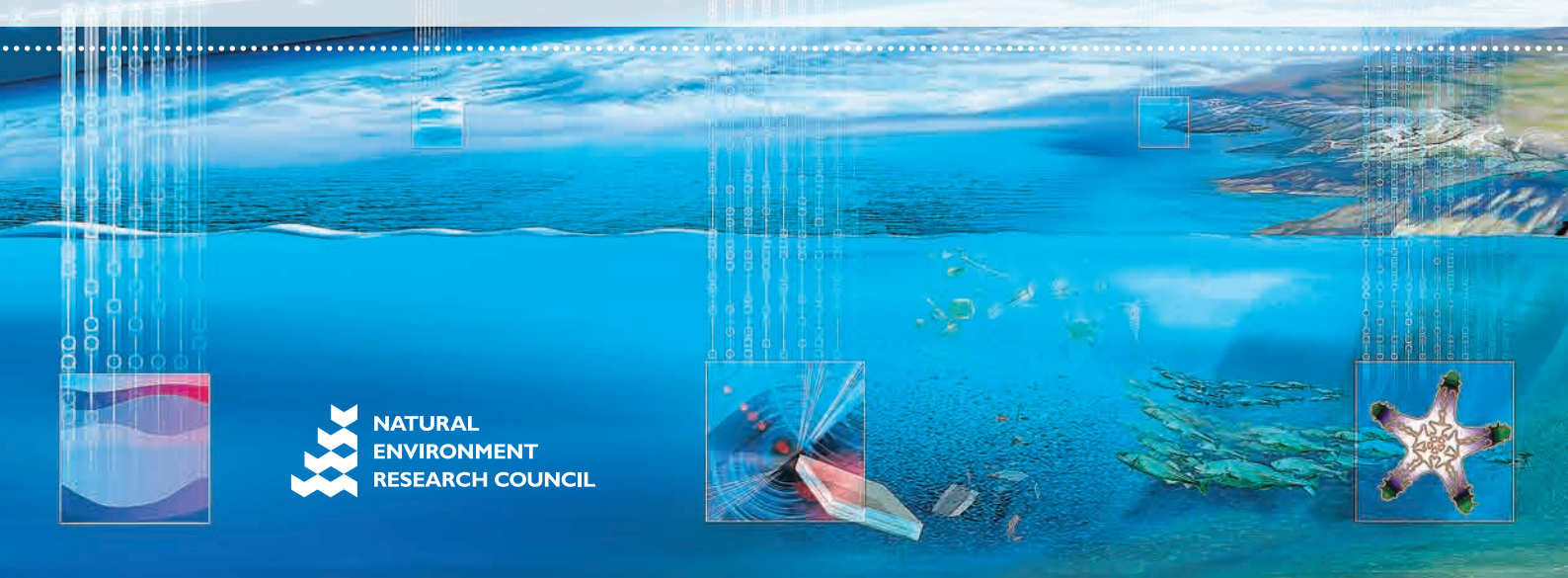




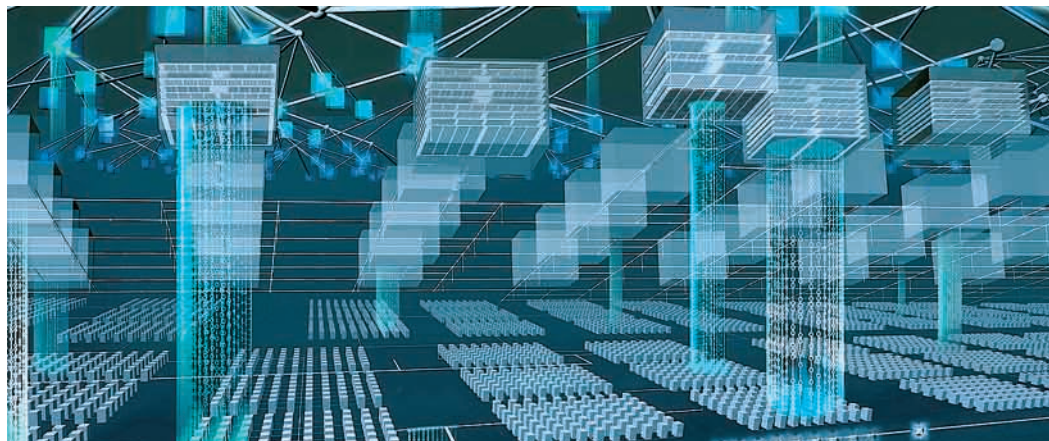
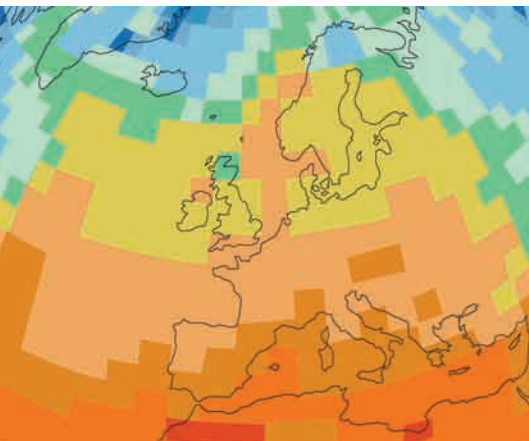
eScience

Harnessing the power of the internet for environmental research



Contents

Introduction	2
Projects	
Environment from the molecular level	4
Do-it-yourself climate prediction	6
Visualisation	8
Integrated Earth system modelling	10
The NERC DataGrid	12
Coastal modelling	13
Sites for all species	14
From weather forecasting to climate prediction	15
eScience centres and training	16
Knowledge exchange and industry engagement	18
Studentships	20
The future of eScience	21



Glossary

Computer cluster

A group of computers working closely together, sharing data and resources, and acting as one larger computer.

Ensemble

Running a computer model many times, each time slightly varying one or more parameters.

Metadata

Data about data. Data that describes datasets, allowing others to find and evaluate them.

Middleware

Software that allows separate computer systems and programs to communicate, often via the internet.

Data Grid

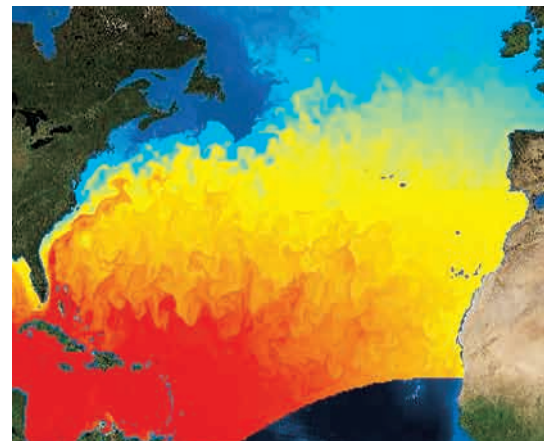
A system that uses the internet to share and manage large amounts of data held on different computers.

Grid computing

Using the resources of many separate but connected computers (usually via the internet) to simultaneously solve a problem.

Access Grid

Using grid technologies to improve online audio and visual applications.



What is eScience?

The internet and other information technologies are revolutionising the way researchers work. They enable scientists to participate in global collaborations and have changed the way researchers access and use computing power, extremely large sets of data and experimental facilities.

The research community, driven by the UK's research councils, has developed new collaborative tools that are breaking down the barriers between institutions. These tools are contributing to new and better science on a global scale.

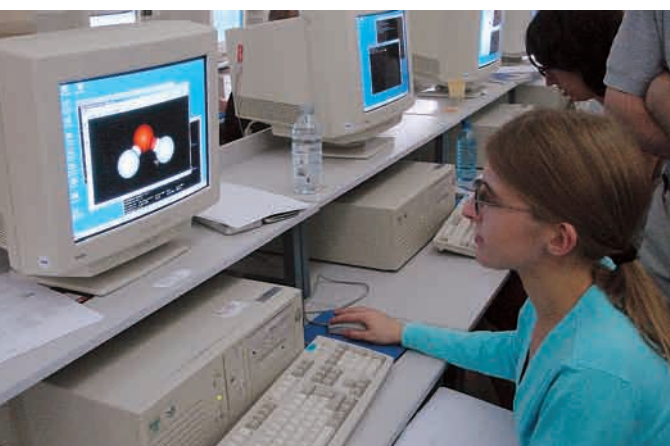
This new way of working is referred to as eScience.

The environment and the internet

The Natural Environment Research Council's (NERC) £14.2 million eScience programme is part of a broad cross council programme involving all the UK research councils. It has funded major collaborations between environmental scientists and computer scientists to deliver world-class environmental research. Their work is described in this document.

Technological advances are particularly important for environmental researchers. They have come at a time when society faces unprecedented challenges in responding to environmental change. Scientists need to produce better predictions of how the environment will change and improve how they quantify uncertainty. The programme has contributed to a shift in the understanding of uncertainty and significantly affected how science is turned into policy.

Much of the work described in this document is still under way, but the eScience programme has already pioneered new ways of doing environmental modelling and developed better ways of sharing environmental data. The research community must build on the successes of the projects and ensure that the new ways of working are integrated into NERC's future activities.



The eScience programme has changed the way we work in quite profound ways.

Professor Martin Dove, director of the National Institute for Environmental eScience

It's significant that the UK is the first country to develop a national eScience grid, which intends to make access to computing power and scientific data repositories as easy as the web makes access to information.

Tony Blair, October 2002

The Natural Environment Research Council

The Natural Environment Research Council (NERC) uses an annual budget of £400 million to fund independent environmental research. Our scientific priorities drive the marine, polar, atmospheric, earth, terrestrial and freshwater science communities.

We run a fleet of research ships and scientific aircraft and have bases in some of the world's most hostile environments. We manage the UK's Earth Observation budget, which uses satellite technology to monitor environmental change on a global scale.

Our research informs policy on local, regional, national and international levels. We collaborate closely with government departments and agencies, Research Councils UK, the Met Office, NASA, the European Space Agency and many other national and international organisations.

The NERC eScience programme

2001-2007

- £14.2 million
- 8 projects:
 - eMinerals
 - climateprediction.net
 - GODIVA: Grid for Ocean Diagnostics, Interactive Visualisation and Analysis
 - GENIE: Grid ENabled Integrated Earth system modelling
 - The NERC DataGrid
 - GCOM: Global Coastal Ocean Modelling
 - CATE: Creating a Taxonomic eScience
 - GCEP: Grid for Coupled Ensemble Prediction
- 2 research centres:
 - National Institute for Environmental eScience (NIEeS)
 - Reading eScience Centre (ReSC)
- 17 studentships
- 2 knowledge exchange projects



Environment from the molecular level

eMinerals

The eMinerals project used computer simulations to understand a range of environmental processes at the molecular scale. The work included analysis of how pollution moves through soils and groundwater; long-term encapsulation of nuclear waste; minerals in the inner Earth; and how minerals and fluids interact. The challenge was to exploit grid computing to manage the deluge of scientific data generated in such studies, and enable close scientific collaboration. Grid computing made it possible to run unprecedented numbers of tasks at once, requiring, for example, thousands of similar computations, with many different temperatures and pressures, or many chemical compounds.

Highlights

Arsenic in the environment

Arsenic often appears in minerals rich in iron and sulfur. Unfortunately, the presence of these minerals near man-made wells can lead to arsenic contamination of drinking water. The team worked out how arsenic is incorporated within a number of iron sulfide minerals such as 'fool's gold'. With this information they can identify the factors that lead to arsenic release in water.

This discovery raises the possibility of stabilising arsenic-containing iron sulfide rock by introducing additives that slow the rate at which it dissolves, thus reducing water contamination.

Toxic molecules in soils

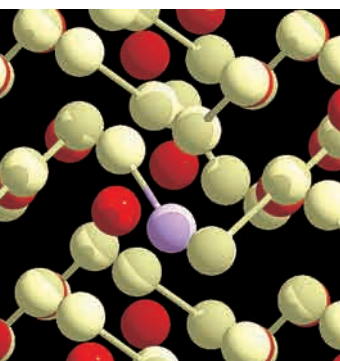
In the past, chlorinated organic chemicals have been used in industrial processes or by farmers in pesticides. Although many are now banned, they remain within polluted soils and groundwater. It is difficult to separate these toxic molecules from the ground they contaminate, but, by using grid computing, the team has a better understanding of the forces

governing how toxins interact with soils.

For example, dioxin molecules bind strongly to clay surfaces. But water competes with dioxins to bind to clay surfaces. The binding ability of a dioxin molecule is a balance between the binding strength of the dioxin to the clay surface, the water to the clay surface, and the dioxin to the water. Understanding how pollutant molecules such as dioxins are held within soils at a molecular level, using atomistic simulation tools, will eventually help develop effective land remediation strategies.

Nuclear waste encapsulation

The UK government wants to know the best ways to encapsulate high-level nuclear waste for long-term storage. The eMinerals team explored encapsulation within ceramic materials by running extremely large simulations of the atomic-scale processes that occur within radioactive materials when they decay. In some cases the ceramic materials remain damaged by these processes for incredibly long times, and in other cases they repair very quickly. Researchers now



Other research teams and institutions are adopting this system.

understand better the major factors that govern these processes, such as atomic bonding, allowing them to say which materials are better than others for nuclear encapsulation.

Why do some materials become softer when compressed?

Although scientists understand how many materials behave, they often come across examples that surprise them. One example is silica glass, which researchers have long known becomes softer when compressed. The *e*Minerals team only came to understand this counter-intuitive behaviour by running a large number of dynamic simulations helped by grid computing tools. The key finding is that the softening occurs because the inherent flexibility of the glass structure reduces when the sample is either stretched or compressed. This mechanism is applicable to a wide range of rock-forming materials.

Tools for grid computing

If scientists want to fully exploit grid computing, they need to submit jobs to multiple computers in different locations. To make this easy from a desktop computer, the team developed job submission systems that work with a wide range of grid resources and are linked to data management and metadata tools.

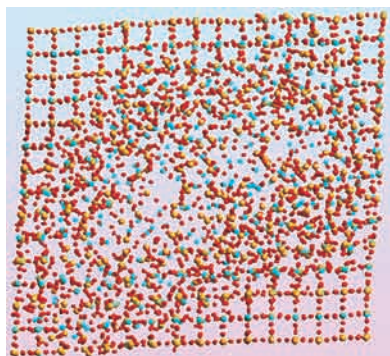
One of the particular features of grids is they allow sampling of model parameters on a scale not widely available before. This involves changing parameters very slightly, for example, temperature or pressure, or slightly altering the chemical under analysis and running the model again, many, many times. The new system has enabled scientists to set up and submit thousands of similar jobs using one simple command, and to use a second simple command to collate the core information provided by all these jobs and present the results to the scientist in an easy form for plotting or any other type of analysis. Other research teams and institutions are adopting this system, including the Diamond Light Source, the ISIS neutron scattering facility, and the UK's Central Laser Facility.

Making collaboration easier

The *e*Minerals project has focused on making collaboration between its members easier. At one extreme, the team has developed tools to enable scientists to extract and share the information hidden within data, particularly for cases where not all partners understand the format in which the data are stored. At the other extreme, they have worked to try to make collaboration between scientists in remote

locations as natural as if they were working in the next office.

The team has explored how scientists can use a wide range of collaborative tools, including desktop video conferencing, application sharing tools and social networking tools (developed by the team) for sharing ideas, new results, data, documents and visualisations.



Key outcomes

- A new understanding of how arsenic is incorporated within iron-bearing minerals.
- Increased knowledge of how soil particles absorb chlorinated pollutant molecules.
- An understanding of how materials for long-term encapsulation of high-level nuclear waste respond to damage caused by radioactive decay.
- Improved understanding of the way that minerals respond to large changes in pressure and temperature, including magnetic minerals in the inner Earth and the softening of some minerals under compression.
- Determining the dynamics of water molecules in the fluid state.
- Development of integrated computing and data grids with new middleware tools.
- Development of desktop grid tools, including tools for running exceptionally large numbers of tasks at once.
- Information delivery tools, particularly integrated with data grids.
- Development of a suite of XML tools and libraries for simulation sciences.
- Tools for automatic metadata analysis and management.
- Creation of the first large-scale UK campus grid.
- Collaborative tools, including tools to support desktop video-conferencing, and social networking tools.

Do-it-yourself climate prediction

climateprediction.net

The goals of the climateprediction.net project are to:

- harness the power of idle computers in homes and businesses to help forecast the climate of the 21st century;
- improve public understanding of the nature of uncertainty in climate prediction.

climateprediction.net is the world's largest experiment to try to produce a forecast of 21st century climate. To do this, the project asked hundreds of thousands of people around the world to download a climate model onto their computers. The climateprediction system uses the idle computing power of home PCs to run many thousands of climate models. For each run, they subtly altered approximations of features, such as clouds.

The project has been extremely successful: over 300,000 people to date have downloaded the climateprediction.net software, allowing much scientific research.

Highlights

BBC Climate Change Experiment

climateprediction.net teamed up with the BBC to launch the BBC Climate Change Experiment as part of the Climate Chaos season of programmes. Over 250,000 people took part in what was the largest climate forecasting experiment in the world. Participants were asked to download a 160-year experiment, from 1920 to 2080, forming a massive ensemble forecast taking into account the effect of greenhouse gas emissions, solar output and volcanic eruptions, as well as variations in the approximations that have to be made in state-of-the-art climate models. This has allowed the team to explore how

climate may change in the next century under a wide range of different scenarios. The BBC Climate Change Experiment still has a strong user

community and participant base, and continues to provide new results for analysis.

Climate change prediction – quantifying uncertainty

Climate models predict significant changes to the Earth's climate in the coming century. But there is a huge range in what they predict. How should we deal with this uncertainty?

If models are overestimating the speed and scale of climate change, society may end up panicking unnecessarily and investing huge amounts of money trying to avert a problem which doesn't turn out to be as serious as the models suggested.

Alternatively, if the models are underestimating



How can scientists assess and reduce this uncertainty?

the change, policy-makers will end up doing too little, too late in the mistaken belief that the changes will be manageably small and gradual.

To cope with this problem scientists need to evaluate their confidence in climate model predictions. In other words, they need to quantify the uncertainty in these predictions. By participating in the *climateprediction.net* experiment people from around the world have helped researchers to quantify uncertainty in a way that would not otherwise be possible.

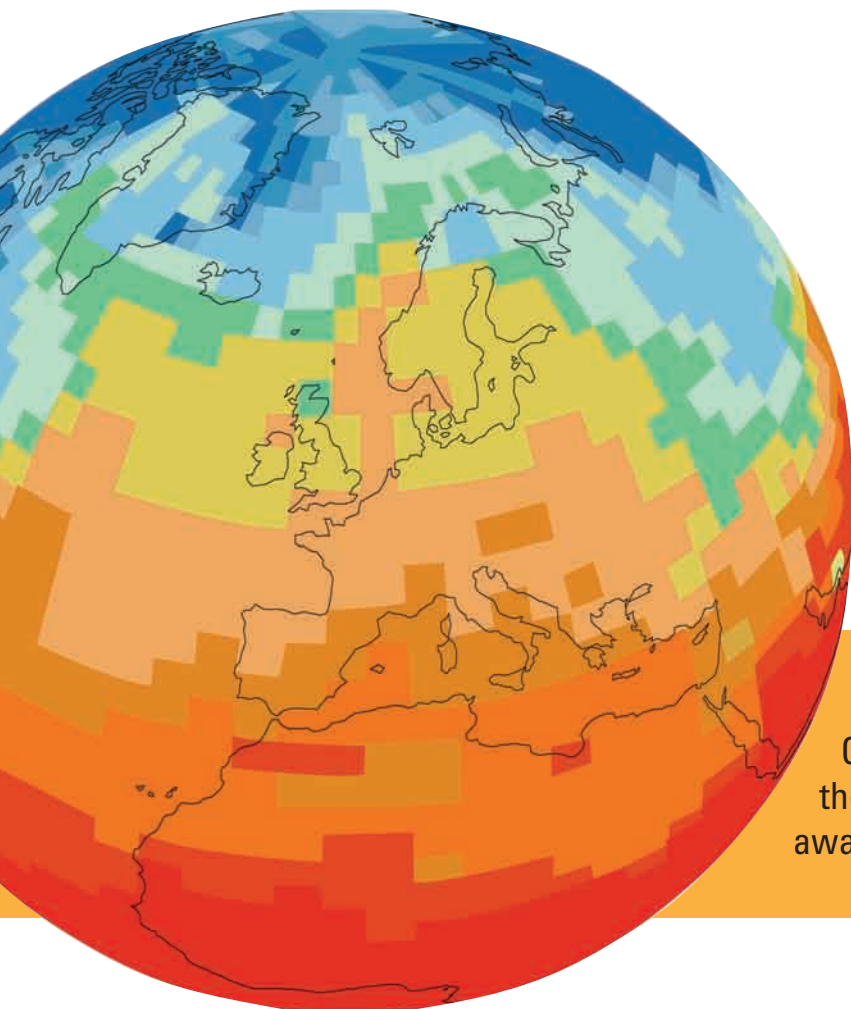
The results give a range of climate responses to increased atmospheric carbon dioxide. They also indicate that uncertainties in the model physics can combine in unpredictable ways to broaden this range even further.

Modelling very small-scale processes

Even with the incredible speed of today's supercomputers, climate models still simplify some small-scale physical processes, such as clouds.

Any one simulation requires a large number of interacting parameters and for many of these we do not know precisely which value is most realistic. This means that any single forecast represents only one of many possible ways the climate could develop.

How can scientists assess and reduce this uncertainty? One practical solution is to use distributed computing which combines the power of thousands of ordinary



Key outcomes

- A new way of modelling the climate using distributed computing.
- A massive data resource for scientists to quantify the uncertainties in climate change.
- Science, maths and geography teaching materials for schools.
- An Open University course to engage the public on the subject of climate change.

computers, with each computer tackling one small but key part of the global problem. By running many model simulations, participants in *climateprediction.net* have enabled the team to explore the effects of varying the values of approximately 30 key model parameters.

Ensemble climate forecasts

Instead of running one complex climate model with a given set of starting conditions, many researchers now believe it is better to run many, many models and slightly vary the starting conditions and other parameters. A key advantage is that the outputs from these ensemble models can give a likelihood or probability of the range of possible climates we can expect in the future.

Teaching and outreach

To help make participation in *climateprediction.net* more rewarding and fun, the team developed educational resources for participants to learn more about what their model is telling them. These include materials for schools, an Open University short course, and a lively, interactive web-based community where participants can compare, discuss, analyse and learn about their model simulations.

climateprediction.net is part of the BBC Climate Change Experiment team that won the Prix Europa, a major European media award, in 2007.

Visualisation

With the ever-increasing sophistication of computer models and the rapid growth in the number of Earth-observing systems, such as satellites, environmental scientists are facing a data deluge. In order to make sense of this large volume and diversity of information, scientists need new ways to explore data visually. Traditional plotting tools do not give the scientist the capability to explore large datasets interactively in three dimensions or easily to combine information from different sources to reveal relationships. NERC eScientists have developed several innovative methods to address these problems.

Highlights

Exploring virtual Earths

New technology is revolutionising how researchers view and share environmental data. Virtual globes such as Google Earth bring all kinds of geographical information, including satellite imagery, aerial photography and ground-based observations, within the reach of everyone, including the general public. Like their powerful (but complex) cousins in the world of Geographic Information Systems (GIS), virtual globes allow researchers easily to combine data from several different independent sources, revealing new information.

British Antarctic Survey researchers are using Google Earth for scientific expeditions on penguin foraging in the Southern Ocean. Scientists at the survey's headquarters in Cambridge can download data from tagged penguins, and combine this with, say, ship and satellite data of chlorophyll concentration on Google Earth. In the future, this system could help determine, for example, where the ship should go next, dictating the science in real-time.

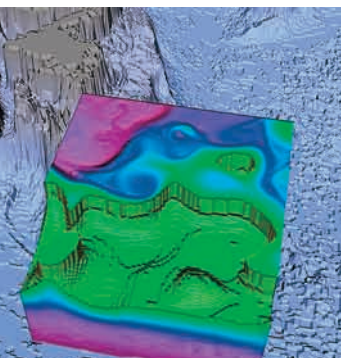
Walking the oceans

Imagine exploring the oceans by walking across vast ocean basins, following the currents and admiring the intricate

structure of the swirls and eddies around you. In the GODIVA project (Grid for Ocean Diagnostics, Interactive Visualisation and Analysis), scientists at the National Oceanography Centre, Southampton, developed the 'wandering cube' technique for exploring the vast datasets produced by the finely detailed global ocean computer model based at Southampton, known as OCCAM (Ocean Circulation and Climate Advanced Modelling). As the user 'walks' through the virtual ocean, a coloured cube displays the portion of the data immediately in front of the user. This technique allows the user to see the data from several angles, for example, the top and side views, simultaneously. The cube of data can be 'eroded' to expose its internal structure, revealing eddies and convection chimneys, giving scientists a truly three-dimensional appreciation of the dynamic ocean.

Browsing giant datasets on the web

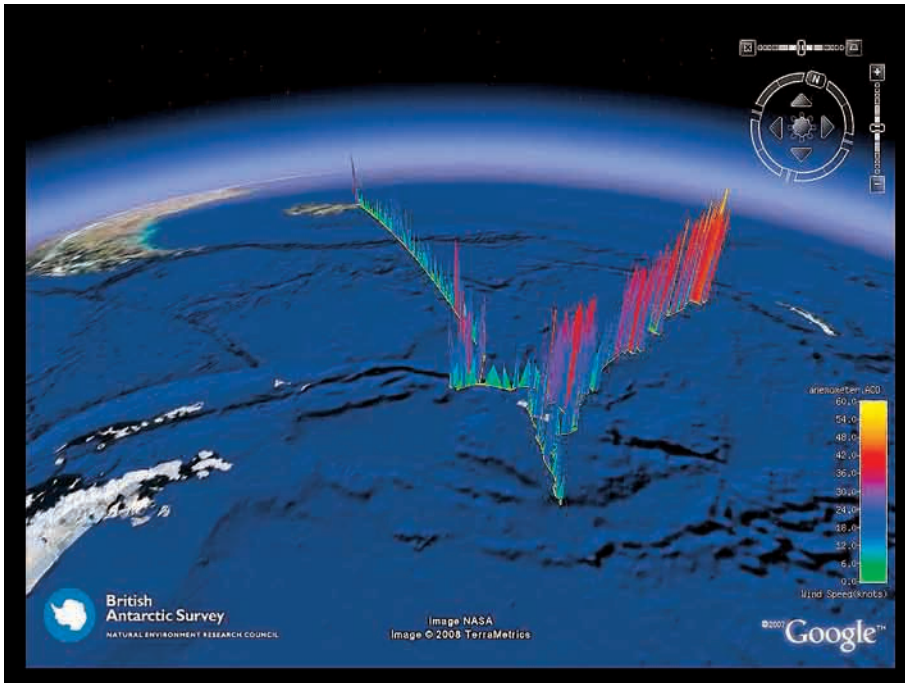
Almost everyone now uses the web to discover and download all kinds of information, but what if you want to hunt through terabytes* of environmental data to find just the feature you're looking for? Or maybe you simply want to get a quick picture of the data to assess its quality?



Imagine exploring the oceans by walking across vast ocean basins.

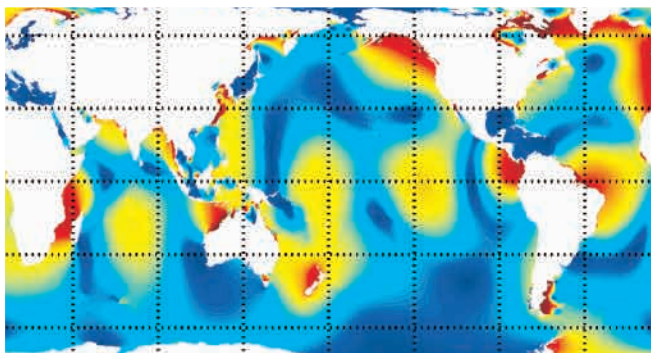
Key outcomes

- New methods for visualising large datasets online without the need to download them.
- Novel uses of virtual globes, such as Google Earth, for combining different datasets and comparing model predictions with observations.
- Bridging the gap between environmental science and Geographical Information Systems through open standards.
- Expansion of the community through special-interest groups.
- Web portals like Google Maps now show daily real-time operational oceanography data from the UK Met Office and several other centres around Europe. This allows easy exploration of the 4D-ocean structure.
- Techniques adopted in major European projects.



Downloading an entire dataset is time consuming and inefficient, but recent advances in web technology allow scientists to interactively visualise large datasets in their web browser – even from internet cafés in far-flung corners of the Earth. The Reading eScience Centre has developed a number of data exploration and visualisation websites, including the Godiva2 site, which builds upon technology developed in the GODIVA project. Godiva2 allows scientists to explore environmental datasets from numerical models and satellites through an interactive, zoomable online map. Other research groups are now using the technology where visualisation of large datasets is critical, including two European Union projects, MERSEA (Marine Environment and Security for the European Area Integrated Project) and ECOOP (European COastal sea OPerational observing and forecasting system).

*A terabyte is a trillion bytes or one thousand gigabytes.



Integrated Earth system modelling

GENIE: Grid ENabled Integrated Earth system modelling

The GENIE project's key challenges were to:

- develop a software framework for flexibly constructing computationally-efficient Earth system models;
- build more comprehensive Earth system models which include the carbon and other nutrient cycles, ice sheets and ocean sediments;
- use these models both to understand past climate change and to predict future change;
- improve confidence in model projections by taking statistically sound approaches to tuning models and better quantifying uncertainty.

The most sophisticated Earth system models include complex interactions between different components of the Earth system: the oceans, atmosphere, land-surface, ice sheets and life. Scientists use specific computer models to represent each of these components. Researchers working on the GENIE project harnessed the greatly enhanced power of distributed computing to develop a new Earth system modelling framework that allows much more flexibility in how these different components interact. This makes it easier to use different configurations of models to study different aspects of the Earth system.

Highlights

Past climates

Over the last 125,000 years a great ice age has come and gone. Understanding the causes of such natural changes in climate, and the corresponding variations in greenhouse gases, is an essential prerequisite to improving confidence in future predictions.

Together with NERC's Quantifying and Understanding the Earth System (QUEST) programme, the GENIE team is using their models and grid computing resources to simulate the climate and carbon cycle through the last ice age cycle. By comparing the models with past climate data, the team can

constrain how the models represent key processes, improving confidence in future projections.

Climate in year 3000

How much warmer will the world be at the start of the next millennium? Could Britain become a subtropical oasis with a reshaped coastline? These were questions addressed by the GENIE team in a study commissioned by the UK Environment Agency through NERC's collaborative centre the Tyndall Centre for Climate Change Research.

The study revealed that if society burns the known reserves



How much warmer will the world be at the start of the next millennium?

Key outcomes

of coal, oil and gas the world will get 6°C warmer with sea levels rising 3-4 metres. If unconventional fossil fuel sources are also exploited it could get over 10°C warmer causing a complete meltdown of the Greenland ice sheet, collapse of the global ocean conveyor, and over 10m of sea-level rise.

The team's models show that by globally achieving a 60 percent reduction in carbon dioxide emissions by 2050 followed by a complete stop in emissions by 2200, we might protect Greenland from melting and restrict global warming to 1.5°C.

To be or not to be circulating

Giant circulation engines in the atmosphere and ocean help drive and regulate Earth's climate. Some of the great circulation engines can be either 'on' or 'off'. The global ocean conveyor is the most famous example. But some recent studies have questioned whether it really has two stable states.

GENIE scientists have used computers from around the UK to map out the possible states of the ocean circulation in Earth system models of differing complexity. In all models there are certain combinations of ocean and atmospheric parameters for which both 'on' and 'off' states of the circulation are possible and stable. That is, for a given set of conditions in the atmosphere and ocean, the circulation could either be moving, or it could have faltered.

To the team's surprise, they found that in the more sophisticated models, where the oceans and atmosphere fully interact, this extends the range of parameters where the ocean exhibits two stable states. In these cases the model showed if the ocean conveyor is switched off, the atmosphere makes it harder for it to switch back on again.

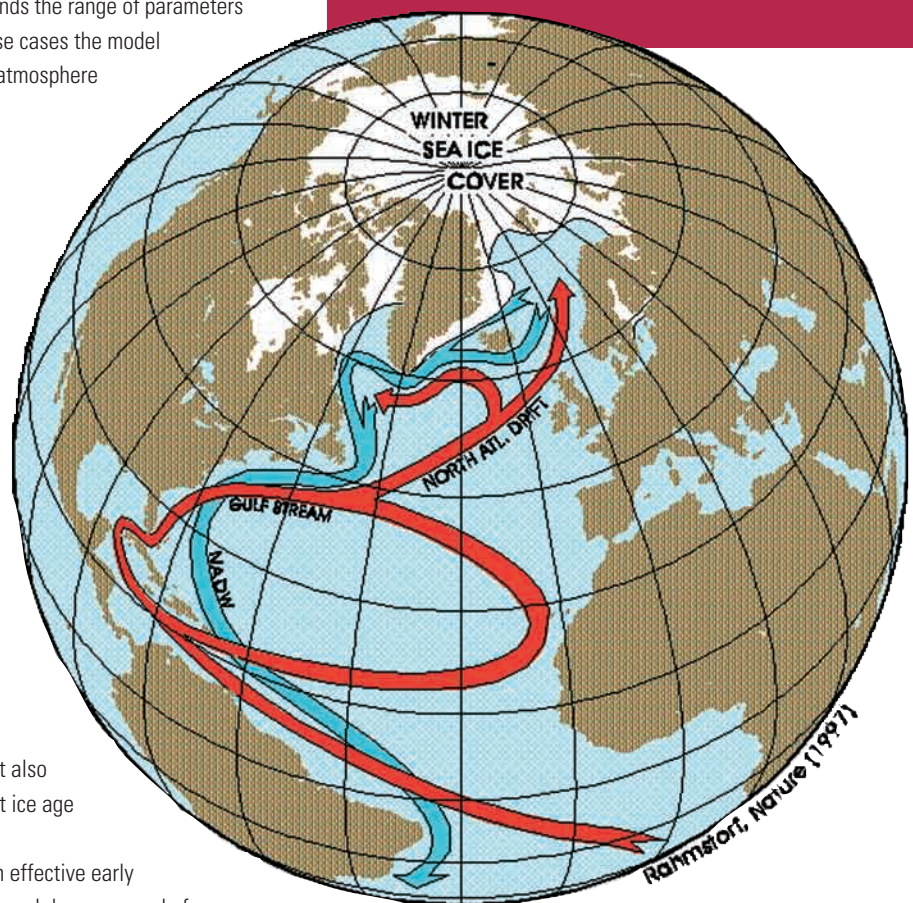
Anticipating the tipping point

Tipping points are when changes in one part of the climate system cause a domino effect, forcing shifts in other parts of the system, for example, the slowdown of an ocean circulation. Can we detect an approaching tipping point before we reach it? If they are detectable, an early warning system would be of obvious value to policy-makers and wider society.

In principle, an early warning is possible because a system approaching a critical transition becomes more sluggish in its response to perturbations, including natural variability. GENIE researchers have developed a method that extracts information about a system's response time from time-series data. In model simulations, the method successfully anticipates a collapse of the ocean conveyor belt. It also detects the transition that marks the end of the last ice age in ice-core records.

However, the GENIE model world reveals that an effective early warning system for the real world would require a much longer record of ocean circulation than currently exists.

- Simulations of climate and the carbon cycle over the last ice-age cycle.
- Projections of climate change and sea level rise on a millennial timescale.
- Quantification of uncertainties in future projections.
- New software to couple together Earth system models.
- Advanced tuning of Earth system models to more closely follow observations.
- User-friendly toolkit to run simulations and analyse the results.



The NERC DataGrid

Computer infrastructure for global access to NERC data

New technologies are driving an explosion in the volumes of environmental data available. From new climate models providing detailed predictions for multiple possible futures to new sensors providing more comprehensive and extensive measurements, the volume, type, and range of data are increasing beyond the scope of any one institution to manage and make available.

The NERC DataGrid team has developed the infrastructure to allow scientists to find, understand, manipulate and visualise data from many institutions around the world. Currently focused on atmospheric and oceanographic data, the evolution of the NERC DataGrid is expected to include all NERC science. Like the web, the NERC DataGrid has no owner or central control, data remains with data providers, and users can access all the data from a single location.

Highlights

Finding and using data

Before the NERC DataGrid, scientists struggled to find out what datasets existed and where they were held. If they could track down a dataset, they still needed to know enough about the data format to be able to use it. With data held in different places and in different formats, efficient exploitation has varied from difficult to impossible.

The NERC DataGrid now makes it easy to find datasets at participating institutions around the world, understand where, when and how they were collected and produced, and select the most appropriate data for the researcher's own requirements. The data are then converted to a common format and downloaded and/or visualised from a single portal. The portal can guide dataset selection by helping with the search terms, for example, the results for a search on 'rain' can also include results for 'precipitation' datasets. It can also respect access restrictions at multiple sites without requiring the user to log in more than once.

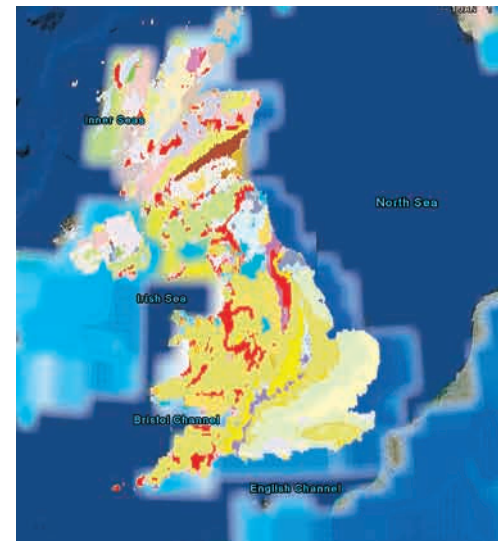
Setting common standards

Although initially limited to the atmospheric and marine sciences, the team developed the NERC DataGrid to allow it to expand to other disciplines and to be used internationally. By using and extending international standards, the NERC DataGrid provides access to some major data holdings in the United States, Germany and Australia. NERC DataGrid technology will also be integral in exploiting NERC data in the context of the new European spatial data infrastructure INSPIRE.¹

¹ *Infrastructure for Spatial Information in Europe.*

Key outcomes

- New techniques, middleware (software that allows applications to communicate with each other, often across a network or the internet) and application software to describe and manipulate data in a distributed manner.
- National and international leadership in data access technologies. This has included setting the agenda within the atmospheric and marine science communities to get networks, operating systems and applications working together effectively and developing technical modifications to international standards to allow efficient data exchange.



Coastal modelling

GCOM: Global Coastal Ocean Modelling

What role do shallow seas play in the global climate? This is a question that global climate models struggle to answer because these seas are too small and too shallow. The Global Coastal Ocean Modelling project brings together oceanographers, ecosystem modellers and computer scientists from four leading UK research institutions to address this challenge.

Why are shelf seas so important for the Earth system? Climate change will alter the marine environment. Coastal areas are where it will most dramatically affect people's lives: 37 percent of the world population live within 100km of the sea. Shelf seas are regions of exceptionally high biological production. Around 14-30 percent of the marine plant growth – the base of the marine food chain – occurs in shelf seas. Furthermore, everything that travels between the land and the deep ocean goes via the continental shelves, be it nutrients, pollutants, carbon or fresh water. On the way, they get mixed physically and transformed chemically.

What makes shelf seas so difficult to model?

While shelf seas only occupy seven percent of the ocean by area, they require 70 times the computer power than the deep ocean because of the complex small-scale processes that occur there, such as currents, tides and mixing. Because of this, we need oceanographers and computer scientists working together to find new ways of looking at shelf seas to increase knowledge of the whole Earth system.

Highlights

Building models of all the continental shelves

For a particular shelf sea, such as the North Sea, researchers have traditionally needed specific models developed from scratch. This often means spending months gathering and configuring the data for each new model shelf sea. The team has developed an automated system to set up models of continental shelves anywhere on the planet, drawing on global datasets of parameters such as water depth, weather conditions and river run-off. This shortens the processes to a matter of minutes, allowing researchers to quickly build models of the 60 or so shelf seas around the globe.

Connecting models of seas together

Most continental shelves (and their related shelf seas) are connected to other continental shelves. What makes the

GCOM system unique is that the various shelf-sea computer models can talk to each other, exchanging information. This allows scientists to make integrated predictions of how coastal regions will respond to environmental change.

Connected supercomputers

Each of the shelf-sea models requires the massive computer power available from large clusters of computers. The team are developing grid-computing tools to spread the considerable workload without affecting the other users of those machines. The tools are easy to use and it looks like the model is running on the user's desktop, even though it is actually running on a remote machine. Large outputs (terabyte per job) are automatically transferred to the scientist's institute without accumulating on the remote computing resource.

Key outcomes

- New ways to explore the effects of climate change on the world's major fisheries.
- Quantification of the effects of Saharan dust on plankton growth in Atlantic up-welling regions off the coast of west Africa.
- Better integration of ocean and ecosystem models.
- New grid computing tools for more efficient model simulations and shared use of computing resources.



Sites for all species

CATE: Creating a Taxonomic eScience

Taxonomy is about the description and identification of all species, their geographical distribution and their relationships. There is a growing and increasingly urgent demand for detailed and reliable taxonomic information on species to increase knowledge of ecosystem function and to understand the consequences of losing biodiversity.

But achieving a standard classification of plants, animals and other organisms has been greatly hampered by fragmented underlying information scattered across the globe in myriad institutions and an enormous literature stretching back hundreds of years. But now the internet is revolutionising taxonomy, making long-distance collaboration dynamic and productive and offering the prospect of something radically new: an agreed classification of organisms which continually improves as research and discoveries advance.

The CATE project is starting this process by publishing online full revisionary taxonomies of two groups: the hawkmoths (insects) and the aroids (flowering plants). Crucially, it will enable wiki-style continuous updating coupled with community peer-review. The online approach allows closer collaboration between taxonomic experts, both professional and amateur, and fundamentally changes the relationship between providers and users of taxonomy.

Key outcomes

- Integration of taxonomists, both amateur and professional, to produce standard online classifications of two major families in rich, illustrated detail, with online identification tools.
- New software that allows update of existing information and peer-review for new contributions.
- The system presents a current consensus taxonomy as well as alternative views.
- Linking the consensus taxonomy to global inventories, such as the Encyclopedia of Life, bar-coding and other molecular sites and specimen databases.
- Exploring the key issues for taxonomic publication on the internet.



Highlights

Comprehensive taxonomy online

CATE-style classifications are unique in being web-based, comprehensive, updatable, peer-reviewed and quality-controlled by experts. The project contributes taxonomic content and technology to the European Distributed Institute of Taxonomy project (EDIT).

International collaboration

CATE has attracted collaboration from major taxonomic institutions in Europe and the USA, for example, the Berlin Botanical Garden and the Missouri Botanical Garden.

Branching out

CATE has secured additional funding from the Global Biodiversity Information Facility (GBIF) and the Royal

Botanic Gardens, Kew, to develop digital biodiversity identifiers and gather biodiversity information on the Araceae of Africa.

Encyclopedia of Life

CATE has been involved, along with the Global Biodiversity Information Facility and other projects, in developing a method of sharing species level information for global use, for example, by the Encyclopedia of Life project.

Compliant software

CATE's software is compliant with major *etaxonomy* initiatives such as the Solanaceae Source (tomato family) and EDIT projects, which widen its application.

From weather to climate prediction

GCEP: Grid for Coupled Ensemble Prediction

Weather forecasts start with observations taken around the world by the National Meteorological Services. Could climate researchers make forecasts a few years ahead in the same way, starting from the current state of the climate? Policy-makers and many private enterprises would gain immense benefit from knowing the probability of a cold winter or a hot summer, or a run of several unusually dry years.

Traditionally, climate models focus on external drivers such as changing greenhouse gases, aerosols and the solar cycle to predict changes on timescales of several decades or longer. The equations for the climate involve the oceans, ice and land surface, as well as the atmosphere. But, if they could all be started from accurate knowledge of the entire state of the Earth system today, could researchers predict shorter-term changes that depend on this initial state? This is a hard problem. Modellers need to improve the way data from different components of the Earth system are incorporated into global models.

The Grid for Coupled Ensemble Prediction (GCEP) project uses grid technology to harness the computing power required to run large numbers (ensembles) of the Hadley Centre's HadCM3 climate model. The aim is to develop new ways to incorporate observations of the Earth system and to investigate the effects of knowing the initial state of these slowly varying properties on climate predictions.

Key outcomes

- Running and managing many large climate simulations over multiple computer clusters accessible to most universities, without accessing high performance computing resources.
- Developing new ways of incorporating observations into climate forecasts from one to ten years ahead, for example, improved El Niño forecasts.
- Improving control of snow cover, soil moisture and sea-ice initial conditions in climate models.
- Studying the impact of changes in the Nordic Seas on the Atlantic thermohaline circulation and heat transport.
- Providing a framework for studying decadal prediction in the current changing climate.

Highlights

Disturbing the North Atlantic Current

The temperature in Western Europe is highly dependent on the heat carried by the North Atlantic Current. However, this current is affected by conditions in the Nordic Seas where it sinks and returns south. By perturbing the amount of salt in the Nordic Seas in climate models, GCEP scientists have shown how signals propagate down the coast of Canada and the US to control the strength of the North Atlantic Current. Better knowledge and observations of these processes could lead to more accurate decadal forecasts of Western European temperatures.

Ocean temperatures and El Niño forecasts

Getting the right ocean temperatures in the tropics, particularly the Pacific, is crucial to forecasting El Niño. Scientists working on the Grid for Coupled Ensemble

Prediction project have shown that in 1998 – the warmest year on record – El Niño could have been forecast as early as October 1996 using assimilation of subsurface ocean temperatures now available.

What if all the sea ice goes?

Sea ice affects how the atmosphere and ocean interact at high latitudes, so can current sea ice conditions tell us about future climate? Our model runs show that sea-ice recovers in one to two years, even from extreme perturbations like the melting of all sea ice in the summer months. However, prolonged warming of the oceans can lead to much longer-lasting impacts on sea-ice. This suggests a more passive role for sea ice than previously thought, and emphasises the importance of obtaining more measurements of Arctic water changes.

eScience centres and training

National Institute for Environmental eScience (NIEeS)

The eScience programme set up the National Institute for Environmental eScience (NIEeS) in Cambridge to concentrate on outreach and training. Through its workshops, events and training programmes the institute became a centre of national capability in environmental eScience.

The NIEeS team set up demonstrations of eScience technologies to train environmental researchers how to use grid tools that range from job submission through to data handling and visualisation. Test cases included earthquake prediction and ice cover computer codes. To enable this to work, the team developed a set of software that allows easy connection to Google Earth.

NIEeS supports a small number of environmental eScience working groups. One of these has helped bring grid technologies to geographic information systems (GIS).



Highlights

Geobrowsers

Geobrowsing technologies, such as Google Earth, present exciting possibilities to view geospatial data in a dynamic, interactive environment on the user's desktop or on the web. Workshops organised by the institute and ReSC brought together tool developers and scientists for both practical training and an exchange of ideas.

Multimedia resources and desktop video conferencing

Video conferencing from any desktop computer can improve collaboration, reduce travel time and minimise carbon emissions. The institute has contributed to a distributed computing system specifically designed to improve communications, including video conferencing, known as the Access Grid. The system makes it easier to coordinate large meetings, collaborative work sessions, seminars, lectures, tutorials and training.

eGenomics

The environmental genomics community worked with NIEeS to organise a series of workshops that brought together biologists and bioinformaticians. This led to great progress in several related projects, including the Genome Catalogue and the Genomic Standards Consortium.

Social networking for scientists: scispace.net

NIEeS has developed scispace.net, a social network site for scientists. It is like MySpace and Facebook, but tailored specifically for the research community's needs. It has features such as individual profiles and tagging to enable potential collaborators to discover each other, blogs, wikis, bookmarks, comments, shared filespace and visualisation tools to document the exchange of ideas and results.

Reading eScience Centre (ReSC)

The Reading eScience Centre develops and applies innovative eScience techniques to support a wide range of applications focusing on ocean and atmospheric science. Through demonstrator projects, often in collaboration with industry and government agencies, the centre is developing tools to help scientists to use eScience infrastructure to analyse very large environment and climate datasets that may exist in different locations around the UK or around the world.

Highlights

Innovative data visualisation

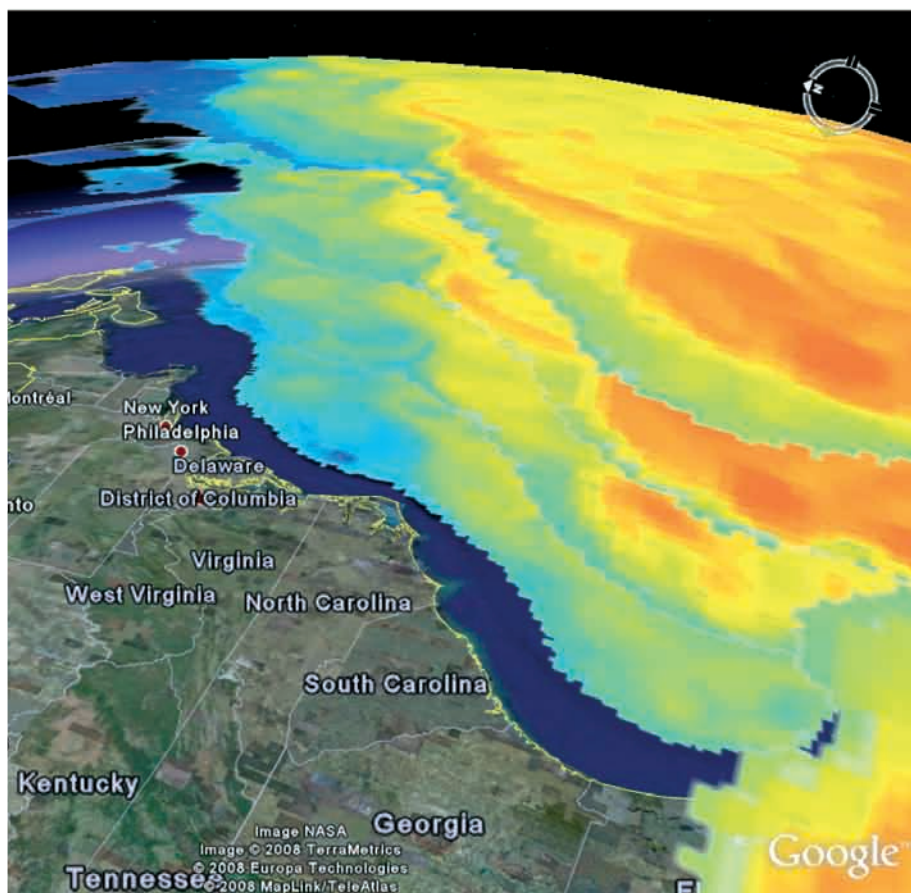
The centre has developed new technology to visualise large, four-dimensional environmental data in web browsers (page 8) and in geobrowsers such as Google Earth (page 16). This technology is being adopted by a number of international projects, and the centre is collaborating with Google to use Google Earth to advertise and distribute scientific data more widely.

Industry engagement

Two important services in the UK – the healthcare sector and the coastguards – could save more lives if there was a better system for providing the specific weather forecasts they need. The DEWS project (Delivering Environmental Web Services) linked the UK Met Office with NERC eScientists and software developers to produce such a data delivery system (see page 18). In 2007, the centre set up a partnership with Schlumberger to enable two-way knowledge exchange in the fields of environmental data management, visualisation and the prediction of hazards due to storms.

Making grids easier to use

The centre's G-Rex software ('Grid Remote Execution') makes it easier for scientists to run complex, data-intensive simulations, such as climate simulations, on



grids of computing clusters. Other eScience projects (GCEP on page 15 and GCOM on page 13) are using this software to run climate and coastal-ocean simulations on a set of computer clusters in various NERC institutes.

Promoting open standards

Effective sharing and visualising of environmental data require data providers, software developers and users to adhere to common geographical information standards. ReSC is working in collaboration with the NERC DataGrid (page 12) to promote the use of open international standards in order to make it easier to integrate scientific data with other environmental information and to make it more usable and accessible by decision makers in government and industry.

Knowledge exchange and industry engagement

The NERC eScience programme has delivered both an improved scientific understanding of the environment and new state-of-the-art techniques for managing and delivering science. But many of the new techniques and tools have applications beyond environmental science in other research fields and in industry. Scientists must work together with industry and government to ensure that this knowledge is widely exploited for the benefit of society. A number of collaborative projects have already emerged.

Saving lives at sea and at home

Many industries and services need accurate forecasts of the weather and the oceans to make decisions. For example, when a yachtsman falls overboard, coastguards use forecasts of the winds and ocean currents to predict the drift patterns of both the stricken sailor and the yacht, enabling a quick rescue. The DEWS project (Delivering Environmental Web Services) brought together the Met Office, NERC eScientists and software developers to produce a data-delivery system that delivers forecasts to the maritime and the healthcare sectors. This is particularly useful for mitigating the effects of medical conditions such as Chronic Obstructive Pulmonary Disease. The condition accounts for five percent of all deaths globally (as many as Aids) and is worsened by specific weather conditions. The DEWS project allows forecasters to provide adequate warnings of approaching conditions that could be deleterious to health.

Scientists can apply the same technology to other cases such as flood prediction and oil-spill mitigation.



Richard Cooke/Alamy.com



Developing new materials

The MaterialsGrid project aims to create a unique database of material properties. This will accelerate the development and use of new custom materials for products such as medical sensors and aircraft. Scientists will use a web portal to search the database rapidly, receiving pre-computed properties where available, or triggering new calculations on the grid that fill in missing data. Users will be able to donate data and grid resources in exchange for access, ensuring the continued growth of the service

The MaterialsGrid project exploits both the grid tools and the knowledge generated in the NERC eMinerals project. It is a collaboration between Accelrys Ltd., IBM, the University of Cambridge, the Science and Technology Facilities Council (Daresbury Laboratory) and the Johann Wolfgang Goethe Universität Frankfurt.*

**Funded by the Department for Business, Enterprise and Regulatory Reform, formerly the Department of Trade and Industry.*

Millennial scale climate prediction

The Environment Agency used the GENIE team's projections of millennial timescale sea level rise to inform thinking about coastal flood risk and long-term waste management.

From research to policy

GENIE scientists contributed to the Stern Review on the economics of climate change with their projections of long-term global warming.

Climateprediction.net and GENIE submitted large ensembles of climate simulations to the UN's Intergovernmental Panel on Climate Change Fourth Assessment Report published in 2007. These models captured the uncertainty in future projections. Both these reports will form the basis of global climate change policy in the coming decades.

Support from Microsoft

The GENIE project has received support from Microsoft to create a virtual framework that links all applications and data, regardless of the system used to create them. The solution is based on Windows Compute Cluster Server 2003, Microsoft® .NET technology, Windows Workflow Foundation, Windows Presentation Foundation and a Microsoft SQL Server™ 2005 database to manage data.

Influence and leadership

One of the by-products of the development of the NERC DataGrid has been the influence on the software architectures deployed within the Met Office and elsewhere in the atmospheric and oceanographic community. The NERC DataGrid architecture is heavily based on the geospatial data access standards defined by the Open Geospatial Consortium (OGC), and, while these standards had existed prior to the NERC DataGrid, they had been largely ignored by the atmospheric and oceanographic community. This is because the standards were not developed with four-dimensional environmental data in mind. The NERC DataGrid project analysed the issues, proposed changes to the relevant standards, and introduced a data modelling philosophy into the community. One consequence is that the Met Office has funded internal projects, inspired by the NERC DataGrid, which have resulted in wide acceptance within the Met Office that OGC techniques are the most effective way to develop and deploy new meteorological products.

The NERC DataGrid is perhaps the most advanced design of its type in the environmental sciences.

Gil Ross, Information Regulation Principle, Met Office.

Studentships

The *eScience* programme funded 17 PhD studentships.

Below are five case studies.

■ Richard Bruin

Richard joined the *eMinerals* project straight from a computer science degree at Bath University. He has developed core middleware to support the project's use of integrated compute and data grids, including systems to select automatically the best available computing resources and enable automatic collection of metadata from output files. Richard's tools have contributed to the overall *eMinerals* toolset which is now being rolled out for use in central-site facilities. Richard used his tools to calculate the adsorption of dioxin molecules on chalk minerals and is now working on the MaterialsGrid project.

■ Richard Myerscough

GENIE student Richard Myerscough has developed a new way of representing how water of different densities mixes in the GENIE ocean model. This encapsulates the insights from NERC-funded oceanographic studies that mixing between water masses of different density depends on the density difference between them. Richard is exploring the implications of this for the different states of the ocean circulation and their stability.

■ Daniel Goodman

climateprediction.net PhD student Daniel Goodman developed Martlet, a computer programming workflow language, to help analyse the vast amounts of data produced by *climateprediction.net* participants. Martlet allows users to write functions without having to worry about what happens when the number of resources their data spans changes, so greatly simplifying the coding and use of such analysis.

■ James Levine

James Levine's PhD explored how gases move between the troposphere (0-12km) and the stratosphere (12-50km) and their relevance to the role that halogens could play in the depletion of stratospheric ozone at low and mid-latitudes. The results suggest the transport of these chemicals to the lowermost stratosphere could make an appreciable contribution to the halogen content of this region.

■ Anne Jones

PhD student Anne Jones has helped develop a seasonal ensemble prediction of malaria in Africa by coupling a climate-driven model of malaria with state-of-the-art seasonal ensemble climate forecasts. This is the first step towards a continental-scale assessment of the potential for seasonal epidemic malaria prediction.



Richard Bruin



Daniel Goodman

The future of eScience

NERC's new strategy *Next Generation Science for Planet Earth 2007-2012* states that NERC's strategic goal is to deliver world-leading environmental research at the frontiers of knowledge:

- enabling society to respond urgently to global climate change and the increasing pressures on natural resources;
- contributing to UK leadership in predicting the regional and local impacts of environmental change from days to decades; and
- creating and supporting vibrant, integrated research communities.

eScience can play a pivotal role in achieving this goal. Society's ability to respond to environmental change requires predictions of the future state of the environment and the impact that human actions will have upon it. To provide accurate predictions we need:

- an understanding of the dynamics of environmental processes;
- computer-based models of these processes;
- high quality datasets to initialise and validate the models;
- the technical infrastructure to run the models, often using very high performance computing, and to manage the ever-increasing volumes of data collected;
- the ability to run the models many times in ensembles to quantify the uncertainty associated with such predictions;
- the ability to visualise and analyse the results.

eScience's ability to effectively harness computing resources and manage, distribute and visualise environmental data makes it a key tool for achieving these objectives.

Better communication

Solutions to the major environmental challenges require more interdisciplinary science and better communication across the disciplines. NERC's new strategy states that: 'NERC aims to remove barriers to help environmental research take place effectively and efficiently.'

eScience is an extremely effective way of achieving this goal. This programme has already developed communication tools and methods for sharing data and computational tools that are radically changing the way that scientists in different institutions and from different disciplines work together.

For the first time, NERC's new strategy identifies technologies as a science theme, recognising the central role that technologies such as eScience will play in advancing environmental science in the future. The eScience programme puts NERC firmly at the cutting edge of these new technologies. The challenge now is for the environmental research community to adopt more widely the new tools and systems developed by the programme.





More information

www.nerc.ac.uk/research/programmes/escience

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, SN2 1EU

