

GENIE: Delivering e-Science to the environmental scientist

M. Y. Gulamali¹, T. M. Lenton², A. Yool³, A. R. Price⁴, R. J. Marsh³, N. R. Edwards⁵,
P. J. Valdes⁶, J. L. Wason⁴, S. J. Cox⁴, M. Krznaric¹, S. Newhouse¹, J. Darlington¹.

<http://www.genie.ac.uk/>

Abstract

The GENIE project aims to deliver a Grid-based, modular, distributed and scalable Earth System Model for long-term and paleo-climate studies to the environmental sciences community. In this paper we address the scientific problem of the vulnerability of the thermohaline circulation to the global climate, and describe our e-scientific solution using a Grid-based architecture involving Condor computational resources, web service oriented data management mechanisms and the employment of a web portal. We find that the scientific results of our e-science efforts are useful to the environmental science community, and provide a means of fulfilling the longer-term aims of the GENIE project.

1 Introduction

The Grid ENabled Integrated Earth system model (GENIE) project is funded by the Natural Environment Research Council (NERC) through their e-Science programme. GENIE aims to deliver both a flexible Grid-based architecture, which will provide substantial long-term benefits to the Earth system modelling community (and others who need to combine disparate models into a coupled whole); and also new scientific understanding from versions of the Earth system model (ESM) developed and applied in the project.

The scientific focus of GENIE is on long term changes to the Earth's climate, particularly the (geologically) recent ice-age cycles and the future response of the Earth system to human activities, including global warming. A realistic ESM for this purpose must include models of the atmosphere, ocean, sea-ice, marine sediments, land surface, vegetation and soil, ice sheets and the energy, biogeochemical and hydrological cycling within and between components.

The e-Science objectives of the project are to develop, integrate and deploy a Grid-based system allowing users to,

- flexibly couple together state-of-the-art components to form a unified ESM,
- execute the resulting ESM on the Grid,
- share the distributed data produced by simulation runs,
- provide high-level open access to the system, creating and supporting virtual organisations of Earth System modellers.

This paper describes our progress, thus far, in delivering the objectives listed above to the environmental sciences community. We begin by outlining a particular line of scientific investigation that our efforts have enabled scientists to explore (Section 2). We then detail the development and operation of a Grid Computing infrastructure that has made such a scientific investigation possible (Section 3), before discussing the scientific results gained from our e-science endeavours (Section 4). We conclude with an outline of our future work (Section 5) and a summary of the work presented herein (Section 6).

2 The scientific problem

2.1 Thermohaline circulation

The thermohaline circulation (THC) of the world ocean, also known as the global ocean *Conveyor Belt*, is a key part of the global climate system (see Figure 1). As its name implies, this large-scale circulation is driven by the geographical distribution of seawater temperature and salinity, and the North Atlantic plays a key role in sustaining the THC. The warm waters of the Gulf Stream, flowing across the Atlantic from the Gulf of Mexico towards Europe, lose heat and moisture to the atmosphere thus becoming denser. In regions of the North Atlantic East of Greenland, the surface water becomes sufficiently dense to sink to great depth in a process of deep convection. The deep water thus formed flows south-westwards until it reaches South America, and then continues down the side of that continent until it reaches the Southern Ocean where it

¹London e-Science Centre, Imperial College London, London, UK.

²Centre for Ecology and Hydrology, Edinburgh, UK.

³Southampton Oceanography Centre, Southampton, UK.

⁴Southampton Regional e-Science Centre, Southampton University, Southampton, UK.

⁵Physics Institute, University of Bern, Switzerland.

⁶School of Geographical Sciences, University of Bristol, Bristol, UK.

circulates eastwards. Branches of deep water flow north-east into the Indian and Pacific Ocean basins where they eventually up-well to the surface, fuelling a return current of surface water. The circulation is maintained by a particular pattern of atmospheric moisture transport, including a net transport from the Atlantic to Pacific basins and a somewhat restricted transport from tropical regions to the high latitude North Atlantic.

The THC is potentially vulnerable to systematic changes in the global hydrological (water) cycle. In particular, the Conveyor Belt mode of a single convection cell in the Atlantic, crossing from one hemisphere to the other, is sensitive to changes in westwards (latitudinal) and northwards (longitudinal) atmospheric transports of moisture. Under global warming the hydrological cycle is expected to intensify [10], with a 0.2 mm/day increase in global mean precipitation (typical for a 2°C global warming) equating to a freshwater transport of 11.8 Sv (1 Sv = $10^6 \text{ m}^3\text{s}^{-1}$) [24]. The THC may respond to such intensification in a non-linear fashion, associated with the existence of multiple equilibria. A previous study using a simple box model of the climate system [22] has demonstrated such multiple equilibria of the THC under different combinations of intra-basin (low to high latitudes of the North Atlantic) and inter-basin (Atlantic to Pacific) atmospheric moisture transport.

With this in mind, the first scientific objective of the GENIE project was to explore whether multiple equilibrium states of the THC exist in a more complex climate model under similar variations in atmospheric moisture transport.

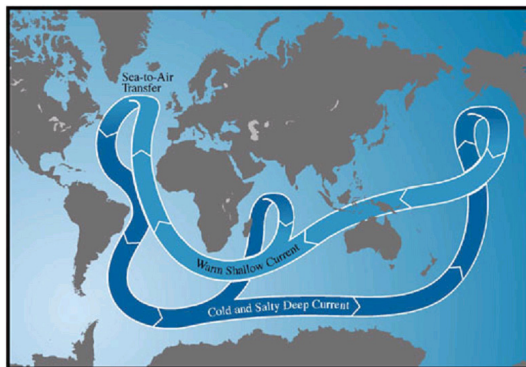


Figure 1: Thermohaline circulation of the world ocean. See text for description

2.2 The computational model

The initial instance of GENIE (-1-C) comprises a 3-dimensional (frictional geostrophic) ocean model coupled to a (dynamic and thermodynamic) sea-ice

model and a 2-dimensional (energy-moisture balance) atmosphere model. The whole model is also known as *c-GOLDSTEIN* and is described in detail elsewhere [3]. The horizontal resolution is 36×36 in coordinates of longitude versus $\sin(\text{latitude})$ and there are 8 depth levels in the ocean plus one for the atmosphere. The model is computationally very efficient, taking approximately 2.5 hours for a 4000 year integration on a typical PC or workstation (i.e. an Intel P3/1GHz processor with 256MB RAM). Integrations of this duration are sufficient for the slowest component (the ocean circulation) to approach equilibrium.

The model takes a number of input parameters which describe various physical quantities in the simulation, including the surface freshwater flux between the Atlantic ocean and the Pacific ocean (hereafter denoted as DFWX); the atmospheric freshwater transport between the the tropics and the extra-tropics of the North Atlantic ocean (hereafter denoted as DFWY); and the overall atmospheric moisture diffusivity in the simulation (hereafter denoted as DIFF). See Figure 2.

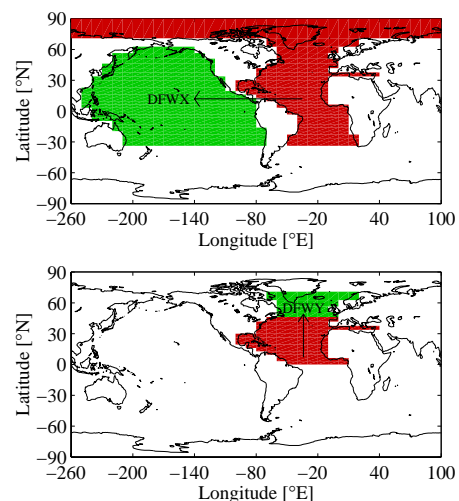


Figure 2: The DFWX anomaly transfers freshwater meridionally between the Atlantic and Pacific basins (−0.3 to 0.3 Sv), while DFWY transfers freshwater zonally between the low and high latitudes of the North Atlantic (−0.3 to 0.3 Sv).

3 Providing an e-Science infrastructure

3.1 Condor as a Grid computing resource

The initial scientific objective demanded the execution of many ensemble experiments, each com-

prising of 961 individual runs of the model, varying two parameters in combination. Our experiments are discussed in more detail in Section 4. Despite the speed of the model, undertaking this many runs on a single processor was clearly too time consuming. Hence a more efficient way of conducting large ensemble experiments was developed.

We used Condor [21] to execute the individual model runs of an ensemble experiment on a large computing resource made up of existing Condor pools. The pools resided at the London e-Science Centre [11], the Department of Computing at Imperial College London [2], and the Southampton Regional e-Science Centre [19]. They were combined into a single computing resource using the flocking features within Condor (e.g. see [4]). The presence of institutional firewalls required the designation and utilisation of port ranges specified by the Condor and firewall administrators at these institutions. The flocked Condor pool was composed of Linux and Solaris platforms, with approximately 200 nodes being available at peak times (i.e. when the host institutions allowed these nodes to be available, and CPU activity on them was low).

In order to use the GENIE model on the flocked Condor pool it was compiled from source on both the Linux and Solaris platforms, with the standard Condor libraries. This allowed us to use the Condor *standard* universe at execution time, granting the additional features of automatic checkpointing and automatic job migration (e.g. see [12]). Furthermore, by utilising the Condor File Transfer Mechanism in our Condor submission scripts we did not have to ensure that each of the Condor pools in the flocked pool provided access to a shared filesystem.

Thus by allowing Condor to schedule and distribute the model runs in an ensemble experiment we were able to execute the individual runs in parallel. This dramatically reduced the total time for the experiment to complete from the calculated ~ 163 days on a single processor, where each model run was being executed sequentially, to ~ 3 days on the Condor resource.

3.2 A web portal for job submission

In order to execute the individual model runs on our Condor resource we had to specify standard input and output files for each run. This task amounted to creating 961 individual input files for each ensemble experiment. As a means of automating this chore, as well as providing an interface to allow scientists to create ensemble experiments, submit them to the flocked Condor pool, and subsequently monitor their progress, we developed a prototype web-based portal based on Apache Tomcat servlets [1].

The servlets for the portal were developed in Java to interface the browser with the underlying Condor system. They make system level calls in order to submit experiments to Condor, as well as retrieve information about the status of jobs in the pool. This is illustrated in Figure 3 where we show how a user may monitor the progress of their ensemble experiment through the portal. The process takes place as follows: (1) A user makes a request through their web browser to view the status of the Condor jobs in their experiment; (2) a Tomcat servlet interprets the request and makes a system level call to the operating system hosting the servlet container; (3) the operating system returns the status of the user's Condor jobs to the servlet container; (4) the container processes the results and formats them to create a response in the user's browser. The submission of an ensemble experiment to the Condor pool takes place in a similar way.

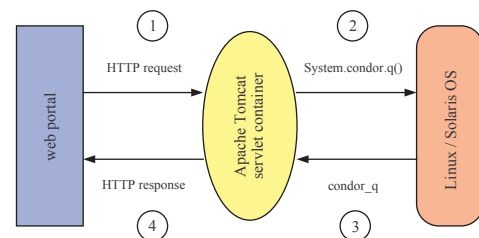


Figure 3: A typical request-response for a user wishing to monitor the progress of their ensemble experiment on the Condor pool through the portal. See text for details.

The portal is accessed by an environmental scientist with their X.509 e-Science certificate. By using certificate-based authorisation and authentication we can ensure that only those environmental scientists on the GENIE project have access to the provided computational resources, while forgoing the requirement that they each have an account on each resource. This is clearly advantageous from the point-of-view of a system or virtual organisation (VO) administrator [5].

The creation of a new experiment in the portal consists of specifying the range of the parameter space to be explored in the experiment, as well as setting numerous other input parameters for each of the model runs in the experiment. Upon submitting the experimental configuration to the portal, the portal will then create the necessary input files for each of the model runs in the experiment, and a Condor submission script, before inviting the user to submit the experiment to the Condor pool.

Once an experiment is submitted to the flocked Condor pool, a user can monitor the status of jobs on the pool. The user is also provided with options

for cancelling the experiment or creating a new one, thus overriding the currently running experiment. At this time the portal can only manage one ensemble experiment at a time.

Finally, when all the model runs in an experiment have completed, the results are archived. The user can then download the results, as well as those from previous experiments, to their local filesystem for further analysis, the mechanisms for which are discussed in the following subsection.

Together, the actions above allow a GENIE scientist to specify their experiment and retrieve the final results without worrying about the technicalities of the underlying Condor pool or the GENIE simulation, granting them more time to focus upon the analysis of their results.

3.3 A Grid-enabled data management system

The increased resources available through Condor require novel approaches to the management of, and access to, large volumes of data generated by the experiments. A single ensemble experiment (consisting of 961 simulations) produces up to 10000 flat files amounting to approximately 825MB of uncompressed data. This data would previously have been managed by the scientist who ran the experiment and would often simply have been archived to CDROM after post-processing and analysis was completed.

In order to facilitate sharing of the data between members of the GENIE project and enable scientists to navigate the data produced by unfamiliar component simulation codes, we have implemented an initial data management system exploiting the Geodise Database Toolkit [9]. This is a generic Grid-enabled data management solution based on open W3C compliant standards technologies [23]. The system exploits database technology to allow additional information (metadata) to be associated with data produced by the GENIE model. This metadata describes the nature of each file and can be queried within the database so that files can be located easily.

The design of the GENIE data management system is presented in Figure 4. The data generated by the GENIE model runs are archived into the data management system in two stages. (1) Data files are first transferred to the file server using GridFTP using a proxy certificate authenticated by a user's X.509 certificate. The data are stored on the file server with a uniquely generated Universal Unique Identifier (UUID) as a file handle; (2) the portal calls web service methods to archive metadata associated with the data. At present this service is hosted

at the Southampton Regional e-Science Centre but future work will extend this to a distributed replicated repository.

Web services have been implemented to provide an interface to the database that stores the metadata. These are conceptually grouped into three services that are used to maintain information about the files archived on the file server(s):

1. the *location service* records details of the host and the file system directory associated with an archived data file. This allows a user to archive their data to any Globus enabled file server on which they have an account.
2. the *metadata service* allows application and domain specific metadata in XML to be recorded with an associated data file. The service records two types of metadata: (a) a standard set of metadata about the characteristics of the data file (i.e. size, creation date, etc.), and (b) any application or domain specific metadata presented as an XML document. This service also provides a query to the underlying database.
3. the *authorisation service* allows registered users to be given access rights to the data archived in the repository. Access permissions can be assigned to data files according to a user's X.509 certificate subject.

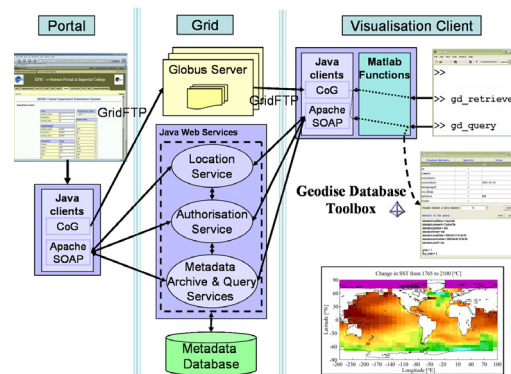


Figure 4: Architectural design of the GENIE data management system.

A familiar tool amongst the environmental scientists on the GENIE project is MATLAB [15]. Consequently we adopted the Geodise Database Client tools to provide an initial interface to the data management system. These tools are Grid-enabled and required users to authenticate themselves using an X.509 certificate. Upon doing so, MATLAB functions from the toolkit can be used to query the metadata and retrieve files from the repository.

The query function operates by instantiating a Java proxy class that calls the query web service attached to the metadata database. A MATLAB function to retrieve data files, by accessing the location web service, is also provided. These mechanisms allowed environmental scientists to easily retrieve, analyse and visualise data from the ensemble experiments within the same environment.

4 Scientific achievements

4.1 (DFWX,DFWY) experiment

In our first experiment, we varied atmospheric moisture transports by adding/subtracting anomalies to the surface freshwater flux in selected ocean regions. The two forcing parameters described in section 2.2, DFWX and DFWY were varied in combination to achieve this. In order to search for equilibrium states of the THC, we constructed an ensemble of independent 4000 year model runs in (DFWX,DFWY) parameter space, varying each of DFWX and DFWY from -0.3 Sv to 0.3 Sv in steps of 0.02 Sv, giving $31 \times 31 = 961$ runs in total. Each model simulation was initialised from a uniformly cold state.

4.2 (DFWX,DFWY) results

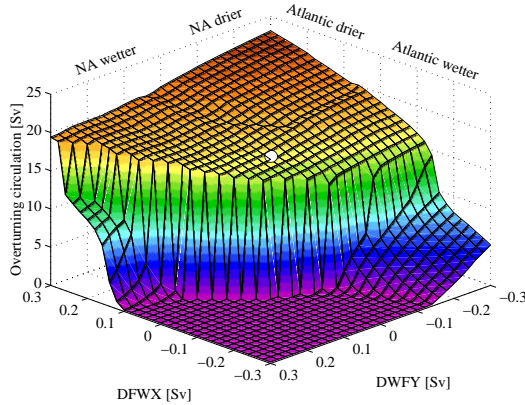


Figure 5: Intensity of the thermohaline circulation as a function of DFWX and DFWY, measured as the maximum northward transport of water mass in the Atlantic.

In our initial experiment we found two distinct equilibrium states of the THC in (DFWX,DFWY) parameter space (Figure 5). These can be characterised as Conveyor *on* and *off* modes. The unperturbed solution (marked with a white dot in the figure), which is our best guess at the present state of the system,

is notably close to the sharp transition from Conveyor *on* to *off* state. In Figure 6 we show plots of the circulation state at opposite extremes of the (DFWX,DFWY) parameter space, which illustrate the contrast between these two modes.

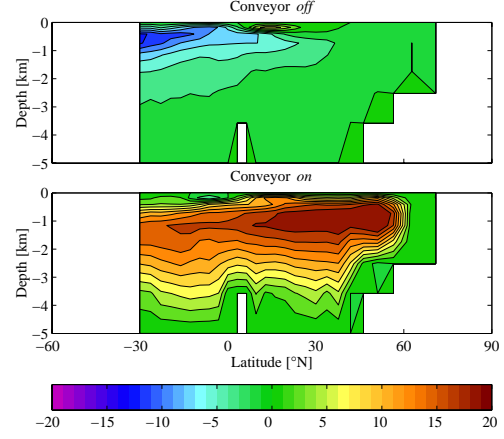


Figure 6: Plots of the two modes of the Conveyor belt: Conveyor *off* and Conveyor *on*, at opposite extremes of (DFWX,DFWY) parameter space.

4.3 (DFWX,DIFF) experiment

Although there is some rationale for using DFWX as a forcing parameter, DFWY is harder to justify. Hence, in a subsequent suite of experiments, we replaced DFWY with the parameter DIFF, which determines atmospheric moisture diffusivity. The main effect of DIFF is on meridional moisture transports: higher values of DIFF support larger poleward moisture transports. Ensembles in (DFWX,DIFF) parameter space involved varying DFWX from -0.3 Sv to 0.3 Sv in 31 steps of 0.02 Sv (as before) and varying DIFF from 5×10^4 to 5×10^6 m^2s^{-1} in 31 equal logarithmic steps of $\log_e \text{DIFF} = 0.06666666$. As in the previous experiment, we performed a series of 961 runs across (DFWX,DIFF) space, starting from a uniformly cold state.

4.4 (DFWX,DIFF) results

Figure 7 shows a comparable plot to that of Figure 5 for our experiment in (DFWX,DIFF) parameter space. In common with the earlier experiment, the parameter space can be divided broadly into Conveyor *on* and *off* modes. On this occasion the unperturbed solution (again marked with a white dot) is further from the transition between the two modes. However, parameter DIFF is not well constrained, and the most realistic solution may lie elsewhere, possibly closer to the transition zone.

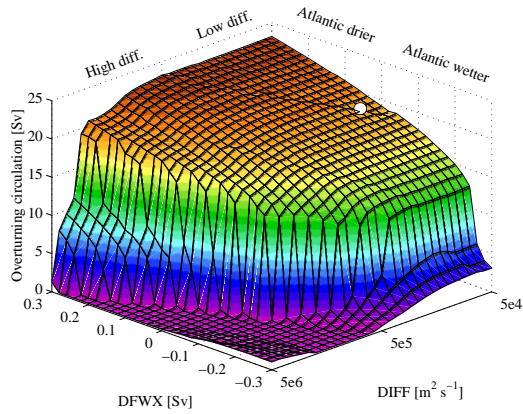


Figure 7: Intensity of the thermohaline circulation as a function of DFWX and DIFF, measured as the maximum northward transport of water mass in the Atlantic.

4.5 Initial conditions experiment

Earlier studies [22] have shown that there are two distinct equilibrium states of the THC as seen above. These are found to be stable for the same pattern of freshwater forcing to the Atlantic, and that the history of the system plays an important role in determining to which stable mode the Conveyor converges. To examine this sensitivity of the Conveyor a series of nine ensembles based upon the (DFWX,DIFF) ensemble were performed. Each of these was identical to that performed previously, except that rather than initialising the model from a uniform cold state, final states from the previous ensemble were used. In terms of (DFWX,DIFF) these were,

$$\begin{aligned} &(-0.3, 5 \times 10^4), (-0.3, 5 \times 10^5), (-0.3, 5 \times 10^6), \\ &(0.0, 5 \times 10^4), (0.0, 5 \times 10^5), (0.0, 5 \times 10^6), \\ &(0.3, 5 \times 10^4), (0.3, 5 \times 10^5), (0.3, 5 \times 10^6). \end{aligned}$$

Again, all model runs were for 4000 years. We also continued the original ensemble experiment for a further 4000 years to provide the restarted experiments with a control run of equivalent duration (i.e. 8000 years in total).

4.6 Initial conditions results

By comparing the strength of the Conveyor across parameter space and between the nine ensembles described in the previous subsection, Figure 8 highlights the region of parameter space across which the Conveyor can converge to quite different stable modes despite identical forcing. This region of multiple solutions corresponds to small shifts in the position of the “cliff” seen in Figure 7 between different ensembles.

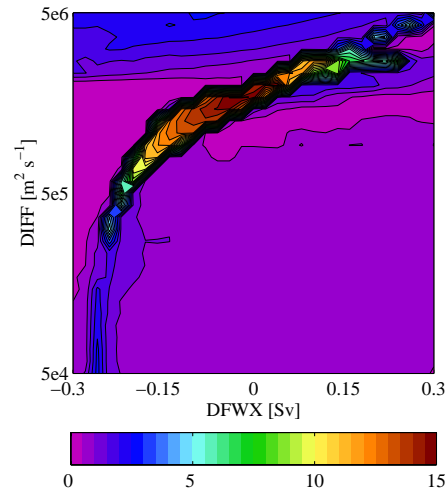


Figure 8: Contour plot of the most extreme MOC difference between the two states of the thermohaline circulation as a function of DFWX and DIFF.

In terms of the consequences for the climate, we find that the collapse of the Conveyor creates a temperature dipole in the Atlantic, with slightly warmer temperatures in the South Atlantic, but a decrease of approximately 2°C in the British Isles [14].

5 Further work

5.1 An ICENI-based web portal

The current portal has a number of shortcomings which limit it from providing the type of flexibility required to fulfil the aims of the GENIE project (cf. Section 1). Firstly, the present portal is only capable of using the version of the GENIE model embedded into it during the development phase. A user is given no choice of using another version of the model. Secondly, the format of the input and output data used in the model is predetermined, and a user cannot perform experiments using different data (i.e. data at higher spatial resolution) despite the ability of the model to support them. Similarly, the parameter space being investigated in the ensemble experiments is fixed and cannot be chosen by the user during the creation of the experiment. Finally, while the use of the flocked Condor pool is clearly beneficial, a user is presently unable to use any dedicated computing resources that they may have access to.

Consequently we are currently attempting to implement other middleware platforms, such as the Imperial College e-Science Networked Infrastructure (ICENI) [7], in order to extend the functionality and scope of the present portal. In particular, the ICENI middleware provides an Open Grid Services Infrastructure (OGSI) -compliant service ori-

ented architecture [8] in order to federate resources among Grid users. It uses a component-based application model [16] to efficiently broker applications on Grid resources.

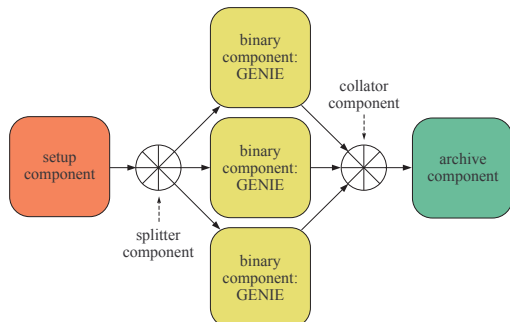


Figure 9: The GENIE ensemble experiment as a component-based application. Arrows describe the direction of control and data flow between components.

In this scenario an ensemble experiment consists of a number of components that encapsulate the workflow of the experiment (see Figure 9). A *setup component* initialises the experiment, creating the necessary input files for the ensemble, using parameters chosen by the user at run time. This passes data and control to a *splitter component* which delegates the data to multiple *binary components* (only 3 are shown in the figure), each of which execute the GENIE model. As each simulation finishes, the *binary component* associated with that simulation passes the resultant data to a *collator component* which passes it to an *archive component* that handles the archiving of the resultant data.

This application model has the advantage of being both flexible and extensible. Different GENIE models can be substituted into the *binary component* by the user at design time (i.e. when the user creates the application). Moreover, the *archive component* or *setup component* can be replaced with one of the user's choosing. The ICENI middleware ensures that the computational model encapsulated in the *binary component* is scheduled and executed on the most appropriate computational Grid resource, be it a Condor pool or dedicated computing resources associated with the user.

We are developing a web portal interface to ICENI, to allow a user to flexibly create and execute an application on their Grid resources. This is made possible by the architecture of ICENI which complies with the Open Grid Services Architecture (OGSA) [6]. Consequently, we are able to describe GENIE components as Grid services within ICENI, and then expose these through a gateway, the *OGSA gateway* [8], to the servlet container that serves the portal. This is illustrated in Figure 10 (cf. Figure 3),

where we show an arbitrary request-response by a user. The process takes place as follows: (1) A user makes a request through their web browser; (2) a Tomcat servlet interprets the request and makes the associated Grid service request to the OGSA gateway as a Simple Object Access Protocol (SOAP) message [17]; (3) the gateway transfers the request as an Remote Method Invocation (RMI) message [20] to ICENI, which (4) returns a response to the gateway in a similar manner; (5) the gateway transfers the response as a SOAP message to the servlet container, which (6) processes it and sends the appropriate response to the user's browser.

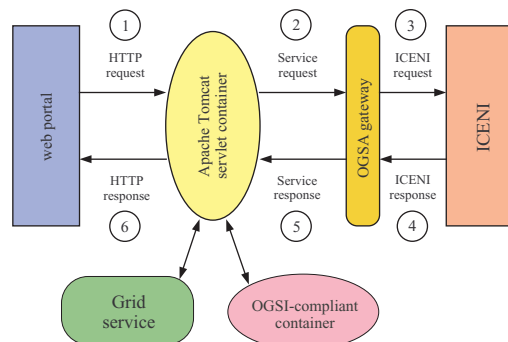


Figure 10: A typical request-response process for an ICENI-based web-portal for GENIE. See text for details.

As a further feature, it is possible for external Grid services and OGSI-compliant containers to communicate with the servlet container and make themselves available to the user.

Thus we envisage providing environmental scientists with the ability to create and execute their GENIE experiments in a flexible and transparent way on the Grid.

5.2 Future database developments

We intend to adopt the netCDF (CF) binary format [18] for the output of data produced by the simulation codes. The query services will be updated to allow in-situ processing of the data and enable slices of the data to be returned without the need to retrieve complete files. We also intend to provide the ability to query for summary information about the metadata itself in addition to data focused queries.

6 Conclusions

Environmental scientists within the GENIE project have used a new web-based portal to test hypotheses concerning the influence of freshwater transport upon the global ocean circulation. In ensemble experiments, each consisting of 961 simulations, they

have been able to investigate the sharp transition between two dominant modes of the ocean thermohaline circulation in unprecedented detail [13]. This result has only become apparent through the use of computational resources such as Condor pools, and user interfaces such as web portals to facilitate them. In particular, summed together the ensemble experiments amount to 46 million years of simulated time. In the absence of e-Science support such a computational effort would take several years to complete, or more likely would not be attempted at all.

Furthermore, we have exploited and enhanced grid and database technologies developed by the Geodise project to implement an initial data management system for the GENIE project. By associating metadata with the data from the GENIE model, and exploiting database technology, we have implemented a system that enables scientists to more easily manage the large volume of data generated from the ensemble experiments. In the future this will facilitate data sharing and re-use in the GENIE virtual organisation of Earth system modellers.

References

- [1] Apache Jakarta Tomcat:
<http://jakarta.apache.org/tomcat/>
- [2] Department of Computing, Imperial College London, London, UK.:
<http://www.doc.ic.ac.uk/>
- [3] Edwards, N. R. and R. Marsh, An efficient climate model with three-dimensional ocean dynamics, *Clim. Dyn.*, submitted.
- [4] Epema, D. H. J., et al., A worldwide flock of Condors: Load sharing among workstation clusters, *Journal on Future Generations of Computer Systems*, **12**, 1996.
- [5] Foster, I., et al., The anatomy of the Grid: Enabling scalable virtual organisations, *International J. Supercomputing Applications*, **15**(3), 2001.
- [6] Foster, I., et al., The physiology of the Grid: An Open Grid Services Architecture for distributed systems integration, Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002. Available at: <http://www.globus.org/research/papers/ogsa.pdf>
- [7] Furmento, N., et al., ICENI: Optimisation of component applications within a Grid environment, *Parallel Computing*, **28** (12), 1753-1772, 2002.
- [8] Furmento, N., et al., ICENI: An Open Grid Service Architecture implemented with Jini, in *SuperComputing 2002*, Baltimore, MD., USA., November 2002.
- [9] Grid Enabled Optimisation and Design Search for Engineering: <http://www.geodise.org/>
- [10] Hulme, M., et al., Precipitation sensitivity to global warming: comparison of observations with HadCM2 simulations, *Geophys. Res. Lett.*, **25**, 3379-3382, 1998.
- [11] London e-Science Centre, Imperial College London, London, UK.:
<http://www.lesc.ic.ac.uk/>
- [12] Litzkow, M. et al., Checkpoint and migration of UNIX processes in the Condor distributed processing system, University of Wisconsin-Madison Computer Sciences Technical Report #1346, April 1997.
- [13] Marsh, R. J., et al., Multiple equilibria of the thermohaline circulation under anomalies in zonal and meridional moisture transport: Early results with a new Earth System Model, EGS-AGU-EUG Joint Assembly, Nice, France, 2003.
- [14] Marsh, R. J., et al., Atlantic thermohaline catastrophe mapped in two-parameter space, *Clim. Dyn.*, submitted.
- [15] The MathWorks MATLAB:
<http://www.mathworks.com/>
- [16] Mayer, A., et al., Meaning and behaviour in Grid oriented components, in M. Parashar, editor, *Grid Computing – GRID 2002: Third International Workshop, Baltimore, MD., USA., November 18, 2002. Proceedings, Lecture Notes in Computer Science*, **2536**, 100-111, 2002.
- [17] Mitra, N., editor, SOAP Version 1.2 Part 0: Primer, W3C Recommendation, 24 June 2003. Available at: <http://www.w3.org/TR/soap12-part0/>
- [18] NetCDF Climate and Forecast (CF) Metadata Convention: <http://www.cgd.ucar.edu/cms/eaton/cf-metadata/index.html>
- [19] Southampton Regional e-Science Centre, University of Southampton, Southampton, UK.:
<http://www.e-science.soton.ac.uk/>
- [20] Sun Microsystems, Java Remote Method Invocation – Distributed computing for Java, *White Paper*. Available at: <http://java.sun.com/marketing/collateral/javarmi.html>
- [21] Thain, D., et al., Condor and the Grid, in F. Berman, A. J. G. Hey, G. Fox, editors, *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley, 2003.
- [22] Wang, H. and G. E. Birchfield, An energy-salinity balance climate model: Water vapor transport as a cause of changes in the global thermohaline circulation, *J. Geophys. Res.*, **97**, 2335-2346, 1992.
- [23] The World Wide Web Consortium:
<http://www.w3.org/>
- [24] Yang, F., et al., Intensity of hydrological cycles in warmer climates, *J. Climate*, **16**, 2419-2423, 2003.