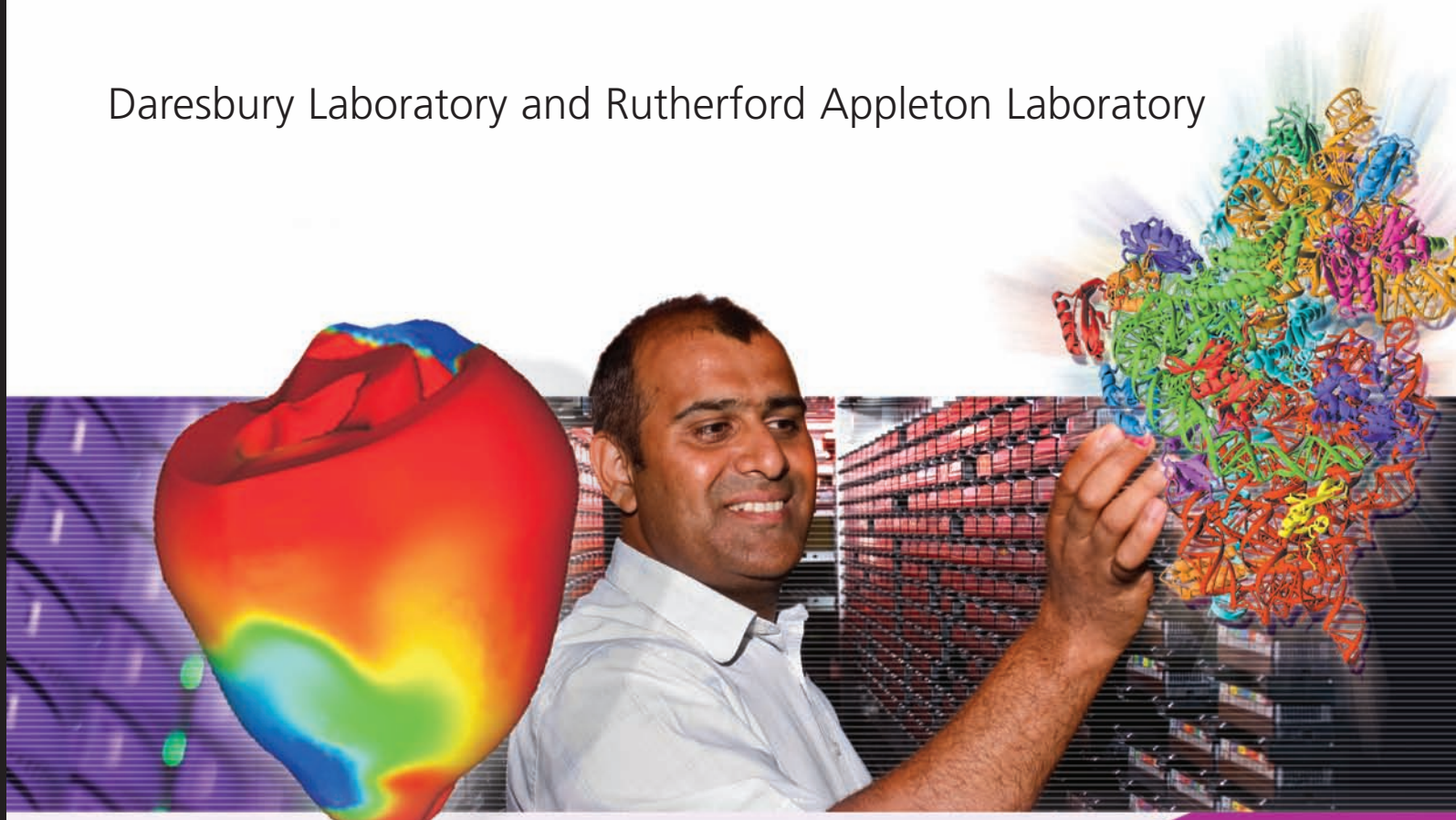
CCLRC

e-Science Centre

Annual Report 2005-2006

Daresbury Laboratory and Rutherford Appleton Laboratory



e-Science

Council for the Central Laboratory of the Research Councils

Photographs and illustrations provided by CCLRC Media Services.

Designed and produced by CCLRC Media Services.

CCLRC

The Council for the Central Laboratory of the Research Councils (CCLRC) is one of Europe's largest multidisciplinary research organisations and supports scientists and engineers world-wide.

The CCLRC is an independent, non-departmental public body reporting to Parliament and sponsored by the Office of Science and Innovation (part of the Department of Trade and Industry). It operates world-class large scale research facilities, provides strategic advice to government on their development and manages international research projects in support of a broad cross-section of the UK research community. In partnership with the other research councils, the CCLRC sets future priorities to meet UK science needs. The CCLRC operates three world-class research centres: the Rutherford Appleton Laboratory (RAL) in Oxfordshire, the Daresbury Laboratory in Cheshire, and the Chilbolton Observatory in Hampshire.

"It's significant that the UK is the first country to develop a national e-Science Grid, which intends to make access to computing power, scientific data repositories and experimental facilities as easy as the web makes access to information."

The Prime Minister, Rt Hon Tony Blair MP, July 2002

"e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it."

John Taylor, Director General of the Research Councils 1999-2003

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

e-Science will be vital to the successful exploitation of the next generation of powerful scientific facilities operated by CCLRC on behalf of the UK research community, and to the UK's effective use of major facilities elsewhere. These facilities – spallatial neutron source (ISIS), synchrotrons (DLS), satellites, telescopes, high power lasers (CLF) - 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 are developing the infrastructure which enables this.

The CCLRC e-Science centre operates from both the Rutherford Appleton Laboratory and the Daresbury Laboratory.

The e-Science centre undertakes a programme of work including hosting 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 CCLRC facilities.

Foreword

It is a delight to introduce the first annual report of the CCLRC e-Science Centre. As this is the first edition, it is perhaps as much an introduction as an annual report. The Centre has now been in existence for five years, growing from an initial staff of three (!) to an interdisciplinary team of 100 e-science professionals. We have expertise in system operations, applications development and information management and are helping to create, disseminate, and preserve scientific data, information, and knowledge across the full range of science and engineering. The growth of the Centre reflects the growing importance of e-Science across CCLRC's activities and puts us in an unrivalled position in the UK with respect to the breadth and depth of expertise that we can bring to bear on challenging problems.

E-Science brings the rapid advances in computing, computer science and communications and information technology to bear on the major challenges in science and engineering. Both physical and biomedical science are moving into an era of increasingly collaborative endeavour. CCLRC is no newcomer to collaboration, but now, our activities frequently cross traditional subject boundaries and depend upon massive distributed datasets with correspondingly massive data processing demands. To fully exploit the potential offered by these endeavours requires new methods of data federation, distributed computation, visualisation and interactive analysis, while working in geographically distributed teams. The e-Science Centre exists to spearhead the exploitation of e-Science technologies throughout CCLRC's programmes, the research communities they support and the national science and engineering base.

Almost all the work of the e-Science Centre is collaborative with a focus on developing integrated solutions to support the broader science programme. Working with CCLRC facilities naturally encourages an outward looking philosophy and encourages collaborations with facility users in UK universities and beyond. Collaborative developments remain important across the full spectrum of our activities and are underpinned by our core computational, data and information services. If you have a challenging problem then we are probably interested in the solution!

This report gives a brief introduction to some of our activities and a snapshot of our technical expertise. The report is necessarily focused on the technical capabilities that we can offer, with expertise in large scale data storage, cluster computing, security, information and data management, visualisation and grid computing. However, ultimately it is by pushing forward the boundaries of science and engineering that our success is measured, and we include some specific examples of deployed solutions.

It has been an exciting five years so far, and we expect no less in the next five. The start up of the LHC, the Diamond Light Source and the ISIS second target station are all imminent, as is an upgrade for the Central Laser Facility. Developing, deploying and supporting the e-Science infrastructure to fully exploit these and other national and international facilities is a challenge we look forward to.

Neil Geddes
Director
CCLRC e-science centre

Contents 2005-2006

Section 1 - Physical Resources

Section 2 - Capabilities

Section 3 - Applications and Service Delivery

Physical Resources

ATLAS Petabyte Data Store

The Atlas Petabyte Store (APS) provides very large amounts of digital data storage which is permanently accessible. It is used by CCLRC facilities and by many users in the UK research community.

This year the APS services have been developed significantly, in preparation for the start up of Large Hadron Collider (LHC) experiment at CERN in 2007, and for the predicted growth in data from CCLRC facilities and the external UK academic community. The in-house Atlas Data Store (ADS) system still provides the basic archive service, currently supporting around 135 users. In addition this year has seen the installation of a Hierarchical Storage Management system, which runs the SGI application Data Migration Facility (DMF). The DMF allows users to keep their data on disc (around 40TB) for ease of access, and to set policies specific to their community, which migrate their files from disc to tape if the data is not accessed after a particular time. Disc access is available via NFS, CIFS and GRIDFTP. This system is being used by ISIS, The Solar Terrestrial Physics World Data Centre, and is being evaluated for the SOLAR-B data archive facility.

The STK 9310 robot which was in action for around 20 years, has been replaced this year by a STK SD8500. The new robot has 10,000 slots, and uses the new generation T10K tape drives.

This upgrade has increased the total archive capacity from 1 to 5 Petabytes, although we expect this limit to be reached by 2008 when we anticipate further upgrades for storage to maintain provision in line with demand.

This robot forms the backbone of the APS services, and is used by the ADS system, the HSM system, and will also be used by CASTOR2.

CASTOR2 is being deployed at RAL to cope with the requirements of LHC – see page 11 for a description of CASTOR. The CASTOR2 system has been developed at CERN, in collaboration with RAL which have developed the SRM compliant interface, which provides data access for the particle physics community. CASTOR2 is scalable to the multi Petabyte level and is capable of handling millions of files. It will provide the foundation for future archive services for APS, and will eventually replace the ADS service. We are currently planning the development of an SRB interface to CASTOR2. This work will take place in the coming year, and will eventually replace the Storage Resource Broker (SRB) interface into the existing ADS system – the SRB is described further on page 14.

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

The STK 9310 robot which has been in action for around 20 years, has been replaced this year by the STK SD8500.

Physical Resources

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 DLS.

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 in Europe and Teragrid in the USA. The objective is to provide seamless access from the desktop to supercomputers.

The first SCARF user workshop was held on the 1st November 2005, and attracted 37 delegates. Since then, the number of registered users has increased to 70, from a total of seven CCLRC departments. We have purchased a 92 CPU core expansion to augment the compute capabilities of the cluster (now 1.5Tflops), together with an 84TB storage cluster that will be shared with the NGS.

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

Tier 1 Service

The Tier 1 service provides world class computing facilities for the UK particle physics community as part of the global research computing infrastructure.

In 2007 the Large Hadron Collider (LHC) at CERN will start taking data and by 2008 will be generating many Petabytes of data each year (Tier-0).The GRIDPP Tier-1 centre is one of approximately 10 large computing centres world wide who will provide the backbone of a massive, global Grid computing infrastructure that will process and analyse the data. It works with Tier-2 sites at UK universities to form GridPP, the UK particle physics Grid, which currently consists of more than 4,000 CPUs and 250 terabytes of disk storage capacity. RAL, and the global Grid are pushing the boundaries of sustained data transfer beyond previous limits.

The RAL Tier1/A service supports analyses and simulations to be run on data sets in parallel which would otherwise take up to 1000 times longer if they were run on a PC one after the other. The service is used by 18 virtual organisations from the LHC and BaBar particle physics community as well as the Mice, SNO and UKQCD experiments and some theory users.

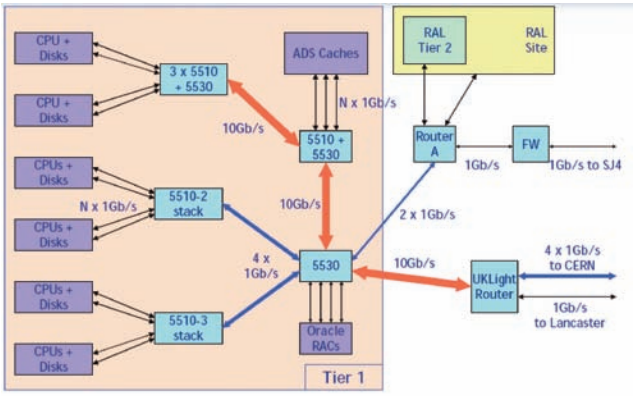
The service is maintained at the leading edge of technology with regular upgrades to be able to achieve the extreme service requirements of the LHC experiments when they begin in 2007.

Four years ago 35 TB of storage was installed which was increased in succeeding years by 40 TB and then 160 TB until it was enlarged by 21 new servers in March 2006 including 168 TB of usable storage. The first 35TB is now becoming hard to find spares for, and will be decommissioned in 2006 when the latest upgrade is operational.
For processing, a similar upgrade path is being followed to maintain a premier service. The service is designed for jobs which use integer rather than the floating point calculations more common on supercomputers, so the conventional high performance benchmarks are not comparable. The service currently consists of over 1000 processors which provided more CPU cycles to the LHC than any other Tier-1 centre in 2005.

Network connectivity is equally undergoing managed upgrades with the 10Gb/s backbone for the Tier1 LAN expected to be completed by May 2006. The 10Gb link to the dedicated optic fibre UKLight, and through that on to CERN is expected to be operational by the summer of 2006. The link from RAL to the UK academic network SuperJanet 5 is expected to be operating at 10Gb/s by autumn 2006.

Security is becoming an increasing concern for Tier1 and similar remote access high performance computing systems. Providing a secure firewall at 10Gb/s is already pushing the limits of available technology. Hacker discussion forums are beginning to focus on breaking into Grid services, requiring counter measures to be put in place against these threats.

The Tier-1 service successfully took part in a challenge to test an international scientific computing Grid under working conditions. During the week-long challenge, the LHC Computing Grid sustained transfer rates of a gigabyte per second. - a world first for a permanent, international Grid using scientific data. Completion of the tests was announced on February 15th 2006. Dr Andrew Sansum, the manager of the particle physics computing centre at RAL, was pleased with the results of the tests, "By receiving nearly 200 megabytes per second from CERN, we went well beyond our target data rate. With the installation of the new academic network, SuperJanet5, later this year, we plan to double these data rates - and then we'll really be approaching the speeds we need when the LHC comes on line in 2007".



RAL Tier-1 connectivity in March 2006.

Capabilities

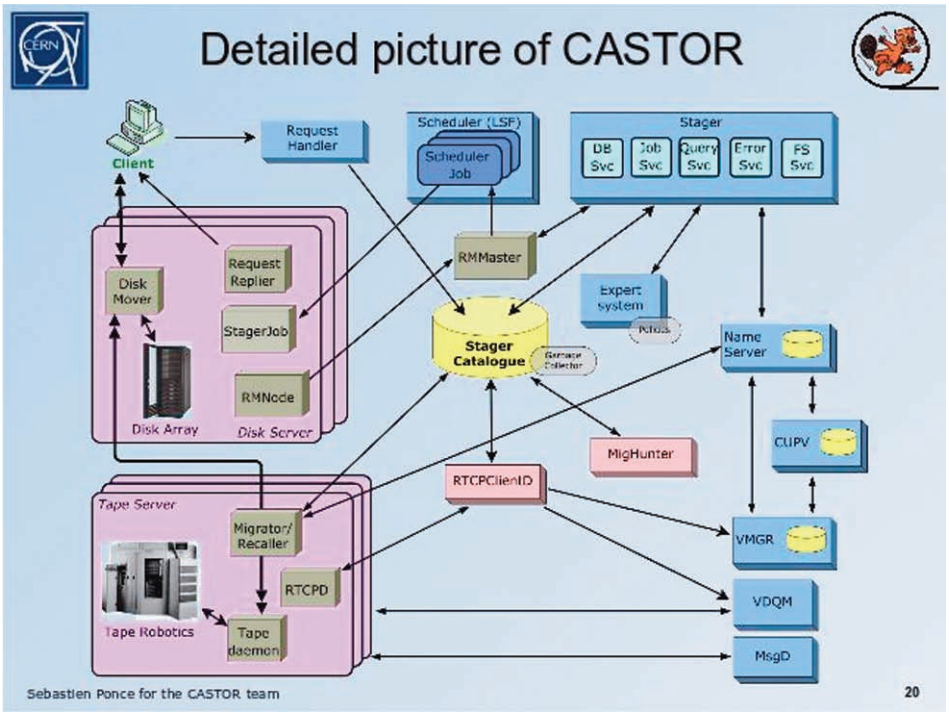
Database Services

The CCLRC e-Science centre provides production quality services of which many have a database as a key component. The Scientific Database Services group play a major part in many systems including Storage Resource Broker, the LHC Computing Grid's Distributed Deployment of Databases project, CASTOR (CERN Advanced STORage)and the National Grid Service.

The data management group has recognised the need to establish a database service within the e-Science centre in CCLRC.
The database services team currently provide systems to many internal and external customers supported by a team of specialist database administrators and developers. Database services provide the backbone for applications such as the Storage Resource Broker (SRB) where the resource broker depends on the database-stored metadata catalogue (MCAT). Capable of storing millions of files in disparate locations, all SRB metadata are stored in the database which interacts constantly with the SRB servers. Database services operate these systems on Oracle 10g Real Application Clusters for maximum performance, availability and manageability.
Database Services are playing a vital role in the development of the Petabyte datastore application CASTOR (CERN Advanced STORage). The backbone of CASTOR is an Oracle database used for all interactions between the clients and the tape store.

As an addition to the compute and data services offered by the National Grid Service (NGS) a database service is supported and administered on the RAL node. The five-node production Oracle cluster hosts services for collaborators across the UK.

Working with the Tier-1 group at RAL, database services have also been involved with the 3D project since it started two years ago. The aim of the project is to stream database data from Tier-0 (at CERN) to various Tier-1 and Tier-2 sites across the world for the Large Hadron Collider. Based on the Oracle Streams technology, a pre-production cluster is currently being installed at RAL in time for the start of the service in autumn 2006.



Contact: Gordon Brown E-mail: G.D.Brown@rl.ac.uk

Data Management Technologies

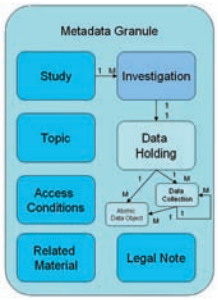
CCLRC's Facilities are producing large quantities of high quality data every year. The volumes are rising exponentially due to the use of new enhanced detector technologies and changes to the facilities that allow for faster turn over times.

The wealth of information encapsulated in this data is seen as real asset for CCLRC, which should be exploited much more in the future to further the advances of science. Data management technologies developed by the e-Science Centre enable easy often automatic management and enhancement of the data captured at Facilities, making it available for access, evaluation and reuse. As many modern research projects include the collaboration of several research groups the technologies are also supporting secure data sharing and collaborative working. Key technologies in the area of data management are SRB - for virtual distributed file storage management and data sharing, Metadata models for data annotation and capture of key scientific parameters, AgentX, Ontology's and RCommands for automatic metadata capture and data exploration, Metadata editor for user access to the metadata, Data Portal for easy and save access to the data across different facilities and company boundaries.

With ever increasing volumes of data to be handled maintenance will only be possible through a strategy based on a suite of well developed technology. Some facilities have very good local data management policies these could be enhanced allowing integration into wider CCLRC wide based systems while maintaining local autonomy. Much data however is still being kept on media which is offline with indexing information and proposal information in handwritten notes and printouts, this situation can cause data loss. Even if some research communities that are served do not have a culture of sharing their data often, they need to rediscover their data for checking and new methods of analysis. There is a danger also that with the loss of skilled staff the key to accessing the data could be lost. While there is a long way to go the e-Science centre is dedicated to developing data management technologies which could improve data practice at CCLRC and beyond.

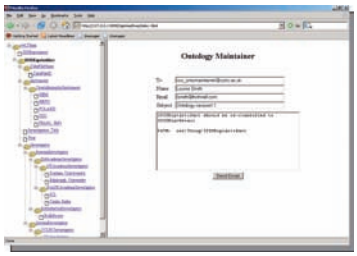
Scientific Metadata Model

Although there are standards for describing scientific research projects (e.g. CERIF recommended by the European Commission), and the data produced in individual scientific fields (e.g. NetCDF recommended by the University Cooperation for Atmospheric Research - UCAR), no general metadata model existed for linking together all the information about a study which was common across the many scientific fields that CCLRC addresses. Such a model is required to support the interdisciplinary use of data from the CCLRC facilities, so one was created by combining the facilities of catalog based systems such as Dublin Core and other low level scientific details inspired from formats such as XSIL with a hierarchical data organisation and multi level indexing support. This CCLRC Scientific Metadata Format is now in its 2nd version (CSMD v2) and has been used as the basis for the ISIS information catalog (ICAT) 20 year back catalog schema, the CLF CAT and future DLS CAT as well as on the e-Minerals and e-Materials e-Science projects. An earlier version was also used as the basis for the MyGrid Information Model.



Ontologies

Ontology has been defined as a the study of existence in order to define the basic categories of things and the relationships between them, or more simply as a 'specification of a conceptualisation'. Ontology provides formalisms for describing concepts and the relationship between them. Ontological descriptions expressed in those formalisms are a crucial step in converting unstructured keywords used in a particular domain into a well defined controlled vocabulary; decreasing ambiguity of terms, showing relationships between terms, and potentially being a building block for mapping terms from one area (e.g. ISIS and neutron scattering) to another (e.g. DLS). As well as developing a set of ontologies for ISIS related to Instrument, Experiment, Investigator and Sample information we are serving these in an application called an Ontology Maintainer which will allow users to comment on the ontologies suggesting amendments and additions to them; for example, in one case facility users are annotating their electronic ISIS proposals with such terms.



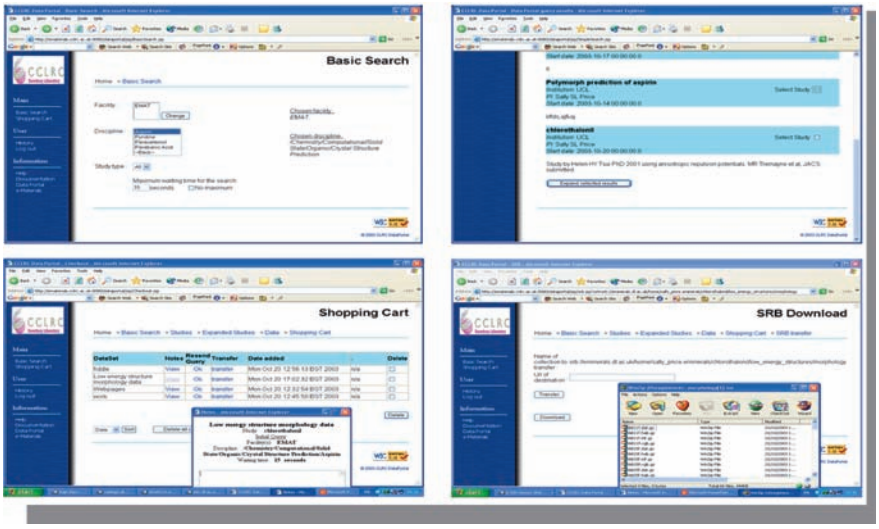
The Ontology Maintainer

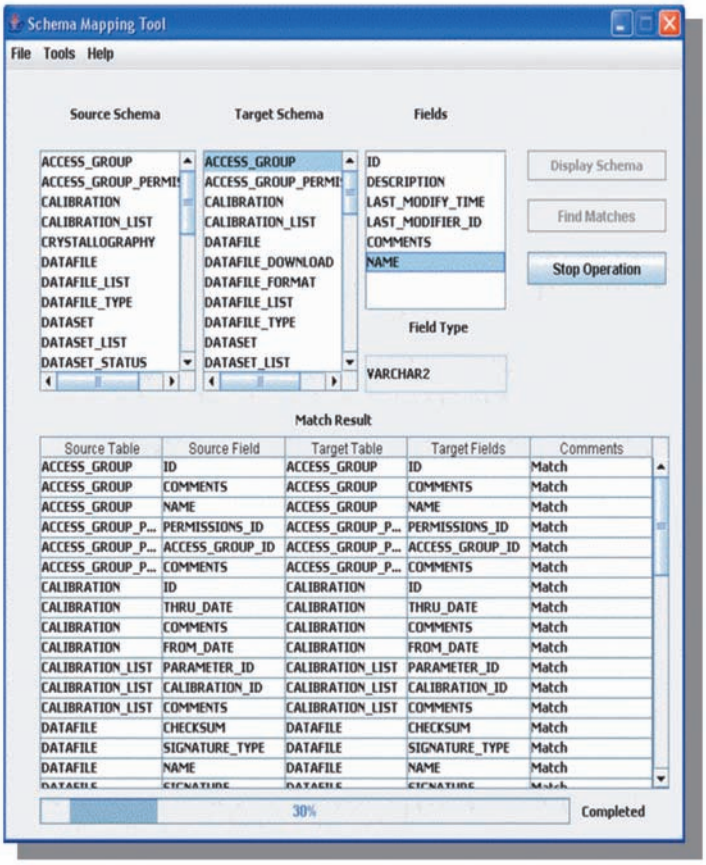
Capabilities

DataPortal

The CCLRC DataPortal aims to be for a 'one stop shop' for searching for and retrieving data produced at CCLRC. It is built using Grid and enterprise middleware to allow advanced federated search and third party transfer of scientific studies and the data that they produce.

Data Portal version 4 is the latest development version whose creation was guided by close work with ISIS with a view to offering the ISIS 20 year back catalog as the primary demonstrator of the technology. Other facilities such as DLS and CLF have stated an interest in integrating with DataPortal.





The user interface to the schema mapping tool from the Metadata Manager.

Proposal Systems and Metadata Managers

The electronic proposal system developed by ISIS is the first stage in metadata ingestion; it is the one time that users are forced to put in metadata about the experiments that are planning to perform; these can, once approved be fed into the various facility catalogs with data associated with them once the experiments are actually performed.

Currently e-Science is helping take the proposal systems developed by ISIS and adapting it for use by other facilities such as CLF and DLS. Also, while the ISIS Metadata Manager for their local catalog meets the user requirements for local search we are adding additional editing functionality to allow the user offices and scientists to edit data in their associated facility catalogs to broaden its user base, and thereby increase the power of the information available to any individual.

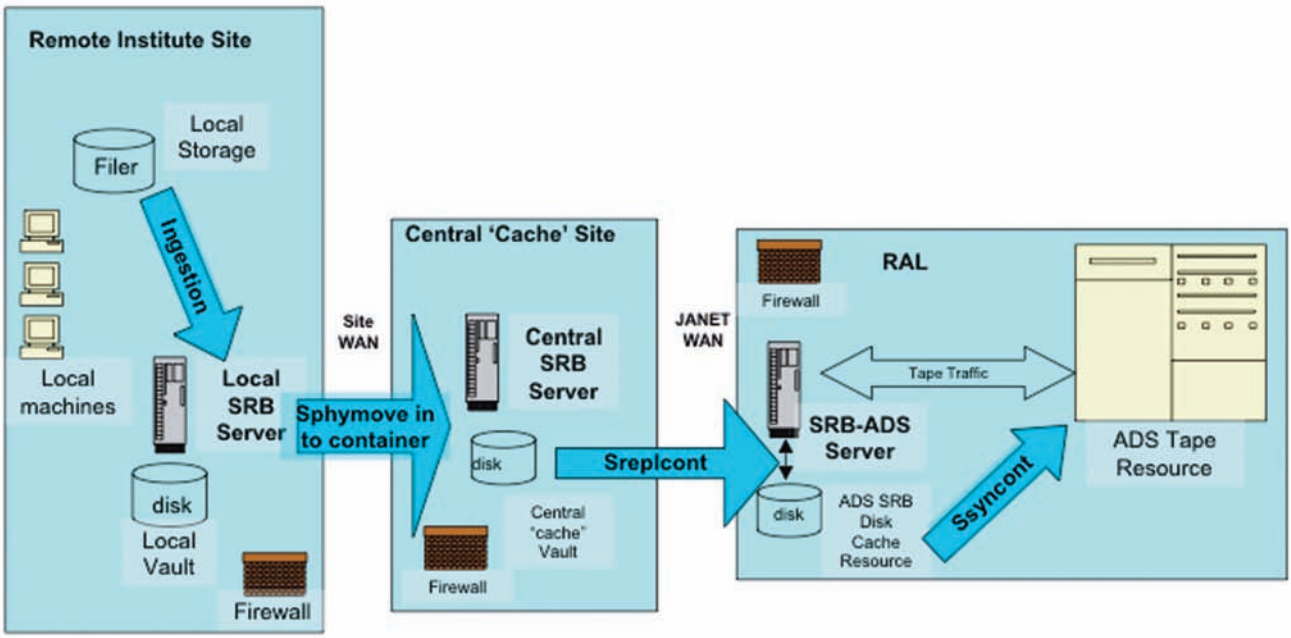
Contact: Kerstin Kleese-Van-Dam E-mail: kleese-van-dam@dl.ac.uk

Data Management with the Storage Resource Broker

The Storage Resource Broker (SRB) is a client-server middleware system originally developed by the San Diego Super Computing Centre (SDSC). The SRB system provides a means of accessing distributed heterogeneous storage resources in a uniform manner.

The MCAT (metadata catalogue) is the heart of an SRB system. The MCAT stores the logical-to-physical file mappings, in addition to the ability to add limited extra user-defined metadata attributes, allow users the ability to arrange data in a project-meaningful manner. The diagram below shows the typical flow of data for an SRB system for one project. The data is stored locally in an SRB managed storage device and then daily moved to a central cache SRB resource. At regular intervals the data are backed up into the RAL Atlas Data Store.

Within the UK the data management group has used the SRB to create data management systems for a variety of projects covering many disciplines including BBSRC, ISIS, NGS, DLS, the Arts & Humanities Data Service (AHDS) and CLF. The e-Science Centre Data Management Group (DMG) group have developed a number of generic tools that are proving to be of value within the data management community: a request tracker system that keeps track of the various stages of the archival process, a suite of unit-tests for testing commonly used SRB features, a system for monitoring the distributed log files and a system for measuring the performance of various SRB features. These tools are proving to be essential tools necessary for building a production quality data management system.



The SRB Architecture

Capabilities

Digital Curation

Curation of scientific data and records is inherently required over a very long timescale. CCLRC already curate data from Her Majesty's Nautical Almanac Office spanning back to 1767.

Examples from the CASPAR project http://www.casparpreserves.eu

The curation of digital scientific data will in future require a multi-disciplinary, collaborative approach, in many aspects of which CCLRC is already playing a leading role.

The Atlas Petabyte Store and the Data Management Group show best practice in the areas of bit preservation and data publication, as has been detailed elsewhere in this report.

The first requirement of Digital Curation is digital preservation by which one means preservation of the information encoded in digital objects. The continued understandability and usability of the digital object is the test of success of preservation. The Open Archival Information Systems Reference Model (ISO 14721) is a fundamental standard which forms the basis of the majority of large digital preservation projects throughout the world. CCLRC played a leading role in its development and is helping to organise its forthcoming review.

Advice and research into techniques to support digital curation are being undertaken in several areas. The Digital Curation Centre, of which CCLRC leads the Development arm, is recognised as one of the main resources in the UK for advice and outreach. The Development team is creating a Representation Information Registry/Repository, with supporting tools, which is a key component for an infrastructure to support digital curation. Interoperability with related efforts internationally is a primary requirement.

The quality of metadata is also the subject of the work of two PhD students which CCLRC is sponsoring.

Internationally there is much activity. Of particular importance is the CASPAR project (Cultural, Artistic and Scientific knowledge for Preservation, Access and retrieval), which is large Integrated Project part funded by the EU, and which is led by CCLRC.

With a total spend of 16 million Euros over three years, and a total of 17 partners including the European Space Agency and UNESCO, CASPAR aims to provide components of a common preservation infrastructure. To do this it will be challenged by digitally encoded information from a wide range of disciplines, and be validated by a set of testbeds in major organisations.

The Task Force for the Permanent Access to the Records of Science brought together key organisations across Europe in order to develop a research programme and strategy for a support infrastructure for European digital preservation. This is being carried forward by a Steering Group in which CCLRC plays an important role.
To underpin all these activities there has been the demand to have a way of evaluating digital repositories, to provide some way of identifying the trustworthiness of an archive. CCLRC has been part of an international team which has released a draft checklist and will lead the production of a full ISO standard on which a full accreditation and certification programme can be built.

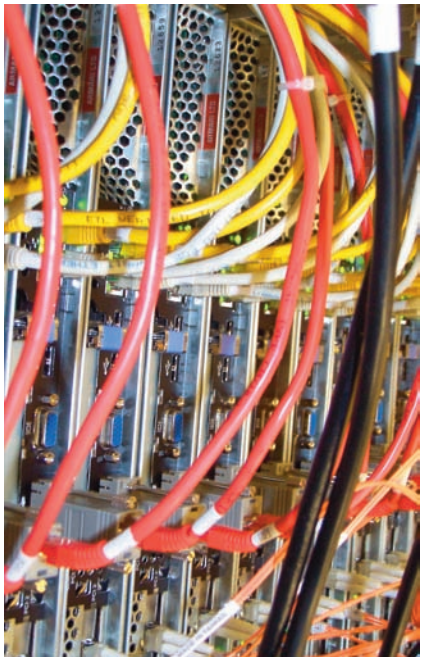
Contact: Dr David Giaretta E-mail: D.L.Giaretta@rl.ac.uk and Dr Juan Bicarregui E-mail: J.C.Bicarregui@rl.ac.uk

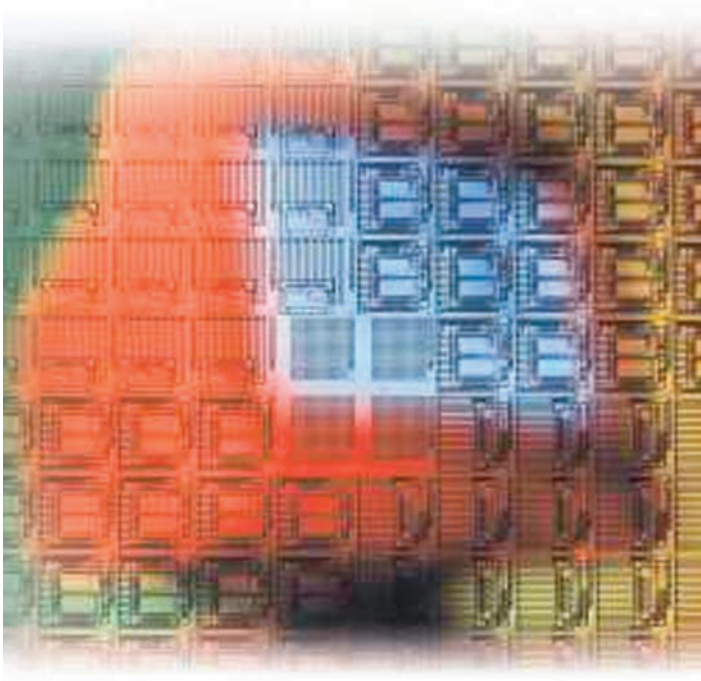
Grid Operations

The National Grid Service (NGS) is the UK's national production level Grid for e-Science.

The NGS forms a central focus for computer and data resources for UK academics, provides a framework for resource sharing and coordination and encourages their use particularly by those academics who ordinarily would not have access to Grid type resources. NGS resources are centrally funded by JISC, EPSRC and CCLRC, and are available free at point of use to UK researchers. The NGS is coordinated by the CCLRC e-science centre.

The NGS currently links the Universities of Bristol, Cardiff, Edinburgh, Lancaster, Leeds (representing the White Rose Consortium), Manchester, Oxford, Westminster and the CCLRC. Membership advantages include being able to share resources and having access to expensive and specialist facilities. For example, if one university department finds itself short on computing power it can log onto the NGS and team up with another university to share the computing load. Specialist research services such as BLAST (genome sequencing software) or FFTW (Fastest Fourier Transform in the West) can be shared across institutions. Access to data is also central to the NGS. NGS researchers will benefit from remote access to an increasing range of national and international data sources, including experimental data from cutting edge research facilities, such as those hosted by CCLRC.





In October 2004 the NGS entered full service. Over the last six months the NGS has brought together a wide diversity of users, ranging from computer scientists to archaeologists and social scientists. Radiologists from opposite ends of the UK have been able to share data, and biologists from Bristol and Edinburgh have been able to work together on molecular dynamics models, without having to be in the same city. What is more, the NGS has encouraged interdisciplinary research , bringing together specialists from different disciplines who might never normally meet.

The NGS hosts the UK e-Science Certification Authority (CA) which issues digital identification certificates for users of Grid resources across the NGS and EGEE. In this role the e-Science centre represents the UK in the global network of Grid CAs – the International Grid Trust Federation (IGTF). The UK e-Science CA is currently the second largest CA in the world having issued over 8000 certificates during its four year existence. The CA acts as a focus for developments in Grid security by CCLRC and partners in the UK and overseas as a result of which CCLRC leads the security policy activities in LCG and EGEE.

The NGS is also working toward integrating with other Grid operations all over the world. One of the biggest challenges is to find compatible software, but gradually institutions from as far afield as the USA, Japan and Australia are being linked up with institutions in the UK. It is an exciting time for the NGS, and the CCLRC is proud to be playing such a fundamental role.

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

Capabilities

Enabling Grids for e-sciencE - EGEE

The EGEE Grid consists of over 20,000 CPUs available to users 24 hours a day, 7 days a week, in addition to about 5 Petabytes (5 million Gigabytes) of storage, and maintains 20,000 concurrent jobs on average.

The map uses Googlemaps to show site availability across the world according to the test results performed by the operations team at CERN

In its second phase from 2006-2008, EGEE is extending its Grid service both to more disciplines, and
to more countries across the world. The CCLRC e-Science Centre is the UK host for EGEE.

The work being carried out within EGEE is organised into three main areas:

- Networking Activities which are the management and coordination of all the communication aspects of the project and also training;
- Service Activities are the support, operation and management of the Grid as well as the provision of network resources;
- Research Activities concentrate on Grid research and development.

The e-Science Centre participates in the EGEE activities relating to Operations and Support. The work undertaken is in collaboration with GridPP, The National Grid Service and also Grid Ireland. Services offered are:

A production service - uses the gLite middleware stack, with new resource centers and new applications being added frequently. This is where applications will run their production work, and should expect (eventually) 24x7 support.

A pre-production service - run in parallel with the production service, but in a way that is as much like a production service as possible. In general this is where the next version of the middleware is first deployed, where applications test that it is useable, and that the site operators verify that it meets their needs too.

The UKI ROC - offers grid support, operation and management and includes tasks such as VO-related services, database services, grid monitoring. The operations structure within the UK is organised into tiers. From RAL not only is there the central UKI ROC co-ordination but also a Tier 1 centre and Tier2 Particle Physics site.

An accounting and monitoring service – the accounting service comprises data collection and reporting based around 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.

GOCDB - A centralized repository of data used by configuration, monitoring, accounting and reporting tools. There is also a web-based user interface for manual entry and retrieval of the data. In February 2006 the LHC Computing Grid using the RAL based UK Tier 1 node announced sustained transfer rates of a gigabyte per second. - a world first for a permanent, international Gridusing scientific data.

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. EGEE is also collaborating with other projects to extend its coverage to the Baltic States, the Balkans, South America and Far East Asia.

Contact: Phillippa Strange E-mail: P.J.Strange@rl.ac.uk

Grid Technology

Portals are gateways to distributed electronic resources for science. These resources include data, computers and applications, publications and information, and even instruments ranging from telescopes to microscopes.

Web-based portals do not require the installation of any additional software on the user's machines, just a normal web browser will do. The resources are therefore available any time, any where, even from a PDA or up-market mobile phone.

The Grid Technology Group at CCLRC are working with Java standards and open source frameworks such as uPortal and Sakai to develop and deploy portal-based tools to meet a range of requirements.

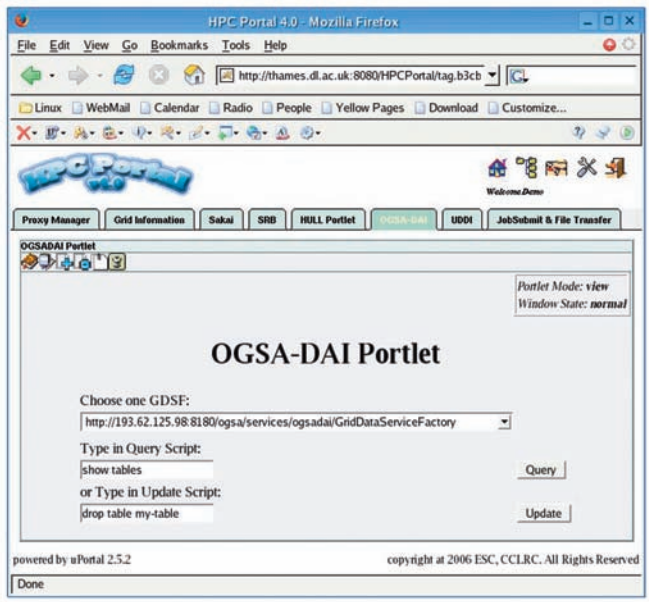
HPCPortal v4.0 is a generic example of a compute portal based on uPortal, which is also used as a testbed for the NGS Portal. See *http://thames.dl.ac.uk:8080/HPCPortal*

An applied example is the NGS Portal which uses HPCPortal portlets, although currently in the StringBeans framework. This is delivering a customized portal interface to the National Grid Service for generic compute resource and job management.

All NGS users can use this portal which is hosted on an IBM BladeCenter at the Daresbury Laboratory. The JISC-funded GridShib project with partners at CCLRC and University of Oxford is developing a Shibboleth security interface for this portal. See *http://portal.ngs.ac.uk*

The e-HTPX Project has now had a portal interface for a couple of years and has had extensive user feedback on both architecture and features. This portal is used to manage the experimental and data analysis workflow for high-throughput protein crystallography and supports both academic and commercial users. Partners at Daresbury Laboratory and universities of Oxford and York have contributed, see *https://hub.e-htpx.ac.uk/hub/*

DataPortal v3.0 is a generic example of how scientific data and metadata can be discovered and accessed in a variety of ways. DataPortal uses the CCLRC Scientific Metadata Model and XSL mappings to underlying data repositories. See *http://dataportal.dl.ac.uk:8080.*



Another example is the Sakai VRE Demonstrator, a JISC-funded Virtual Research Environment. The Sakai open-source framework was originally developed as a collaborative learning environment, but is being adapted for research purposes in this project. A range of on-line collaboration and Grid tools are being made available. Partners in this project are Daresbury Laboratory and universities of Lancaster, Oxford and Reading, see *http://www.grids.ac.uk*

Portal technology comprises: (1) a presentation layer which renders the content via HTML and cascading style sheets; (2) a control layer; and (3) the logic layer which communicates with the resources. By using standards such as JSR-168, WSRP and Java Beans these can be kept separate, thus enabling an architecture to be developed on top of a set of distributed re-usable services – a Service Oriented Architecture. Such services can be called from a programming function library, enabling more traditional applications such as GUI visualization tools to use data and compute resources on the Grid. Our JISC-funded GROWL project is doing this, see *http://www.grids.ac.uk/GROWL*

We encourage e-Science projects to develop their own portals for their Virtual Organisations. We will in the future supply myNGS – a portal toolkit on CD based on HPCPortal v4.0 with tried and tested services, which can be installed on a laptop or server. We hope to discuss your project requirements and how we can work together to create application-specific tools.

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

Applications and Service Delivery

Applications and Visualisation

The e-science centre provides expertise and support to improve the performance and scalability of scientific applications and include advanced higher dimensional visualisation and data analysis.

The Grid Applications Group within e-Science Centre bring together expertise in computational methods, grids and Web Services technology, advanced visualization techniques for very large datasets and real-time data analysis. Members of the group work closely with scientists working in a range of scientific specialisations from material science to biological systems to understand human disease processes, both within CCLRC and our academic colleagues nationally and internationally.

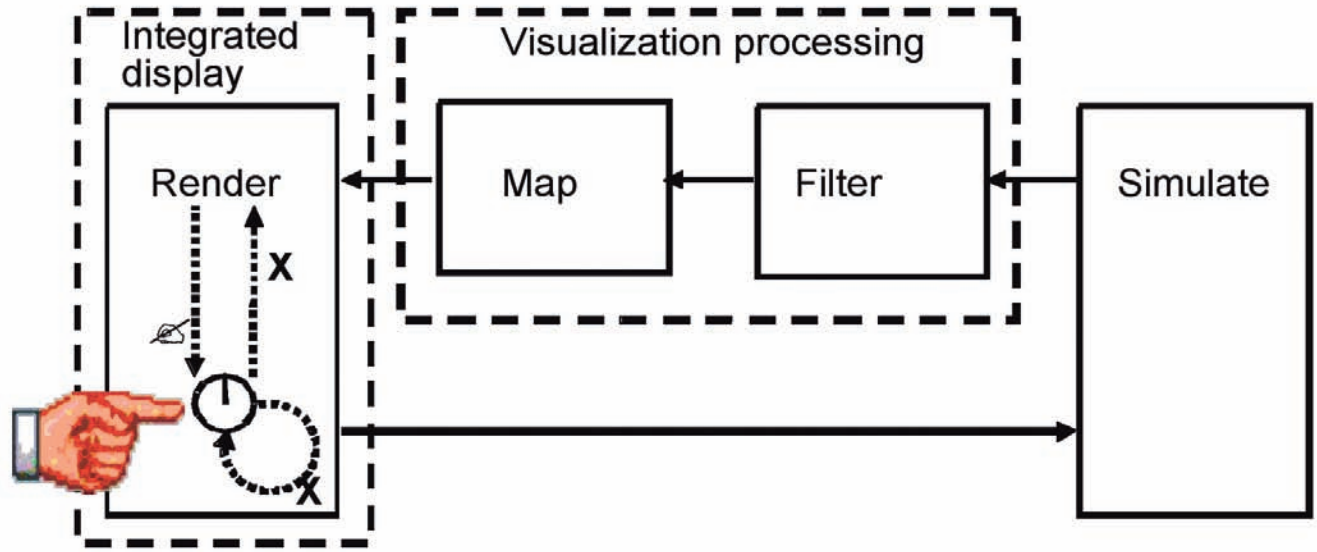
The Group operates a 17 node dual processor single core visualization server based on Linux operating system and Chromium technology for high resolution distributed visualization for real-time interactive data analysis on the grid.

The group excels in taking output from major national and international e-Science projects and embedding it in scientific applications to deliver performance, scalability, productivity and modernity for scientific data exploration. Grid enabled Tobyfit and SXD2002 are examples from ISIS of collaboration results from the group.

The Group has successfully embedded computational steering into simulations running on the Grid with control through a desktop client. The novelty is the approach is based on image based steering thereby supporting a range of available steering libraries. This provides unprecedented freedom for application developers to choose steering libraries based on knowledge and collaboration and the users to steer using parameter types, providing enhanced HCI with the simulation on the Grid.

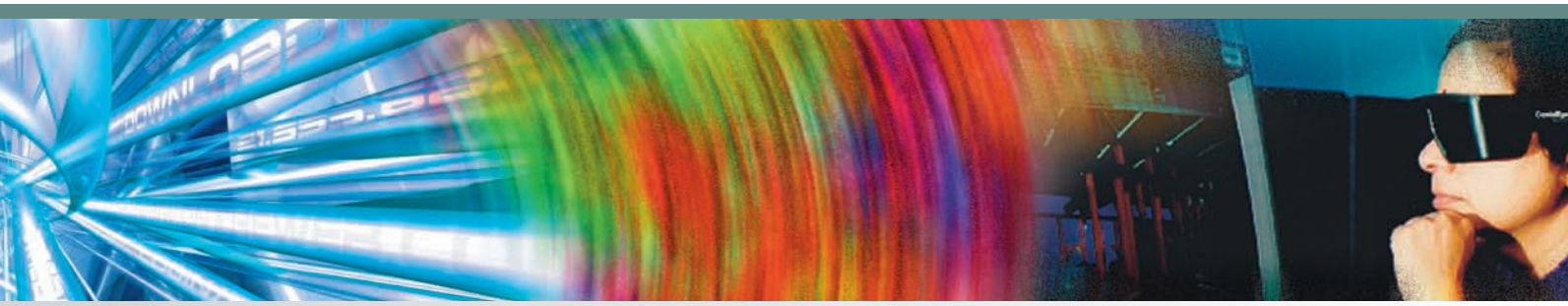
The visual feedback of the results from the simulation is streamed from the server using open source technology.

Image based steering is seamlessly included in the data analysis core toolkits on a range of projects and applications.



The process of visualising a simulation.

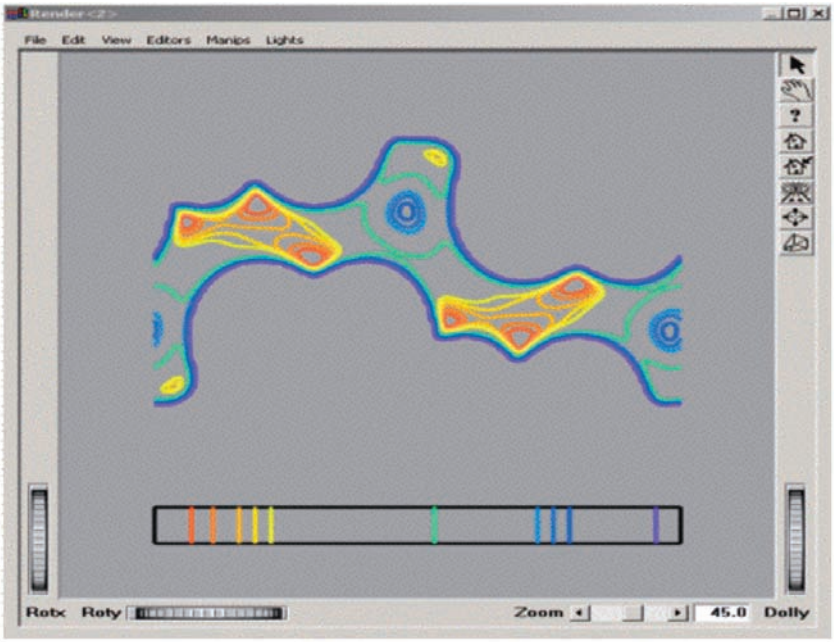
Contact: Dr Lakshmi Sastry E-mail: M.Sastry@rl.ac.uk

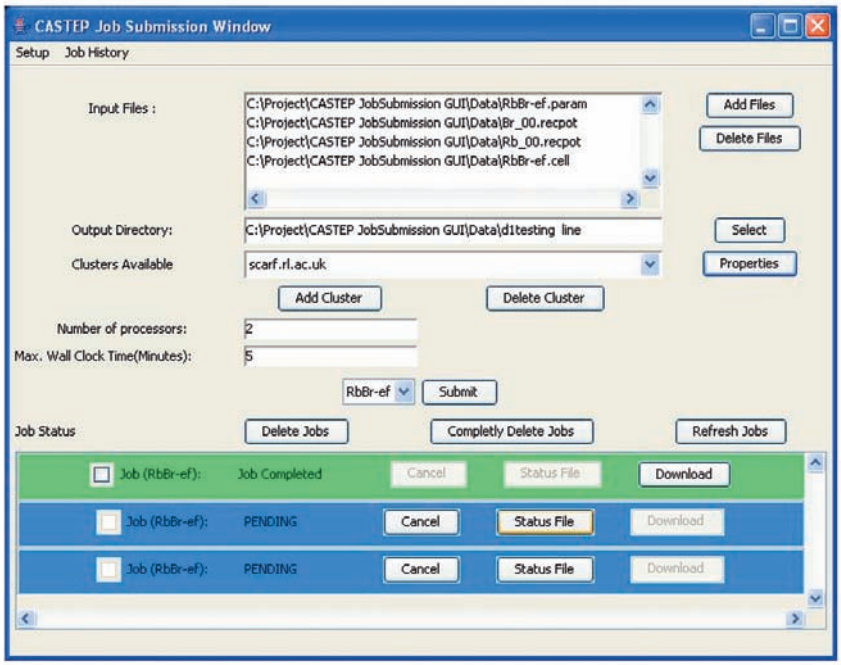


In addition, the group is building services and have embeded these successfully in a variety of toolkits to reduce data transfer latency for interactive applications on the grid.

At the elementary level, the group provides portals and services for job submission on the grid for community software such as CASTEP. An SRB interface to a variety of commercial and public domain toolkits such as Matlab.

At the R&D level we are developing advanced visualisation algorithms for higher and multi-dimensional data visualisation for very large datasets (>1GB per dataset).





A user interface for the submissions for jobs to the CASTEP molecular modelling software over the Grid.

Applications and Service Delivery

Environmental e-Science

NERC DataGrid

The NERC DataGrid (NDG) project (funded under the UK e-Science programme) is developing an infrastructure to integrate access to a range of distributed environmental data across the UK. It will make data discovery, delivery and use much easier than it is now, facilitating better use of the existing investment in the curation and maintenance of quality data archives. The project is coordinated by the British Atmospheric Data Centre (BADC), with the CCLRC e-Science Centre contributing on information modelling aspects.

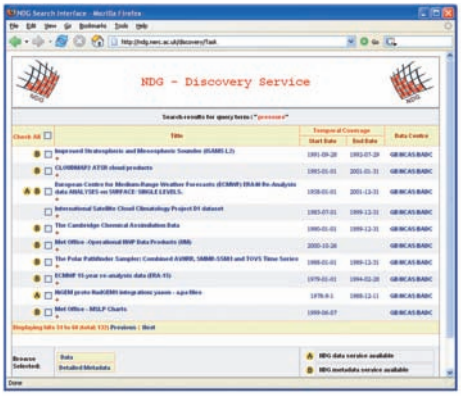
The focus in the first funding phase (2002-2005) was on data curated by the BADC and British Oceanographic Data Centre – two of the NERC designated data centres. With a successful bid for renewal funding, the second phase of NDG is engaging also with the Plymouth Marine Laboratory's Remote Sensing Data Analysis Service, and the National Oceanography Centre, Southampton.

Highlights from the past year include:

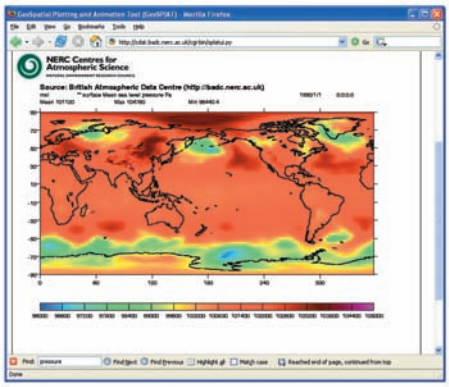
- Deployment of an NDG data discovery service (http://ndg.nerc.ac.uk/discovery/). This harvests metadata from NDG partners and various international collaborators, providing one central point of discovery for the combined data resources.

- The implementation of a novel role-based security framework. This enables data providers to define mappings between their access roles – thus allowing a BODC user (for instance) to access certain BADC data and vice versa.

- Growing international interest in the NDG data model ("Climate Science Modelling Language", CSML). This provides a conceptual view of data (rather than files and database tables) and is a state-of-the-art application of emerging ISO standards for geospatial information. An initial code suite has been developed for CSML generation and parsing.

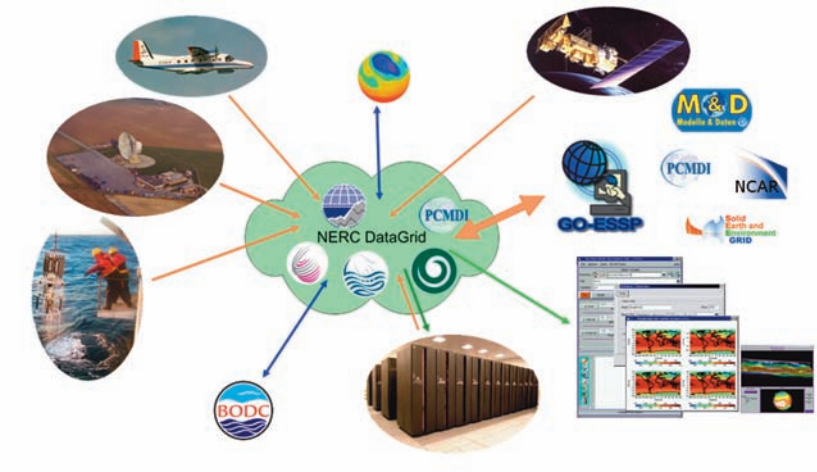


- A rich schema has been developed ("Metadata Objects for Links in Environmental Science", MOLES) for expressing relationships between data-gathering activities, generating instruments, and observation stations.



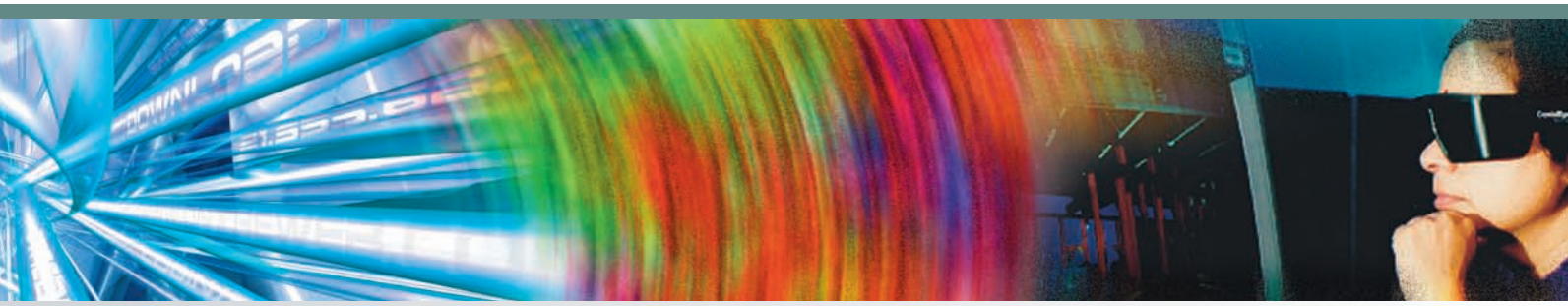
- Prototype data access and visualisation services are being trialled.

e-Science Centre expertise is also playing a major role in a new European Directive on harmonisation of environmental data (INSPIRE), by contributing to the development of Implementing Rules.



NERC DataGrid will simplify access to a wide range of environmental data

Contact: Dr Philip Couch E-mail: p.a.couch@dl.ac.uk

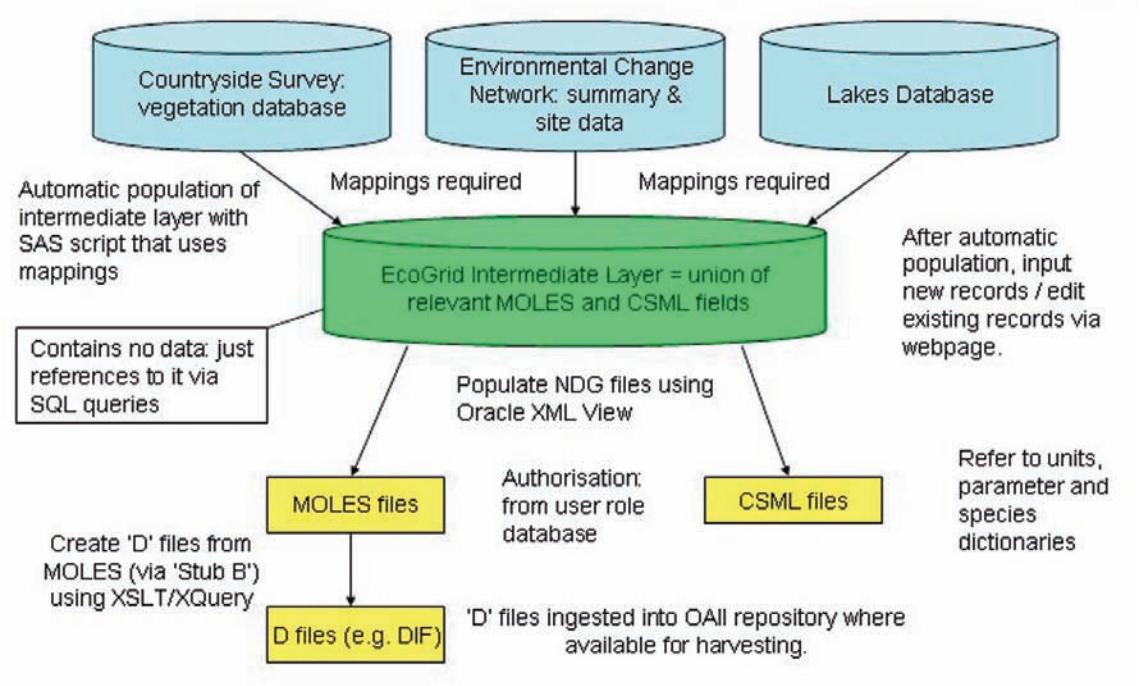


Ecological Data-Grid (Eco-Grid)

The Centre for Ecology and Hydrology (CEH) is the leading UK body for research, survey and monitoring in terrestrial and freshwater environments. CEH holds various databases collectively representing a valuable environmental research resource. However, their use inside and outside CEH is constrained by lack of data accessibility and interoperability. This project has focused on three test-bed datasets held at the Lancaster Environment Centre which form a good example of the diversity of CEH terrestrial and freshwater data. Metadata systems and access tools have been constructed in collaboration with the NERC Data Grid (NDG) to provide users with Grid services linking data discovery to dataset delivery.

To integrate the EcoGrid data into NDG requires mapping to the NDG models for metadata (MOLES) and data (CSML). This is done first by creating an 'intermediate layer' that describes all CEH data in a consistent way. This intermediate layer will resolve conflicts in the base data by concentrating on semantics. Thus, data with different names but similar meaning will map to the same intermediate element, and data with similar names but different meaning will map to different intermediate elements.

Once intermediate layer records are created for all EcoGrid datasets, these can be mapped to the NDG. Furthermore, if one wishes to create EML (Ecological Mark-up Language) records from EcoGrid in the future, this can be done relatively easily, either from the intermediate layer or the NDG records.



EcoGrid architecture

Applications and Service Delivery

The environment from the molecular level-eMinerals

The eMinerals project is a NERC funded e-Science pilot project focused on fundamental science problems associated with key environmental issues such as nuclear waste storage and pollution.

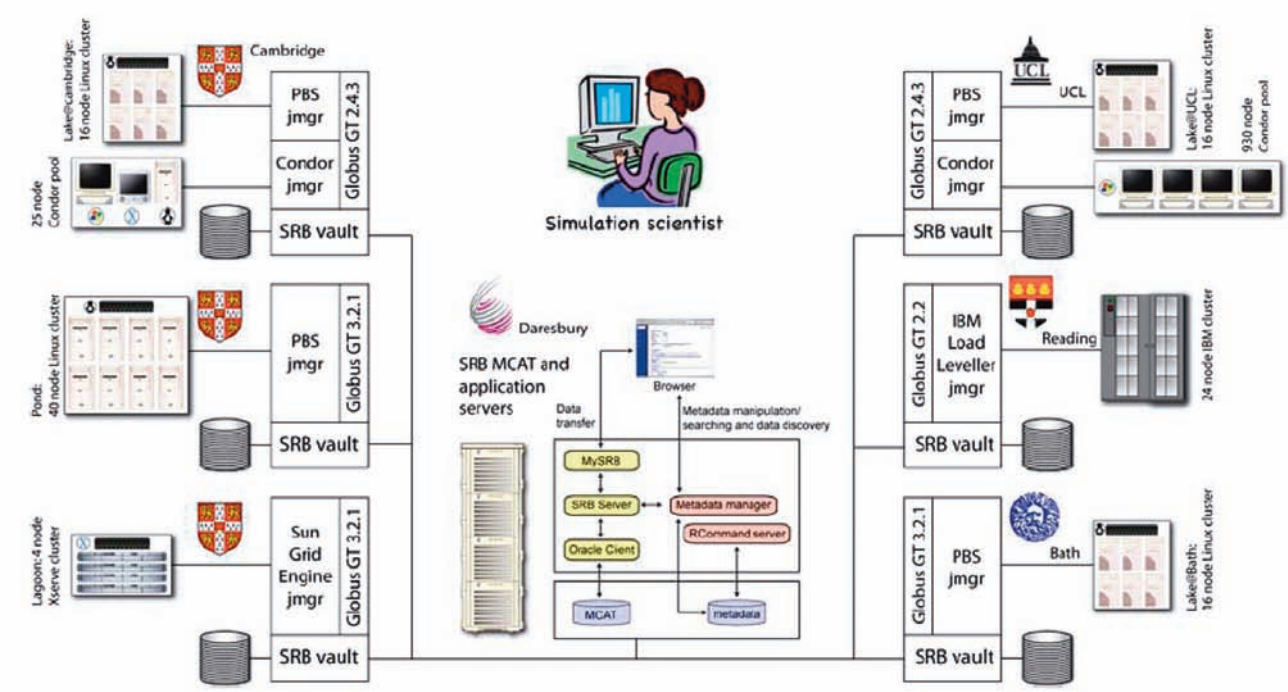
The project started in 2002, and by 2005 had a miniGrid running simulations of the adsorption of pollutants on mineral surfaces. The project has proven the benefits of a Grid to run the many calculations required by a combinatorial study in parallel and demonstrated how a usable human computer interface can make Grid resources accessible to environmental researchers.

The eMinerals project involves the collaboration of environmental scientists, scientific code developers and computer scientists from Bath University, Cambridge University, CCLRC Daresbury Laboratory, Reading University, the Royal Institute and University College London. The science requirements of the project drive both the e-Science activities and the scientific code development, creating a synergy between the three teams within the project.
The scientific research is challenging both in terms of the computational power required to tackle realistic system sizes with the required accuracy and the data management issues related to handling large amounts of data over a distributed virtual organisation.

Hence the use of Grid computing, together with associated data management technology and collaborative working environments, provide enticing opportunities to facilitate and enhance this work.

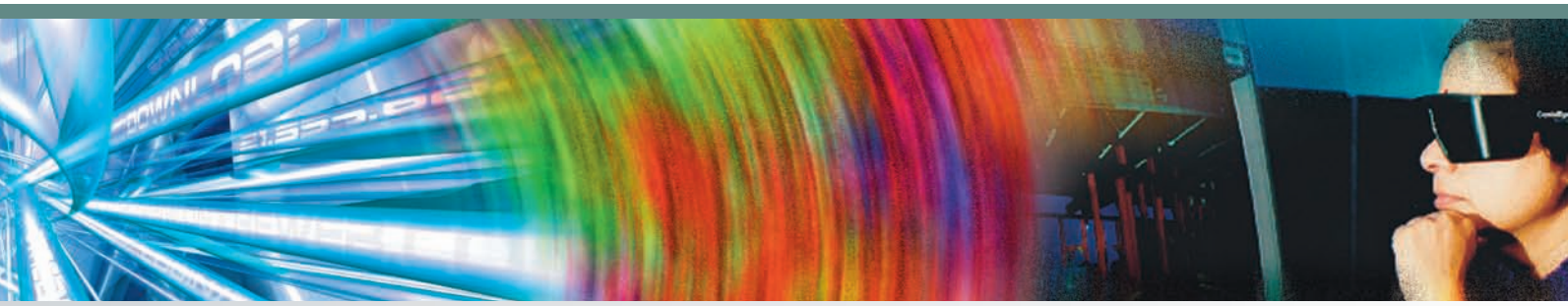
The aim of the e-Science technology is to provide an enabling infrastructure which supports, enhances and facilitates this collaborative working and the underlying science applications. With this in mind, the eMinerals miniGrid was created in close collaboration with the environmental scientists.

As the computational requirements vary significantly between both the different simulations techniques, so the compute components of the miniGrid comprise both HPC and HTPC resources. In addition to these eMinerals specific resources, use is also made of larger resources such as the NGS and HPCx.



Schematic view of the computer, data and metadata components of the eMinerals miniGrid.

Contact: Dr Andrew Woolf E-mail: a.woolf@rl.ac.uk

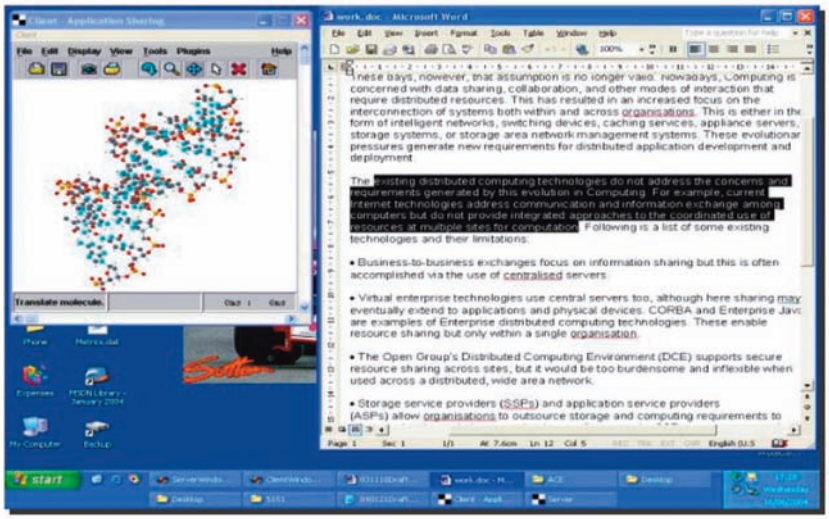


The computational simulations generate data files that are, at least for this area of science, of unprecedented size and complexity. Accordingly the project has set up a data grid and associated metadata infrastructure to facilitate distributed data management and data sharing within the virtual organization.

The data grid has been built using the Storage Resource Broker (SRB).

Each of the compute resources within miniGrid also hosts an SRB storage vault with the data grid currently providing around 4 Tb of storage capacity for the project.

As well as the data, metadata and compute tools, the project also makes heavy use of collaborative working technologies such as the Access Grid.



The Multicast Application Sharing Tool (MAST) allows the sharing of arbitrary applications within the Access Grid framework.

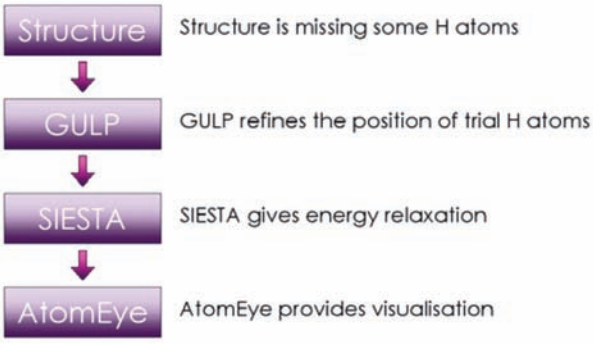
Minigrid Client Tools

The project has pursued multiple tracks with respect to client tools in an effort to provide an interface to the miniGrid that is suitable and usable by all the members of the virtual organization, who naturally have a range of IT competencies, as well as different preferred working environments. Hence there are command line tools and portal based interfaces to the miniGrid being used in production.

There is, in addition, a Business Process Execution Language (BPEL) workflow engine in development. The key to all the different client tools is to ability to enact a workflow that combines functionality from the compute, data and metadata components of the miniGrid with an interface that is simple enough to be controlled by the scientists themselves. This workflow allows the results of simulations run on the miniGrid to be automatically uploaded to the data grid with suitable annotations and entries simultaneous being entered into the metadata database.

Data Interoperability

Due to the range of systems of interest to the environmental scientists within eMinerals, the project works with a wide range of simulation codes. Accordingly, the project has made heavy use of XML and semantic technology in order to enable data to be transferred seamlessly from application to application. The use of XML based data formats together with rich metadata is essential to ensure data derived from a particular simulation is accessible to all scientists within the project.



A scenario that requires the passing of data from through a sequence of applications.

Information Exchange in Computational Chemistry

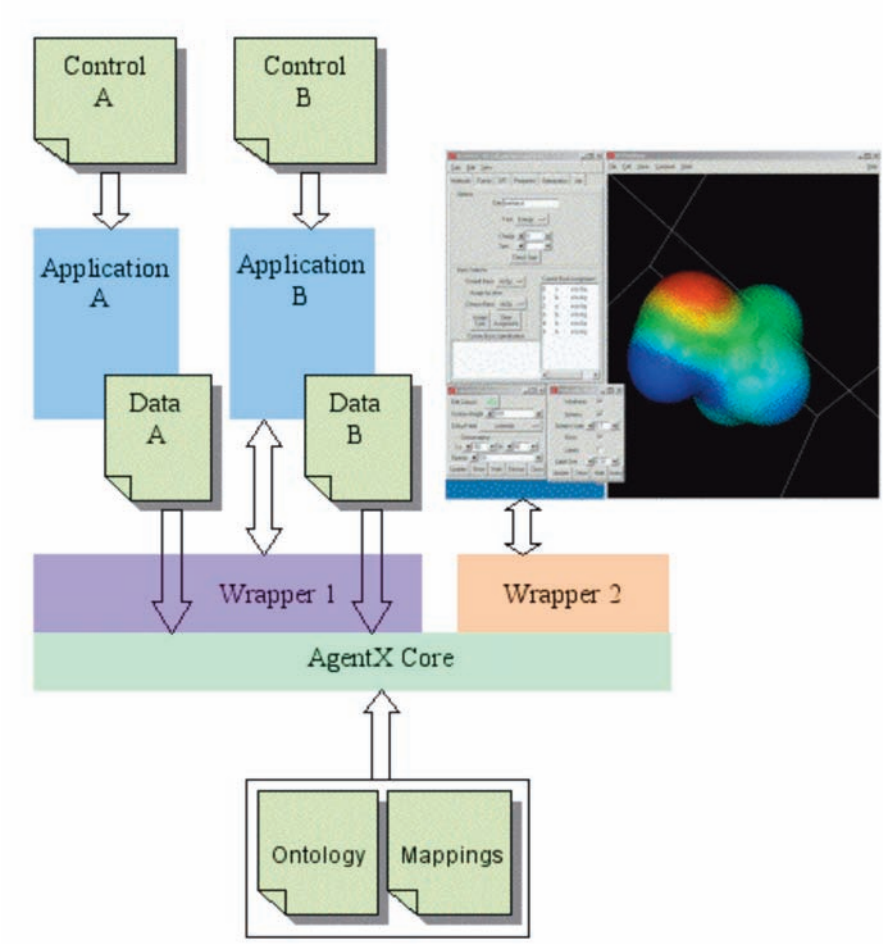
Increasingly, complex scientific problems must be addressed through the application of several computational codes used in close cooperation.

The eCCP project has been established to investigate approaches that facilitate information exchange between computational codes. An ontology based framework that addresses the key difficulties is under development - AgentX. The framework aims to allow efficient representation of the data sets while enabling complex relationships and context to be explicitly expressed.

The W3C Web Ontology Language (OWL) is used to specify terms that can be used to provide the context of data, such as Molecule, Atom and x-Coordinate. These terms are then related to data using a set of mappings. Currently, the framework is used with XML data and, where possible, representations use existing standards such as the Chemical Markup Language (CML).

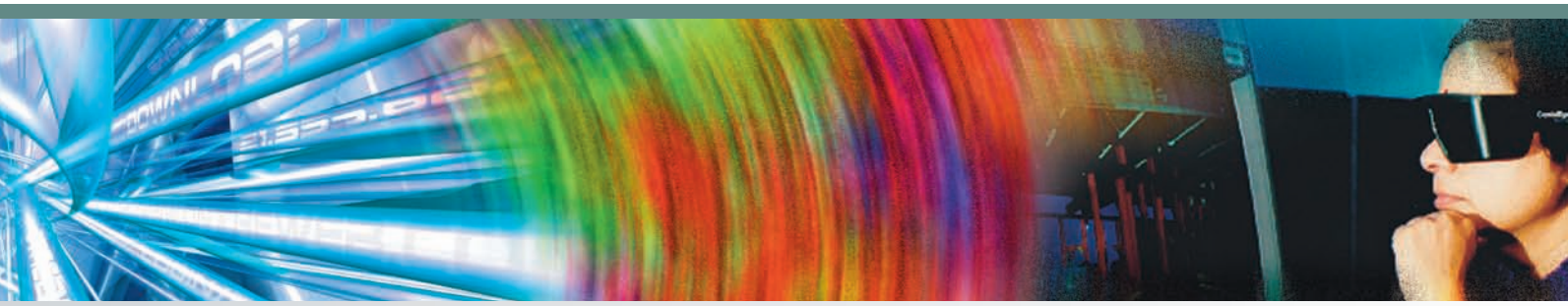
Each data format will have its own set of mappings and in this way AgentX can work with data conforming to different data models. The framework provides several methods to express complex relationships between data sets; these are based on the W3C XLink and XPointer standards.

A library is under development that can be used to query data documents and extract information based on the terms in the ontology. The core library is written in C and there are wrappers to allow access from a range of other languages: Python, Perl and Fortran. AgentX presents a simple API and has few dependencies on other software (built on top of libxml2).



The Agent-X Architecture

Contact: Dr Philip Couch E-mail: p.a.couch@dl.ac.uk



e-HTPX: A Service Oriented Architecture For Protein Crystallography

The volume of data coming from structural genome projects has generated a demand for new methods to obtain structural and functional information about biological proteins and macromolecules.

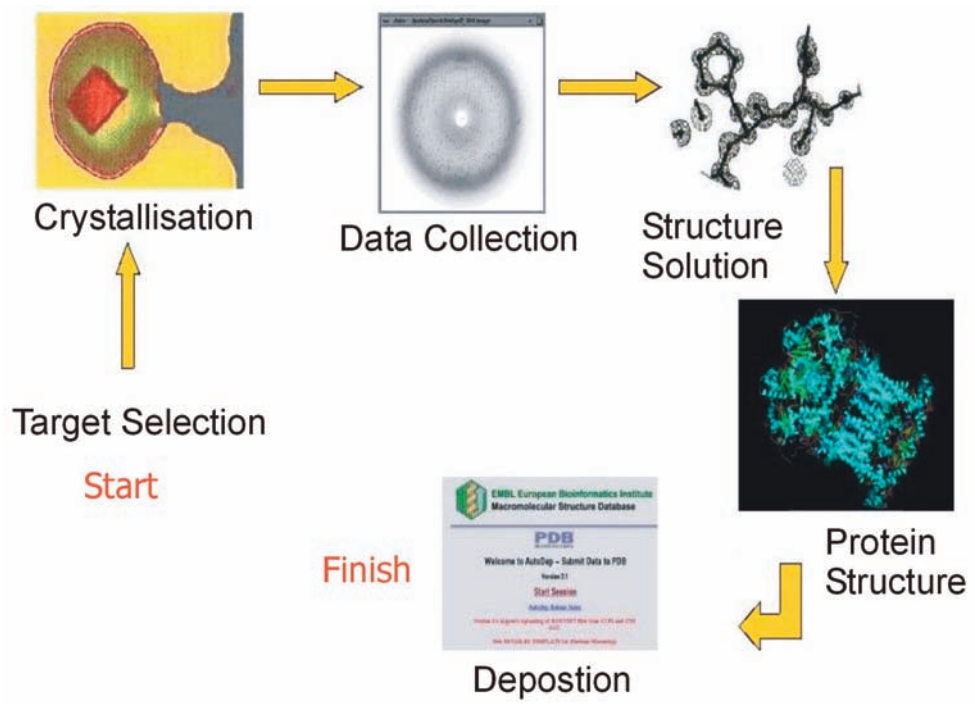
This has led to a demand for high throughputtechniques to determine the structure of important proteins. The e-HTPX project is a Service Oriented Architecture to help structural biologists remotely plan, initiate, monitor and manage collected experimental data from a synchrotron facility fromtheir home laboratories. The project covers all stages of the protein crystallography experimental pipeline, from target selection to (digital) solution of the final protein structure and deposition into public databases.The e-HTPX pipeline is now in use atseveral laboratories in the UK for experiments carried out at the ESRF (European Synchrotron Radiation Facility).

The CCLRC e-Science Centre is leading the design and implementation of the computational techniques and informatics that are required to build the e-HTPX system, and integrate a collection of remote services into a Service Oriented Architecture (SOA) for protein crystallography in the UK.

Key services for e-HTPX have been developed by leading UK e-Science, synchrotron radiation and protein manufacture laboratories, including the CCLRC e-Science Centre, Synchrotron Radiation Facility (Daresbury Laboratory), Oxford Protein Production Facility (OPPF), the European Synchrotron Radiation Facility

(ESRF), York Structural Biology Department (YSBL), CCP4 (Daresbury Laboratory) and the European Bioinformatics Institute (EBI).

In addition to providing technical contributions, the e-Science Centre and e-HTPX project are also playing an important role in both forging and enabling



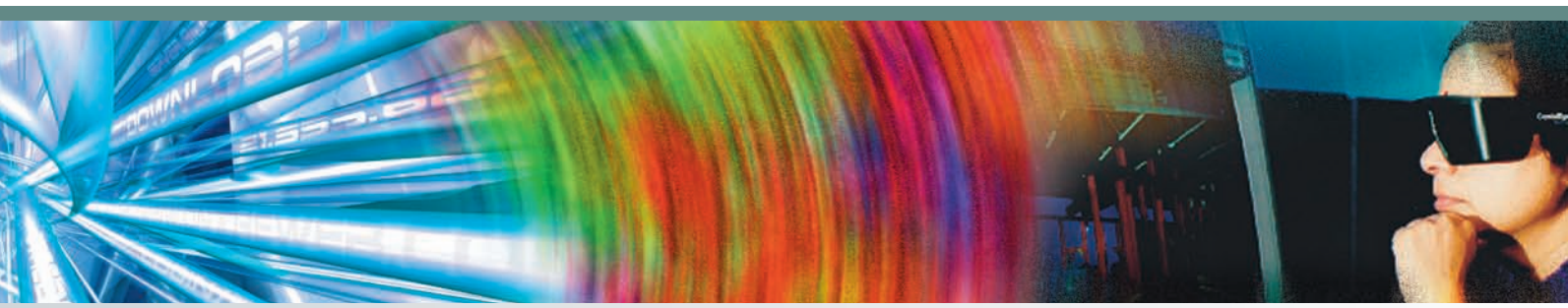
The Multicast Application Sharing Tool (MAST) allows the sharing of arbitrary applications within the Access Grid framework.

collaboration and interoperability between different academic and industrial protein-biology laboratories and synchrotron radiation facilities. This has largely been achieved through implementation of remote access e-Science technologies, especially with development of the e-HTPX XML data model. In doing this, the project has identified and documented important variations in client and experimental data requirements across different institutions. This has proved invaluable for the standardisation and production of a well constrained experimental pipeline, which is critical for high throughput operation.

Data Model

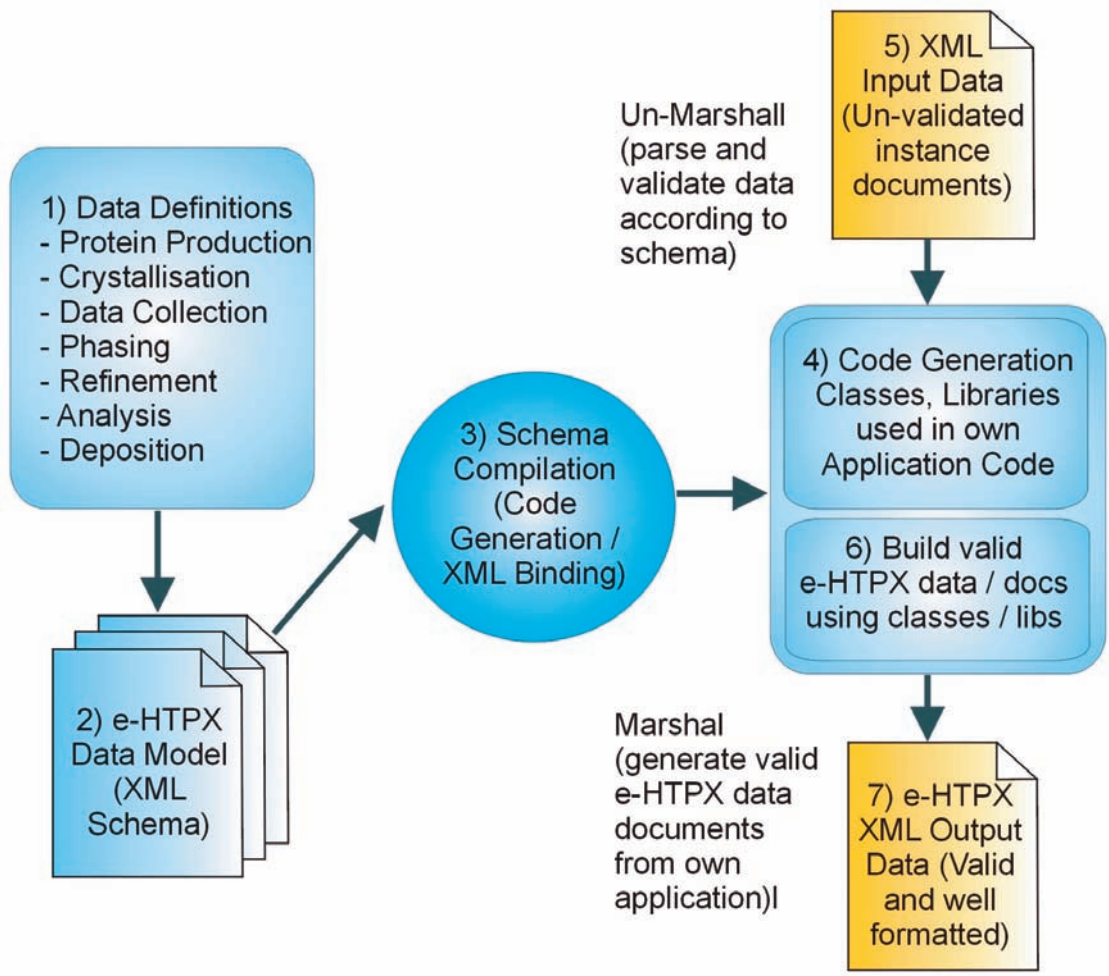
The e-HTPX data model, described and constrained in XML schema, provides an open and agreed standard for communication and data-exchange between the different partners involved in the project (clients and services). A key aim of the Data Model (also known as the Protein Production Data Model - PPDM) is to describe the data required to both reproduce the samples and experiments involved in protein production, and to inform subsequent work. The data model is particularly important for the successful implementation and future maintenance of the system. Through agreed standardisation of the data model, the e-HTPX services and data-formats are better suited for widespread adoption which encourages interoperability across multiple institutions. The e-HTPX data model allows users to automatically generate their own code libraries using standard schema compilation tools in languages of their choice. These libraries can be used to build and validate data-documents which help ensure that data is recorded in a fully annotated and well structured format. The data model describes a wide variety of data from experimental beam-line collection parameters and diffraction plans, to administrative information.

Contact: Michael Gleaves E-mail: m.t.gleaves@dl.ac.uk



Design, Implementation and Management.
The e-Science Centre is playing an important role in supporting collaborating partners in such areas as provision andimplementation of HPC (High Performance Computing) and data-processing via Grid accessible resources, provision of storageand data archiving resources (including the ATLAS Data-store) and best practices in software engineering and computation(providing technical expertise).

The e-HTPX pipeline is now in use at several laboratories in the UK when when carrying out protein crystallography experimentsat the ESRF. It will also be made available at Diamond Light Source.



e-HTPX Data Model Binding

Publications 2005

A Akram (CCLRC), R J Allan (CCLRC), O F Rana (University of Cardiff)
Virtual Peer Communities and the Community Coordinator
Proc. *1st International Conference on Semantics, Knowledge and Grid*

A Akram, D Chohan, X Dong Wang, X Yang, Rob Allan
A Service oriented architecture for portals using portlets
Proc. *Uk e-Science All Hands Conference 2005*

A Akram (CCLRC), D Chohan (CCLRC), D Meredith (CCLRC), XD Wang (CCLRC), R Allan
CCLRC Portal infrastructure to support research facilities
Global Grid Forum 14 (GGF 14), USA, 25-30 Jun 2005

A Akram (CCLRC)
Converting HelloWorld Web Service into WSRF based HelloWorld
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial1.pdf

A Akram
Destroying the WSRF Resource (Immediate Destruction)
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial5.pdf

A Akram (CCLRC)
Destroying the WSRF Resource (Scheduled Destruction)
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial7.pdf

A Akram (CCLRC), D Meredith (CCLRC), R Allan (CCLRC)
Evaluation of BPEL for Scientific Workflows
6th IEEE International Symposium on Cluster Computing and the Grid, CCGrid 2006 (CCGrid 2006), Singapore, 17-21 May 2006

A Akram (CCLRC)
Implementing WSRF based HelloWorldin Object Oriented Way
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial2.pdf

A Akram (CCLRC)
Querying the WSRF Resource
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial6.pdf

A Akram (CCLRC)
WSRF based HelloWorld with Factory Design Pattern
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial4.pdf

A Akram (CCLRC)
WSRF based HelloWorld with multiple instances
http://esc.dl.ac.uk/WOSE/tutorials/WSCore/Tutorial3.pdf

A Akram (CCLRC), R Allan (CCLRC), R Crouchley (CCLRC)
WSRP Reincarnation of Service Oriented Architecture
All Hands Meeting 2005 (AHM 2005), Nottingham, UK, 17-21 Sep 2005

Rob Allan, Dharmesh Chohan, Xiao Dong Wang, Xiaobo Yang, Rob Crouchley, Adrian Fish, et al
Virtual Research Environments: Sakai Demonstrator
Proc. *UK All Hands Conference 2005*

M Baker, H Ong, RJ Allan
Grid Services in the UK and Beyond
UK OGSA Testbed Workshop, Westminster, 22-23 Mar 2005

Bruce Beckles (University of Cambridge), Se-Chang Son (University of Wisconsin, Madison), John Kewley (CCLRC)
Current methods for negotiating firewalls for the Condor system
Proc. *4th UK e-Science All Hands Meeting (AHM2005), Nottingham, UK*, in Proceedings of the UK e-Science All Hands Meeting 2005, eds. Simon Cox, David W Walker, ISBN 1-904425-53-4 (EPSRC) (2005)

N D Bennett (CCLRC), et al (NERC DataGrid)
NERC Data Grid Authorisation Architecture
Proc. *UK e-Science All Hands Meeting 2005 (AHM05)*

N Bennett (CCLRC), R Cramer (BODC), S Kondapalli (BODC), R Lowry (BODC), G Drinkwater (CCLRC), M Gutierrez (CCLRC), et al (NERC DataGrid)
NERC Data Grid Authorisation Architecture
Proc. *UK e-Science All Hands Meeting 2005 (AHM 2005), Nottingham, UK, 19-22 Sep 2005*

N Bennett (CCLRC), R Scott (CEH), M Brown (CEH), M Lane (CEH), K Kleese-van Dam (CCLRC), K O'Neill (CCLRC), et al (NERC DataGrid)
NERC Ecological Data Grid
Proc. *UK e-Science All Hands Meeting 2005 (AHM 05)*

LB Casely-Hayford (CCLRC)
Environments, Methodologies and Languages for Supporting Users in building a Chemical Ontology
MSc Bioinformatics/Computational Biology Thesis

D Chohan (CCLRC), A Akram (CCLRC), R Allan (CCLRC)
Building of a Grid Portal Infrastructure
3rd International Workshop on Middleware for Grid Computing (MGC 2005), Grenoble, France , 28 Nov 2005 - 29 Nov 2006

D Corney (CCLRC), P Berrisford (CCLRC), T Folkes (CCLRC), J Jensen (CCLRC), J Mencak (CCLRC), D Ross (CCLRC), et al
Experiences with Supporting Mass Storage for Multiple Communities
e-Science All Hands Meeting,, Nottingham, UK, 19-23 Sep 2005

P Couch (CCLRC), P Sherwood (CCLRC), S A Sufi (CCLRC), I Todorov (Cambridge University), R Allan (CCLRC), P Knowles (Cardiff University), et al
Towards Data Integration for Computational Chemistry

S J Cox (CSIRO Exploration and Mining), A Woolf (CCLRC), L Wyborn (Geoscience Australia), R Woodcock (CSIRO Exploration and Mining), R Atkinson (Social Change Online), J Esterle (CSIRO Exploration and Mining)
Harmonisation of Grid and Geospatial Services Standards in the Earth and Environmental Sciences
Proc. *2005 Joint Assembly of the American Geophysical Union, New Orleans, 23-27 May 2005, Eos Trans. AGU, Joint Assembly Supplement, 86(18) (2005)*

Rob Crouchley (University of Lancaster), Ties van Ark (University of Lancaster), John Pritchard (University of Lancaster), John Kewley (CCLRC), Rob Allan (CCLRC), Mark Hayes (University of Cambridge), et al
Putting Social Science Applications on the Grid
in Proceedings of the First International Conference on e-Social Science (National Centre for e-Social Science) (2005)
R Crouchley (University of Lancaster), A Fish (University of Lancaster), T van Ark (University of Lancaster), R J Allan (CCLRC), D Chohan (CCLRC), X D Wang (CCLRC), et al
Virtual Research Environments and e-Social Science
1st Int. Conf. on e-Social Science

G Drinkwater (CCLRC), S A Sufi (CCLRC)
Data Mining using the Data Portal for the NESSI Project
Proc. *The 2005 International Conference on Grid Computing and Applications (GCA'05), Las Vegas, Nevada, USA*

M Hayes (University of Cambridge), L Morris (University of Cambridge), R Crouchley (University of Lancaster), D Grose (University of Lancaster), T van Ark (University of Lancaster), R Allan (CCLRC), et al
GROWL: A lightweight grid services toolkit and applications
Proc. *4th UK e-Science All Hands Meeting (AHM2005), Nottingham, UK,* in Proceedings of the UK e-Science All Hands Meeting 2005, eds. Simon Cox, David W Walker, ISBN 1-904425-53-4 (EPSRC) (2005)

L Huang (University of Cardiff), A Akram (CCLRC), D W Walker (University of Cardiff), R J Allan (CCLRC), O F Rana (University of Cardiff), Y Huang (University of Cardiff)
A Workflow Portal supporting Multi-language Interoperation and Optimisation
SuperComputing 2005 Grid Computing Environments Workshop: Grid Computing Environments (GCE05), Seattle, November 2005, March 2006

L Huang (Cardiff University), A Akram (CCLRC), D Walker (Cardiff University), R J Allan (CCLRC), O Rana (Cardiff University), Y Haung (Cardiff University)
A workflow portal supporting multi-language interoperation and optimisation
Concurrency and Computation: Practice and Experience (2005)

T Kirkham (CCLRC), D Mac Randal (CCLRC), J Gallop (CCLRC), B Ritchie (CCLRC)
Akogrimo:a work in progress on the delivery of a next generation grid.
Proc. *SOAS 2005*

T Kirkham (CCLRC), D Mac Randall (CCLRC), B Ritchie (CCLRC), J Gallop (CCLRC), P Ritrovato (CRMPA), G Cirillo (CRMPA)
An Akogrimo approach to securing virtual organizations within mobile GRID computing environments
ERCIM NEWS (2005)

T Kirkham (CCLRC UWB), T Varsamidis (UWB)
Introducing BDIFS
Proc. *EEE 06*

K Millard (HR Wallingford Ltd), A Woolf (CCLRC), G Ross (Met Office), F van der Wel (Royal Netherlands Meteorological Institute), R Longhorn (IDG Ltd)
European developments in GIS standards for met/ocean data
Proc. *11th EC GI & GIS Workshop, Alghero, Sardinia, 29 Jun - 01 Jul 2005*

Keiran Millard, Rob Atkinson, Andrew Woolf, Roy Lowry, Pieter Haaring, Francisco Hernandez, et al
Using XML Technology for Marine Data Exchange
Proc. *4th EuroGOOS Conference, Brest, France, 06-09 Jun 2005*

J.Y. Nief (ccin2p3), W. Kroeger (SLAC), A. Hasan (CCLRC) (BABAR Computing Group)
BaBar Data Distribution using the Storage Resource Broker
Computing in High Energy Physics meeting

Publications 2005

A S Shaon (CCLRC)
Long-term Metadata Management & Quality Assurance in Digital Curation
MSc, University of Reading (2005)

B Strong (CCLRC), D R Corney (CCLRC), P Berrisford (CCLRC), T Folkes (CCLRC), C Moreton-Smith (CCLRC), K Kleese van Dam (CCLRC)
Key Lessons in the Efficient Archive of Small Files to the CCLRC MSS Using SRB
Symposium on "Data Interoperability - Challenges and Technologies", Sardinia

S A Sufi (CCLRC)
DynamicWrappers: a rule based metadata conversion and integration system for e-Science applications
Proc. *Fourth All Hands Meeting, 19 - 22nd September 2005 Nottingham*

J F Wheeler (CCLRC)
Introduction to XFS filesystem
UK HEP System Managers Meeting (HEPSYSMAN), Oxford, England

A Woolf (CCLRC), B Lawrence (British Atmospheric Data Centre), R Lowry (British Oceanographic Data Centre), K Kleese van Dam (CCLRC), R Cramer (British Oceanographic Data Centre),

M Gutierrez (British Atmospheric Data Centre), et al
Climate Science Modelling Language: Standards-based markup for metocean data
Proc. *The 85th Meeting of the American Meteorological Society: Building the Earth Information System (AMS 2005), San Diego, USA, 09-13 Jan 2005*

A Woolf (CCLRC), B Lawrence (British Atmospheric Data Centre), R Lowry (British Oceanographic Data Centre), K Kleese van Dam (CCLRC), R Cramer (British Oceanographic Data Centre), M Gutierrez (British Atmospheric Data Centre), et al
Data integration with the Climate Science Modelling Language
Proc. *European Geosciences Union General Assembly 2005, Vienna, Australia, 24-29 Apr 2005, Geohysical Research Abstracts, Volume 7, 08775, 2005 (2005)*

A Woolf (CCLRC), K Millard (HR Wallingford Ltd), G Ross (MetOffice), F van der Wel (Royal Netherlands Meteorological Institute)
European developments in GIS standards for met/ocean data
Proc. *European Geosciences Union, General Assembly 2005, Vienna, Austria, 24-29 Apr 2005, Geohysical Research Abstracts, Volume 7, 08919, 2005 (2005)*

A Woolf (CCLRC), R Cramer (British Oceanographic Data Centre), M Gutierrez (British Atmospheric Data Centre), K Kleese van Dam (CCLRC), S Kondapalli (British Oceanographic Data Centre), S Latham (British Atmospheric Data Centre), et al
Semantic Integration of File-based Data for Grid Services
Proc. *Semantic Infrastructure for Grid Computing Applications Workshop, Cluster Computing and Grid Symposium (SIGAW, CCGrid 2005), Cardiff, UK, 09-12 May 2005*

X Yang (CCLRC), D Chohan, X D Wang (CCLRC), R Allan (CCLRC)
A Web Portal for the National Grid Service
Proc. *UK e-Science All Hands Meeting 2005 (AHM2005), Nottingham, UK, 19-22 Sep 2005*

X Yang (CCLRC), X D Wang (CCLRC), R Allan (CCLRC)
WSRP Support Investigation of Selected Open-Source Portal Frameworks
Poster at GCE 05: Grid Portals Workshop, held in conjunction with SCl05, 18 November 2005, Seattle, USA

Further Information

Laura Johnston	DL e-Science Secretary	dlescsec@dl.ac.uk
Sue Griffiths	Administrative Assistant	s.griffiths@rl.ac.uk
Denise Small	Personal Assistant	d.small@rl.ac.uk
Andrew Kaye	Business Manager, e-Science Centre	a.kaye@rl.ac.uk
John Gordon	Deputy Director, e-Science Centre	j.c.gordon@rl.ac.uk
Neil Geddes	Director, e-Science Centre	n.i.geddes@rl.ac.uk

Group Leaders

Andrew Sansum	Tier 1	a.sansum@rl.ac.uk
Lakshmi Sastry	Applications and Visualisation	l.sastry@rl.ac.uk
Andrew Richards	Grid Operations	a.j.richards@rl.ac.uk
Peter Oliver	HPC Services	p.m.oliver@rl.ac.uk
Kerstin Kleese-van-Dam	Data Management	k.kleese@dl.ac.uk
David Corney	Petabyte Storage	d.corney@rl.ac.uk
Rob Allan	Grid Technology	r.j.allan@dl.ac.uk

Web sites

http://www.cclrc.ac.uk	CCLRC
http://www.rcuk.ac.uk/escience	UK e-Science Programme
http://www.e-science.cclrc.ac.uk	e-Science Centre

Addresses

CCLRC Rutherford Appleton Laboratory
Chilton
Didtot
Oxon OX11 0QX
Tel: 01235 445000
Fax: 01235 445808

CCLRC Daresbury Laboratory
Keckwick Lane
Daresbury
Warrington WA4 4AD
Tel: 01925 603000
Fax: 01925 603100

eScience