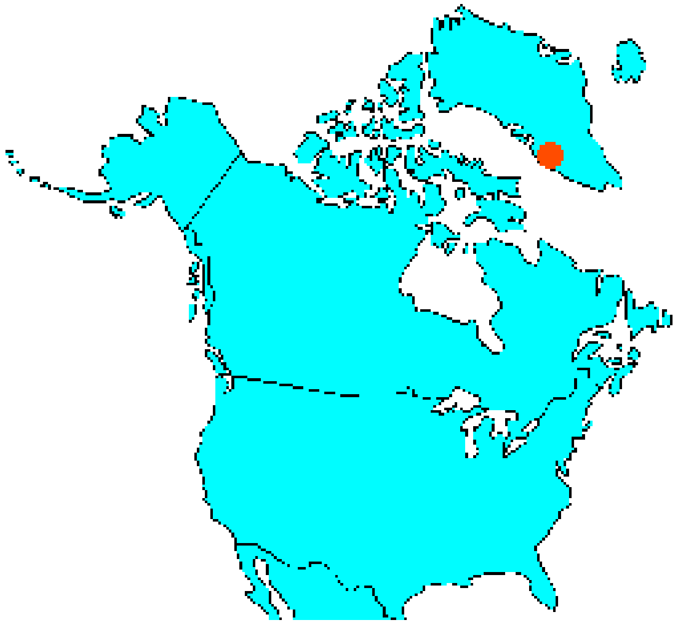
- Regional leadership and facilitation of this well-established and valued supercomputing symposium series.
- Well known for computational science education and training.
- Culture of collaboration between academia, industry, and government.
- International leadership in modeling/predicting extreme weather events.

Act Locally; Engage Globally

- "engage globally" broadly defined to include geography, inter-disciplinary, inter-institutional, integrated research and education.
- supercomputing as a major resource (often the centerpiece) of broader distributed knowledge communities supported by CI.
- Working in global reach virtual organizations.
- Supercomputing in the age of virtualization; relaxed constraints of distance and time.

The University of Michigan Upper Atmospheric Research Collaboratory (UARC)

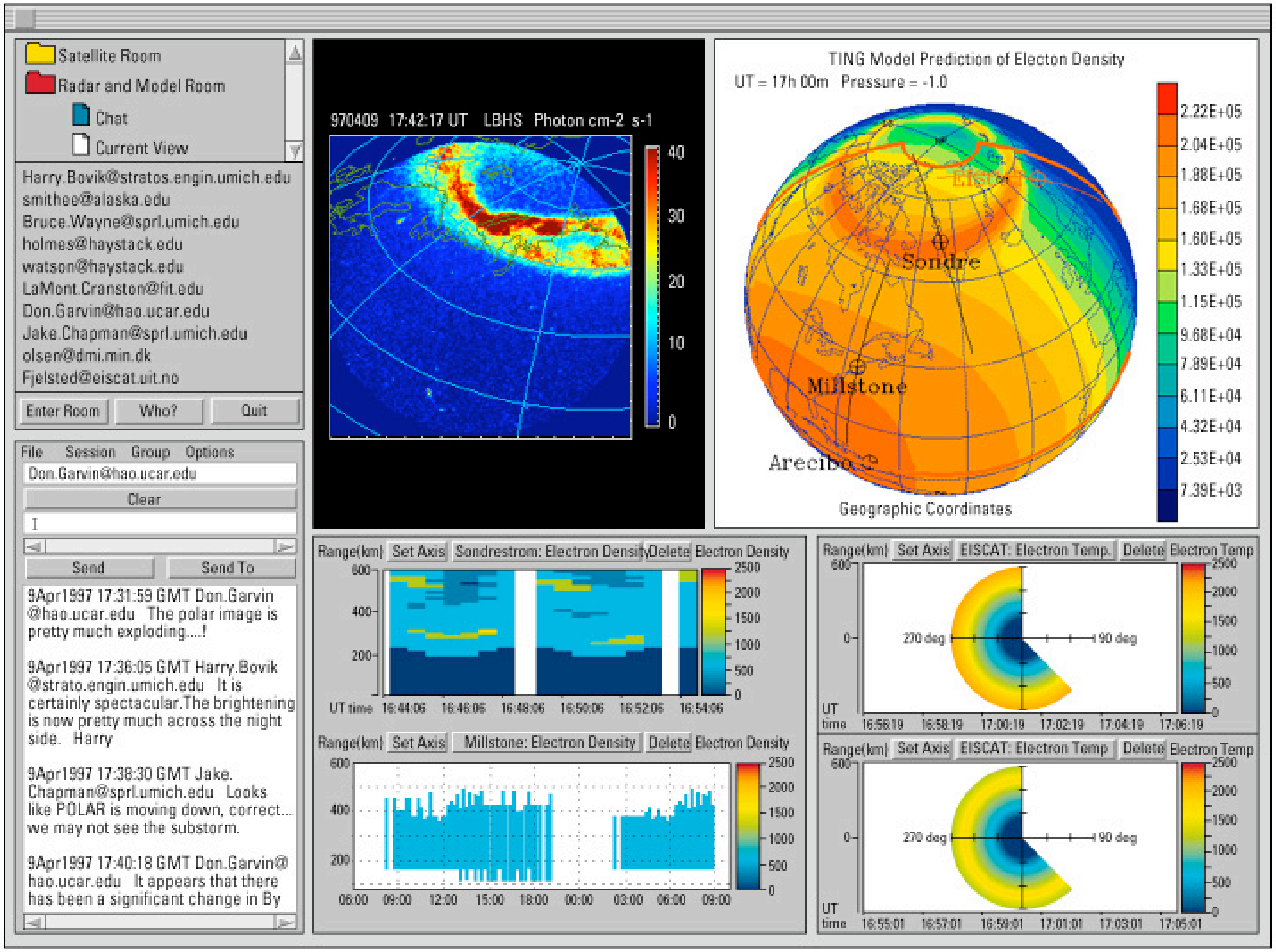
The Initial Facility at Sondrestrom, Greenland



UARC Interface (circa 1988)

Real-time instruments ***computational models***

dynamic work rooms



annotation

team chat

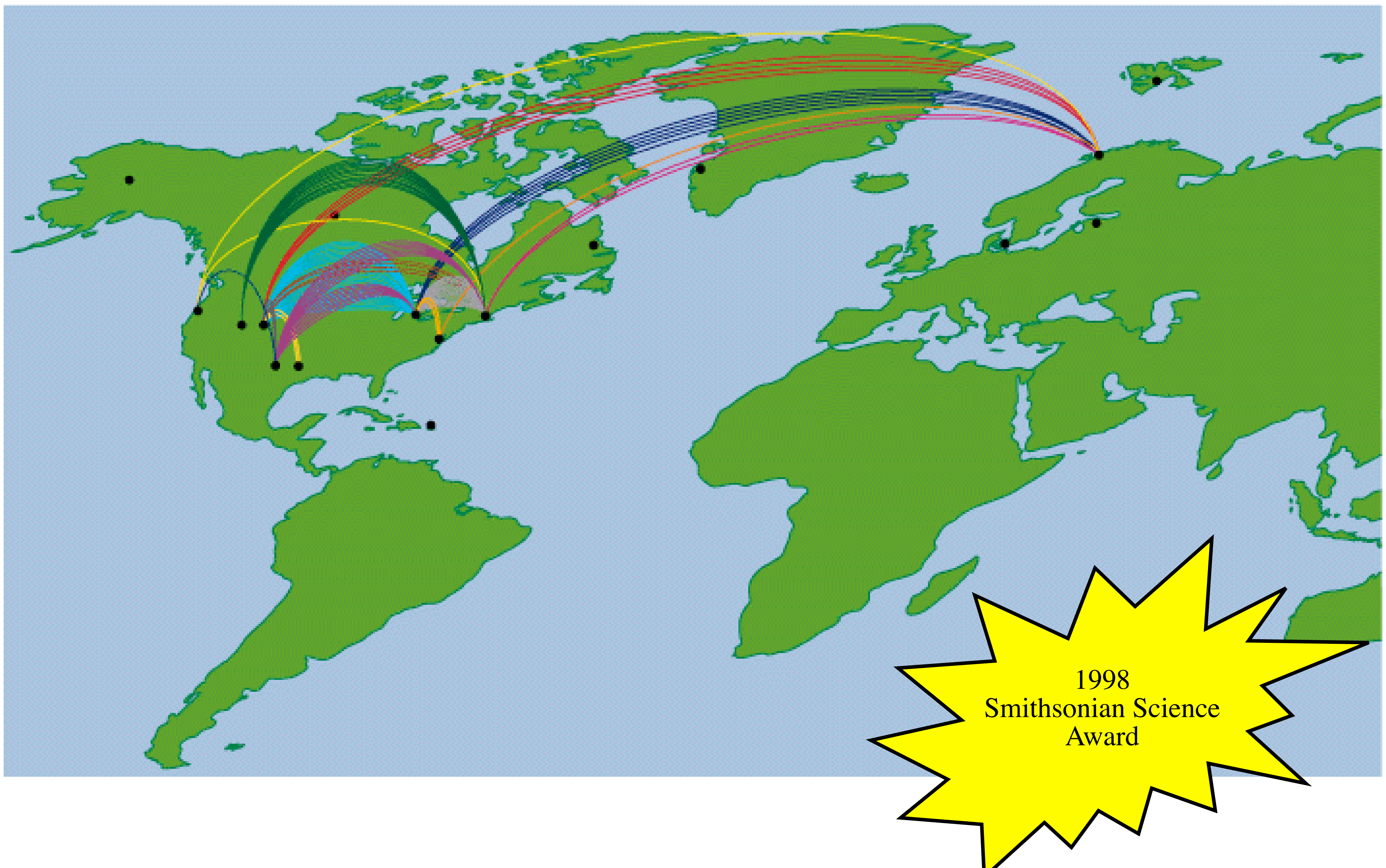
Session replay

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.

Nomenclature

ACLS-Mellon Study: Cyberinfrastructure for Humanities
http://www.acls.org/cyberinfrastructure/

HASTAC Project
http://www.hastac.org/

research & learning

science

engineering

rapid response

humanities

world universities

CI-enhanced | enabled

Cyber science

e-research

e-science

e-infrastructure

cyberinfrastructure (CI)

e = electronic | enhanced | enabled

Dualities

cyberinfrastructure ⇄ enables ⇄ research & development

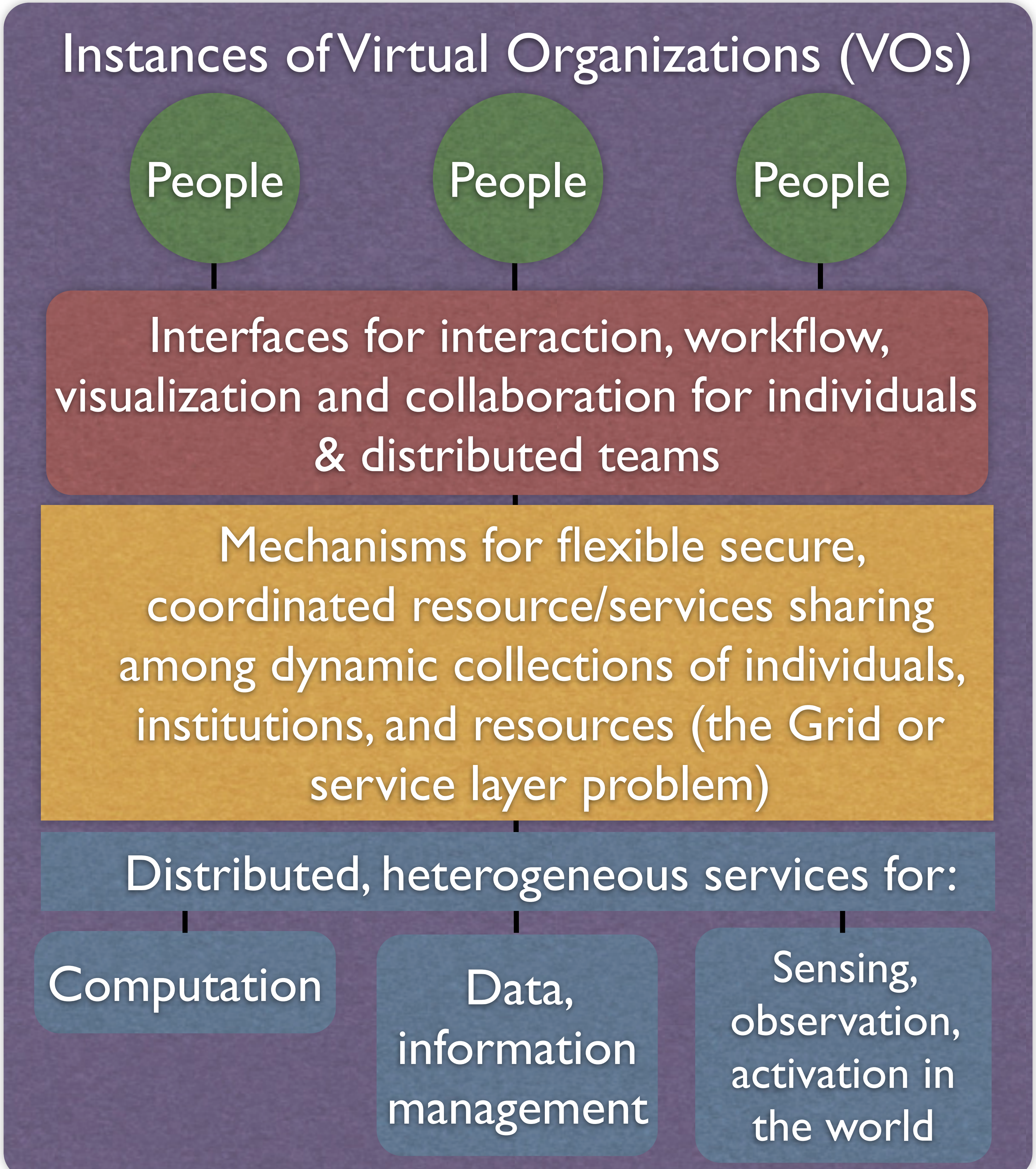
CI is both an object and means for R&D

collaboration ⇄ enables ⇄ CI environments

Multi-stakeholder collaboration required to create, provision, and apply CI; CI supports collaborations across time and distance (geographic, disciplinary, institutional)

learning | education ⇄ enables ⇄ CI environments

Learning and workforce development initiatives required to create and use CI; CI enables/enhances learning/education

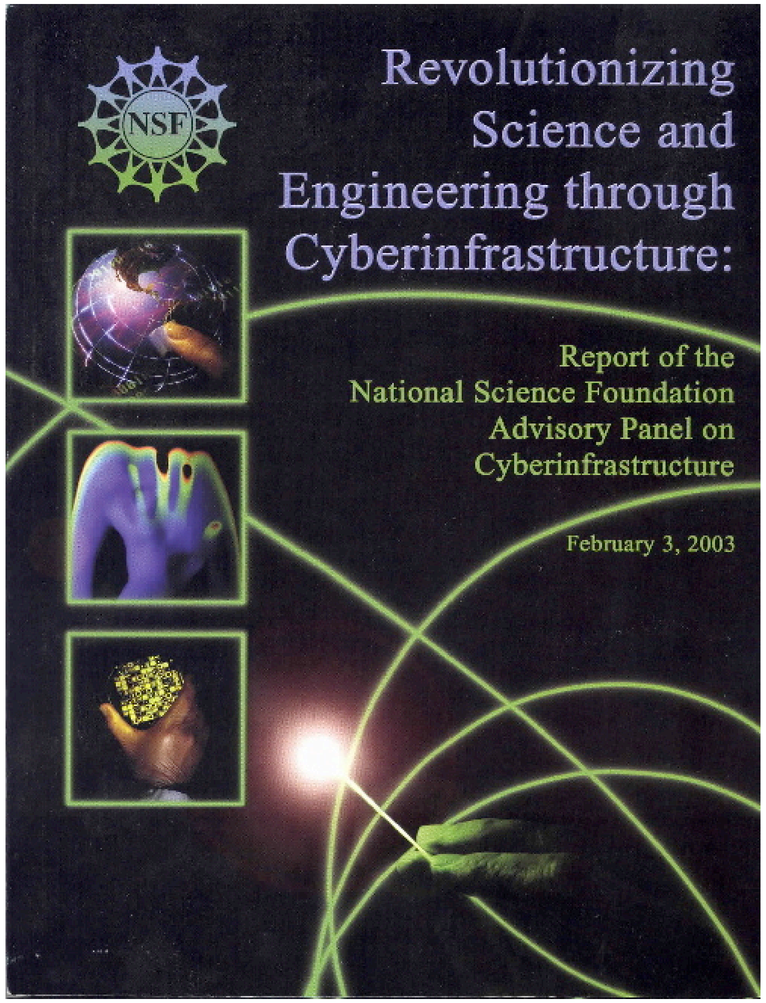


Alternate Names for Instances of VOs:

- Co-laboratory
- Collaboratory
- Grid (community)
- Network
- Portal
- Gateway
- Hub
- Virtual Research Environment (VRE)
- Cyberinfrastructure Collaborative
- Other?

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."



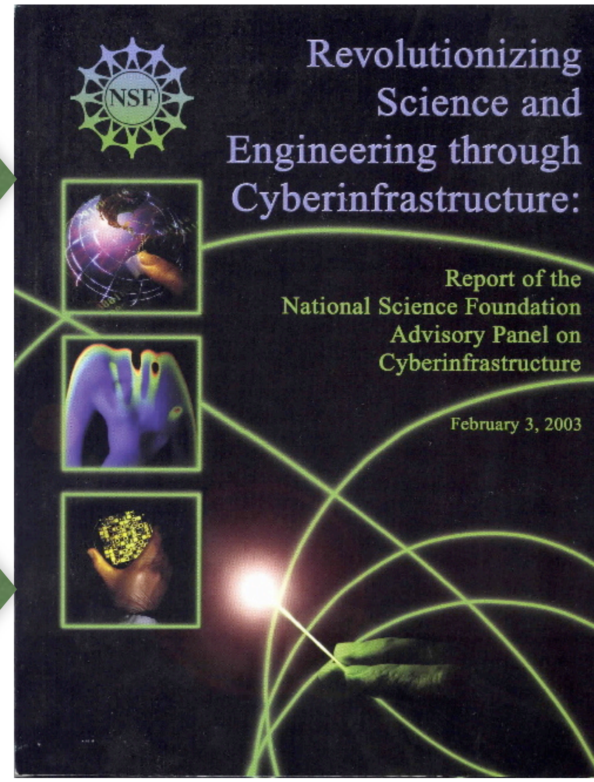
Daniel E. Atkins, Chair
University of Michigan
Kelvin K. Droegemeier
University of Oklahoma
Stuart I. Feldman
IBM
Hector Garcia-Molina
Stanford University
Michael L. Klein
University of Pennsylvania
David G. Messerschmitt
University of California at Berkeley
Paul Messina
California Institute of Technology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

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

Vision and Activities Based on Broad and Diverse Community Engagement

Advances in components of CI-systems for S&E R&E

Complex, multi-scale, multidisciplinary S&E research challenges

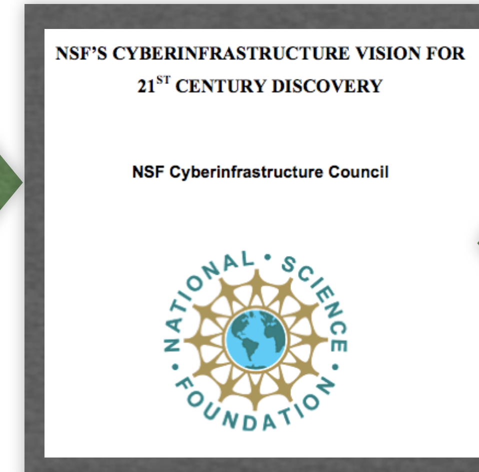

Revolutionizing Science and Engineering through Cyberinfrastructure:

Report of the National Science Foundation Advisory Panel on Cyberinfrastructure

February 3, 2003

30+ disciplinary workshops on CI vision & impact

NSF internal working groups


NSF'S CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY

NSF Cyberinfrastructure Council

NSB & Community Input

CI Council, Directorate/Office CI Activities, OCI, ACCI

Vision Framework

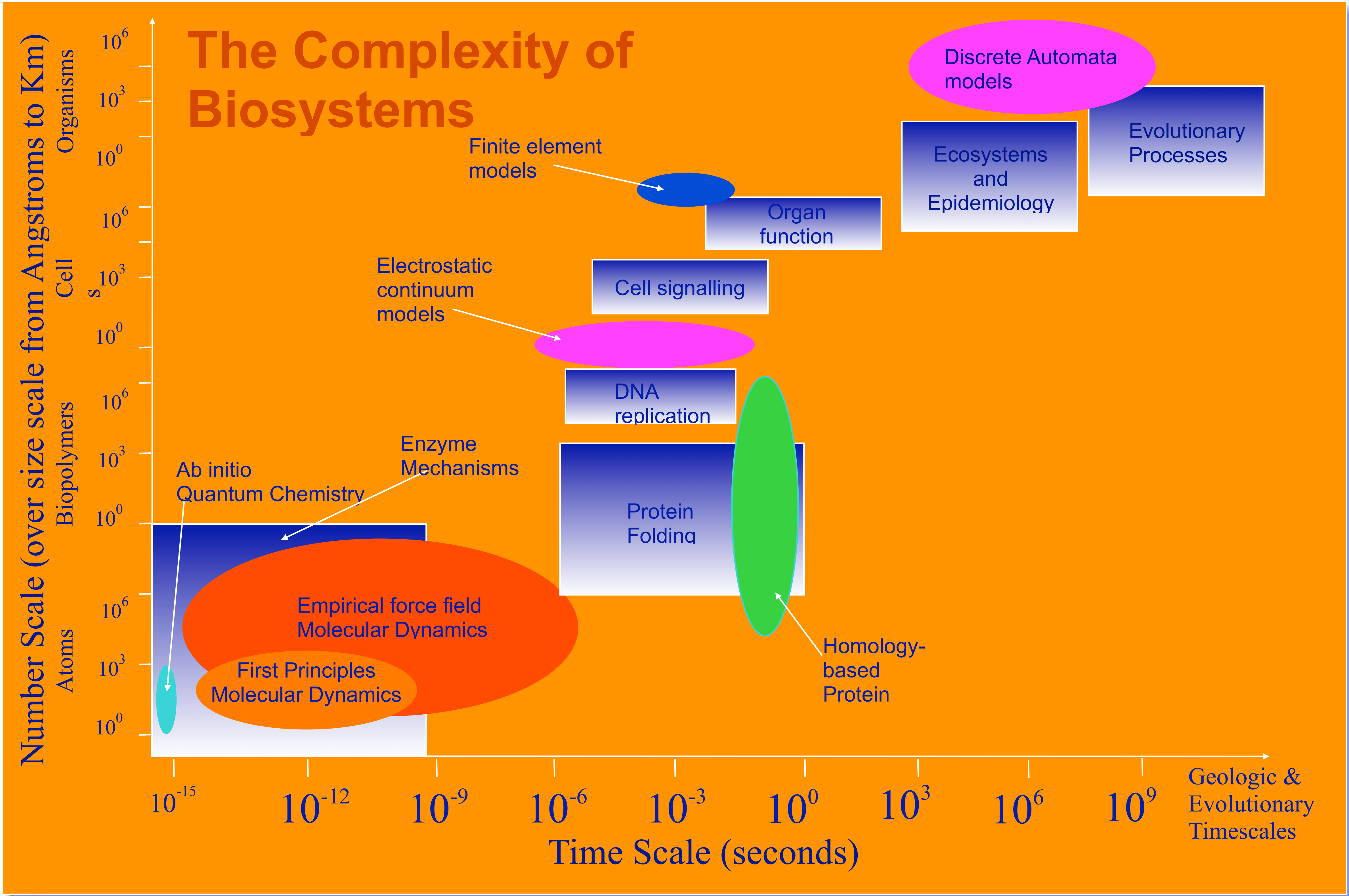
- High Performance Computing
- Data, Data Analysis & Visualization
- Virtual Organizations
- Learning & Workforce Development

- *All directorates and offices support cyberinfrastructure.*
- *Science-driven partnerships between creation, provisioning and use of CI*
- *Supports integrated research and education and broadened access and participation.*

Some Science Drivers

- Inherent **complexity and multi-scale** nature of todays frontier science challenges.
- Requirement for **multi**-disciplinary, **multi**-investigator, **multi**-institutional approach (often international).
- High **data intensity** from simulations, digital instruments, sensor nets, observatories.
- Increased value of data and demand for data **curation & preservation** of access.
- Exploiting infrastructure **sharing** to achieve better stewardship of research funding.
- Strategic need for **engaging more students** in high quality, authentic science and engineering education.

Biosystems Need for Multiscale Computational Modeling



Office of Cyberinfrastructure

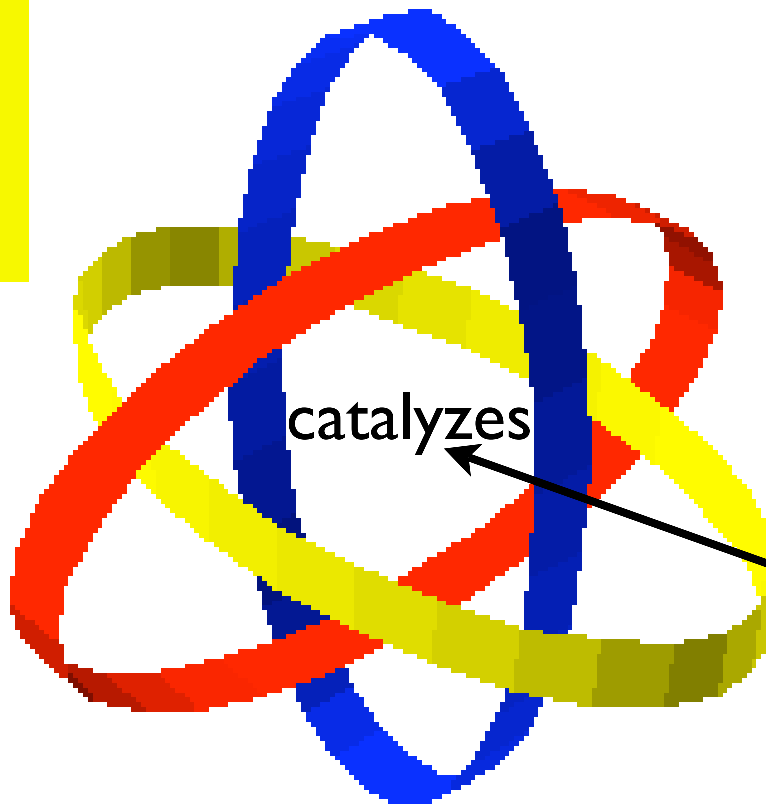
From J. Collins, NSF/BIO AD

D. E. Atkins

Achieving the NSF CI (e-science) Vision requires synergy between 3 types of activities

Transformative Application - to enhance discovery & learning

Provisioning - Creation, deployment and operation of advanced CI



provides shared and connecting CI

Office of Cyberinfrastructure

Borromean Ring: The three rings taken together are inseparable, but remove any one ring and the other two fall apart. See www.liv.ac.uk/~spmr02/rings/

R&D to enhance technical and *social* effectiveness of future CI environments

NSF CI FY07 Budget Request Total of $600M in CI Funding with $182M in OCI

Cyberinfrastructure Funding

(Dollars in Millions)

	FY 2005 Actuals	FY 2006 Current Plan	FY 2007 Request	Change over FY 2006 Amount	Change over FY 2006 Percent
Biological Sciences	$77.00	$84.00	$90.50	$6.50	7.7%
Computer and Information Science and Engineering	45.32	63.00	68.00	5.00	7.9%
Engineering	52.00	52.00	54.00	2.00	3.8%
Geosciences	71.35	71.35	75.00	3.65	5.1%
Mathematical and Physical Sciences	56.52	59.30	63.56	4.26	7.2%
Social, Behavioral and Economic Sciences	20.39	20.54	20.54	-	-
Office of Cyberinfrastructure	123.28	127.12	182.42	55.30	43.5%
Office of International Science and Engineering	0.22	1.00	1.05	0.05	5.0%
Office of Polar Programs	25.38	26.24	26.24	-	-
Subtotal, Research and Related Activities	471.47	504.55	581.31	76.76	15.2%
Education and Human Resources	20.27	15.02	15.52	0.50	3.3%
Total, Cyberinfrastructure Funding	$491.74	$519.57	$596.83	$77.26	14.9%

Totals may not add due to rounding.

www.nsf.gov/oci/

NSF'S CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY

NSF Cyberinfrastructure Council

NATIONAL SCIENCE FOUNDATION

Several Active Solicitations Posted

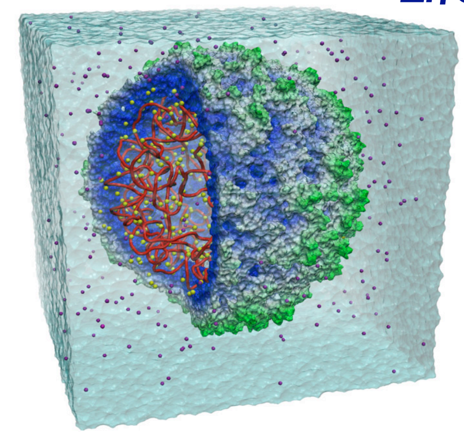
Seeking more program officers.

www.nsf.gov/od/oci/ci-v7.pdf

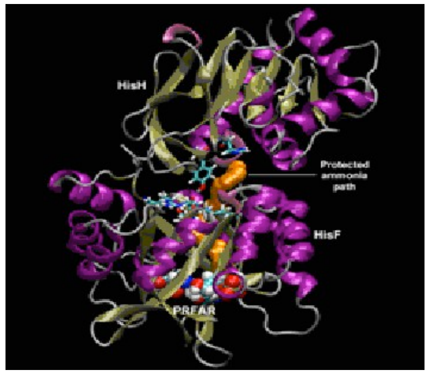
High Performance Computing

increasingly important tool for understanding

Life

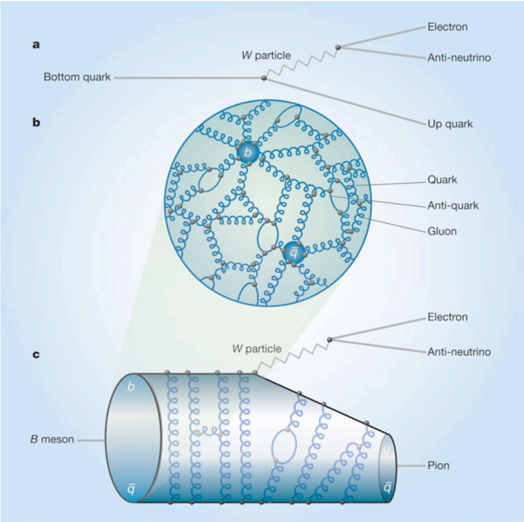


Satellite tobacco mosaic virus, P. Freddolino et al.

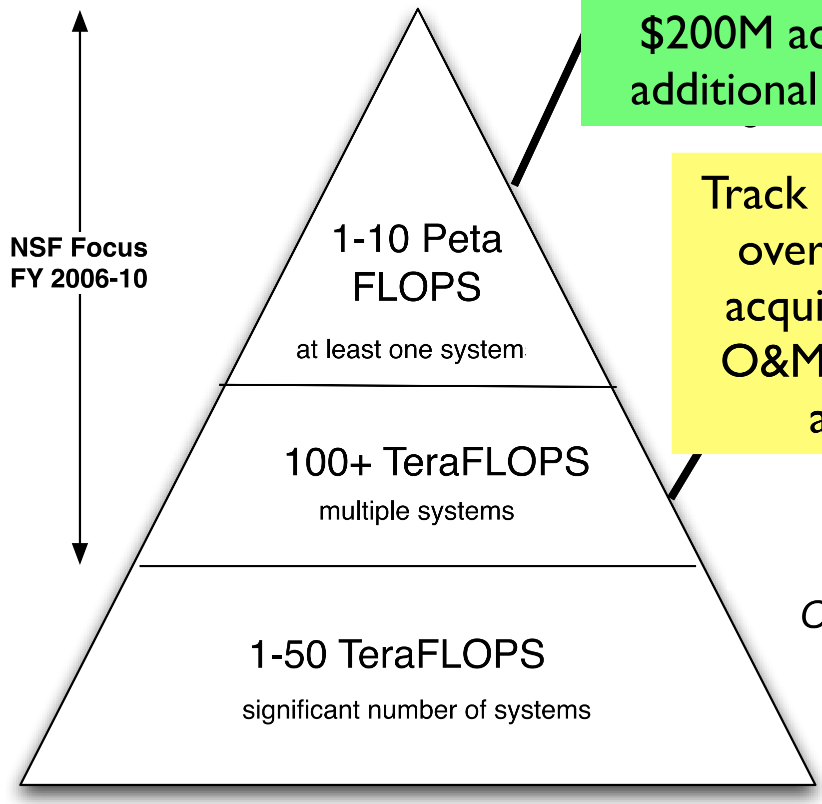


Aldehyde dehydrogenase, T. Wymore and S. Brown

Matter



I. Shipsey

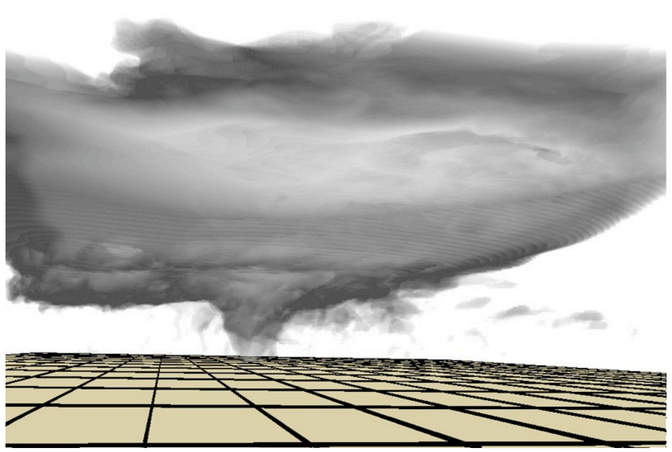


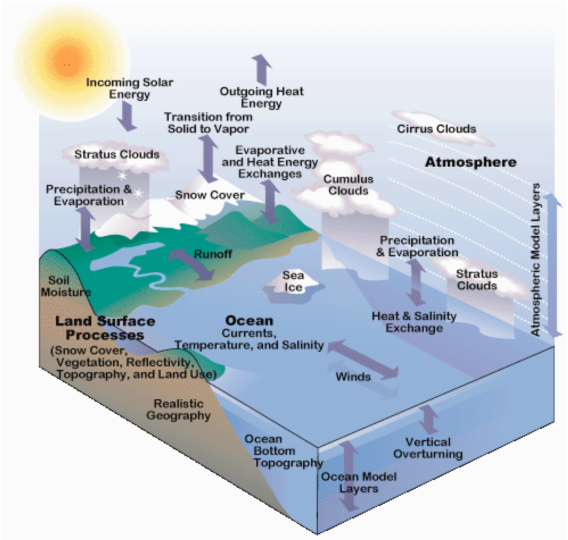
Track 1: One solicitation funded over 4 years: \$200M acquisition + additional O&M cost.

Track 2: Four solicitations over 4 years: \$30M/yr acquisition + additional O&M cost. First track 1 approved 8-07

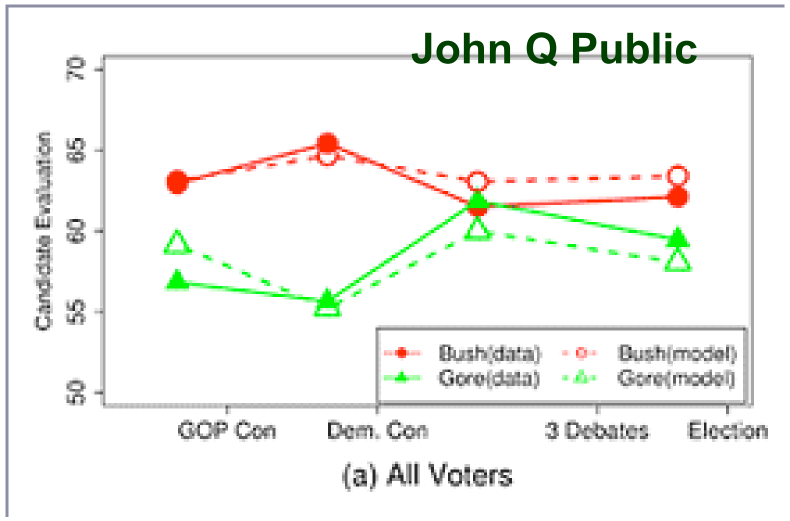
Campus Level

The Environment





Society



S.-Y. Kim, M. Lodge, C. Taber.

National Science Foundation

WHERE DISCOVERIES BEGIN

SEARCH

NSF Web Site

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT | FastLane

News

- News
- For the News Media
- Special Reports
- Research Overviews
- NSF-Wide Investments
- Speeches & Lectures
- Multimedia Gallery
- *NSF Current* Newsletter
- News Archive

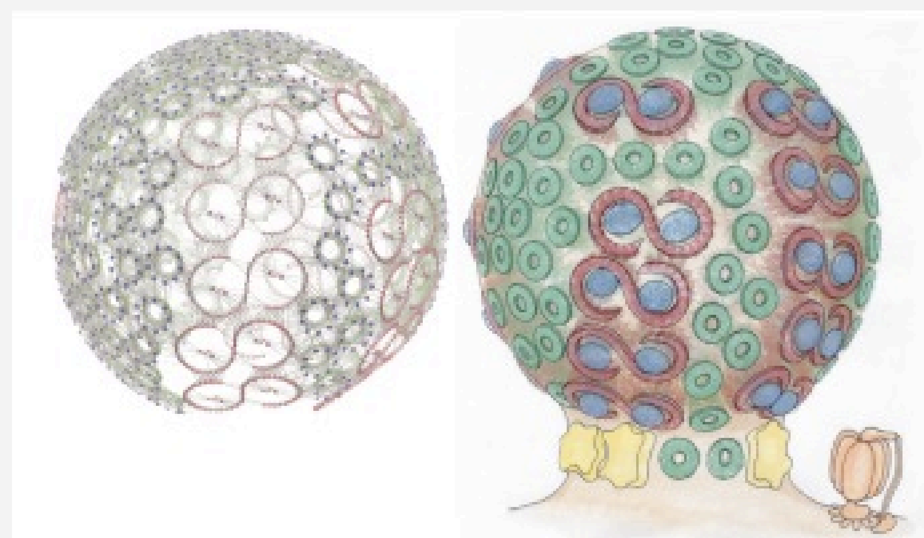
News by Research Area

- Arctic & Antarctic
- Astronomy & Space
- Biology
- Chemistry & Materials

Press Release 06-137

National Science Foundation Awards Texas Advanced Computing Center $59 Million for High-Performance Computing

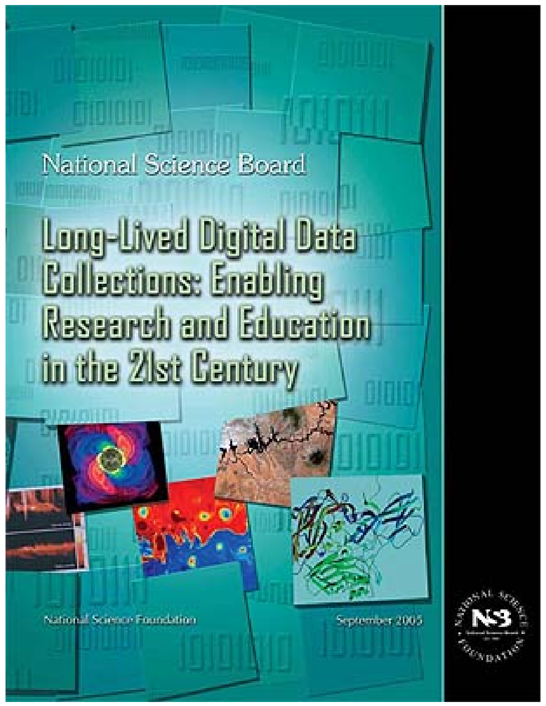
University and industry consortium to deploy powerful general-purpose computing system



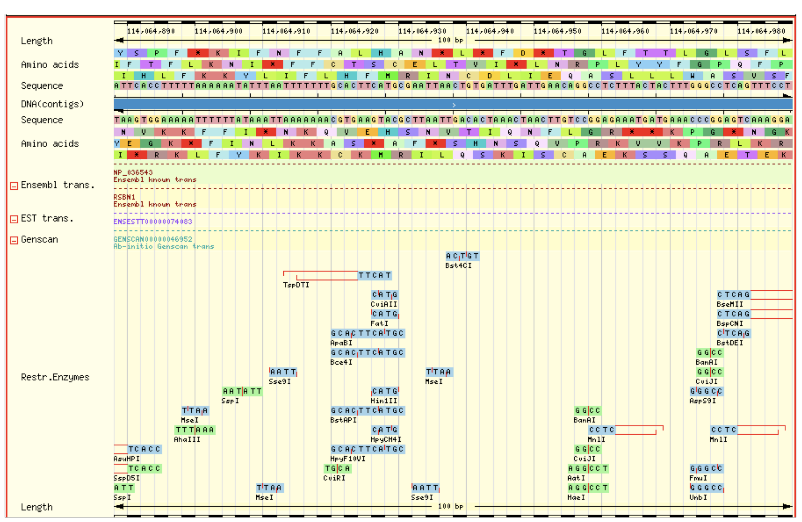
Scientists will use the TACC computer to simulate the 10 milion atoms in this bacterial organelle.
Credit and Larger Version

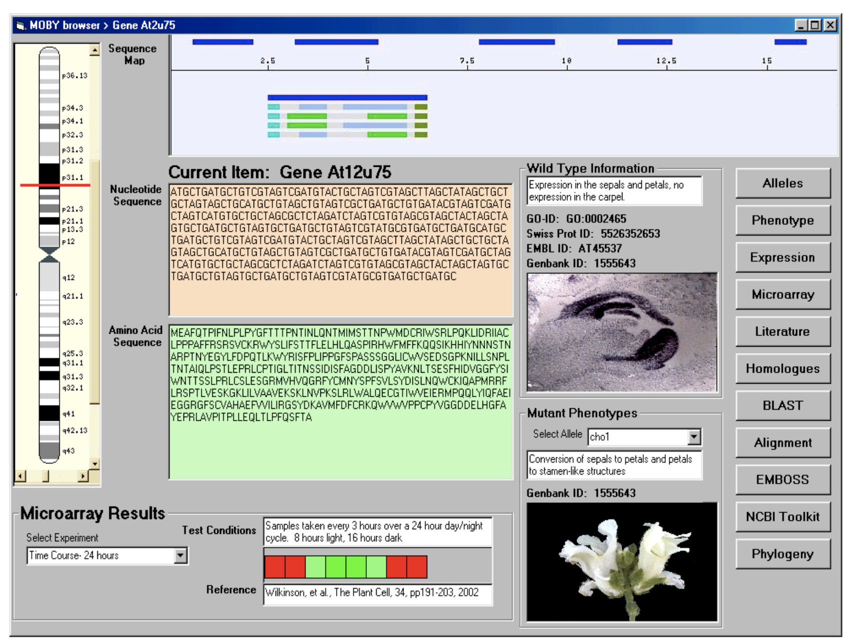
http://www.nsf.gov/news/news_summ.jsp?cntn_id=108051&org=NSF&from=new

Data, Data Analysis & Visualization



- Challenges: **increased scale, heterogeneity, and re-use value** of digital scientific information and data. Inadequate digital preservation strategy of long-lived data.
- Taking initial steps to **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.)
- Complemented by a **new generation of tools** and services to facilitate data mining, integration, analysis, visualization essential to transforming data into knowledge.
- NSF Leadership for OSTP/Interagency Working Group on Digital Data



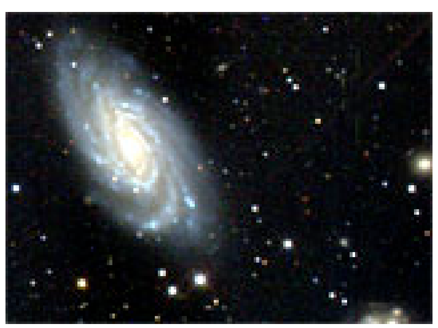


Virtual Organizations

- Distributed virtual organizations are **based upon CI** that provides flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources.
- **Resources and services** include HPC, data/information management, sensor-nets/observatories, linked through global networking and middleware, and accessed by people through web portals and workflow environments.
- Increasing numbers of **virtual organizations are required** by S&E research and education communities. Referred to by many names, e.g. *collaboratory, co-laboratory, grid, gateway, portal,. hub, ...*
- **Challenges** being address include tools for more rapid building and ease of use, interoperability/middleware, high performance, end-to-end networking, and dynamic reconfiguration, social issues, assessment of impact, and economic and technical sustainability.

NVO

LEAD

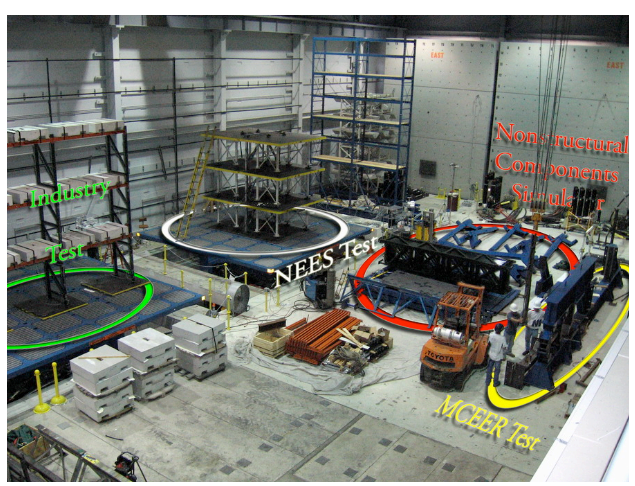


iVDgL

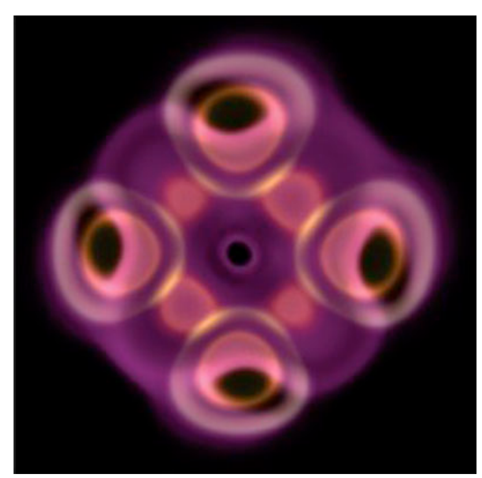
TeraGrid

Enabling Grids for E-sciencE

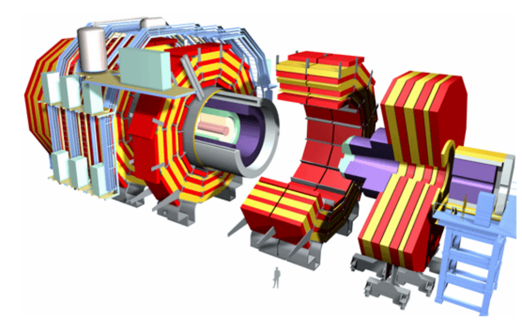
Open Science Grid



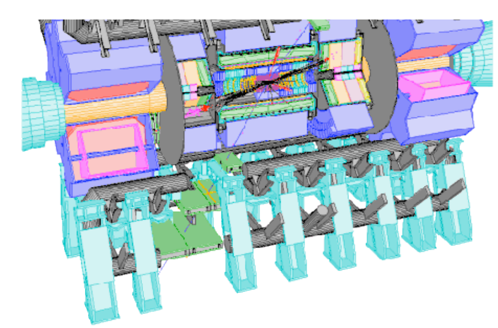
NEES



NanoHub



CMS



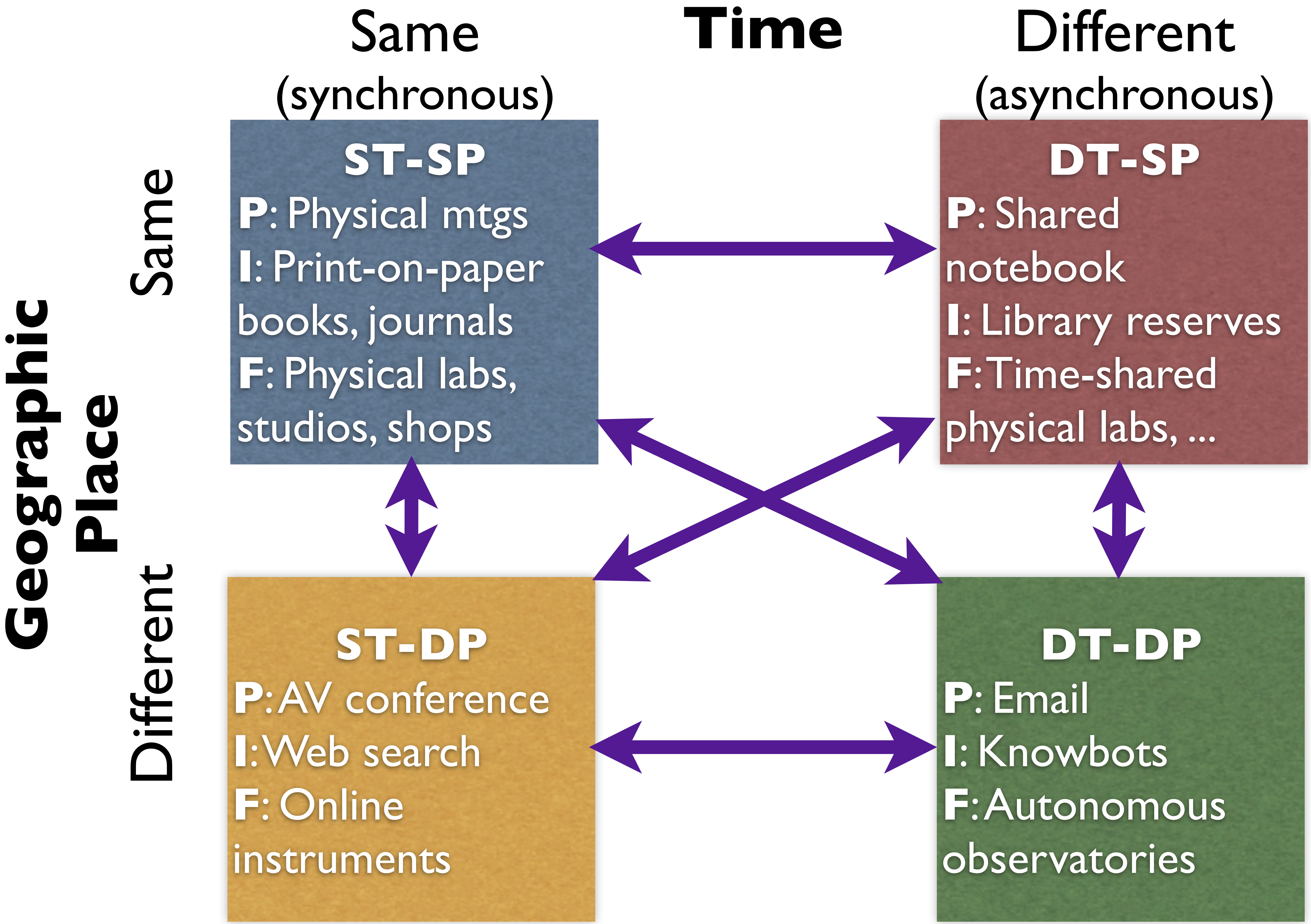
ATLAS

Virtual Organizations offer additional modes of interaction between People, Information, and Facilities

Geographic Place	Time: Same (synchronous)	Time: Different (asynchronous)
Same	**ST-SP** **P**: Physical mtgs **I**: Print-on-paper books, journals **F**: Physical labs, studios, shops	**DT-SP** **P**: Shared notebook **I**: Library reserves **F**: Time-shared physical labs, ...
Different	**ST-DP** **P**: AV conference **I**: Web search **F**: Online instruments	**DT-DP** **P**: Email **I**: Knowbots **F**: Autonomous observatories

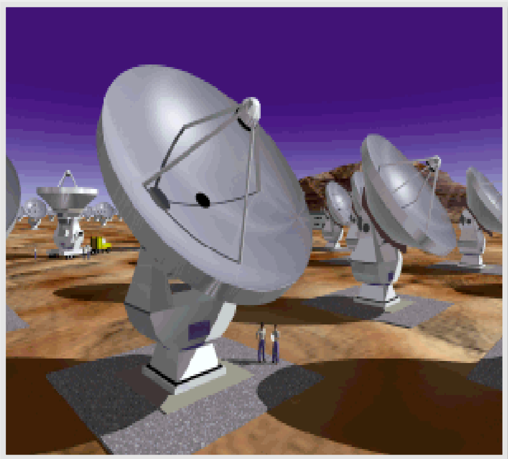
P: people, **I**: information, **F**: facilities, instruments

Virtual Organizations offer additional modes of interaction between People, Information, and Facilities

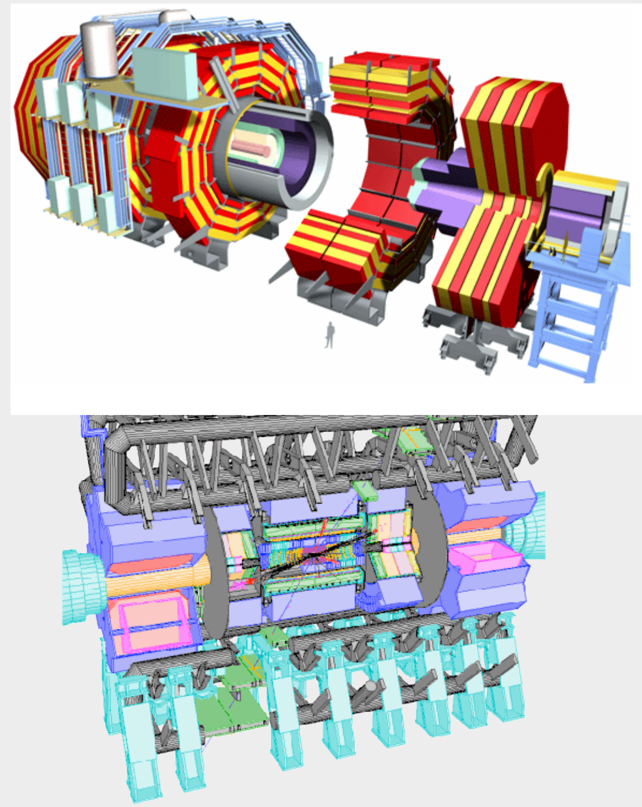


P: people, **I**: information, **F**: facilities, instruments

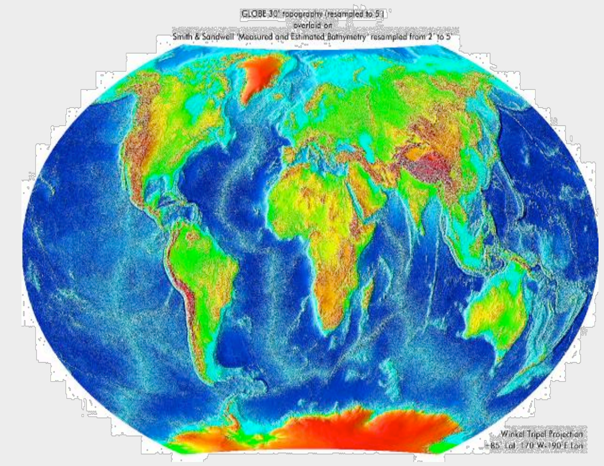
CI/VO Enabled Science



NVO and ALMA



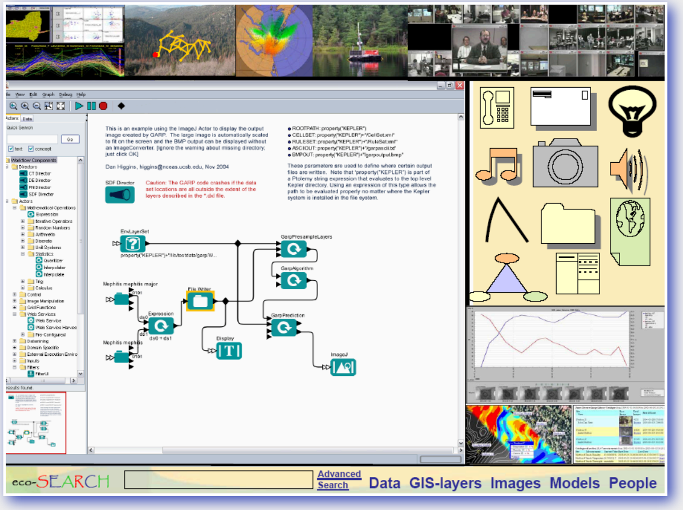
ATLAS and CMS



Climate Change

LIGO

United States virtual observatory



NEON

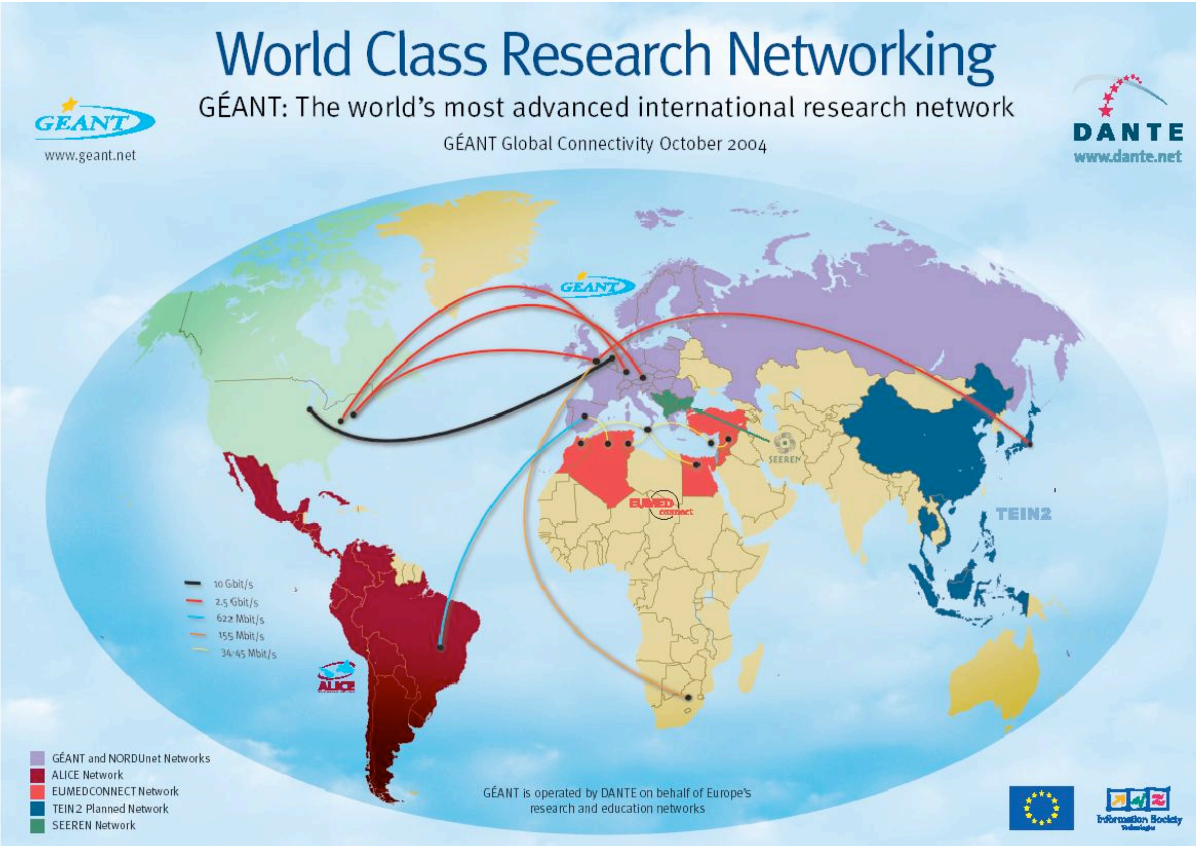
The number of nation-scale projects is growing rapidly!

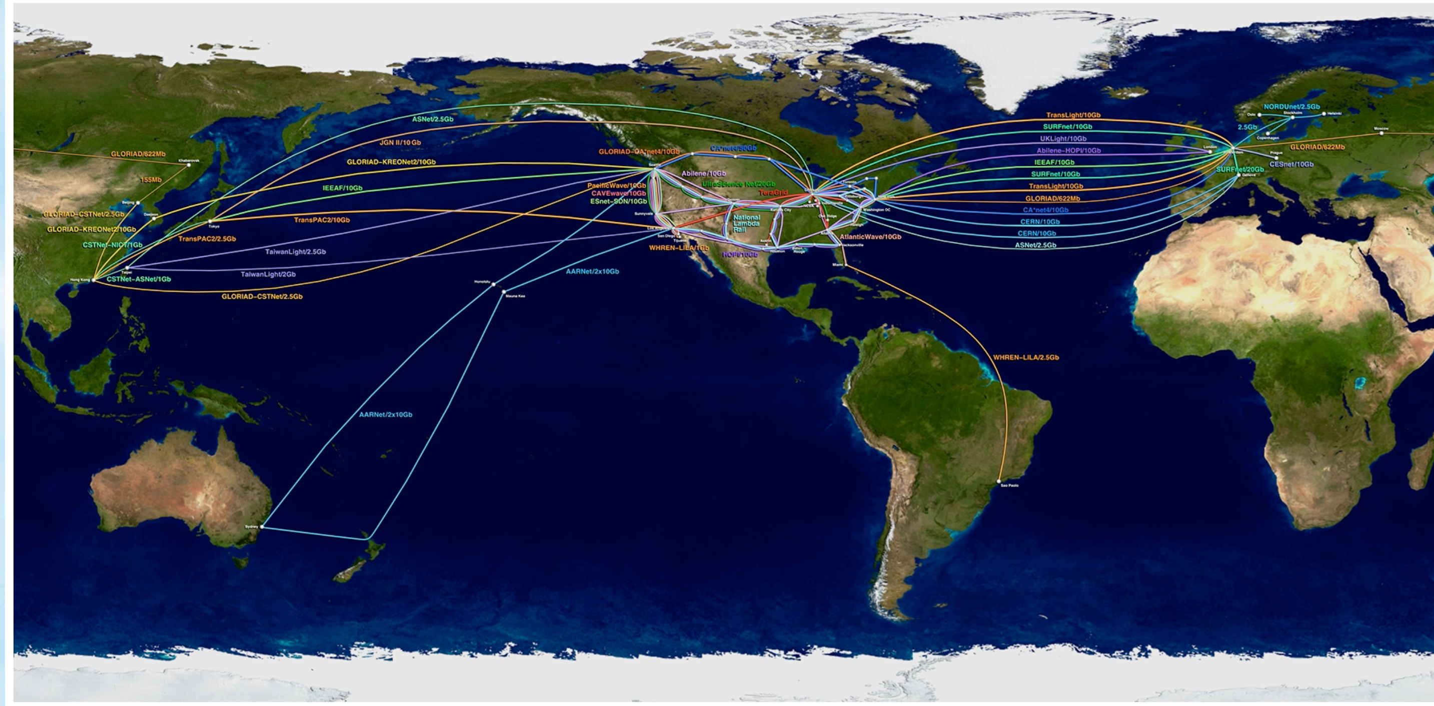
Need a socio-technical approach: Realizing the potential of e- | cyber science to support effective VOs requires attention to *institutional infrastructure*

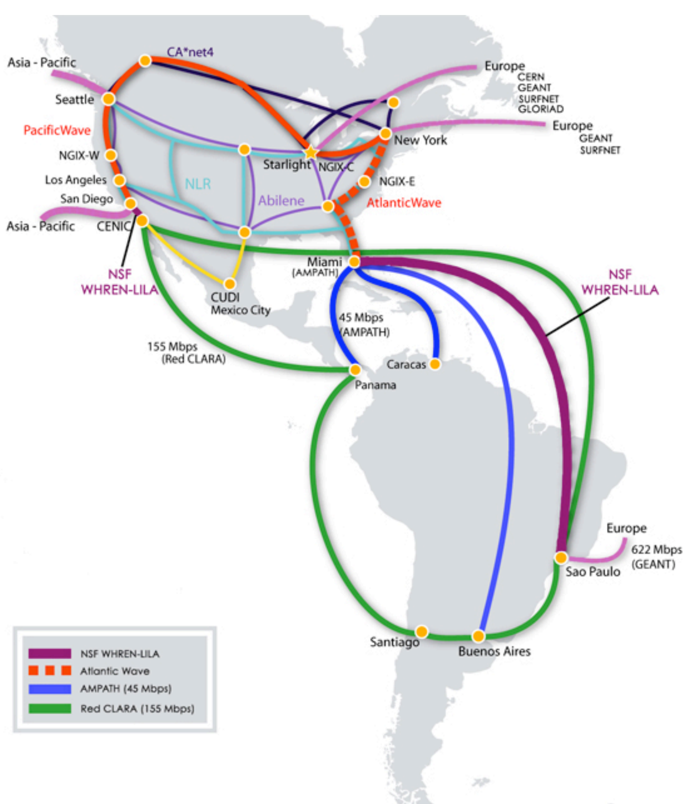
- Need more than good ICT systems and tools for individuals and organizations. (Technology determinism alone is not enough).
- No less important is the *institutional contexts* (i.e. norms of practice and rules) to facilitate collaboration within science and technical research communities.
- The institutional and organizational environment of e-science encompasses a wide and diverse array of interrelated social, economic, and legal factors that
  - create incentives for, and constraints upon individual and collective action; and
  - thereby shape the production, utilization, consumption, and governance of e-science capabilities and products.

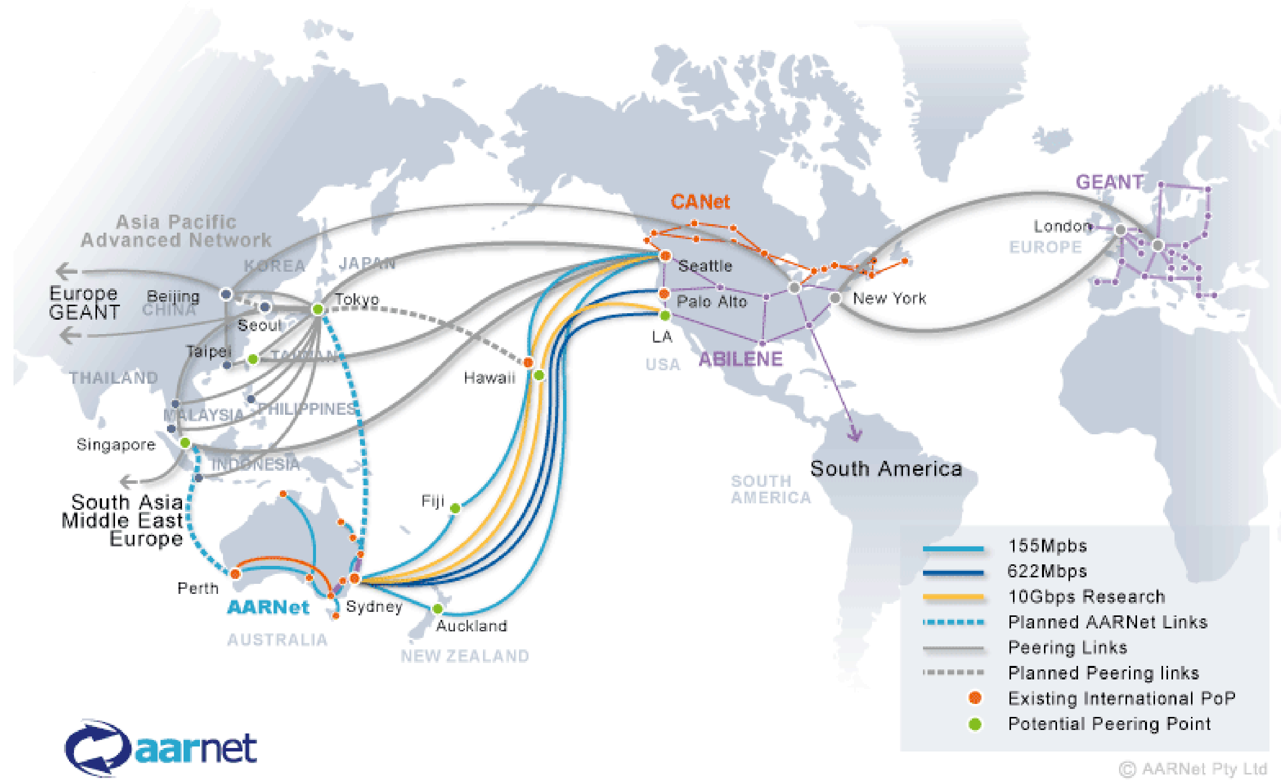
Adapted from Paul David, see www.oii.ox.ac.uk/resources/publications/RR2.pdf

VO-substrate: International R&E Networking





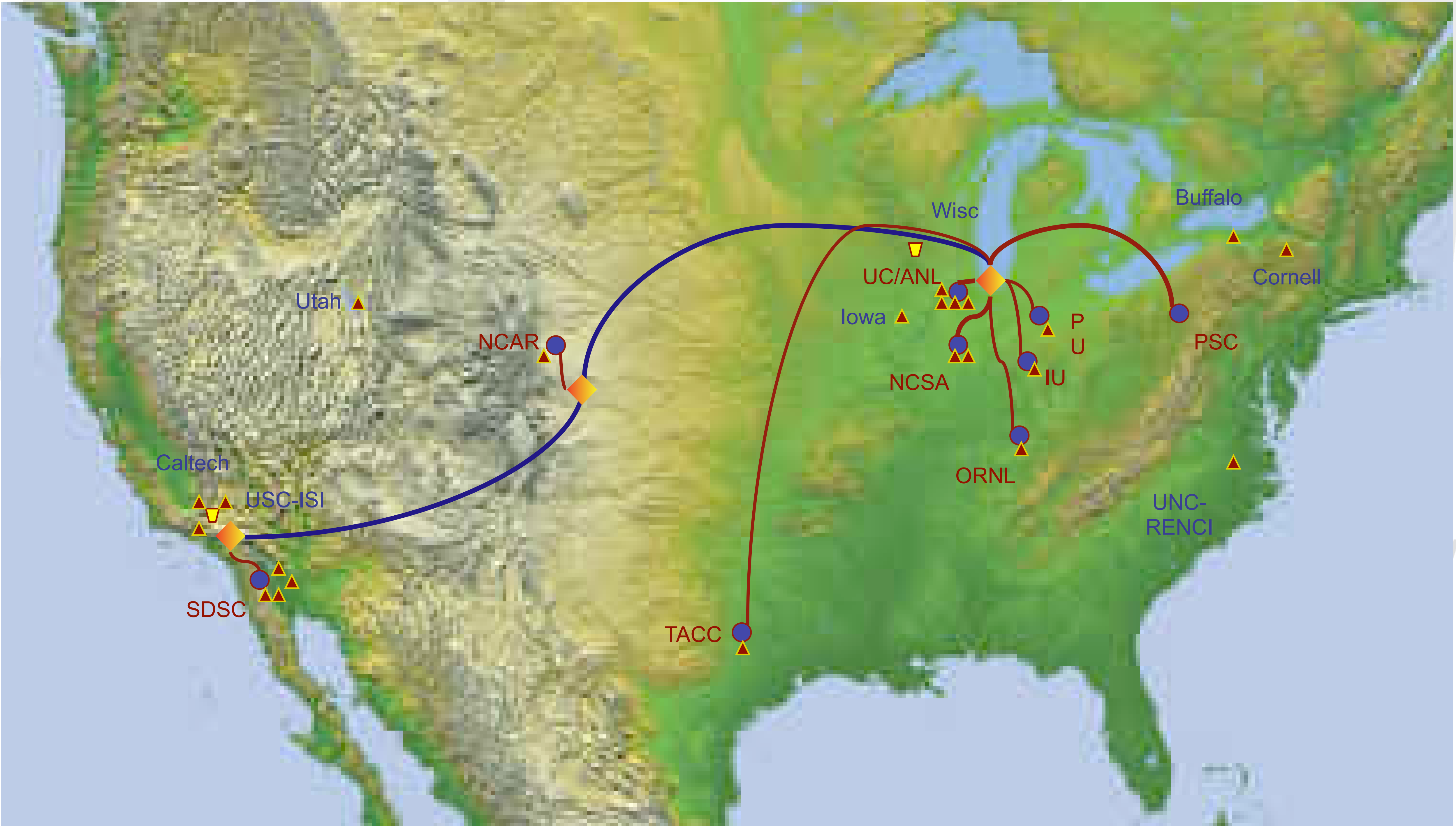




Office of Cyberinfrastructure

D. E. Atkins

TeraGrid: Integrating NSF Cyberinfrastructure



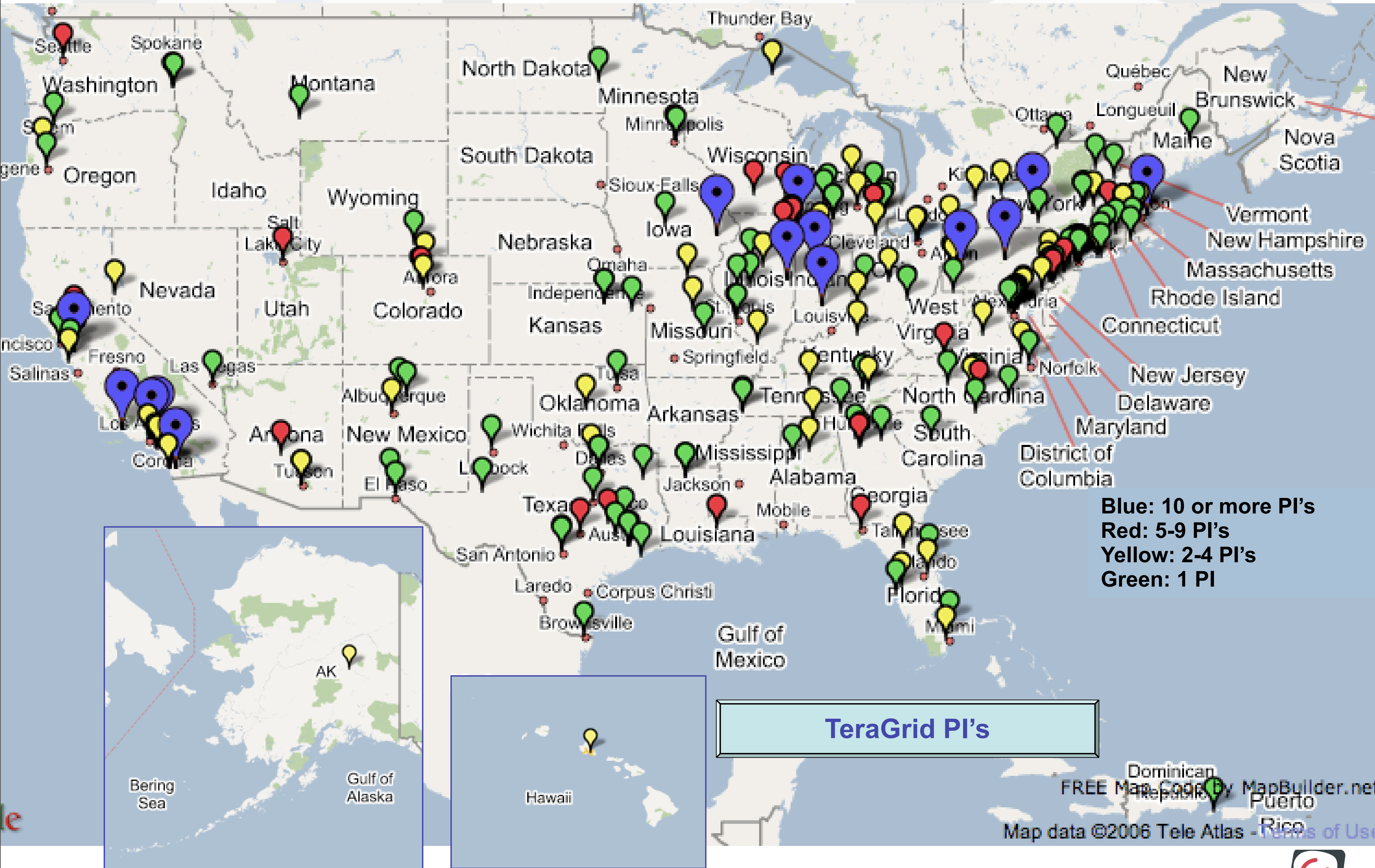
TeraGrid is a facility that integrates computational, information, and analysis resources at the San Diego Supercomputer Center, the Texas Advanced Computing Center, the University of Chicago / Argonne National Laboratory, the National Center for Supercomputing Applications, Purdue University, Indiana University, Oak Ridge National Laboratory, the Pittsburgh Supercomputing Center, and the National Center for Atmospheric Research.

Charlie Catlett (cec@uchicago.edu)

June 2006

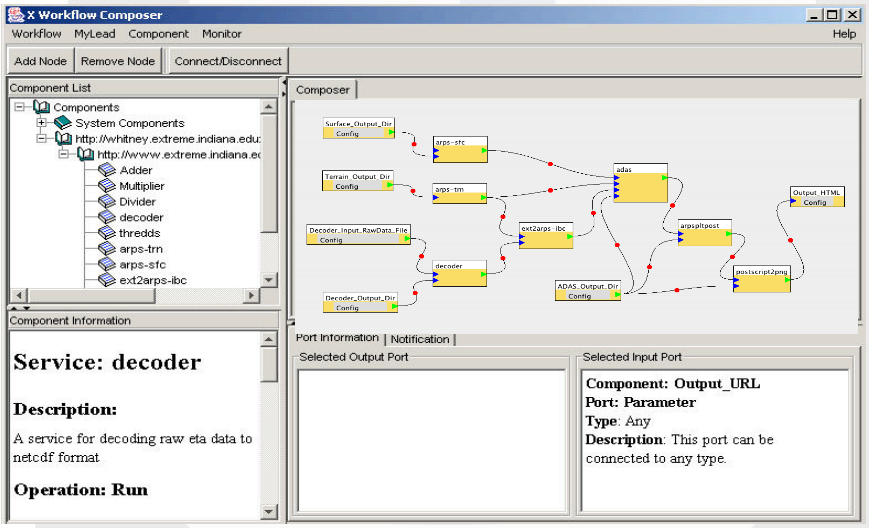
TeraGrid™

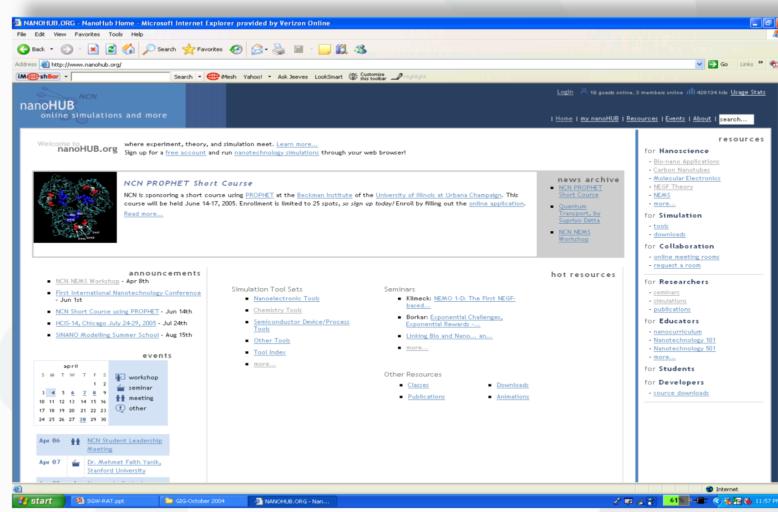
TeraGrid PI's By Institution as of May 2006

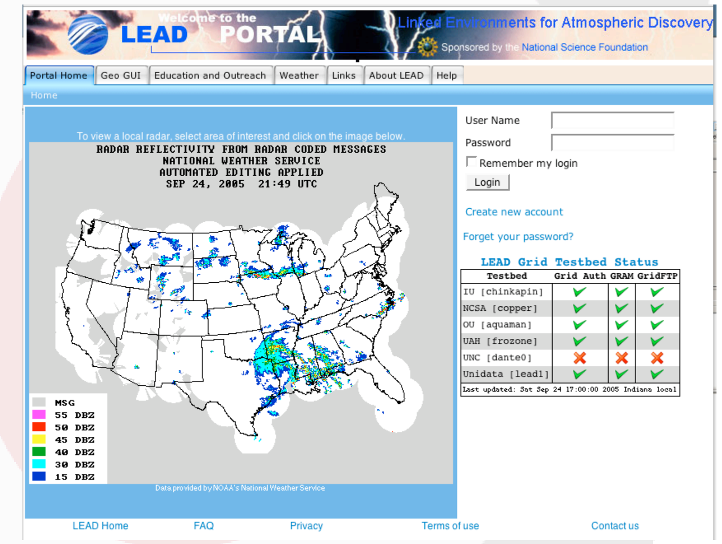


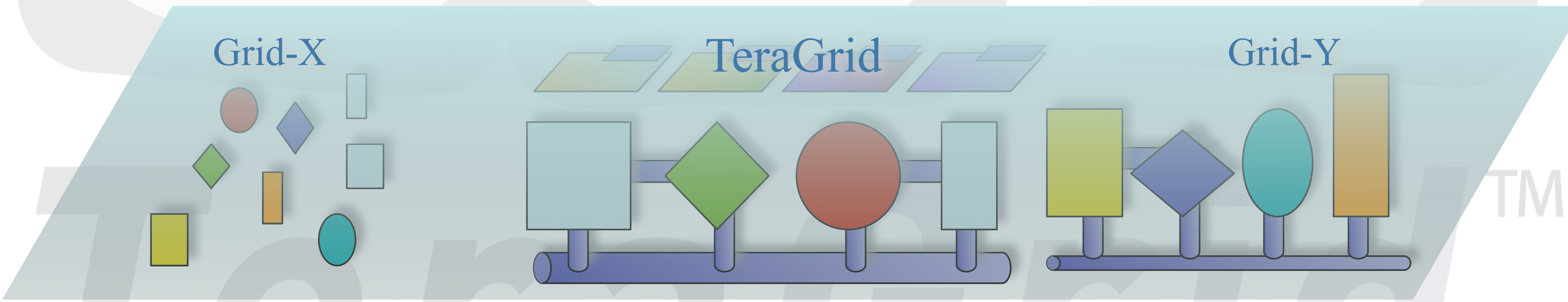
Charlie Catlett (cec@uchicago.edu)

TeraGrid Science Gateways Initiative: Community Interface to Grids









- **Common Web Portal or application interfaces (database access, computation, workflow, etc).**
- **"Back-End" use of TeraGrid computation, information management, visualization, or other services.**
- **Standard approaches so that science gateways may readily access resources in any cooperating Grid without technical modification.**

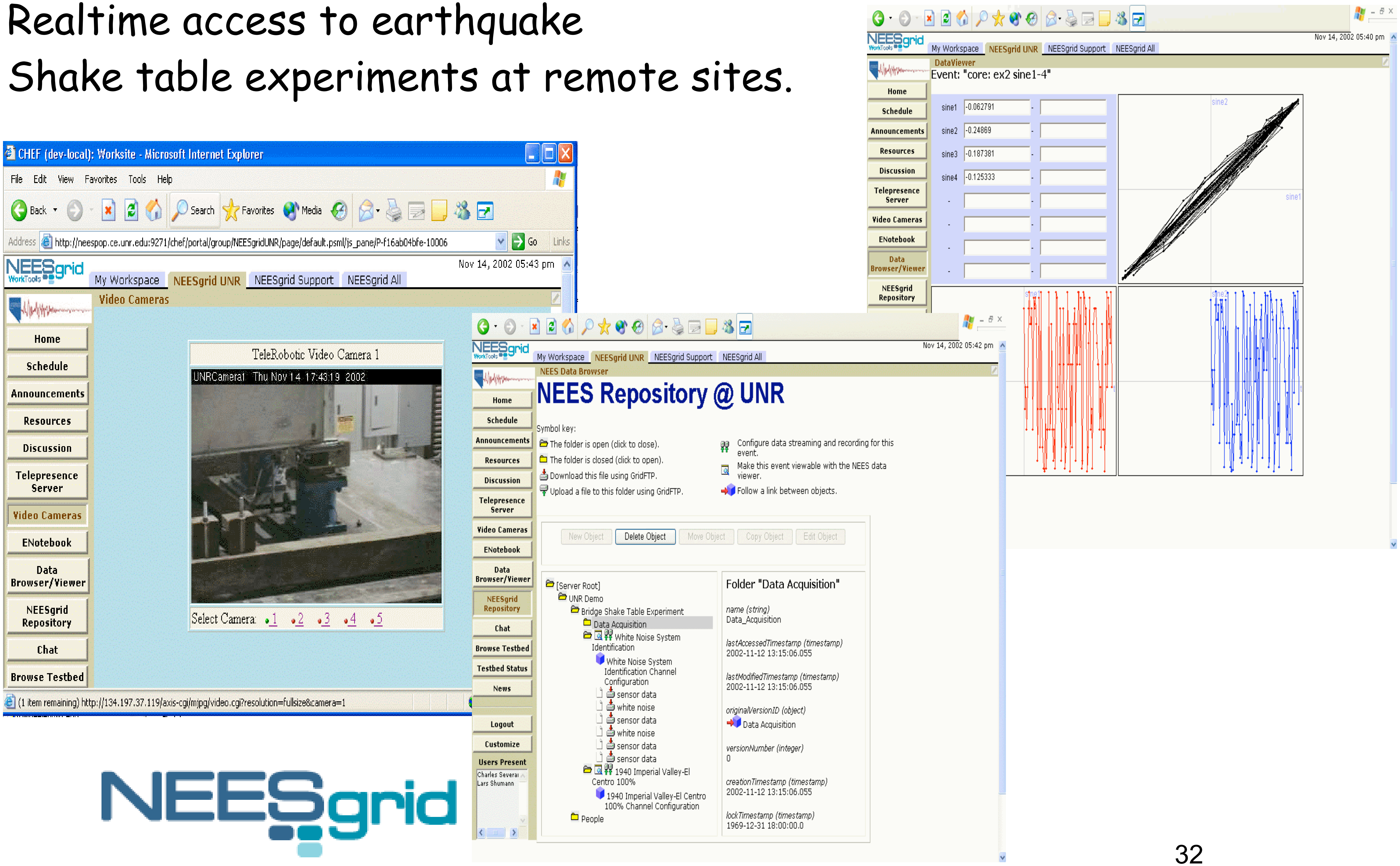
Let’s look at a few real example Grid Science Gateways

(about a dozen … many more exist!)

These example slides courtesy of D. Gannon

NEESGrid

Realtime access to earthquake
Shake table experiments at remote sites.



BIRN – Biomedical Information

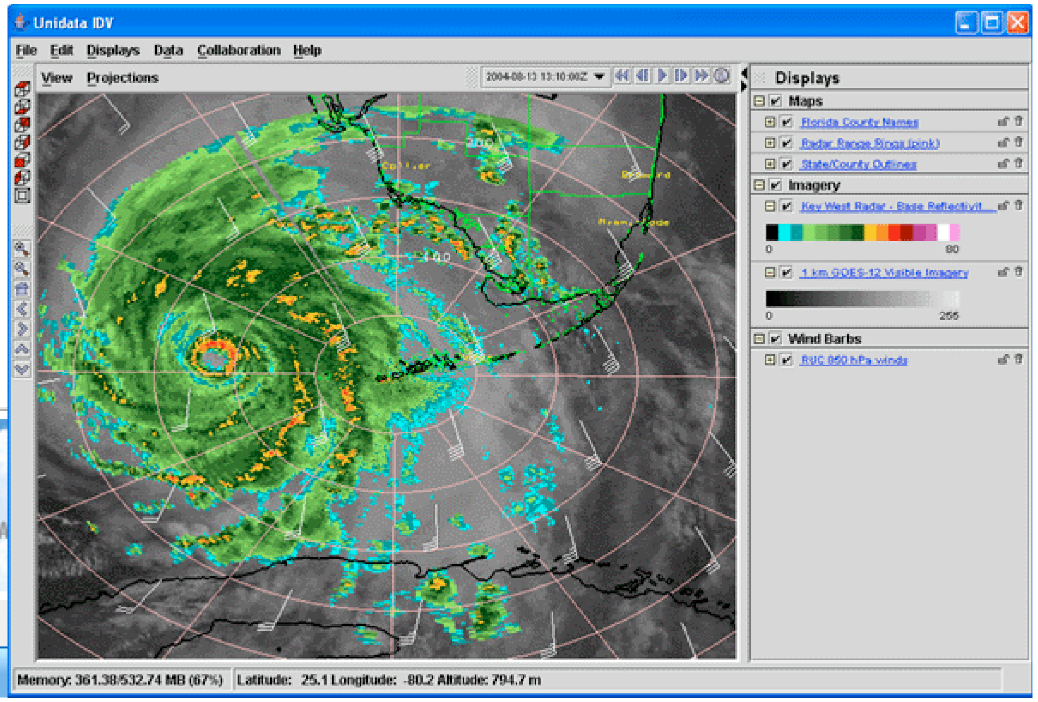
33

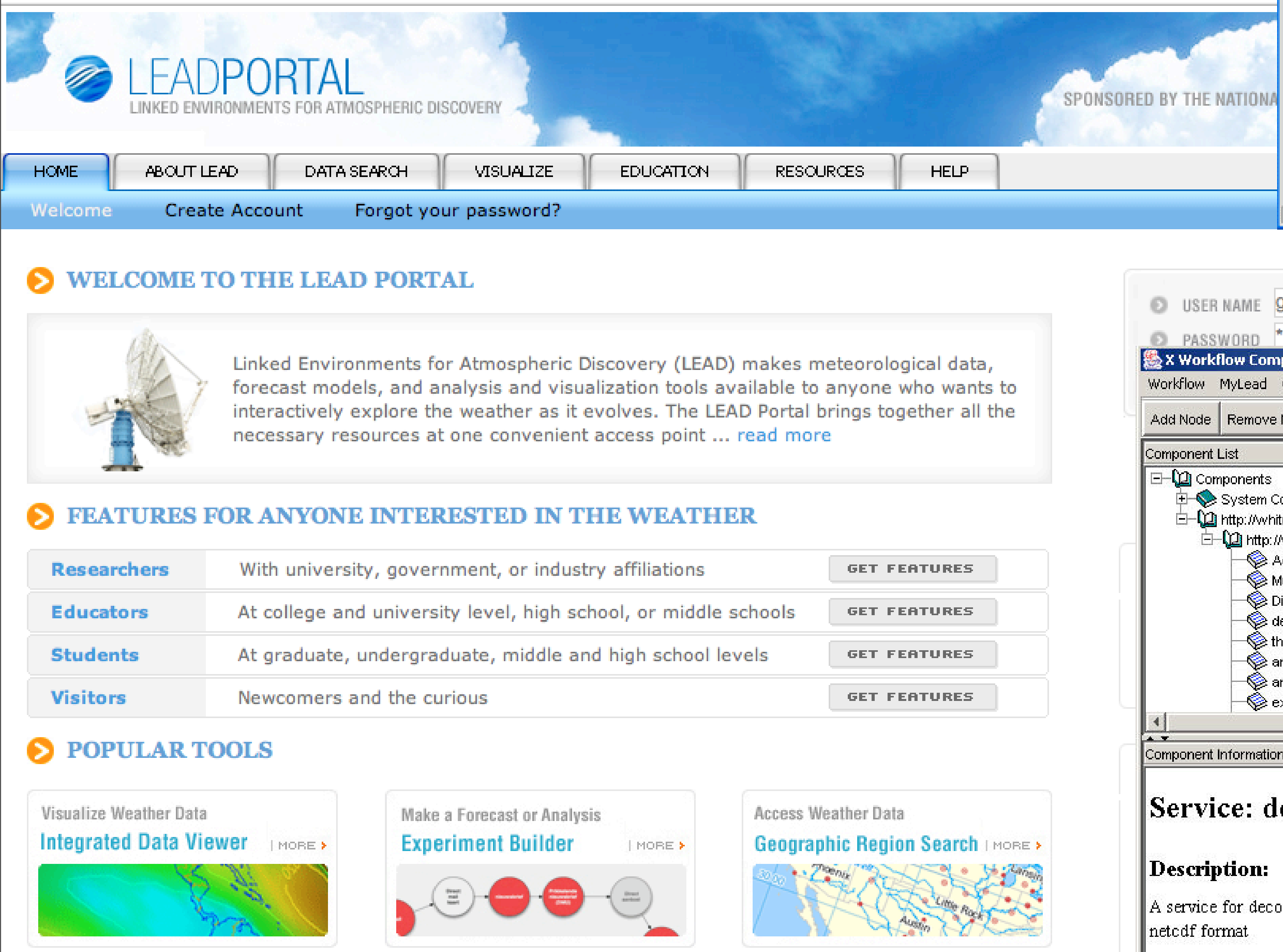
Geological Information Grid Portal

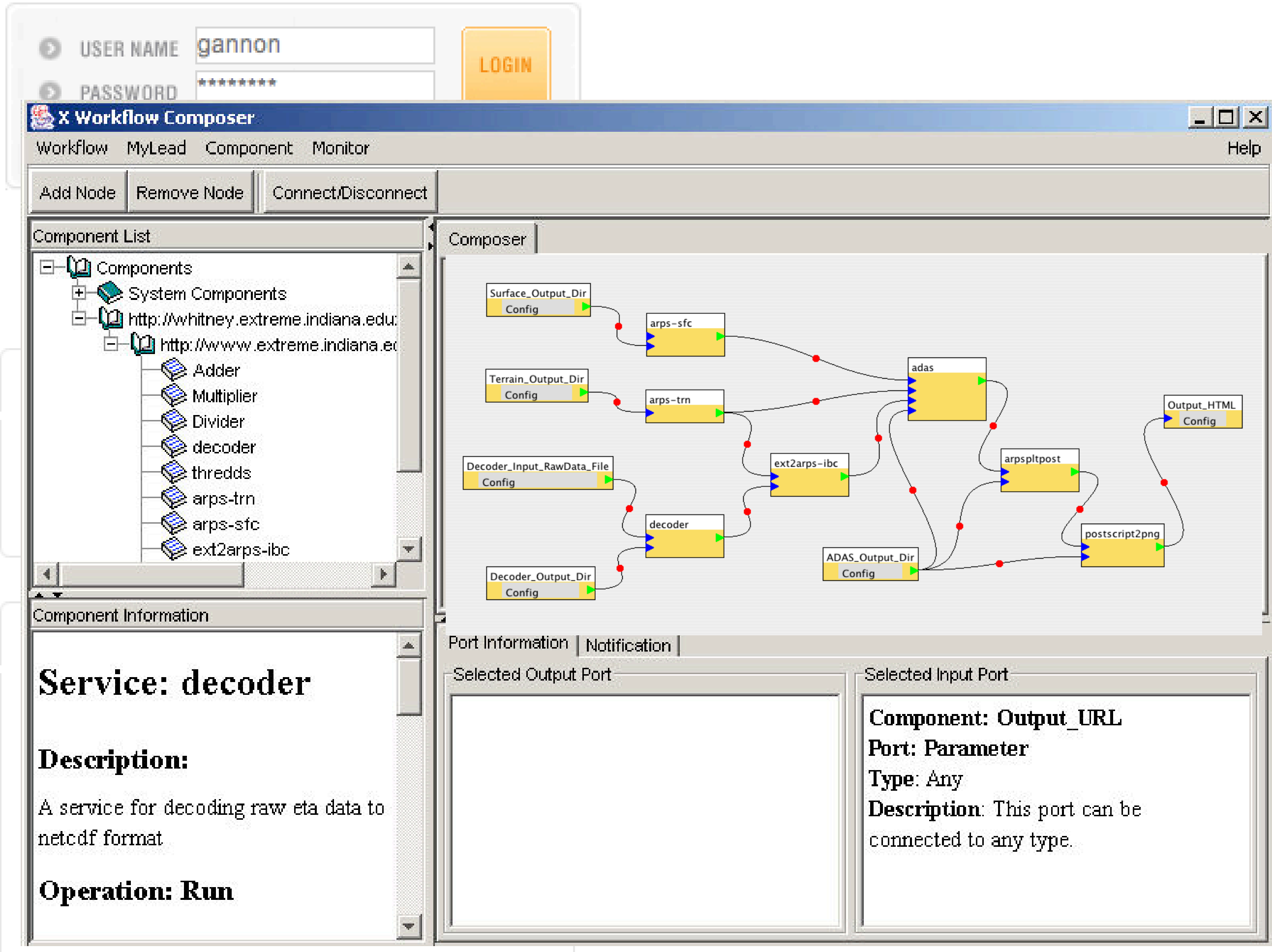
Mesoscale Meteorology

NSF LEAD project - making the tools that are needed to make accurate predictions of tornados and hurricanes.

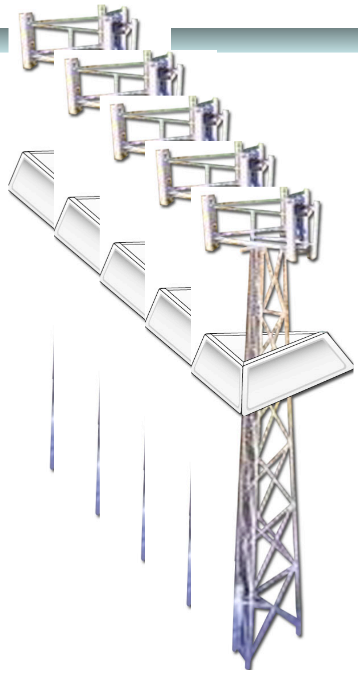
- Data exploration and Grid workflow

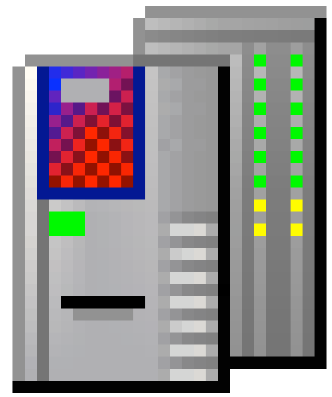


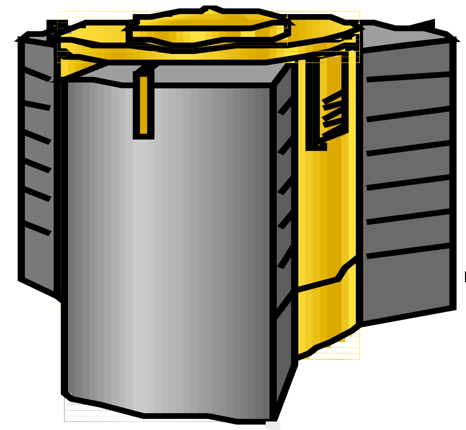


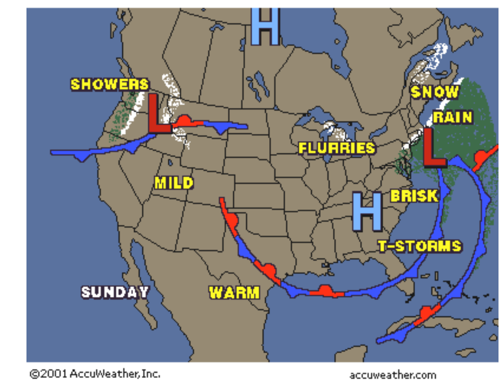


The LEAD Vision: Adaptive Cyberinfrastructure









DYNAMIC OBSERVATIONS

Analysis/Assimilation

Quality Control
Retrieval of Unobserved Quantities
Creation of Gridded Fields

Prediction/Detection

PCs to Teraflop Systems

Product Generation, Display, Dissemination

Models and Algorithms Driving Sensors

End Users

NWS
Private Companies
Students

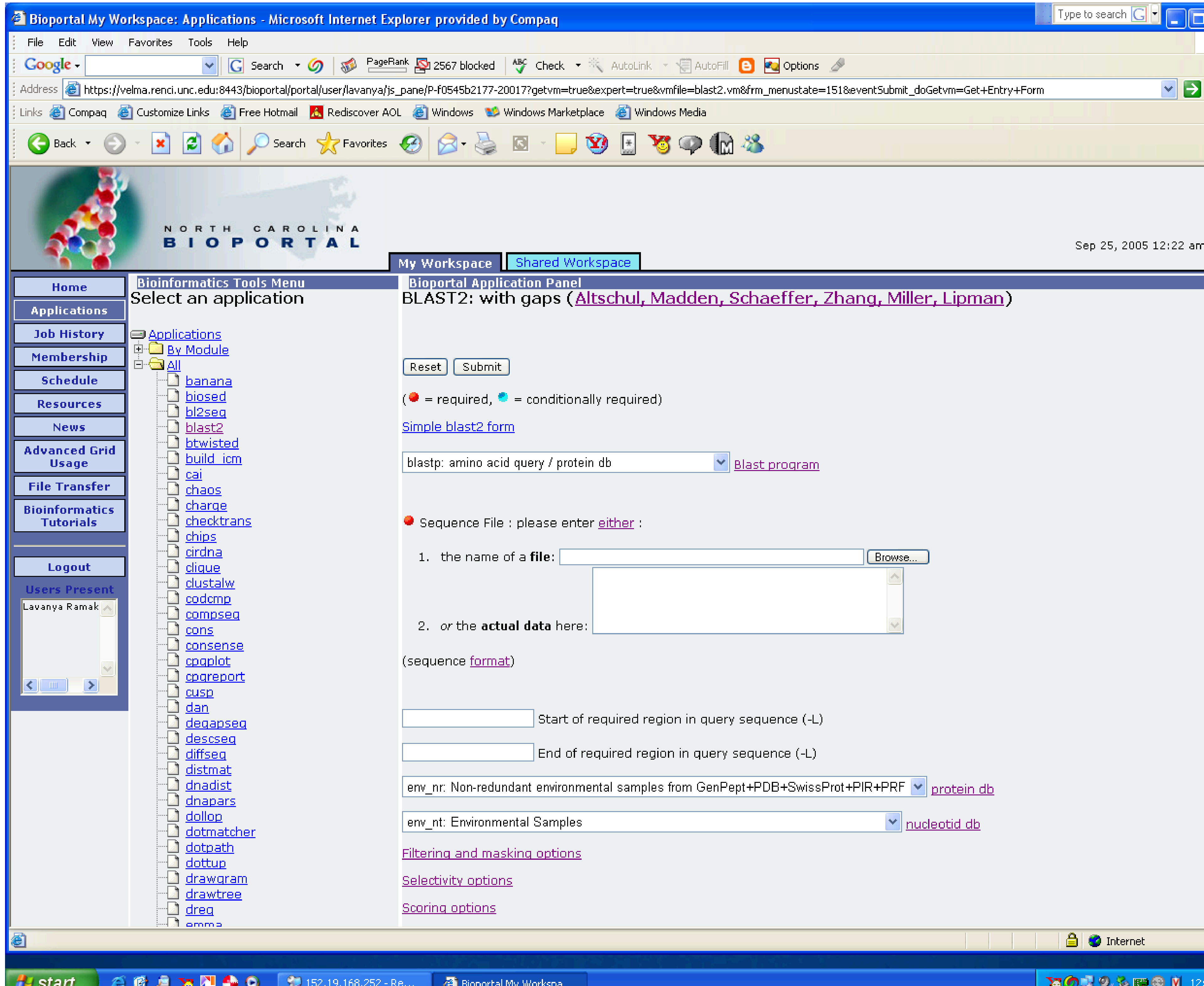
The CS challenge: Build cyberinfrastructure services that provide adaptability, scalability, availability, useability, and real-time response.

From D. Gannon

Renci Bio Portal

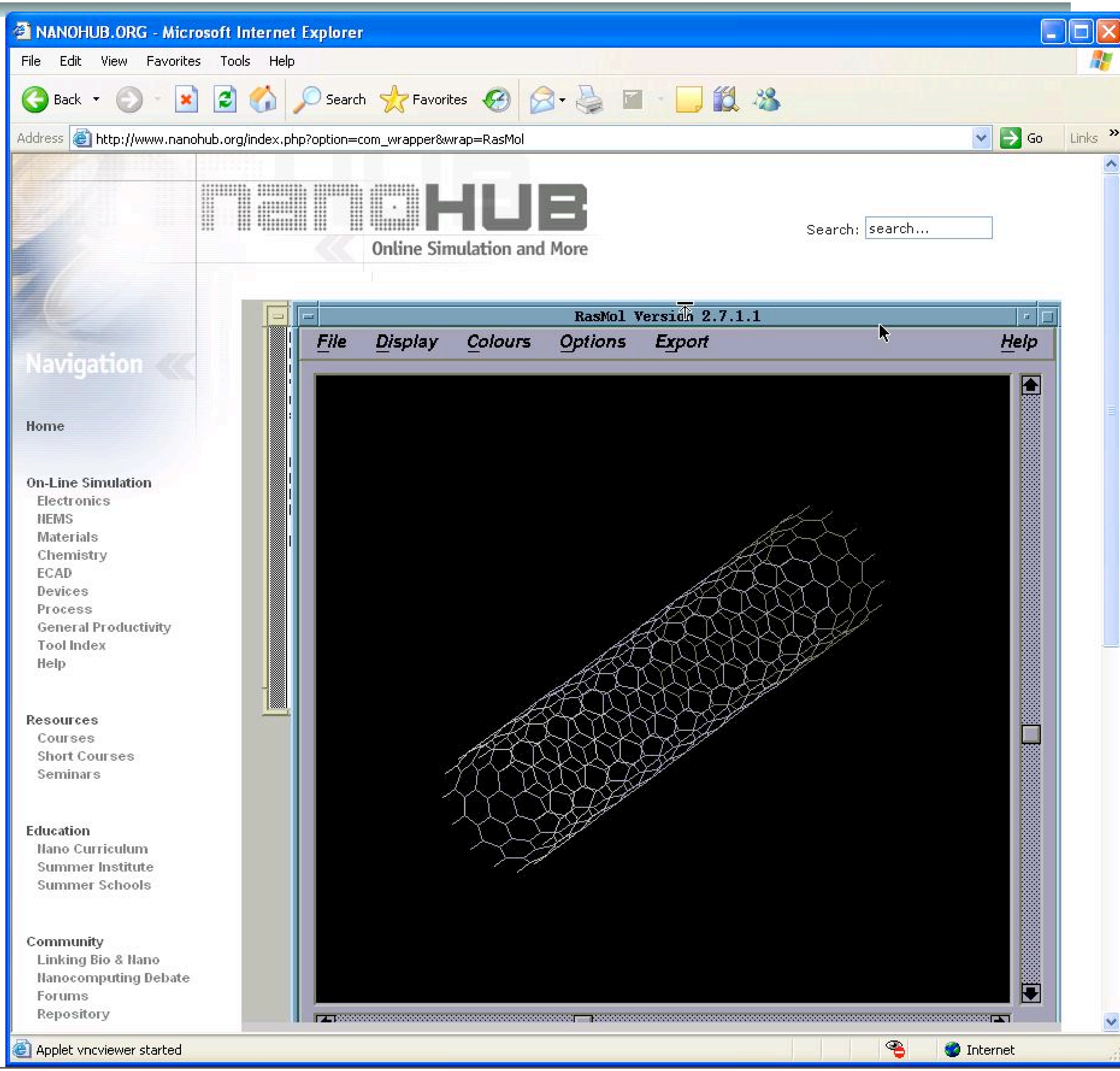
Providing access to biotechnology tools running on a back-end Grid.

- leverage state-wide investment in bioinformatics
- undergraduate & graduate education, faculty research
- another portal soon: national evolutionary synthesis center



NESCent

Nanohub - nanotechnology



X-Ray Crystallography

C.I.M.A.
Common Instrument Middleware Architecture

Welcome to the Crystallography Portal

Username:
Password:
Login
Remember me on this computer
Login Help

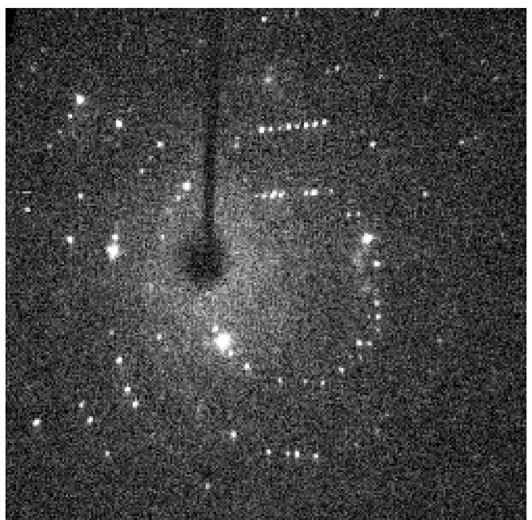
Home | Current Status | Data Repository | About

- Home
- IUB IUMSC
- IUB Myers Hall
- **Purdue Crystallography Center**
- CSAF Sydney, Australia
- Minnesota X-ray Lab
- ChemMatCARS - Univ. of Chicago at APS
- Other collaborators
- NCS Southampton, UK

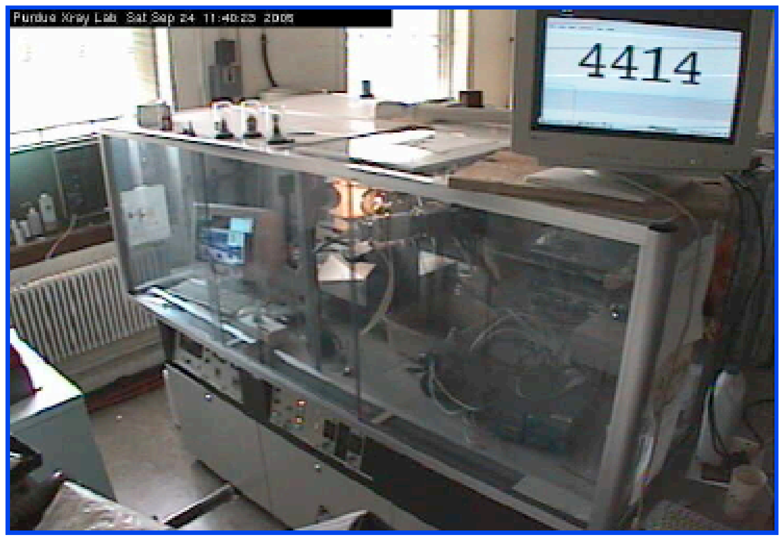
The Purdue Chemistry Crystallography Center

The Purdue Chemistry Crystallography Center

Disable your browser's cache to get the live stream!



Data from Nonius Kappa CCD detector
(Under development!)
Total Number of jpg: 10
Frame: s01f0010.jpg
All available jpg images
Browse the 20 latest jpg images



<□>

Streaming video from the lab showing the Nonius instrument

<□>

Streaming video from the crystal microscope on the Noniu diffractometer

Local date/time: 2005-09-24 11:36:54

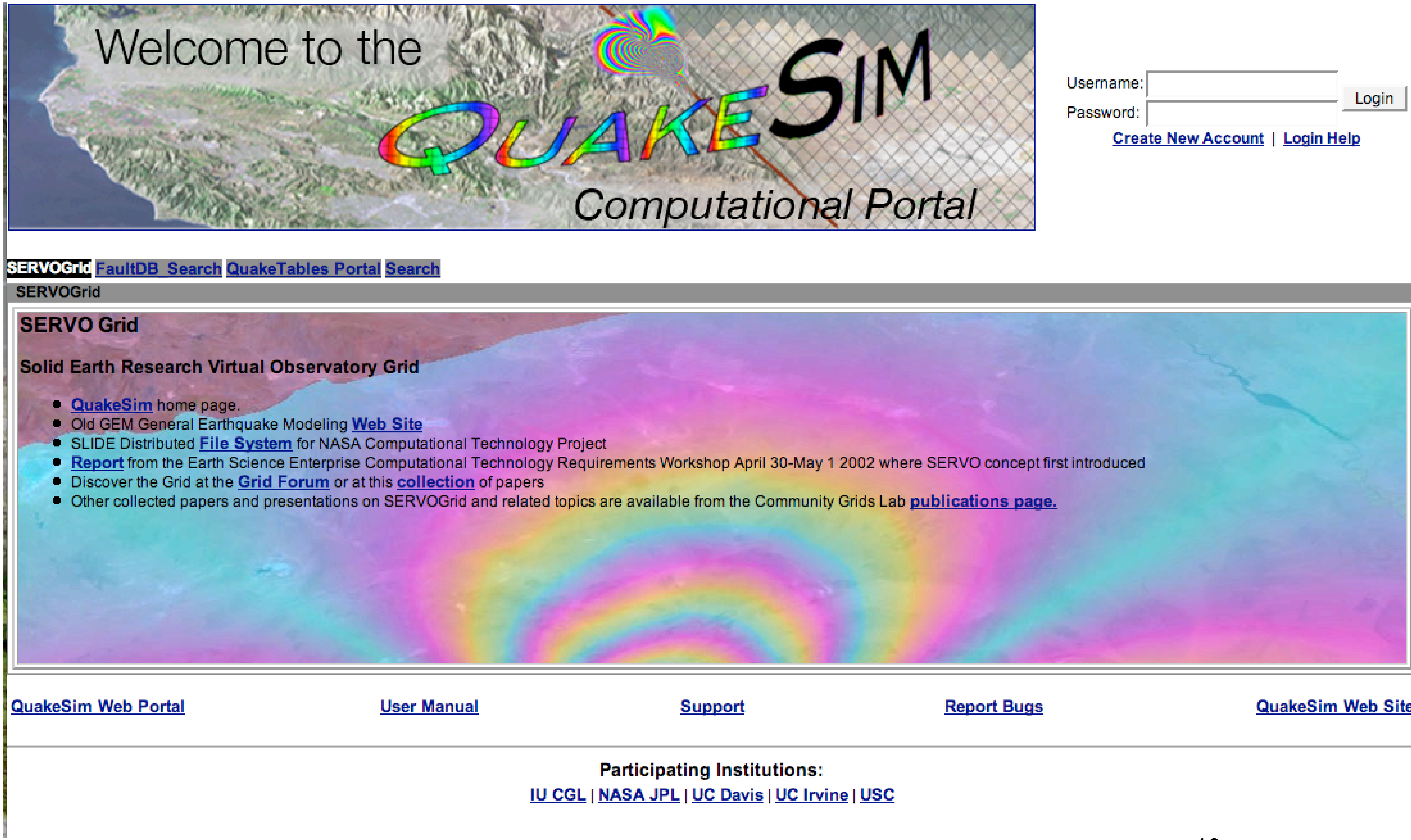
These values are updated approx. every 60 sec.

Times in UTC

LabJack U12

Instrument Enclosure Temp. & Humidity:	23.4 C Rel. Humid. 43.1 %	2005-09-24 16:35:59
Chill Water In:	16.4 C	2005-09-24 16:36:25
Chill Water Out:	19.3 C	2005-09-24 16:36:25
Generator Relay Voltage: All previous voltages	3.42 X-ray Generator is: OFF	2005-09-24 16:36:48

ServoGrid Portal



40

Belfast Gene Grid Portal

41

MyGrid - Bioinformatics

myGrid

Navigate

- Home
- About
- Downloads
- Components
  - Component Overview
  - Research Components
- Using myGrid
- Research Using myGrid
- Links
- Publications
- Contact

Log In

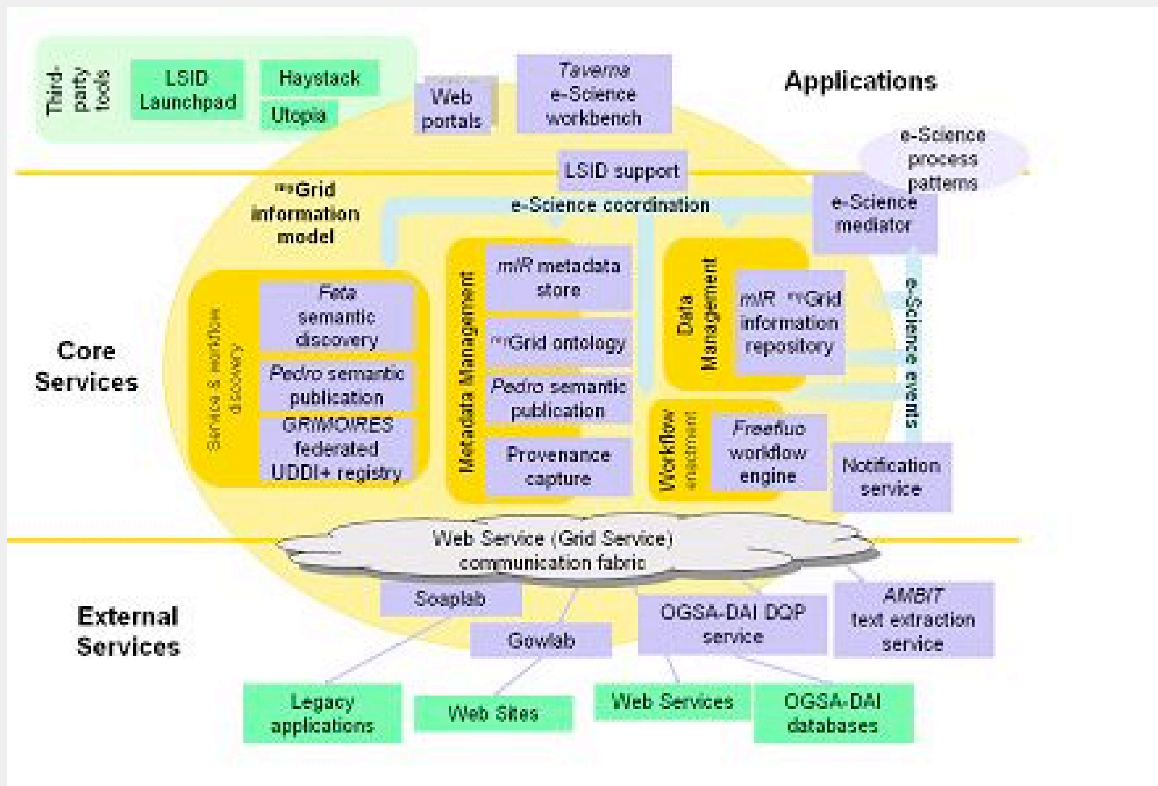
Username

Password

Log In

New Account Signup
Forgot Password

myGrid Architecture



myGrid components - overview

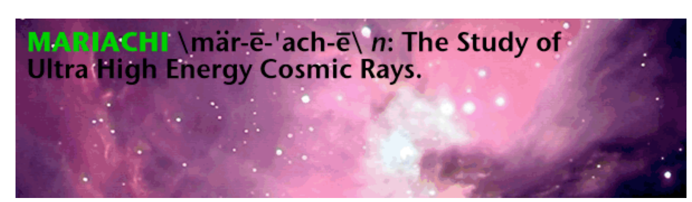
myGrid is a collection of services and components that allows the high level integration of biological applications. The architecture provides the infrastructure necessary, in a web service environment, t e-science workbench that actively supports the scientific lifecycle. Each component or service contri system that allows the e-scientist to perform complex in-silico experiments across distributed bioinfo resources.

Learning & Workforce Development

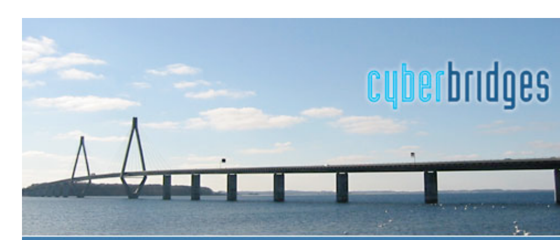
- Learning **supported by** CI. (cyber-enabled learning).
- Workforce development **to create and use** CI for S&E research and education.
- **Broadened participation**: Exploit the new opportunities that cyberinfrastructure brings for … people who, because of physical capabilities, location, or history, have been excluded from the frontiers of scientific and engineering research and education.
- Explore CI support for **integrated research and education**.

BIOINFORMATICS CI INSTITUTE

EPIC



MARIACHI

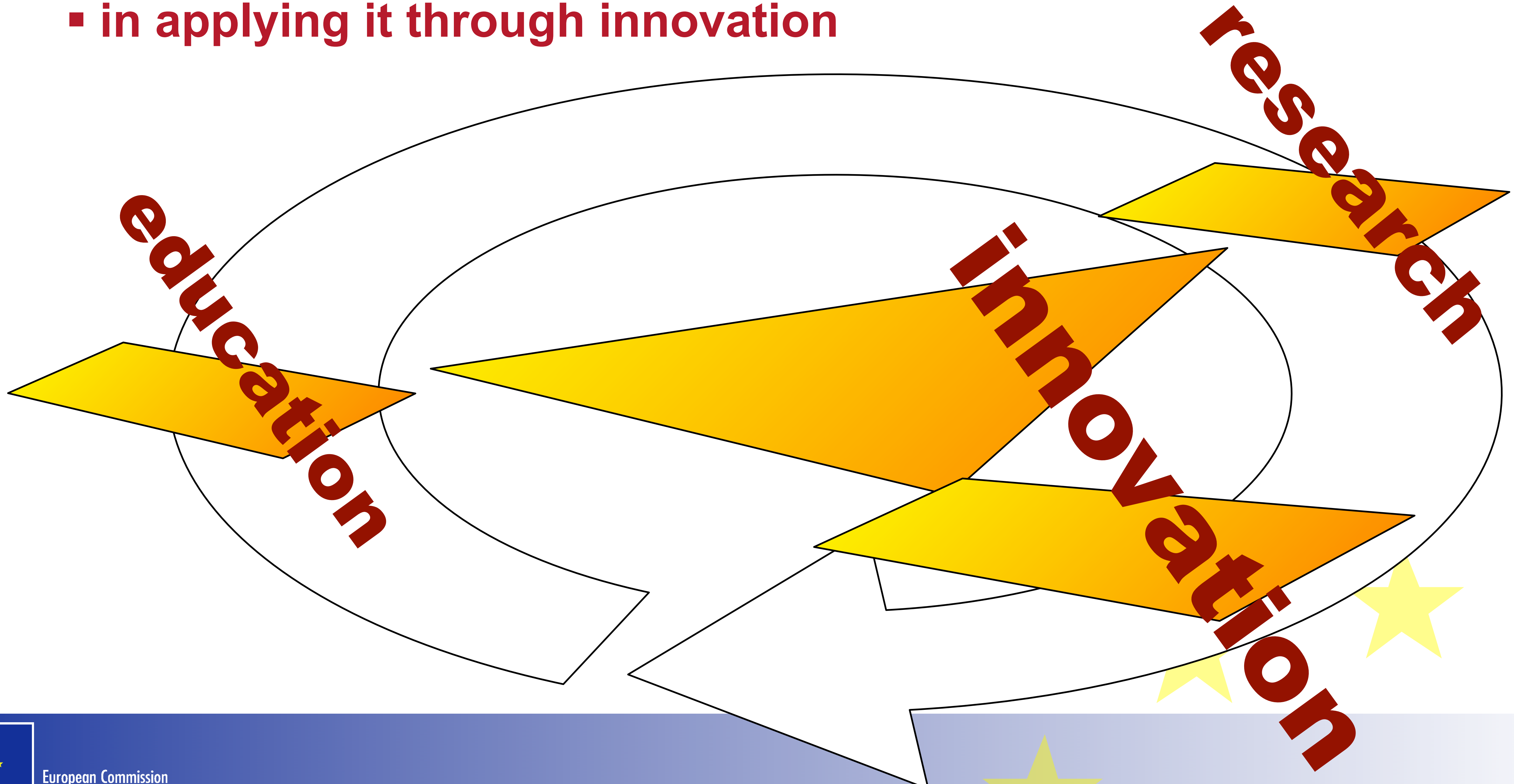


CyberBridges

FP7 - Putting the knowledge triangle at work

To be a genuinely competitive knowledge economy, Europe must be better

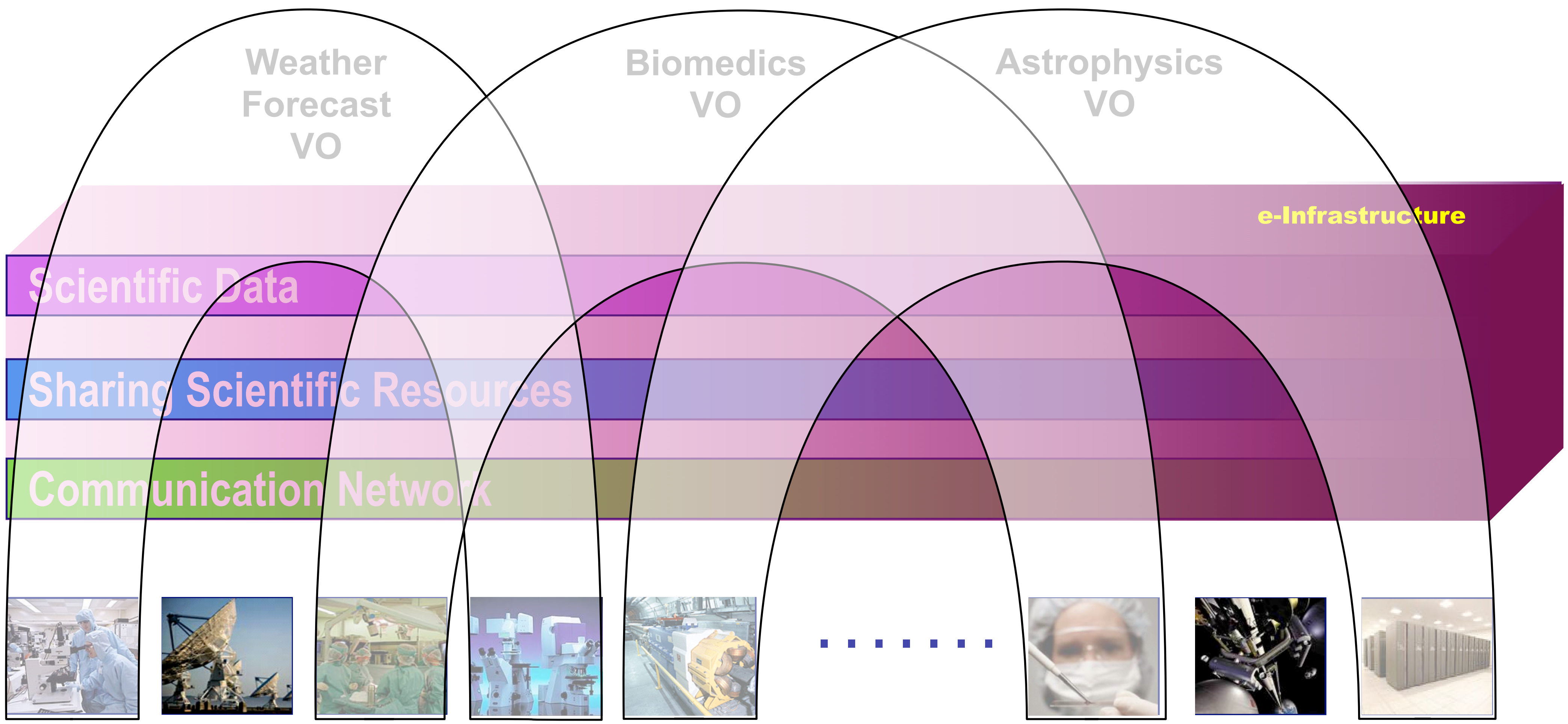
- **in producing knowledge through research**
- **in diffusing it through education**
- **in applying it through innovation**



e-Infrastructures in FP7 - strategy - Virtual Organizations (VO)

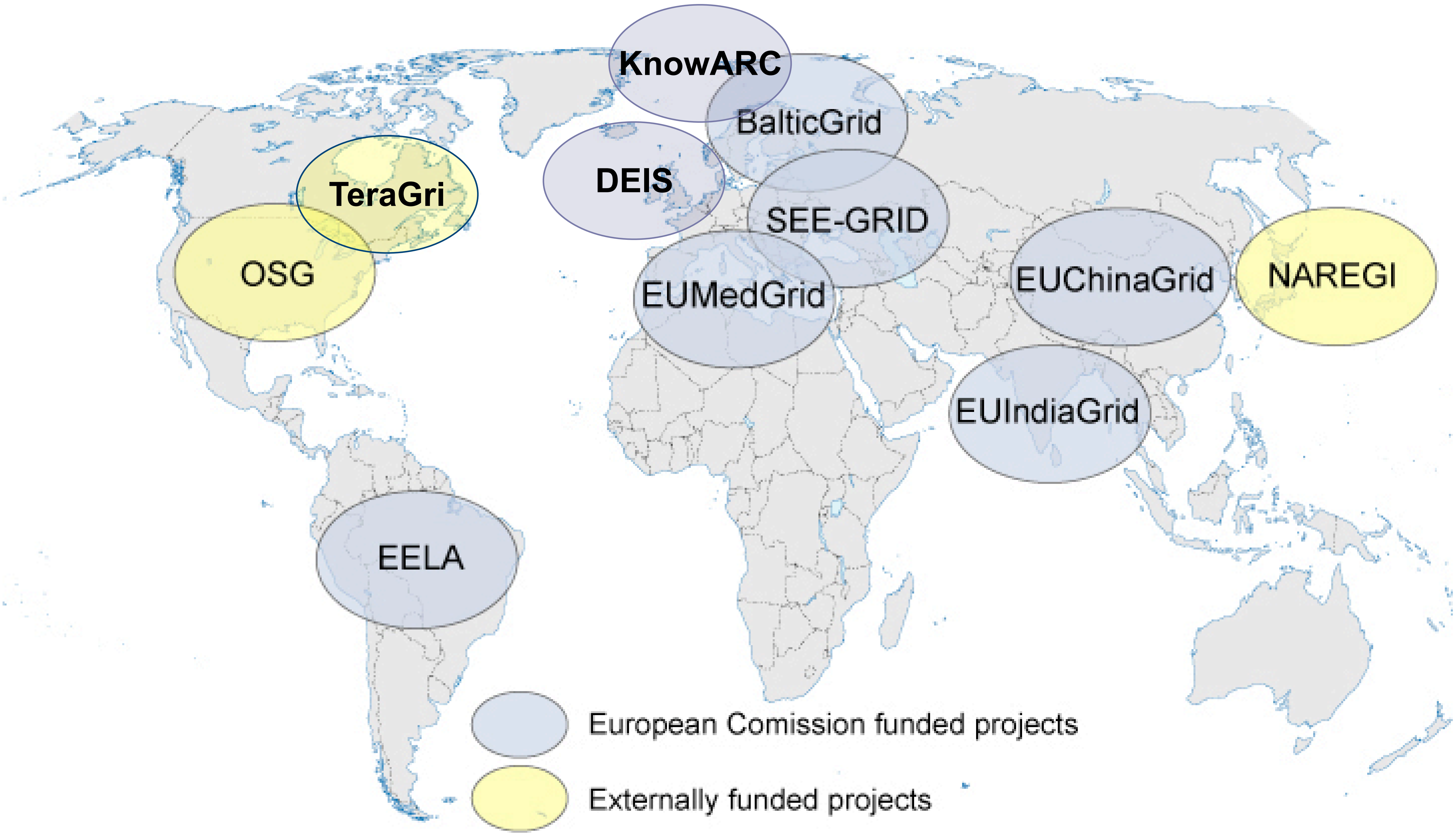
Bringing the best brains together

Sharing the best scientific resources



Producing the best science

Collaborating Infrastructures



Potential for linking ~80 countries by 2008

Some Existing & Potential Interactions

International Science Projects, e.g. ATLAS, CMS

Funding & science collaboration

U.S. Investments

- U.S. part of international science/engineering research projects.

- Open Science Grid (OSG)
- TeraGrid & Science Gateways
- Grid Interoperabilty Now (GIN)
- GLOBUS
- Condor Technologies
- Virtual Data Toolkit (VDT)
- NMI Build and Test
- Shibolleth
- GridShib
- Other?

EGEE

- gLITE
- Experience with large, production, international Grid operation
- Other?

staff R&D interactions

use of components

shared development

Other National/ Regional Grid Projects

New Opportunities

"Cyberinfrastructure-enhanced knowledge communities offer the potential for enabling a new wave of global-scale collaboration across multiple disciplines, geography, and institutions. It could empower a revolution in ***what*** *science explores,* ***how*** *it is done, and* ***who*** *participates.*

Realizing this potential will, however, also required ***a new wave of commitment to collaboration*** *between the complex array of stakeholders necessary to create, deploy, sustain, and apply cyberinfrastructure in transformative ways.*

Cyberinfrastructure both enables and requires a new wave of collaboration."

D. E. Atkins, Keynote for EDUCAUSE Australasia, Auckland, NZ, April 5-8, 2005

Questions & Discussion