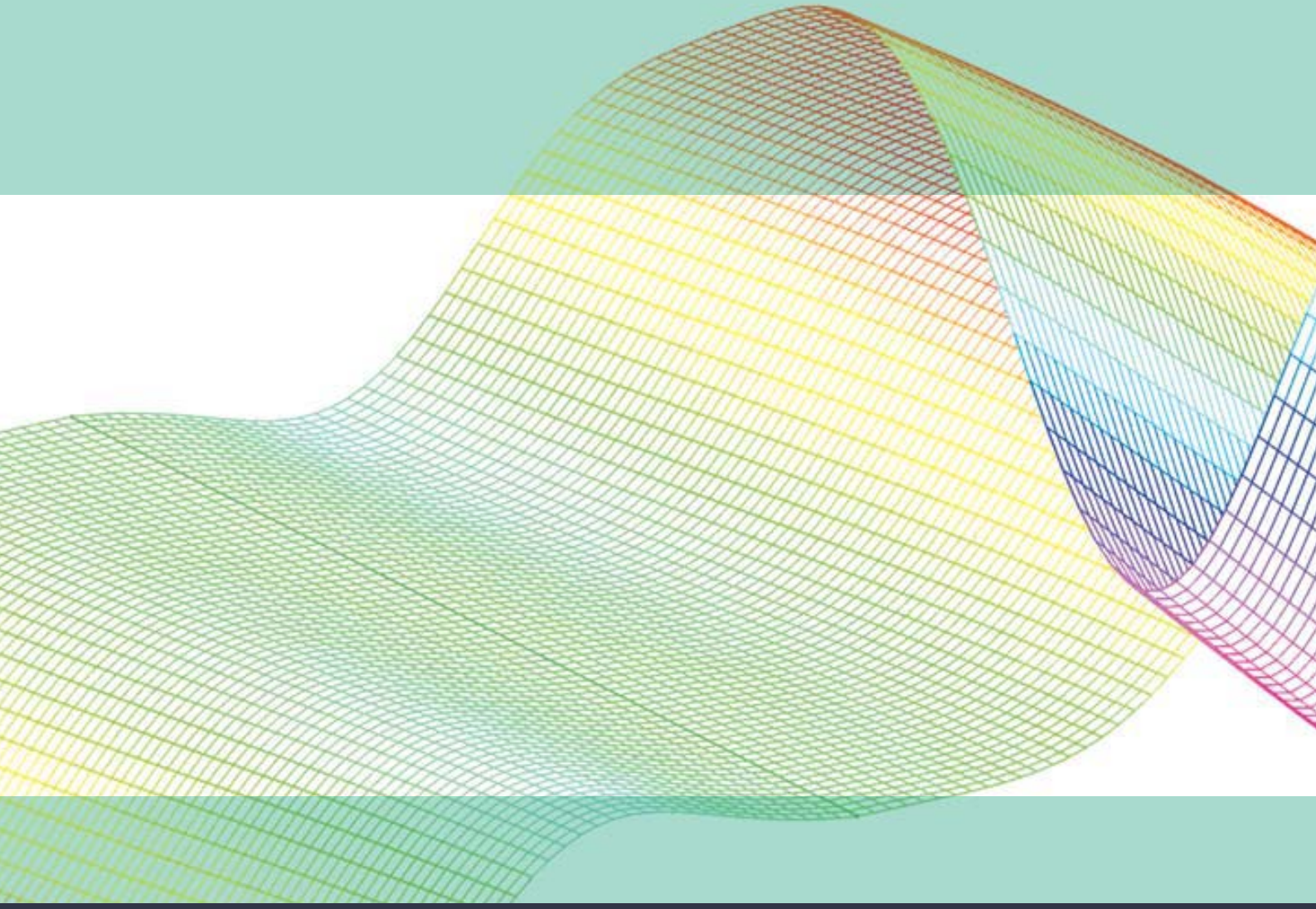
EPSRC

Engineering and Physical Sciences Research Council

# e-Science 2005: Science on the Grid



# Introduction

By enabling scientists to access distributed databases, computing resources and instrumentation, Grid computing technologies are creating opportunities for faster, better or different research. Challenging problems are opening up to scientific scrutiny that otherwise would have remained out of reach, or even been impossible to study.

The UK e-Science Programme, a £250 million five-year coordinated initiative involving all the Research Councils and the Department of Trade and Industry, began in 2001 with the aim of making this vision a reality. It set out to develop generic e-Science technologies and test their use in many different applications. The EPSRC funded six major pilot projects in key application areas to ensure that these new technologies meet the needs of real users. Two further pilot projects started in 2004.

This year the first six pilots – CombeChem, DAME, Discovery Net, GEODISE, myGrid and RealityGrid – are finishing. All have developed and demonstrated e-Science tools and used them to achieve some real scientific results. All six have further funding from industry, the DTI or the EPSRC, to enable them to promote the take up of these techniques elsewhere in academia and industry. The two newer projects – GOLD and Integrative Biology – have also made strides during the past year.

# Science on the Grid

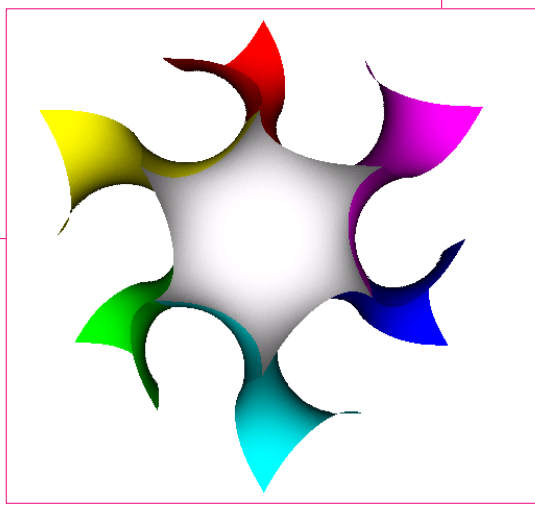
**Discovery Net** and **myGrid** have developed workflows to enable researchers to cope with and make best use of the data deluge now engulfing them. myGrid has rapidly become the bioinformatician's tool for extracting information and knowledge from the wealth of data now stored in databases all over the world, often in incompatible formats. It is also finding use in many other fields, including research in psychiatry, chemistry and engineering. Discovery Net is enabling researchers to cope with the time-critical data generated by high throughput devices. Researchers in the pharmaceutical industry, geohazard modelling and systems biology are among those who are now using it.

**CombeChem** has shown that Grid computing can transform the way in which chemistry is done, from the writing of a laboratory notebook to the publication of data and results. The National Crystallographic Service has adopted its methods and other EPSRC-funded chemistry services are developing plans to do so.

**RealityGrid** has developed Grid technologies to steer and view near real-time simulations of complex condensed matter systems, for example oil invading water-saturated rock or the development of a liquid crystalline gyroid phase. RealityGrid simulations are revealing new insights into several scientific problems and so guiding the design of future experiments. The technologies are being taken up by several other leading-edge e-Science projects.

**GEODISE** and **DAME** have developed tools for the engineer. GEODISE enables design engineers to share knowledge by working in virtual organisations with access to widely distributed software, computing power and databases. The technology has been demonstrated in several engineering applications, for example aircraft wing design and is being taken forward in real applications. DAME has demonstrated the use of Grid technologies to make sense of the vast amount of data returned by sensors on aircraft engines during flight. DAME technology is now being further developed for use by real aircraft maintenance crews.

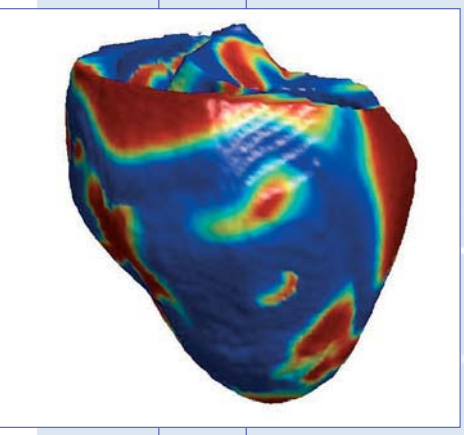
During their first year, **GOLD** and **Integrative Biology** have taken some important steps. GOLD has developed the basic architecture needed to set up a virtual organisation in the chemical process industry and Integrative Biology has expanded its team to include a large international contingent and has demonstrated a prototype heart modelling infrastructure.



# Integrative Biology

## An integrated approach to health

Cancer or heart disease kills six out of every ten people in the UK. e-Science might one day change that statistic for the better through the Integrative Biology project. The project's aim is to produce detailed, accurate computer simulations of cancer tumours and of the heart, based on research at the molecular and cellular level. Such simulations could eventually lead to new treatments based on a clearer understanding of how organs function and the conditions that result when they go wrong.



In the heart modelling field, Integrative Biology is building on other work. This includes models of cardiac electrophysiology developed by Denis Noble's group in Oxford; the extensive work underpinning computational modelling of the whole heart by Peter Hunter's group in Auckland; and Grid software already developed by several of the project's partners, particularly the Council for the Central Laboratory for the Research Councils (CCLRC). In cancer modelling, the project is building on the experience of renowned research groups in Oxford, Nottingham and Birmingham to determine ways of modelling colorectal cancer.

Complex, multi-scale computer models of whole organs require enormous computing power, and better accuracy and greater detail than used in modelling so far. To achieve this, the Integrative Biology project has put together an international, multi-institutional, multi-disciplinary team driven by the e-Science vision that Grid technology will allow computational biologists to carry out collaborative *in silico* experiments within a distributed virtual laboratory, or 'collaboratory'.

**A global infrastructure**

Since the project started in February 2004, the scientific consortium has expanded to include biomedical and computing experts who are uniquely qualified to develop the necessary global computational infrastructure. The original partners have been joined by eight more groups, each with complementary expertise. Within the heart modelling field, these include the University of Graz, Austria; Utrecht University, the Netherlands; the University of Tulane, New Orleans, US; and the University of California, San Diego, US. Researchers in Japan are also showing an interest. The cancer modelling group has expanded to include more partners within the UK and US, including groups at the University of Oxford and at the University of Arizona.

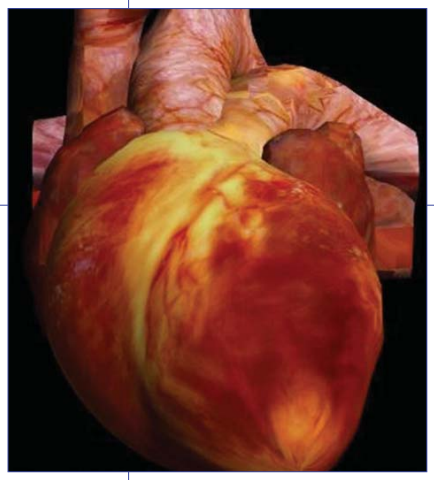
Integrative Biology is a second-generation e-Science project building on the products of some first-round projects and integrating these and the latest developments in Grid technologies into a customised Grid framework for running large scale, whole organ simulations. The team is working in the following areas:

- **Job Management.** Collaboration with the CCLRC on portal interfaces to services, and with the Open Middleware Infrastructure Institute (OMII) to identify the needs for generic job management services.
- **Computational steering.** Collaboration with the RealityGrid (*see page 12*) and gViz projects.
- **Data management.** Collaboration with CCLRC, Daresbury which has implemented Storage Resource Broker (SRB) capability on the National Grid Service (NGS).
- **Visualisation.** Collaboration with the CCLRC and the gViz and RealityGrid projects.
- **Workflow and collaborative working.** Collaboration with the myGrid (*see page 10*) project.
- **Data finder and security.** Collaboration with the eDiaMoND breast cancer screening project.
- **Model and Simulation management within the Matlab environment.** Collaboration with the GEODISE project (*see page 8*).

These services will be embedded into a 'Virtual Research Environment' to support the entire research process from experimental and simulated data generation, acquisition, analysis and curation, through access to IT, HPC and experimental resources, to project management, administration, and learning and teaching support tools.



**www.integrativebiology.ox.ac.uk**



The project will have to face several e-Science challenges, such as finding ways to provide transparent, co-scheduled access to appropriate combinations of the distributed resources needed to run sophisticated whole organ simulations. In addition, it must exploit these resources efficiently using computational steering, workflow, visualisation and other techniques. Another important issue will be how to develop solutions that the members of such a diverse community find relatively easy to use. The project is exploring whether myGrid workflow technology could help with usability via a parallel project called myIB (*see page 18*).

**The first prototype**

A two-pronged approach is enabling scientists to access compute and data management services for research now, while development continues on an early prototype that exploits simple job and data management capability on the NGS. Simulations were run and visualised at a demonstration of the first prototype of the IB portal at the EPSRC e-Science conference in Edinburgh in April 2005. Refinements will be made following feedback from new users.

With such tools at their disposal, researchers will be able to model some of the most complicated biological systems. For example, the plan is to 'grow' virtual tumours through the crucial stages of early development and so gain a better understanding of the many complex biochemical and physiological processes that take place in tumour growth. This could eventually lead to development of a drug to stop tumour growth in its tracks. The project will also make possible global collaboration among biomedical researchers so they can cooperatively control, analyse and visualise simulation results. And in doing this, it will need to address important issues of data security without inhibiting the science.

**Original project partners**

University of Oxford (Professor David Gavaghan and Professor Denis Noble) Lead partner, heart, cancer and molecular modelling, HPC, security, project management
web.comlab.ox.ac.uk/oucl/research/

Council for the Central Laboratory of the Research Councils (Dr Lakshmi Sastry) Grid services, visualisation, data management, HPC, project architect
www.e-science.clrc.ac.uk/web/

University of Nottingham (Dr Helen Byrne and Professor John King) Cancer tumour modelling
www.maths.nottingham.ac.uk/TMsec/Research/Research.html

University College London (Professor Peter Coveney) Modelling, computational steering, HPC
www.chem.ucl.ac.uk/ccs/

University of Leeds (Professor Ken Brodlie) Data visualisation, computational steering
www.comp.leeds.ac.uk/vis/

University of Birmingham (Dr Eamonn Gaffney and Professor Marta Kwiatkowska) Modelling chemotherapeutics
www.mat.bham.ac.uk/research/applied/research.htm

University of Sheffield (Professor Clare Lewis) Experimental studies of cancer tumours
www.shef.ac.uk/dgm/op/path/research.htm

University of Auckland (Professor Peter Hunter)
Expertise and codes for whole organ heart modelling
www.bioeng.auckland.ac.nz/home/home.php

IBM (Professor David Watson) Data management,
Grid data services (OGSA-DAI)

Website
www.integrativebiology.ac.uk

Principal Investigator
Professor David Gavaghan
(gavaghan@comlab.ox.ac.uk)

Lead institution
University of Oxford

Project dates
February 2004 – January 2008

Contact
Dr David Boyd (d.r.s.boyd@rl.ac.uk)