



postnote

July 2007 Number 286

GRIDS AND e-SCIENCE

Governments and industry worldwide are spending billions of pounds developing Grid computing - an evolution of the Internet that pools computer resources to process, store and access large amounts of data. Grids already have widespread applications in research and potential in industry and government. They are also an important part of e-science (science using advanced information and communications technologies). This note describes e-science and Grids and details key UK and international projects. It examines the potential impact of the forthcoming Comprehensive Spending Review on e-science, as well as wider policy issues such as Grid security.

Background

Government, scientific research and industry are dealing with growing amounts of data, whether analysing protein structure or creating film special effects. Grids are increasingly used to handle these data within science and industry (Box 1). They let users access hundreds or thousands of computers or data sets at different locations, as if they were a single large facility. SETI@home, a computer program that searches for extraterrestrial signals using spare capacity on home and office computers, is based on a similar concept. Grids can serve many purposes (Box 2) but all Grids need three levels of technology:

- **Hardware:** the computers themselves, and a fast network to connect them. Grids can run over the Internet or over private networks.
- **Applications:** software (computer programs) that use Grids. Functions range from processing payroll information to analysing human genome data.
- **Middleware:** software to run a Grid. Middleware examines the resources available on a Grid and decides where to obtain or store the data or run programs, handling security and authorisation on behalf of the user.

Box 1: What are Grids?

While the World Wide Web lets people see information on other computers using the Internet, a Grid also shares resources such as data, storage, experiments and computer processing. The term "Grid" comes from comparison with the electricity grid - a user plugging into the network has access to all its resources, without needing to know where they are or who owns them.

The concept of combining computers is not new - clusters of computers have been used for over a decade. However, these are often restricted to one department and can be busy at some times but unused at others.

Grids aim to increase efficiency by sharing the same computers more widely. They can also include computers owned by different organisations, and the resources available fluctuate as computers join and leave.

To date, Grids linking computers owned by different organisations are mostly used in scientific research. However, the House of Lords Science and Technology Committee in their 2002 report *Chips for Everything*, noted that, "*the world-wide web was first developed to support scientific research. It may not be long before Grid technology advances from serving scientists to providing benefits available to every Internet user.*"

E-science and UK funding

The main UK government support for Grids has been through its £230 million e-science programme. e-Science uses advanced computing, information processing and networking to enable better research. Grids are one approach used in e-science; the programme has funded ~100 projects (Box 3), including sophisticated videoconferencing systems that let researchers share data and 'virtual research environments' to help scientists to manage their research. The funding also includes a network of national and regional e-science centres.

e-Science funding was first announced in the 2000 Spending Review. Funds from the then Office of Science and Technology were allocated through the research councils under the leadership of a national e-Science Director. In April 2006, the programme was restructured. Ring-fenced funding for e-science was removed, raising concern from some members of the scientific community (see page 3).

Grids in science

Many distinct Grids exist, in different nations and disciplines. Two of the largest UK science Grids are:

- GridPP, which has £65m funding from the Science and Technology Facilities Council to run a UK Grid for particle physics. Its 17 UK sites and 5000 computers are part of an even larger international Grid that will analyse data from Europe's next particle accelerator, the Large Hadron Collider.
- National Grid Service (NGS), the UK's Grid for academics from any discipline, with around 500 users from over 30 institutes. It currently pools resources from nine locations across the UK.

Grids in industry

Industry is becoming increasingly involved in Grid computing as a user and supplier. Many major computer companies, such as Sun, IBM and Oracle, provide Grid products, either through supplying hardware or software. Amazon, the online book shop, often has spare capacity on its computers, enabling it to sell computer storage and processing to other companies as a metered service (a 'utility Grid').

Businesses in many sectors use Grid-based products to speed up computing, including oil companies analysing seismic data and pharmaceutical companies modelling new drugs. These 'enterprise Grids' tend to be restricted to a single company. Google employs a Grid of an estimated 450,000 computers to process search requests from millions of Internet users, while HSBC uses Grids to calculate derivatives trades. However, much use of Grids in industry has been to speed up calculations or make cost savings. Some analysts say there needs to be more exploration of what new business models the Grid might create to justify any widespread adoption in industry.

Grids for government

Grids and e-science have applications within government. For instance, Grid products can be used to share cardiology and radiology images, while a group at the University of Leeds has used Grid computing to draw together information for town planning. The EU Cyclops project is examining Grids for civil protection work, while Grids of data sensors could allow traffic redirection to avoid congestion. There is also military interest: the US Department of Defense is developing a \$21bn Global Information Grid. In addition, many government departments have large numbers of computers that could form enterprise Grids, although they have not yet used this capacity.

Box 2 What do Grids do?

Roles for Grids include:

- Finding, accessing and integrating information from large data sets stored on different computers. This is a **data Grid**. An example is AstroGrid, a 'virtual observatory' that lets astronomers use data from many different telescopes and their archives.
- Distributing complex calculations across many computers in order to process them faster or at larger scales, thereby addressing questions that could not previously be tackled. These are known as **compute Grids**. An example is e-Minerals, led by Cambridge University, which runs complex molecular simulations. Industry also uses Grid technologies to co-ordinate computing power across a company, while universities join together computing clusters from different departments to form a **Campus Grid**.
- **Volunteer computing** uses spare capacity on home and office computers. Projects include drug development to combat malaria and searching for extraterrestrial signals. Climateprediction.net (featured on the BBC) is run by Oxford University. Its 400,000 users worldwide have simulated over 20 million years of climate modelling.

Grid management and regulation

Regulation

At present, like many other research tools, Grids are self-regulated. However, should shared Grids become more widely used outside scientific research, some form of regulation may be necessary. Parallels can be drawn with the Internet, which was initially self-regulated but attracted government attention as it became more fundamental to daily life.

Security

Grids can potentially be much more secure than the Internet. Researchers on an academic Grid must have a "digital certificate"; an electronic identification that is trusted between Grids and countries. They are also typically members of "Virtual Organisations", used to restrict or prioritise access to computers at individual sites. This is particularly important in areas such as healthcare that deal with personal data. Grid security standards are used widely in academia, but have not yet converged with industry.

However, if Grid security is compromised there could be serious implications due to the wide user access given to large numbers of computers. To date, there has not been a major Grid security incident, but many predict this is likely at some stage. Grid security is also now a terrorist issue: Sun was initially required by the US Government to restrict its utility computing service to users in the US only, due to concerns over malicious use - for example using the Grid's thousands of computers to decode secure data.

Standards and licensing

The middleware used to run Grids ranges from open source software to commercial products. For example, the UK NGS and GridPP (see above) use slightly different middleware. This causes some duplication of activity, while programs that work on one Grid may not run on

others. The Organisation for Economic Co-operation and Development suggests that interoperability requirements should be built into Grids from the start.

Licensing is also a key issue for Grids, with many commercial software products requiring a separate licence for each computer. A single Grid user could run a given application on any computer from a pool of thousands. To avoid the high costs associated with current licensing models, many scientific Grids use open source software with free licences. However some software companies say these are more costly to implement and manage in the long term.

Box 3 Example UK e-science projects

DAME (Distributed Aircraft Maintenance Environment) is a system to access and analyse maintenance data from aircraft engines in flight. These diagnostics provide huge amounts of data, which are modelled and analysed using Grid-based programs. DAME was a £3.5m collaborative project between four UK universities and three companies. Its follow-on project, BROADEN, is deploying the tools within Rolls-Royce and collaborators.

myGrid is a bioinformatics and e-science project led by the University of Manchester, with £5.5m UK e-science funding. It has developed a virtual 'workbench' that helps scientists make use of distributed resources, such as analysing genome data. For example, studying data related to sleeping sickness in African cattle revealed that cholesterol has a protective role – and that this is also true for people in intensive care.

Gridcast researched how to share broadcast material and resources across the BBC's distributed network using Grid technology. It was a £0.5m collaboration between Belfast e-Science Centre and the BBC.

Data creation, access and curation

Traditionally, scientific data were stored using lab books. Data are now often digital from their creation, needing new approaches. The origin and reliability of data have to be recorded, as well as making the data accessible for reuse and analysis by others (in many disciplines new research relies on this). In addition, as old media for storing electronic data become more fragile, high quality archiving and curation are needed to prevent data being lost. A 2007 Office of Science and Innovation (OSI) working group recommended reviewing government legislation to ensure long-term protection and accessibility of data.

Using Grids

Existing Grids can be difficult to use, often having no simple user interface, and requiring a high level of user knowledge to extract data or run programs effectively. The middleware needed to add computers to the Grid can be hard to install and maintain, requiring much time and expertise. In particular, it is challenging to combine different types of resources, so some of the largest Grids require all their sites to use the same operating system.

There is work ongoing to make Grids easier to use, but some have suggested the need for a 'social Grid', that allows researchers to experiment and form loosely coupled collaborations without the need to install complex middleware. Researchers in Manchester are working on 'myExperiment', a collaborative site based along similar lines to the website 'mySpace'. It will let scientists share data, which they can rate and label with useful information, as well as allowing social networking.

e-Science research and funding

Sustainable funding for e-science

The UK e-science programme was restructured in 2006, moving to a model where e-science is funded by individual research councils and the Joint Information Systems Committee (JISC), which runs university computing. This aims to shift e-science from a one-off programme, to embed it in the day-to-day running of research. An e-Science Envoy was appointed to support UK e-science and act as its ambassador.

JISC has been allocated £10m for e-infrastructure¹ over three years. However, this is time-limited and does not include funding from Scotland or Northern Ireland. JISC believes it is capable of bringing the results of the e-science programme to a broader user community, and is seeking ongoing funds in the Comprehensive Spending Review (CSR). It also stresses the importance of e-infrastructure to support university teaching and learning, as well as research.

Some argue that it is difficult to encourage scientists or organisations to use, or contribute to, a shared Grid whose future is uncertain. They also suggest that e-science projects increasingly require a long-term, reliable, well-managed e-infrastructure. Similar issues arise in EU projects, which are often funded on a short-term basis. EU and national governments are currently considering how to take this forward.

Future co-ordination

Within the research councils there is no longer ring-fenced funding for e-science, so most will support e-science within their normal programmes. Supporters of this approach argue that e-science must stand on its own merits as an integral part of the science programme, without needing its own funding. Some also question the programme's value for money so far. However critics of the new approach, concerned about loss of overall co-ordination, suggest that:

- Other countries, such as the US, have co-ordinated programmes based on the UK model (see Box 4 for details of some international programmes).
- Without a formal method for encouraging scientists to work together, there will be little incentive for collaboration or pooled resources.
- Without co-ordinated support and leadership, researchers may use different standards, so that tools and data will not be compatible.
- Researchers will 're-invent the wheel', reproducing work that has been done elsewhere.

In February 2007, an OSI working group published *Developing the UK's e-infrastructure for science and innovation*, with a range of recommendations. It suggested that greater investment is required in e-infrastructure, as well as, "strong coordination between government, funders, research and development agencies, service providers, and universities and research centres." It is not yet clear how these recommendations are being taken forward in a co-ordinated way, although they have influenced bids made by departments to the CSR.

The role of industry

As Grid technology matures, commercial systems may be able to meet research needs. For example, universities may rent time on commercial clusters of computers, rather than buying their own cluster. However, apart from cost issues, the US Office of Cyberinfrastructure¹ says that science and technology have distinctive needs, unlikely to be met by industry without government help. In contrast, Tony Hey, Microsoft's Vice-President for Technical Computing (formerly the UK research councils' e-Science Director), argues that scientists are not best placed to develop computer systems to handle and analyse vast quantities of data. He suggests that researchers should look to industry and the expertise of computer scientists to provide basic infrastructure.

Box 4 International Grid projects

The **EU's Sixth Framework Programme**, which ended in 2006, spent €275m on Grid research and infrastructures, funding projects including Grids for science, small business, security, e-health and e-learning. 'Enabling Grids for E-science', one of the largest projects, brings together computers from more than 40 nations to form a Grid which runs up to 50 thousand programs per day. Under the **Seventh Framework Programme**, €600m has been allocated for e-infrastructure.¹

The **US Office of Cyberinfrastructure**¹ created in June 2005, is part of the National Science Foundation. It awards grants for developing cyberinfrastructure that will be of use to the science and engineering communities; its budget request for 2007 was \$182m.

Public funding in **France** is governed through the national research body, Agence Nationale de la Recherche. There are three main Grid research programmes, as well as programmes for technology transfer and infrastructure. Similar projects exist in Germany (DGrid) and Italy (Grid-IT) among others.

Technology transfer

Technology developed by the e-science programme has been used in successful industrial applications, including aviation (Box 3) and a Newcastle University spin-off company to identify new drugs against the MRSA "superbug". Originally part of the e-science programme, industry/academic R&D projects in Grids are now funded by the Government's Technology Programme, which is providing £12m to support nine collaborative Grid projects. The Technology Programme has also established a Knowledge Transfer Network with funding of £1m, to promote Grid use to industry.

In addition, individual research councils run knowledge transfer activities. However these are small-scale; the e-Science Envoy suggests that more sustained efforts are required to produce sufficient engagement between industry and academia. JISC notes that Grids themselves can also be used to aid collaboration, as a platform for joint research. Further, universities could use Grids to sell services, allowing companies to use specialised equipment remotely or to rent time on university computers.

Encouraging collaboration

e-Science is interdisciplinary by its nature, with biologists or chemists needing to work alongside computer specialists. However, some point out that interdisciplinary research is notoriously difficult: peer review panels are rarely balanced between the disciplines, there is little funding for trans-disciplinary work and communication between fields may be difficult because of the different jargon used.

Grid computing also requires collaboration between universities and departments. Funding bodies, universities and researchers may be reluctant to buy computers that will be shared with others, or to share data – particularly if it cannot be specified who will have access. University funding structures may also work against such collaboration. Similar concerns apply in industry, where departments may not want to share their data across the company, or may be wary of relying on other peoples' computers to run vital programmes.

Summary

- Grids can be used to deal with increasing amounts of data produced by science, business and government.
- There are a number of large Grids running in science and industry, but for them to develop further, issues such as security, usability and software licensing may need to be addressed.
- UK e-science funding has been restructured. Some question the impact of this on its long-term sustainability; others say it will help integrate e-science into wider research.

Endnotes

¹ Cyberinfrastructure (in the US) and e-infrastructure (in Europe) are collective terms used to describe the computers, software and networks needed to conduct e-science.

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of public policy issues that have a basis in science and technology.

POST is grateful to Dr Sarah Pearce for researching this briefing, and to all contributors and reviewers. For further information on this subject, please contact Dr Chandy Nath, at POST.

Parliamentary Copyright 2007

The Parliamentary Office of Science and Technology, 7 Millbank, London, SW1P 3JA; Tel: 020 7219 2840; email: post@parliament.uk

www.parliament.uk/parliamentary_offices/post/pubs.cfm