

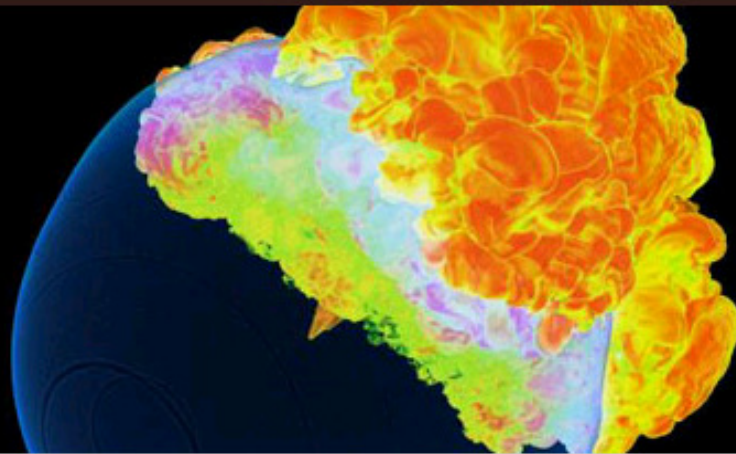
# Technology for Distributed Collaboration

Ian Foster

Computation Institute  
Argonne National Laboratory  
University of Chicago  
[foster@mcs.anl.gov](mailto:foster@mcs.anl.gov)



# Computation Institute



## CI BLOG

**Tapping Private Sector Innovation**  
NASA, with as strong a history of technical innovation as any Federal agency, has been making the news recently initiating partnerships with Google, Inc. . Originally announced late last year without specific details, one of their first joint projects was detailed this week - applying Google search technology to help scient...

## Computing Resources

Contact Us

## >> EVENTS

*National Laboratory*

LOCATION: Eckhart 133, UChicago

[\[more info\]](#)

January 31, 2007

Computations in Science Seminars

**"Learning Networks from Biology, Learning Biology from Networks"**

SPEAKER: Chris Wiggins, *Columbia University*

LOCATION: Kersten Physics Teaching Center,

KPTC 206, UChicago

[\[more info\]](#)

## >> HIGHLIGHTS



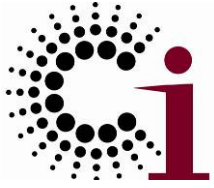
12.11.06

**Argonne's William Gropp  
Recognized by the  
Association for  
Computing Machinery**

[\[read more\]](#)

[www.ci.uchicago.edu](http://www.ci.uchicago.edu)

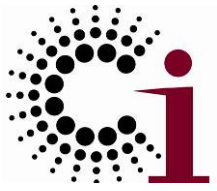
[www.ci.anl.gov](http://www.ci.anl.gov)



## In the Next 50 Years, We Must ...

- Increase energy production by 5, while reducing GHG emissions by 2 or more
- Mitigate and adapt to climate change
- Address increasingly drug resistant diseases
- Provide meaningful livelihoods for 9B people

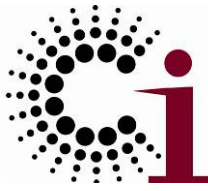
**→ Innovation**



## We Must Get **Smarter** ...



Maxwell Smart (NBC, 1960s; Warner 2008)



# The Three Dimensions of **Smart**



Technology

Biology

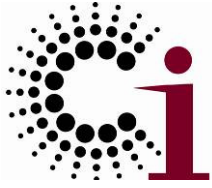
Culture



# Problem Solving as “Thinking Aloud”

- “What if I try A?”
- “I wonder how I do B?”
- “What do others know about C?”
- “Hey, I’ve just learned how to do D!”

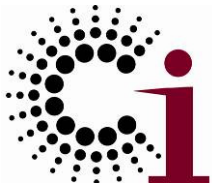
→ How do I reduce cycle time?



# Thinking Aloud: Reducing Cycle Time

- “What if I try A?”
  - Design, modeling, fabrication tools
- “I wonder how I do B?”
  - Wikis, design databases, conversation
- “What do others know about C?”
  - Databases, search tools, conversation
- “Hey, I’ve just learned how to do D!”
  - Publication, conversation, education

(Distributed) collaboration is a crosscutting theme

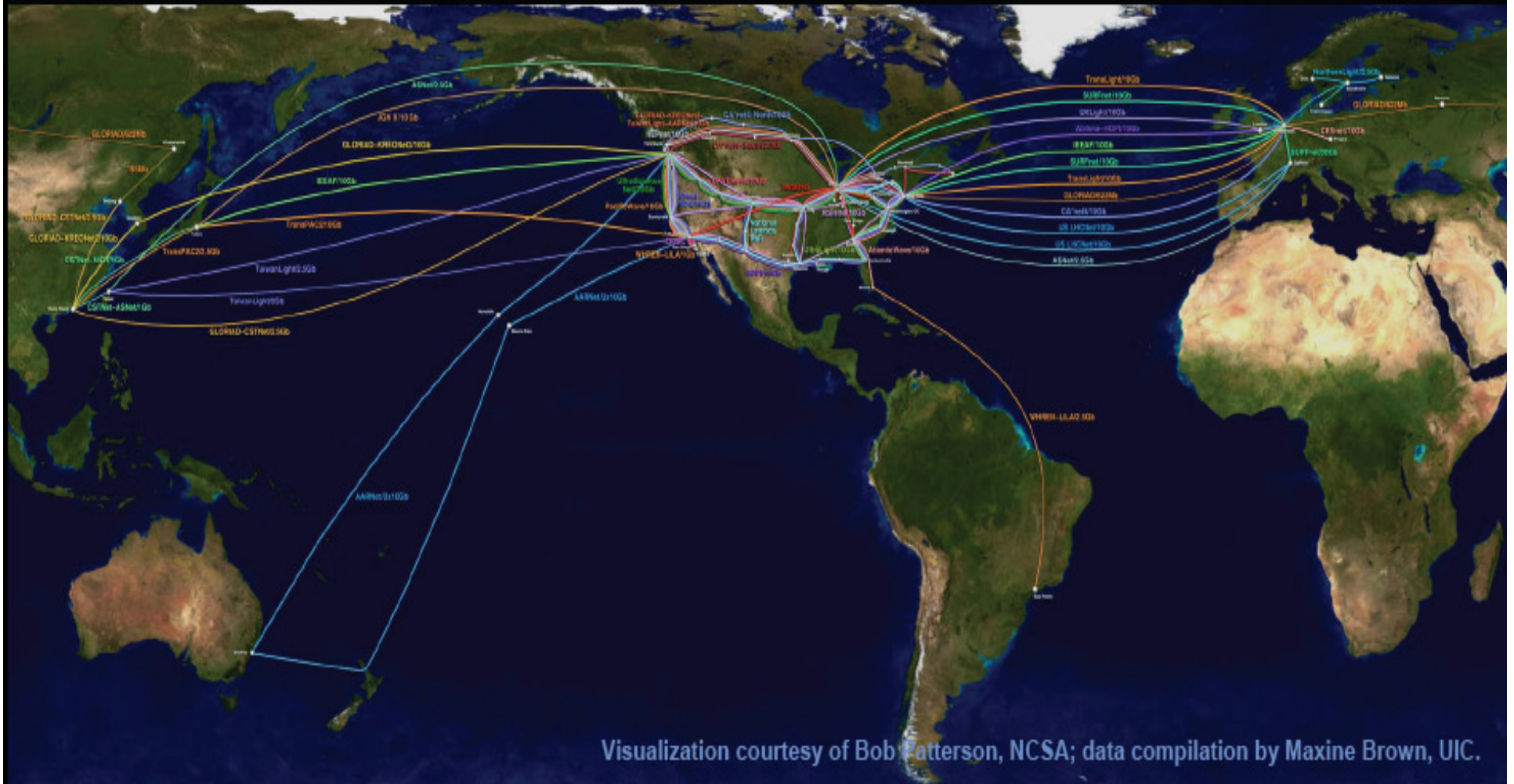


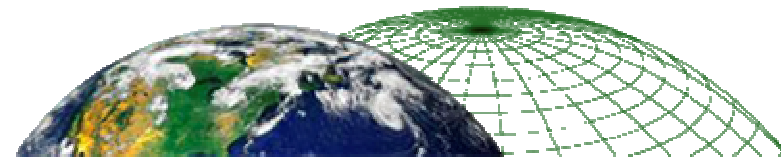
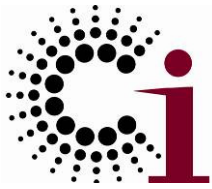
# Technologies for Distributed Collaboration

- Conversation
  - ◆ Post
  - ◆ Fedex
  - ◆ Telephone
  - ◆ Email, IRC, ...
  - ◆ Instant messaging
  - ◆ Videoconference
- Immersive
  - ◆ MUDs
  - ◆ Access Grid
  - ◆ Second Life
- Data publication
  - ◆ FTP, Gopher, ...
  - ◆ Web
  - ◆ Blogs
  - ◆ Semantic Web
- Federation
  - ◆ Collaborative bookmarking
  - ◆ Grid computing
  - ◆ Service-oriented architecture



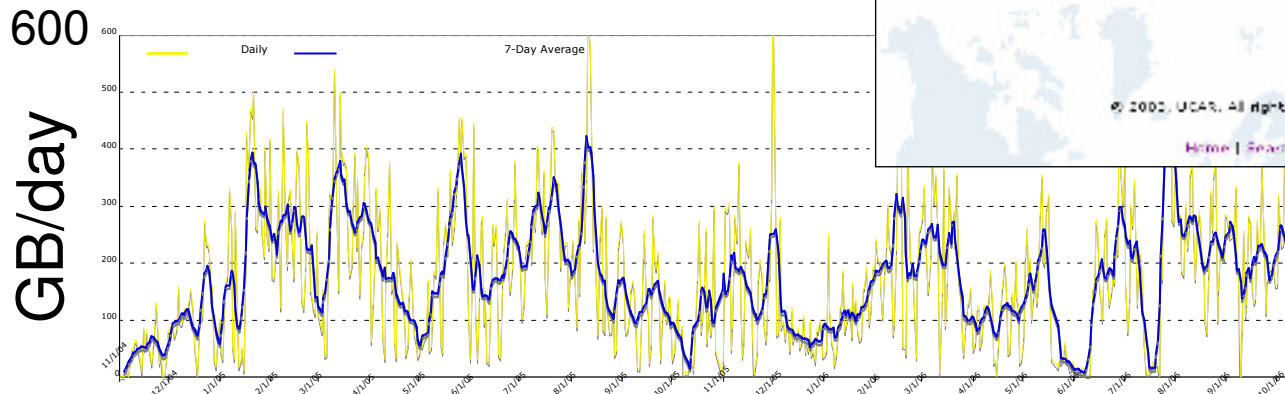
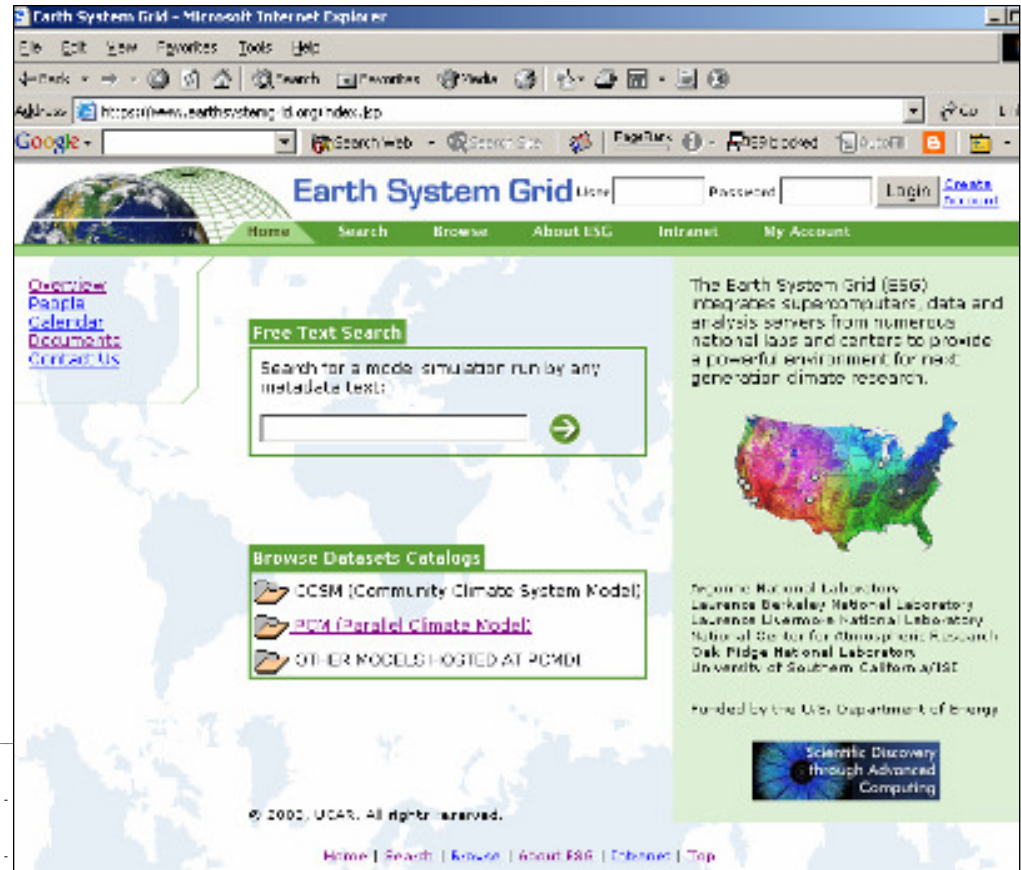
National Research & Education Networks who voluntarily share optical networking resources and expertise to develop the *Global LambdaGrid* for the advancement of scientific collaboration and discovery



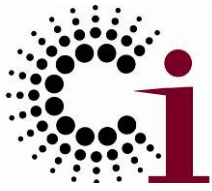


# Earth System Grid

- Provides access to all IPCC data
- >150 TB data downloaded
- >300 scientific papers written







# Integrating Data and Computing, on Demand

**Public PUMA Knowledge Base**  
 Information about proteins analyzed against ~2 million gene sequences

gi 23499780 gnl REF_tigr BRA0013	gi 16080253 ref NP_391080.1	44.27	253	131	1	15	257	8	2603.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 123098409 ref NP_691875.1	43.48	253	133	2	16	258	5	2573.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 48637187 ref ZP_00294182.1	44.92	256	126	2	14	256	7	2591.1	e
gi 23499780 gnl REF_tigr BRA0013	gi 52008400 gb AA025342.1	44.75	257	126	2	15	258	3	2561.9	e
gi 23499780 gnl REF_tigr BRA0013	gi 48664015 ref ZP_00317908.1	44.49	245	134	1	13	257	5	2476.1	e
gi 23499780 gnl REF_tigr BRA0013	gi 30348891 gb AA028934.1	39.53	253	138	3	18	257	5	2522.0	e
gi 23499780 gnl REF_tigr BRA0013	gi 19655222 gb AA093939.1	40.64	251	138	1	17	256	10	2602.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 12735806 gb AA007757.1	43.03	251	130	4	18	256	11	2602.5	e
gi 23499780 gnl REF_tigr BRA0013	gi 112597924 gb AA038999.2	46.70	162	96	1	62	243	5	1836.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 46363318 ref ZP_0026079.1	39.58	240	136	2	14	253	6	2361.8	e
REF_tigr BRA0013	gi 39933731 ref NP_946007.1	34.90	255							
REF_tigr BRA0013	gi 48782600 ref ZP_00279106.1	35.92	245							
REF_tigr BRA0013	gi 41407534 ref NP_960370.1	36.09	266							
REF_tigr BRA0013	gi 48851585 ref ZP_00305793.1	32.39	247							
REF_tigr BRA0013	gi 15966306 ref NP_386659.1	36.50	263							
REF_tigr BRA0013	gi 17548526 ref NP_521866.1	36.36	264							
gi 23499780 gnl REF_tigr BRA0013	gi 51891730 ref VP_074421.1	38.87	247	136	7	18	256	1	2403.4	e
gi 23499780 gnl REF_tigr BRA0013	gi 145881 gb AA023739.1	33.87	246	147	3	13	253	3	2404.4	e
gi 23499780 gnl REF_tigr BRA0013	gi 25029334 ref NP_739388.1	35.20	250	147	4	15	256	6	2485.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 21220953 ref NP_636732.1	36.52	257	138	6	12	255	5	2545.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 46314029 ref ZP_00246385.1	33.86	254	153	2	12	259	3	2485.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 41406852 ref NP_959688.1	35.61	238	149	2	16	253	2	2309.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 115644471 ref NP_229523.1	35.69	255	144	5	12	256	2	2469.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 23470090 ref ZP_00125423.1	35.20	250	145	4	12	253	3	2439.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 24935279 gb AA064237.1	34.63	257	146	4	12	257	4	2499.8	e
gi 23499780 gnl REF_tigr BRA0013	gi 48647651 ref ZP_00303815.1	36.05	256	145	9	12	257	4	2531.3	e
gi 23499780 gnl REF_tigr BRA0013	gi 28851510 gb AA054587.1	36.40	250	142	4	12	253	3	2431.3	e
gi 23499780 gnl REF_tigr BRA0013	gi 12737873 ref NP_770312.1	36.25	251	143	3	14	255	3	2491.3	e
gi 23499780 gnl REF_tigr BRA0013	gi 1708936 sp P50198 LINDX_PSEPA	34.23	260	143	4	12	257	4	2491.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 33594146 ref NP_381792.1	34.17	240	148	5	18	256	6	2363.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 33594116 ref NP_381759.1	34.17	240	148	5	18	256	6	2363.7	e
gi 23499780 gnl REF_tigr BRA0013	gi 33283006 ref NP_232830.1	34.23	241	143	5	18	256	6	2363.7	e

**Bioinformatics Group MCS, Argonne**  
 PUMA2 -- Evolutionary Analysis of Metabolism

NCBI gi number: 16124111 - putative autotransporter protein [Yersinia pestis]

NCBI related proteins: 15981892, 25511357  
 TrEMBL: Q8Z436  
 FIR-NREF: NFD0798375  
 NCBI Accession: CAC93445.1  
 Source Organism: Yersinia pestis CO92  
 Taxon ID: 214092

Sequence length (1070 aa)

Chromosomal Comparison: The SEED, Similarity -- Global, BLAST vs. nr, Fast3 vs. UniProt, Blocks-Blast, PhyloBlast, BLink, Similarity -- Local, InterPro, Blocks, DART, Protein families, COGs, TIGRFAMs

Domains: Pertactin domain, Autotransporter beta-domain, Outer membrane autotransporter barrel

Similarity results: BLAST vs. nr, InterPro (10945158, 22127716, 33599810, 33595429, 33595157, 33599439), Protein families (16119581, 17938938, 33591734, 34496292, 23500862, 33504427)

**Back Office Analysis on Grid**  
 Millions of BLAST, BLOCKS, etc., on OSG and TeraGrid

Global Observation Database (View) C

# Social Informatics Data Grid

VCR-Style Control Panel

Animated Text Transcript (Paragraph Representation)

Tag Transcript Editor

Animated Avatar Representation

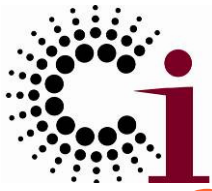
Animated Graph Panes

Video Displays

Video List

Bennett Berthenthal et al., [www.sidgrid.org](http://www.sidgrid.org)

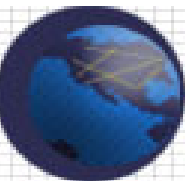
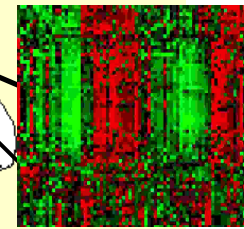
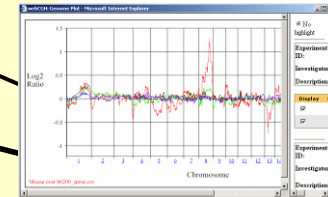
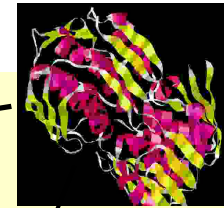
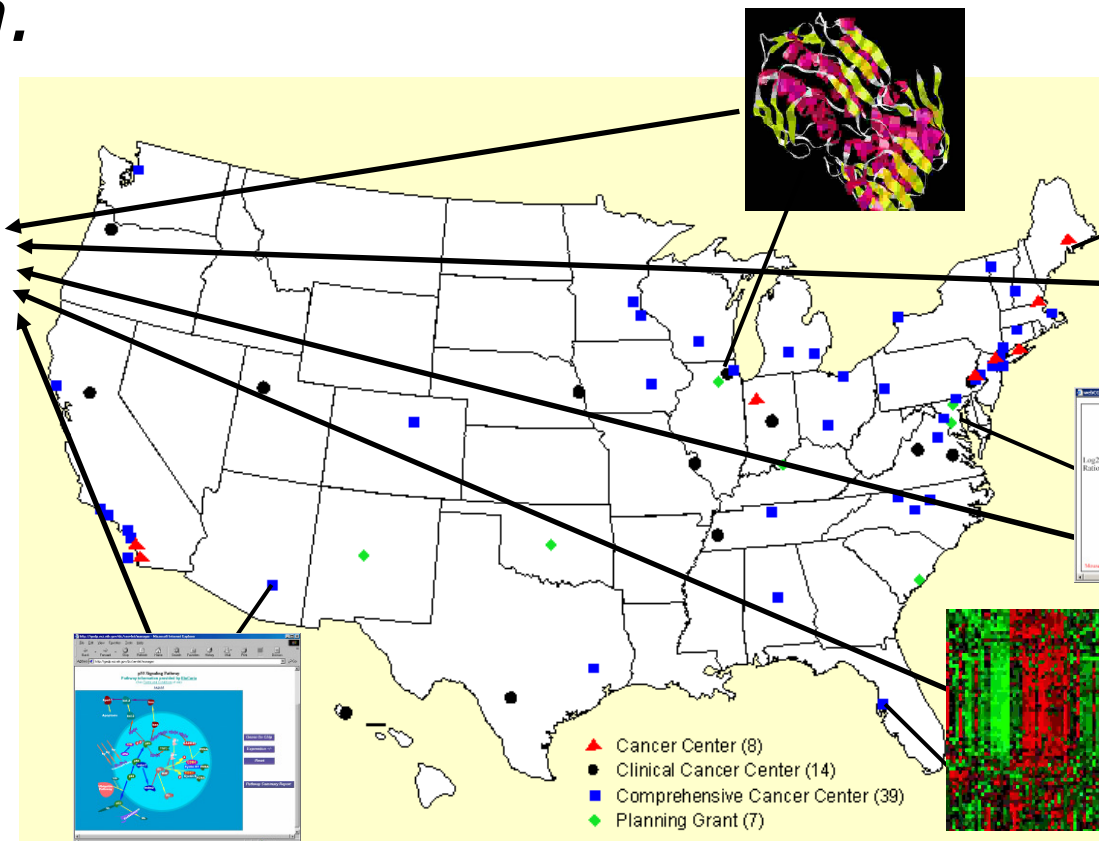
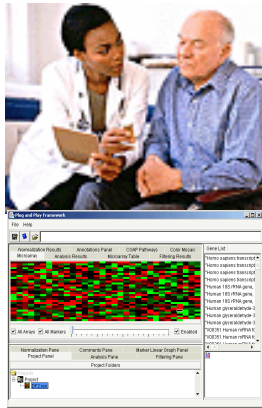




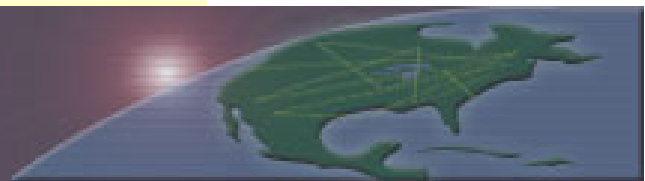
# NIH's

# Cancer Biomedical Informatics Grid

*caBIG: sharing of infrastructure, applications, and data.*



**caBIG** cancer Biomedical Informatics Grid



# Medical Education over Access Grid



Credit: Jonathan Silverstein, U.Chicago

# Access Grid and SARS





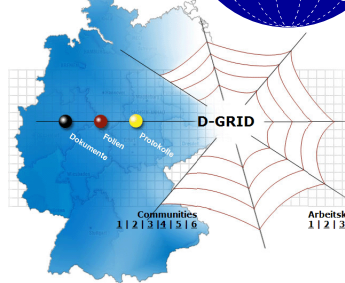
# Global Communities



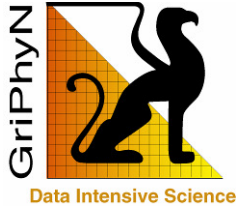
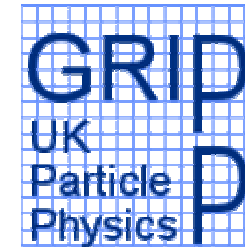
GRID.it  
project



K\*GRID

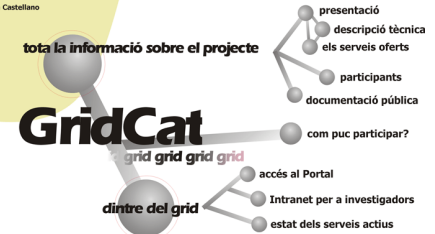


National Grid  
NG  
SINGAPORE



CERN

openlab for DataGrid applications  
Developing Solutions for the Data-Intensive Science of the Large Hadron Collider



eGee  
Enabling Grids for E-science in Europe

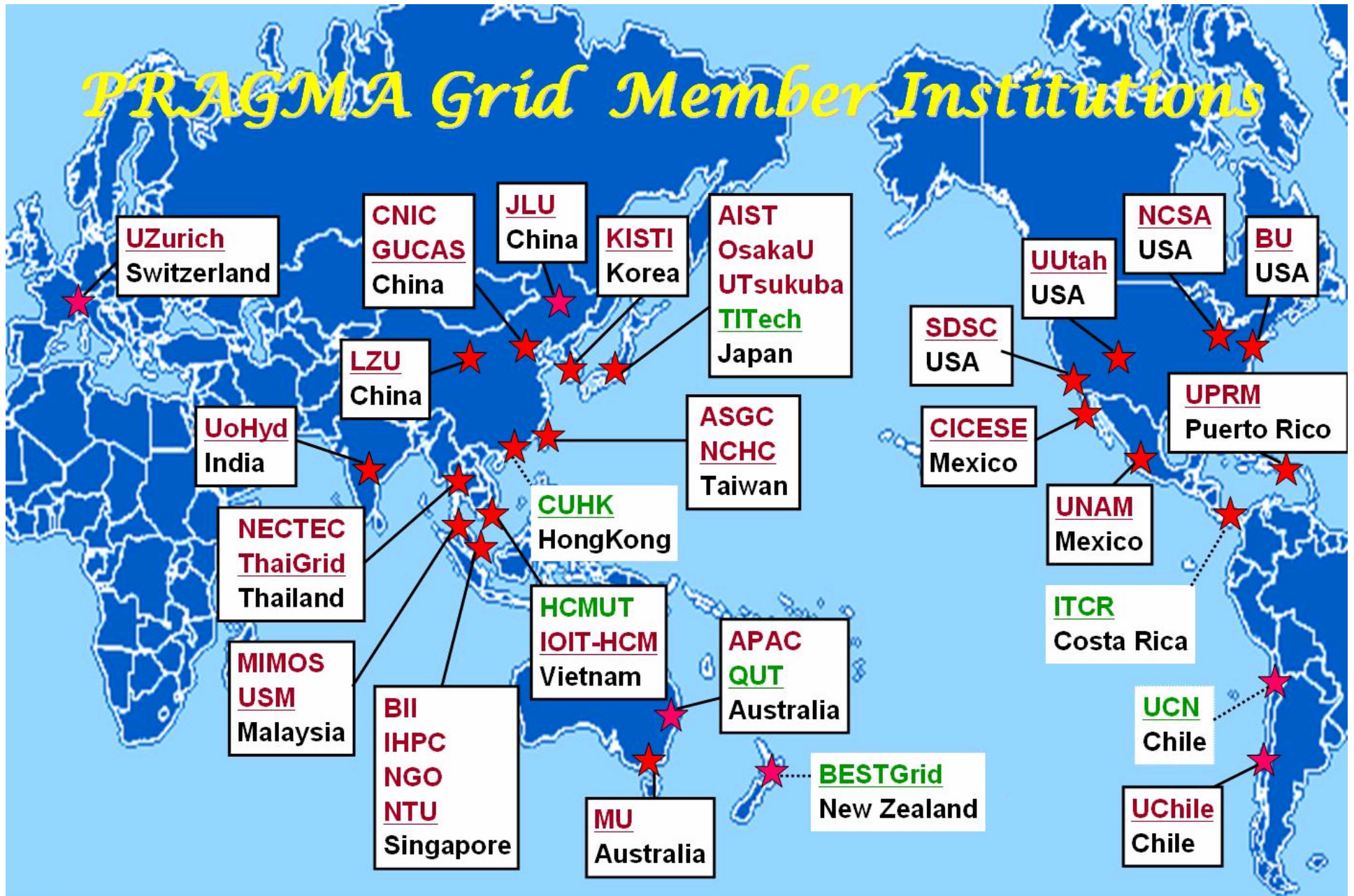


国立情報学研究所グリッド研究開発推進拠点 NII -The National Institute of Informatics





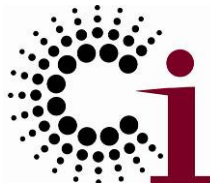
# PRAGMA Grid Member Institutions



31 institutions in 15 countries/regions (+ 7 in preparation)

Last update: 5/30/2007





# Community Website

## Software Downloads, User-contributed Content, Hardware Reference, & More



**Navigation**

- ▣ Home
- ▣ Software
- ▶ Hardware
- ▣ Documentation
- ▣ Community
- ▣ Events
- ▣ Mailing Lists
- ▣ News
- ▣ Search
- ▣ Users

**User login**

**Username:**

**Password:**

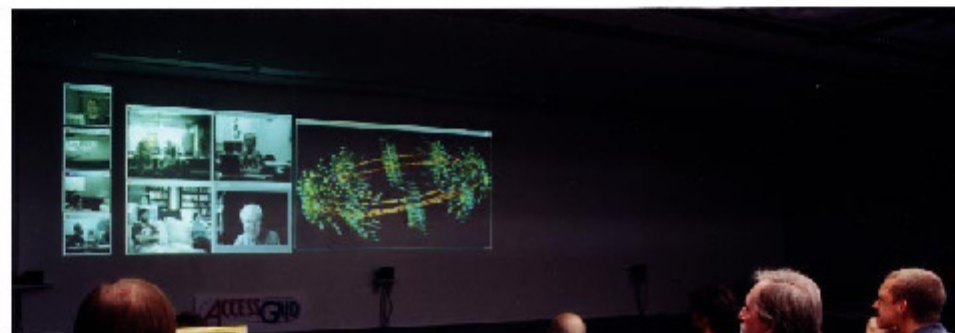
- [Create new account](#)
- [Request new password](#)

[Home](#)

### Welcome to AccessGrid.org

Submitted by [Tom Uram](#) on Thursday, October 5, 2006 - 22:41

The Access Grid® is an ensemble of resources including multimedia large-format displays, presentation and interactive environments, and interfaces to Grid middleware and to visualization environments. These resources are used to support group-to-group interactions across the Grid. For example, the Access Grid (AG) is used for large-scale distributed meetings, collaborative work sessions, seminars, lectures, tutorials, and training. The Access Grid thus differs from desktop-to-desktop tools that focus on individual communication. The Access Grid developers have issued over 20,400 certificates to users across 56 countries. Each institution has one or more AG nodes, or "designed spaces," that contain the high-end audio and visual technology needed to provide a high-quality compelling user experience. The nodes are also used as a research environment for the development of distributed data and visualization corridors and for the study of issues relating to collaborative work in distributed environments.



**Recent News**

- [Access Grid 3.0.2 Release](#)
- [Hardware Reference online](#)
- [Access Grid 3.0.1 Release](#)
- [Access Grid 3.0 Released](#)

**Browse archives**

« November 2006

Su	Mo	Tu	We	Th
			1	2
5	6	7	8	9
12	13	14	15	16
19	20	21	22	23
26	27	28	29	30

**Search AG Mailing Lists**

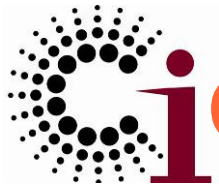
Search:

Match:  ▾

Format:  ▾

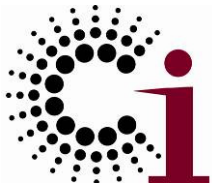
Sort by:





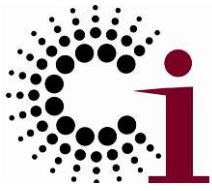
# Globus Downloads Last 24 Hours





## Lessons Learned

- The power of diversity & scale
  - ◆ Open Science Grid: 80 sites, 30K CPUs
  - ◆ World Community Grid: 700,000 CPUs
  - ◆ Access Grid: several thousand nodes
  - ◆ Wikipedia, Flickr, CiteULike, Connotea, ...
- The challenges of heterogeneity
  - ◆ Bandwidth, hardware, interests, trust, understanding, meaning, timezone, ...
- The challenges of scale
  - ◆ Participants, data, computing, ambition
- Everything is still far too complicated!



# Access Grid:

## The Power of Context



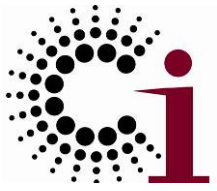
The screenshot shows the 'Venue Client' window for 'Argonne National Laboratory'. The address bar displays the URL: <https://vv3.mcs.anl.gov:8000/Venues/000000ff6f312bfe008c00dd00>. The interface is divided into several sections:

- My Venues:** A list of venues including Argonne Lobby, Big Horn, Bridgeport, Dantooine, Full Sail, Institution Lobby, Jack Frost, Kamar, Lucky Labrador, Talus, Test Room, and Windmer.
- Participants:** A list of names with profile icons: Eric Olson, Joe Insley, Mark Hereld, Michael E. Papka, Susanne Lefvert, Thomas D. Uram, and Ti Leggett.
- Data:** A list of files: DataSnapshot01.jpg, DataSnapshot02.jpg, MeetingSlides.ppt, and Proposal.pdf.
- Services:** A section for services.
- Application Sessions:** A list of sessions: Shared Movie Viewer - 12:01:04 PM Jun and Shared Presentation - 10:15:59 AM Jun.

At the bottom, a chat window shows the following text:

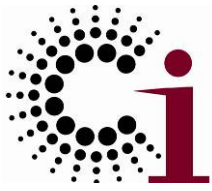
-- Entered venue Argonne National Laboratory (Mon, 18 Sep 2006, 12:21:38)  
Venue for Argonne National Laboratory, information at <http://www.anl.gov>.  
**Thomas D. Uram:** Compare the snapshots in the venue with the latest snapshots here: <http://www.mcs.anl.gov/fl/research/snapshots>  
**Mark Hereld:** There's an interesting change in the third set of time values  
**Susanne Lefvert:** That's likely due to a build-up of nickel as the front expands

A 'Display' button is located at the bottom right of the chat window.



# “Thinking Aloud” (for Science or Invention): 10 Year From Now

- On-demand access to powerful data, design, analysis, & fabrication resources
  - ◆ Service-oriented science & engineering
  - ◆ Deep analysis of vast quantities of data
  - ◆ Commoditization of design & analysis
- Communities of 2, 20, 200, 2K, 2M can self-identify easily within a sea of billions
  - ◆ To share information, converse, discover
- We understand innovation & collaboration far better than today



## Some Key Challenges

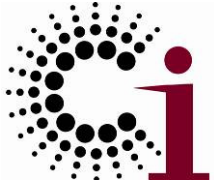
- Enable smooth scaling in many dimensions
  - ◆ Number of participants (K-, M-, G-persons?)
  - ◆ Internet capabilities (0 to Tbit/sec)
  - ◆ Physical resources
  - ◆ Amount of data (megabytes to exabytes)
  - ◆ Complexity of questions asked & answered
  - ◆ Degree of trust, shared language, etc.
- Integration with the physical world
  - ◆ Active sensors
  - ◆ Automated experimental protocols
  - ◆ Integrate manufacturing and problem solving



## Current Activities

- Access Grid 3.0
  - ◆ Conversation, context, scale, ease of use
  - ◆ Dozens of sites
- Collaborative tagging for scientific data
  - ◆ Collaborative creation of data exegesis
- Resource federation in virtual organizations
  - ◆ Grid protocols and software





# We Can Contribute to a Democratization of Science

- Personalized manufacturing (FabLab)
- Personalized reporting (blogosphere)
- Personalized innovation (“Global Knowledge Environment”?)

“So much ingenuity my generation has,  
and no place to put it” — Charlie Leduff