

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 <u>datkins@nsf.gov</u>



http://egee-technical.web.cern.ch/egee-technical/conferences/EGEE06/index.html



What I know about the... Oklahoma Supercomputing Context

- Regional leadership and facilitation of this wellestablished 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, interinstitutional, 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)



annotation

Session replay

team chat



Archival data

Journals

D. E. Atkins

5

Evolved into a Network of Instruments (one global instrument)





UARC Patterns of Communication

Pattern of Communication, UARC Campaign, April 9, 1997



Office of Cyberinfrastructure

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 precollege 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



Dualities



Instances of Virtual Organizations (VOs)

People

People

People

Interfaces for interaction, workflow, visualization and collaboration for individuals & distributed teams

Mechanisms for flexible secure, coordinated resource/services sharing among dynamic collections of individuals, institutions, and resources (the Grid or service layer problem)

Distributed, heterogeneous services for:

Computation

Data, information management Sensing, observation, activation in the world 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."

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



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

Vision and Activities Based on Broad and Diverse Community Engagement



- •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.

Cyberinfrastructure

Some Science Drivers

- Inherent complexity and multi-scale nature of todays frontier science challenges.
- Requirement for multi-disciplinary, multiinvestigator, 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



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

Transformative Application - to enhance discovery & learning

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



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



Office of Cyberinfrastructure



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

Cyberinfrastructure Funding

(Dollars	s in Millions)				
		FY 2006		Change	eover
	FY 2005	Current	FY 2007	FY 20	006
	Actuals	Plan	Request	Amount	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/

HOME FUNDING AWA	RDS DISCOVERIES NEWS	PUBLICATIONS	STATISTICS	ABOUT	FastLane
National Scient OFFICE OF Cyberinfrastru	nce Foundation cture		SEARCH NSF Web S	ite	:
OCI Home OCI Fundi	ng OCI Awards	OCI Discoveries	OCI New	s	About OCI
Cyberinfrastructure - s advances in 21st centu science and engineerin	timulating ry g	1145			
About OCI	Special Announcements			Quick Lir	iks
View OCI Staff Directory Search OCI Staff Directory	Career Opportunities - The Office of Cyberinfrast	Dear Colleague Letter	a nationwide	Reports an Relating to Cyberinfra Its Impact	nd Workshops o structure and s
	search to fill a number of	Program Director positio	ons. For more	Publicati	ons <u>See All</u>
Career Opportunities Advisory Committee	Petascale Acquisition F	orum, Mar 24, '06	OCI&nsf_org=OCI	Report of Advisory F Cyberinfra	Blue-Ribbon Panel on structure
Budget Excerpt	NSF Invites Prospective P Discussion of Plans for a l	roposing Institutions and Petascale HPC Acquisition	Vendors to a	Other Sit	e Features
How to Prepare Your Proposal	As indicated in the Preside	ent's FY 2007 Budget Re	uest, NSF is	Research (ports Overviews
Grant Proposal Guide	planning for the acquisitio	n of a petascale high-pe	rformance	Multimedia	a Gallery
Frequently Asked Questions	computing (HPC) system. Subject to the availability of funds, NSF expects to begin funding the resulting multi-year acquisition project in FY07. The petascale HPC system to be acquired will permit science and engineering communities to address some of their most			Classroom	Resources
Other Types of Proposals				NSF-Wide	Investments
Regional Grants Conferences	computationally challengin	ng research needs.	one of their most		
How to Manage Your Award	HPC system vendors and (organizations who, either	potential resource provid r separately or in collabo	er organizations ration with others,		
Grant Policy Manual	wish to propose to manag	e the development, depl	oyment, and		
Grant General Conditions	engineering research com	munity) are invited to m	eet with each		
Cooperative Agreement Conditions	other and with NSF staff t	to discuss the time-line a	and strategy for 24, 2006, from		
Special Conditions	9:00a.m 11:00a.m., at	the National Science Fo	undation, 4201		
Federal Demonstration Partnership	Wilson Blvd., Arlington, V this meeting should send	A, 22230. Those interes email with their name ar	ted in attending nd affiliation to		
	HDC Input@nof gov no in	tor than March 30, 2006	E Enaco lo		

NSF'S CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY

NSF Cyberinfrastructure Council



Several Active Solicitations Posted Seeking more program officers.

<u>www.nsf.gov/od/</u> <u>oci/ci-v7.pdf</u>





Nor web site	NEE Web Site	
	Nor web site	

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



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











iVDgL







Open Science Grid



- 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.



NEES



NanoHub



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

Geographic Place Same

Different

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

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



Geographic

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

CI/VO Enabled Science



NVO and ALMA





Climate Change



LIGO



NEON

The number of nation-scale projects is growing rapidly!



United virtual States observatory

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.



Office of

vberinfrastructure

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

VO-substrate: International R&E Networking











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.



TeraGrid PI's By Institution as of May 2006



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.



Charlie Catlett (cec@uchicago.edu)

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

G · D

Home

- 🖹 🙆 🏠 🔎 👷 🌒 🏵 🍰 🖂 📃 🦓 🔂

Event: "core: ex2 sine1-4"

My Workspace NEESgrid UNR NEESgrid Support NEESgrid All

Realtime access to earthquake Shake table experiments at remote sites.



Į _ ð ×

Nov 14, 2002 05:40 pm

BIRN – Biomedical Information



Geological Information Grid Portal

GEONgrid P	ortal		CYBERINFRASTRUCTURE FOR THE GEOSCIENCES
			Welcome Dennis Gannon: Logout
ortalHome GEONsearch myGEON	GEONscience System UserProfile	MapIntegration	
EON Search GEON Ontology GEON Resou	rce Registration		
° ?	GEON Searc	h	
O was determined at	Soloct a Subject to Show I	Decources	
Metadata Related: Chassa recourse type:		lesources	
	Biological oceanography	Chemical oceanography	Cryology
<all resource="" types=""> 💌</all>	Ecology	Education	Environmental science
Choose subjects:	Forestry	Geochemistry	Geologic time
<all subjects=""></all>	Hydrology	Mineralogy or petrology	Natural bazards
Optional keywords:	Other	Paleontology	Physical geography
	Physical oceanography	Soil science	Structural geology
Spatial Coverage:	Technology		
Type a place name:	(Those subjects will be reen	appiaed coop by comothing ci	miliar to the elactification
GO>	GeoRef.)	ganizeu soon by someuning si	iniliar to the classification
or select an area on the map:	Pasources in Coology	-6 17 files b	
	Kesources in Geology 1-5	or 17 mes V	
	Title: Arizona Geology Map	d.	
	Format: shapefile		
	Dataset Id: GEON-25dfb3d	b-e710-11d8-b226-ab22ed76 27 Eact: 100.04 South: 21.2	581c0 2 Wort: 114 82
· · · · · · · · · · · · · · · · · · ·	Temporal Coverage: any	37 East109.04 South. 31.3	5 West, -114.02
	Description: This is a geolo	ogy map of Arizona in USA.	
Å- 10	Semantic Annotations: see	e details	

Mesoscale Meteorology

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

- Data exploration and Grid workflow





S WELCOME TO THE LEAD PORTAL



Linked Environments for Atmospheric Discovery (LEAD) makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves. The LEAD Portal brings together all the necessary resources at one convenient access point ... read more

> FEATURES FOR ANYONE INTERESTED IN THE WEATHER

Researchers	With university, government, or industry affiliations	GET FEATURES
Educators	At college and university level, high school, or middle schools	GET FEATURES
Students	At graduate, undergraduate, middle and high school levels	GET FEATURES
Visitors	Newcomers and the curious	GET FEATURES

> POPULAR TOOLS



Make a Forecast or Analysis
Experiment Builder







The LEAD Vision: Adaptive Cyberinfrastructure



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



File Edit View	Eavorites Tools Heb	and any second	
		PageRank 🖏 2567 blocked 🔤 Check 🔹 🎋 Suitolijsk 😴 🗐 SuitoFill 💽 🍋 Options	
Address Address	velma renci unc edu:8443/bioportal/user/lau		
Links 🎒 Compan 🌶	🔊 Customize Links 🖉 Free Hotmail 📕 Redisci	anya/is_bala/P=00-502177-200179get/m=d/dedexpert=d/dedwinine=biast2.vindrim_mendstate=151devent3dbinit_dodet/m=det+tind y=n orm	
G Back 🔹 🕞) 🕆 📕 🛃 🎧 🔎 Search 🤺 Fav	orites 🚱 🖾 - 🥌 🖾 - 🛄 🥸 🔛 🦉 🕼 🥸	
	NORTH CAROLINA BIOPORTAL	My Workspace Shared Workspace	Sep 25, 2005 12
Home	Bioinformatics Tools Menu	Bioportal Application Panel	
Applications	Select an application	BLAST2: with gaps (<u>Altschul, Madden, Schaeffer, Zhang, Miller, Lipman</u>)	
Job History	Applications		
Membership	By Module		
Schedule	banana	Reset Submit	
Resources	biosed	(单 = required, 🍨 = conditionally required)	
News	blast2	Simple blast2 form	
Advanced Grid	btwisted		
Usage	build icm	blastp: amino acid query / protein db 🛛 💙 <u>Blast program</u>	
File Transfer			
Bioinformatics	checktrans	Coquence File : places enter either :	
Tutoriais		- Jequence File , please enter <u>eluter</u> ,	
Loneut	Cigue	1. the name of a file: Browse	
Users Present			
Lavanya Ramak 🔨			
		2. or the actual data here:	
	Consense Dispapilat	(sequence format)	
<u> </u>		(sequence <u>romat</u>)	
	- Cusp - Didap		
		Start of required region in query sequence (-L)	
	descseq	End of required region in guery sequence (1)	
	dnadist	env_nr: Non-redundant environmental samples from GenPept+PDB+SwissProt+PIR+PRF 💌 protein db	
		any nt: Environmental Samples	
	dotmatcher		
	dotpath dottup	Filtering and masking options	
	drawgram	Selectivity options	
	dreg	Scoring options	

🖬 stari

Nanohub - nanotechnology



X-Ray Crystallography

Victor Instrument Victor Victor Victor Victor	
Indic Date Reporting Product IUB IUMSC The Purdue Chemistry Crystallography Center IUB Myers IUB Myers Hall Disable your browser's cache to get the live stream! Purdue Crystallography Crystallography Center Minnesota Vinix.of ChemMatCARS Vinix.of Chicago at APS Vinix.of Other Other	Login
IUB IUMSC The Fundee Chemistry Crystallography Center IUB Myers The Purdue Chemistry Crystallography Center Purdue Crystallography Center Disable your browser's cache to get the live stream! CSAF Sydney, Australia The Purdue Chemistry Crystallography Center Minnesota X-ray Lab Chemistry Carter The Purdue Chemistry Crystallography Center Other Other Chemistry Crystallography Center The Purdue Chemistry Crystallography Center	
Crystallography CsAF Sydney, Australia Minnesota X-ray Lab ChemMatCARS - Univ. of Chicago at APS Other	
Collaborators Data from Nonius Kappa CCD detector (Under development!) Total Number of jpg: 10 Frame: s01f0010.jpg Image: Streaming video from the lab showing the Nonius instrument NCS Southampton, UK All available jpg images Streaming video from the lab showing the Nonius instrument Image: Streaming video from the crystal microscope Local date/time: 2005-09-24 11:36:54	n the Noni
These values are updated approx. every 60 sec.	es in UT
LaDJack U12	16.25.5
Rel. Humid. 43.1 %	10.55.5
Chill Water In: 16.4 C 2005-09-2	16:36:2
Chill Water Out: 19.3 C 2005-09-2	16:36:2
Generator Relay Voltage: 3.42 2005-09-2 All previous voltages X-ray Generator is: OFF 2005-09-2	16:36:48

ServoGrid Portal



Username:		L a sin [
Password:		Login		
Create New Account Login Help				

SERVOGrid FaultDB Search QuakeTables Portal Search SERVOGrid

SERVO Grid

Solid Earth Research Virtual Observatory Grid

- QuakeSim home page.
- Old GEM General Earthquake Modeling Web Site
- SLIDE Distributed File System for NASA Computational Technology Project
- Report from the Earth Science Enterprise Computational Technology Requirements Workshop April 30-May 1 2002 where SERVO concept first introduced
- Discover the Grid at the Grid Forum or at this collection of papers
- Other collected papers and presentations on SERVOGrid and related topics are available from the Community Grids Lab publications page.



Belfast Gene Grid Portal



MyGrid - Bioinformatics



Navigate

<u>Home</u>

About Downloads Components Component Overview Research Components Using myGrid Research Using myGrid Links Publications

Contact

Log In



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, the e-science workbench that actively supports the scientific lifecycle. Each component or service contring system that allows the e-scientist to perform complex in-silico experiments across distributed bioinfor 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.









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

Enabling Grids for E-sciencE



Potential for linking ~80 countries by 2008

Some Existing & Potential Interactions



EGEE

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

staff R&D interactions use of components shared development

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?

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."



Office of

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



Questions & Discussion



