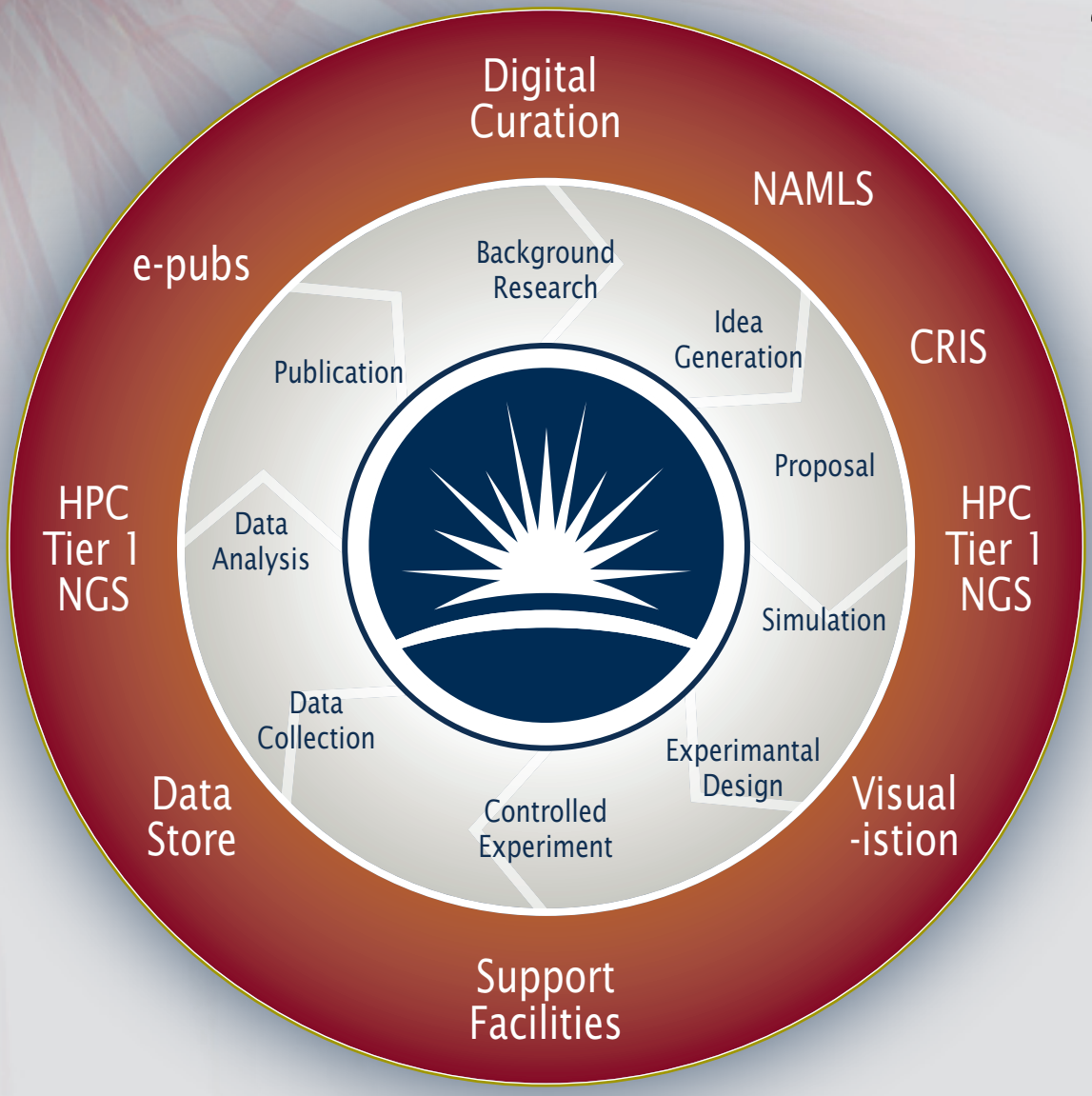
Science & Technology Facilities Council

E-SCIENCE SERVICES 2008



The STFC e-science centre provides a range of services to the UK research community. Some of these are available for individuals to apply for and use, while others are available to organisations.

A brief description of the available services is presented here for you to identify those which meet your needs.

National Grid Service (NGS)

The NGS is the UK's national production level Grid for e-Science. The NGS provides compute and data resources for UK academics 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.

Individual applications to use the service can be made on-line at:
http://www.ngs.ac.uk/apply.html

Atlas Petabyte Storage Service

The digital data store provides 5 Petabytes of on-line storage accessible through a Storage Request Broker (SRB) and other grid interfaces.

The data store is available to universities, consortia, and research laboratories through medium term contracts.

Applications from organisations should be made to:
David Corney
E-mail: D.R.Corney@rl.ac.uk
Tel: **01235 445993**

Tier 1 Computing Service

The Tier 1 computing service provides computing and data storage facilities to the UK particle physics research community for LHC and other experiments.

The service is centrally funded by PPARC and CCLRC to be free at the time of use.

Applications for use of the Tier 1 Computing Service should be made through the GridPP collaboration:
http://www.gridpp.ac.uk/tier1a/

HPC Support Service

The HPC Services Group can house compute clusters and other large computing resources in a secure, controlled environment in purpose built machine rooms.

These resources will be professionally managed according to an agreed service level definition. We anticipate that these resources will be made available using grid technologies allowing interoperability with local, organisational and national resources.

Resources can be hosted by the HPC services Group on behalf of universities, collaborations and research councils.

Applications for the HPC Services should be made to:
Dr Peter Oliver
P.M.Oliver@rl.ac.uk
Tel: **01235 445164**

National Academic Mailing List Service - JISCmail

JISCmail provides a mailing list service to support collaborative activities of UK academics. Users may join existing mailing lists or establish new ones.

JISCmail is operated on behalf of UKERNA to be free at the time of use to UK academics.

Access to JISCmail is available at:
http://www.jiscmail.ac.uk/

Application Development

Expertise and support is available to improve the performance and scalability of scientific applications and include advanced higher dimensional visualisation and data analysis.

To apply for collaborative activities requiring application development apply to:

Dr Lakshmi Sastry
M.Sastry@rl.ac.uk
Tel: **01235 446892**

Scientific Database Service

Relational database design and support services are available.

Contact:
Gordon Brown
g.d.brown@rl.ac.uk
Tel: **01235 778028**

Information Management Support

The service provides expertise in the management of data and publications to maximise their impact through the scientific lifecycle.

The service includes the provision of access through portals to the archives of data from STFC's facilities at ISIS and Diamond:
https://kisumu.esc.rl.ac.uk:8181/dp/faces/logon.jsp

The service also includes access to a repository of academic publications by STFC staff:
http://epubs.cclrc.ac.uk/index

To apply for collaborative activities requiring information management development apply to:
Dr Brian Matthews
b.m.matthews@rl.ac.uk
Tel: **01235 446648**

Data Curation

We are building on the Atlas Petabyte Storage Service and the Data Curation Centre to develop a curation service for scientific data in collaboration with partners.

Contact:
Dr Juan Bicarregui
J.C.Bicarregui@rl.ac.uk
Tel: **01235 445710**

For more information contact

e-Science Centre
STFC Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
Oxfordshire OX11 1QX
UK

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

Science & Technology Facilities Council

e-SCIENCE
FOR STFC FACILITIES

The STFC is one of Europe's largest multidisciplinary research organisations supporting scientists and engineers world-wide. It operates world-class large-scale research facilities, provides strategic advice to the government on their development and manages international research projects in support of a broad cross-section of the UK research community.

This article describes the current work to integrate e-Science capabilities into the large-scale research facilities operated by the STFC.

The Facilities

The STFC has started a work program to integrate key e-Science services within the following large scale research facilities:

ISIS is the world's leading pulsed neutron & muon source situated at the STFC Rutherford Appleton Laboratory (RAL). It supports an international community of around 1600 scientists who use neutrons and muons for research in physics, chemistry, materials science, geology, engineering and biology.

Diamond Light Source (DLS) is a new third generation synchrotron currently being built at the Harwell site in Oxfordshire. Synchrotrons produce ultra-bright light which is made up of a combination of ultraviolet light and x-rays. This is used to investigate the internal structure of materials, having a number of different applications including: biomedical science, medical research, environmental science, agriculture, mineral exploration, material science and forensics.

Central Laser Facility (CLF) work involves development and application of advanced, world-leading laser facilities. These are used for ultra-fast time-resolved infrared spectroscopy and to provide high repetition rate ultra-high intensity laser irradiation (petawatt). The facilities enable a broad base of rapidly expanding areas of research to be investigated with applications in inertial confinement fusion, astrophysics simulations, bioscience, accelerator science and molecular physics.

Aims of the Project

The e-Science Centre was founded with the distinct aim to 'Grid enable' the facilities. This was defined as developing, deploying and running e-Services for experimental, computing and data facilities, to further the Science carried out at these facilities. E-Science is aiming to achieve this goal by:

- Working collaboratively with Facilities and their users
- Creating a powerful, long lasting scientific knowledge resource for UK academia.
- Enabling users to get rapid access to their current and past data, related experiments, publications etc., leading to improved analysis through more complete information.
- Providing an integrated Infrastructure for Data Management and Advanced Analysis

ISIS and DLS Facilities at RAL

Joint Achievements

In order to produce a Facilities infrastructures on common core components agreement with the facilities have been reached on:-

- Production of CSMD a common metadata format.
- ICAT will be used as metadata catalogue
- NeXus data format will be used for all future data
- Using the same underlying Proposal system
- Single Sign On between all facilities and services - one user id for all
- SRB and Petabyte DataStore for data location management and long-term storage.
- Using SCARF for advanced Data Analysis
- Oracle databases for operational use, largely maintained and operated by the Scientific Database Group

Supporting Services & Tools

The functionalities of the e-Infrastructure are supported by a variety of Services within the STFC e-Science Centre, these are at present: Scientific Database Services, SRB Services, Petabyte DataStore, SCARF, National Grid Service, North West Grid, UK Certificate Authority.

The facilities work is also based on software tools developed within the e-Science centre e.g.

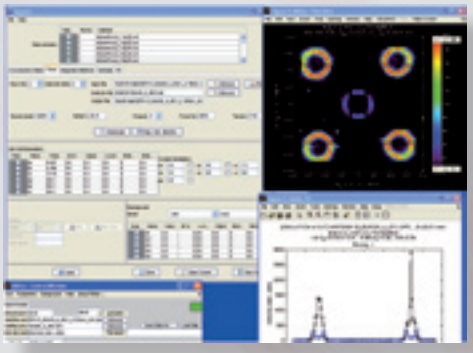
- AgentX, My Condor Submit (MCS), Remote My Condor Submit (RMCS), Information Catalogue (ICAT) and Data Portal

The Future for e-Science Facilities.

Globally, we are currently going through a golden age for big science. There are increasing numbers of large scientific projects like Diamond, ISIS and CLF being built. Also, there are also other new facilities on the horizon for example 4GLS; the UK Muon and Neutrino factory; and European Square Kilometres Array.

e-Science has an important role in providing computing infrastructure that allows the scientist to use the facility more efficiently. By providing new ways to store, manage and analyse the incredible amounts of data that these these facilities will create. e-Science will allow the scientist to concentrate on doing better science.

The e-Science Facilities projects are at a development stage and initial versions of the e-Science infrastructures are currently being deployed at DLS, ISIS and CLF for feedback from the scientific community.



Interactive interface for Toby Fit

ISIS back catalogue

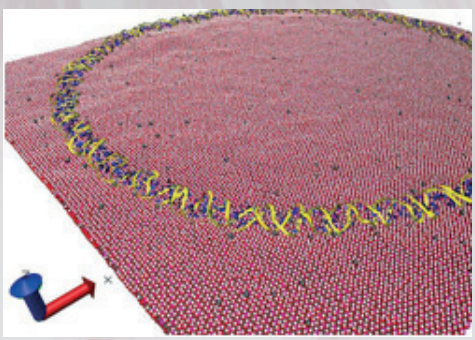
Data Portal for DLS

Science & Technology Facilities Council

RESEARCH HIGHLIGHTS 2008

The STFC e-science centre exists to bring the rapid advances in computer science and information technology to bear on the major challenges of science and engineering and thereby support the range of the STFC science programme.

A sample of some of the science resulting from e-science provides insight into the broad range of research applications.

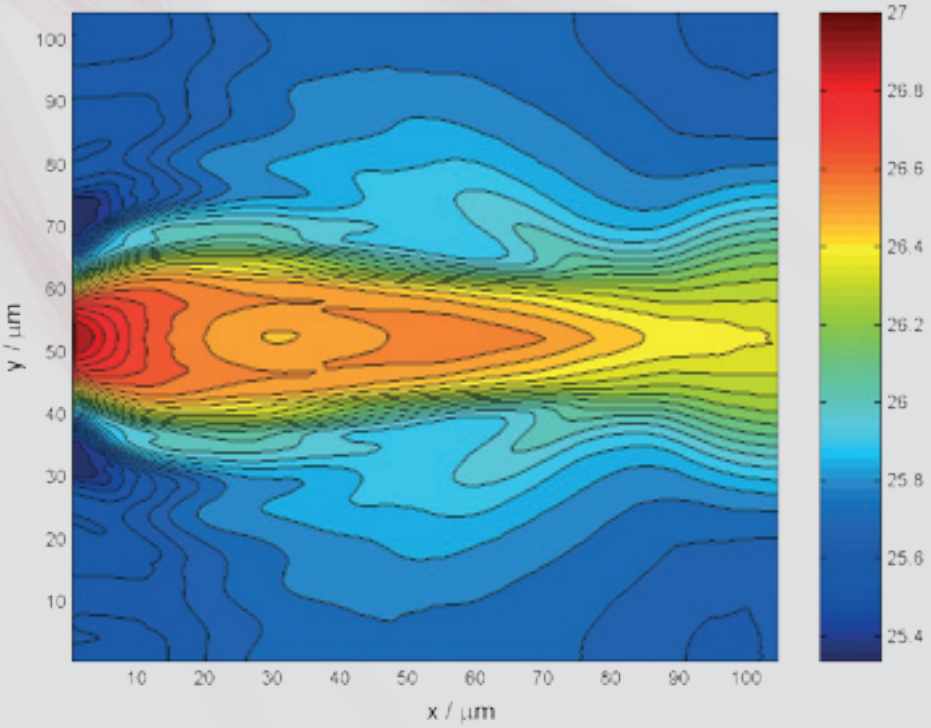


Growing DNA molecules in layered clays – is this how life evolved ?

Professor Peter Coveney and his colleagues at UCL published two papers on research done using the National Grid Service (NGS) in the Journal of the American Chemical Society in 2008. Using Teragrid and EGEE as well as the NGS Prof Coveney and his colleagues have showed how DNA plasmids may remain stable within clay sheets around a hydrothermal vent. This research has important implications for the origin of life on Earth.

Deep ocean hydrothermal vents have long been suggested as possible sources of biological molecules, such as RNA and DNA, but it was unclear how they could survive the high temperatures and pressures that occur round these vents. The team have used computer simulation to provide insight into the structure and stability of DNA while inserted into layered minerals.

The still image shown from a molecular dynamics simulation of a 480 base pair DNA plasmid sandwiched between two sheets of a clay mineral by Mary-Ann Thyveetil has also won a Wellcome Image Award 2008.



Moving closer to economic fusion energy

A problem with using nuclear fusion to efficiently generate power is to ensure that more energy is generated than was put into the reaction. Electron beams can be used to ignite a fusion reaction in very dense deuterium, but the energy used to generate the electron beam has to be sufficiently low to make the approach worthwhile. Such electron beams can be generated by hitting a metal target with a powerful laser beam.

Drs A.P.L.Robinson & M.Sherlock, with Prof.P.A.Norreys of the plasma physics group at the STFC Central Laser Facility (CLF) have been using the SCARF-Lexicon computer cluster hosted by e-Science to simulate how to magnetically focus relativistic electron beams so that they don't spread out. A poorly focused electron beam requires more than 50kJ for ignition, whereas their simulations in 2007 showed a technique to reduce the energy required down to less than 20kJ.

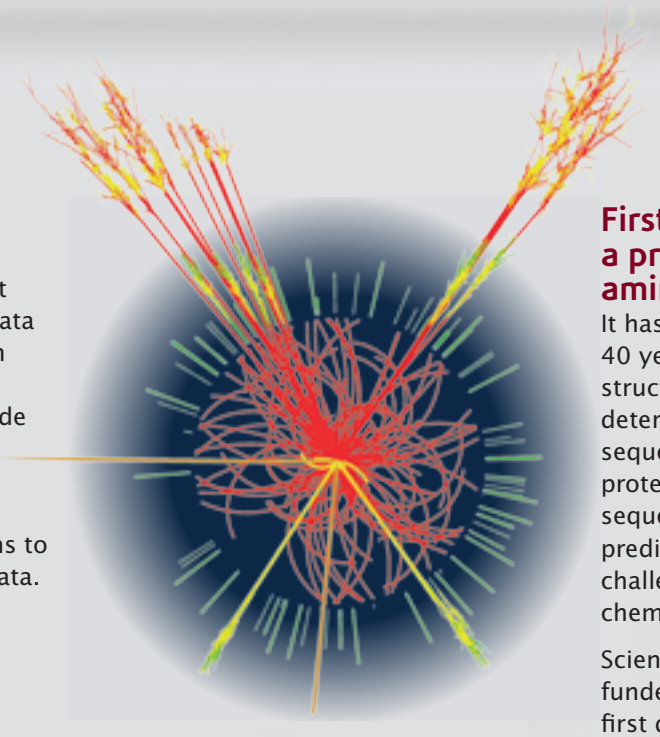
The simulated technique is expected to be experimentally tested at the CLF during 2008, and if it can be successfully demonstrated there, it might be adopted in large fusion projects such as the European HiPER project (it will also have an impact on other large projects in the US and Japan), to move a step closer to producing fusion energy efficiently enough to be economically viable.

LHC computing grid sets records

The Large Hadron Collider (LHC) at CERN is nearly ready to produce data from particle collisions detected in its four experiments: Atlas, Alice, CMS, LHCb. However, the worldwide LHC computing grid that has been developed to analyse the data has already been operating for several years, running collision simulations to prepare for the interpretation of data.

In these simulations, some initial conditions are transformed by a physics generator to produce descriptions of the energy and momentum of the hundreds or even thousands of particles that could arise from a collision in one of the four experiments. These descriptions are then fed into a detector simulator which will simulate the results that the detector will identify in such a situation (e.g. trajectories of particular particles, changes of energy in calorimeters etc...). These detector results are then fed into an experiment reconstruction programme which will simulate the initial conditions that gave rise to these experimental observations. By comparing these simulated initial conditions with the actual initial conditions provided at the start of the simulation, the error and efficiency of the detectors and analysis can be calculated, allowing clearer interpretation of real results.

The 2008 combined computing readiness challenge included sending simulated data from the four experiments at CERN to the 11 Tier 1 sites around the world, of which RAL is the Tier 1 site for the UK, Ireland, Estonia and Finland. All experiments exceeded the required data transfer rates of 1.3 GB/s for extended periods, and simultaneously for the four LHC experiments. A rate of 2.7GB/s was sustained on several days. All Tier 1s achieved (or exceeded) their target acceptance rates.



Simulation of a Higgs event as seen by the Atlas Experiment

RAL has provided the 3,139 CPUs and the 1,920 TB of disk space that it has pledged to the LHC computing grid by 2008. The project as a whole has available 92% of the 53,413 CPUs, and 76% of the 25,770 TB of disk space pledged by supporting countries by 2008.

By May 2008 the global LHC computing grid was running 340,000 computing jobs per day, 4% of which were run at RAL. In the six months to May '08 the RAL Tier 1 achieved the required reliability targets, operating at between 91% and 98% availability.

For more information contact

e-Science Centre
STFC Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
Oxfordshire OX11 1QX
UK

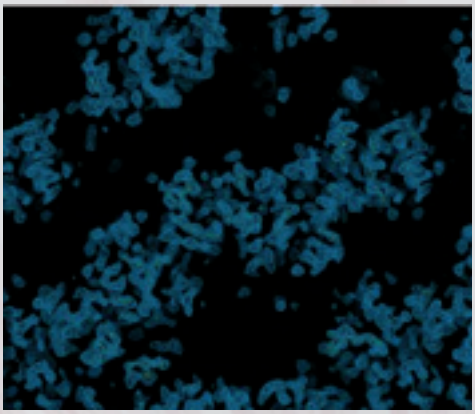
T: **+44 (0)1235 446084**
F: **+44 (0)1235 445945**

First validated prediction of a protein structure from an amino acid sequence

It has been known for more than 40 years that the three-dimensional structures of proteins are completely determined by their amino acid sequences. The prediction of protein structure from amino acid sequence—the "de novo" structure prediction problem—is a long-standing challenge in computational biology and chemistry.

Scientists from Newcastle University, funded by BBSRC, have solved the first de novo crystal structure of a protein using diffraction data from Diamond Light Source. The researchers successfully crystallised a protein called RsbS from the bacterium Moorella thermoacetica, and solved the crystal structure to 2.5 Å resolution.

The data from the Diamond light source is managed on the Atlas Petabyte Data Store by STFC e-Science.



Section of the 3.5 Å experimentally phased electron-density map produced by the program RESOLVE.

Maureen Quin, Joseph Newman, Susan Firbank, Richard J. Lewis and Jon Marles-Wright **Crystallization and preliminary X-ray analysis of RsbS from Moorella thermoacetica at 2.5A° resolution**, 2008 Acta Crystallographica Section F, Structural Biology and Crystallization Communications