The monthly newsletter from the National e-Science Centre

NeSC News

Issue 72 August 2009 www.nesc.ac.uk

National e-Science Centre

A New Age for the Oldest Science

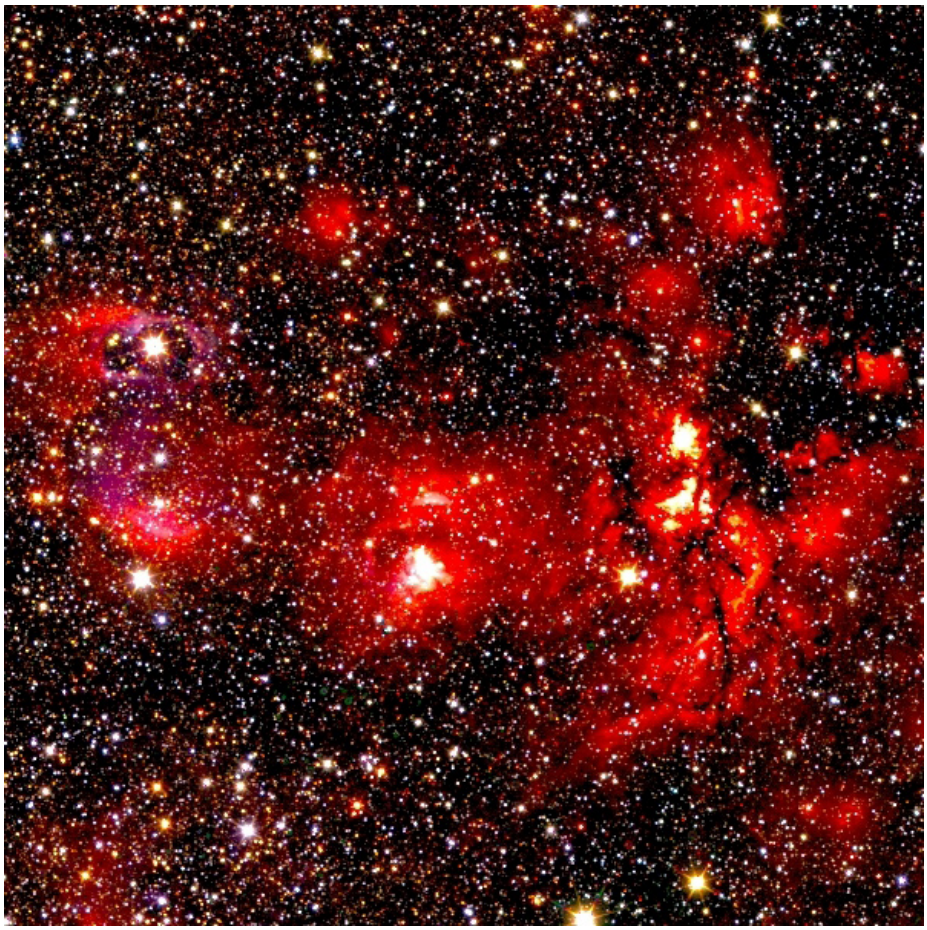
By Iain Coleman

For millennia, astronomy meant looking at the night sky and sketching what you saw. In this way the first star maps were made, with the relative brightnesses of stars estimated by eye and the routes of the wandering planets traced against the celestial sphere.

Even the advent of the telescope didn't change this much. Sure, astronomers could see fainter objects like the Galilean moons of Jupiter, and resolve point-like planets into disks with structures of colour and shade. But the human visual system remained an integral part of the instrument. This not only limited data gathering to what could be seen and sketched by a human observer in real time, it also gave rise to optical illusions like the fabled Martian canals.

A greater revolution even than the telescope was the photographic plate. Photography doesn't just allow more precise and objective measurements to be made of the position and brightness of objects in the sky: with long exposures, it can also reveal fainter and more distant objects than could ever be seen by eye. But this still involves some human effort in extracting the data, from developing the plates to reducing the data into standard form.

In recent years, all the mucking about in darkrooms has been superseded by digital photography. At first, this might have seemed just a cleaner and more convenient way of doing the same science, but in time it has led to a huge change in the practice of astronomy.



Infrared image of a supernova remnant from the UK Infrared Deep Sky Survey

Digital astronomy started out as a host of small programmes targeting a few individual objects. The data, manually reduced, would typically end up in the astronomer's desk drawer. But as digital detectors became larger and cheaper, they gave birth to a new kind of astronomy: fewer surveys, but on a much larger scale, mapping vast areas of the sky at a time. Data reduction is automatic, and the data ends up in a queryable database that astronomers all around the world can make use of. Cold nights in the dome have been replaced by the warm glow of the computer screen.

These changes, driven by policy, economics and technology, are the subject of the new e-Science Institute research theme, "Next Generation Sky Surveys: Astronomical Opportunities and Computational Challenges". The Theme Leader, Dr Bob Mann, gave a public lecture at

A new age for the oldest science

Continued

eSI on 7 July in which he outlined the goals and ambitions of the theme to come.

Imagine making a survey of the whole sky. How much data would you generate? Well, the atmospheric interference that makes stars seem to twinkle sets a limit on the resolution you can observe from a ground-based telescope of about half an arcsecond, or just over one ten-thousandth of a degree. Dividing the whole area of the sky by this pixel size, and allowing 2-4 bytes per pixel to give an acceptable dynamic range for measurements, puts the size of a whole-sky survey at a few tens of terabytes. Call it 20 TB, for the sake of argument.

In the old days of photographic plates, it might take 60 years of observing time, and another ten years of digitisation, to produce 20 TB. Current digital sky surveys can produce 20 TB in a year. The new generation of sky surveys will produce 20 TB per night for a decade. As data volumes increase dramatically, the importance of computation also increases.

Survey science is statistical in nature. This can mean trying to characterise populations, such as clustering of galaxies or types of stars within galaxies – for which you need large samples. It can also mean detecting outliers from those populations, such as very distant quasars or very low mass stars – which are rare. In either case, you need to sample a large volume of the cosmos to produce the scientific results.

This leads to computing challenges. For example, users will want to analyse data sets that are too large to practically download. Hence the data analysis code must be run at the data centre. Also, the relational database model does not support all forms of astronomical data analysis well. This has led to the development of SciDB, a generalisation of the relational model based on multidimensional arrays.

A further challenge is in real-time follow-up studies of transient events. When a star explodes, for example, a range of telescopes can promptly swing round to examine the explosion at different wavelengths, revealing a wealth of scientific information – but only if they're told about it in time. If a sky survey is to be able to issue an alert within one minute of detecting such a transient effect, its data reduction system needs to achieve a data rate of around 2 TB / hour.

So it's clear that astronomers need to work with computing scientists. But is the reverse true?

Astronomical data can be very useful to computing scientists who want to develop data handling and data mining algorithms. The data has no commercial value and does not refer to human beings, so there are no issues of commercial confidentiality or personal privacy, unlike in pharmaceuticals or healthcare.

The data is real, well-documented, and multi-dimensional, with both spatial and temporal data from multiple instruments, places and times. And there is an awful lot of it. Computational astronomy is an excellent sandbox for data mining algorithms, and an effective way to teach both astronomy and computer science.

Over the next year, the theme will investigate how the new generation of sky surveys demands new computational techniques and engage the computing science community in solving them. The oldest science is entering a bright new age.

Slides from this event can be downloaded from http://www.nesc.ac.uk/esi/events/993/

SAGA tutorial

A SAGA tutorial and training day will be held at eSI on 3-4 September 2009. The Simple API for Grid Applications (SAGA) provides a simple, uniform, integrated interface for performing distributed functionality.

SAGA has been used to develop distributed applications and tools that support distributed applications. Application Scientists, Tool Developers and Resource Providers will find this event of great use.

Day one of the tutorial and training event will provide an overview of the landscape of distributed applications, followed by a detailed introduction to SAGA and some basic exercises aimed at becoming familiar with the API and deploying SAGA. Day two will involve further examples and simple programming exercises. Be sure to bring along your application or tooling needs, because the final half of the second day will be devoted to discussions of how SAGA can help your requirements.

For more information see http://www.nesc.ac.uk/esi/events/945.

Edikt Technical Workshop

A technical workshop will be held on September 16 in Edinburgh, to allow edikt2 participants to share information about the technical aspects of the various activities being funded through edikt2. The meeting is open to other interested parties.

The edikt (eScience Data, Information and Knowledge Transformation) project has been running since May 2002 and is using computational science to extract knowledge from vast datasets and simulation models. edikt is funded by the Scottish Funding Council.

For more information on edikt, see the project web site at http://www.edikt.org.uk.

Rapid portlets are a hit with chemists

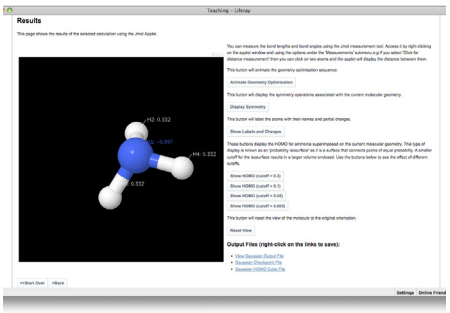
By Simon Hettrick and Gillian Law

Portlets make inaccessible technology accessible, because they run from within a browser – a familiar interface for even the most technophobic researcher. To encourage the use of portlets, it is necessary for them to be easy to develop. This led OMII-UK to fund Rapid, an easy-to-use portlet development tool.



Chemists from the Universities of Edinburgh and St. Andrews (EaStCHEM) have recently used the software to create portlets that have allowed access to computational-chemistry software by over 140 students.

EaStCHEM use computational-chemistry software, such as Gaussian 03, for research and teaching activities. However, technological barriers prevented many chemists from using the software. To overcome these barriers, the Rapid project showed EaStCHEM how to create portlets with Rapid. This training led to one chemist creating four portlets without the help of a software developer.



One of the four portlets is now used to teach over 140 chemistry students. This user-friendly portlet hides the complexity of the tutorial (such as, command line options, authentication protocols and job submission commands) from the student, so that they can concentrate on learning the software. Importantly, the only complexity that isn't hidden, is the code needed to learn how to drive the computational-chemistry software. Learning this code is a skill that the students will need throughout their careers as chemists.

Portlets do not just make applications more user friendly. Future work with EaStCHEM will make more resources available to the chemists, such as those supplied by the National Grid Service. This will overcome difficulties in securing time on the resources available within the universities and will harness more computational power for the chemists.

Elsewhere, Rapid is being used in a wide range of areas, including brain imaging, seismology, and fire safety engineering.

Rapid is being used to design, build and deliver a portal to help brain imaging experts perform tasks using tools developed by the SFC Brain Imaging Research Centre.

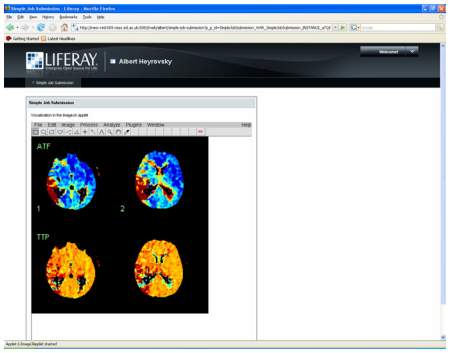
Likewise, the Rapid Portals for Seismological Waveform Data project aims to encourage the seismology community to use an application that analyses seismic waveform data.

Potential users tend to be put off by using the waveform application because they are daunted by the prospect of installing and even understanding the application. They also tend to worry about the difficulty in transporting large amounts of data needed for analysis, and how to visualise the results

.

Self-contained web portlets will be directly embedded in the community gateway and link to the data available in the Orfeus Data Center; the primary European centre for this kind of data. All the data and computing will be managed through Rapid, and the results presented via an existing web portal.

The development tool is also being used in the Firegrid project at the University of Edinburgh. FireGrid aims to establish a cross-disciplinary collaborative community to pursue fundamental research for developing real time emergency response systems, using the Grid, beginning with fire emergencies.



Deadline extensions for IEEE and JOGC

The deadline for submissions to the fifth IEEE e-Science conference has been etxended to August 7, while the editorial committee of the EGEE special issue of the Journal of Grid Computing has decided to extend the deadline for paper submission, which was previously set for July 31, to August 24.

The IEE conference will be held in Oxford, UK from Dec 9-11, coinciding with the UK e-Science All Hands Meeting that will be held from Dec 7-9th, 2009. More information can be found at http://www.oerc.ox.ac.uk/ieee

The EGEE project, in collaboration with the Journal of Grid Computing (JOGC), is preparing a special issue to be published in September 2010, with the theme "EGEE applications and supporting grid technologies." Submissions are invited for 15-20 page papers describing original and significant work involving applications on EGEE, and the use and development of key grid technologies. Interested authors should submit their papers to the JOGC specifying on the submission "Paper submitted for the special EGEE edition". Further information is available at http://tinyurl.com/lagwzk

Open Source: the next generation

By Simon Hettrick, OMII-UK.

The Google Summer of Code (GSoC) is a global programme that pays students to write code for open-source projects over their summer holidays. Since its inception in 2005, the programme has brought together nearly 5000 students and mentors from 98 countries – all for the love of open-source code. For the second-year running, OMII-UK has been selected to participate as a mentoring organisation.



One of the most successful of OMII-UK's GSoC students from last year was Michael Micelli, an undergraduate student from Louisiana State University. Michael took MapReduce, a programming model used to perform data processing on large clusters, and made it available to run through SAGA – OMII-UK's programming interface for Grid resources. '[Michael's work] was a very important proof of concept which showed that SAGA has the correct specificity to work with a number of applications' said Shantenu Jha, Michael's project mentor and head of the SAGA project. It appears that the reviewing panel for CCGrid thought so too, as they published two of Michael's papers on the subject.

GSoC is a win-win situation for the open-source community. For the mentors, the benefit comes from training the next generation of developers, as described by Shantenu 'Training people is the most important requirement for any successful project, so the GSoC is exceptionally useful'. The students get to partake in a real project, developing code that will be used by the open-source community. 'It was a great experience overall' says Micheal 'I wrote an application from scratch, and developed it all the way to a working state'. They also get to meet a wide group of student developers and work with some of the leading figures in the open-source world 'Working with Shantenu was great' says Micheal, 'he helps you to do things that you couldn't normally do'.

This year, the interest in GSoC has continued. There have been several thousand hits on OMII-UK's GSoC webpage and we've managed to attract roughly the same number of students as last year – even though the application process for students has become more involved. We look forward to the output from this year's students and we hope that they will all be as successful as Michael.

www.tinyurl.com/43ddpk



Forecasting UK city crime rates using the NGS

Nick Malleson, a final year PhD student at the University of Leeds is using the NGS to forecast crime rates in the city of Leeds through the use of sophisticated modelling of possible crime scenarios that would be impossible to perform on a desk top PC.

The occurrence of crime is a complex matter and can result from the interactions of a massive number of environmental factors as well as complex human behaviours. Nick uses an agent-based model in which many of these factors can be taken into consideration and modelled using NGS compute resources. An immense number of runs are required to provide sufficient material for analysis but each run was taking several days to perform on a desktop PC before Malleson switched to using the NGS.

Malleson explained "This is where the NGS was essential for the project to be feasible. Without access to NGS resources, the project would not have been able to continue. I can now achieve hundreds of results in a few days where I would have previously only produced one".

So far Malleson's model has thrown up some interesting results such as some areas of Leeds having much lower crime rates than expected despite environmental and social factors suggesting otherwise. The model will also be able to look into the future and predict how major development schemes in the city will impact on crime figures.

For example, the current EASEL regeneration scheme (http://www.leeds.gov.uk/Business/Regeneration/easel.aspx) in Leeds involves many



physical changes to the environment such as new houses, green spaces etc as well as social changes (rented and "affordable" accommodation etc) but what affect will this have on crime figures? Malleson's model could help to identify potential crime hotspots or other problems areas produced by these schemes that would be impossible to see otherwise.

You can read more about Nick's research in his blog - http://crimesim.blogspot.com/

Updated MyProxy Upload Tool

An updated version of the MyProxy Upload Tool has now been released for user comment. It is now available from the NGS Labs section of the NGS website under the title "VOMS enabled Certificate Management Wizard".

There has been an overhaul of the general layout of the tool which we would welcome feedback on. The main new feature is the capability of generating proxies with VOMS attribute certificates for those that need them.

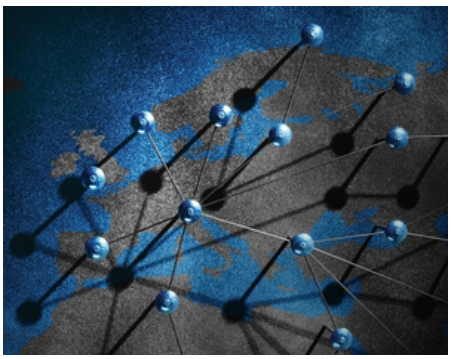
Please note that this should be considered a Beta rather than a production release and also that it depends on Java 1.6 (although we may be able to extend support to 1.5 shortly).

Please send any comments on this tool to David Meredith (david.meredith@stfc.ac.uk) and Will Rogers (will.rogers@stfc.ac.uk).

New login software

Are you looking for a new way to log into grid resources? The NGS at RAL are trialling a new system that allows users to login to grid resources using any SSH client, and we are inviting users to come along and have a play. Simply put, all you need to do is to create a MyProxy credential (easily done, with the MyProxy Upload Tool) and then use virtually any SSH client (such as Putty, WinSCP or Nautilus) to login to ngs.rl.ac.uk with the username and password to your MyProxy credential. It's a very simple way to get connected to a grid resource. Further details are available at http://wiki.ngs.ac.uk/index.php?title=MEG_User_Guide

First EGI council meeting held in Amsterdam



The EGI (European Grid Initiative) held its first Council on July 9, in Amsterdam, and appointed an interim EGI.eu director and an editorial board.

The Council took place after 24 NGIs (National Grid Initiatives) had signed the EGI Memorandum of Understanding (MoU).

Several important steps took place in the first quarter of this year, towards the establishment of the EGI: in January, the EGI Policy Board endorsed the EGI Blueprint, the document defining the roadmap for the construction of a new European grid organisation, and in March, the city of Amsterdam was chosen to host EGI.eu, the coordinating body for the EGI.

By signing the MoU, the NGIs confirmed that they are ready to implement the EGI, as defined in the Blueprint, and to commit financially to the establishment of EGI.eu.

Steven Newhouse, director of the EGEE project at CERN, has been appointed interim director, while Laura Perini, of INFN Italy and Charles Loomis, of LAL/N2P3, will form the editorial board. The editorial board will lead the preparation of the EGI proposals to the European Commission.

An EGI Steering Committee was also formed to supervise the proposal work and the formation of EGI.eu.

Digital Curation Centre team Face to Face meeting

The DCC held its six-monthly Face to Face meeting for all partners at Gleddoch House Hotel in Langbank on 8 & 9 July and managed to remember to take a staff photo. The regular meetings provide an opportunity for all members of the DCC from each of the four partners to get together, report on progress and plan the next period.



Standing L to R: Graeme Pow, Edinburgh (photographer); Graham Pryor, Edinburgh (Deputy Director); Colin Neilson, UKOLN, Bath (SCARP Project); David Giaretta, STFC, (Associate Director); Angus Whyte,Edinburgh (SCARP); Steven Rankin, STFC; Bridget Robinson, UKOLN; Alex Ball, UKOLN; Chris Rusbridge, Edinburgh, (Director); Martin Donnelly, Edinburgh; Manjula Patel, UKOLN; Sarah Jones, HATII, Glasgow; Joy Davison, HATII (Acting Associate Director); Matt Dunckley, STFC; Patrick McCann, HATII; Brian McIlwrath, STFC; Paul Lyden, consultant.

Seated: Sarah Higgins, Edinburgh; Liz Lyon, UKOLN (Associate Director); Florance Kennedy, Edinburgh; Emily Nimmo, HATII; Mags McGeever, Edinburgh.

IWPLS'09 International Workshop on Portals for Life Sciences

14 September, 09 10:00 AM - 15 September, 09 03:30 PM

New technologies and experimental methods in the field of life sciences have enabled researchers to create huge amounts of data in extremely short times. In turn, analysing this data demands a rise of compute power, storage, and transmission. An enormous number of complex and sophisticated algorithms and tools have been developed to aid research. The demand for compute power and storage is met with a rise in provision of infrastructure and High-Performance Computing facilities and has resulted in the popularity of novel utility computing infrastructures, such as grid and cloud computing. Despite the increasing available resources, the broad uptake of these facilities and infrastructures is not following the same pace. The main reasons for this lie in the complexity of the human-computer interface to these computing infrastructures. Partly these infrastructures are only accessible and popular in local or specific communities.

An approach to offer easy and intuitive access to computing infrastructures irrespective of their location is a portal. It is essentially a website specific for a community that shares resources, e.g., scientific data, scientific tools and underlying computing infrastructure. Through a portal, essentially a thin-client graphical interface, it is much easier to give a scientific community access to an underlying infrastructure as one can hide all the complexities of the technology involved.

The workshop aims to be a forum of discussion therefore we also encourage contributions describing work in progress and partial solutions. We invite the submission of papers related to various aspects of molecular and systems biology and portals.

IWPLS'09 will bring together scientists of the field of life sciences, bioinformaticians and computer scientists. The aim is to exchange experience, formulate ideas and introduce up-to-date technological advances in molecular and systems biology in the context of portals.

Additionally to participation at the workshop, there are three alternatives for researchers to present their work: talks, lightning talks, and poster sessions. Submission of a paper may result in a talk or a presentation of a poster, submission of an abstract may result in a lightning talk.

All abstracts will double-blind peer reviewed and evaluated on quality and relevance. Accepted submissions will be given a 10-minute time slot (plus 5 minutes for questions) in a lightning talk session. The schedule is planned with eight lightning talks.

Abstracts will be accepted electronically in PDF format only and are to be submitted through our EasyChair conference site. The abstract length must not exceed 200 words and the abstracts must be written in clear and readably English. As the workshop is hosted at the e-Science Institute in Edinburgh, every abstract will acknowledge support from the e-Science Institute. Every accepted submission must register and pay for at least one person. We expect one of the authors to attend the workshop and present the work.

All papers will be double-blind peer reviewed and evaluated on quality and relevance. All submissions are held in confidentiality prior to publication in the proceedings. Accepted papers will be published as open access, online-only workshop proceedings under CEUR-WS.org and will result in fully citable articles. The proceedings will be available online approximately one week prior to the workshop opening.

The presentation of accepted papers is dependent on number and quality of submitted papers. The workshop schedule is planned with the best 10 papers for oral presentation. Presenters will be given a 20-minute time slot (plus 5 minutes for questions) to provide a summary and update to their work. About 10 papers will be presented in poster sessions.

Papers will be accepted electronically in PDF format only and are to be submitted through our EasyChair conference site. They have to be formatted according to the layout style required by the OUP bioinformatics journal. Formatting requirements can be found at: http://www.oxfordjournals.org/bioinformatics/for_authors/submission_online.html. The paper length must not exceed eight pages and the papers must be written in clear and readably English. As the workshop is hosted at the e-Science Institute in Edinburgh, every paper will acknowledge support from the e-Science Institute. Every accepted submission must register and pay for at least one person. We expect one of the authors to attend the workshop and present the work.

Cluster 2009 Workshops

IEEE Cluster 2009 will hold four workshops in New Orleans, two on Monday, August 31, and two on Friday, September 4.

Monday: High Performance Interconnects for Distributed Computing, (contact: Ada Gavrilovska), and Workshop on Parallel Programming on Accelerator Clusters (PPAC), (contacts: Dimitris Nikolopoulos, Cal Ribbens)
Friday: Interfaces and Abstractions for Scientific Data, (contact: Rob Ross), and Web 2.0 on e-Research Infrastructure, Services and Applications, (contact: Mark Baker)

Each workshop will have its own requirements and schedule for submissions (but the registration fee for the workshops will be included as part of the conference registration fee). Workshops are expected to present work that is more cutting-edge and preliminary or that has more practical content than the more mature research presented in the main conference. Ultimately, the goal of the workshops will be to broaden the content of the week's presentations at IEEE Cluster 2009 by extending the topics of interest beyond those of the main conference. The workshops may also simultaneously deepen the week's presentations at IEEE Cluster 2009 by focusing on specific topics or bring together a critical mass of researchers in selected areas of interest.

More information is available at http://www.cluster2009.org/

Forthcoming Events Timetable

August			
18	eSI visitor seminar - Istvan Juhos	eSI	
31-4	Lie algebras, vertex algebras and automorphic forms	ICMS	http://www.icms.org.uk/workshops/lievertex
31-4	Cluster 2009	eSI	http://www.cluster2009.org/workshops.php
September			
3-4	SAGA Tutorial and training	eSI	http://www.nesc.ac.uk/esi/events/945/
8-10	Multicore Workshop - Exploiting Multicore Processors: Challenges and Programming Models	ICMS	http://www.nesc.ac.uk/esi/events/991/
14-15	IWPLS'09 International Workshop on Portals for Life Sciences	ICMS	http://www.nesc.ac.uk/esi/events/1000/
30-1	Mapping Information with and without Geography: Approaches to Data Visualization and Structure in the Arts, Humanities and Social Sciences	eSI	http://www.nesc.ac.uk/esi/events/1005/

This is only a selection of events that are happening in the next few months. For the full listing go to the following websites:
Events at the e-Science Institute: http://www.nesc.ac.uk/esi/esi.html
External events: http://www.nesc.ac.uk/events/ww_events.html

If you would like to hold an e-Science event at the e-Science Institute, please contact:
Conference Administrator,
National e-Science Centre, 15 South College Street, Edinburgh, EH8 9AA
Tel: 0131 650 9833 Fax: 0131 650 9819
Email: events@nesc.ac.uk

This NeSC Newsletter was edited by Gillian Law.
Email: glaw@nesc.ac.uk
The deadline for the September 2009 issue is August 21, 2009