



## **E-SciNet Project Report 2007**

Dave Berry, National e-Science Centre  
Clare Gryce, University College London  
Alex Hardisty, Cardiff University  
Cliff Addison, University of Liverpool  
Mark Birkinshaw, University of Bristol  
Ian Bland, University of Reading  
Andrew Parker, University of Cambridge

**06-03-2008**

### **Abstract:**

This document reports the activities of the EPSRC-funded E-Science Network project (E-SciNet), covering the period from January 2007 to December 2007. It describes the work to set up the project, the events supported during this period and progress against members' project plans.

The project has had a successful first year. We have supported members attending meetings of the e-Science Directors' Forum, of the Campus Grids Special Interest Group and the e-Science All-Hands Meeting, as well as other events of interest.

The infrastructure to support the network is based primarily around the web site and wiki, which were set up early in the year. It took us longer than expected to put in place all the financial and contractual procedures but these are now completed.

## **UK e-Science Technical Report Series**

### **Report UKeS-2008-01**

Available from [http://www.nesc.ac.uk/technical\\_papers/UKeS-2008-01.pdf](http://www.nesc.ac.uk/technical_papers/UKeS-2008-01.pdf)

Copyright © 2008 The University of Edinburgh. All rights reserved.

# TABLE OF CONTENTS

<b>1</b>	<b>Abstract .....</b>	<b>3</b>
<b>2</b>	<b>Background.....</b>	<b>3</b>
2.1	Overview and Goals .....	3
2.2	Project Structure .....	3
<b>3</b>	<b>Membership.....</b>	<b>3</b>
3.1	Funded and Non-Funded Members.....	3
3.2	Multiple Members per Institution .....	4
3.3	Initial Members .....	4
3.4	Additions and Deletions .....	5
<b>4</b>	<b>Infrastructure .....</b>	<b>5</b>
4.1	Web Site .....	5
4.2	Wiki .....	5
4.3	Mailing List.....	5
<b>5</b>	<b>Events .....</b>	<b>6</b>
5.1	E-Science All-Hands Meeting.....	6
5.2	E-Science Directors' Forum .....	6
5.3	Campus Grids SIG .....	6
5.4	Other E-SciNet Sponsored events .....	6
5.5	Other Events .....	6
<b>6</b>	<b>Management.....</b>	<b>7</b>
6.1	Collaboration Agreement and Sub-Contracts .....	7
6.2	Action Plans .....	7
6.3	Operating Procedures .....	7
6.4	Management Meetings .....	8

<b>6.5</b>	<b>Publicity .....</b>	<b>8</b>
<b>6.6</b>	<b>Finances.....</b>	<b>8</b>
<b>6.7</b>	<b>Project Deliverables .....</b>	<b>8</b>
<b>7</b>	<b>Future Plans .....</b>	<b>8</b>
<b>8</b>	<b>Progress Reports from Members .....</b>	<b>9</b>
<b>8.1</b>	<b>UCL .....</b>	<b>9</b>
<b>8.2</b>	<b>Cardiff .....</b>	<b>11</b>
<b>8.3</b>	<b>Reading.....</b>	<b>12</b>
<b>8.4</b>	<b>Liverpool .....</b>	<b>14</b>
<b>8.5</b>	<b>Birmingham and Warwick .....</b>	<b>15</b>
<b>8.6</b>	<b>Cambridge .....</b>	<b>15</b>
<b>8.7</b>	<b>Bristol .....</b>	<b>15</b>
<b>9</b>	<b>Project Roles.....</b>	<b>17</b>
<b>9.1</b>	<b>Dave Berry - Principal Investigator.....</b>	<b>17</b>
<b>9.2</b>	<b>Clare Gryce - Co-Investigator .....</b>	<b>17</b>

# 1 Abstract

This document reports the activities of the EPSRC-funded *E-Science Network* project (E-SciNet), covering the period from January 2007 to December 2007. It describes the work to set up the project, the events supported during this period and progress against members' project plans.

The project has had a successful first year. We have supported members attending meetings of the e-Science Directors' Forum, of the Campus Grids Special Interest Group and the e-Science All-Hands Meeting, as well as other events of interest.

The infrastructure to support the network is based primarily around the web site and wiki, which were set up early in the year. It took us longer than expected to put in place all the financial and contractual procedures, but these are now completed.

## 2 Background

### 2.1 Overview and Goals

The E-Science Network (E-SciNet) seeks to embed e-Science as a standard practice for research advancement across Higher Education Institutions (HEI's) and across disciplines. It aims to develop and share best practice in encouraging take-up of e-Research methods and technologies and in bridging the gap between applications and e-Infrastructure.

A particular focus is on sharing strategy and experience regarding outreach within HEI's and the provision and exploitation of 'Campus Grids'. The network also aims to channel user requirements back to e-Infrastructure providers so that the provision and development of e-Science services can be more effective at meeting the real needs of researchers and their applications.

The network builds on the activities of the e-Science Core Programme (eSCP), including events run by JISC, NeSC, eSI, NGS, OMII-UK, the Campus Grids SIG and the All-Hands Meeting. We also run our own network-specific events.

### 2.2 Project Structure

E-SciNet is run jointly by NeSC and UCL. Membership is open to any HEI in the UK; institutions may join during the operation of the network. Funded members may claim travel expenses for attending E-SciNet events and other workshops of relevance to the network.

The network provides information via a web site, wiki and mailing list.

## 3 Membership

### 3.1 Funded and Non-Funded Members

For operational purposes, the network is divided into two classes. Funded members are eligible for financial support from E-SciNet. Non-funded members are those with existing sources of funds, such as EPSRC awards for e-Science centres, which cover travel to e-Science events. This ensures that the limited funds of the network are directed to those institutions that most need them.

To qualify for funding, members must submit a summary of their plans for increasing the uptake of e-Research methods within their institution. These are used to track progress against goals and are made available to other members on the project web site. Funded institutions are also required to sign a sub-contract agreement with the University of Edinburgh, as detailed in the Management section of this report.

## 3.2 Multiple Members per Institution

As the year progressed, it became clear that several institutions wished to send different members of staff to different meetings. This is because the meetings reflected a range of activities and these activities are the responsibility of different people. Therefore we adopted a policy whereby several people may be members from the same institution. All such members are eligible for travel funding (where applicable). We restrict the number of funded delegates from each institution per meeting, so ensure that all institutions get a chance of funding.

The original member for each institution remains responsible for the oversight of network activity within that institution, so that we have a single point of contact for contractual and other matters.

## 3.3 Initial Members

The initial members were as specified in the project proposal.

Funded members:

- Dr. Clare Gryce (UCL & co-Investigator)
- Prof. Andrew Parker (University of Cambridge)
- Mr. Alex Hardisty (University of Cardiff)
- Prof. John Darlington (Imperial College)
- Prof. Mark Birkinshaw (University of Bristol)
- Mr. Tony Linde (University of Leicester)
- Dr. Georgios Theodoropoulos (University of Birmingham)
- Dr. Stephen Jarvis (University of Warwick)
- Prof. Keith Haines (University of Reading)
- Dr. Jon Blower (University of Reading)

Non-funded members:

- Prof. Malcolm Atkinson (UK e-Science Envoy)
- Dr. Neil Geddes (STFC)
- Dr. Steven Newhouse (OMII UK)
- Prof. Carole Goble (University of Manchester)
- Prof. Andrew Keane (University of Southampton)
- Prof. John Brooke (University of Manchester)
- Prof. Paul Watson (University of Newcastle)
- Prof. Ron Perrott (Queen's University, Belfast)
- Prof. Robert Crouchley (University of Lancaster)
- Prof. Paul Jeffreys (University of Oxford)
- Dr. Joanna Schmidt (University of Leeds)
- Prof. Richard Sinnott (University of Glasgow)
- Prof. Saleem Bhatti (University of St. Andrews)

We have not yet received institutional plans from Imperial College or Leicester. Therefore members from these institutions have not been able to claim travel expenses from the network. Tony Linde left the university at the end of the year, so Leicester's participation in the network appears doubtful.

## 3.4 Additions and Deletions

The following institutions have joined the network during the first year.

Funded members:

- Dr. Clifford Addison (University of Liverpool)

Non-funded members:

- Mr Matthew Dovey (JISC)
- Dr. Stuart Dunn (King's College London)
- Dr. Robert Fletcher (University of York)

We are proactively inviting other institutions to join the network. Currently we are actively talking to two other institutions about potential membership.

Steven Newhouse left academia and we therefore cancelled his membership of the network.

## 4 Infrastructure

### 4.1 Web Site

The web site is hosted by NeSC and provides information about the project. It is partly driven from a database and includes web forms that people can use to apply for membership or funding. There is a corresponding administration site that the investigators use to process requests.

The public area of the web site describes the network and lists the members and events. It includes links to reports of key events and a link to the wiki. It also includes a web form for people who want to join the network.

The members-only area contains members' institutional plans, minutes of management meetings and a form to request travel support.

The development of the web site took place hand-in-hand with the development of our management procedures, e.g. for processing applications.

### 4.2 Wiki

The E-SciNet wiki provides an interactive web space where members can share information. By request from the membership, there is no password on the wiki other than the need to sign up before editing.

Currently the wiki is used mainly by events. It includes a list of events and detailed programmes and reports from key events. The wiki has not been used for general discussion, although we would like this to happen.

### 4.3 Mailing List

We operate a Majordomo mailing list. Members are automatically added when they are accepted into the network. They are sent the usual confirmation e-mail to protect against bogus subscriptions. Only members may post to the list.

So far, the mailing list has only been used by the investigators to announce events and other news. However, the list is open to members and could be used for general discussions.

## 5 Events

One purpose of E-SciNet is to support relevant events and to pay the expenses of funded members to attend relevant events.

### 5.1 E-Science All-Hands Meeting

The AHM is a central event for the e-Science community and the E-SciNet proposal specifically commits to supporting members to attend.

- 10th - 13th September 2007 [The UK e-Science 2007 All Hands Meeting \(AHM\) - Nottingham.](#)

### 5.2 E-Science Directors' Forum

These meetings are for directors of e-Science centres and similar institutions. Again, the E-SciNet proposal specifically commits to supporting members to attend.

- 4th July 2007 [e-Science Directors' Forum - NeSC.](#)
- 30th October 2007 [e-Science Directors' Forum - Newcastle.](#)

### 5.3 Campus Grids SIG

Campus Grids are a key means for increasing the uptake of e-Science methods. E-SciNet supports the activities of the Campus Grids SIG by sponsoring meetings and paying travel expenses of network members. Reports of these events can be found on the web site and wiki.

- 16th April 2007 [Campus Grids Special Interest Group - First meeting, Oxford.](#)
- 28th June 2007 [Campus Grids Special Interest Group - Second meeting, Access Grid.](#)
- 10th September 2007 [Campus Grids Special Interest Group - Third meeting, Nottingham \(co-located with the AHM\)](#)

### 5.4 Other E-SciNet Sponsored events

E-SciNet sponsors some other key events. Reports of these events can be found on the web site and wiki.

- 27th - 30th November 2007 [High Throughput Computing Week - NeSC.](#)
- 10th December 2007 [Workshop for UK e-Science Educators - NeSC.](#)

### 5.5 Other Events

E-SciNet also provides support for members to attend other events of interest. The PI and co-I maintain a list of possible events on the web site, including those shown here. We also ask members to suggest worthwhile events. As an example, we part-funded Alex Hardisty's trip to the EGEE 2007 conference, in return for which he provided a report that is available on the web site.

- 14th-15th June 2007 [Support for e-Research: Filling the Library Skills Gap - NesC.](#)

- 27th - 28th September 2007 [Grid Computing and the National Grid Service Induction - NeSC](#).
- 11th - 12th October 2007 [An Introduction to Grid Data Services using OGSA-DAI - NeSC](#).
- 19th October 2007 [Data and Storage on the National Grid Service - NeSC](#).
- 31st October 2007 [NW-Grid Campus Grids Workshop - University of Liverpool](#).
- 1st - 2nd November 2007 [The Chris Date Seminar: Data and Knowledge Modelling for the Geosciences - NeSC](#).

## 6 Management

### 6.1 Collaboration Agreement and Sub-Contracts

The E-Science Network grant is held by the University of Edinburgh, with funds transferred to UCL under a collaboration agreement. This agreement, although straightforward, took a long time for the two universities to process. It was finally completed in November 2007.

Each funded member of the network is required to sign a letter of sub-contract with the University of Edinburgh (UoE). This sub-contract sets out the procedures for members to claim payment for expenses. This procedure is complicated by the fact that EPSRC will only pay 80% of these costs. In order to ensure auditable accounts, network members are required to claim 100% of their expenses from UoE; UoE will then claim back 20% of these costs at the end of the project. This unusual arrangement has caused delays as the contract management departments at some institutions have questioned the process.

### 6.2 Action Plans

As noted previously, each funded institution is required to submit a brief plan that summarises how they will advance the goals of the E-Science Network within their institution. UCL, Reading, Birmingham and Warwick, Cardiff, Cambridge and Liverpool have all written their plans.

We encountered some management problems, as some institutions pointed out that they could not commit these plans to writing without the support of their institution's computing services. This caused a small delay before the plans could be uploaded to the web site. Also, one site decided to revise their plan when they realised that it would be publicly available.

### 6.3 Operating Procedures

We have defined the following operational procedures. :

- Requests for non-funded membership
- Requests for funded membership
- Agreeing support for events
- Requests for travel funding to events
- Processing of expense claims
- Members leaving the network

Where appropriate, the web site includes forms for applicants to use and also private forms for the investigators to process applications.

We have given two options for processing expense claims. Claimants may submit expense forms directly to UoE. Alternatively they may process expenses forms locally and have their institution bill UoE.

## 6.4 Management Meetings

The investigators have held 22 management meetings over the course of the year. The minutes from these meetings are available via the members' section of the web site.

## 6.5 Publicity

To publicise the network, we have produced advertising "postcards", which have been included in delegates' packs at the AHM and other relevant meetings. We have a logo for the network which has been included in material for our sponsored events. We have also reported on project activities to the E-Science Directors' Forum and have approached potential new members directly.

## 6.6 Finances

The project has a total of £90,000 to fund members' travel. We have provisionally allocated approximately £9,000 to each funded institution, on the assumption that we began with 8 funded members and hoped to attract 2 more during the lifetime of the project. Actual takeup has been much lower than expected so far, both in terms of the number of institutions and the amount claimed per institution.

<b>Institution</b>	<b>Expenses to date</b>
UCL	£0
Cardiff	£1,900
Reading	£856

We expect this takeup to increase in the second year of the project. We will use some of the underspend to increase funding for key events (e.g. speakers' expenses). If the underspend continues, we expect to request a no-cost extension for the network.

## 6.7 Project Deliverables

The only deliverable required by this project is the annual report, i.e. this document. Other outputs include the content of the web site and wiki; the members' institutional plans; and event reports.

# 7 Future Plans

For the coming year, we plan to continue to support members to attend a range of events, including those organised by the Campus Grids SIG and by NeSC. We would also like to support regional events organised by members of the network.

If the AHM2008 takes similar format to previous years, we will run a booth for E-SciNet funded members, thus giving them booth space at the event that they would not otherwise be able to afford.

We are actively planning a number of events. These include a 4-day event on SRB/iRODS at NeSC, a similar event on Web 2.0 technologies, and a one-day workshop on Computing Services Support for e-Science. We have several other topics in mind for future events.

We will look to slowly grow the membership of the network.

## 8 Progress Reports from Members

### 8.1 UCL

UCL's e-Science activity is supported at the institutional level by the Research Computing team within Information Systems. For 2007, the activity of this team was dominated by the procurement and early installation of a new SRIF3-funded HPC system. This project has consumed many more man hours and energy than was expected and, regrettably, the team has not been able to progress some of its e-SciNet objectives as far as had been hoped.

Nonetheless, UCL Research Computing remains committed to supporting e-Science at the institutional level and there have been some useful engagements during the year that have increased our understanding of the needs of some of our heaviest users of e-Science technologies and methods. Additionally, we have increased our understanding as to why, historically, it has not been possible to support some more novel computational paradigms and models as requested by these users.

Continued lobbying has resulted in a bid for a new Research Computing staff post being submitted to College. If funded, it is envisaged that this post lead on projects to develop UCL's e-Science service infrastructure through projects such as those outlined above, in addition to providing a point of liaison into the e-Science technology development community.

All this bodes well for our ability to develop a sustainable platform for the support of e-Science technologies and methods going forward.

A report on progress against each of our stated objectives (which are given in italics) is provided below:

#### ***1. To increase usage of the UCL Condor Pool.***

*The UCL Condor Pool includes over 1400 W2K machines, each with up to 1.5 GB RAM. These machines are deployed across a large number of student cluster rooms. To date, the UCL Condor Pool has been run on a best-efforts basis by research staff. In order to develop the pool as a more 'mainstream' resource and develop a robust set of services around its use, the Research Computing team should assume full responsibility for the Condor Pool.*

*Realisation of this objective is dependent on:*

- *modest funds for new hardware (new Central Manager and Submission Machine(s)),*
- *Staff time as above.*

#### Action Plan:

- (i) *Procurement of appropriate hardware (by Aug '07)*

This was completed by Aug '07.

- (ii) *Migration of Systems Administration responsibility to core Research Computing team (by Sept '07)*

This is in progress but has not yet been completed. The current estimated date of completion is March '08.

- (iii) *Incorporate Condor support into central User and Applications Consultancy service (by Sept '07)*

As for (ii)

- (iv) *Outreach within UCL and proactive/assisted migration of tasks to Condor Pool where appropriate (ongoing from Sept '07)*

This has not yet commenced and is dependent on the successful conclusion of the HPC project for staff resource.

- (v) *Further development of Condor service as appropriate (ongoing from Sept '07)*

As for (iii)

- (vi) *Evaluation:*

(vii) *Success will be measured as the number of projects active on the Condor Pool per month.*

(viii) *This metric is currently available and a baseline therefore readily established.*

The number of projects being supported by the Condor Pool has remained static as has the overall utilisation. Increased uptake is dependent on the above actions.

**2. *To provide more usable access to distributed Research Computing and other resources, within and external to UCL.***

*At present, access to the various Research Computing systems at UCL is provided by direct login to each individual machine. There is no uniform interface that provides seamless access to multiple such resources internal or external to UCL. In order to facilitate the integrated and co-ordinated use of such distributed resources, the deployment of an appropriate, usable and uniform interface for access should be pursued.*

*Realisation of this objective is dependent on:*

- *identification of an appropriate and ready technology,*
- *possible modest funds for new hardware (server?),*
- *staff time as above.*

*Action Plan:*

- (i) *Investigation of available technologies (to Sept '07)*
- (ii) *Development of implementation plan for chosen technology (Sept to Dec '07)*
- (iii) *Deployment of chosen technology (Jan '08 to March '08)*
- (iv) *Assimilation of user support for deployed interface into existing Research Computing staffing resources (Feb '08 to March '08)*
- (v) *Outreach within UCL and promotion of assisted usage of interface (ongoing from March '08)*

*Evaluation:*

- *Success will be measured as:*
  1. *The successful deployment of the proposed interface.*
  2. *The proportion of accesses to available systems made through the deployed interface, per month.*
  3. *User feedback regarding use of the interface.*
- *An appropriate baseline for (2) and (3) will be established on deployment.*

This activity has not yet been formally initiated though as explained above, we believe that we have increased our understanding of some user requirements and possible deployment issues that will assist in our progressing this activity when time permits.

Staff time continues to be a limiting factor and to this end, as explained above, a bid for an additional Research Computing staff post for Application Development has been submitted to College. If funded, it is envisaged that this post would lead on this project and others.

## 8.2 Cardiff

Cardiff's e-Science activity at the institutional level has been "in transition" throughout 2007. The year can be characterised as one in which Cardiff has commenced the journey of transferring the technology of e-Science out of the research domain and into the service domain. This will not be accomplished overnight, nor is it completely clear at this moment what that journey will look like in its entirety. More strategic formulation is needed in that regard.

As indicated in our submitted Action Plan, the University is establishing a new University Division, "Advanced Research Computing@Cardiff" (ARCCA) with the mandate to transform Cardiff University's approach to Advanced Research Computing. Over the medium-term this includes the adoption of relevant e-Science / e-Research / e-Infrastructure techniques as they reach a level of maturity sufficient for deployment.

Professor Martyn Guest, formerly of STFC Daresbury Laboratory, has been appointed as Director of ARCCA and took up his post in spring 2007. Since then, Martyn has had his hands full initiating the new Division, moving staff into new office accommodation, undertaking procurement of new SRIF-3 funded computing infrastructure, and building a brand new machine room to house the equipment when it arrives (due March 2008). At the present time, ARCCA expects to start providing its high-end computing and storage service to the University's research community in April 2008. Based on one of the first major installations of Intel quad-core technology in Europe and with state-of-the-art water cooling, this service will be one of the greenest research computing services in the UK. Alongside the computing service, Bull (the SRIF3 technology provider) are partnering with ARCCA and Intel to establish a High Performance Computing "Centre of Excellence" at Cardiff in combination with ARCCA's own dedicated support engineers. This will provide a resource for the provision of support and expertise at all levels.

In addition to the SRIF3 cluster service, ARCCA are also responsible for developing and running the University's Condor Pool. With over 1600 Windows PCs providing spare CPU cycles, this is one of the largest Condor systems in the UK, being more heavily utilised both in terms of cycle provision and number of different users than any other UK-based Condor solution. ARCCA and WeSC, working together throughout 2007, have made some of the capacity of this pool available to the National Grid Service (<http://www.grid-support.ac.uk/content/view/305/136/>), thus reinforcing Cardiff's commitment to being an NGS "Partner".

Cardiff University is also undertaking a substantial multi-million pound modernisation of its more general purpose IT environment under the heading of the "Modern Working Environment" (MWE) (<http://www.cf.ac.uk/insrv/futures/mwe/index.html>). This work is being carried out by the Information Services Division (INSRV). This multi-year programme, utilising portal technology and personalisation, aims to provide access "any time, any place, any device" to the University's information systems for all staff (academic and administrative) and students. During 2007, WeSC, INSRV, ARCCA and School of Computer Science held a joint workshop to explore the opportunities offered by the MWE programme, and the existing and emerging tools for e-Research that may be of use to Cardiff University. As the MWE and e-Science agendas move forward we expect to identify priority areas for development that will include "e-Science" portlets for access to Condor, new ARCCA service and the National Grid Service.

e-SciNet funding has been and continues to be important to the success of the journey being described. Funding has been utilised not only to sustain engagement of core e-Science staff with the national activity but also as a means of introducing staff, both from the service provision side of the house and from the research community itself, to e-Science. Engaging staff in e-SciNet activities leads, we believe, to raised levels of awareness, realisation of the benefits and upskilling through training. We see e-SciNet as an effective mechanism for bridging gaps that will continue to be important for the foreseeable future. During 2007, staff have been funded to attend; AHM07, e-Science Director's Forum, CampusGrid SIG, High Throughput Computing week and the EGEE conference.

Late in 2007, WeSC and ARCCA jointly submitted a bid to EPSRC's call for platform grant funding for e-Science, with the main objective of establishing a 5 year platform for further embedding e-Science into Cardiff University. At the time of writing of this report we are presently awaiting outcome of the EPSRC evaluation process.

From these foundations, ARCCA, INSRV, WeSC and the University will be positioned to move forward with the further formulation and implementation of our strategy for e-Science in Cardiff.

### 8.3 Reading

Reading proposed two areas in its Action Plan: to improve access to Reading Campus Grid and NERC Cluster Grid, and to work with Oxford University towards the development of a Thames Valley Grid. Supporting both of these developments has been the appointment of David Spence, formerly of STFC-RAL, to the post of e-Science development officer at the University. The post is split, half with Reading and half with the Oxford e-Research Centre.

The following plan references the Action Plans for each of these areas.

#### *1a. Improve Access to Reading Campus Grid*

One of David's first tasks has been to map e-science activities in the University with a view to getting more grid users onto the growing campus grid. This has resulted in a number of proposals and trials being run on the grid. In particular, the grid has been used to help support Cybernetics research into intelligent game playing by analysing chess positions for particular players in an attempt to develop a statistical comparison between these engines and the relative human strength of a player (iii). This project has been useful to exercise the grid and has taught both researcher and grid administrator in the process.

As according to our plan, Reading has succeeded in transferring administration of the grid over to its IT Services department (i). This has put the master node on new hardware (iv) and increased the number of nodes in the grid. The grid now come under the normal support of IT Services and features in their service catalogue with David providing Applications Consultancy alongside others in ITS and Systems Engineering (ii). Reading is continuing to use the CoLinux approach, which provides users with a Unix (Linux) environment using the wealth of Windows based machines and the technology to provide clients has also been transferred, although development continues to rest with the IT team in Systems Engineering. This approach continues to prove popular with researchers, who continue to present problems in the Unix domain. An increasing requirement is support of Matlab based code and there have been some attempts to use the Matlab compiler to distribute these in the grid. Unfortunately, these have not proven particularly successful as the combination of Matlab compiled code and Condor have created segmentation violations in the Matlab code. Development work aimed at identifying the root cause of this continues (vi). Progress developing a repository has not formally progressed but the team is now looking at iRODS and leveraging research expertise from elsewhere in the University (v). The use of NFS locally has been particularly successful and so this has been a disincentive to move to a more formal storage solution, although the inevitability of this is recognised, especially if this is to be shared over an inter-institutional grid.

#### *1b. Improve Access to NERC Cluster Grid*

##### *(i) Development of Styx Grid Services software*

A big step forward was made by the introduction of a new Java grid middleware system called G-Rex ("Grid Remote Execution"). G-Rex is the successor to Styx Grid Services, and from the user's perspective is almost identical. The main differences are in the software design framework and the data transfer mechanism. G-Rex is a Web application that runs inside a servlet container such as Apache Tomcat. It uses HTTP instead of the Styx protocol for data transfer, which enables the developers to use popular, open-source

Java libraries such as Apache Commons and Ageci Security. This, and the fact that G-Rex is built using the Spring application framework, makes the G-Rex code much smaller and easier to develop and maintain than Styx Grid Services. G-Rex is undergoing continuous development, the developers benefiting from the experience of setting up and maintaining reliable grid services that are routinely used by scientists.

- (ii) *Outreach within the NERC modelling community to increase awareness of the benefits of sharing computational resources (ongoing from June '07)*

The NERC Cluster Grid project has benefited from Reading e-Science Centre (ReSC) developers joining the NERC Cluster Compute Forum (CCF). This group, composed mainly of cluster administrators, aims to promote cluster computing at NERC institutes and help scientists to use clusters to achieve their objectives. After a presentation about G-Rex and the NERC Cluster Grid at the last meeting, CCF members expressed support for this approach as a way to increase utilisation of NERC institute clusters. Awareness of G-Rex within the ocean and climate modelling community is growing, with recent requests from potential users for services running on HECToR, HPCx and the QUEST cluster in Bristol.

- (iii) *Deployment of NEMO and POLCOMS ocean models as SGS services on all grid nodes (June-July '07)*

NEMO and POLCOMS have been deployed as G-Rex services on the three clusters currently in the Grid. The NEMO services are being used routinely by scientists, and frequent use of the POLCOMS services is expected to start soon following recent progress in the GCOMS project.

- (iv) *Deployment of additional scientific applications as SGS services on all grid nodes (from July '07)*

No new applications have been deployed yet, but there are plans to set up HadCM3 model services in connection with the GCEP project.

- (v) *Implementation of SRB grid storage management system (Jan-Mar '08)*

SRB implementation has dropped down the list of grid development priorities. G-Rex users are benefiting from the automatic transfer of output data from remote clusters to storage facilities at their own institutes. Data transfer takes place while models are running in order to minimise the data footprint of G-Rex services on remote clusters, and this feature has reduced the need for a separate grid storage system such as SRB.

- (vi) *Expansion of grid to include computational resources belonging to other collaborating organisations (from Jan '08)*

ReSC is currently discussing the possibility of G-Rex deployment with system administrators at HECTOR, HPCx and Bristol University. There has also been a recent enquiry from the NGS.

2. *To work with Oxford towards the development of a Thames Valley Grid.*

Reading has applied for affiliate status to the NGS. It has configured its gateway to support the necessary submission protocols and has passed the acceptance tests (iii). Management of security and support contacts has been established within IT Services. It is envisioned that this will be the first step in joining the two campus grids based in Oxford and Reading into a wider Thames Valley grid. The approach also encourages users in both institutions to have a common interface to the NGS: the need to preserve local authentication to grid resources has been identified (iv). David's unique role between the two institutions is especially helpful with outreach to both communities and he will be using his experience gained at the NGS to help both communities to use the interfaces and work-flows available (vi).

## 8.4 Liverpool

Liverpool joined e-SciNet in September 2007 and this report concentrates on activities in the last 4 months of 2007.

Liverpool has an active Condor / Condor-G service. The Campus Windows Condor Pool provides access to approximately 300 systems spread across Teaching Centre and other limited access rooms. These systems are configured so that they can be powered on remotely when required for a Condor job and then are powered off after a certain time period if there is no logged in user. This means that systems do not need to be left on overnight and at weekends just to provide Condor support.

The Condor-G service provides a relatively easy-to-use interface to several of our local Beowulf clusters, including systems that form part of NW-GRID. Files for these jobs can be staged in and out via a local SRB service. Users with a UK e-Science certificate can also use the local SRB service to store the results of Condor runs from the Windows Condor Pool.

Much of the Liverpool e-Science activity is being run through the NW-GRID consortium, which in addition to Liverpool includes STFC Daresbury and the Universities of Lancaster and Manchester.

NW-GRID flyers and a rolling presentation were used at the AHM 2007.

At the end of October, NW-GRID sponsored a one-day workshop on Campus Grids at Liverpool. Approximately 30 participants attended the event. PowerPoint presentations of the talks are available at <http://www.nw-grid.ac.uk/?q=nodes/uoli/CampusGrid>. Liverpool presentations included an overview of our Condor service by Dr. Ian Smith and an eye-opening presentation by Dr. Zafar Chaudry on the data issues at Liverpool Women's Hospital and how virtualisation is helping to cope with an explosion in storage requirements.

Although most of the attendees to the Campus Grid workshop were academics, a one-day industrial seminar co-sponsored by NW-GRID at this year's Machine Evaluation Workshop identified a strong interest amongst some industrial users in exploiting Condor-like techniques to coordinate internal computational power to support production activities with reasonably sharp time constraints.

Some preliminary contact was also made (at Liverpool workshop on Image Processing) with Radiologists at Liverpool University and at Clatterbridge Hospital (Wirral) and Christie's Hospital (Manchester) on their data and computation needs. It is hoped that follow-up discussions with both of these external groups can be arranged in 2008.

Lancaster and Liverpool tendered for new systems as part of NW-GRID, with orders placed just before Christmas. It is hoped that expanded facilities at both locations will be available to users by the end of March.

Finally, NW-GRID has successfully applied to join the National Grid Service. A certain number of node hours from all 4 NW-GRID sites should be available to NGS users by the end of the first quarter 2008.

NW-GRID related activities in 2008 will concentrate firstly on running a special Test Bed with a focus on multi-site coordination and real-time interaction with computations, secondly on making the NW-GRID infrastructure more attractive to new users and finally on how the NW-GRID infrastructure can be sustained when the funded part of the project ends in March 2008.

Liverpool specific activities will focus on getting Liverpool users to exploit the Grid when planning, coordinating and running computational experiments. An Engineering project with extensive data and computation requirements is just starting. The School of Tropical Medicine and School of Veterinary Sciences are two other groups that we want to contact with regards to Condor-work. Both these groups successfully completed large scale parameterisation studies in 2007 using the Windows Condor Pool, but we believe their efforts were more labour intensive than they could have been.

## 8.5 Birmingham and Warwick

No report received.

## 8.6 Cambridge

Recent progress highlights against our stated E-SciNet objectives are as follows:

1. *"To maintain and expand the use of e-Infrastructure on a permanent, self-sustaining basis"*

- Office space in the Department of Physics is being made available to accommodate the new Cambridge Centre for Scientific Computing (CSC), incorporating resources from CeSC, the University Computing Service and HPC service.
- The uptime and usage figures on CamGrid (our local campus grid) remain constant. Further efforts to promote use of CamGrid are planned.

2. *"To increase the depth and diversity of e-Science uptake within Cambridge and beyond."*

- E-SciNet funds enabled Mark Calleja to attend the recent High Throughput Computing week at NeSC, where he gave presentations on the eMinerals project and CamGrid.
- Michael Simmons is organising several outreach events in the next few months, including a potential workshop at the June OGF meeting on "creative industries and search".
- Many recent publications cited use of CamGrid and support from CeSC (4 already in 2008). See <http://www.escience.cam.ac.uk/projects/camgrid/publications.html> for further details.

3. *"To take a leading role in educating students in computational science"*

- CeSC is now working in partnership with the grassroots "scicomp@cam" group (see <http://www.scicomp.org.uk/>) to shape its programme of events. Membership of scicomp@cam includes students and post-docs who find it a useful forum for networking and sharing expertise. The CSC is also developing an MPhil course in computational science.

## 8.7 Bristol

Bristol's efforts are concentrated on increasing e-Science awareness in the University and advancing the University "Exabyte Initiative" in relation to the new HPC facility under construction. Specific required actions, and progress, are:

*Action 1.1: Maintain access to Condor pool*

Progress: The Condor pool remains available, with modest user levels.

*Action 1.2: Provide assistance with grid systems*

Progress: This is currently done through GridPP. Activities in that area continue to be strong, though some disruption is anticipated as the main user group plans for relocation in early 2009.

*Action 1.3: Provide training in e-Science systems.*

Progress: Training is undertaken solely through NeSC at present. There has been little activity here, except in relation to training under action 2.2 below.

*Action 1.4: Improve e-Science infrastructure*

Progress: Little activity here at present, though planning is being done for maintaining access of Physics users after their relocation.

*Action 2.1: Complete existing Exabyte Initiative projects*

Progress: A project in the automated recognition of objects in infrared astronomy has been completed by the Masters student who took it on. Other projects relating to dataset classification, bird recognition, and so on continue.

*Action 2.2: Integrate HPC and existing e-Science operations*

Progress: Most activity over the past period has concentrated in this area, and has centred on establishing the HPC facility, under the leadership of Dr. Ian Stewart. A partial installation has been completed, and this is in heavy demand, while the physical plant to support the full facility has now been completed. Training of students and staff to use the facility is ongoing.

## 9 Project Roles

### 9.1 Dave Berry - Principal Investigator

Dr. Dave Berry is Deputy Director for Research & e-Infrastructure Development at the National e-Science Centre (NeSC). He is project manager for the EPSRC NeSC project, overseeing the efficient running of the NeSC services. He leads the NeSC Edinburgh Research & e-Infrastructure Development team, which currently numbers 12 staff. He leads the technical content for the TSB-funded Knowledge Transfer Network, *Grid Computing Now!*

Dave is co-investigator of several e-Science research projects, including GridQTL (BBSRC), FireGrid (DTI) and NanoCMOSGrid (EPSRC). He is also co-chair of the OGSA Data working group of the Global Grid Forum. Dave has organised many e-Science workshops and meetings at NeSC and elsewhere.

Project responsibilities include:

- Deciding on requests for membership and discretionary funding (with the co-I)
- Producing project reports
- Outreach to new potential members
- Liaison with the advisory board
- Scheduling and running management meetings
- Reporting to the E-Science Directors' Forum

### 9.2 Clare Gryce - Co-Investigator

Clare Gryce is the Research Coordinating Manager at UCL, working within Information Systems, part of UCL Education and Information Support Division. She was previously the Manager for the UCL e-Science Centre of Excellence in Networked Systems based in the Department of Computer Science.

She is responsible for the delivery and development of Research Computing services to the UCL research community including computational services, user and applications support and training. She is also responsible for community development associated with the use and management of the above services.

Within UCL she is also a member of the Research Computing Sub-Committee and the Computational Resource Allocation Group. Within the UK she is Secretary to the UK High-Performance Computing Special Interest Group.

Other relevant posts held include Organisation Chair for the International Conference on Software Engineering (ICSE) 2004, held in Portland USA.

Project responsibilities include:

- Organising network-specific events
- Overseeing the writing of event reports
- Maintaining the list of supported events
- Outreach to new potential members
- Deciding on requests for membership and discretionary funding (with the PI)