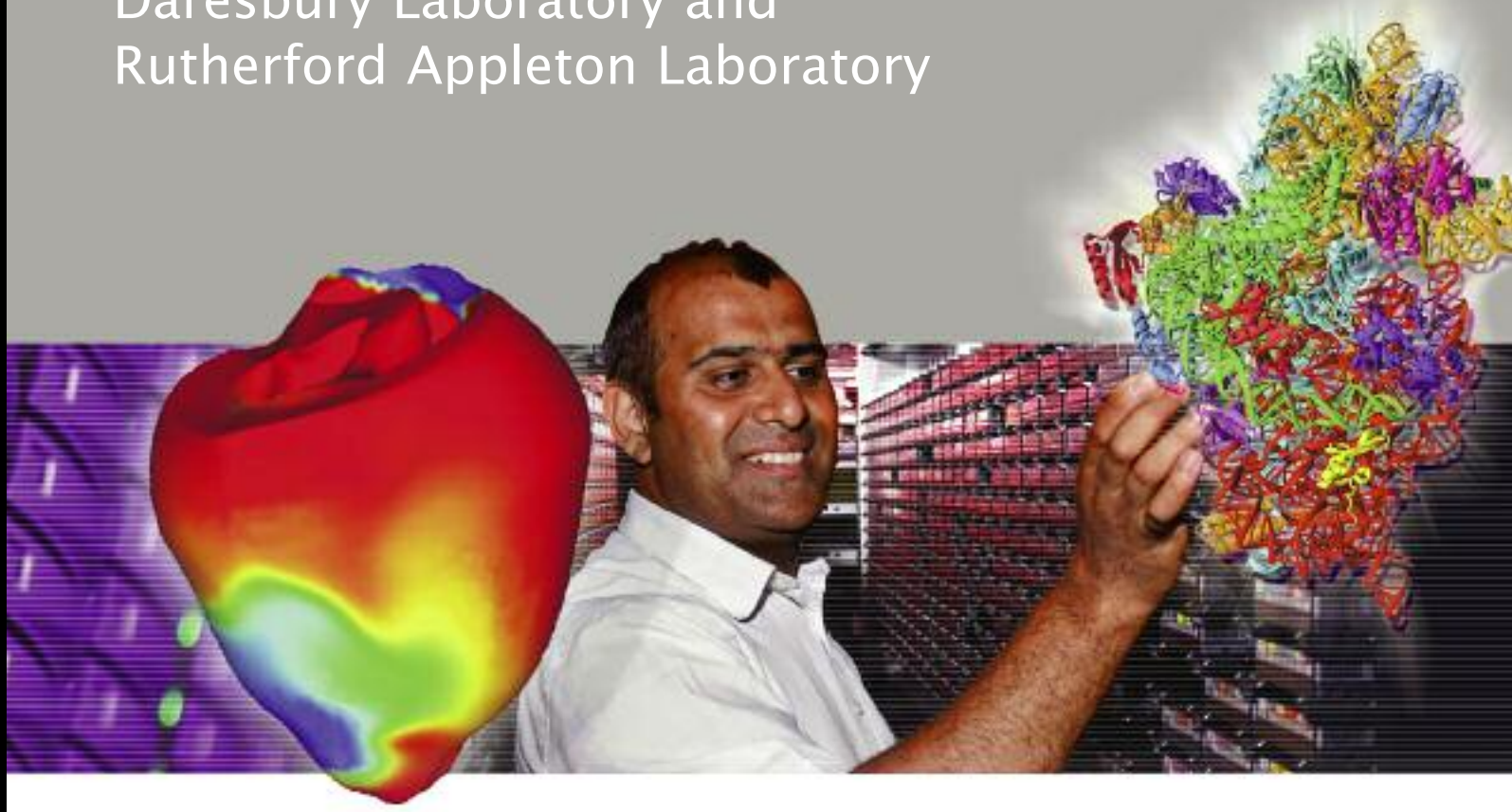
e-Science Centre
Annual Review 2006-2007

Daresbury Laboratory and Rutherford Appleton Laboratory



Science & Technology Facilities Council
e-Science

Further Information

Laura Johnston	DL e-Science Secretary	dlescsec@dl.ac.uk
Sue Smith	Administrative Assistant	s.smith@rl.ac.uk
Denise Small	Personal Assistant	d.small@rl.ac.uk
Andrew Kaye	Business Manager, e-Science	a.kaye@rl.ac.uk
Derek Vokins	Finance	d.vokins@rl.ac.uk
Juan Bicarregui	Head of Scientific Applications Support	J.C.Bicarregui@rl.ac.uk
John Gordon	Deputy Director, e-Science	j.c.gordon@rl.ac.uk
Neil Geddes	Director, e-Science	n.j.geddes@rl.ac.uk

Group Leaders

Rob Allan	Frameworks	r.j.allan@dl.ac.uk
Gordon Brown	Database Services	g.d.brown@rl.ac.uk
David Corney	Petabyte Storage	d.corney@rl.ac.uk
Kerstin Kleese-van-Dam	Data Management	k.kleese@dl.ac.uk
Simon Lambert	Information Services	s.c.lambert@rl.ac.uk
Brian Matthews	Information Management	b.m.matthews@rl.ac.uk
Peter Oliver	HPC Services	p.m.oliver@rl.ac.uk
Andrew Richards	Grid Operations	a.j.richards@rl.ac.uk
Andrew Sansum	Tier 1 Computing Resource	a.sansum@rl.ac.uk
Lakshmi Sastry	Applications and Visualisation	sastry@rl.ac.uk
Penny Windebank	JISCmail	p.a.windebank@rl.ac.uk

Web Sites

Science and Technology Facilities Council	http://www.stfc.ac.uk
UK e-Science programme	http://www.rcuk.ac.uk/escience
STFC e-Science Centre	http://www.e-science.stfc.ac.uk

Addresses

e-Science Centre
Science and Technology Facilities Council
Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
OX11 0QX

e-Science Centre
Science and Technology Facilities Council
Daresbury Laboratory
Daresbury Science and Innovation Campus
Daresbury, Warrington
WA4 4AD

Foreword

Welcome to the 2007 annual report, the first for the e-Science Centre in its new home: the Science and Technology Facilities Council. The creation of the Council, bringing together the programmes of PPARC and CCLRC is a very logical development from the perspective of our Centre. The CCLRC facility and PPARC communities cover the major components of our work. Bringing these communities together strengthens our ability to exploit common developments across the research programmes in support of UK e-Science. Work to embed novel ICT solutions into everyday support for researchers is a theme which has struck a chord worldwide. Often going under the name "e-Infrastructure" in Europe or "Cyberinfrastructure" in the USA, these activities are recognised as key to optimising total return on investment, to supporting multi-disciplinary research, and to improving research and researcher effectiveness. In Europe, the importance of coordinated investment in research infrastructure has been captured in reports from the European Strategic Forum for Research Infrastructure *(ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf)* ands the e-Infrastructure Reflection Group *(http://e-irg.eu/roadmap/eIRG-roadmap.pdf)* which have helped to formulate the 7th Framework Programme (FP7). In the USA, the NSF has created a new Division of Cyberinfrastructure to deliver on its vision for the 21st century *(http://www.nsf.gov/od/oci/ci_v5.pdf)* and here in the UK the OSI e-Infrastructure Steering Group has published its own report on developing the UK's e-Infrastructure for Science and Innovation *(http://www.nesc.ac.uk/documents/OSI/report.pdf).*

A key challenge in the development of an e-infrastructure is to ensure that it delivers for science, prepares us for the interdisciplinary challenges of the coming decades and, in short, is fit for purpose. The need for close connection of technological innovation with science led projects recurs throughout these reports and has always been at the heart of the STFC e-Science Centre mission as illustrated in this report.

Within the department a major event in the last year has been the start-up of the Diamond Light Source. Ensuring that the data from the first beamlines can be recorded, used, and re-used, has clearly been a high priority. The lessons learned from the early DLS deployments are now being exploited for the new beamlines and in work to deploy a data infrastructure for the Central Laser Facility at RAL. Another highlight was the Integrative Biology project showcase held at RAL in March 2007. This event demonstrated the tools and infrastructure developed by the project to support computational biology. The event was attending by a wide ranging audience from Oxford and collaborating Universities. Such was the demand that the event had to be repeated twice through the day to accommodate everyone who wanted to attend. The day showed strong interest in the possibilities opened up by the computational and data management advances on display.

Neil Geddes Director e-Science

The Council for the Central Laboratories of the Research Councils (CCLRC) merged with the former Particle Physics and Astronomy Research Council (PPARC) and the Engineering and Physical Sciences Research Council's (EPSRC) Nuclear Physics Programme on 01 April 2007 to form the Science and Technology Facilities Council.

Formed by Royal Charter, the Science and Technology Facilities Counci is one of Europe's largest multidisciplinary research organisations supporting scientists and engineers world-wide. The Council operates world-class, large scale research facilities and provides strategic advice to the UK government on their development. It also manages international research projects in support of a broad cross-section of the UK research community. The Council also directs, coordinates and funds research, education and training.

The period reported here is the year before the change in organisation from CCLRC to STFC. However, since the e-Science centre moved from one body to the other without significant change, reference is made to both bodies in this report.

The e-Science centre includes staff based at both the Daresbury Laboratory in Cheshire, and the Rutherford Appleton Laboratory in Oxfordshire. CCLRC also operated the Chilbolton Observatory in Hampshire, which was joined in STFC by sites in Edinburgh, the Canary Islands, Hawaii and the head office in Swindon.

The mission of the e-Science centre is to spearhead the exploitation of e-Science technologies throughout the Council's programmes, the research communities they support and the national science and engineering base.

e-Science is the third pillar of research, along with theory and experimentation, in which collaborative simulation and analysis are performed. e-Science will be vital to the successful exploitation of the next generation of powerful scientific facilities. These facilities will collectively generate many terabytes of data every day. Their users will require efficient access to geographically distributed leading edge data storage, computational and network resources in order to manage and analyse these data in a timely and cost effective way. e-Science will provide the infrastructure which delivers this.

The e-Science centre undertakes a programme of work including the hosting of physical computing and storage resources; provision of national services to access resources; collaborative development with UK and international researchers of technologies that can be used in future services; and the integration of e-Science technologies into the Council's facilities.

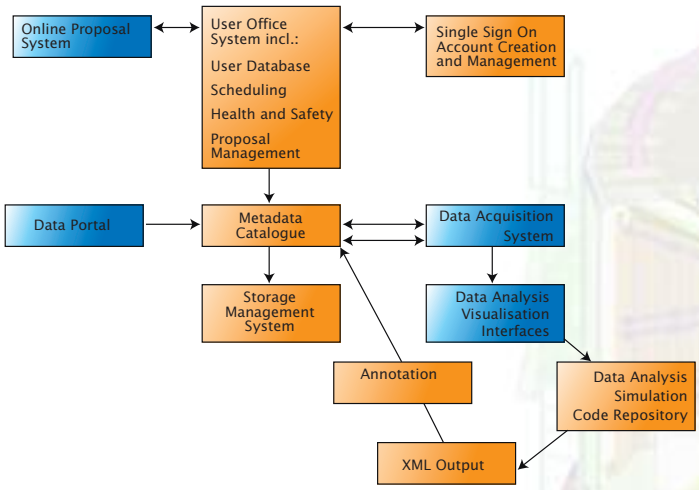
The Council's facilities include light sources (Diamond Light Source & SRS) and a neutron & muon source (ISIS) for determining the structure of materials, lasers (CLF), telescopes, atmosphere radars, and several minor facilities.

Contents

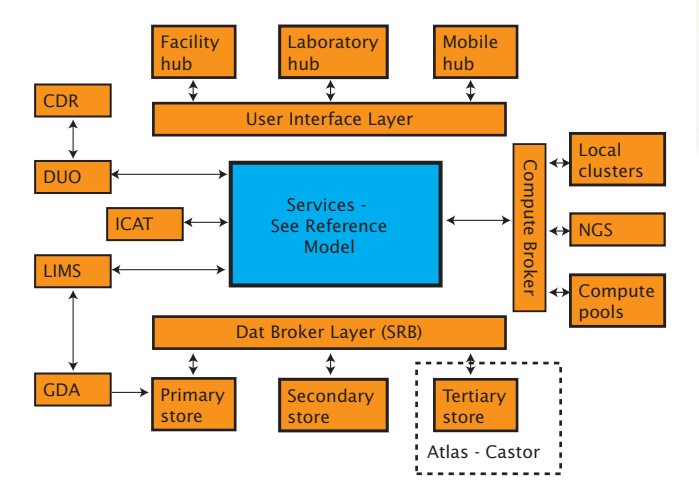
e-infrastructure for **scientific facilities**

The unique role of the STFC e-Science Centre is to 'grid enable' the United Kingdom's large scientific facilities. This is done by developing, deploying and running e-services for experimental, computing and data facilities, in order to further the science carried out by facility users.

Since 2001 the e-Science Centre has developed tools, technologies and services for an integrated data and analysis e-infrastructure in collaboration with the main UK research facilities: ISIS, CLF and SRD. More recently the Diamond Light Source, SNS at the Oak Ridge National Laboratory in the USA, and facilities scientists in Australia have also contributed to the joint development. This e-infrastructure supports the full data and information lifecycle from the electronic proposal system, through the automatic capture and annotation of raw data and meta-data at the experiment, the analysis of the data, to the final publication of data and results.



e-Infrastructure for research facilities supports the scientists through their whole interaction with the facility.



Architecture of the Integrated e-Infrastructure for Research Facilities - Acronyms: CDR - Corporate User ID Repository, DUO - Business System at Facilities, ICAT - Integrated Metadata Catalogue, LIMS - Laboratory Information System, GDA - Data Acquisition System.

Application and Preparation - The users apply for time at the facility through their online proposal system, they are supported in this by access to past proposals, past experiments and their results and publications on past proposals. Once successful they will be issued with a permanent user identifier on their first access to the facility, which they will be able to use for any interaction with or at the facility. The proposal information will be passed on into the Information Catalogue (ICAT). The user office will add information about health and safety assessments and experiment scheduling. The principal investigator can add additional collaborators to grant them access to the information. It will be possible to upload limited information and data required for the experiment.

Experiment - Information relevant to the proposal will be available at the instrument at the time when the project is planned to take place. The data acquisition system has access to this information and will automatically add the relevant project links to the data generated. Further information on the experimental set up and environment are then added to the data files. The raw experimental data itself will be moved in copy to a secure storage system. Once the data is catalogued it can be accessed by collaborators off site. The process is fully automated and hidden from the user.

Analysis - Basic local and advanced grid based systems will be available for the data analysis. The system automatically records and annotates the process, and links the analysis results to the proposal information. Additional support is offered for the more advanced analysis in particular of large data volumes by providing access to large grid based cluster for which applications have been parallelised, reducing execution times from weeks to minutes. Similarly the interactive visualisation based analysis of larger data volumes is supported through libraries which allow for the analysis and rendering of the data to be carried out on a remote cluster displaying the results on the users' local desktop.

Result Publication - When the experimental data is analysed and a publication submitted the users are encouraged to upload their final data and their publication details to finalise the experimental record.

Contact: Kerstin Kleese-Van-Dam E-mail: *k.kleese-van-dam@dl.ac.uk*

ISIS Pulsed Neutron and Muon Facility

ISIS has been the first and longest collaborating facility because ISIS preserved all raw data ever produced by users on the facility in its own data format. In collaboration with ISIS scientists the Information CATalogue 2 (ICAT2) was developed describing all ISIS raw data, which can be searched through the ISIS Online system and the DataPortal. The system went live to ISIS Scientists in March 2007; it includes 92,000 separate investigations and references around 1.8 million files.

In the meantime further development work was carried out to allow the usage of ICAT by other facilities (ICAT3), for easier application access to ICAT3 (ICAT API) and for more advanced and flexible search functions (DataPortal5). These will be deployed in September 2007. At present new studies and their associated raw data are added at the end of each experimental period, which will change to real time availability in autumn 2007.

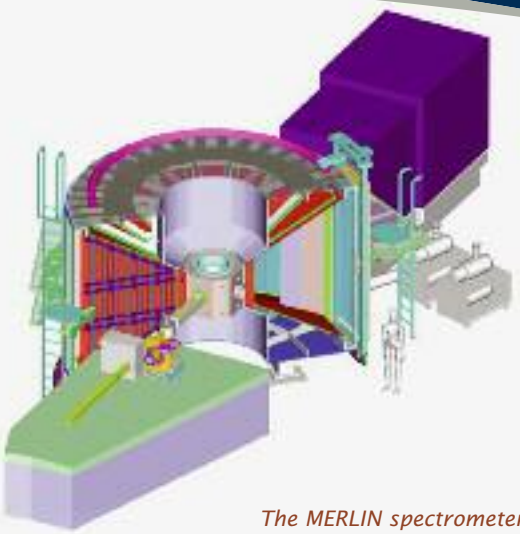
To aid the search and classification of the ISIS data an ontology based on the key words collected in ICAT was created, reducing 10000 keywords from a sample selection of data to 43 classes. More work is planned for the future since the complete catalogue contains at present over 800,000 distinct keywords. Following on from this work, an ontology describing the ISIS Online Proposal System (IOPS) was developed. IOPS is a system that enables users of the ISIS facility to electronically create, submit and collaborate on applications for beam-time. The ISIS ontologies aid the mapping of concrete manifestations of familiar terms in one domain as well as related concepts in different domains.

The NeXus archive format was further developed as part of the collaboration to map and ingest ISIS data into the NeXus format which is an international standard for Neutron and X-Ray facility data.

CCLRC was part of the Neutron Science Software Initiative (NeSSI) which is a collaboration of ISIS, SNS - Oakridge National Laboratory, Argonne National Laboratory and J-Parc. In this collaboration ISIS, SNS and e-Science have collaborated on the further development of the NeXus data format and the usage of ICAT. SNS have adopted ICAT2 and ICAT3.

Simulation and Data Analysis for MAPS

In an experiment on one instrument, such as the MERLIN spectrometer, pulses of neutrons are fired at the sample and the position and time of flight of scattered neutrons are measured in the surrounding detector banks. This data must be processed to extract information about the sample. The results are sampling a four dimensional space, the scattering function S(Q.Ω). The computational problem is to relate this data to models of the processes that cause this scattering. One way to do this is to fit the data to a parameterized model of the scattering processes, within the confines of the actual instrument.



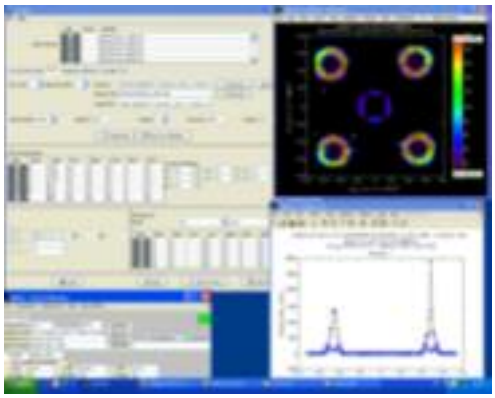
The MERLIN spectrometer

Fast analysis of the data is needed so that scientists can visualise and explore the results in near real time to select the most important areas of interest. Such analysis is used to guide an experiment and maximise the useful science possible on the expensive instruments. This requires access to the large compute and data resources made available via the grid.

An analysis for the MAPS instrument uses the three packages HOMER, for initial data reduction, MSLICE for data inspection, and TOBYFIT, for model fitting. These packages are serial codes and so will not scale to the increasing data demands of ISIS with the result that model fitting is often done on only part of the data set.

The e-Science Centre worked with ISIS scientists to adopt Tobyfit for the grid. The 55000 lines of code were adapted for a 64-bit architecture, wrapped in Globus communication protocol to be used over the grid, and modules were added to stage and marshal input and output huge data files through the Storage Request Broker (SRB).

The re-implemented, grid-enabled Tobyfit is one of the most sophisticated applications on the grid. The new software has been not only demonstrated, but has been released to users of MAPS nationally. The application shows near linear performance improvement for up to 64 processors. The software design is primed to support interactive analysis. The use of SRB enables the facility e-infrastructure to link it to the ISIS data management infrastructure seamlessly and create an Experimental Visualisation Environment in future.



The Matlab/Java interactive interface for Tobyfit provides an integrated environment for using MSLICE to select input data and visualise output

DIAMOND Light Source (DLS)

The Diamond synchrotron light source at the Harwell Science and Innovation Campus welcomed its first users in February 2007. Diamond will ultimately host up to 40 cutting edge research stations, called beamlines, supporting the life, physical and environmental sciences.

Initial user requirements for an e-infrastructure were signed off by Diamond in May 2006 which defined the use cases for users collecting data from the following experimental stations: Protein Crystallography (I01 I02 and I03), Microfocus (I18) and Extreme Conditions (I15). As part of the user requirements the internal business processes behind the User Liaison Office were also examined.

This user requirement document was used as the template for a proposal of an e-Infrastructure for Diamond Light Source. Phase 1 focussed firstly on integrating the e-Science infrastructure with existing back office applications used by Diamond to manage proposal submission, User Liaison Office and GDA; then secondly automatically collecting data and metadata relating to experiments carried out at the beamline; then thirdly storing data securely in the Atlas DataStore; then fourthly, data is made available to the scientist through web based cataloguing system. Coding of this part of the infrastructure is planned to be complete in April 2007.

Central Laser Facilities (CLF)

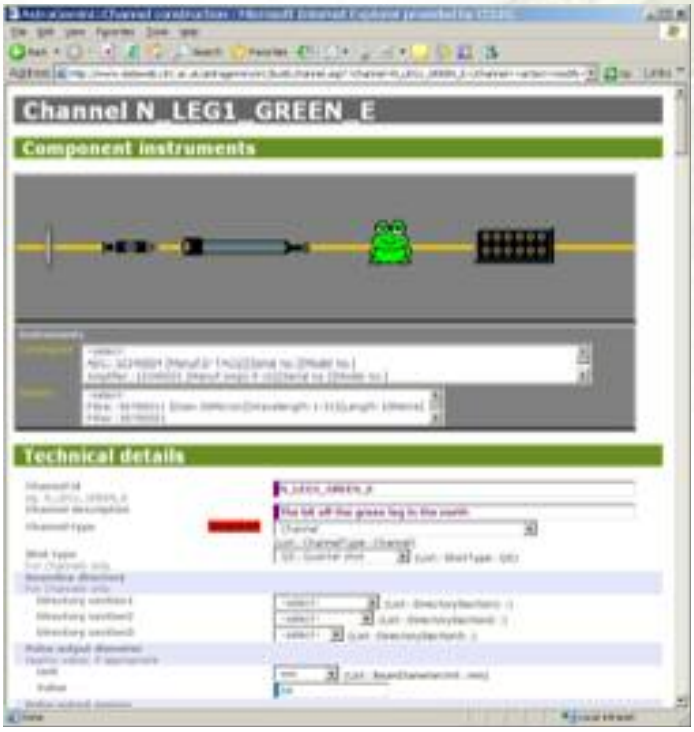
CLF decided to adopt the e-Infrastructure for their facility in 2005, with an initial deployment onto the new Astra/Gemini Laser which will be going into operation in July 2007. It was decided that in phase one of the deployment laser diagnostics and monitoring information would be collected, but not yet the user related experimental data. This information is than used by the users to analyse their collected data. The requirements of the various areas where collected and documented.

Based on the requirements and available resources it was decided that the diagnostic data will be collected and catalogued. Grid enabled applications will be provided to analyse the performance of the laser. All the information will be available to users through the DataPortal. The users and station scientists will be able to access all information using single sign on.

During the reporting year much of the preparation work has been carried out. CLF decided to adopt the NeXuS data format, the initial content of data and metadata for these files was defined as well as the information source. A tool was developed to convert the initial CLF raw data format into the NeXus format. In relation to this a tool 'The CLF Laser Configuration System' (CLCS) was created to capture and describe the important components of the laser, which can be used to create a description of the actual set up for each experiment. The tool allows scientist to easily add new components and their description if required.

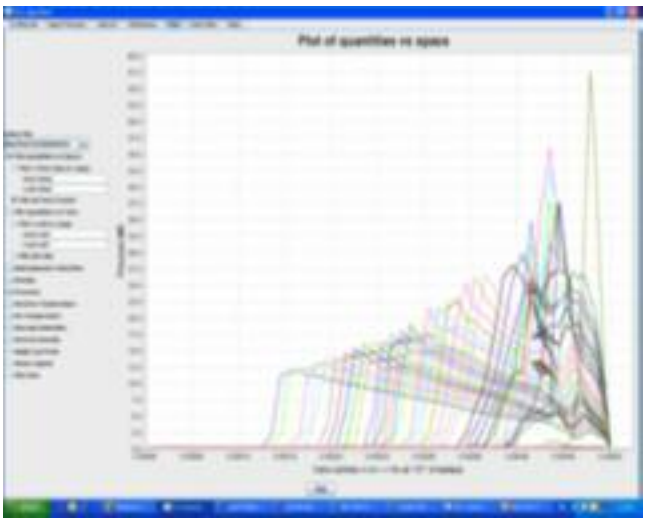
Aerial view of the Diamond Light Source.

The CLCS also captures the link between the different hardware components and the data created as well as a description of the data itself. The CLCS builds the basis for the automatic collection of laser diagnostic data and metadata, by delivering a complete description of data producing components, data streams and formats, which need to be sampled for a particular experimental set up.



A screen image of the CLF Laser Configuration System.

Infrastructure has been set up to collect both the data (into SRB) and the metadata (into the CLF ICAT). For the CLF ICAT the existing schema was pre-populated with the relevant laser specific parameters and CLF specific information.



The e-Science centre worked with scientists from the CLF theory group to adapt the one-dimensional Lagrangian fluid modelling MED103 hydrodynamic code for the Grid. MED103 was originally designed to model ICF experiments but has been extended to model X-ray laser experiments. This code executes on SCARF and other distributed clusters, with a master process overseeing the job farming and collation of output data for visualisation. The interactive interface for job composition, submission, data management and visualisation is shown above.

Synchrotron Radiation Department (SRD)

The Synchrotron department at Daresbury Laboratory will cease operation in December 2008 and it was decided that is would be most useful for the users to catalogue the data that has been kept at the facility, make it available for searching and store both securely, similar to the other facilities. In collaboration with the SR department we developed a simple system that extracts all available metadata from the data files stored in the SR generic data format, moves the data into SRB and stores the metadata. In addition a simple search and access interface has been developed, where users can search through the catalogue. As the metadata available from the files is very limited an additional function was added to the interface that will allow users to add more information. This will make the archive more useful for them and allow them to link their old SR data to new investigations carried out on other facilities.

The SR data access tool.

Scientific Database Services

The e-Science Centre provides quality services of which many have a database as a key component. The Scientific Database Services group delivers production class databases for scientific users.

These databases provide the backbone for systems including Storage Resource Broker systems for facilities such as Diamond Light Source and ISIS, the LHC Computing Grid's Distributed Deployment of Databases project, CASTOR (Cern Advanced STORage) and the National Grid Service. Specialising in Oracle's relational database, version 10g Real Application Cluster systems are implemented as well as single-instance databases for non-clustered solutions.

Many features of the database are explored in the configurations including the provision of Resource Description Framework (RDF) triple stores, XML databases, Oracle Text and Spatial systems as well as exploring the java capabilities. Other database technologies are supported other than Oracle systems and research is carried out with users to explore the technologies that would assist them with their scientific work. The Database Service group works with many fields of science providing beam-line databases for Diamond Light Source, user databases and applications for ISIS, an RDF triple store database for Chemists at Southampton University and replicated Large Hadron Collider conditions data through replication technology from CERN to the Tier-1 centre in the UK.

The group has spent the last year improving its production systems and has revamped its monitoring infrastructure and backup strategy so that all databases are administered and perform to the highest level
As well as providing production databases the continues to work with users to look at advancing database technology to provide systems for their applications and scientific research.

National Grid Service

The UK's national production level grid, the National Grid Service (NGS), entered its third year of operation and its second phase this year.

With over 500 users and nine sites, the NGS is rapidly expanding, with a mission to provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location. The growth in numbers of NGS staff reflects that expansion and the continued commitment to support and develop the NGS. With the integration of another partner site at Queens University Belfast, and more sites in the middle of the joining process, the NGS is set to expand further in the coming years. Resources are funded by EPSRC, JISC and STFC and the NGS is co-ordinated by the STFC e-Science Centre.

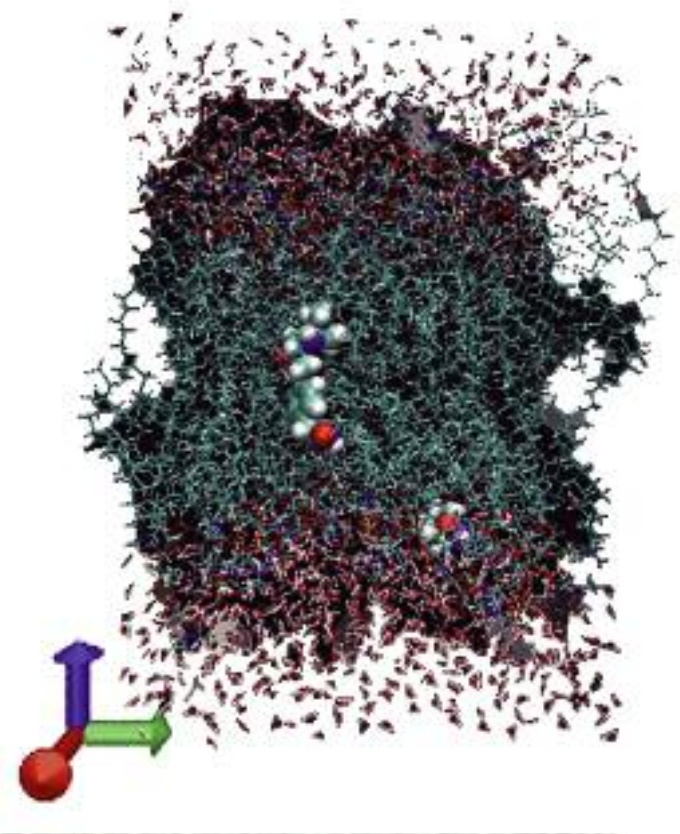
All academics in the UK are eligible to apply for a free account on the NGS. The research achieved using the NGS covers a diverse range of subjects, from the permeation of drugs through a membrane to the welfare of ethnic minority groups in the United Kingdom. Recent applications include cases for modelling the coastal oceans, modelling of HIV mutations, research into cameras for imaging patients during cancer treatments and simulating galaxy formation.

Research and development of the NGS is continuous and shows in tools such as the NGS User Accounting System and the Applications Repository, released in early 2007 as a more user-friendly method of accessing the grid. The Applications Repository has been well received both by the NGS users and the international Grid community at the OGF20 conference. The User Accounting System went into production in October 2006, automating the process of monitoring, accounting and policing user accounts.

The past year has been an exciting time for the NGS, with the start of the second phase in October 2006. All four core sites have taken delivery of new hardware and June saw the launch party for NGS-2 at the Third NGS User Forum. The amount of training given to both new and experienced users has increased with NGS-2, led by the National e-Science Centre in Edinburgh and supported by the NGS Support Centre at the Rutherford Appleton Laboratory.

During the three years of production, the NGS has grown to a national service so involved in the HPC and e-Science communities that all STFC High Performance Computing calls for project proposals now make it a requirement that systems should be prepared to integrate with the NGS. The proposal for the new UK supercomputer also contained requirements that the successful bid be a partner of the NGS and so when it comes online in October 2007, the NGS will gain another partner site – the High End Computing Terascale Resources: HECToR.

With its integration into the UK e-Science community, its growth in users and the second phase well under way, the NGS has come a long way since entering production. And with the continued support from the community, the users and the funding bodies, it is intending to go much further.



Drs Brian Cheney and Jonathan Essex of the University of Southampton are investigating what physical and chemical features make a molecule a good or bad at permeating a membrane. Simulation performed on the NGS of atenolol permeating through a membrane.

Contact: Dr Andrew Richards E-mail: *A.J.Richards@rl.ac.uk*

EGEE Grid Service

The Enabling Grids for E-sciencE project brings together scientists and engineers from more than 90 institutions in 32 countries world-wide to provide a seamless Grid infrastructure for e-Science that is available to scientists 24 hours-a-day.

Conceived from the start as a four-year project, the second two-year phase started on 1 April 2006, and is funded by the European Commission. The EGEE Grid infrastructure is ideal for any scientific research where the time and resources needed for running the applications are considered impractical when using traditional IT infrastructures.

The EGEE Grid consists of over 30,000 CPU available in addition to about 5 Petabytes (5 million Gigabytes) of storage, and maintains 30,000 concurrent jobs on average. Having such resources available changes the way scientific research takes place. The use depends on the users' needs: large storage capacity, the bandwidth that the infrastructure provides, or the sheer computing power available.

The EGEE-II Project is organised into ten "activities", which come under three main areas:

- Networking Activities (NA) are the management and coordination of all the communication aspects of the project
- Service Activities (SA) are the support, operation and management of the Grid as well as the provision of network resources
- Research Activities (JRA) concentrate on Grid research and development
- e-Science participates in the activities relating to operations and support. The services offered are:

A production service - uses the gLite middleware stack and must be run as a stable, well-supported infrastructure. This service is where applications will run their production work.

A pre-production service - where the next version of the middleware is first deployed, applications are tested, and where the site operators verify the middleware meets their needs.

An information database - The GOCDB is a centralized repository of data used by configuration, monitoring, accounting and reporting tools within EGEE and NGS. The GOCDB has undergone major redevelopment in the latter half of 2006 and is due for launch in Q2 2007. During 2006 the GOCDB became increasingly relied on and is classed as a core service within EGEE.

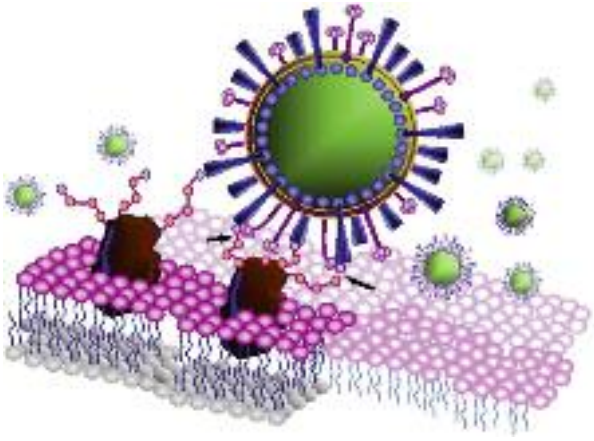
An accounting and monitoring service - data collection and reporting services, using a large centralised database which collects and aggregates CPU usage information from compute sites across the grid. The monitoring service performs tests of the core grid services: Workload Management System, Storage and the Gatekeeper.

A user and operational support service - The UKI ROC helpdesk is operated and managed within the NGS Support Centre.

Expanding from originally two scientific fields, high energy physics and life sciences, EGEE now integrates applications from many other scientific fields, ranging from geology to computational chemistry. Two notable successes this year in the life sciences have addressed potential new drugs to treat malaria and avian flu.

A round of virtual screening to identify the most promising drug compounds for fighting malaria was completed between October 2006 and January 2007. The compounds were screened using in silico docking techniques, whereby researchers computed the probability that potential drugs will 'dock' or interact with a target protein of the malaria parasite. A total of 140 million docking arrangements between drug compounds and target proteins were processed using the equivalent of 420 years of a single computer's computing power.

During April 2006, 300,000 possible drug components against the avian flu virus H5N1 were analysed. The goal was to find potential compounds that can inhibit the activities of an enzyme on the surface of the influenza virus, the so-called neuraminidase, subtype N1. For the docking of 300,000 compounds against 8 different target structures of Influenza A neuraminidases, 2000 computers were used during 4 weeks in April – the equivalent of 100 years on a single computer. Over 200 of these jobs were processed at RAL. Using the Grid to identify the most promising leads for biological tests could speed up the development process for drugs against the influenza virus. Taking advantage of the experience acquired in the previous data challenge on malaria, the Grid-enabled in silico process was implemented in less than a month on three different grid infrastructure, paving the way for a virtual drug screening service on a large scale.



Contact: Dr Andrew Richards E-mail: *A.J.Richards@rl.ac.uk*

Information Management

The information management group is extending the capability of the e-Science Centre to develop the next generation of tools supporting e-Science.

Supporting Virtual Organisations

New forms of scientific collaboration are a key aspect of the emerging Grid infrastructure, opening new possibilities for the way that we do science. The Information Management Group is exploring new ways of supporting collaboration in "Virtual Organisations" which allow scientists in different organisations to work together, sharing their data and working on common tasks.

A key aspect of practical virtual organisations is ensuring that the collaboration is safe and secure from outside interference. In the projects TrustCom and GridTrust the group is developing tools to analyse the key security features of working in virtual organisations, and to make the collaboration safe by producing the appropriate policies and controls to Grid infrastructure itself is changing with ever higher demands on it, and it can be hard to set up Grids which work effectively in a changing environment.

It would be desirable to be able to get all the features of the Grid "out of the box", so that we could rapidly deploy the Grid and give us an efficient infrastructure, The new European project XtreemOS is developing a Grid operating system, based on Linux, with the features of the Grid built in as standard. e-Science is working with this project to support new ways of securely collaborating in Virtual Organisations.

Supporting Scientific Vocabularies

A key aspect of the effective sharing of data and information is working out the common language which people use to describe the key features of their work. For example, in Chemistry the names of chemicals and their properties, and in Atmospheric Science features such as temperature, humidity and wind direction. In order to discover and share information, users need to be able to use a common vocabulary, and the information systems they use need to use this vocabulary to describe resources.

Information Management Group is working with the World-Wide Web Consortium to develop the Simple Knowledge Organisation System (SKOS), a framework for capturing and sharing common vocabulary. This work will define a common standard which scientists and librarians to work more effectively in classifying their data and publications and make it easier for collaborators in the community to find and reuse their work.

Collaborative Research Environments

Earlier this year we completed the first full-scale evaluation and additional development of the Sakai Collaborative Research Environment. This is just one of the frameworks we are using to bring distributed computing, data management and collaboration tools (including Web 2.0) to many practicing scientists in a wide range of disciplines. Sakai started as a Web-based Collaborative Learning Environment at University of Michigan and was based on an earlier project called CHEF. Indeed Sakai is the name of a famous Japanese chef. Sakai spread quickly as an open-source open-standards development programme and is now being used across many of the main US teaching universities and from there around the world. The community Web site is at *http://www.sakaiproject.org.*

In the UK, two projects evaluating Sakai for research were funded under JISC's Virtual Research Environments Programme. Work at Cambridge has focussed on Sakai as a tool for research into distributed teaching methods. See *http://www.jisc.ac.uk/whatwedo/programmes/programme_vre.aspx .*

At Daresbury, in collaboration with Universities of Oxford, Lancaster and Portsmouth (now Reading) we focussed on Sakai for management of distributed projects using the Grid. Over two years, we have extended the Sakai framework significantly so that it now has all the capabilities of a full HPC Grid and Data Portal with additional collaboration tools built in and the possibility of using the additional teaching tools for training purposes. In the coming year it is our goal to deploy Sakai as a framework for the following communities:

1. Social Scientists via the ESRC e-Infrastructure Project;
2. Scientists using Grid-based applications on the NW-GRID;
3. Scientists undertaking experimental data collection and analysis on the Diamond Light Source.

A map showing scientific installations in the UK in a mashup tool created using the Yahoo Map! Service.

Contact: Dr Brian Matthews E-mail: *B.M.Matthews@rl.ac.uk* Contact: Dr Robert Allan E-mail: *r.j.allan@dl.ac.uk*

Securing the Grid for research

The Council's grid security work spans a very wide range of topics: from daily operational security and incident response to high-level trust building; from single sign-on to the global Grid public key infrastructure; from users in academia to business applications.

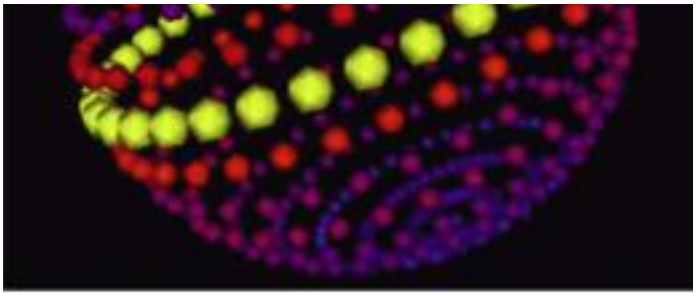
Perhaps ironically, in operational security, no news is good news. None of the e-Science or GridPP clusters have had any serious security incidents in the past year, despite increased use and visibility. A coordinated approach to security ensures that best practices are shared and monitoring is integrated between the GridPP systems, the National Grid Service clusters, the SCARF cluster, and the other HPC clusters.

Security is often seen as an antonym of usability: security "gets in the way" of legitimate users who have to remember multiple passwords and keep typing them in. Thus, paradoxically, strong security measures can weaken the security infrastructure because users start sharing accounts, reusing passwords in different contexts, or writing their passwords on pieces of paper that they leave on their desks. Clearly much can be gained if we can improve usability without compromising the effective security.

Single sign-on can deliver such improved usability. If users have a single account, can securely use a single password, and ideally only have a single log in (per day, week or session), then the infrastructure becomes more usable without compromising security. The JISC Shibboleth deployment aims to deliver a single password – users use their home institution password to authorise access to remote web-based resources. The e-Science centre has worked in the ShibGrid project to deliver single sign-on to the Grid in general, and the NGS in particular. Using MyProxy to manage Grid credentials, ShibGrid delivered the ability to access the Grid via both portals and "thick clients" (i.e., desktop Grid applications). Apart from Grid services, the e-Science centre is also working on Shibboleth-enabling JISCmail – the challenge here is mainly to manage users outside the federation in parallel with those within, including migration, and to adapt to Shibboleth's more anonymous approach to logging in.

The e-Science centre has also been working with the facilities to deliver other aspects of single sign-on for facilities' staff and users. Although this will primarily benefit users of the Council's databases, innovations are made available for the wider community. For example, by using MyProxy again to transparently and securely convert site credentials into Grid credentials, we developed a patch to make MyProxy use keys on a secure key token, thus improving security and the trustworthiness of this part of the infrastructure.

Scientific collaborations are optimally managed via Virtual Organisations (VOs). The e-Science Centre has a research programme on Grid Security tackling open issues in securing



VO management; research funded mainly by the EU FP6 Programme through the projects TrustCoM, GridTrust, XtreemOS and Akogrimo.

The TrustCoM project in which e-Science is the programme co-ordinator ends in May 2007. TrustCoM has integrated the workflow management offered by business process languages with the constraints offered by policy languages for security and quality. This combination of technologies enables the enforcement of contractual obligations on collaborating partners within VOs and commercial supply chains. This technology provides the assurances needed to manage the trust required between organisations to sustain business relationships. The technologies are based on open web services standards to maximise adoption and interoperability. The results of the project will be exploited in products from the commercial project partners Atos Origin, BAE Systems, BT, IBM, Microsoft, and SAP who have applied for 14 patents on technologies developed in the project. Such projects transfer knowledge acquired in scientific research to companies, but also lead to the competitive commercial provision and support of the technologies required to provide the e-infrastructure for science.

The GridTrust project is developing the technology to manage trust and security for Next Generation Grids (NGGs). GridTrust has a vertical approach tackling issues of trust, security and privacy from the requirement level down to the application, middleware and foundation levels; making emphasis on models and tools to assist in reasoning about trust and security properties along the NGG architecture. The XtreemOS project aims at investigating and proposing new services that should be added to current operating systems to build Grid infrastructure in a simple and secure way. The Akogrimo project aims at designing a grid middleware capable of supporting mobility.

On a global scale, the e-Science centre continues to run the UK e-Science Certification Authority (CA). As a member of the International Grid Trust Federation (IGTF), representing the UK, it is trusted internationally, and continues to provide certificates for users for national and international Grid access. As the second-largest Grid CA in the world, with about 1200 distinct "live" user certificates and 2000 hosts, it continues to innovate, usually "behind the scenes," and it is the first CA to implement "robot" certificates. The CA has also contributed much work to the expansion of the global Grid PKI by working with other countries to get their national CAs accredited by the IGTF.

Contact: Dr Jens Jensen E-mail: *J.Jensen@rl.ac.uk* Contact: Dr Alvaro Arenas E-mail: *A.Arenas@rl.ac.uk*

Digital Curation

The ongoing phenomenal growth in the availability of computing resources, together with the increasing sophistication of modern software in everyday life, means that today's practicing scientist increasingly expects to be able to move seamlessly between different scientific activities without concern for how the data and software infrastructure that supports these activities is provided. Meeting this growing expectation in an efficient and effective manner is a major challenge for the provision of scientific infrastructure and in particular for the development of an information infrastructure providing end-to-end support for the scientific process.

However, interoperating with data from a variety of sources poses a number of significant challenges, both technical and sociological. On the technical side, legacy software systems with proprietary data formats must coexist with each other and with generic technology built upon agreed standards. Historical results from defunct systems must be compared to contemporary data. Sociologically, different communities of experts with established practices built upon particular technology must be integrated with each other and with experts from other disciplines. Data from one discipline should be available to be overlaid with data from another, to enable new forms of analysis and the testing of new cross disciplinary hypotheses.

The rapid availability of data from many different types of experimental measurement is absolutely crucial in today's studies of increasingly complex materials and systems. Scientists need to be able to overlay different views - 'Google Earth' at the scale of atoms and molecules. The e-Science department is undertaking some world leading projects in the areas of Data Curation and Integration.

CASPAR. The e-Science Centre leads the CASPAR Project which is a large European project in digital preservation. CASPAR, standing for Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval, involves 17 partners across Europe and will make major contributions to the theory and practise of digital curation by rigorously applying and simultaneously testing, the OAIS Reference Model (ISO 14721) with the aim of creating key components needed to support a global preservation infrastructure. *(http://www.casparpreserves.eu)*

Digital Curation Centre. The DCC provides a national focus for research and development into curation issues and promotes expertise and good practice, both national and international, for the management of all research outputs in digital format. The e-Science Centre is one of four partners, the others being the Universities of Bath, Glasgow and Edinburgh, amongst which we lead the development activities including the development of a Representation Information Registry Repository. *(http://www.dcc.ac.uk)*

Claddier. Data is becoming recognized as an important output of science in its own right. In the JISC funded Claddier project, the e-Science Centre is working with the British

Atmospheric Data Centre to investigate how to publish data and allow it to be referenced, similar to the references to conventional publications. This requires that a unique name can be assigned to a collection data, recording what it is, who collected it, and where it can be accessed. The project is also considering how references to data can be represented in publication repositories, and building a demonstrator in the STFC institutional repository ePubs. *(http://claddier.badc.ac.uk)*

Software Repositories. Many research outputs come in the form of software and its associated artefacts such as documentation, install scripts, test cases, tutorials, etc. Tools and guidelines for the preservation of software research outputs is a JISC funded project investigating issues around the deposition and preservation of software artefacts in repositories. It will develop guidelines for preservations of software research outputs, in particular by monitoring and analysing in detail the activities around a particular thematic software repository in the software engineering domain. The project will focus on four areas: the classification of software artefacts to enable their retrieval and use in new contexts; the preservation of software artefacts as their associated environment evolve; the role of standards in preservation of software; and the process of depositing software as part of the research project workflow to enable validation of analyses undertaken in the course of the research lifecycle.

Contact: Dr David Giaretta E-mail: *D.L.Giaretta@rl.ac.uk* Contact: Dr Juan Bicarregui E-mail: *J.C.Bicarregui@rl.ac.uk*

Library and Information Services

The library and information services joined the e-Science Centre in 2006, a move that brought skills which complement the centre's expertise in the management, organisation, and curation of data and information.

Library functions are an essential part of the research lifecycle, and in an increasingly electronic world this is just as true as ever. Broadly speaking, the service has two aspects: provision of information to users, for example through subscriptions to journals, the book collection, and access to databases, such as British Standards; and facilitating access for the wider community to the Council's research outputs.

There are libraries at Daresbury Laboratory (the Chadwick Library) and Rutherford Appleton Laboratory. Most of the library's users are Council staff on these sites, but relations with external organisations continue to be strong. A new Service Level Agreement has been negotiated with the Diamond Light Source, continuing the successful business partnership which provides full library access for Diamond staff, and a service is provided for the Daresbury Science and Innovation Campus.

The Library has 1,774 registered customers, including 166 Diamond staff, 112 visitors and 42 members from the Daresbury Innovation Centre and Cockcroft Institute.

CCLRC's institutional repository, ePubs, continues under development and its content is constantly expanding - more than 700 new entries were made in the year. ePubs is a publicly accessible system that presents the research output

of the Council to the world at large, by recording publications written by members of staff, collaborators and facility users. It has achieved a high visibility and is harvested by search engines such as Google Scholar, thereby raising the profile of the Council's work worldwide. It stores 23125 bibliographic records of which 880 have the full text of the item attached locally, although 68% link to copies of the work available online somewhere. The Library and Information Services has also been involved in projects advancing the research lifecycle. The JISC-funded project CLADDIER (Citation, Location and Deposition in Discipline & Institutional Repositories) has been investigating how to publish data, encapsulating it and making it possible to cite data sources in the same way as citing scientific publications.

National Academic Mailing List Service (JISCmail)

JISCmail is funded by JISC and operates as a JANET service to provide a national electronic mailing list service for UK academics. It joined e-Science in 2006, bringing with it the Council's largest user population.

This year the JISCmail community is as vast and as diverse as ever, supporting more than 6,000 lists and a total of 729,831 subscriptions to those lists. It is crucial in enabling vital research and collaboration across the UK, Europe and Internationally. Our newsletter provides us with an opportunity to share some of this work with our community.

In 2007 the JISCmail service joined the UK Access Management Federation, as an early service provider of the Shibboleth software. JISCmail is supporting the Federation whose aim it is to reduce the need for academics to use more than one set of credentials to access information resources whether doing research, teaching or learning. JISCmail has for some years enabled remote communication across collaborations and communities and sees the Federation as a natural progression into further improvements to enhance the sharing of information.

Among this year's highlights is a list on forced migration, FMLIST, which is designed to exchange information and promote linkages within the forced migration research community. An example, of this is illustrated by the picture of some of the 4,000-8,000 Sudanese refugee women waiting for aid distribution in Djoran, Birak canton.

JORUM is a free online repository service for teaching and support staff in UK Further and Higher Education Institutions, funded by JISC and managed by the two national data centres, EDINA and MIMAS, helping to build a community for the sharing learning materials. JORUM says their mailing list has been pivotal in promoting the Jorum Outreach and Training Tour.

Scientific Cluster Computing Facilities

Scientific Computing Application Resource for Facilities (SCARF)

SCARF provides a large scale computing resource with rapid access and turn round exclusively for users of CCLRC, its facilities, and Diamond. Access to SCARF is via grid technologies which are interoperable with National projects such as the National Grid Service (NGS) and international projects such as EGEE and Teragrid. The objective is to provide seamless access from the Desktop to Supercomputers.

The SCARF service continues to expand with ~85 users covering 11 disciplines. Over 700 processor cores are now available giving access to approximately 3.5 TFlops of computing power. This combined with the large application portfolio and ~60TB storage makes SCARF an attractive resource for scientists.

National Service for Computational Chemistry Software (NSCCS)

A new machine (SGI Altix 4700) was commissioned for the NSCCS during 2006-7 giving them access to 224 Itanium processors and 896GB of memory in a single large machine. This has enabled the NSCCS to continue to have a significant impact in Computational Chemistry with over 40 publications in 2006-7.

Minerals and Ceramics Consortium (MOTT)

The Minerals and Ceramics Consortium is a collaboration between 11 UK Universities to perform research under the heading of "Simulation on the Edge". The service has ~100 users and consumed over 1.7M CPU hours during 2006-7. During 2006-7 over 70 peer reviewed papers have been published.

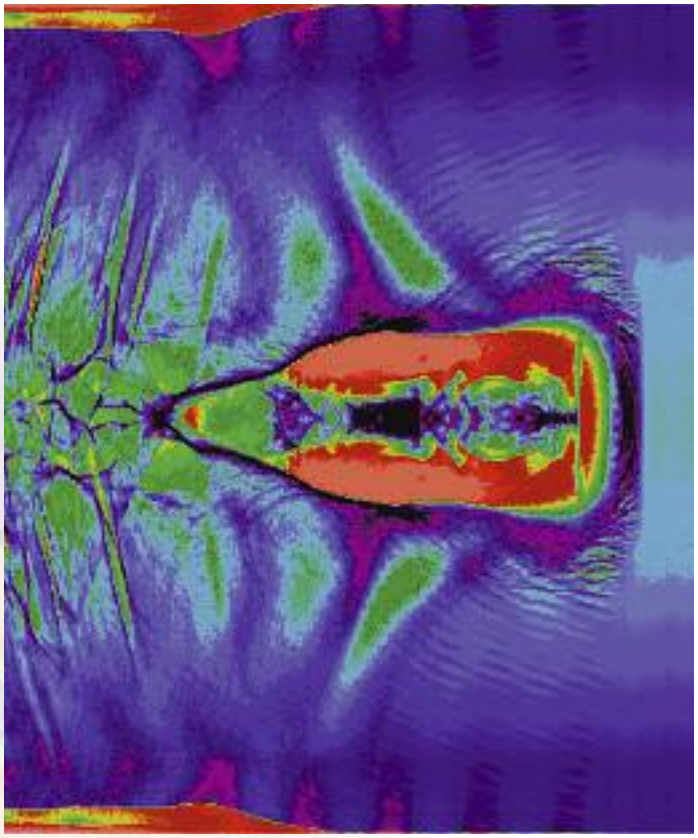
National Grid Service – Core Site Resources (NGS)

The CCLRC-RAL NGS core site node has been upgraded and for the first time is a compute as well as a data resource. In total 256 processor cores, 640GB of memory and 72TB of storage is now available to NGS users.

Complex Infrastructure

In total over 1700 CPU Cores, 4TB memory and 150TB of disk are available requiring ~150kW power and weighing ~12 Tons delivering over 15M CPU hours per year. In order to achieve the same level of computation your desktop PC would have to be running for ~1700 years!

This is a complex infrastructure to manage and during 2006-7 a new management system was deployed to enable the e-Science Centre to expand the computing infrastructure whilst maintaining efficiency.



Scientific Delivery Example - Laser Wakefield Acceleration of Electrons, Dr Raoul Trines, Central Laser Facility

For many years, researchers have been developing plasma-based electron acceleration as a way to obtain very high electron energies (1-10 GeV and beyond) even within modestly sized institutions. In this scheme, a driver (either a laser pulse or a high-energy particle beam) propagates through tenuous plasma and creates a wake wave, which is then used to accelerate an electron bunch. Recently, an important breakthrough was achieved with the discovery of the so-called bubble regime, in which a very intense driver sweeps almost all plasma electrons out of its way to create a single-period wake wave, the so-called "bubble". This structure has proved particularly advantageous for the production of electron bunches having not only a high mean energy but also a low energy spread. The numerical study of this "bubble" regime requires two- or three-dimensional (preferably) large-scale particle-in-cell simulations, requiring massively parallel computing facilities. We are using SCARF to study the details of bubble formation and electron acceleration by bubbles, and to investigate how the final electron energy scales with various laser and plasma parameters.

Contact: Dr Peter Oliver E-mail: *P.M.Oliver@rl.ac.uk*

ATLAS Petabyte Data Store

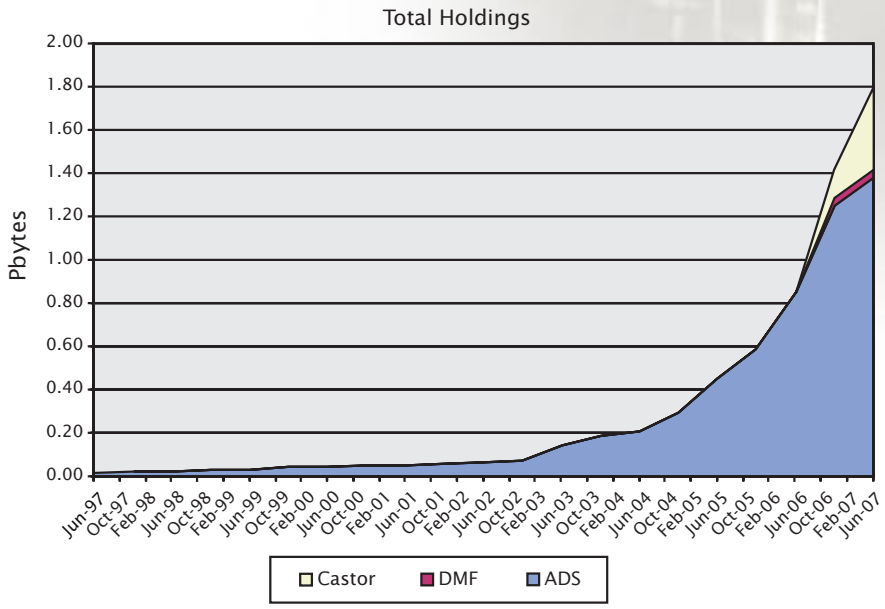
The Atlas Petabyte Data Store now runs a series of services for a growing and diverse community of scientific users, all of whom want access to data storage. The service provides a safe data archive with multiple optional copies of users' data, including fire safes copies and copies stored safely and securely off-site ten miles away.

This year we have continued to operate the ADS which has proved its reliability for many years. However with the total volume of data stored now at 1.8 Petabytes we know that the ADS is not scaleable to the level necessary to cope with expected demand for the next few years. Consequently considerable effort has gone into deploying and managing the CASTOR service this year. CASTOR is a complex data archive system built around concepts of Hierarchical Storage Management (or HSM), and was developed by CERN in collaboration with others, including this group at RAL who developed the GRID interface to CASTOR. CASTOR is scalable to the data volumes we envisage. At present we only provide operational CASTOR services for the experiments of Large Hadron Collider community, as part of the UK's Tier1 centre. Within a year or two we expect to replace the ADS with CASTOR for all our users. Until then we will continue to work with CERN to improve and develop the required functionality.

In addition to the ADS and CASTOR we are also running a commercial HSM system based on SGI's Data Migration Facility software. This is used by groups who require efficient access to (hundred's of) thousands of small files. Users currently include ISIS, The Solar Physics World Data Centre and SOLAR-B, the latter of which is collaboration with Mullard Space Science Laboratory.

In addition to the three "backbone services", we have also (in conjunction with the Data Management group in e-Science), implemented additional services which use Storage Resource Broker (SRB) as a layer on top of the ADS. In November 2006, BBSRC and RAL signed a contract to provide the BBSRC with a ten year archive for all its institutions, and some estimated six thousand scientists across the UK. This service uses SRB backed into the ADS. SRB as a layer above the ADS is a cornerstone in the provision of services being rolled out to STFC facilities this year, including Diamond Light source, who have just such a test service deployed this year.

Currently, the combined "back bone" services have around 350 registered users. Of these LHC have 50 active users, with around 960 Terabytes of data stored. Other users within STFC include Space Science with around 100 Terabytes and ISIS with around 10 Terabytes.



Performance achievements this year include the CMS data challenge, held in October 2006. This involved pushing data from CERN in Geneva to the archive service at RAL, as part of emulating and testing the actual running conditions of the operational LHC. The test ran continuously for almost month, achieving a sustained data rate over that time of 200 Mbytes/sec to tape.

Data growth showing new services: CASTOR and DMF deployed this year.

Contact: David Corney E-mail: *D.R.Corney@rl.ac.uk*

Applications in Particle Physics

The particle physics users of the Tier-1 computing resource operated by e-Science at RAL collaborate within the GridPP and EGEE projects. Within these they are organised into virtual organisations representing different experiments. The Tier-1 computing resource is mainly used by virtual organisations (VO) addressing four experiments on the LHC at CERN (Atlas, CMS, LHCb and Alice) and two experimental facilities in the USA (Minos and BaBar).

Hunting the Higgs boson

The large masses of W and Z particles raise inconsistencies in the Standard Model. In the 1960s British physicist Peter Higgs proposed that the universe is filled by a Higgs field which is distorted when particles pass through it, with the result that they are leant mass. Different theories predict one or more Higgs particle(s) which carry this mass.

Although physicists working at CERN's LEP saw traces of particles that might fit the right pattern in August 2000, the evidence was still inconclusive when the LEP was closed down in the beginning of November, 2000. The Atlas and CMS experiments in the LHC at CERN are designed to detect the Higgs boson among the many particles that will be created in the high energy proton-proton collisions. The two experiments will be able to verify each other's results, once the LHC becomes operational, which is currently planned for May 2008.

There are several sorts of event which might produce Higgs particles, resulting in different decay patterns, and different background noise. During the year, teams from the Atlas and CMS experiments have been using the Tier-1 computing resource to simulate and reconstruct different decay modes of the Higgs boson and the various backgrounds within different models in order to establish the sensitivities of the experiments in preparation for real data.

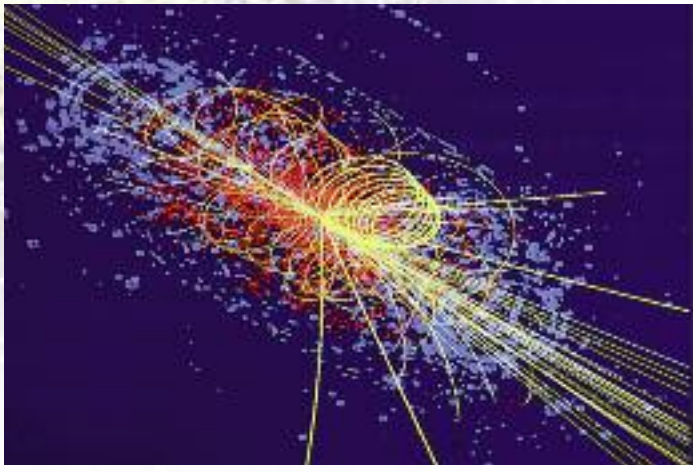
The first microsecond after the Big Bang

During the first microsecond after the Big Bang before the transition to protons, neutrons and related particles, there existed a form of matter which may only exist today in the cores of dense, distant stars – the quark-gluon plasma. The Alice experiment in the LHC will detect the high pressure and high density results of collisions of ions of heavy metals such as lead or gold that are expected to produce the quark-gluon plasma. These experiments will allow testing of the theory of the strong force at higher energies than previously attained. During the year, the Tier-1 computing resource has been used to simulate and reconstruct different collisions to optimise the software and to build up the computing infrastructure in readiness for real data from the LHC.

CP violation – why is there more matter than antimatter?

The simplest model of the Big Bang suggests that equal quantities of matter and antimatter would be produced, which would then cancel each other out, leaving nothing. Clearly this is false and more matter appears to have been produced than antimatter. The discovery of CP violation in 1964 in the decays of neutral K-mesons resulted in the Nobel Prize in Physics in 1980 for its discoverers James Cronin and Val Fitch. In 2001 the BaBar experiment at the Stanford Linear Accelerator Centre (SLAC) observed CP violation for the first time in a heavier particle – the B-meson. The computing resources at RAL contributed to the analysis of data which led to this discovery.

During this year, the BaBar B-factory VO has continued to use RAL computing resources, although now they use the Tier-1 computing resource – but to the US led B-factory VO it is termed a Tier-A centre. Experiments continue at SLAC to generate electron-positron collisions producing B-mesons which decay through various channels to produce different secondary particles. The models of CP violation continue to be probed through comparing simulations with an increasing data sample.



Simulation of a particle collision on the LHC.

The BaBar experiment is a precursor to the LHCb experiment at CERN which will investigate B-meson and anti-B meson decays resulting from proton-proton collisions at higher energies than are available at SLAC. LHCb will study additional types of B meson and will produce 1000 million such particles each year and look for new physics which will help to refine theories of the matter-antimatter imbalance.

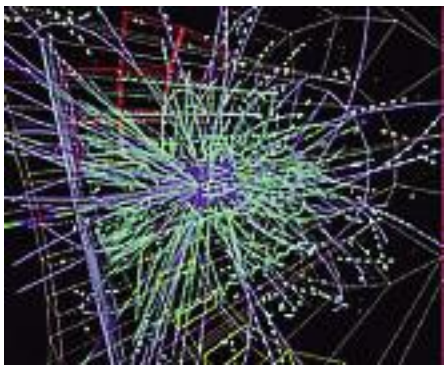
The LHCb VO has been conducting a year long "data challenge" in which large samples of the different B meson decays (there are over 250 modes) have been simulated, reconstructed and analysed. This has provided important information on the performance of the hardware and software of the experiment in advance of data taking in 2008.

Contact: Dr John Gordon E-mail: *J.C.Gordon@rl.ac.uk*

The missing mass in the universe

In 1998, the Super-Kamiokande collaboration announced the discovery of oscillations of neutrinos, as evidence that they had mass. This made the neutrino the first serious particle physics dark matter candidate actually known to exist.
In Feb 2006 the Minos experiment created muon neutrinos in a particle accelerator at the Fermi National Accelerator Laboratory (Fermilab) in Illinois, US and then passed a high intensity beam of these particles through 450 miles of the Earth to a detector in the disused Soudan iron mine in Minnesota. The results on neutrino disappearance are consistent with neutrino oscillations.

The activity this year has been to determine the detailed nature of the oscillations in order to more precisely measure the mass differences of the neutrino flavours. The Tier-1 computing resource has been a major contributor to Monte-Carlo generation critical to data analysis for the Minos VO.



Computer simulation of a central collision as perceived by the ALICE detectors.

LHC Tier 1 Computing Resource

The Tier-1 computing resource is the main e-Science resource for the particle physics community. Following an upgrade of its CPU capacity by 50% in January with the addition of 64 Intel dual core "Woodcrest" servers, it consisted in March 2007 of:

- 1300 CPUs running Linux
- 900TB of available disk space
- Tape storage using a Sun robot system - up to 5 PB

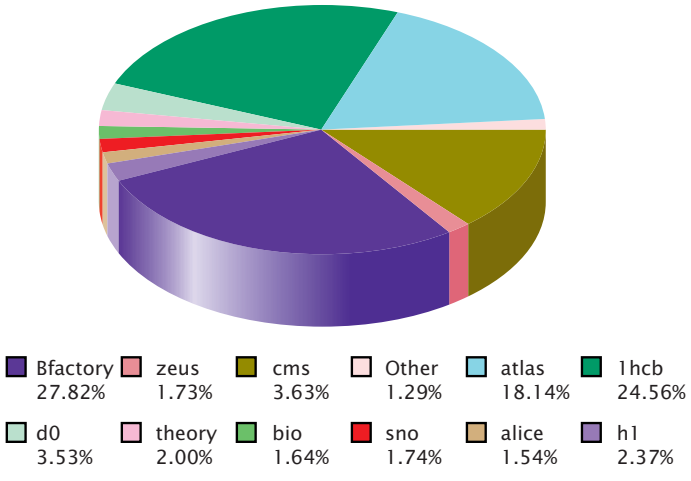
The number of registered users increased to 2726 by the end of 2006, and although the number of active users increased from 150 to 260 over the year, this is still a low percentage of the registered users since jobs for many VOs are only entered by a few individuals.

Much of the technical activity related to the Tier 1 resource has been directed at ensuring that the service can cope with the data produced by the LHC when it becomes operational. The role of the Tier 1 centre is to take the data from CERN (the tier 0), store it, and make it available to the Tier 2 centres in UK universities - London, NorthGrid, ScotGrid and SouthGrid. This provides the requirement by 2008 to receive data and store it to tape at a rate of 40 MB/s whilst also allowing the Tier 2 sites to retrieve other data and process it.

Each of the LHC experiments has run a series of Data Challenges to assess if the requirements will be met. These are large and complex computing exercises using simulated data to test and develop middleware and individual experiment Grid applications.

For example, in the 45 day CMS Oct '06 computing, software and analysis challenge the RAL Tier 1 resource had the target of 10MB/s transfer from CERN to tape at RAL, which was exceeded by a sustained performance of 33MB/s for the last 15 days of the challenge. Further site challenges and middleware validation tests are needed in order to improve overall Grid efficiency.

The Tier 1 resource is also used by other disciplines, including biomedical data challenges for drug discovery, as part the EGEE European grid project described earlier in this report.



CPU usage by virtual organisation in 2006.

Applications in Materials Science

Applications to determine the structure and properties of materials are among the most common uses of the ISIS facility, and also frequent on other facilities. Here are two examples of such applications that have involved e-Science in different roles.

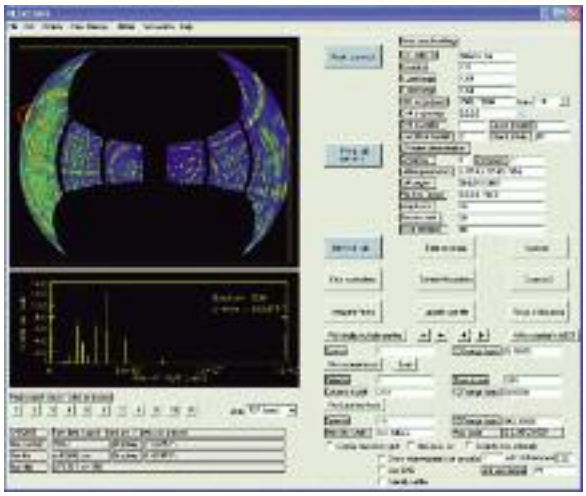
Enabling Single Crystal Diffractometer Analysis to exploit the grid

One of the instruments available at the ISIS facility is the Single Crystal Diffractometer (SXD) which accesses large 3-D volumes of reciprocal space in a single measurement. Applications of SXD have included routine structural studies and diffuse scattering.

The instrument's user-friendly software, SXD2001, allows the experimenter to extract normalized Bragg intensities or corrected diffuse scattering data from SXD measurements. The graphical user interface to the program uses the IDL graphical display package. Each frame of data is searched to identify the positions of possible Bragg peaks, given in terms of an (x, z, t, n) coordinate, where x and z identify the horizontal and vertical positions of the pixel of detector n containing the possible peak at time-of-flight t. Trial UB matrices are then generated until one is found to account for some or all of the possible peak positions based on the known or expected crystal unit cell.

Performance profiling of this code indicated that grid enabling peak search module will provide very quick gain to demonstrate while retaining the rest of the interface and codes intact. The peak search module was rewritten in Fortran, parallelised and wrapped in grid communication protocols. This is then seamlessly interfaced into the IDL interface. Use of just four processors produces nearly ten times faster performance.

In the screen shot below, the top left-hand panel shows a time integrated representation of the scattering collected by the six equatorial detectors looking from the moderator along the neutron beam, and clearly showing the Laue-like nature of the data collection.



Screen capture of the SXD2001 data-reduction graphical user interface.

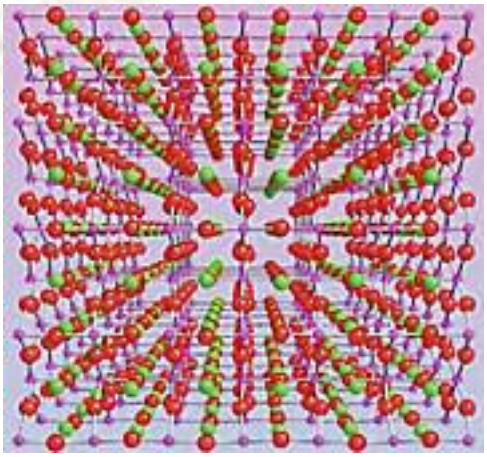
The bottom left-hand panel shows the scattering in one detector pixel as a function of neutron time-of-flight and with the orders of a Bragg reflection separated in time-of-flight. The pixel being displayed corresponds to the circled bright spot at the far left of the Laue plot above (i.e. at the high-angle edge of high-angle detector 1). The data are from a sample of SrF2, measured on SXD at 20 K.

MaterialsGrid: Large scale computer simulation of physical properties of materials

The e-science department is part of a consortium led by Accelrys as part of the DTI technology programme in Inter Enterprise Computing. The project also involves scientists and grid specialists from IBM, University of Cambridge, the Cambridge e-science Centre and University of Frankfurt.

The aim of the MaterialsGrid project (2005 to 2008) is to create a pilot dynamic database of materials properties (such as elastic stiffness, dielectric constants, optical properties, heat capacity, electronic band gap) based on quantum mechanical simulations run within grid computing environments, and thereby creating a new industry service.

The e-science department have the task of devising a relational database schema to hold the predicted materials and their properties and manage the database infrastructure. In addition we have developed an XML to SQL parser for the simulation code CASTEP to directly enter predicted crystals and properties into the database. Furthermore we host and maintain the application servers and relevant software for the project at Daresbury Laboratory.



The perovskite family, with general formula ABO3, is one of the most important class of materials for modern technological applications, not least because its physical properties can be tuned through changing chemical composition.

Contact: Lisa Blanshard E-mail: *L.J.Blanshard@dl.ac.uk*

Applications in Computational Chemistry

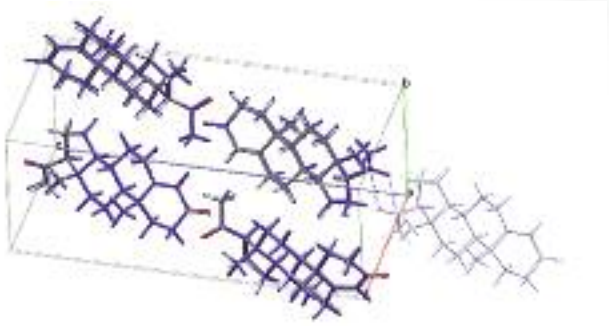
Many of the applications of the CCLRC and now the STFC facilities are used by computational chemists. Two example applications show how s-Science staff work with researchers to improve their productivity and the quality of their science.

CPOSS (Basic Technology Programme Control and Prediction of the Organic Solid State)

This is a major four-year project funded by the Basic Technology Programme of the Research Councils UK, which started in October 2003. Underpinning this project are software applications that were developed by the e-science department for the e-materials project 2002-2005. The scientific aims of the project relate to the development of a computational method of predicting the polymorphism of organic molecules. Previously, their scientific data was distributed across a multitude of sites and the scientists had very limited support for accessing, managing and transferring data. In a grid environment that spans numerous sites and organisations, it is essential to ease and automate many of these data management processes.

The e-science department developed and provided a number of data management tools for files as well as devising relational database schemas to hold metadata relating to their simulation output files and another to hold predicted crystals and their properties. The database schemas are held on the National Grid Service (NGS) Oracle database. The applications provided by the e-science department (except where indicated) are:
CPOSS Database of Computed Crystal Structures and its web user interface Crystal Navigator, Storage Resource Broker (developed by San Diego Supercomputer Center) for virtualized file storage and CPOSS Data Portal for retrieval via web interface CPOSS Metadata Manager, web interface for editing file related metadata

Professor Sally Price of University College London, who leads the Basic Technology project says “NGS hosts our database of Computed Crystal Structures and the portal that stores the output files from many different computational studies. The NGS has enabled us to store many hundreds of data files and allow our users easy, secure access to this data via the Crystal Navigator web interface.”

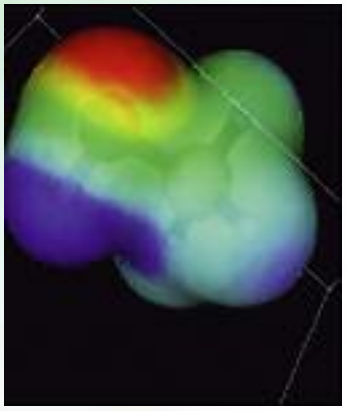


Overlay of the novel crystal structure of racemic progesterone (coloured element), formed by mixing occuring progesterone with its synthesised mirror image, and its prior computational prediction (coloured blue).

Computational Science and Engineering Department

The eScience Centre has worked closely with the Computational Science and Engineering Department (CSED) to help them to utilize grid technologies and assess the suitability of various approaches to facilitate their scientific research. This work centered around both data management/sharing and job submission, the use of which in CSED is becoming more tightly coupled and therefore more effective throughout the interaction between our departments.

The computational chemistry group (CCG), part of Collaborative Computational Projects - 1 Computational Chemistry (CCP1), has implemented several different job submission solutions in the CCP1 GUI. Thus far this includes RMCS/eMinerals framework, Growl Scripts/Globus, and NorduGrid job submission framework, while the AHE is also being looked at. This allows users to submit jobs running any of several different quantum chemistry programs to a grid resource. Users are thus able to select the approach to job submission that is most appropriate to the way that they work. A similar effort is under way with the Computational Materials Science (CMS) group, part of CCP3, concerning DL_Visualize.



A number of projects were enabled by e-Science technology, including the Quantum Directed Genetic Algorithm (QDGA) project with the CCG and Prof. Marcus Durant at Northumbria University; study of high Tc compounds by the Band Theory Group, part of CCP9; computational materials science studies of Aluminium Florides; the global Coastal Ocean Modeling project (formerly Proudman Ocean Labs gCOMS).

The use of AgentX has been an important component of many of these projects, integrated with RMCS for seamless metadata capture, and used in the CCP1 GUI as well as other GUIs to trivially expand the number of codes that the GUI can be used with. Additionally, several CSED developed codes now support XML input/output.

Applications in Social Science

The technologies that e-Science uses and the tools we develop are applicable to many areas of research.

Social scientists are now very active in helping to increase the uptake of e-Science technology by providing requirements and feedback on how people use the tools which have been provided, sometimes in surprising ways. Whilst not straying too far from our original goals, we therefore now have several activities supporting these new communities of Grid users. Several projects are briefly outlined below, but much more information can be found from the relevant Websites.

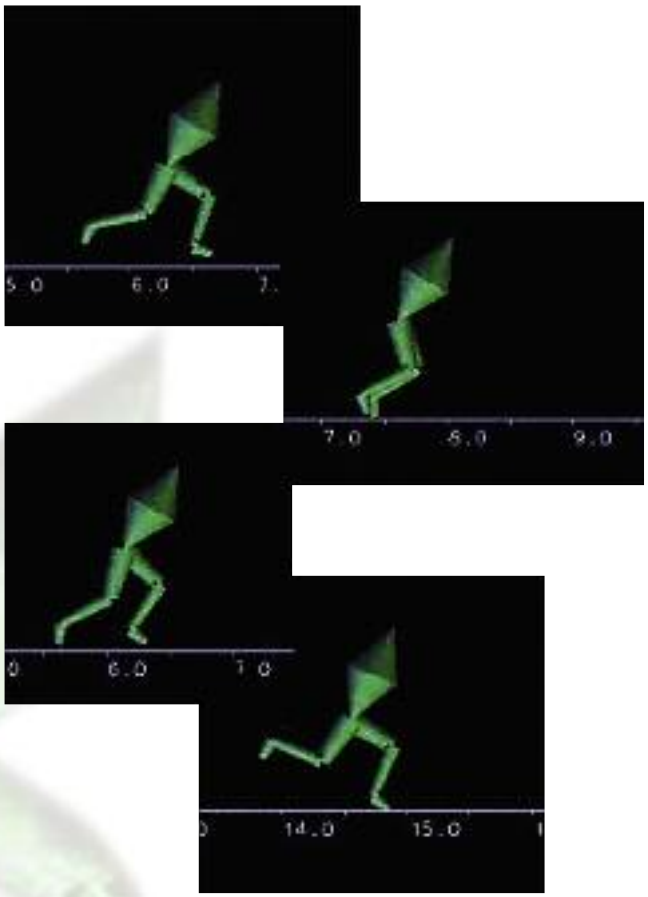
ReDRESS: Resource Discovery for Researchers in e-Social Science

A large set of on-line training and awareness resources funded by JISC and ESRC. This contains many links to Web-based material and video material from workshops and conferences about computational technologies enabling established researchers to carry out self-motivated learning to update their skills. It is not confined to social science and discussions are under way to combine it with similar resources from the arts and humanities, see *http://redress.lancs.ac.uk*

CQeSS: Collaboratory for Quantitative e-Social Science

This was the first node of the National Centre for e-Social Science funded by JISC, see *http://www.ncess.ac.uk* . It is a collaboration between University of Lancaster and Daresbury Laboratory to apply Grid computing and data management technology to quantitative research in statistical analysis and modelling. High-performance applications such as SABRE (Software for Analysis of Binary Recurrent Events) are applied to studies of educational or work attainment and financial modelling. ESRC e-Infrastructure. Funding from the ESRC has been provided over three years to deploy an e-Infrastructure linking data collections to Grid resources, applications and user interfaces. This is lead by the NCeSS hub at University of Manchester, but involves several of the nodes working together in a programme of deployment using software which has been developed in previous projects. Daresbury Laboratory will be deploying a portal framework based on Sakai.

APEMAN - a consortium of academics based around the North West Grid (NW-GRID) is building a collaboratory to develop evidence-based models for research into primate and human evolution. They are firstly simulating locomotion and carrying out gait analysis to understand the energy costs and locomotor capabilities of extinct and extant species using metric evidence from fossil remains and observational fieldwork. They are also beginning to develop agent-based models of the Pleistocene Out-Of-Africa migration event that led the African hominins to spread to other parts of the world. Several on-line movies show preliminary results of this simulation work, performed on the NW-GRID: *http://www.liv.ac.uk/premog/premog-nw-grid.htm.*



Simulated locomotion of homo species created on the North West Grid as part of the APEMAN project.

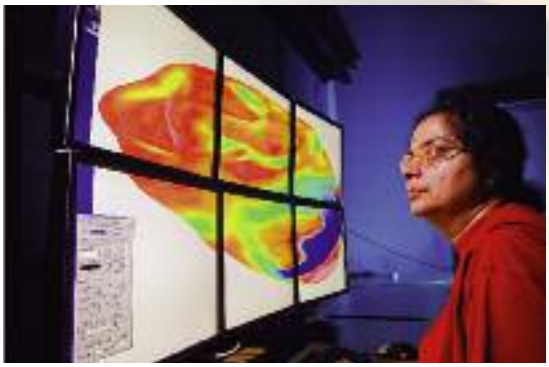
Contact: Dr Robert Allan E-mail: *r.j.allan@dl.ac.uk*

Applications in Biomedical Science

Applications in the biomedical sciences are few at present but they are expected to increase in significance as the Diamond Light Source starts to operate with more users in this area.

Integrative Biology

The Integrative Biology (IB) project, funded by EPSRC, is nearing completion with the building of a customized Grid framework. This is being used to run large multi-scale models, from cellular to whole organ simulations, to manage a growing database of in-vitro experimental data and simulation results, and to support advanced visualization for interactive data analysis with comparison and assimilation of experimental and observed data. The services offered by the computational framework are based on the requirements of two application areas: arrhythmia and cancer. Computational and experimental biologists are using the prototype infrastructure, thereby aiding the STFC computer scientists in improving the framework and its services.



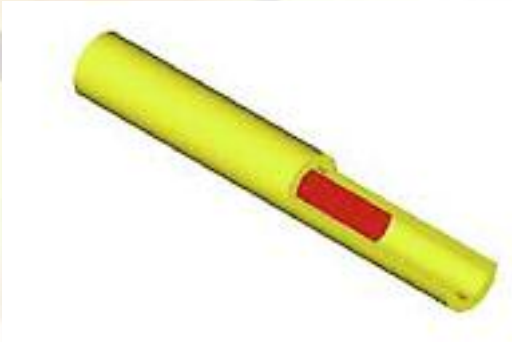
Heart disease and cancer represent the two biggest diseases in the West, and hence have become the focus of intense research within the biomedical community. Computer simulation of whole organs offers the potential to improve our understanding of the causes of these conditions, and eventually to develop new treatment regimes and drugs to reduce their threat to life. The Integrative Biology (IB) project brings together a team uniquely qualified to tackle these problems. It includes researchers from the universities of Oxford, Auckland, the STFC, Sheffield, Birmingham and Nottingham, University College London, and Leeds. One aspect of this work is the computer simulation of whole organs based on molecular- and cellular-level models, and this requires the application of large-scale computational and data management resources. Another major aspect is the provision of advanced visualization services that allow real-time interactive data analysis on the Grid.

The e-Science centre has developed advanced visualization services are built to handle extremely large datasets close to where the data is stored or generated. These include building 3D images from 2D MRI and histology data, 3D isosurfaces, server-side animation generation and viewpoint-dependent animations. Tailor-made generation and rendering services can render up to 0.5GB of data in high resolution and in close to real-time. The results can be sent to the user's desktop using public-domain open-software stacks.

A Medical Application of the Diamond Light Source

Blocked arteries are commonly treated using angioplasty techniques in which special catheters are inserted through femoral artery to clear blockages and the arterial section is strengthened by positioning a stent in the cleared section. Stents are increasingly being used in complex cases where the arterial geometry necessitates new designs. This vital area of design engineering is one of the many experiments being performed using the Diamond JEEP instrument.

The e-Science Centres Applications Group is working with Diamond scientist to take the tomography images to reconstruct 3D geometry which can then be used in simulations to further improve the design based on clinical results.



Artery model and sketch of a stent being expanded in an artery.

BBSRC and MRC support

The e-Science centre supports the BBSRC in two main ways: by providing data storage for all BBSRC institutes on the Atlas Petabyte Store, and by providing consultancy on the management of cluster computers.

The only support provided for the MRC is the very small, unfunded service of the MRC Psycholinguistic Database. This is one of the first publicaly available datasets which has continuously been supported since 1986. It records data from previous experiments that characterise properties of words, and is used by psychologists and linguists to construct wordlists used as cues in experiments. It was ported to the Web in 1994 and has been used to test several Web technologies as they have been designed, including both XML and RDF.

Applications in Environmental Science

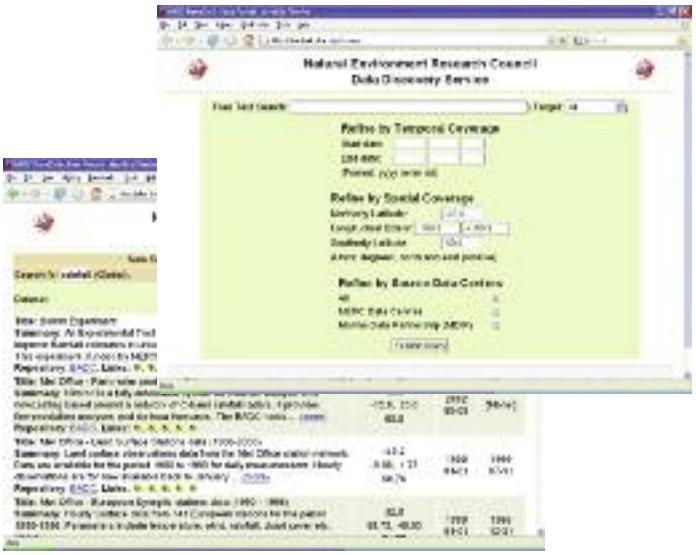
e-Science is developing a range of applications for environmental scientists and contributing to both standards and policies to facilitate their wide adoption.

NERC DataGrid (NDG)

NERC DataGrid (NDG) is developing a framework for integrating access to a wide range of environmental data within the UK research community. During the current phase of NERC e-Science funding, data access is being integrated to the curated archives of the British Atmospheric Data Centre (BADC, hosted at RAL), the British Oceanographic Data Centre (BODC, co-located with the Proudman Laboratory, University of Liverpool), the National Oceanographic Centre Southampton (NOCS), and the Remote Sensing Data Analysis Service (Plymouth Marine Laboratory).

The integration approach has adopted a service-oriented architecture, and is based on all providers committing to common metadata formats (for discovery and evaluation/browse), data models, and service interfaces.

The e-Science Centre has continued to develop NDG's metadata and data models, and contributed to the definition of service interfaces. These are now being rolled out across providers in the final stages of NDG's core funding period.



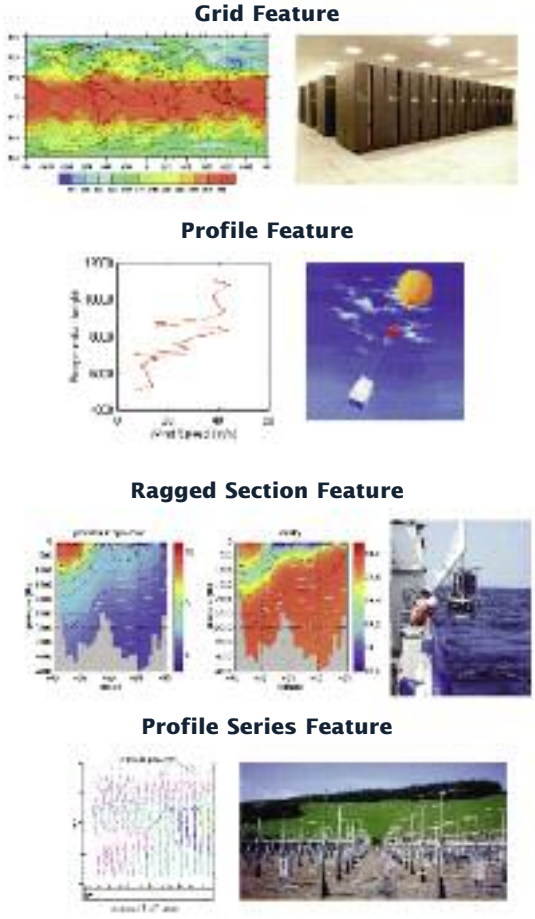
NDG Discovery Portal with search results for 'rainfall' query.

Climate Science Modelling Language (CSML)

The Climate Science Modelling Language (CSML) was first developed by the e-Science Centre as the common data model for NERC DataGrid. It abstracts-away data storage details, and provides instead a more semantically meaningful conceptual 'view' of data. CSML development is based on the 'model-driven approach' of emerging ISO standards for geographic data integration. Under this approach, a conceptual data model is developed using a formal schema language (UML), including classes for key information types ('feature types'), and capturing their properties and relationships. This model is then serialised into an XML encoding using the Geography Markup Language (GML, ISO 19136) and a defined mapping process. In addition, CSML provides a novel mechanism for linking to large legacy persistence formats.

CSML development has continued, with a new version providing greater compliance to emerging standards, a richer set of information classes, the modelling of behaviour as well as structure, and a clearer separation between the core conceptual model and the storage-related elements.

CSML continues to attract the interest of external organisations, including the Norwegian Met Office, University of Kiel, the Australian Antarctic Data Centre, and Karadeniz Technical University (Turkey).



CSML feature types.

INSPIRE

A new European Directive entered into force on May 15, 2007, establishing an INfrastructure for SPatial infoRmation in Europe (INSPIRE). This will provide policymakers and the general public with harmonised access to environmental information across all member states and 34 thematic areas (including landcover and use, soil, geology, hydrography, atmospheric, meteorological, oceanographic data, species distributions etc.)

The e-Science Centre is contributing to the development of statutory Implementing Rules on Data Specifications for INSPIRE.

MOTIIVE

The European FP6 project, MOTIIVE (Marine Overlays on Topography for annex II Valuation and Exploitation) is developing technology in support of INSPIRE implementation. The project is evaluating the cost benefits of deploying standards-based integration approaches for marine data interoperability. A key contribution is the recognition that significant benefit results from re-use of information models – this enhances interoperability and reduces development costs. An essential operational component to enable re-use is a 'feature-type catalogue' (or 'semantics repository') that enables data models to be formally registered and queried. While an emerging standard (ISO 19110) defines a conceptual model for such a data model register, there is no consensus on implementation best practice.

MOTIIVE is developing an ebRIM implementation of the ISO 19110 Feature Catalogue. The e-Science Centre is contributing to this development – in particular developing the MOTIIVE Scenario and Use Case, and validating the draft INSPIRE Methodology for developing Data Specifications. In addition, a MOTIIVE data model for tidal observations and simulations has been developed by extending the NDG data model, CSML.

INSPIRE will provide an information fabric for the European initiative 'Global Monitoring for Environment and Security' (picture courtesy of www.gmes.info).

Standards activities

Many of the technical contributions of the e-Science Centre in the area of environmental data integration are based on an emerging standards framework for interoperability of geospatial data. While these standards have a heritage in traditional Geographic Information Systems (GIS), their scope is much broader, and will provide the basis for implementing the INSPIRE Directive, for instance.

However, problems have been identified by the e-Science Centre with some of these standards. For instance, vertical referencing may be based on atmospheric pressure rather than height; and gridded datasets may have irregular spacing – aspects not yet supported by the emerging standards. Strategic activity has been undertaken in order to remedy some of these shortcomings – for instance a new work item (ISO 19111-2) has been approved by the International Standards Organisation (ISO) to extend vertical reference systems, and a GML Change Request is being formally considered by the Open Geospatial Consortium (OGC) to implement irregular grids.

Brief News

UK e-infrastructure plans

In February '07 the Office of Science and Innovation (OSI) published a report entitled Developing the UK's e-infrastructure for science and innovation which sets out the requirements for a national e-infrastructure to help ensure the UK maintains and indeed enhances its global standing in science and innovation in an increasingly competitive world. Several members of e-Science contributed: Alveno Arenas, Juan Bicarregui, Neil Geddes, Jens Jensen, David Kelsey and Brian Matthews.

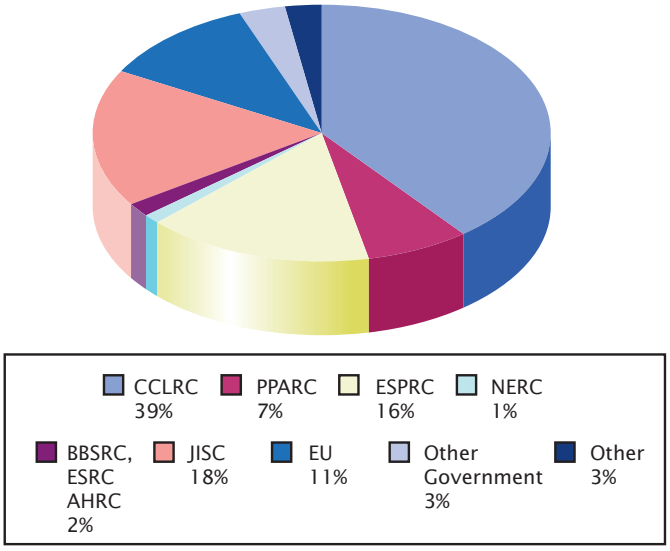
New library spaces

During the year the Chadwick Library at Daresbury was relocated and refurbished and is now better suited to support the research and development activities on the Campus.

A similar refurbishment exercise for the RAL library was also begun.

e-Science financial summary

e-Science income has risen by 40% since the previous financial year mainly because of the incorporation of the staff from the old CCLRC BITD department at the beginning of the year which has increased the EU funding for research projects with industry, and included CCLRC funding for the library. There was also an increase in the JISC funding to cover a large capital investment for new equipment for the NGS.



The sources of the £10.8m income of the e-Science centre in 2006/7.

e-Science receives 61% of its income from sources outside the CCLRC as a result of competitive grant applications and tenders, which indicates the quality of the work undertaken. This income includes contributions from seven of the research councils, which shows that the infrastructure provided through the main funds is added to by contributions from a broad range of disciplines who use it.

Innovation Award Winner

Kirsten Kleese van Dam, pictured with the Daresbury based part of her group, was awarded the British Female Inventors and Innovators Network (BFIIN) Silver Award 2006 in the Information Technology and Engineering category for her work on Effective Management of Scientific Data.

Events during the year

The UK e-science stand at the international supercomputing conference held in Tampa, Florida, USA in November '06, was managed by CCLRC.

The speakers at the conference on XML Access Languages held at the Rutherford Appleton Laboratory in September '06, organised by XML UK and the W3C Office for the UK and Ireland, which is hosted by e-Science. From left to right, Andy Seaborne of HP Research Labs, Dr Brian Matthews of CCLRC, Liam Quin of W3C, Michael Kay of Saxonica, Mark Birbeck of x-port.net, Chris Wallace of the University of the West of England.

The e-science centre's Prof Michael Wilson presented several invited keynote speeches at conferences during the year, as well as addressing the ITRE committee of the European Parliament in May '06 to summarise the arguments for a 700 million budget for Future and Emerging Technologies research in the EU FP7 ICT programme.

Andrew Richards at the NGS stand at the JISC Conference in March '07. The NGS presented the services that it offers to the UK research community.

STFC chief executive Prof Keith Mason visited e-Science in Daresbury in March '07.

World LHC Computing Grid collaboration workshop at CERN, Geneva in Jan '07 where e-Science's Neil Geddes and John Gordon each presented reports on aspects of the UK's status in the WLCG.

In October '06, e-Science organised an international workshop in Pisa as part of the EU funded Challengers project which produced a report as guidance for European Union strategic planning entitled Interaction of European and International Grid Communities.

Publications 2006

A Akram, D Meredith
Approaches and best practices in Web Service Data Binding
in Securing web services: practical usage of standards and specifications, Eds. P Periorellis , ISBN 1599046393 (Idea Group Inc, USA), (2006)

A Akram, D Meredith, R Allan
Best practices in web service style, data binding and validation for use in data-centric scientific applications
Proc. *UK e-Science All Hands Conference 2006 (AHM 2006), Nottingham, UK, 18-21 Sep 2006*

A Akram, D Chohan, D Meredith, R Allan
CCLRC Portal Infrastructure to support Research Facilities
Concurrency and Computation: Pract. Exper. 2007; 19:751-766 (Special issue on Science Gateways)

A Akram, J Kewley, R Allan
Data Centric approach for Workflows
Proc. *The Tenth IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2006), Hong Kong, China, 15-20 Oct 2006*

A Akram, J Kewley, R Allan
Modelling WS-RF based Enterprise Applications
Proc. *The Tenth IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2006), 15-20 Oct 2006*

A Akram, R Allan
Organization of Grid resources in communities
Proc. 4th Workshop on Middleware for Grid Computing (MGC 06), Melbourne, Australia, 27-29 Nov 2006

A Akram, R Allan
Organization of Virtual Communities
Proc. *2 nd International Conference on Semantics, Knowledge and Grid (SKG 2006), Guilin China, 31 Oct - 03 Nov 2006*

A Akram
Resource sharing among multiple Web services, Part 1: Implement resource sharing using WSRF
IBM developerWorks

A Akram, R Allan
Virtual Peer Communities and the Community Coordinator
ISSN Number: 1738-6438

RJ Allan, A Akram, D Meredith
Best practices in distributed workflows and web services
Workflow Optimisation in Distributed Environments

RJ Allan, R Crouchley, M Daw
e-Collaboration Workshop: Access Grid, Portals and other Virtual Research Environments for the Social Sciences
Workflow Optimisation in Distributed Environments October 2006

RJ Allan, R Crouchley, CS Ingram
Workshop on e-Research, Digital Repositories and Portals
October 2006

AE Arenas, JC Bicarregui
Applying Unifying Theories of Programming to Real-Time Programming
Proc. *Integrated Design and Process Technology Conference: Critical Systems (IDPT-2006), San Diego, California, USA, 25-30 Jun 2006*

AE Arenas, JC Bicarregui
Applying Unifying Theories of Programming to Real-Time Programming
Journal of Integrated Design and Process Science **10** (4) *69-88 (2006)6*

Ed. A E Arenas
State of the Art Survey on Trust and Security in Grid Computing Systems
RAL Technical Reports, RAL-TR-32006-008 (2006)

AE Arenas, JC Bicarregui, T Margaria (U Potsdam)
The FMICS View on the Verified Software Repository
Journal of Integrated Design and Process Science **10** (4) *47-54 (2006)*

N Beagrie (British Library), R Boulderstone (British Library), L Estelle (Joint Information Systems Committee), J Giles (British Geological Survey), H Hockx-yu (JISC), M Jones (Digital Preservation Coalition), *et al*
e-infrastructure strategy for research: report from the DTI presentation and curation working group
Report of the Preservation and Curation working group of the DTI e-Infrastructure Steering Group

N Bennett, R Scott (CEH), M Brown (CEH), KD O'Neill, M Lane (CEH), A Woolf, *et al*
Application of the NERC DataGrid Metadata and Data Models in the NERC Ecological Data Grid
Proc. UK e-Science All Hands Meeting 2006 (AHM2006), 18-21 Sep 2006

JC Bicarregui, JCP Woodcock (York), M Butler (Southampton), D Crocker (Escher Technologies), S King (York), CB Jones (Newcastle), *et al* (Verified Software Repository)
Mondex Case Study First Workshop
Proc. *First Workshop of Mondex Project of VSR-net*

JC Bicarregui, JCP Woodcock (York), L Freitas (York), G Schellhorn (Augsburg), T Ramananandro (École Normale Supérieure), J Bowen (London South Bank), *et al* (Verified Software Repository)
Mondex Case Study Second Workshop
Proc. *Mondex Case Study Second Workshop, 25-26 May 2006*

JC Bicarregui, CAR Hoare (Microsoft Research), JCP Woodcock (York University)
The Verified Software Repository: A Step Towards the Verifying Compiler
Formal Aspects of Computing (2006)

L J Blanshard, L Roberts, P Couch, A Walkingshaw (University of Cambridge), P Murray-Rust (University of Cambridge), M Dove (University of Cambridge), *et al*
Dynamic Capture of Materials Properties
2nd German Conference on Chemoinformatics: CIC-Workshop: Chemoinformatics (GDCh-CIC), Goslar, Germany, 12-14 Nov 2006

L Blanshard, R Downing, G Drinkwater, D Hanlon, K Kleese-van-Dam, L Roberts, *et al*
The Storage Resource Broker and E-science in the UK
Proc. *SRB Users meeting to exchange ideas and provide feedback*

JM Brooke (Manchester Metropolitan University), J Marsh (Manchester Metropolitan University), S Pettifer (Manchester Metropolitan University), L S Sastry
The importance of locality in the visualization of large datasets
Concurrency and Computation: Practice and Experience (2006)

LB Casely-Hayford, S Sufi
The ISIS Facilities Ontology and OntoMaintainer
In: Proceedings of the UK e-Science All Hands Meeting 2006, In: Proceedings of the UK e-Science All Hands Meeting 2006

M Conrad (NARA), A Hasan, R Marciano (SDSC), R Moore (SDSC), C Wojcik (Michigan Dept of History, Arts and Libraries)
PAT Project Lessons Learned Part 2: The IT Professional's perspective
Newsletter of the Society of American Archivists

SY Crompton, BM Matthews, WA Gray (Cardiff U.), AJ Jones (Cardiff U.), RW White (Cardiff U.), JS Pahwa (Cardiff U.)
OGSA-DAI and Bioinformatics Grids: Challenges, Experience and Strategies
Proc. *6th IEEE International Symposium on Cluster Computing and the Grid: Computational Data and Information Grid (CCGrid06), Singapore, 16-19 May 2006*

R Crouchley, RJ Allan
The Future of Social Research on the Internet
The Future of Social Research on the Internet

M Dixon (London Metropolitan U.), JR Gallop, SC Lambert, L Lardon (INRA), JV Healy (London Metropolitan U.), J-P Steyer (INRA)
Data mining to support anaerobic WWTP monitoring
Control Engineering Practice (in press 2006)

MT Dove (U. Cambridge), TOH White (U. Cambridge), AM Walker (U. Cambridge), RP Bruin (U. Cambridge), KF Austen (U. Cambridge), E Artacho (U. Cambridge), *et al*
Computational Grids for Mid-Sized Collaborative Projects: The eMinerals Experience
Proc. *Second IEEE International Conference on e-Science and Grid Computing 2006, Amsterdam, The Netherlands, 04-06 Dec 2006*

R Downing, A Weise (Reading U.), C Koebernick (Reading U.), A Hasan
Some tools for supporting SRB Production Services
SRB Users meeting to exchange ideas and provide feedback

Z Du (Birkbeck College), VN Alexandrov (U. Reading), M Alfredsson (UCL), E Artacho (U.Cambridge), KF Austen (U.Cambridge), ND Bennett, *et al*
A Virtual research organization enabled by eMinerals minigrid: an integrated study of the transport and immobilization of arsenic species in the enviornment
eScience All Hands Meeting 2006

GT Folkes, E Auden (MSSL), P Lamb (MSSL), M Whillock (MSSL), J Jensen, M Wild
SolarB Global DataGrid
UK e-Science Programme All Hands Meeting 2006 (AHM2006), Nottingham, England, 18-21 Sep 2006

JC Garcia-Ojeda (UNAB), AE Arenas, JJ Perez-Alcazar (EACH)
Paving the Way for Implementing Multiagent Systems: Integrating Gaia with Agent-UML
6th International Workshop on Agent-Oriented Software Engineering (AOSE 2005), Utrecht, The Netherlands, 25-26 Jul 2005, in Agent-Oriented Software Engineering, eds. J P Muller, F Zambonelli, Lecture Notes in Computer Science, vol. 3950, ISBN 0302-9743 (Springer-Verlag), p179-189 (2006)

NI Geddes
The National Grid Service of the UK
Proc. *2nd IEEE International Conference on e-Science and Grid Computing (e-Science 2006), Amsterdam, 04-06 Dec 2006*

D Golby (BAE Systems), M D Wilson, L Schubert (HLRS), C Geuer-Pollmann (Microsoft Innovation Centre)
An assured environment for collaborative engineering using web services
Proc. *13th ISPE INTERNATIONAL CONFERENCE ON CONCURRENT ENGINEERING: Research and Applications (CE '06), Antibes, France, 18-22 Sep 2006, Concurrent Engineering '06; IOS Press*

D Grose (U. Lancaster), R rouchley (U. Lancaster), T van Ark (U. Lancaster), J Kewley, R Allan, A Braimah, *et al*
sabreR: Grid-enabling the analysis of multi-process random effect response data in R
Proc. *Second International Conference on e-Social Science.*

M Hankel (U. Queensland), SC Smith (U. Queensland), RJ Allan, SK Gray (Argonne National Laboratory), GG Balint-Kurti (Bristol U.)
State-to-state reactive differential cross sections for the H + H2 -> H2 + H reaction on five different potential energy surfaces employing a new quantum wavepacket computer code: DiffRealWave.
Journal of Chemical Physics (accepted 2006)

J Jensen, DR Spence, MJ Viljoen
Grid Single Sign-On in CCLRC
Proc. *UK e-Science All Hands Meeting 2006 (AHM2006), Nottingham, UK, 18-21 Sep 2006*

T Kirkham, T Varsamidis (UWB)
A business service network to aid collaboration between small to medium enterprises
Proc. *8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services*

T Kirkham, D Mac Randal, B Ritchie, I Johnson, J Gallop, T Varsamidis
Providing location sensitive e-business integration for the small to medium enterprise
Proc. *IEEE WETICE 2006*

T Kirkham, T Varsamidis (UWB)
Supporting E-Business integration in North Wales
Proc. *ECHALLENGES 2006*

T Kirkham
Burning down the eBusiness barriers facing Welsh small to medium enterprises
IEE Young Members Presentation Evening 2007 St Asaph

T Kirkham, J Gallop, T Varsamidis
Using User Driven Dynamic Grid Services and Temporary Virtual Organisations to integrate new users and old systems in e-science facilities
Proc. *eChallenges 2007*

T Kirkham, T Varsamidis (UWB), J Gallop, MD Wilson, J Kirkham
A framework to provide eBusiness support for African SMEs
Proc. *IST AFRICA 2007*

C Koebernick (Reading U.)
Performance Measurement System for a Storage Resource Broker
MSc, U. Reading (2006)

SC Lambert
Coordinating IST research and development across Europe: the CISTRANA outlook
Proc. *8th International Conference on Current Research Information Systems (CRIS 2006), Bergen, Norway, 11-13 May 2006*

S Lloyd, L Sastry
Integrative Biology - The Challenges of Developing a Collaborative Research Environment for Heart and Cancer Modelling
Future Generation Computer Systems (accepted 2006)

D Lowe, A Woolf, B Lawrence, P Cooper, R Lott
Standards Challenges in the Climate Sciences AGU Fall Meeting
San Diego, Dec 2006. [Eos Trans. AGU, 87(52), Fall Meet. Suppl., Abstract IN43C-0916]

D Lutz (HLRS), T Kirkham, P Mandic (HLRS), J Movilla (TID)
Identity in a mobile grid environment
Proc. *Submitted to UK e-Science 2007 All Hands Meeting*

T Mahler (NRCCL), AE Arenas, L Schubert (HLRS)
Contractual Frameworks for Enterprise Networks and Virtual Organisations in E-Learning
Proc. *e-Challenges 2006 (e-2006), Barcelona, Spain, 25-27 Oct 2006*

A Miles, BM Matthews, MD Wilson, D Brickley (W3C)
SKOS Core: Simple Knowledge Organisation for the Web
New Technology of Library and Information Service 1 (1) 3-9 (2006)

K Millard, A Woolf, J Tandy
Data Models - A Met and Marine Perspective
Proc. *EuroSDR/EuroGeographics Workshop on 'Feature/Object Data Models, Munich, Germany, 24-25 Apr 2006.*

S Naqvi, P Massonet (CETIC), AE Arenas
A Study of Language for the Specification of Grid Security Policies
CoreGRID Technical Reports, TR-0037 (2006)

S Naqvi, P Massonet (CETIC), AE Arenas
Scope of Forensics in Grid Computing – Vision and Perspectives
Proc. *International Symposium on Parallel and Distributed Processing and Application (ISPA 2006), Sorrento, Italy, December 2006, in Frontiers of High Performance Computing and Networking, Lecture Notes in Computer Science, vol. 4331, ISBN 9783540498605 (Springer), p964-970 (2006)*

S Naqvi, O Poitou (CETIC), P Massonet (CETIC), AE Arenas
Security Requirements Analysis for Large-scale Distributed File Systems
Proc. *CoreGRID Workshop on Grid Middleware, Dresden, Germany, 29 Aug - 01 Sep 2006, CoreGRID Series (Springer) (2006)*

S Naqvi, P Massonet (CETIC), AE Arenas
Security Requirements Model for Grid Data Management Systems
Proc. *International Workshop on Critical Information Infrastructures Security (CRITIS 2006), Samos Island, Greece, 30 Aug - 02 Sep 2006, in Critical Information Infrastructures Security, Lecture Notes in Computer Science, vol. 4347, ISBN 9783540690832 (Springer), p30-41 (2007)*

V Page (Kingston U.), M Dixon (London Metropolitan U.), I Choudhury (London Metropolitan U.)
Security Risk Mitigation for Information Systems
BT Technol J 25 (1) 118-127 (2007) [DOI: 10.1007/s10550-007-0014-8]

R Piotter (HLRS), T Kirkham, J Gallop, D Mac Randal, B Ritchie, I Johnson
Enhancing Grid computing applications that use distributed mobile services, by developing service to service understanding
IEEE CCGrid - Context-Awareness and Mobility in Grid Computing Workshop, Rio De Janero

LEC Roberts, L J Blanshard, K Kleese Van Dam
Enabling Effective Collaboration through a Web-Enabled Data Infrastructure
Proc. *The Fourth IASTED International Conference on Knowledge Sharing and Collaboration: Applications and Technologies (KSCE 2006), St Thomas, USVI, 29 Nov - 01 Dec 2006*

LEC Roberts, LJ Blanshard, K Kleese Van Dam, L Price (UCL Chemistry), SL Price (UCL Chemistry), I Brown (UCL Chemistry)
Providing an Effective Data Infrastructure for the Simulation of Complex Materials
Proc. *UK e-Science Programme All Hands Meeting 2006 (AHM 2006)*

G C Silaghi (STFC/U Coimbra), AE Arenas (STFC), Luis Silva (U. Coimbra)
Reputation-Based Trust Management Systems and their Applicability to Grids
CoreGRID Technical Reports, TR-0064 (2007)

DR Spence, NI Geddes, J Jensen, AJ Richards, MJ Viljoen, A Martin (Oxford U.), *et al*
ShibGrid: Shibboleth Access for the UK National Grid Service
Proc. *2nd IEEE International Conference on e-Science and Grid Computing (e-Science 2006), Amsterdam, The Netherlands, 04-06 Dec 2006*

DR Spence, NI Geddes, J Jensen, AJ Richards, MJ Viljoen, A Martin (Oxford U.), *et al*
ShibGrid: Shibboleth Access for the UK National Grid Service
Proc. *2nd IEEE International Conference on e-Science and Grid Computing (e-Science 2006), Amsterdam, The Netherlands, 04-06 Dec 2006*

PJ Strange, T Antoni (Karlsruhe), F Donno (CERN), H Dres (Karlsruhe), G Grein (Karlsruhe), G Mathieu (IN2P3/CNRS), *et al*
GLOBAL GRID USER SUPPORT: THE MODEL AND EXPERIENCE IN LHC COMPUTING GRID
Computing in High Energy and Nuclear Physics (CHEP06), Mumbai, India, 13-17 Feb 2006

A Svirskas (Vilnius U.), J Bedzinskas (Vilnius U.), B Roberts (Kingston U.), I Ignatiadis (Kingston U.), MD Wilson
Adaptive Business Protocol Support in Service Oriented B2B Collaborations
Proc.*PINCET, Manchester, UK, 26-28 Jun 2006*

A Svirskas (Vilnius U), MD Wilson, R Roberts (Kingston U)
Role and Application of Pluggable Business Service Handlers in Web Services Choreographies
Baltic DB&IS'06, Vilnius, Lithuania, 04-06 Jul 2006, in Databases and Information Systems V, eds. Olegas Vasilecas, Frontiers in Artificial Intelligence and Applications (IOS Press) (2006)

A Svirskas, B Roberts, MD Wilson
TOWARDS ARCHITECTURE OF MANAGED DYNAMIC VIRTUAL ORGANISATIONS FOR E-BUSINESS COMMUNITIES
Int. J. Web Based Communities 2 (2) (2006)

A Svirskas, B Roberts, I Ignatiadis, MD Wilson
VIRTUAL ORGANIZATION MANAGEMENT USING WEB SERVICE CHOREOGRAPHY AND SOFTWARE AGENTS
Proc. *PROVE'06, Helsinki, Finland, 25-27 Sep 2006*

A Terracina (DATAMAT), S Beco(DATAMAT), J Gallop, I Johnson, T Kirkham, DMac Randall, *et al*
Orchestration and workflow in a mobile Grid environment
Proc. *WSGE'06*

RP Tyer, K Kleese van Dam, PA Couch, IT Todorov, RP Bruin (Cambridge U.), AM Walker (Cambridge U.), *et al*
Automatic metadata capture and grid computing
Proc. *AHM2006, 18-21 Sep 2006*

AM Walker (Cambridge U), MT Dove (Cambridge U), L A Sullivan (Cambridge U), TOH White (Cambridge U), K Trachenko (Cambridge U), RP Bruin (Cambridge U), *et al*
Anatomy of a grid-enabled molecular simulation study: the compressibility of amorphous silica
Proc. *UK e-Science All Hands Meeting 2006 (AHM2006), Nottingham, UK, 18-21 Sep 2006*

XD Wang, M Gleaves, D Meredith, RJ Allan, C Nave
e-Science Technologies in Synchrotron Radiation Beamline - Remote Access and Automation (a case study for High Performance Protein Crystallography
Macromolecular Research 14 (2) 140-145 (2006)

XD Wang, X Yang, R Allan
Flexible Grid Portlets to Access Multi Globus Toolkits
Proc. *International Workshop on Collaborative Virtual Research Environments(CVRE06) in GCC2006 (CVRE06), Changsha, China, Oct. 2006*

XD Wang, X Yang, R Allan
Portlet Development and Deployment under Portal Frameworks
VRE Developers Workshop, Portsmouth, UK, Jan. 2006

XD Wang, X Yang, R Allan
Top ten questions to design a successful grid portal
Proc. *The 2nd International Conference on Semantics, Knowledge and Grid (SKG2006), Guilin, China, 30 Oct - 03 Nov 2006*

A Weise (Reading U.)
Investigation and Development of Monitoring Tools for a Storage Resource Broker
MSc, U. Reading (2006)

A Woolf, B Lawrence (British Atmospheric Data Centre), R Lowry (BODC), K Kleese van Dam, R Cramer (BODC), M Gutierrez BODC), *et al*
Data integration with the Climate Science Modelling Language
Advances in Geosciences 8 (1) 83-90 (2006)

A Woolf, B Lawrence, J Tandy, K Millard, D Lowe, S Pepler
'Feature types' as an integration bridge in the climate sciences
AGU Fall Meeting, San Diego [Eos Trans. AGU, 87(52), Fall Meet. Suppl., Abstract IN53C-02], San Diego, December 2006

A Woolf, N Bindoff, I Cumming (Insight4 Pty Ltd), G Hyland, J Roberts, S Phipps (Tasmanian Partnership for Advanced Computing (TPAC))
Exploring OGC service wrappers for integrating distributed earth-system science resources
European Geosciences Union General Assembly 2006 (EGU 2006), Vienna, Austria, 02-07 Apr 2006, Geophysical Research Abstracts, Volume 8, 10100, 2006 (2006)

X Yang, R Allan
A Deep Look at Web Services for Remote Portlets (WSRP) and WSRP4J
GCE'06: Second International Workshop on Grid Computing Environments, Tampa, FL, USA, 12-13 Nov 2006

X Yang, R Allan
Collaboration in a Virtual Research Environment
Poster at Second International Conference on e-Social Science

X Yang, XD Wang, R Allan, M Dovey (JISC), M Baker (Reading U.), R Crouchley (Lancaster U.), *et al*
Integration of Existing Grid Tools in Sakai VRE
Proc. *International Workshop on Collaborative Virtual Research Environments (CVRE06), Changsha, China, 21 Oct 2006*

X Yang, XD Wang, R Allan
JSR 168 and WSRP 1.0 - How Mature are Portal Standards?
Proc. *WEBIST 2006 (WEBIST 2006), Setubal, Portugal, 11-13 Apr 2006*

M van Assem (Vrije Universiteit Amsterdam), V Malaise, A J Miles, G Schreiber
Method to Convert Thesauri to SKOS
Proc. *3rd Annual European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 11-14 Jun 2006*

Photographs and illustrations provided by STFC Media Services.

Designed and produced by STFC Media Services

CCLRC

Science & Technology Facilities Council
e-Science

Further Information
e-Science Centre
STFC Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot, Oxfordshire 0X11 0QX
UK

T: +44 (0)1235 446084
F: +44 (0)1235 445945

www.e-science.stfc.ac.uk