SCIENCE AND IMPLEMENTATION PLAN

Introduction

The cross-cutting review of science and research undertaken by the Government in 2000 identified e-Science as one of three priority research areas, allocating a total of £98M for this programme across the Research Councils. This funding will allow scientists in the UK to build on existing world-class strengths and to develop the knowledge and expertise to build the new industries of the future.

The Science Budget defines e-Science as *'science increasingly done through distributed global collaborations enabled by the Internet, using very large collections, terascale computing resources and high performance visualisation'*. Strategically, it is recognised that many areas of leading-edge science are facing major challenges in the processing, communication, storage and visualisation of ever increasing amounts of data. Solving these problems will be a global effort, and if the UK is to stay at the forefront of many key disciplines it needs to invest in the relevant technologies and infrastructure. Within this programme, the way forward will be through investment in solving the problems of individual disciplines and in generic core technologies common to all disciplines.

EPSRC has been awarded £15M. support for an e-Science 'core programme', which *'will work across all of the different Council activities, developing and brokering generic technology solutions and generic middleware'*. This will be supplemented by DTI funding of a further £20M. This may be available for projects of generic e-Science that span the interests of more than one research council and must be regarded as a major funding opportunity for the NERC e-Science community, in particular to collaborate with industry, computer scientists and other science areas in the development of generic technologies.

As an outcome of the Science Budget 2001-4, NERC received £7M over the three years to engage in the cross-Council e-Science initiative. This was augmented in SR2002 by a further £8M over 3 years, from which NERC has funds to help support the modelling and data assimilation needs of QUEST (Quantifying and Understanding the Earth System) and those of other NERC programmes. QUEST is a new programme which will use an early focus on carbon budget and dynamics to both underpin international negotiations under the Kyoto Protocol and develop a structural framework for addressing science priorities which require an Earth System approach.

Objectives

NERC promotes and supports Earth System Science. It has a five-year strategy to prioritise and deliver world-class environmental science to support the development of Earth System models. The aim of these models would be to improve both the understanding of the complex interactions within and between the biosphere, the geosphere, the atmosphere and the hydrosphere and the ability to confidently predict and provide sustainable solutions to environmental problems.

e-Science is seen as an important and timely initiative on the road to realising this objective, since it offers opportunities for addressing the issues that are currently impeding progress. Progress could be accelerated by enhancing interactions between the science and information technology community, using the novel computing techniques offered by Grid technologies to address excellent science questions. NERC e-Science investments will provide an opportunity to do this.

The over-arching objective is that the e-Science funding should be used to underpin high quality science consistent with NERC's mission.

Framework

- In terms of research themes, NERC is not prescriptive and interprets the government's definition of e-Science widely.
- Whenever and wherever appropriate, the NERC programme should seek collaboration with end users in the private sector, particularly small and medium sized enterprises (SMEs).
- The NERC programme should also seek collaboration with other Research Councils and government departments having relevant programmes wherever and whenever appropriate.
- There is no expectation that NERC e-Science funding will support the provision of high performance computing. (EPSRC has received a separate funding line towards the procurement of high performance computing capability for the entire e-Science programme.)
- Similarly, there is no expectation that NERC e-Science funding will support the provision of research infrastructure to ensure GRID connectivity. Funding lines for this activity will be through the Science Research Investment Fund (SRIF) for HEIs and the Research Centre Infrastructure allocation for NERC Research Centres.
- As well as improving understanding and predictive capability, the NERC e-Science programme must focus on producing deliverables as the basis for the solution of problems. As such, it is vital that the NERC programme has demonstrable deliverables by year 3 at least and preferably by year 2.
- e-Science funding opportunities for bioinformatics exist within the NERC Environmental Genomics programme. The expectation is that applications in this area will be addressed by this funding line.
- Reflecting the cross RC management of the programme, information on bids is shared across RCs.

Programme Challenges and Deliverables

Four key topics have been identified:

- **Scaling : from the molecular to the global**

Techniques should address, or should have the potential to address, issues on a global scale. An obvious methodology for addressing this scale is the model. The key challenge in this area is that of scaling observations; moving both up and down scales. Processes modelled are frequently non-linear. Additionally, in aggregating data, researchers face problems in combining data with varying spatial and temporal scales in a meaningful manner. Another challenge was seen as defining the relevant 'landscape' scale for the model and using the right model/methodology to accommodate observations that may range from point data to estimates or observations on a global scale. In modelling the fine scale on a global basis, two elements are required: nested models and/or massive computing power.

Deliverables:

- Methodologies for changing scales including (1) 'process understanding' and data brought together at one scale and (2) algorithms for large-scale computing.
- A tool to identify data sets (and models) at different scales.

- Identification of data/knowledge requirements in relation to scaling.
- Accessible Earth systems models

- **Improving access to data and models**

There is a need for a strong emphasis on the mechanics of improving access to data and on issues relating to metadata (descriptive 'data about data').

A fundamental problem is that NERC datasets are very diverse in nature and cover all spatial and temporal scales.

In defining the resources required to address the creation of a virtual NERC database whose data centre structure appears seamless to the outside world, NERC should take a lead in:

- Delivering an enhanced meta-data record of archived data
- 'dictionary' building
- building systems to translate data and link databases
- integrating computer and natural science communities

Deliverables:

Desktop access to multi-centre datasets/allied software coupled with an ability to integrate such data sets within three years, and comparable access and integration of models within 10 years.

- **Linking Data to Models**

This key theme is the issue of simplifying access to large heterogeneous datasets, given problems that include the disparate nature of the data involved (differing formats, and different ways of ordering the data etc).

Deliverables:

- the ability to generate a single query across multiple datasets (in different catalogues) returning both metadata and data.
- the ability to acquire large datasets in near real time (NRT)
- the automatic production of metadata, both by models, and where possible, by observing systems.

Links between academia, commercial activity, agencies such as the Met Office and other UK data centres; and external links to data from the US, Europe etc. will need to be improved.

To address the technical challenges, it is considered necessary:

- that it would be desirable to farm ensemble members out to idle machines in an automatic manner
- that metadata needs to be comprehensive
- there is a need to address policy issues, protecting access and intellectual property while facilitating collaboration etc.
- that in an e-Science environment there would be a necessity for a "central

broker" of expertise to achieve efficient e-science development and synergy between NERC-funded projects.

- That there is a requirement for visualization tools to deliver NERC understanding of Earth Science (ES) to the outside community. Delivering both knowledge and data to the community would improve public understanding of science.

A non-technical issue to be resolved is how the commercial sector can be involved in NERC science (in particular commercial data users)

In terms of scientific challenges, e-Science should address the grand challenge of "*Understanding the combined role of physics and biogeochemistry in understanding, simulating, and predicting climate system variability"* with application to seasonal prediction, climate prediction and environmental monitoring (at regional and global scales). Such activity should be undertaken in synergistic partnership with other relevant agencies, such as the Met. Office and the Environment Agency. Validation of models/data and assimilation of heterogeneous data into an Earth system science model would help to achieve this objective.

Suggested e-Science projects:

- Further development of marine and atmospheric data assimilation leading to the assimilation of complex biogeochemical data.
- Production of a NERC universal data and metadata catalogue providing open access to data.
- A gridified model framework, including automatic capture of model information into metadata.

All projects should ensure that requirement against the E-Science core programme are well articulated. Equally, it will be important to ensure that information flows from other GRID projects to NERC and vice-versa.

Resource issues:

- The need for clarification on the eligibility of infrastructural elements (computing, networks, database development, development of grid interfaces to NERC databases) under the NERC e-Science programme.
- access to training and central support. Two levels of training are considered necessary; training to help develop new grid tools and training to use new grid tools.
- Easier network access to Met Office data
- People. Given the demands on UK e-science, there will be considerable competition for staff.
- Collaboration tools and resources of necessary capacity. may be core development.

Within the three-year timeframe of the programme, NERC should also demonstrate what can already be done, communicating existing knowledge and distribute existing techniques outside the NERC community.

- **Model to model coupling**

With the vision that Earth system modelling would answer the key question 'how is the global environment going to evolve over the next 100 years', the group suggested two possible solutions: the ensemble approach and a coupled first principle model.

A climate prediction study as an e-Science programme is envisaged. The example discussed involved running the Hadley Centre HADCM3 model (with perturbations) up to two million times, using as data input observations collected between 1950 and 2000 and models to guide forecasts. Analysis would be a distributed exercise with output 'posted' across the Internet. It is considered that this novel approach would predict both the most likely climatic outcome and extreme events. The technical issues associated with this experiment would include the scheduling/control of inputs/outputs, model output interrogation and distributed analysis, and model visualisation. Given the size of the experiment, this would be an excellent use of GRID technology.

Deliverables:

- Code (which must be portable and compiled to agreed standards)
- Documentation on how to use the coupler
- The development of distributed centres
- Mining and the visualisation of model output
- Making the coupler both efficient and user friendly

In terms of resources in years 1 to 3, these are seen as a core team of up to five people and a distributed team of a similar size. Deliverables in this time scale are seen as one to three demonstrators, each addressing a pressing problem and involving communities under represented in present GCM activity.

This approach would both advance UK capability in this area and also provide a technique/tool that was applicable across many fields of modelling (socio-economic, business, environment etc.)

Funding Modes

The NERC programme will be open to the entire NERC community (HEIs, NERC Research Centres and academic analogues). NERC anticipates that this will generate truly inter-disciplinary programmes and consortia, particularly new partnerships.

NERC has identified a need for the development of a programme focused on a few large projects, rather than a more diffuse activity. In common with other Research Council programmes, the aim is to support consortia-based proposals from partnerships involving application scientists and engineers, computational scientists and information technology experts including those from industry and commerce.

These consortia are expected to focus on real world challenges and should be able to demonstrate, within the span of the programme, both solutions to scientific issues and working systems of Grid technologies. Proposals should be based on open source software systems (e.g. Globus).

The NERC programme will provide e-Science opportunities in the following areas:

- Major awards
- Studentships

- Integration awards: Consideration is also being given to the provision of integration awards. If implemented, these are envisaged as modest awards, which address NERC-science specific core technologies not addressed by the generic technologies element of the e-Science core programme.

Data Management

NERC, along with the other research councils, regards the data generated from the research it funds as a valuable long-term, public-good resource. To ensure the data can be fully exploited in support of the research that they were collected for, and to enable them to be available for effective, longer-term, post-programme exploitation, it is NERC's policy that data must be managed effectively from the time of generation onwards.

NERC has established a network of Designated Data Centres that are the exclusive points of deposit for and management of NERC data. These centres are responsible for ensuring that data generated via NERC funded research are properly managed. They provide within programme advice and support in data management and crucially support the long-term, post-programme management of data for the public-good. It is NERC policy that copies of any data generated from NERC funded research must be offered to the data centres.

Whilst NERC supports long-term, post-project data management through direct funding to the Data Centres, the funds for undertaking the management of data during the life of the programme are included within the budget of each scientific programme. It is the responsibility of the programme to resource and support data management activities during the life of the programme. Data management is central to e-Science and so a key element of many proposals will be data management.

Ownership and exploitation of Intellectual Property (IP)

Ownership of IP and IP Rights arising from NERC awards lies with the grant holding body (university, NERC Centre/Survey or other NERC-approved institution). Whilst all recipients of NERC funding must endeavour to ensure that the outcomes of their research are used to the advantage of the UK, NERC retains the right, for a limited period, to exploit IP in partnership with grant recipient organisations to the benefit of the UK and the organisations. This is to avoid circumstances where fragmentation of IP would reduce the likelihood of exploitation (eg for datasets collected by several research groups).

Programme evaluation

Each project will be evaluated for scientific and technical excellence. Where links are proposed to particular programmes, such as QUEST, the strength and importance of these links will also form part of the evaluation. The overall evaluation of the NERC programme follows the overall e-Science programme.

NERC's evaluation plan for the programme is attached below.

Programme Responsibilities

The NERC e-Science **Steering Committee** is responsible to NERC and the Core Programme for directing and integrating the Programme science, for recommending applications to be funded, and for promoting the transfer of outcomes to the user community.

The NERC ***Programme Manager*** is **Dr Andrew Kaye** , Tel: 01793 411977, e-mail: aka@nerc.ac.uk.

Programme administration and co-ordination is the responsibility of the ***Programme Co-ordinator,*** **Dr Frances Collingborn,** Science Programmes, NERC Polaris House, North Star Avenue, Swindon SN2 1EU. Tel: 01793-411737; e-mail: frco@nerc.ac.uk.

NERC e-SCIENCE EVALUATION PLAN 2004

	OST OBJECTIVES	NERC ACTIVITY	EVALUATION MECHANISMS
1	To solve major research challenges in processing, communication and storage of very large volumes of valuable data	To support e-science activity within the NERC community and cover the major areas of NERC remit	▪ Number & quality of applications ▪ Number of awards ▪ Outcomes – case studies ▪ Number of community engaged ▪ RAE rating of PI CoI institution
2	To provide generic solutions to meet the needs of individual disciplines and applications	To successfully demonstrate the potential of e-Science and GRID technology across the NERC remit	▪ Number of published papers ▪ Number of grid enabled solutions developed (software and middleware)
3	To establish best practice across disciplines	Ensure co-ordination between relevant groups funded by research councils	▪ Number & level of people trained ▪ Interactions between NERC e-Science and industry/academia ▪ Attendance at training & research events ▪ PIs and CoIs – numbers associated with 5* departments ▪ Survey of community and stakeholders.
4	To provide infrastructure and facilities needed for the next major stages of international collaborative research	Support infrastructure necessary to ensure exploitation of the UK investment in e-Science Establish National Institute for Environmental *e*Science	▪ Usage of facilities ▪ Provision of e-science expertise in NERC Centres/ Key Universities ▪ Number & importance of international partnerships
5	To build on the leading international role established following SR2000	Engage in international collaborations across science based projects	▪ Number, value & importance of international partnerships ▪ Publications with an international partner
6	To engage with industry	Attract industrial funding into the UK programme NIEeS industrial outreach programme	▪ Number & value of commercial partnerships ▪ Engagement with industry/government/end users ▪ SBRI take up