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 become (quasi-)ubiquitous
We Can Access Computing on Demand

**Public PUMA Knowledge Base**

Information about proteins analyzed against ~2 million gene sequences

**Back Office Analysis on Grid**

Millions of BLAST, BLOCKS, etc., on OSG and TeraGrid

Natalia Maltsev et al., http://compbio.mcs.anl.gov/puma2
We Can Invent New Tools

Bennett Berthenthal et al., www.sidgrid.org
We Can Work in New Ways: Access Grid and SARS
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”

---

**Grid:**
Unifying Concept & Technology
Technology, Infrastructure, & Standards

interoperability

Applications

User-level Middleware and Tools

System-level Common Infrastructure

Resources

For example
Emergence of New Problem Solving Methodologies

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
“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**

Declan Butler, Nature
Service-Oriented Science: E.g., Virtual Observatories

Figure: S. G. Djorgovski
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!

→ 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.
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

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
New communities:

- CNIC GUCAS China
- JLU China
- KISTI Korea
- AIST OsakaU UTsukuba TITech Japan
- SDSC USA
- ASGC NCHC Taiwan
- CUHK HongKong
- HCMUT IOIT-HCM Vietnam
- APAC QUT Australia
- MU Australia
- BII IHPC NGO NTU Singapore
- MIMOS USM Malaysia
- NCSA USA
- UUtah USA
- UPRM Puerto Rico
- UNAM Mexico
- ITCR Costa Rica
- UCN Chile
- UChile Chile

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
Globus Downloads Last 24 Hours
Towards Open eScience

Climate change

Natural disasters

Disease

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