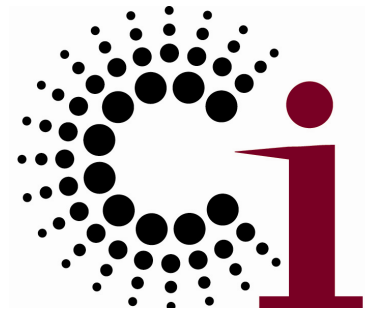


Grid

Enabling Open Science

Ian Foster

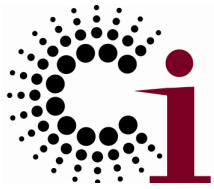


Computation Institute
Argonne National Lab & University of Chicago



Abstract

Rapid advances in both science and information technology are driving the emergence of "eScience." Grid technologies play a crucial role in eScience by enabling resource and service federation across organizational boundaries, supporting on-demand access to computing resources, and allowing the formation and operation of distributed, multi-organizational collaborations. eScience and Grid also require new tools, infrastructure, and policies. I will discuss opportunities, achievements, and challenges in these related areas.



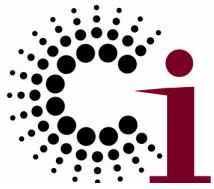
What is the Grid?



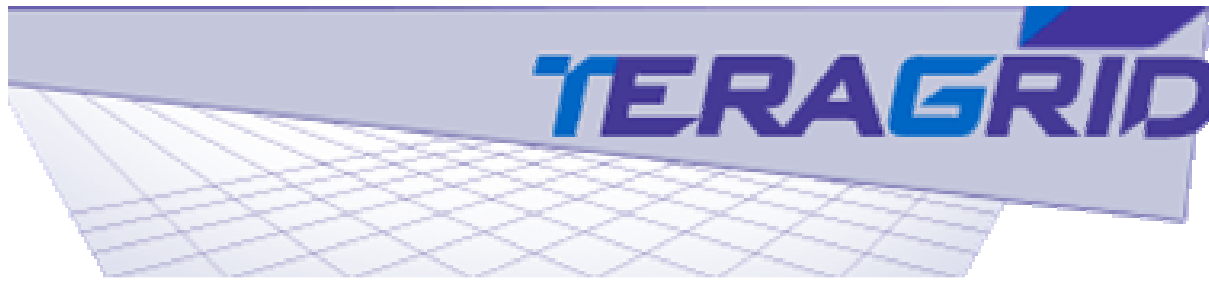
“**The Grid** is an international project that looks in detail at a terrorist cell operating on a global level and a team of American and British counter-terrorists who are tasked to stop it”

Gareth Neame, BBC's head of drama



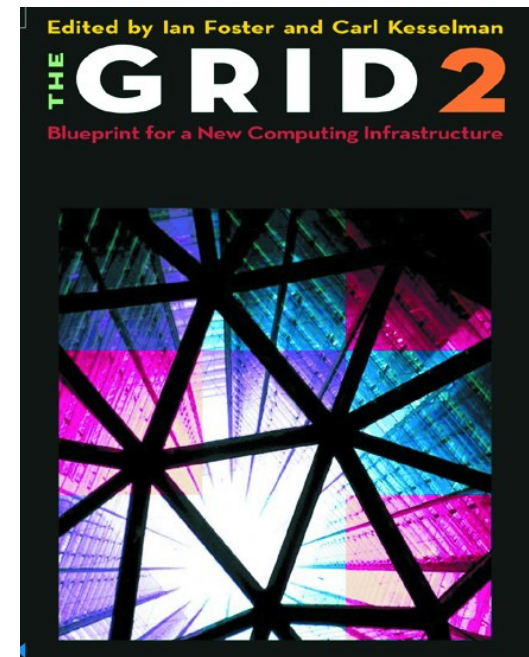
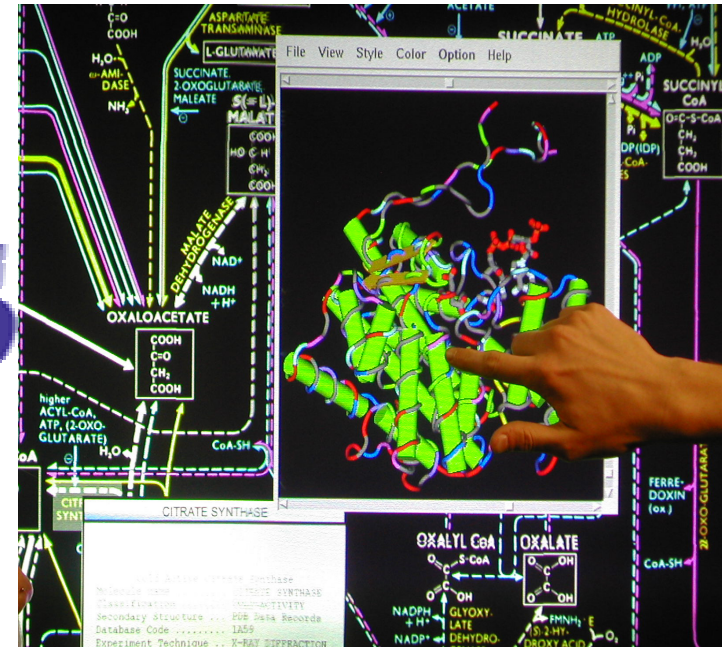


Well, Not Exactly!

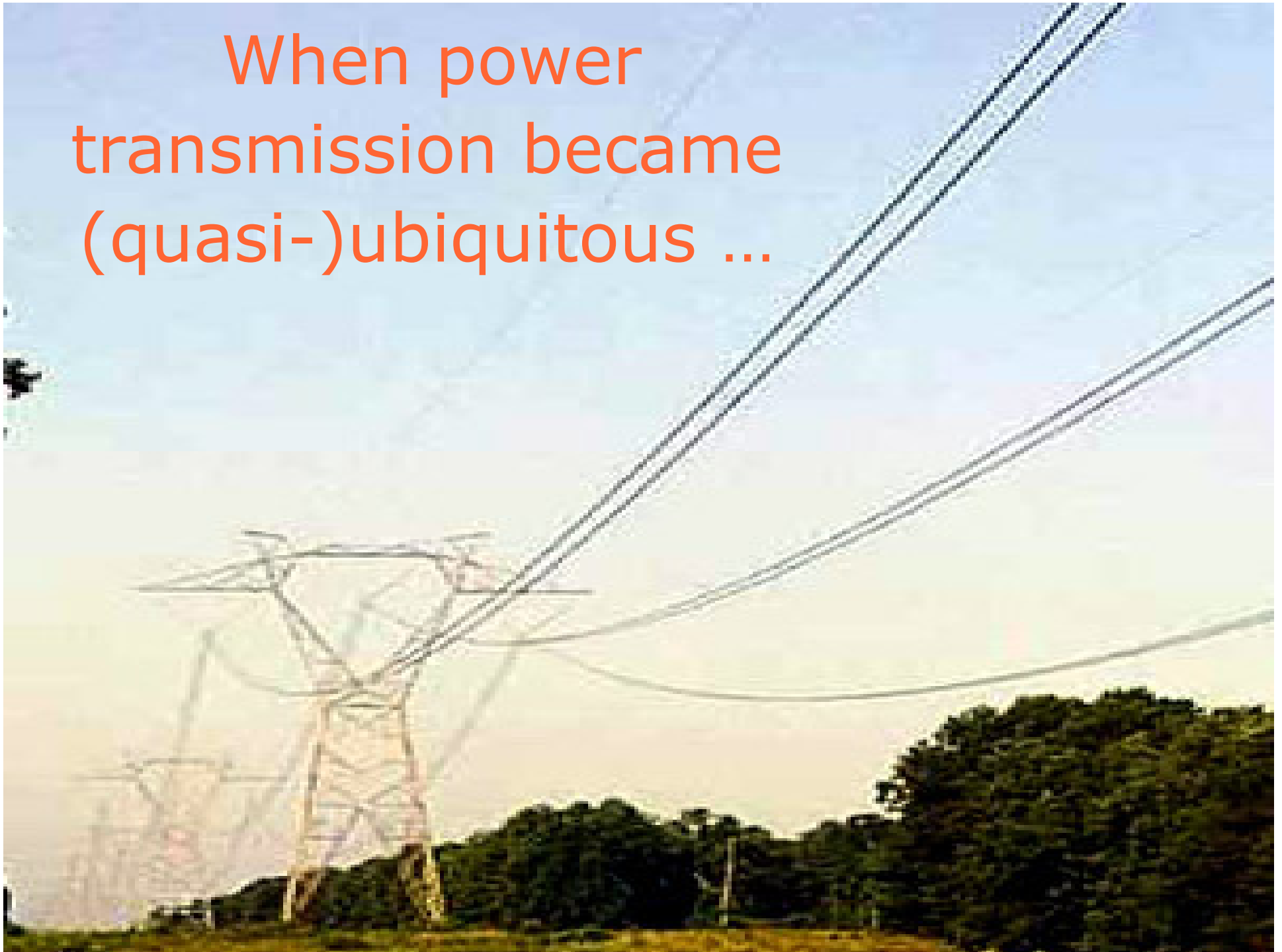


“The Grid is an international project that looks in detail at scientific collaborations operating on a global level and a team of computer scientists who are tasked to enable it”

At least, that’s where it started ...

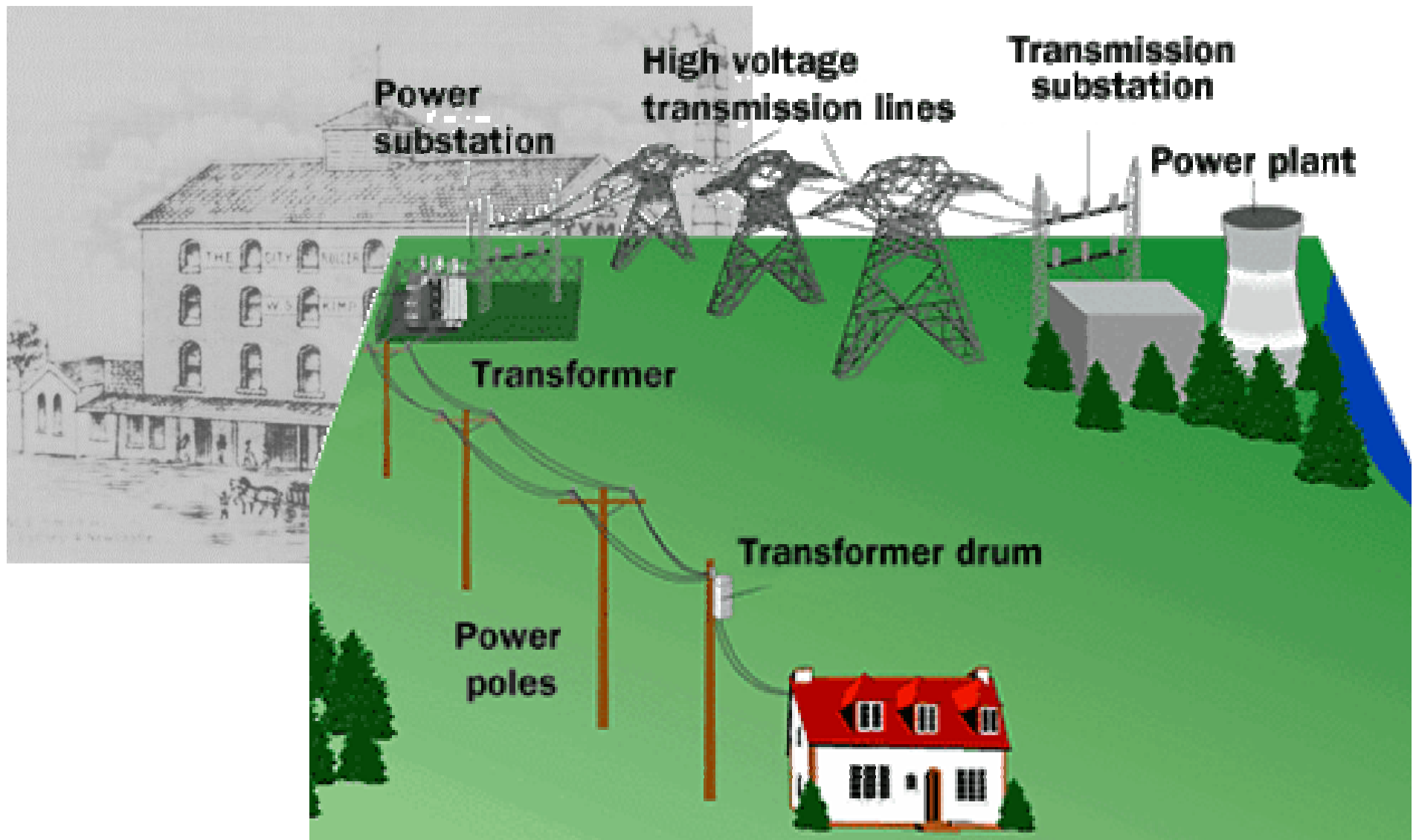


When power
transmission became
(quasi-)ubiquitous ...





We No Longer Had to Travel to Power Plants





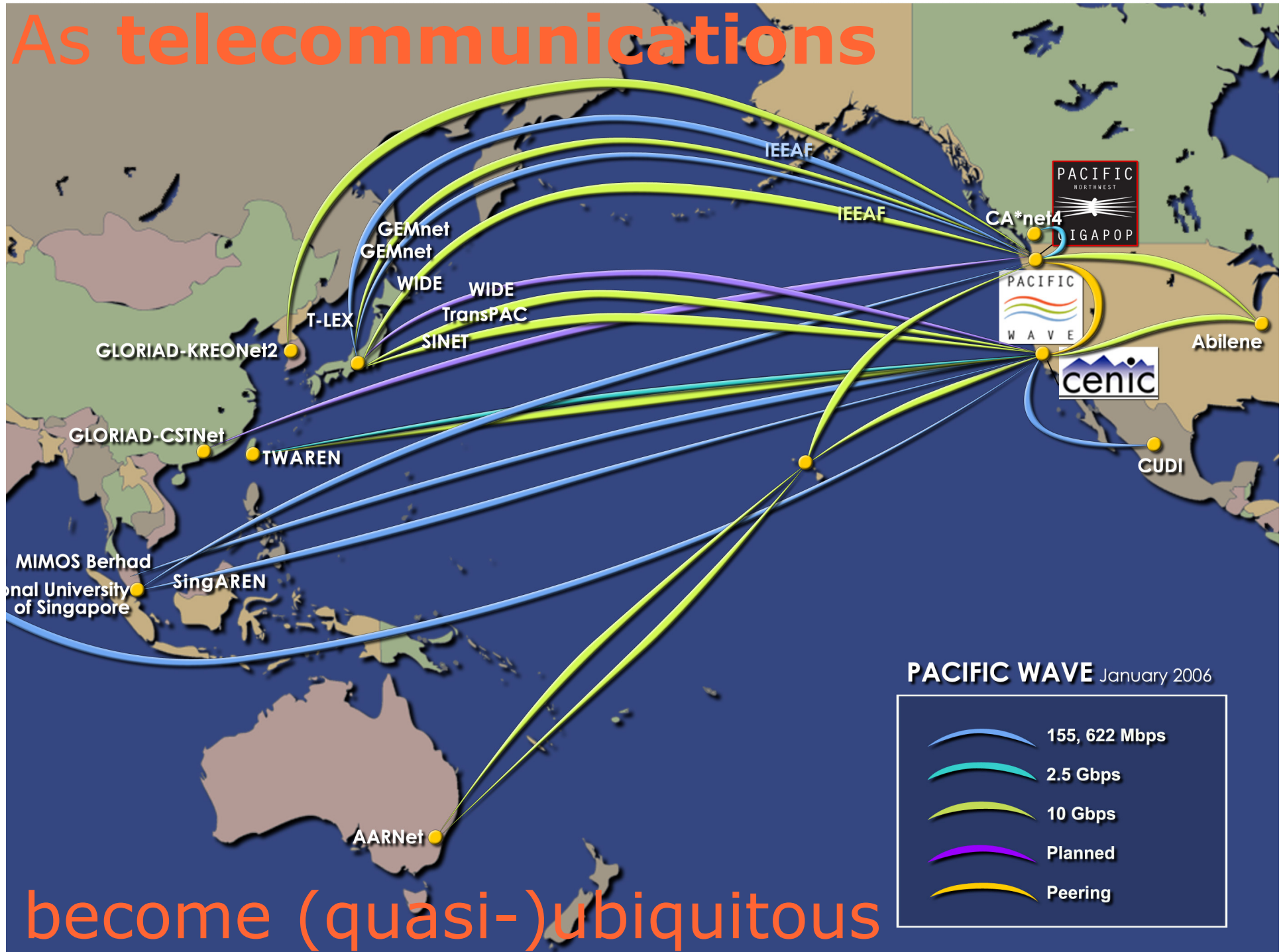
We Invented New Tools





We Worked in New Ways

As telecommunications

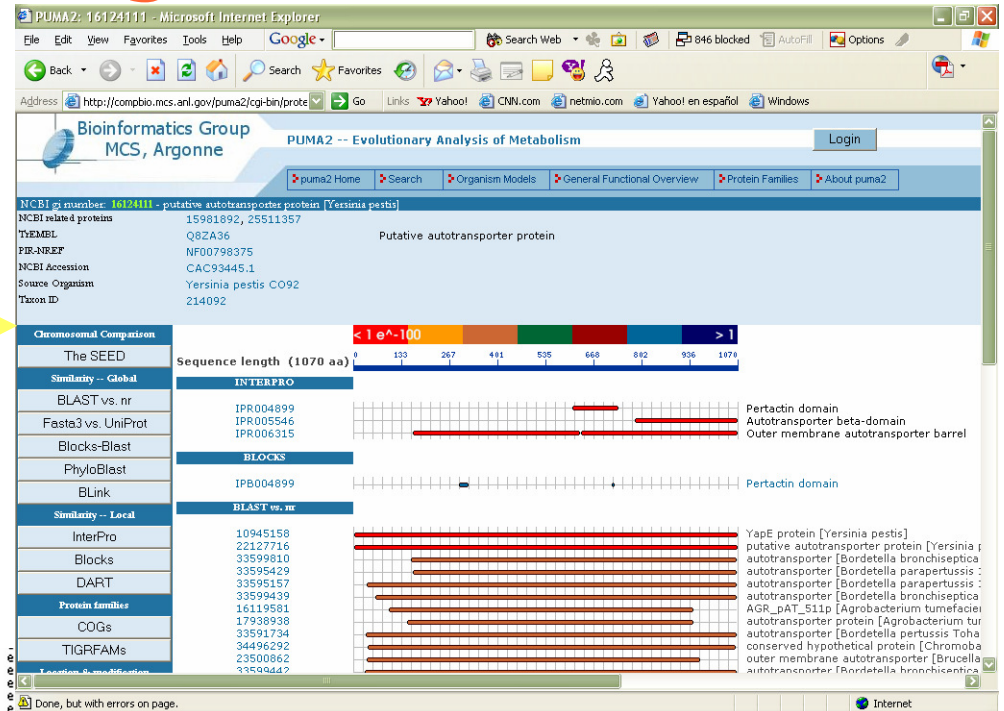




We Can Access Computing on Demand

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 23098409 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 27358806 gb AA007757.1	43.03	251	130	4	18	256	11	2602.5	e
gi 23499780 gn REF_tigr BRA0013	gi 12597924 gb AA018599.2	46.70	162	96	1	62	243	5	1856.8	e
gi 23499780 gn 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						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-30 137.1	
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 AA023739.1	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_636732.1	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 115644471 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_00303115.1	36.05	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	3	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	231	143	5	18	256	6	2363.7	e



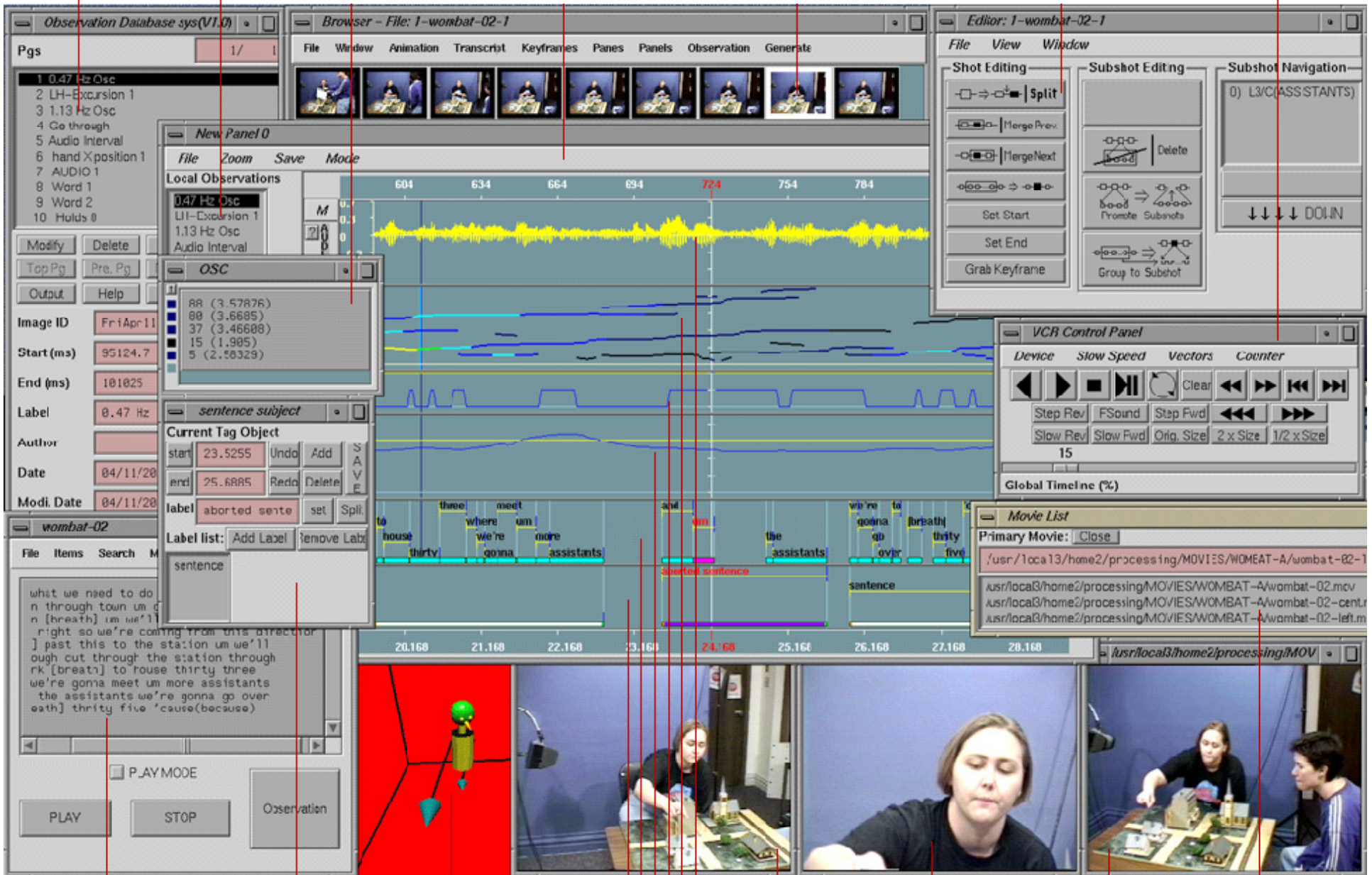
Back Office Analysis on Grid
Millions of BLAST, BLOCKS, etc., on OSG and TeraGrid

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

Global Observation Database (View) Graph-E Observ:

We Can Invent New Tools

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

We Can Work in New Ways: Access Grid and SARS

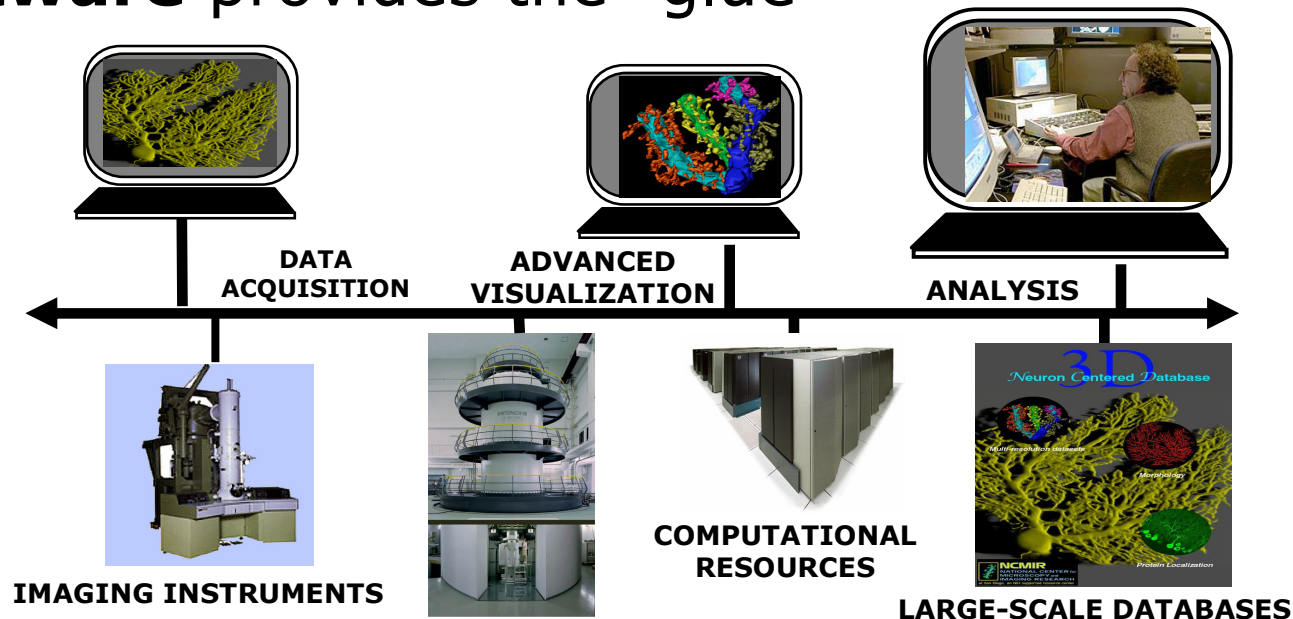




Grid: Unifying Concept & Technology

Enable **on-demand access** to, and **federation** of, diverse resources

- Computers, storage, data, people, ...
- Resources can be distributed, heterogeneous
- **Networks & protocols** provide the connectivity
- **Software** provides the “glue”



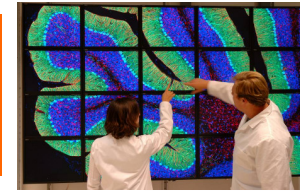
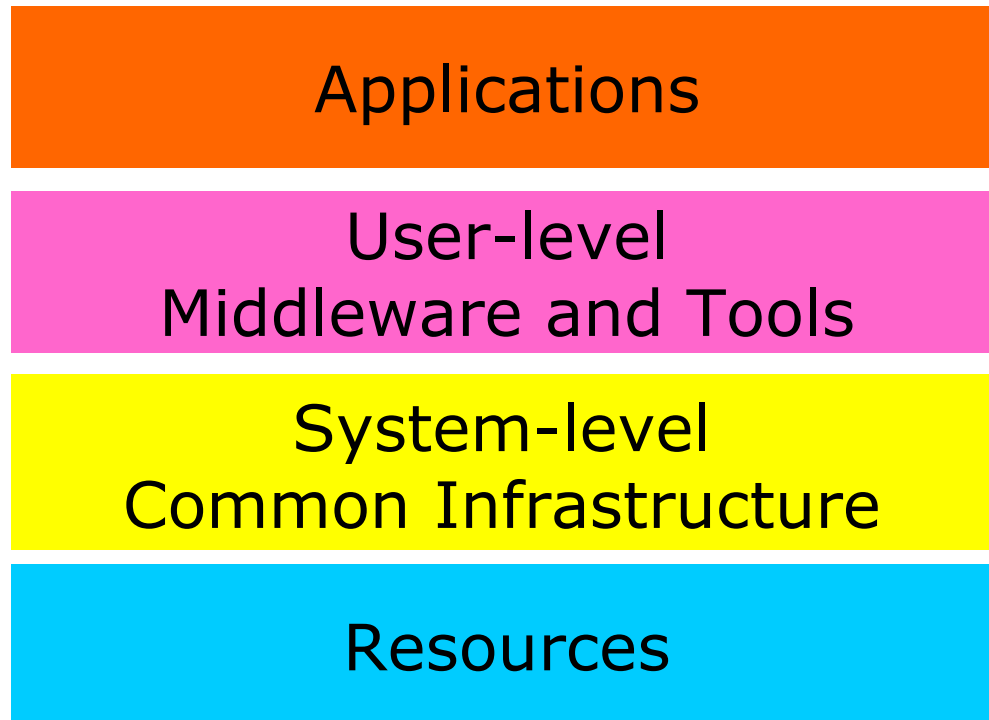


Technology, Infrastructure, & Standards

interoperability



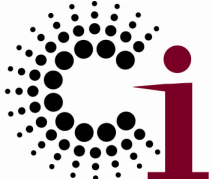
integration



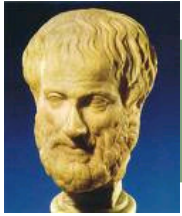
TeraGrid™
EMPOWERING DISCOVERY



For example 14



Emergence of New Problem Solving Methodologies



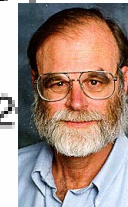
Empirical



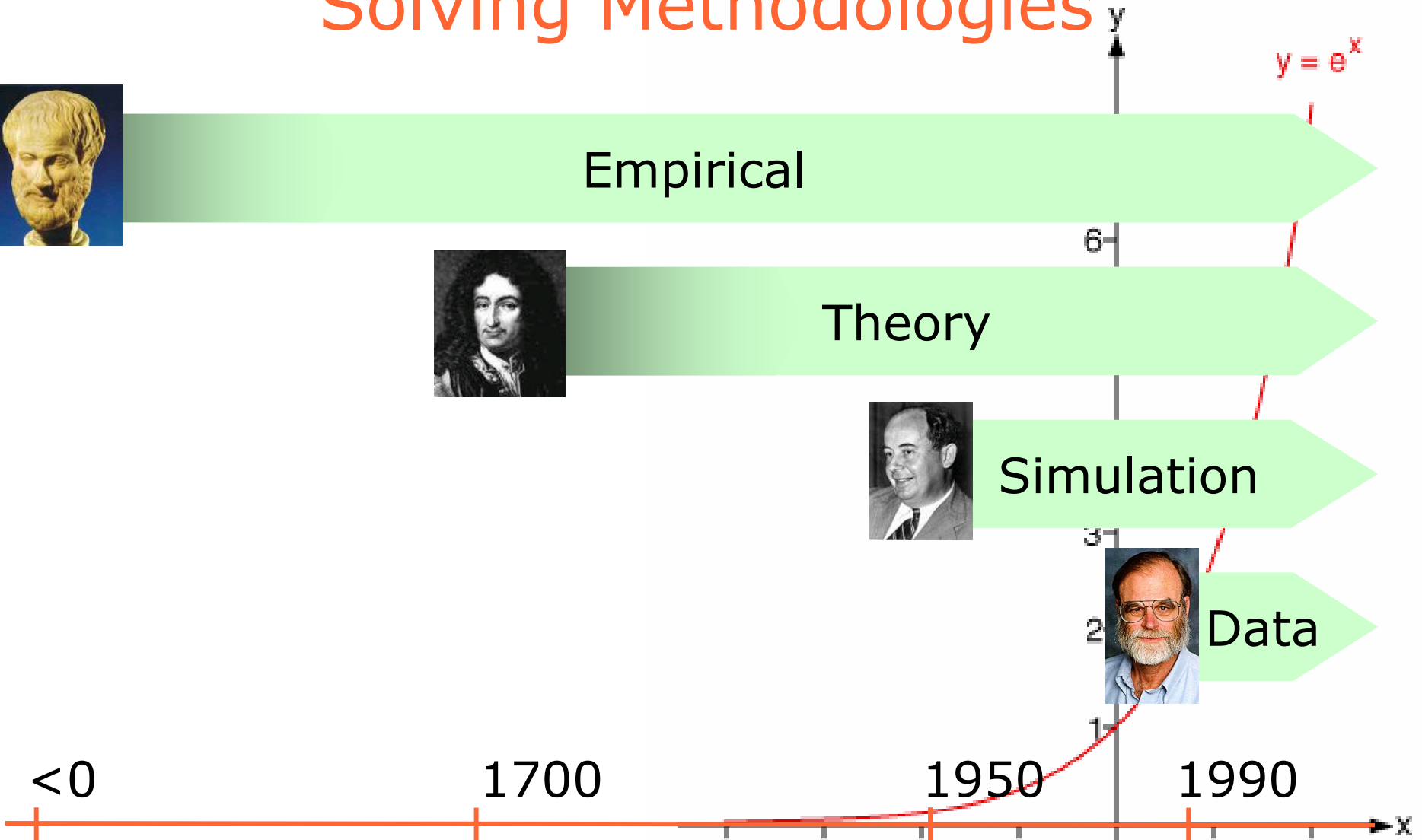
Theory



Simulation



Data

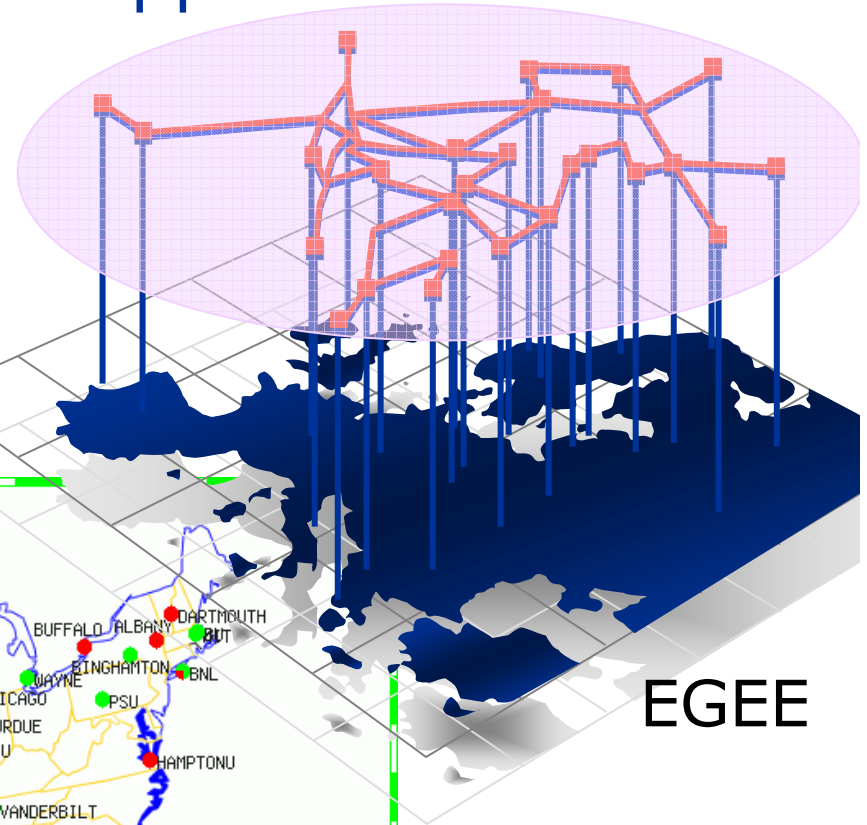
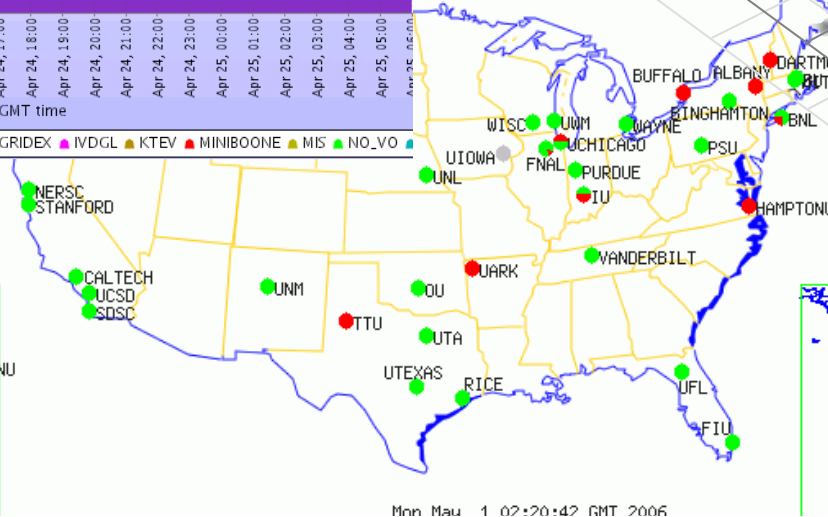
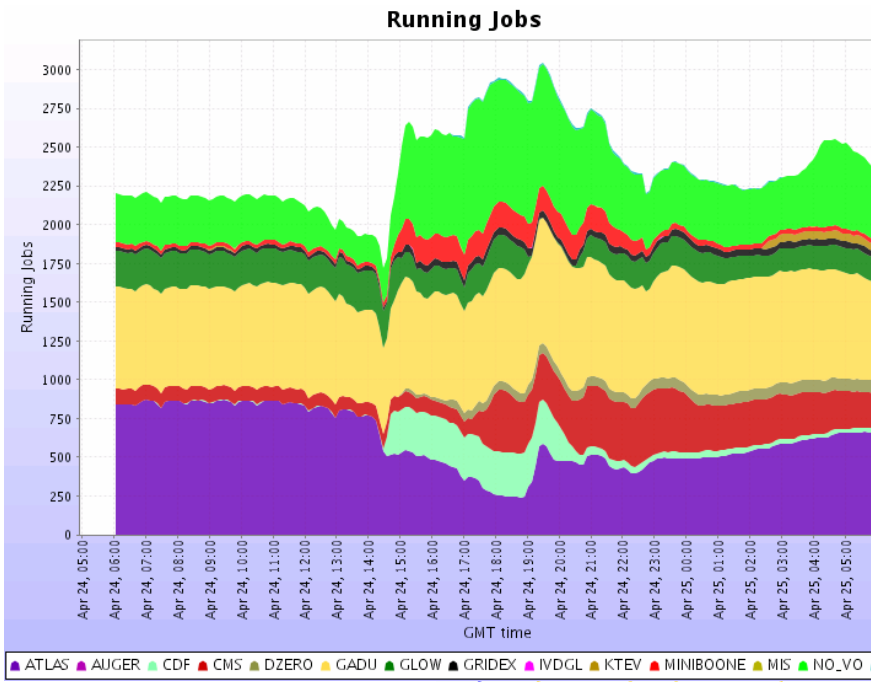


eScience: When brute force doesn't work anymore (Szalay)



First Generation Grids: Batch Computing

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





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

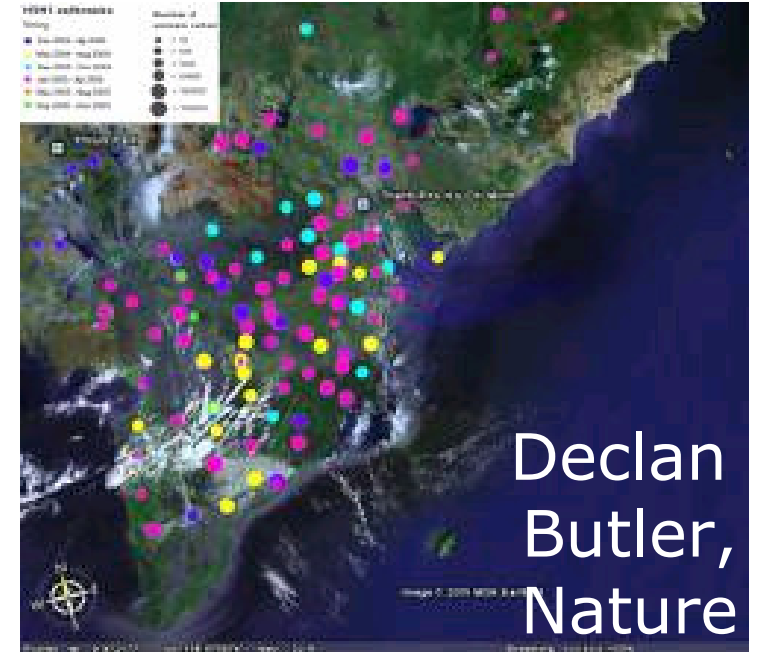


“Service-Oriented Science”, *Science*, 2005



“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

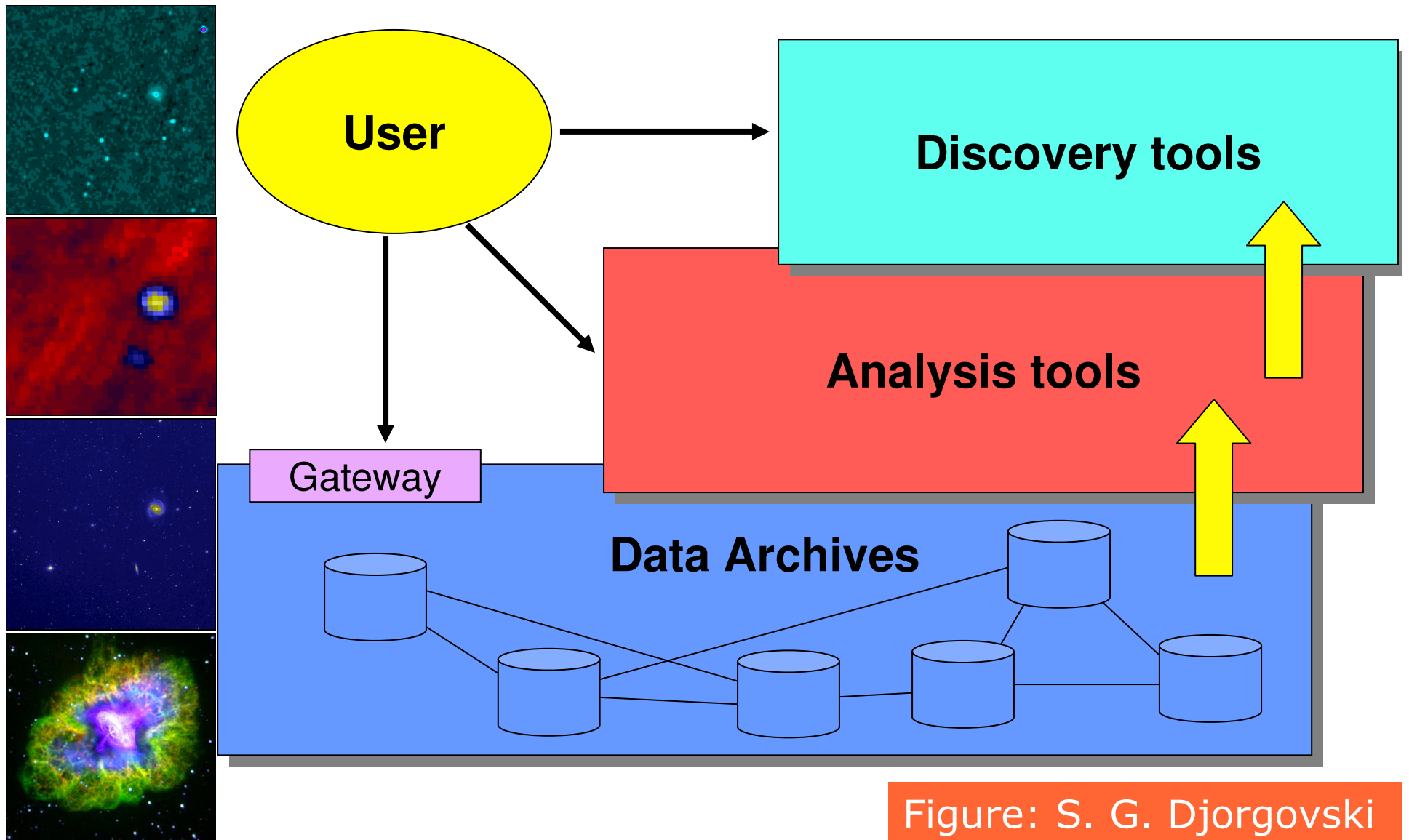
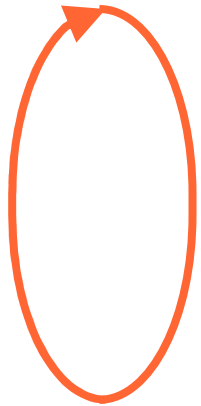


Figure: S. G. Djorgovski



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!



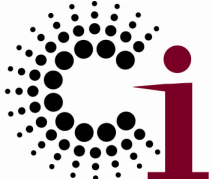
TeraGrid™
EMPOWERING DISCOVERY



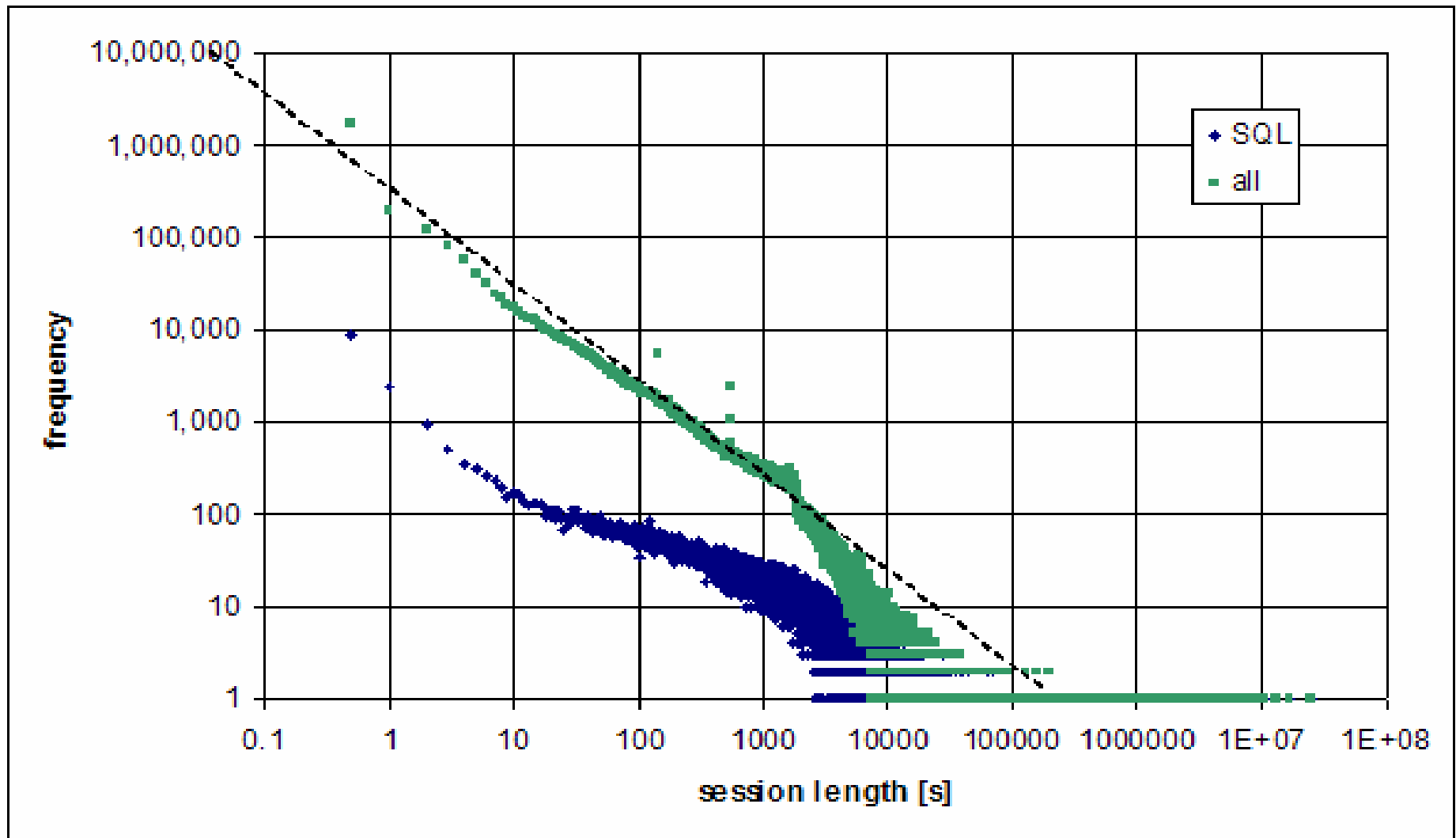
→ I hope that this "someone else" can
manage security, reliability, scalability, ...

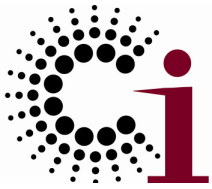


"Service-Oriented Science", *Science*, 2005



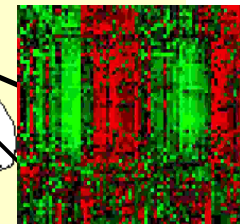
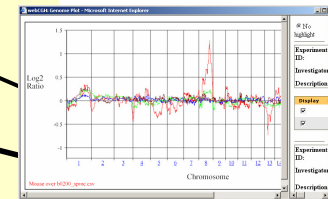
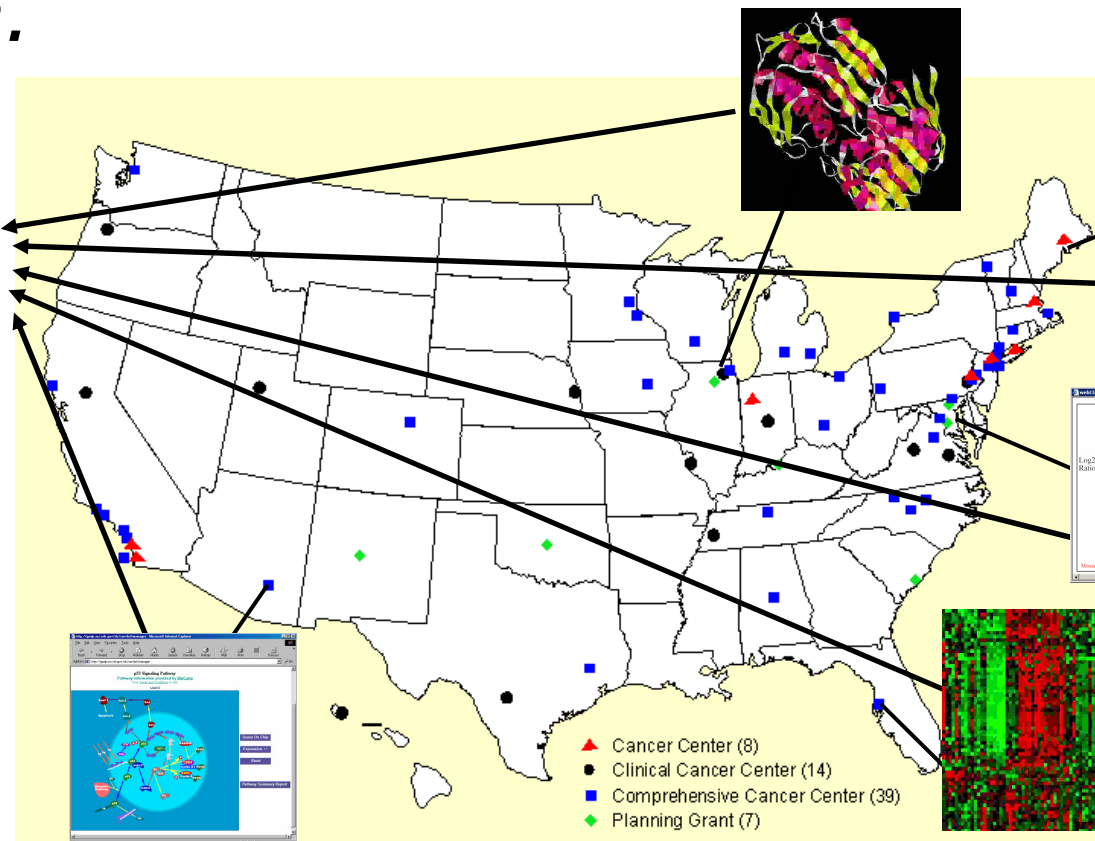
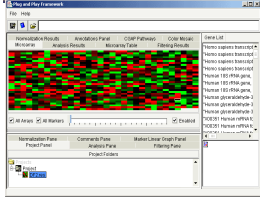
Skyserver Sessions (Thanks to Alex Szalay)



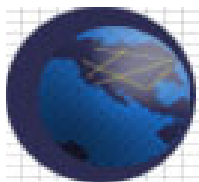


Service-Oriented Science & Cancer Biology

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

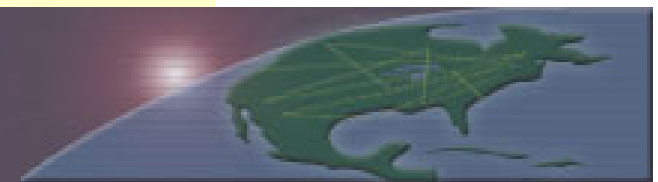


**Data
Integration!**



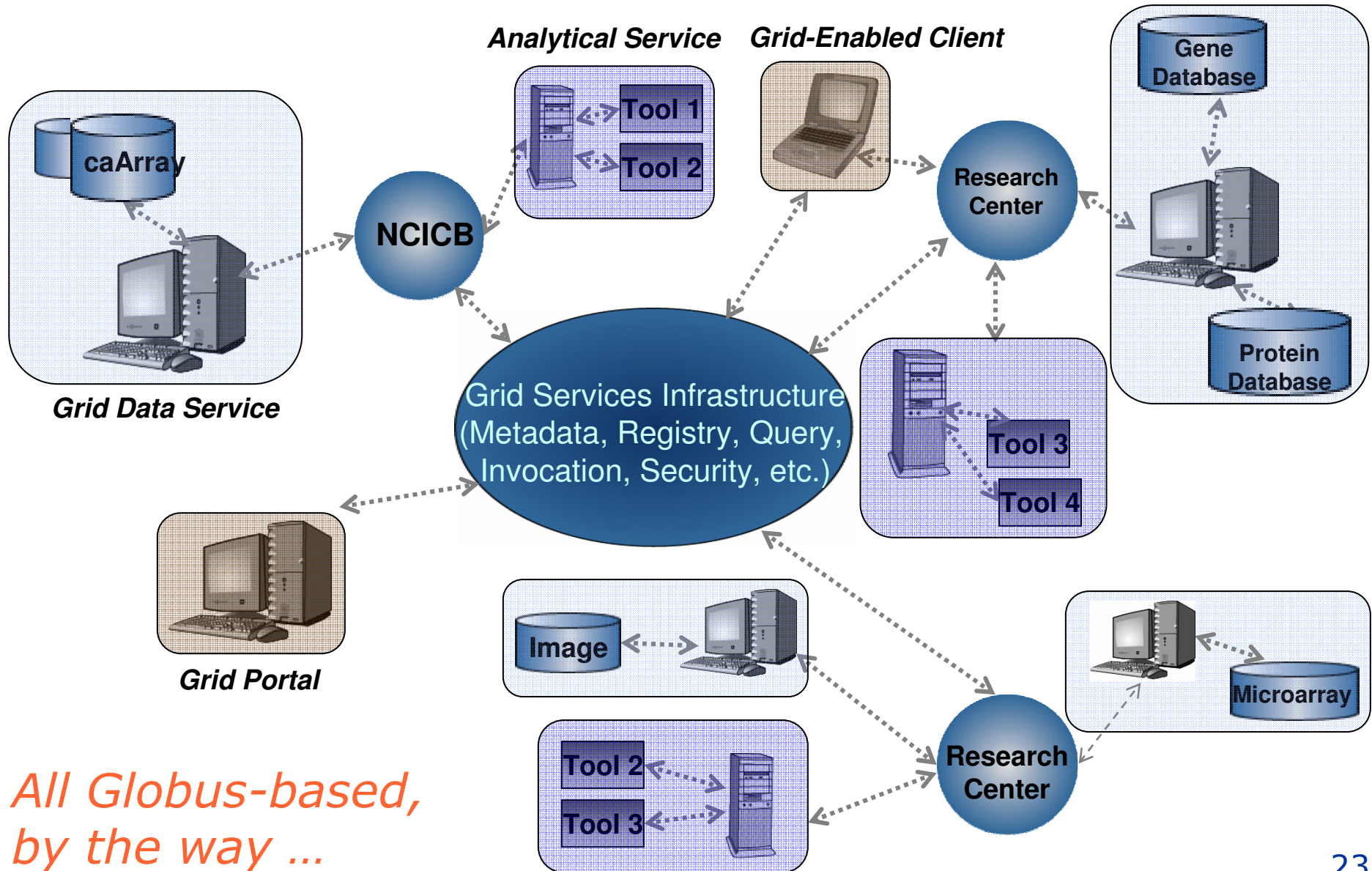
caBIG

cancer Biomedical
Informatics Grid





Cancer Bioinformatics Grid



*All Globus-based,
by the way ...*

Thailand joins the grid

By [Don Sambandaraksa](#), Bangkok Post

Wednesday, May 16 2007 11:34 AM



TALKBACK



PRINT STORY



EMAIL STORY

Bangkok Post

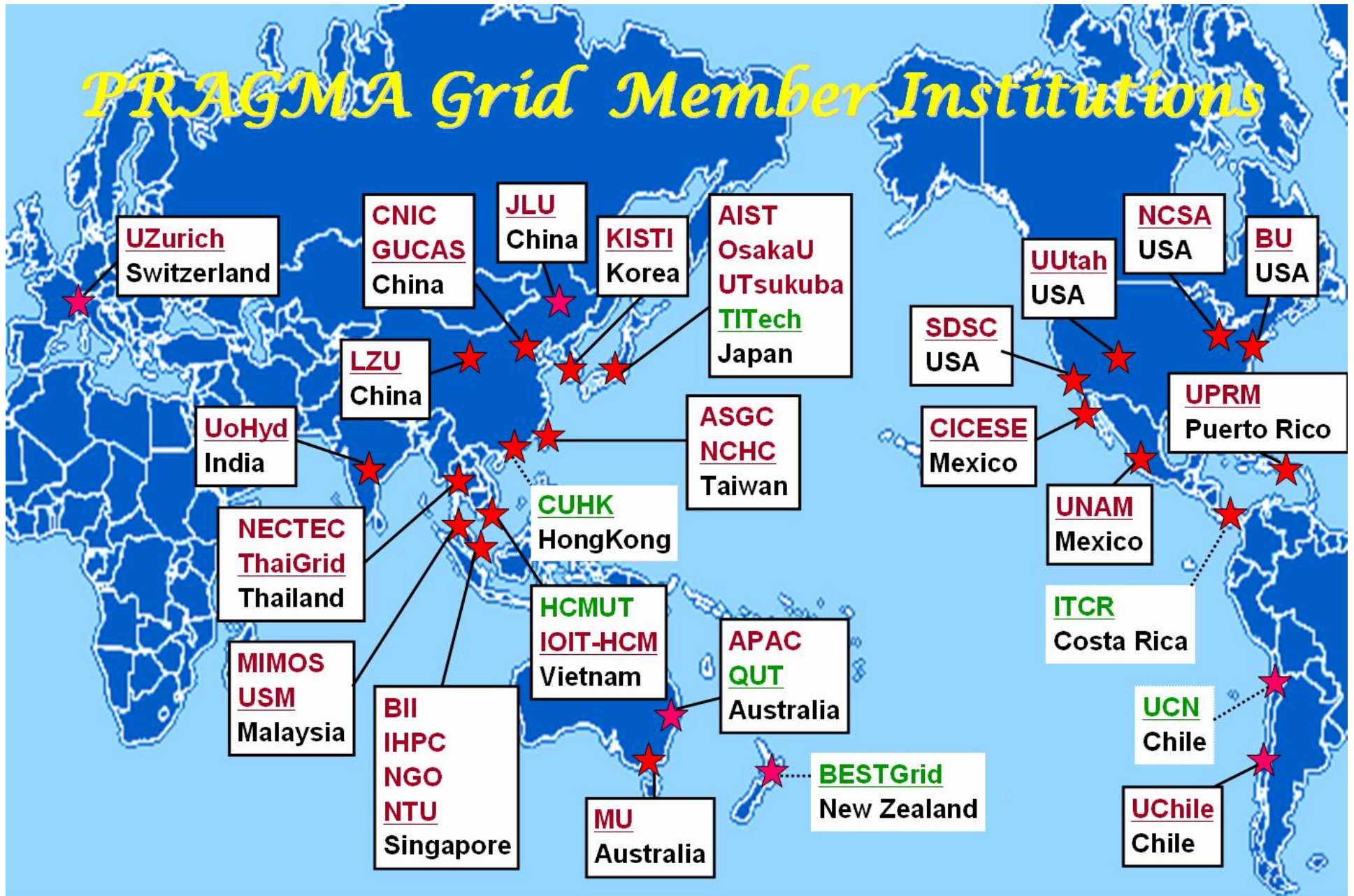
www.bangkokpost.com

Grid computing is no longer just about universities getting together to pool and share their computing power, but increasingly it is about grids of sensors, information and of experts that are shared and leveraged across the world. And now Thailand is playing a major part in this new wave, having recently hosted the 12th meeting of the Pacific Rim Application and Middleware Assembly (Pragma).

“Grids are not just communities of computers, but communities of researchers, of people.”

— Peter Arzberger

PRAGMA Grid Member Institutions



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

Last update: 5/30/2007





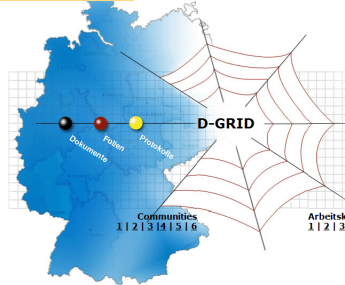
Grids are Communities ...

- Based on (technology-mediated) trust
 - ◆ Common goals
 - ◆ Processes and policies
 - ◆ Reward systems
- That share resources
 - ◆ Computers
 - ◆ Data
 - ◆ Sensor networks
 - ◆ Services
- Supported by software and standards

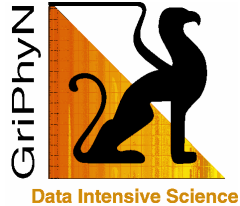
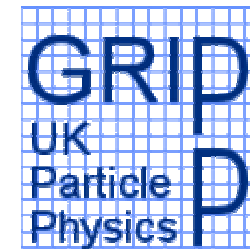
Global Community



GRID.it project

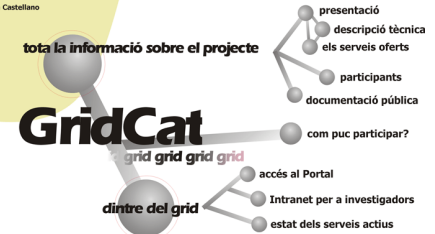


National Grid
NG
SINGAPORE



CERN

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



NAREGI

超高速コンピュータ網形成プロジェクト
National Research Grid Initiative

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

Grid Applications

Grid Middleware

Networking



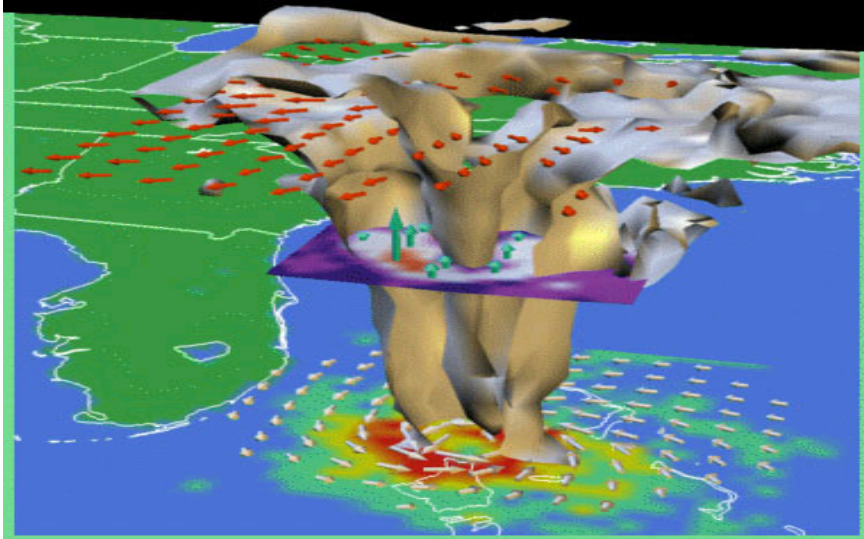


Globus Downloads Last 24 Hours

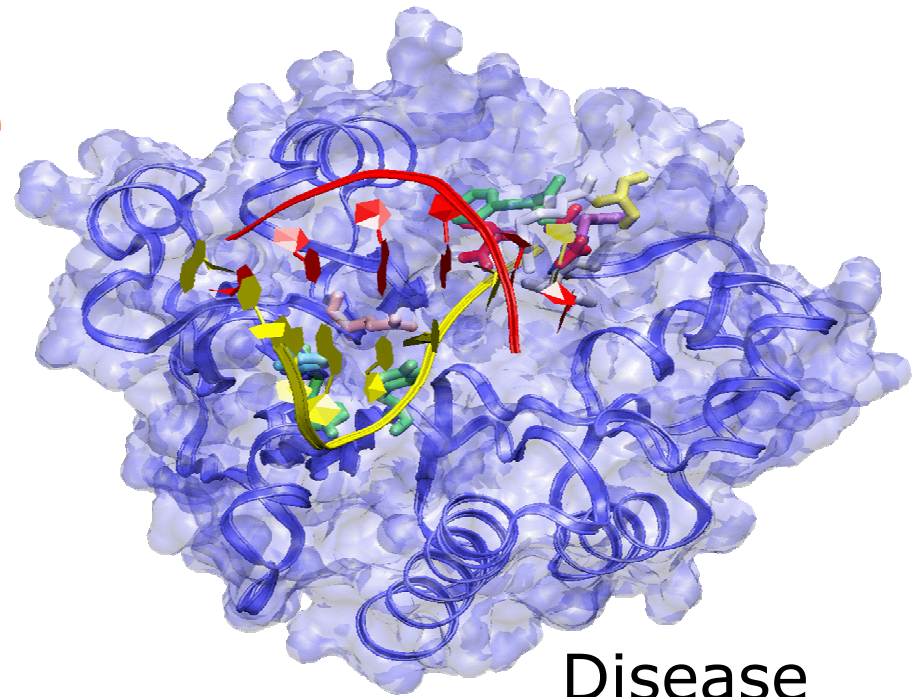




Towards Open eScience



Climate change



Disease



Natural disasters



Sustainable energy



Science 1.0 → Science 2.0

- Megabytes & gigabytes → Terabytes & petabytes
- Tarballs → Services
- Journals → Wikis
- Individuals → Communities
- Community codes → Science gateways
- Supercomputer centers → Campus & national grids ...
- Makefile → Workflow
- Computational science → Science as computation
- Mostly physical sciences → All sciences (& humanities)
- 1000s of computationalists → Millions of researchers
- Government funded → Government funded



Thanks!

- DOE Office of Science



- NSF Office of Cyberinfrastructure



- Colleagues at Argonne, U.Chicago, USC/ISI, and elsewhere

- Many members of the German DGrid community



Summary

- Technology exponentials are transforming the nature of research
 - ◆ Data-, compute-, & communication-intensive approaches are increasingly influential
- Grid is a unifying concept & technology
 - ◆ Federation & on-demand access to resources
- An enabler of “service-oriented science”
 - ◆ Transforms how we conduct research & communicate results
 - ◆ Demands new reward structures, training, & infrastructure

For more information: <http://ianfoster.typepad.com>