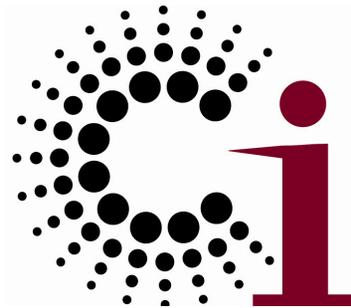




Working Smarter

Ian Foster



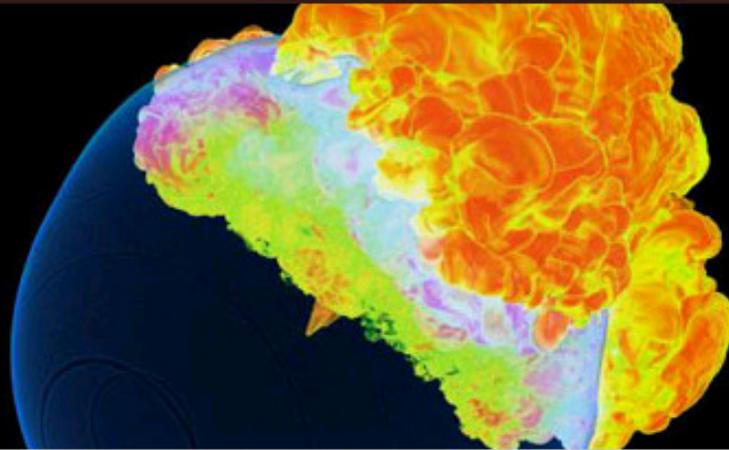
Computation Institute

Argonne National Lab & University of Chicago

<http://ianfoster.typepad.com>



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

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**



Biology

Technology

Culture



J.C.R. Licklider Reflecting in 1960 on Where His Time Went



About 85 per cent
of my “thinking”
time was spent
getting into a
position to think,
to make a decision,
to learn something
I needed to know



For Example ...

“At one point, it was necessary to compare six experimental determinations of a function relating speech-intelligibility to speech-to-noise ratio.

No two experimenters had used the same definition or measure of speech-to-noise ratio. Several hours of

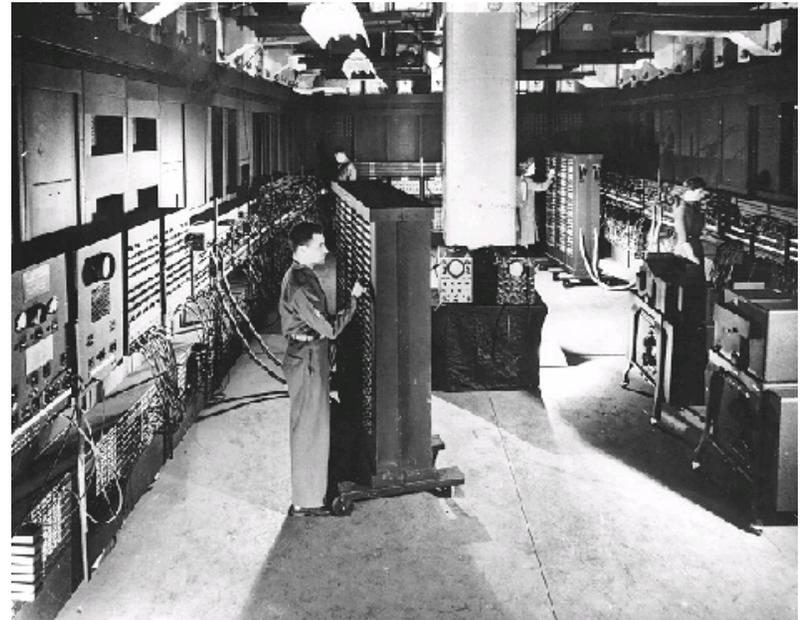
calculating were required to get the data into comparable form. When they were in comparable form, it took only a few seconds to determine what I needed to know.”





Man-Computer Symbiosis

“... will involve very close coupling between the human and the electronic members of the partnership ... Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them.”





Licklider's Vision

- He imagined
 - ◆ Low-cost personal computers
 - ◆ Organized, easily searchable data stores
 - ◆ Intuitive user interfaces
 - ◆ Computer aided reasoning
 - ◆ Interactive numerical simulation

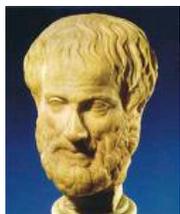


Fast Forward to 2007

- He imagined
 - ◆ Low-cost personal computers ✓
 - ◆ Organized, easily searchable data stores ✓/?
 - ◆ Intuitive user interfaces ✓/?
 - ◆ Computer aided reasoning ✓/?
 - ◆ Interactive numerical simulation ✓/?
- We also recognize the need for
 - ◆ Standard representations of knowledge
 - ◆ Collaborative problem solving
- And we have coined the term “eScience”



Emergence of New Problem Solving Methodologies



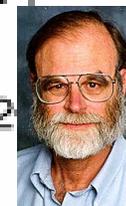
Empirical



Theory



Simulation



Data

<0

1700

1950

1990

y

$$y = e^x$$

6

37

2

1

x

eScience: When brute force doesn't work anymore (Szalay)



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?



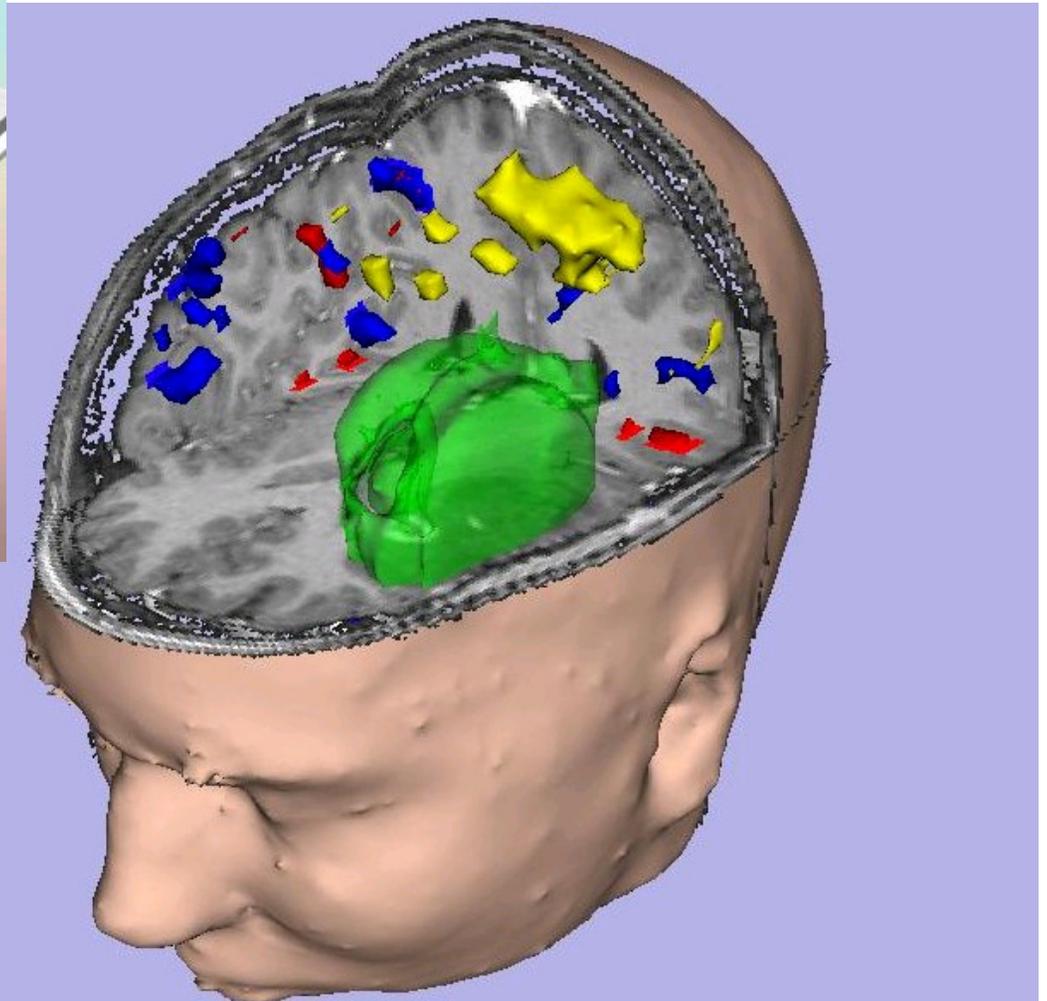
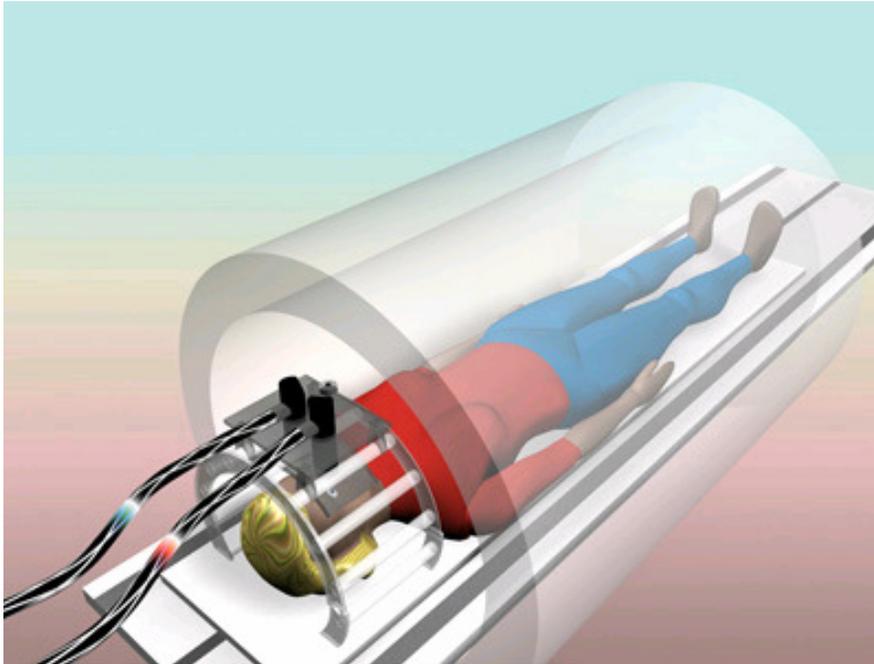
Thinking Aloud: Reducing Cycle Time

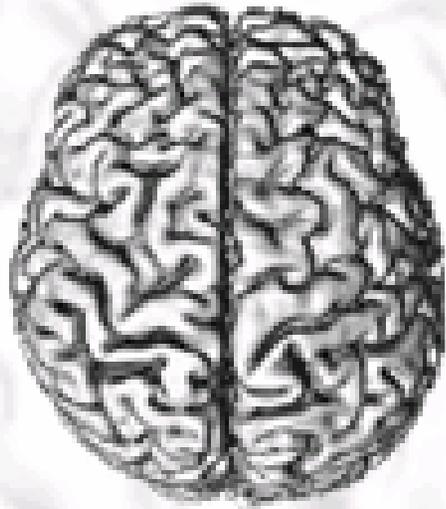
- “What if I try A?”
 - Design, modeling, fabrication tools
- “I wonder how I do B?”
 - Wikis, design databases, conversation
- “What do others know about C?”
 - Databases, search tools, conversation
- “Hey, I’ve just learned how to do D!”
 - Publication, conversation, education

(Distributed) collaboration is a crosscutting theme



Functional Magnetic Resonance Imaging (fMRI)





A public repository of peer-reviewed fMRI studies and their underlying data.

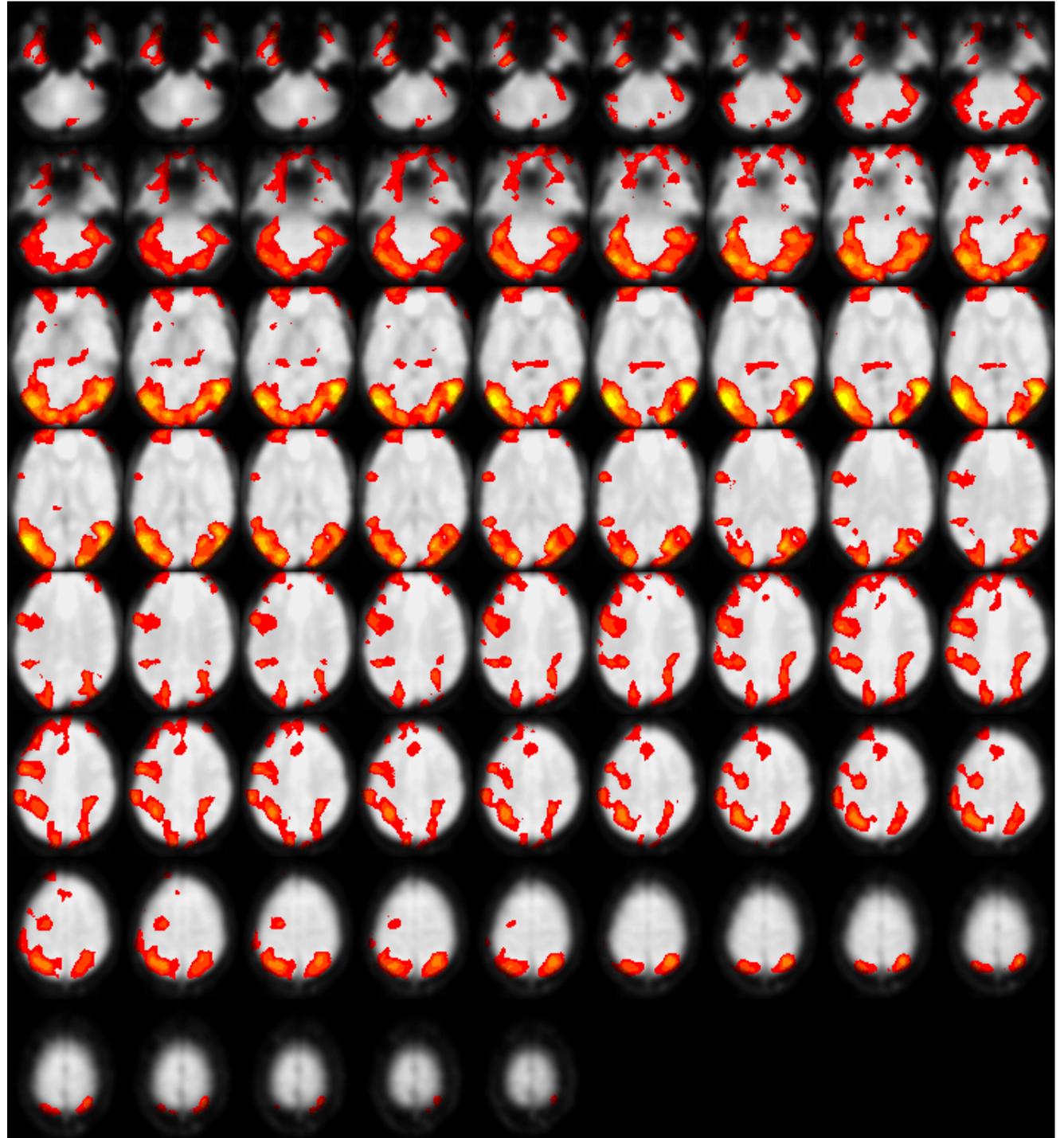
The fMRI Data
Center
(www.fmridc.org)



- A typical study comprises
 - 3 groups,
 - 20 subjects/group,
 - 5 runs/subject,
 - 300 volumes/run
 - 90,000 volumes, 60 GB raw
 - 1.2 million files processed
- 100s of such studies in total

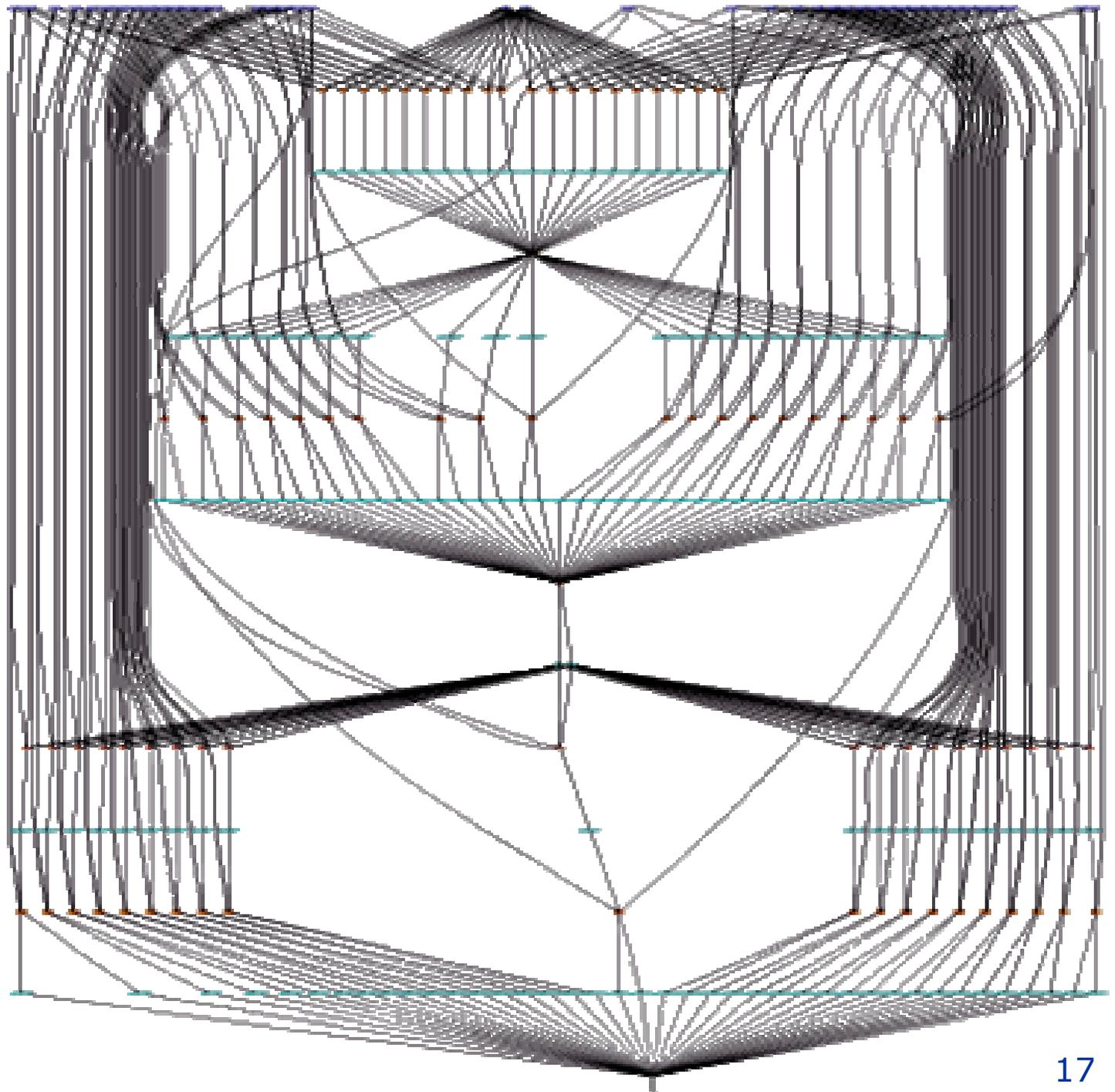


Functional MRI Analysis





A Simple Analysis





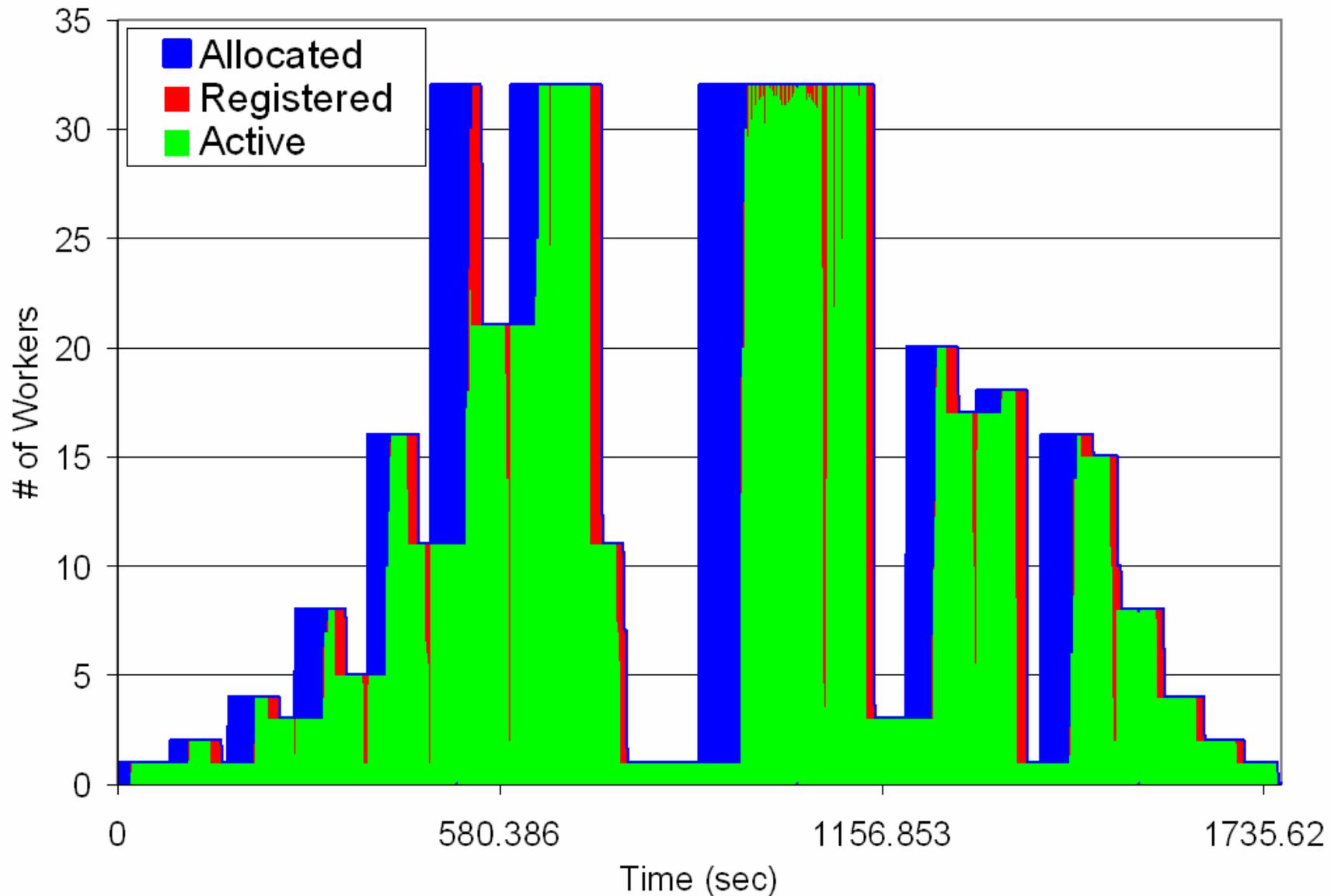
SwiftScript

```
(Run snr) functional ( Run r, NormAnat a,  
                      Air shrink ) {  
  Run yroRun = reorientRun( r , "y" );  
  Run roRun = reorientRun( yroRun , "x" );  
  Volume std = roRun[0];  
  Run rndr = random_select( roRun, 0.1 );  
  AirVector rndAirVec = align_linearRun( rndr, std, 12, 1000, 1000, "81 3 3" );  
  Run reslicedRndr = resliceRun( rndr, rndAirVec, "o", "k" );  
  Volume meanRand = softmean( reslicedRndr, "y", "null" );  
  Air mnQAAir = alignlinear( a.nHires, meanRand, 6, 1000, 4, "81 3 3" );  
  Warp boldNormWarp = combinewarp( shrink, a.aWarp, mnQAAir );  
  Run nr = reslice_warp_run( boldNormWarp, roRun );  
  Volume meanAll = strictmean( nr, "y", "null" )  
  Volume boldMask = binarize( meanAll, "y" );  
  snr = gsmoothRun( nr, boldMask, "6 6 6" );  
}
```

```
(Run or) reorientRun (Run ir,  
                      string direction) {  
  foreach Volume iv, i in ir.v {  
    or.v[i] = reorient(iv, direction);  
  }  
}
```

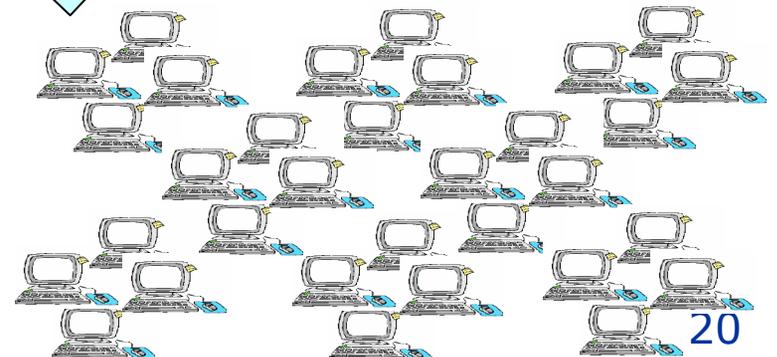
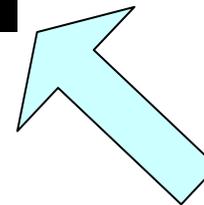
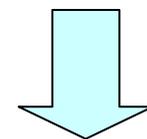
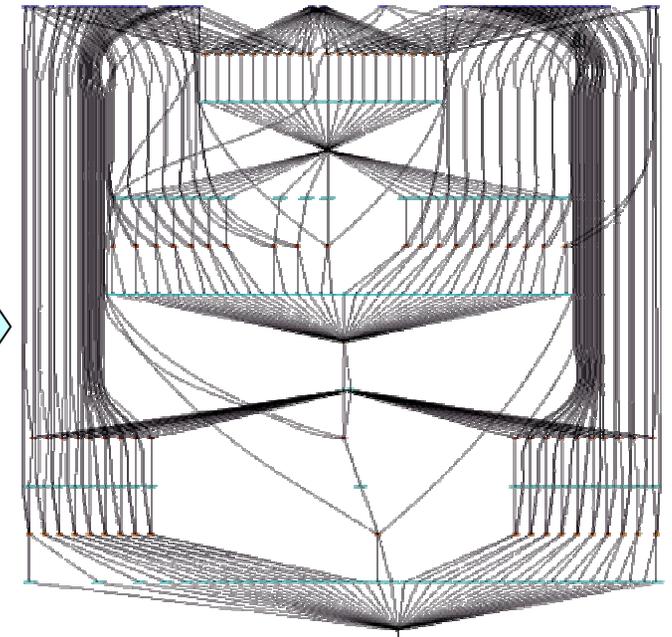
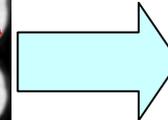
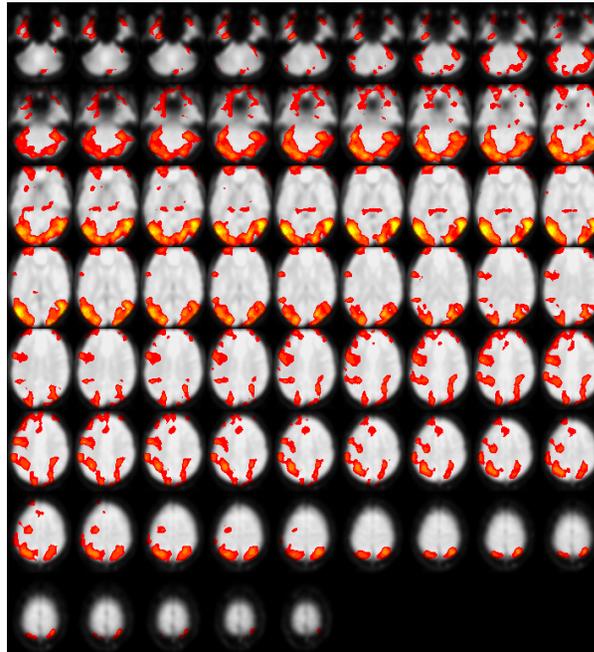


Dynamic Allocation for Dynamic Workloads





Many Users Analyze fMRI Data in Many Different Ways

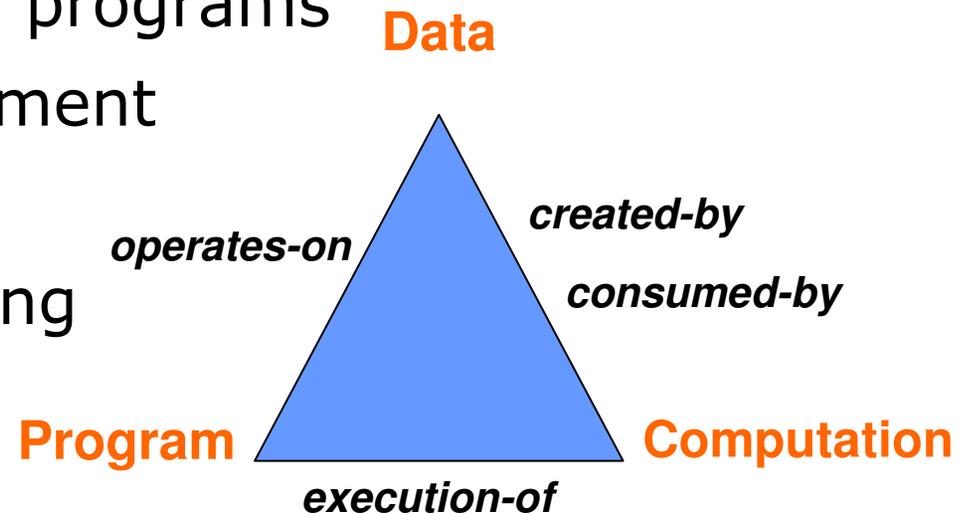


- Wide range of analyses
 - ◆ Testing, interactive analysis, production runs
 - ◆ Data mining
 - ◆ Parameter studies



The Virtual Data Concept

- Capture information about relationships among
 - ◆ Data (varying locations and representations)
 - ◆ Programs (& inputs, outputs, constraints)
 - ◆ Computations (& execution environments)
- Apply this information to:
 - ◆ Discovery of data and programs
 - ◆ Computation management
 - ◆ Provenance
 - ◆ Planning and scheduling
 - ◆ Performance optimization





Provenance Model

- Temporal aspect

- ◆ **Prospective** provenance

- Recipes for how to produce data
- Metadata annotations about procedures and data

- ◆ **Retrospective** provenance

- Invocation records of run time environments and resources used: site, host, executable, execution time, file stats ...

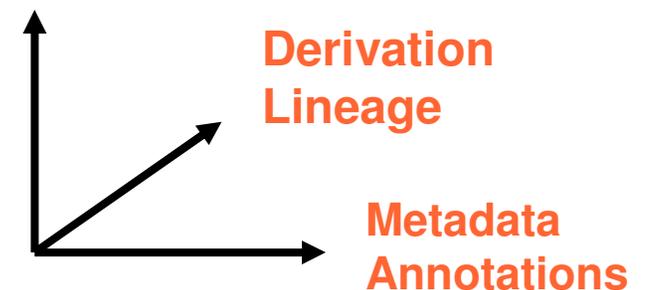
- Dimensional aspect

- ◆ **Virtual data** relationships

- ◆ **Derivation** lineage

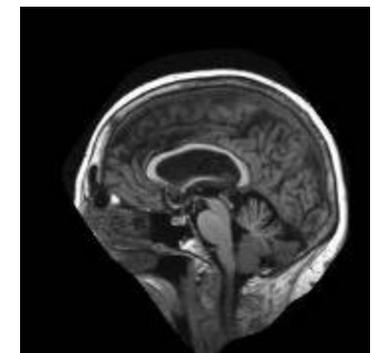
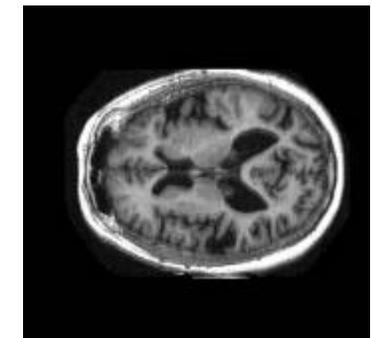
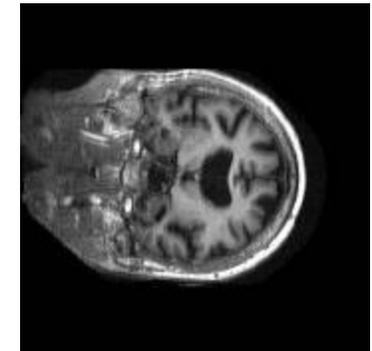
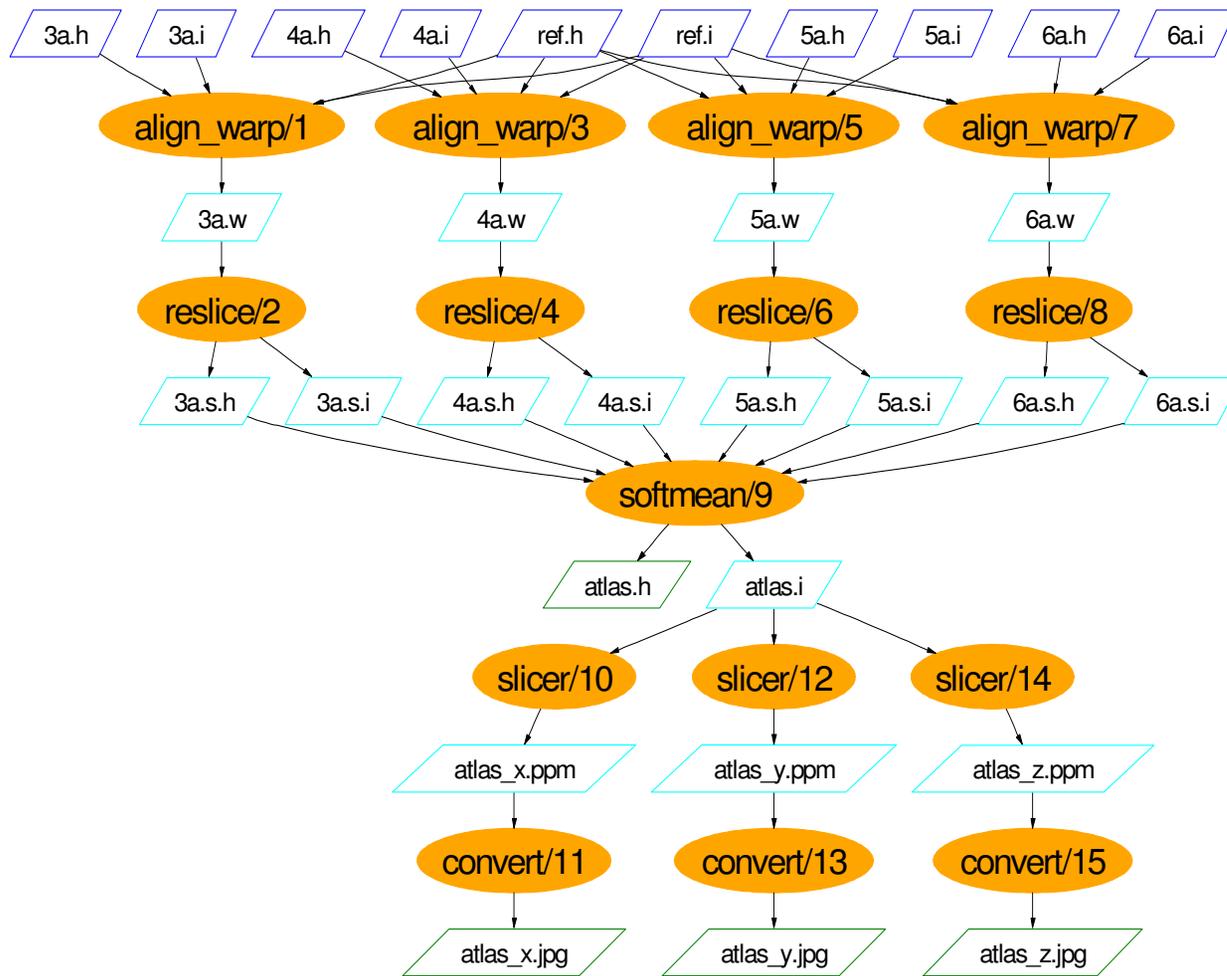
- ◆ **Metadata** annotations

Relationships





Query Context: fMRI Analysis



First Provenance Challenge, <http://twiki.ipaw.info/>



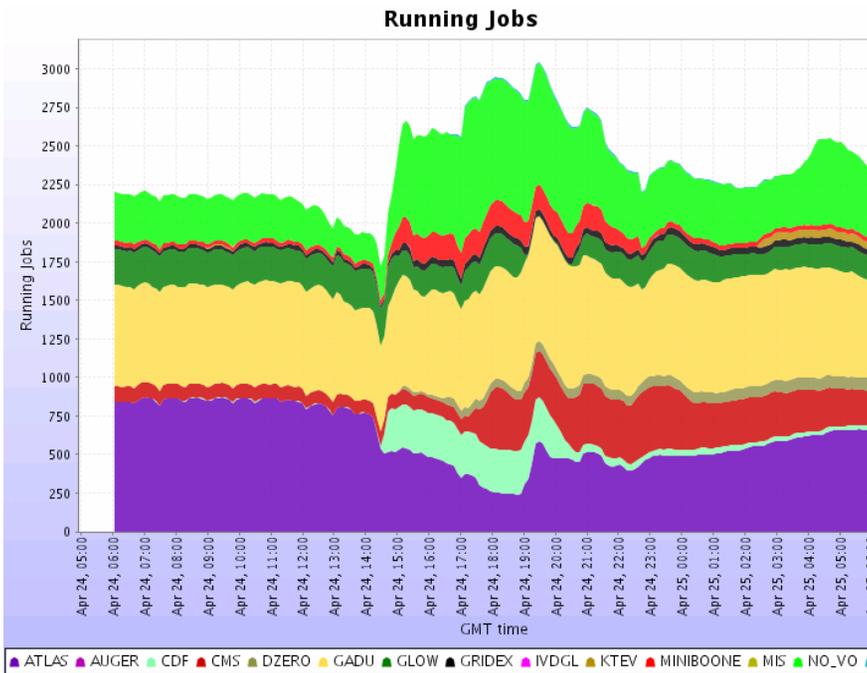
Query Examples

- Query by procedure signature
 - ◆ Show procedures that have inputs of type *subjectImage* and output types of *warp*
- Query by actual arguments
 - ◆ Show *align_warp* calls (including all arguments), with argument *model=rigid*
- Query by annotation
 - ◆ List anonymized subject images for young subjects:
 - Find datasets of type *subjectImage* , annotated with *privacy=anonymized* and *subjectType=young*
- Basic lineage graph queries
 - ◆ Find all datasets derived from dataset '5a'
- Graph pattern matching
 - ◆ Show me all output datasets of *softmean* calls that were aligned with *model=affine*
- Multi-dimensional query

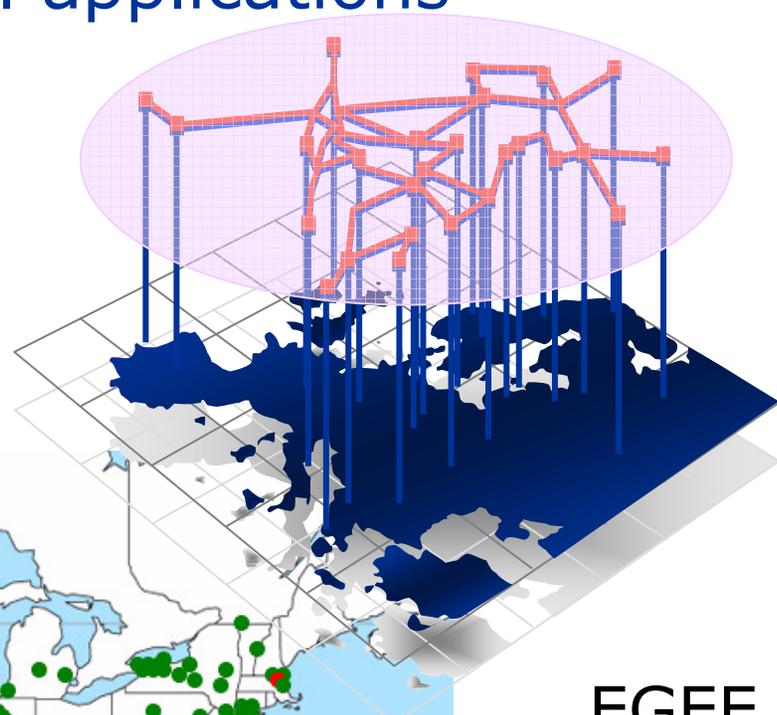
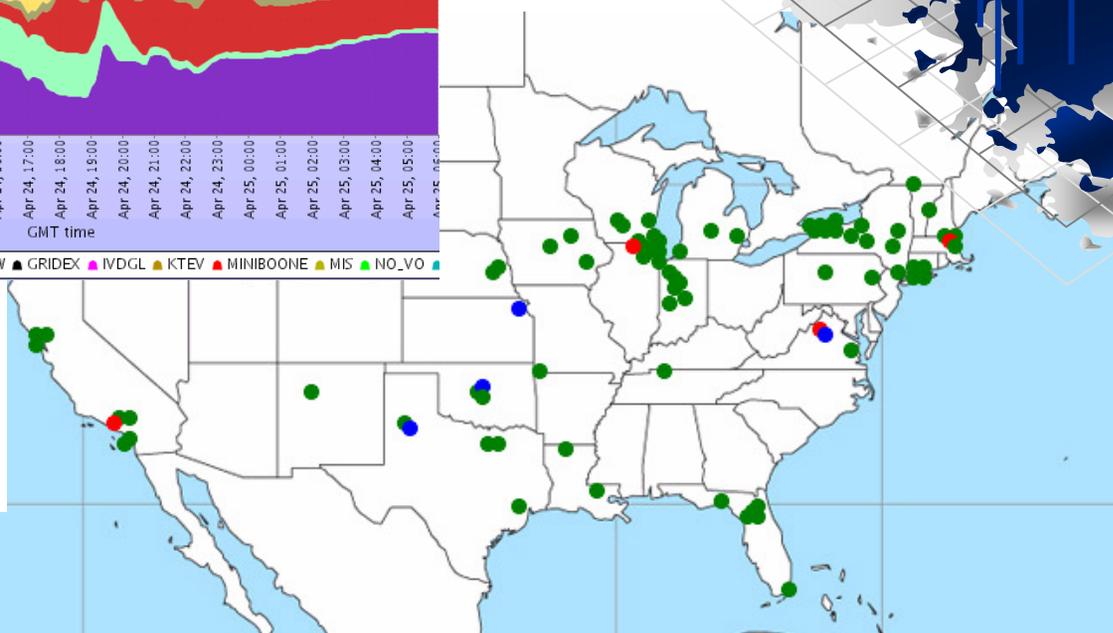


First Generation Grids: On-Demand/Batch Computing

Focus on aggregation of many resources for
massively (data-)parallel applications



Open Science Grid



EGEE



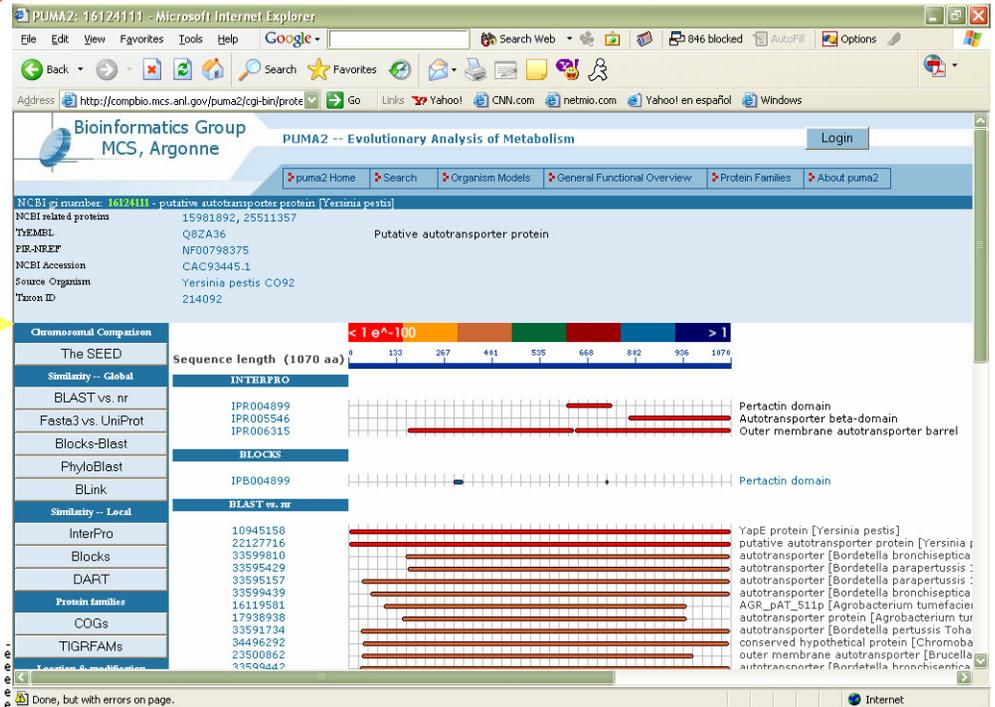
Globus₂₅



Integrating Data and Computing, on Demand

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

gi 23499780 gn REF_tigr BRA0013	gi 16080253 ref NP_391080.1	44.27	253	131	1	15	257	8	2603.7	e
gi 23499780 gn REF_tigr BRA0013	gi 123098409 ref NP_691875.1	43.48	253	133	2	16	258	5	2573.8	e
gi 23499780 gn REF_tigr BRA0013	gi 48637187 ref ZP_00294182.1	44.92	256	125	2	14	256	7	2591.1	e
gi 23499780 gn REF_tigr BRA0013	gi 52008400 gb AA025342.1	44.75	257	126	2	15	258	3	2561.9	e
gi 23499780 gn REF_tigr BRA0013	gi 48664015 ref ZP_00317908.1	44.49	245	134	1	13	257	5	2476.1	e
gi 23499780 gn REF_tigr BRA0013	gi 30348891 gb AA028934.1	39.53	253	138	3	18	257	5	2552.0	e
gi 23499780 gn REF_tigr BRA0013	gi 19655222 gb AA033939.1	40.64	251	138	1	17	256	10	2602.7	e
gi 23499780 gn REF_tigr BRA0013	gi 12735806 gb AA007757.1	43.03	251	130	4	18	256	11	2602.5	e
gi 23499780 gn REF_tigr BRA0013	gi 112597924 gb AA018599.2	46.70	162	96	1	62	243	5	1835.8	e
gi 23499780 gn REF_tigr BRA0013	gi 46363318 ref ZP_0026079.1	39.58	240	135	2	14	253	6	2361.8	e
REF_tigr BRA0013	gi 39933731 ref NP_946007.1	34.90	255						e-33 142.9	
REF_tigr BRA0013	gi 48782600 ref ZP_00279106.1	35.92	245						e-32 141.4	
REF_tigr BRA0013	gi 41407534 ref NP_960370.1	36.09	266						e-32 139.4	
REF_tigr BRA0013	gi 48851585 ref ZP_00305793.1	32.39	247						e-32 139.0	
REF_tigr BRA0013	gi 15966306 ref NP_386659.1	36.50	263						e-31 137.9	
REF_tigr BRA0013	gi 17548526 ref NP_521866.1	36.36	264						e-31 137.1	
gi 23499780 gn REF_tigr BRA0013	gi 51891730 ref VP_074421.1	38.87	247	136	7	18	256	1	2403.4	e
gi 23499780 gn REF_tigr BRA0013	gi 145881 gb AA23739.1	33.87	246	147	3	13	253	3	2404.4	e
gi 23499780 gn REF_tigr BRA0013	gi 25029334 ref NP_739388.1	35.20	250	147	4	15	256	6	2485.7	e
gi 23499780 gn REF_tigr BRA0013	gi 21220953 ref NP_536732.1	36.52	257	138	6	12	255	5	2545.7	e
gi 23499780 gn REF_tigr BRA0013	gi 46314029 ref ZP_00214635.1	33.86	254	153	2	12	259	3	2485.7	e
gi 23499780 gn REF_tigr BRA0013	gi 41406852 ref NP_959688.1	35.61	238	149	2	16	253	2	2309.8	e
gi 23499780 gn REF_tigr BRA0013	gi 11564471 ref NP_229523.1	35.69	255	144	5	12	256	2	2469.8	e
gi 23499780 gn REF_tigr BRA0013	gi 23470090 ref ZP_00125423.1	35.20	250	145	4	12	253	3	2439.8	e
gi 23499780 gn REF_tigr BRA0013	gi 24935279 gb AA064257.1	34.63	257	146	4	12	257	4	2499.8	e
gi 23499780 gn REF_tigr BRA0013	gi 48647651 ref ZP_0030315.1	35.05	258	145	9	12	257	4	2531.3	e
gi 23499780 gn REF_tigr BRA0013	gi 28851510 gb AA054587.1	36.40	250	142	4	12	253	3	2431.3	e
gi 23499780 gn REF_tigr BRA0013	gi 12737873 ref NP_770312.1	36.25	251	143	3	14	255	7	2491.3	e
gi 23499780 gn REF_tigr BRA0013	gi 1708836 sp P50198 LIDX_PSEPA	34.23	260	143	4	12	257	4	2491.7	e
gi 23499780 gn REF_tigr BRA0013	gi 33594146 ref NP_381792.1	34.17	240	148	5	18	256	6	2363.7	e
gi 23499780 gn REF_tigr BRA0013	gi 33594116 ref NP_381759.1	34.17	240	148	5	18	256	6	2363.7	e
gi 23499780 gn REF_tigr BRA0013	gi 3328306 ref NP_232830.1	34.20	241	143	5	18	256	6	2363.7	e



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

Natalia Maltsev et al., <http://compbio.mcs.anl.gov/puma2>



Second Generation Grids: **Service-Oriented Science**

- Empower many more users by enabling on-demand access to **services**
- Grids become an enabling technology for **service oriented science** (or business)
 - ◆ Grid infrastructures host services
 - ◆ Grid technologies used to build services



*Science
Gateways*

TeraGrid™
EMPOWERING DISCOVERY

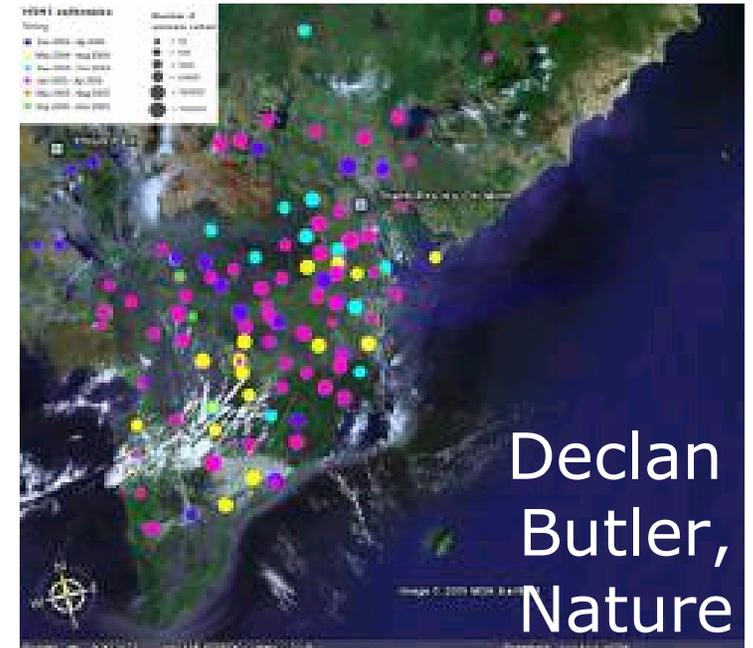


“Service-Oriented Science”, *Science*, 2005



“Web 2.0”

- Software as services
 - ◆ Data- & computation-rich network services
- Services as platforms
 - ◆ Easy composition of services to create new capabilities (“mashups”)—that themselves may be made accessible as new services
- Enabled by massive infrastructure buildout
 - ◆ Google projected to spend \$1.5B on computers, networks, and real estate in 2006
 - ◆ Many others are spending substantially
- Paid for by advertising





Service-Oriented Science: E.g., Virtual Observatories

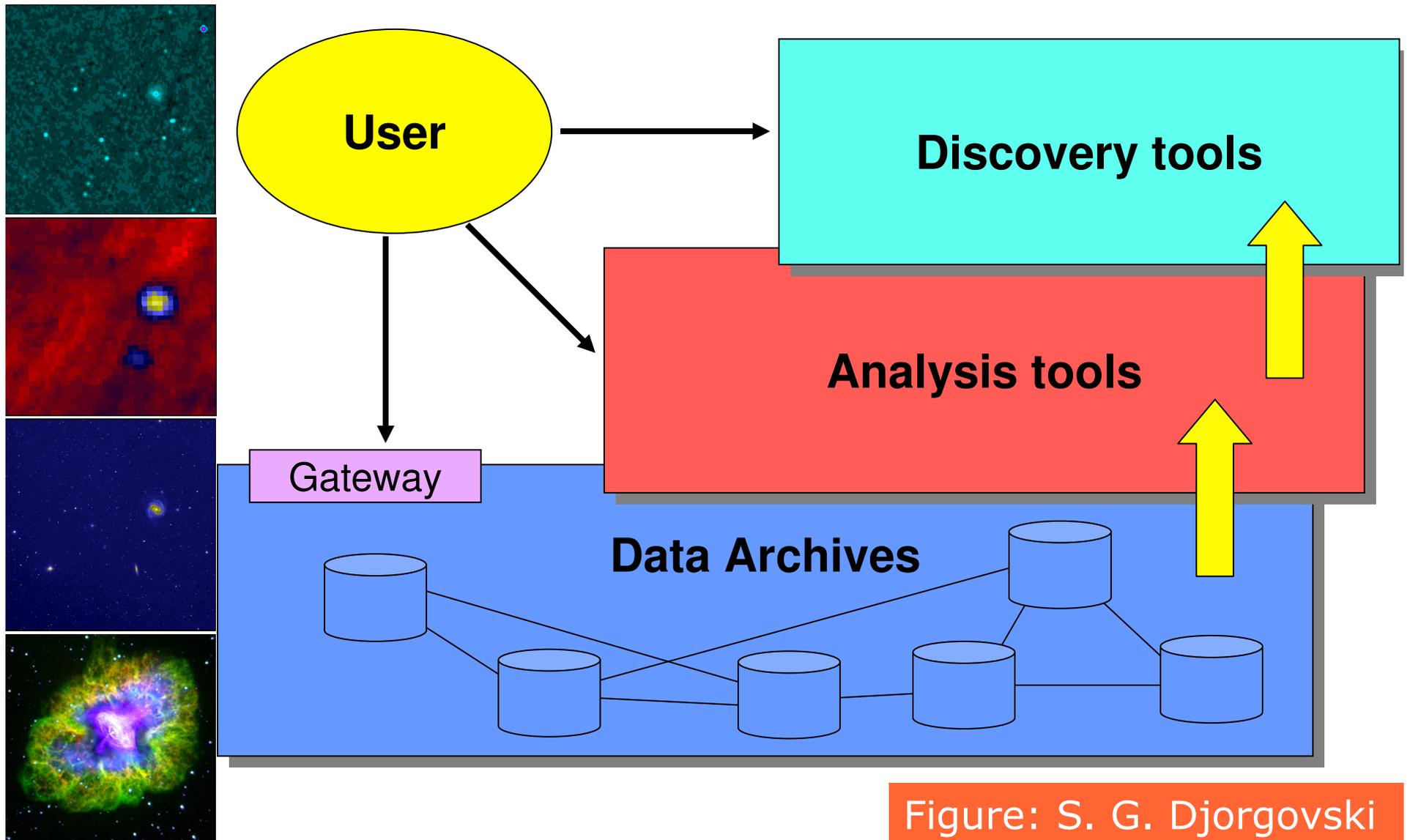
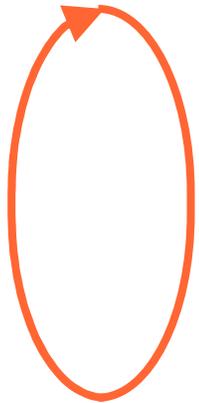


Figure: S. G. Djorgovski



Service-Oriented Science



People **create** services (data or functions) ...
which I **discover** (& decide whether to use) ...
& **compose** to create a new function ...
& then **publish** as a new service.

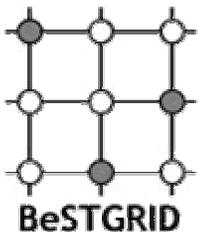
→ *I find "someone else" to **host** services,
so I don't have to become an expert in
operating services & computers!*



TeraGrid™
EMPOWERING DISCOVERY



→ *I hope that this "someone else" can
manage security, reliability, scalability, ...*

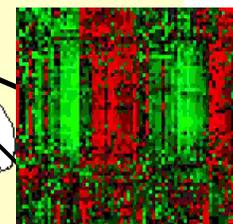
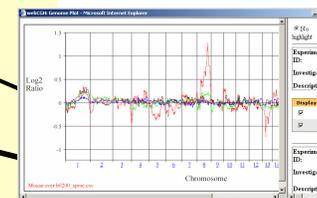
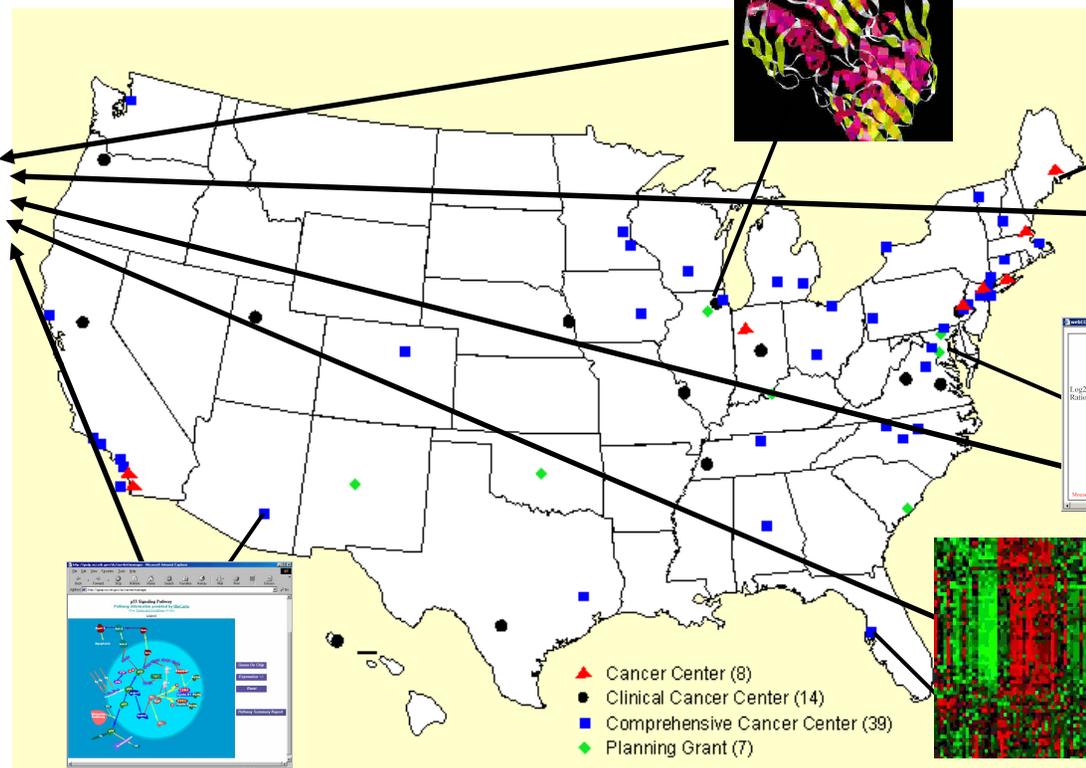
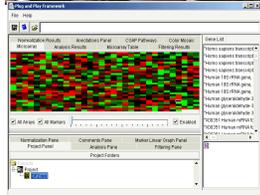


"Service-Oriented Science", *Science*, 2005



Service-Oriented Science & Cancer Biology

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



**Data
Integration!**

- ▲ Cancer Center (8)
- Clinical Cancer Center (14)
- Comprehensive Cancer Center (39)
- ◆ Planning Grant (7)



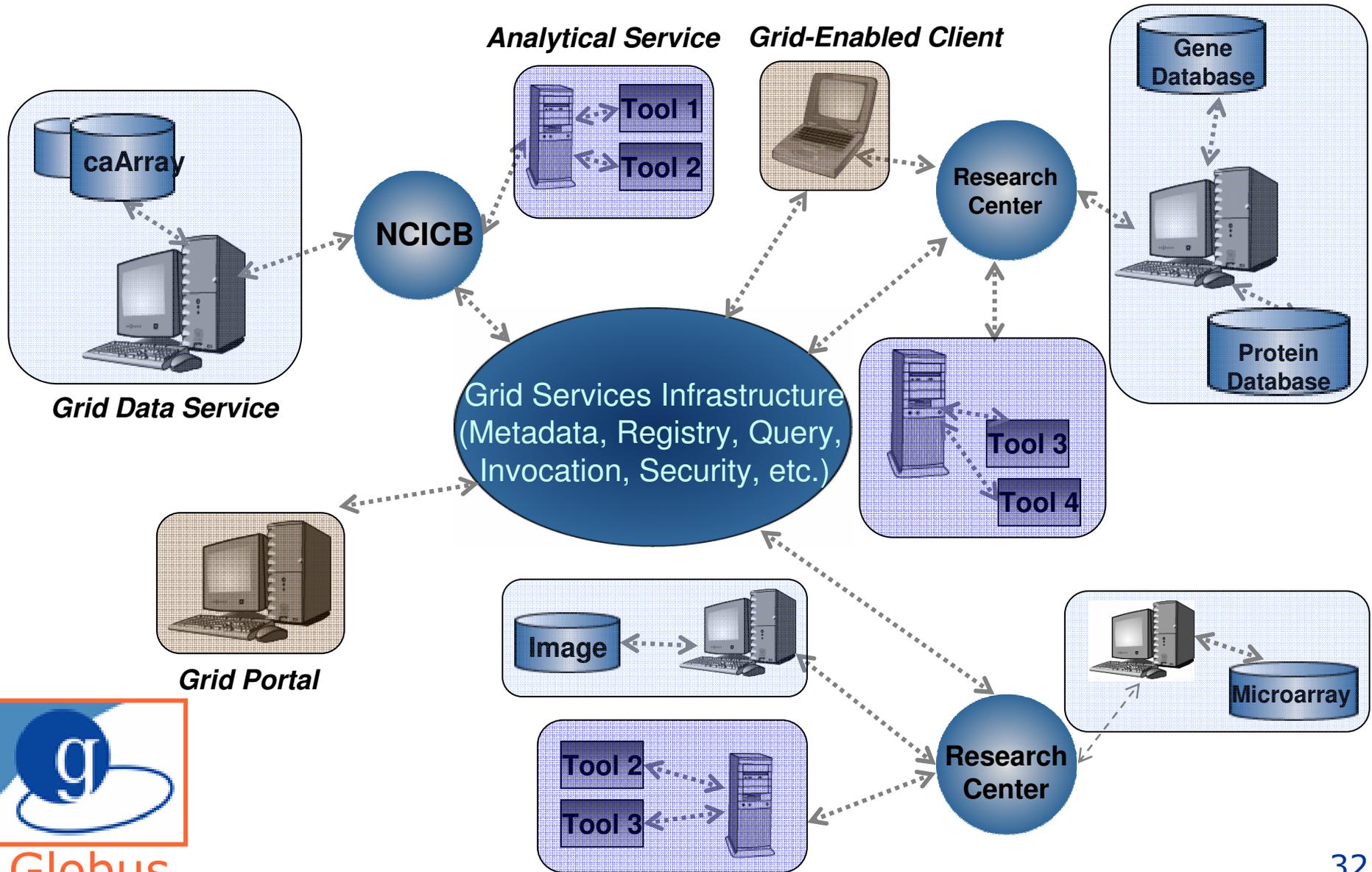
caBIG

cancer Biomedical
Informatics Grid





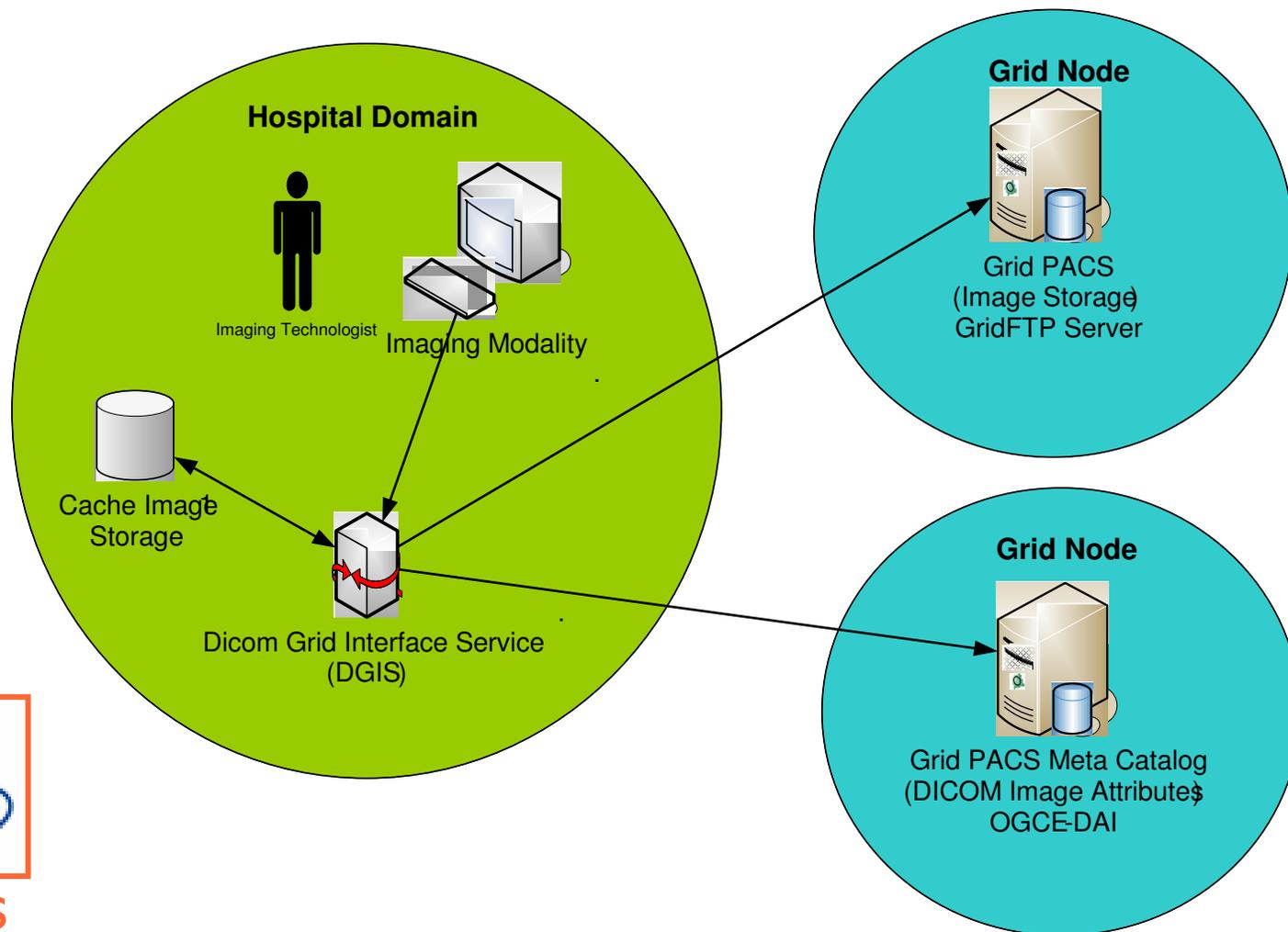
Cancer Bioinformatics Grid



Globus



MEDICUS: Management of DICOM Images

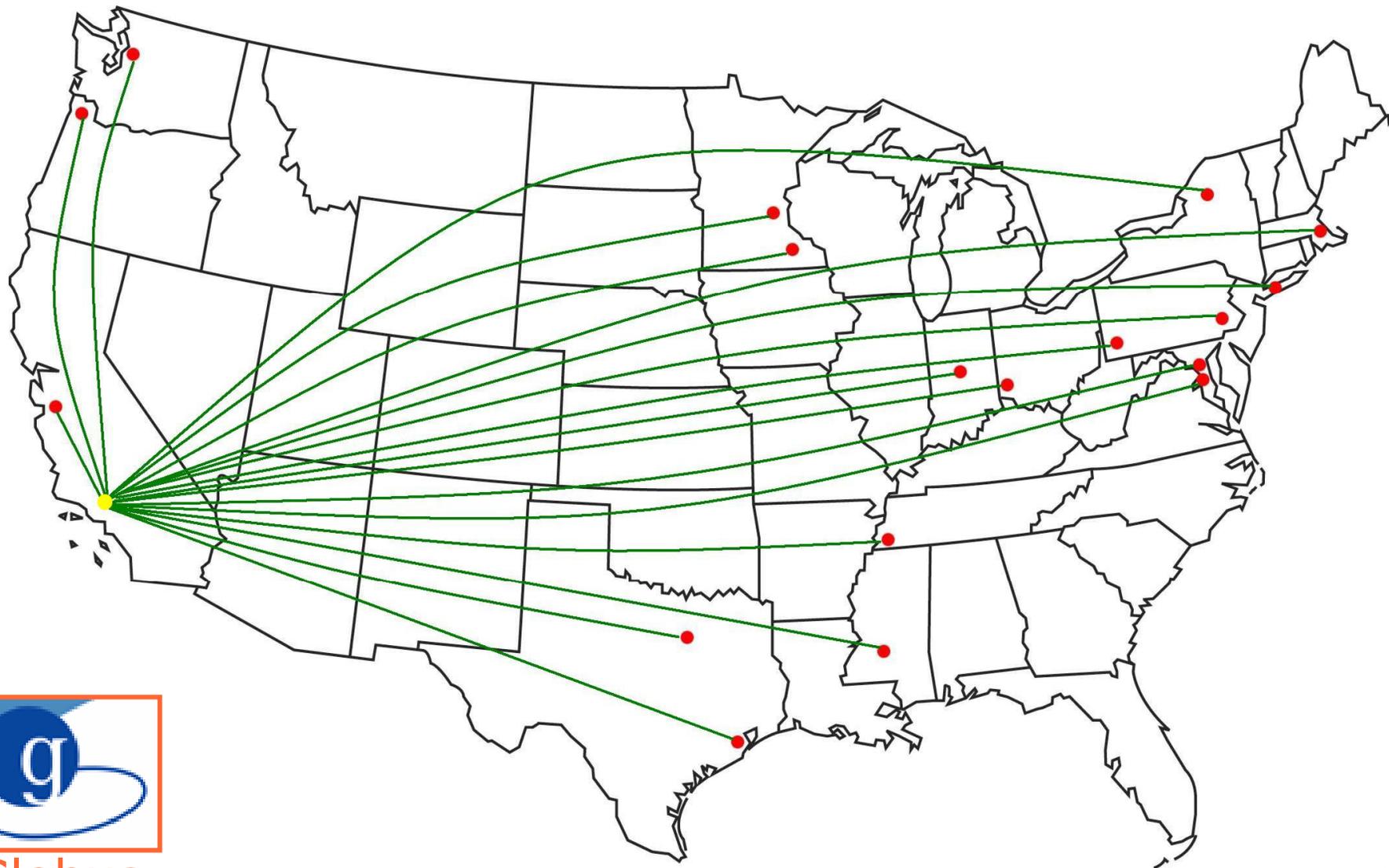


Globus

Stephan Erberich, Manasee Bhandekar, Ann Chervenak, et al.



Children's Oncology Grid: A MEDICUS Deployment



Globus



MEDICUS Under the Covers

Globus Toolkit Release 4

- DICOM images
 - ◆ Send (publish)
 - ◆ Query/Retrieve (discover)
 - Grid Archive
 - ◆ Fault tolerant
 - ◆ Bandwidth
 - Security
 - ◆ Authentication
 - ◆ Authorization
 - ◆ Cryptography
 - Access
 - ◆ Web portal
 - Applications
 - ◆ Computing
 - ◆ Data Mining
- *DICOM Grid Interface Service (DGIS)*
+
Meta Catalog Service (OGSA-DAI)
- *Data Replication Service (DRS)*
- *X.509 Certificates*
+
MyProxy Delegation
- *Grid Web Portal, OGCE / GridSphere*
- *GRAM, OGSA-DAI*

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

Five reasons to use Connotea

- [Save and organize](#) links to your references
- Easily [share](#) references with colleagues
- Access references from [any computer](#)
- [One click](#) is all it takes
- [Easy to use](#). Start creating your library today

[Learn more](#)

[Watch a short video](#) (2m 41s)

**Free online reference management
for all researchers, clinicians and
scientists**

Completely free, no download

[sign up](#)

[nature.com](#)

How it works

1 You find a reference

[Molecular structure of nucleic acids: a](#)
W JD, FH CRICK - Nature, 1953 - [View on Nature](#). 1953 Apr 25;171(4356):737-8 ...
[Web Search](#)

2 One click saves it on Connotea

Comments [Add To Connotea](#)

ature of nucleic acids: a

3 Add your own keywords for easy retrieval

DNA, genetics, geno

[genomics](#)

Explore

[Home](#)
[Recent Activity](#)
[Popular Links](#)
[Popular Tags](#)

Beginner's guide

[About Connotea](#)
[How It Works](#)
[Getting Started](#)
[Introductory Videos](#)

Account details

[My Account](#)
[Advanced Settings](#)
[My Library](#)
[My Community Profile](#)

Help

[Site Guide](#)
[FAQs](#)
[Mailing Lists](#)
[Contact Us](#)

Advanced

[Browser Buttons](#)
[Connotea Code](#)
[Connotea Web API](#)
[Community Pages](#)



Access Grid: The Power of Context



AG Venue Client

Venue Tools Navigation Help

https://vv3.mcs.anl.gov:8000/Venues/000000ff6f312bfe008c00dd00: Go

Argonne National Laboratory

My Venues

- Argonne Lobby
 - Big Horn
 - Bridgeport
 - Dantooine
 - Full Sail
 - Institution Lobby
 - Jack Frost
 - Kamar
 - Lucky Labrador
 - Talus
 - Test Room
 - Windmer

Participants

- Eric Olson
- Joe Insley
- Mark Hereld
- Michael E. Papka
- Susanne Lefvert
- Thomas D. Uram
- Ti Leggett

Data

- DataSnapshot01.jpg
- DataSnapshot02.jpg
- MeetingSlides.ppt
- Proposal.pdf

Services

Application Sessions

- Shared Movie Viewer - 12:01:04 PM Jun
- Shared Presentation - 10:15:59 AM Jun

-- 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

Display

Medical Education over Access Grid



Credit: Jonathan Silverstein, U.Chicago

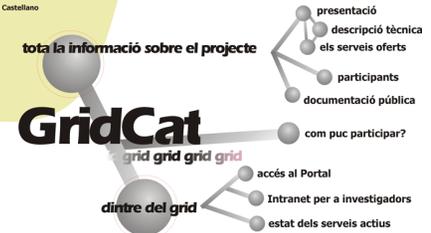
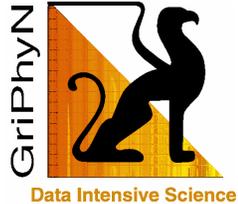
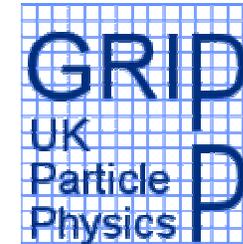
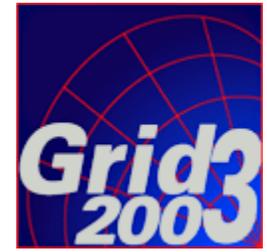
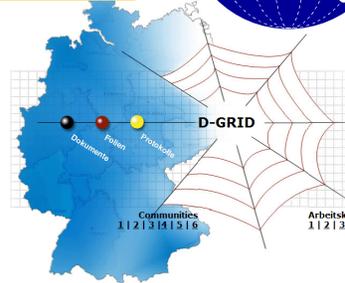
Access Grid and SARS



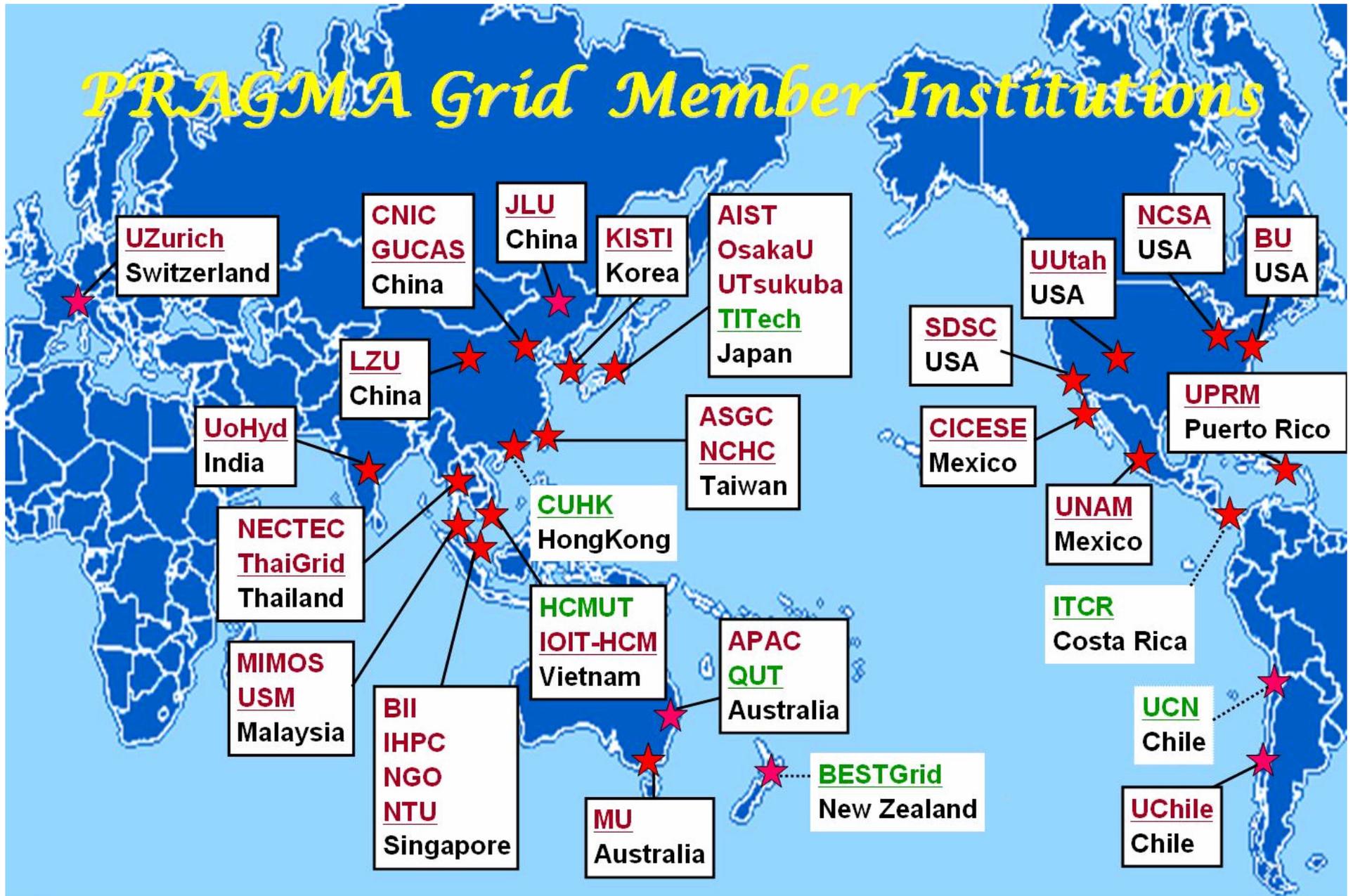
Global Communities



GRID.it
project




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:

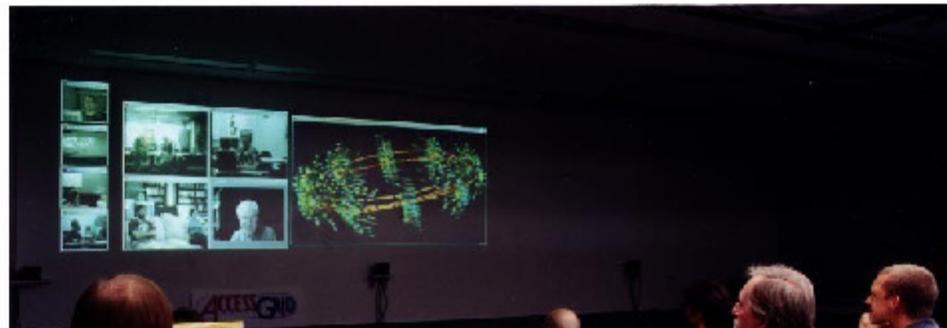
- [Create new account](#)
- [Request new password](#)

[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:



iGlobus Downloads Last 24 Hours

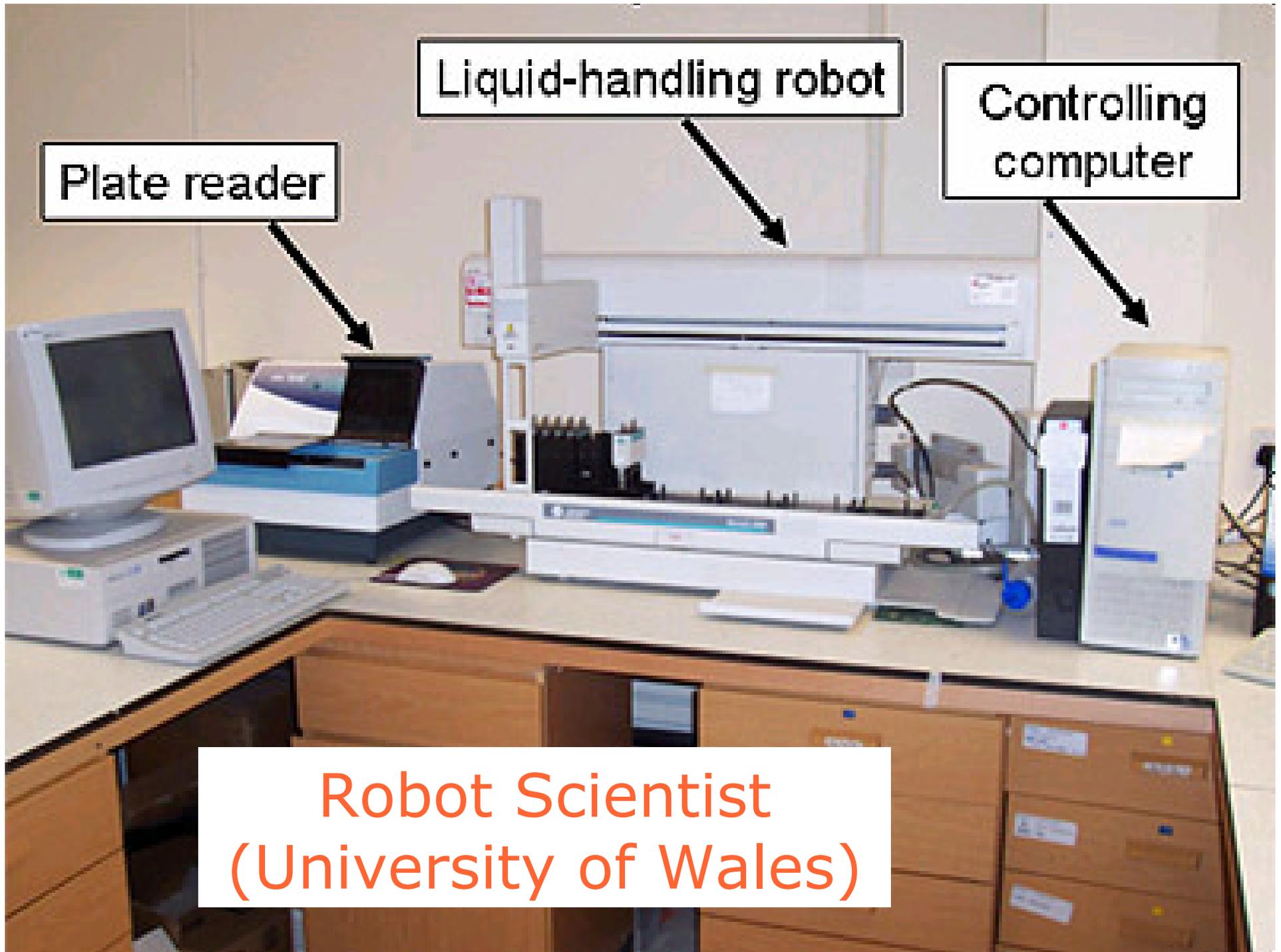




Automating Science Protocols

Protocol: “a predefined written procedural method in the design and implementation of experiments. This should establish standards that can be adequately assessed by peer review and provide for successful replication of results by others in the field.”

(Wikipedia)



Liquid-handling robot

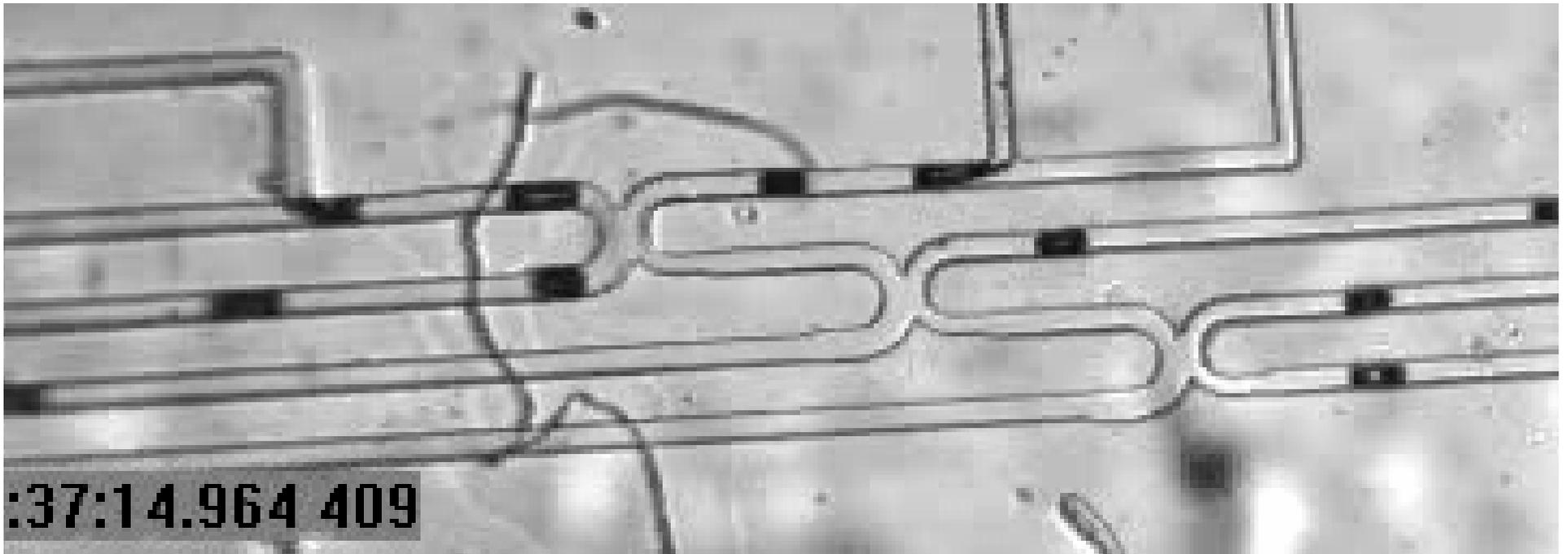
Controlling
computer

Plate reader

Robot Scientist
(University of Wales)



Microfluidic Bubble Logic (Prakash and Gershenfeld)





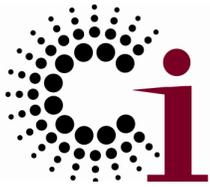
Lessons Learned

- The power of diversity and community
 - ◆ 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!



“Thinking Aloud” 10 Years 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
 - ◆ Individual, team, community, citizen science
- 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 ... Tbit/sec)
 - ◆ Physical resources
 - ◆ Amount of data (MB ... EB), data quality
 - ◆ Complexity of questions asked & answered
 - ◆ Complexity of models
 - ◆ Degree of trust, shared language, etc.
- Integration with the physical world
 - ◆ Active sensors
 - ◆ Automated experimental protocols
 - ◆ Integrate manufacturing and problem solving



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