Grid-Ireland and Irish e-Research Strategy. A Review for KAREN and BeSTGRID

Antony P. Gerdelan
Institute of Information and Mathematical Sciences
Massey University, North Shore City 102-904, New Zealand, 0620
Email: a.gerdelan@massey.ac.nz
Tel: +64 9 414 0800 Fax: +64 9 441 8181

September 2008

Contents

List of Figures

Acknowledgements

I would like to thank KAREN, BeSTGRID, and the Institute of Information and Mathematical Sciences at Massey University for providing the funding and assistance which allowed me to conduct this study, and Grid-Ireland for hosting my research placement with their group. In Ireland special thanks go to the following people for generously giving their time and detail their research activities: *Brian Coghlan, Eamonn Kenny, Stephen Childs, David O'Callaghan, John Walsh, Peter Lavin, Kathryn Cassidy, Geoff Quigley, Neil Simon, Stephane Dudzinski, Gabriele Pierantoni, and John Ryan* at Trinity College Dublin *John Morrison, Philip Healy, and Dave Power* at University College Cork, and *Andrew Shearer* from National University of Ireland, Galway. Special acknowledgements also go to: *Chris Messom, Michelle Campbell, Colleen van Es and Jane A. Kennedy.*

Abstract

Establishing a functioning computational Grid in New Zealand that both provides a useful service to researchers and scientists, and at the same time fully takes advantage of the great resource that is KAREN is proving

to be a considerably difficult challenge. There are many organisational, technical and manpower challenges that must be addressed by BeSTGRID before it can emerge from its fledgling status and grow into a mature vehicle for facilitating a greater range of e-Research within New Zealand and an important player in large international scientific projects. This report presents a case-study of Grid-Ireland; an organisation that has grown from similar roots to BeSTGRID, and has overcome many of the challenges currently facing New Zealand's e-Research community in an identical ICT climate to make a name for itself as a pioneer in the field and a key contributor to international *Big Science*. Indeed, much of the middleware technology model that we already make use of as part of the Grid in New Zealand, we have unwittingly inherited from innovations made by Grid-Ireland. E-Research in New Zealand can benefit greatly by keeping a close eye on the activities and progress of this role model on the other side of the globe.

1 Introduction

In a recent international report on ICT capability by country, New Zealand and The Republic of Ireland placed 22nd equal of 127 countries surveyed for *Networked Readiness* by the World Economic Forum [1]. The Republic of Ireland has the same population size to New Zealand, similar national research interests, and within the e-Research community a Grid organisation that is not only under the same type of Virtual Organisation umbrella but was actually the initiator of the particular Grid *gatekeeper* model [2] that we have been using in New Zealand to provide access to our Grid resources [3]. The proposed travel was an opportunity to learn from the considerable experience of e-Research providers working in a very similar national climate how best to improve uptake of the Grid as e-Research infrastructure for our scientists and students, and how to improve access to grid resources over KAREN (*Kiwi Advanced Research and Education Network*) to increase national e-Research capability. The guiding questions behind the visit were:

- "What key steps did Grid-Ireland take to increase the use of its broadband research networking backbone through grid technology?" *and*
- "How can BeSTGRID in New Zealand emulate this process?".

Ireland's computational Grid project is an excellent case study for New Zealand's expanding e-Research community. Much of Ireland's progress in e-Research is clearly due to the forward-thinking academic policies of that country, supported by Ireland's recent economic boom [4]. However, many of the advances made in Grid Computing have not been solely dependent on the advance of the *Celtic Tiger*, but rather are a result of a very effective organisational reform under strong leadership. The directors of Grid-Ireland have managed to steer the organisation from a decentralised collaborative model (the same as our current organisational model in New Zealand) that effectively gave individual Grid sites little incentive to share resources to a more cohesive model with a focus on international collaboration and resource ownership that eliminates internal competition and encourages shared access to computational and mass-storage resources on a national level over the high-bandwidth HEAnet

(Higher Education Authority Network) broadband research network - Ireland's equivalent to KAREN, so well in fact, that during my visit Grid-Ireland had its network pipe share restricted - the complete inverse situation to the current network uptake by Grid in New Zealand.

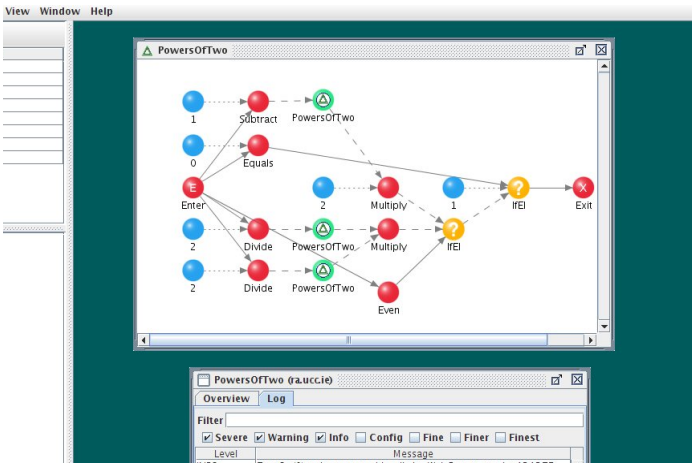

Figure 1: Modelling a workflow in a WebCom-G IDE.

On a technical level I was also interested in middleware being used and developed by Grid-Ireland, having had a brief introduction to Webcom-G [5] (see Figure 1) during a day-trip to University College Cork last year. I was aware that Grid-Ireland were using some significantly more advanced middleware and user tools, and I have documented some of these as snapshots of what can be done with the Grid. Some of the powerful user-level tools should be of interest to e-Researchers in New Zealand, as Grid-Ireland is already making use of the kind of Workflow job creation, submission and monitoring tools (some of which are becoming run-time interactive) and other MPI facilities that our e-Researchers and HPC users in New Zealand have been requesting that we have capability for, the demand for which was raised at the Australasian cooperation panel at the APAN26 conference this August.

I have also interviewed Dr. Andy Shearer from National University of Ireland, Galway, who is instrumental as a director of e-Inis; Ireland's leading e-Research initiative, and is heavily involved in a number of Irish and international e-Research projects that have a particularly New Zealand-relevant flavour; seabed mapping, health care and virus modelling, and astronomy; some of which will provide researchers in these areas in New Zealand with some ideas for how Grid Computing and advanced networking resources might be further utilised.

2 Background to e-Research in Ireland

2.1 e-Inis

The Irish National e-Infrastructure organisation *e-Inis* - a clever play on words as a combination of acronym and Irish language term roughly meaning *electronic-Ireland* - is a Higher Education Authority project involving Irish academic in-

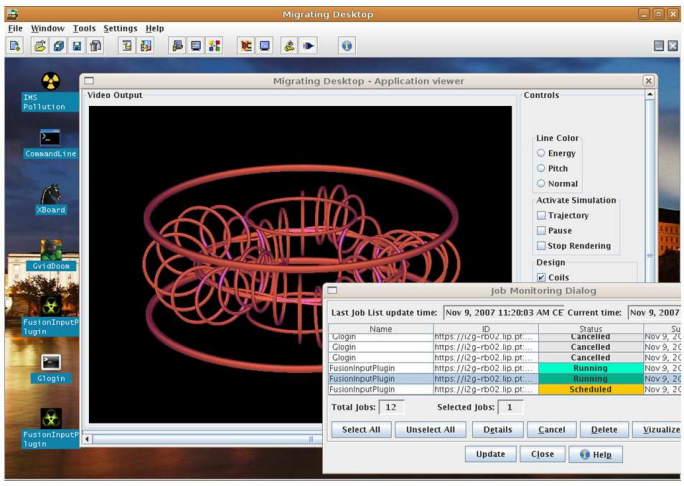

Figure 2: The Migrating Desktop grid-enabled visualisation and computational steering tool

.

Figure 3: e-Inis - Irish e-Infrastructure cooperation.

stitutions and ICT providers led by the Dublin Institute for Advanced Studies (DIAS). e-Inis is a steering force behind e-Infrastructure in Ireland and is a federation of three chief partner organisations; ICHEC, Grid-Ireland and HEAnet which are backed by a large group of academic institutions. Figure 4 provides an illustration of how e-Inis is comprised.

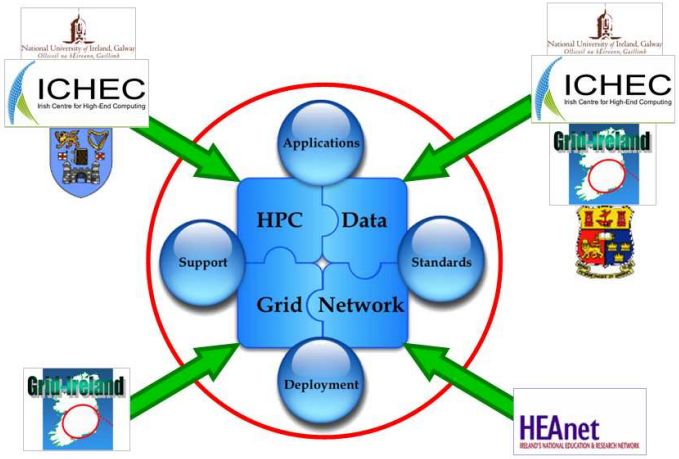

Figure 4: A breakdown of e-Inis member organisations.

The role of e-Inis is to bring together the different e-Research related organisations (mostly academic consortiums) so that the providers and users of e-research can match requirements and the overall direction of e-Research in Ireland can be driven with a common focus with a sustainable national e-infrastructure. A large part of e-Inis deals with filling the gaps that are commonly left by individual e-research groups by providing training and support staff as part of the national e-infrastructure.

2.2 Grid-Ireland

There are three primary Grid sites and Grid research centres in the Republic of Ireland. These are at Trinity College Dublin, University College Cork, and National University of Ireland, Galway (see Figure 5). A parallel can be drawn here to New Zealand's initial three member Universities (Massey University at Albany, the University of Auckland, and the University of Canterbury). From these original members the Grid in Ireland has grown substationally - Grid-Ireland now has an effective Grid-deployment mechanism that it has used to deploy Grid sites at 18 facilities and institutions nationwide, and to Grid sites in Northern Ireland as well (see Figure 6). Prospective Grid member institutions no longer need to provide and train specialised staff to establish Grid gatekeepers as this process is now almost entirely automated, and even maintenance and upgrades can be administered remotely from the Grid Operations Centre (*OpsCentre*) at Trinity College Dublin.

As well as being a research topic in its own right involving more than ten PhD students and numerous other researchers, the Grid in Ireland facilitates some interesting national projects that require either the stringent security pro-

Figure 5: The Grid-Ireland logo, with the initial 3 member locations connected by a stylised HEAnet (red line).

Figure 6: Grid-Ireland Member Locations (from a G.O.C monitoring tool).

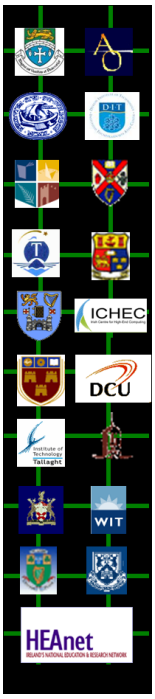

Figure 7: Grid-Ireland participants.

tocols inherent in the Grid or capitalise on the storage management and mass data transfer facilities that allow large amounts of sensitive data (such as in the national seabed mapping project) to be transfered to Grid-enabled storage resources at remote locations via HEAnet in a manner that prevents access to prying eyes from industry or elsewhere. Grid-Ireland has also leveraged the Grid and networking resources to become involved in massive international projects, such as the Enabling Grids for E-sciencE (EGEE) projects (see Figure 8), on a relatively small scale in terms of CPU-hours compared to larger nations but with a high-impact on the development of these projects and great visibility of collaboration - pushing forward Ireland's status as an important international collaborator that punches well above its weight.

3 Irish Strategy

Dr. Brian Coghlan is one of the directors of Grid-Ireland and is based at Trinity College Dublin. I asked Brian what Grid-Ireland's key strategies were to make better use of HEAnet and encourage grid deployment in Ireland, and he gave me three key strategies that Grid-Ireland had taken, and which we could think about as potential strategies to adapt in New Zealand to enhance our grid-deployment capability and KAREN usage:

- Securing a central, Grid-dedicated compute cluster.
- Establishing a National Data Centre.

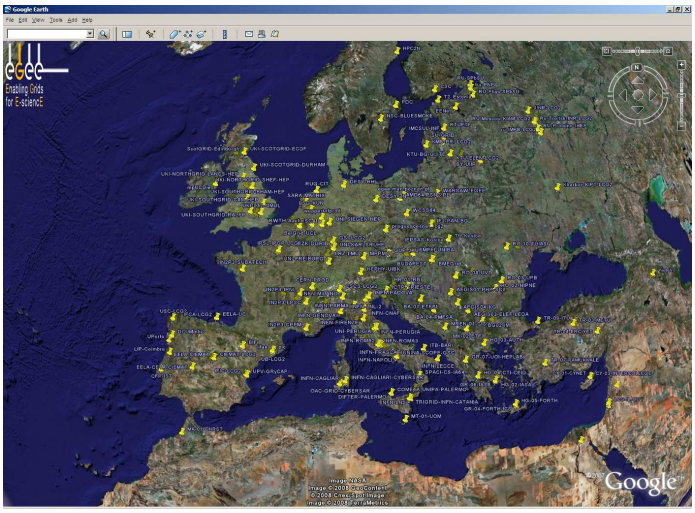

Figure 8: Map of EGEE international participation.

.

- Participating in massive-scale international projects.

3.1 Dedicated Grid Cluster

In the first major funding iteration of Grid-Ireland, the three main sites had the same fundamental problem that we now have in New Zealand. The resources were purchased under Grid funding but owned by individual institutions. This meant that, despite an effective Grid middleware setup, each institution was swamped by demand from cluster users within their own institution (who typically have unlimited demand for CPU resources) who naturally expect preferential allocation of resources, and as a result little resources were left for collaborative reseach between sites utilising HEAnet or for Grid-access from outside the campus. International project participation *via* the Grid was out of the question.

A major breakthrough came when Dr. Coghlan decided that it was necessary to fund a Grid-dedicated cluster. The cluster is now based at Trinity College Dublin (the Grid cluster is pictured in Figure 9) and is comparable in capacity (and presumably budget requirements) to the new cluster at Massey University. Having this cluster owned and administered by the Grid organisation has allowed the group to avoid the politics of individual college control of resources, provide supercomputing access (by way of Virtual Organisations) to researchers around the country that may not be based on a campus with access to a local cluster, and free processing cycles that have allowed Ireland to participate in prestigious international projects.

The result of this directorial steering, although certainly a challenge to organisational structure, is that Grid-Ireland has now activated a whole range of collaborative projects that make much better use of HEAnet and that otherwise could not have happened. I have reported on a selection of these new cross-site, collaborative Grid-enabled projects in Section 5.

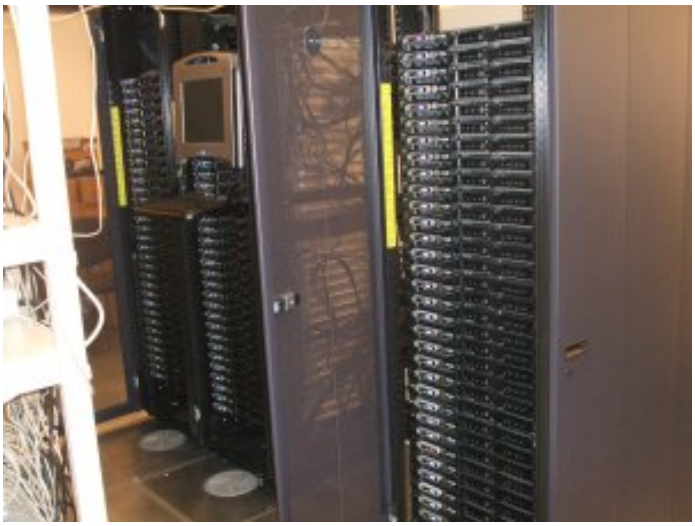

Figure 9: The Grid-dedicated CPU and storage cluster based at Trinity College Dublin.

3.2 National Data Centre

As part of an EU project Grid-Ireland manages a number of storage machines (at the time of my visit this was being expanded to more than 200TB of storage) at Trinity College Dublin with their Grid cluster. The usage strategy taken by Grid-Ireland was to enhance the capability of this storage with the latest meta-data technology and Grid security tools and make this storage available to researchers nationally. Whilst Grid-Ireland could have reserved this storage as scratch space for cluster users, what this move effectively did was enable some very valuable new projects to take place outside of the member institutions.

The Irish national seabed mapping project is a collaboration between all of the major Irish universities and several other institutions to model the seabed of much of Ireland's vast coastline and is of immense value to the national economy. Mapping the ocean is an incredibly data-intensive process and the storage is simply not available to be replicated at all of the mapping sites. What was needed was a nationally central storage location that could be accessed over HEAnet from any major research participant in the county, and this facility needed to be incredibly secure, as such information could be a potential gold-mine to tempted parties from the oil and fisheries industries. Marine Grid (see section 5.3) was created to link this research to the available resources around the country, but while having a large data centre at Trinity College provides a very valuable resource to this initiative, what is really needed is a much larger national datastore.

Having the grid-enabled data facility is also valuable to international projects. As CERN's LHC facility has come on-line Ireland is already receiving a large stream of data over the international HEAnet link. The early stages of this project participation were the cause of the bandwidth choke of HEAnet during my visit and the throttling back of bandwidth share to Grid-Ireland on the Trinity College Dublin campus. Having a nationally-accessible data facility is

certainly key to better usage of available broadband research networks.

Now the e-INIS consortium are about to establish such a National Datastore. The details have yet to be decided but it is likely that the TCD datastore will be integrated within this.

3.3 International Collaboration

Grid-Ireland is a member of the world's largest Grid - the EGEE, which comprises about 50,000 CPUs over more than 250 Grid sites. Grid-Ireland is very active in EGEE phase III, and aside from the prestige of participation in such a massive-scale international project, has used the international link-up to *shake out* any bottlenecks in the National Grid Infrastructure (NGI) in Ireland and improve the overall stability and capability of Grid-Ireland for national users as well.

Participation in these top-level international projects means that Grid-Ireland is not only contributing to large-scale scientific endeavour, but is also abreast of the latest available tools and technology being developed by other groups. The result of this interaction means that Grid-Ireland has developed its own functional layer very rapidly. Receiving and having to cope with massive data transfer over the Grid from Europe has brought out many of the weaker areas of e-infrastructure for improvement. In this way international collaboration has already proven of dual-benefit to Grid-Ireland.

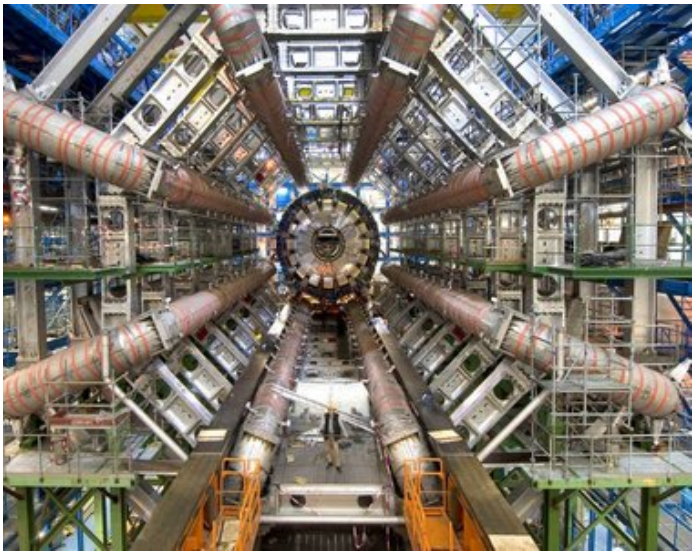

Figure 10: Part of the Large Hadron Collider (LHC) experiment apparatus at CERN.

Grid-Ireland participates in EGEE as a tier-2 member of the ATLAS (A Toroidal LHC ApparatuS) *Cloud* (a Cloud being the next level grouping above grids) Virtual Organisation in the Netherlands, and as a tier-2 member of LHCb (LHC Beauty, which examines the bottom quark, sometimes called the beauty quark). This LHCb tier-2 operation is in direct support of a particle physics group in University College Dublin (UCD), led by Dr. Ronan McNulty, that

is an physics participant in LHCb. As the CERN LHC facility (see Figure 10) is now active, Grid-Ireland is preparing for massive data transfer and full-scale utilisation of Grid and HEAnet resources.

The Grid-Ireland cluster at Trinity College Dublin is already processing CERN's data from ATLAS and LHCb via the Grid and has already computed over 300,000 CPU-hours of data. In a recent interview with Irish news service SiliconRepublic.com [6] Dr. Stephen Childs from Grid-Ireland gave three reasons for participation in the project: "There are three main benefits to our involvement with the LHC. Firstly, we are making a valuable contribution towards this significant global experiment. Secondly, we can directly support Irish scientists working to analyse LHC data. Thirdly, we are gaining valuable experience with grid technology, which we can then pass on to Irish researchers from all disciplines".

4 Grid Deployment and Middleware Management

Grids are composed of multiple physical sites in different administrative domains. The different sites in New Zealand have their own administrators and researchers, and different hardware and equipment. The challenge to Grid-enable each of these sites has required staff at each of these sites to self-train, and individually assemble a Grid middleware stack, working to some agreed guidelines in order to approach a consistent Grid interface and functional layer throughout the country. This process is extremely time-expensive, and the tinkering with and hacking of middleware and additional scripts and modules required to meet functional consistency is fraught with pitfalls, and difficult to keep consistent with updates as staff are not always available to man Grid administration duties at each site. Support and maintenance of Grid systems post-deployment is also a problem, as project budgets and time allowances do not always extend to allow for an administrator, Certificate Authority (for signing up new Grid users) and training and support staff for each individual administrative domain. The issue has already been encountered by Grid-Ireland, and they have developed a set of solutions to overcome these management and infrastructure problems. The tactics used by Grid-Ireland to counter these problems target software deployment and maintenance, manpower, and ownership of resources:

- Using the automated Quattor system to deploy and manage distributed infrastructures, and developing a transactional deployment system to support this
- Centralising manpower resources and employing remote administration techniques
- National-level ownership of Grid gatekeeper machines

Grid-Ireland are using a system called Quattor to manage their Grid gateway software. The Quattor system sits on top of a version management system, so multiple developers at various locations can be working on different parts of the next iteration of the software. The Gateways are all Xen host machines with various Grid servers hosted on these as virtual machines. When new gateway

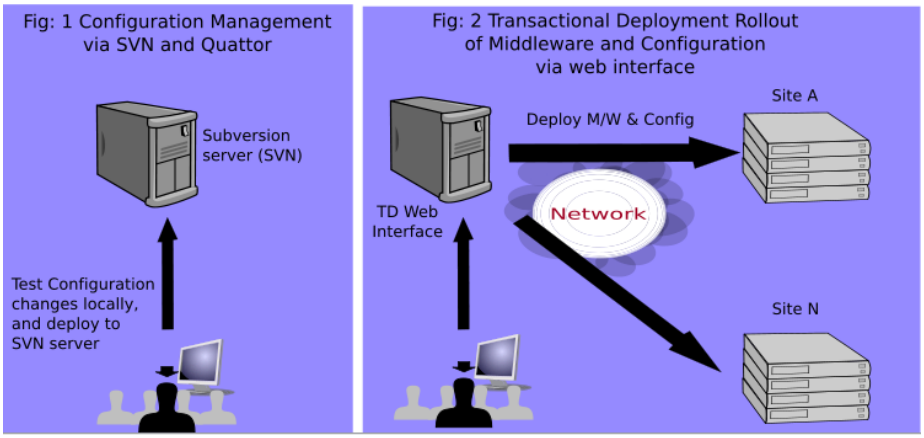

Figure 11: Transactional grid gateway rollout using quattor.

software is ready, the rollout is sent to a virtual *test grid* [7] where the software can be tested in a safe environment against a grid setup mimicking the real Grid (see the first panel in Figure 11). Once the Grid developers are happy with any software changes the new image can be automatically propagated to *all of the grid gateways in Ireland* [8] [9] [10] (see the second panel in Figure 11 and Figure 12 for a view of the remote deployment tool).

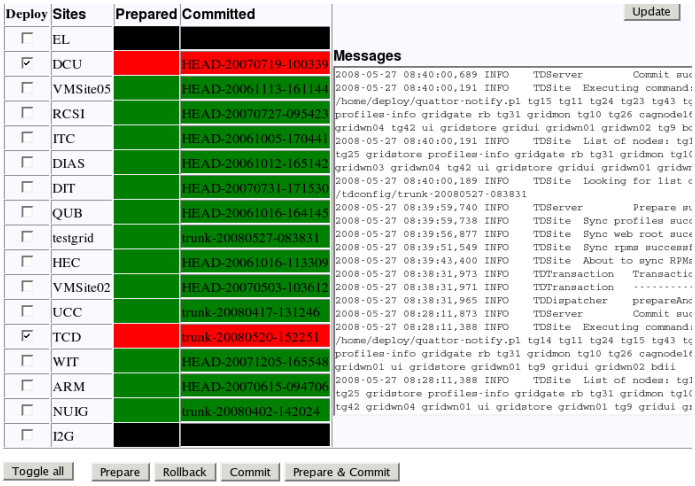

Figure 12: Deploying gateway software profiles from TestGrid to real Grid sites.

Deploying any new Grid gateway is also done via the Quattor system. Once a new site has installed some basic Quattor templates and connected to the network, the entire gateway system [11]; Xen host and virtual server machines can be set up in about 20 minutes. The deployment mechanism is becoming increasingly stable even with heterogeneous physical machines [12].

The effectiveness of this Grid-deployment and maintenance mechanism hinges on two additional factors however. Firstly, Grid-Ireland owns the physical gateway machines that are deployed at each grid site. This means that the Grid

Figure 13: Attending a Grid Operations Centre meeting at Trinity College Dublin. *Pictured (left to right): Eamonn Kenny, Stephen Childs, David O'Callaghan, Geoff Quigley, John Walsh, and John Ryan*

developers do not need to contact anyone for permission or notification before any maintenance, which considerably speeds up deployment and maintenance time, and reduces issues with firewalls and differing local IT policy. The Grid gateway machines at each site are fitted with remote administration equipment so that almost all maintenance happens remotely from the Grid Operations Centre at Trinity College Dublin. Secondly, the Grid development and administration staff are almost all centralised at the Trinity College Dublin campus. This is a much more efficient use of manpower and training, but quite a significant organisational difference to BeSTGRID. Grid-Ireland establishes their new gateways by sending one engineer to all of the new sites to set up the physical machines and uses Quattor to manage the software installation including the entire Operating System with required virtual machines.

To maintain the Grid in Ireland, the entire Grid-Ireland group at Trinity College Dublin sits in the Operations Centre once a week and monitors the status of all of the machines on the Grid (see Figure 13 for my experience in one of these meetings) via a collection of Grid Monitoring (GridMon) utilities (see Figure 14 for an example GridMon tool) that are displayed on a series large monitors. Grid administrators can then amend any required changes remotely or contact staff at local sites to inspect for any local problems. This approach frees up a lot of staff from administration and maintenance for more productive research and development.

5 Projects of Interest to NZ e-Research

5.1 WebCom-G

WebCom-G is a new Grid Operating System that offers an expressive programming model that automatically handles load-balancing, fault tolerance, and task allocation without burdening the application designer. WebCom-G supports a credential-based security model based on Keynote [13] credentials. Resource

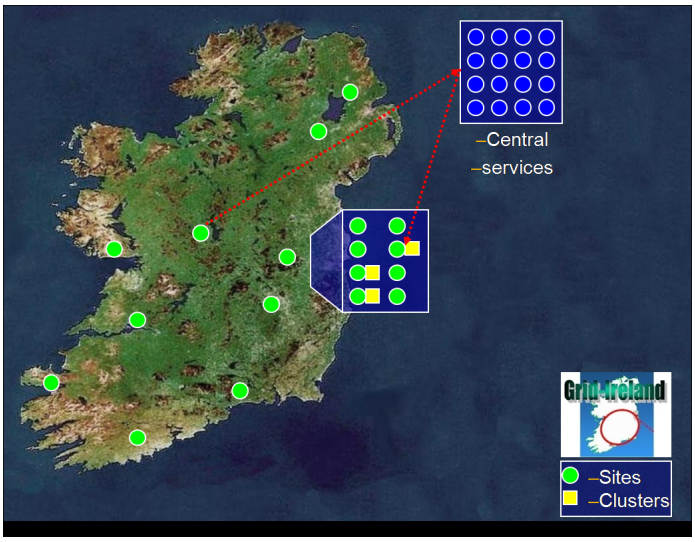

Figure 14: One of Grid-Ireland's Grid Monitoring utilities, from which the status of all Grid machines in Ireland can be observed.

WebCom-G
UCC TCD NUIG

Figure 15: WebCom-G - a collaborative middleware architecture project between 3 sites in Ireland

matching is also supported through an implementation of Condor Classads [14] so that the system can help the user find the best resources to match job requirements. These qualities of WebCom-G allow non-specialists to design complex, multi-stage, parallel and distributed workflows based on a graph methodology. The workflows constructed for Webcom are based on Condensed Graphs [15], which are architecture-independent and therefore can be transfered between heterogeneous environments [16] [17] [18]. Webcom-G is being used as an additional layer of Grid middleware above the existing Grid-Ireland middleware architectures.

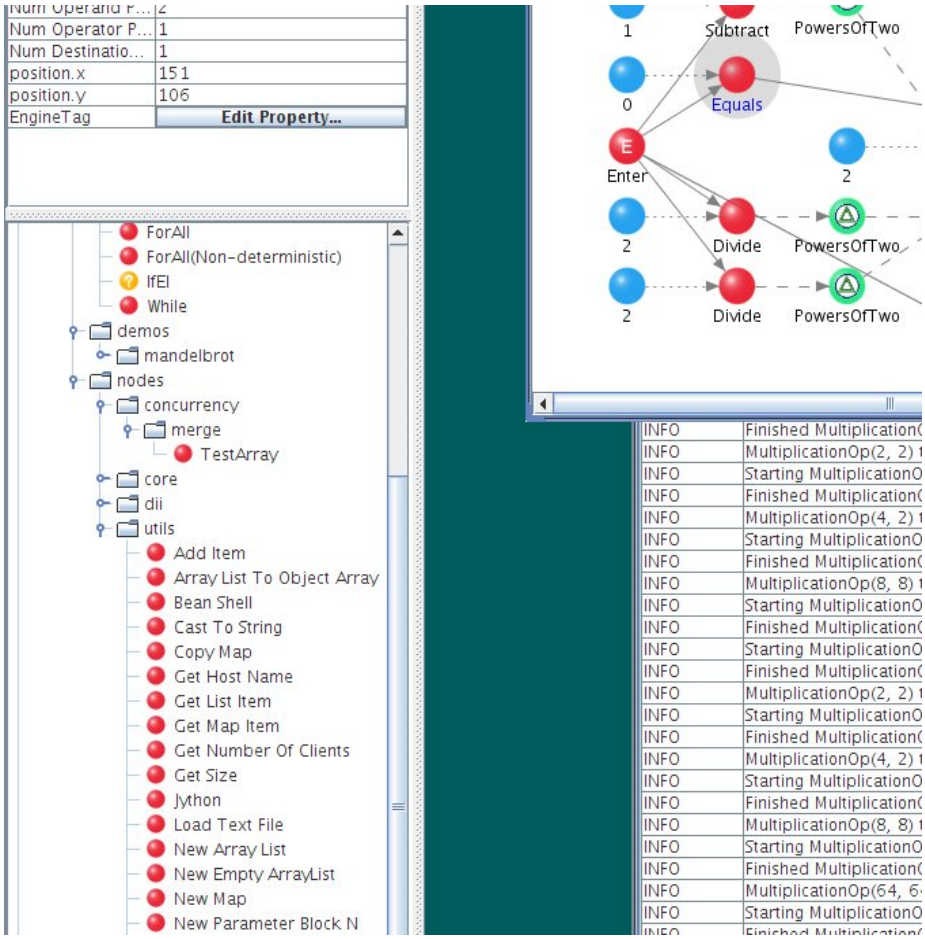

Figure 16: Changing the properties of a node in a WebCom-G IDE.

Figure 16 depicts the graph design process in an IDE that has been built for WebCom-G. Using this software, a user can *drag and drop* a graph representing their workflow from a variety of prefabricated and custom components; a list of which can be seen in the left-hand panel. Nodes in this graph are linked to represent communication between synchronous components. Multi-staged workflows can be created by introducing nodes into the graph that represent Boolean logic or mathematical operations; inculcating a flexible quality into the design of WebCom-G workflows. The whole workflow can be graphically de-

signed, and a designer can dictate that the workflow may change course entirely depending on the output or combined output of some of its operations.

A WebCom-G user can actually design and submit the whole job from with the graphical IDE, and even monitor its progress and fetch its results - bundling the entire process into one central point of design and control.

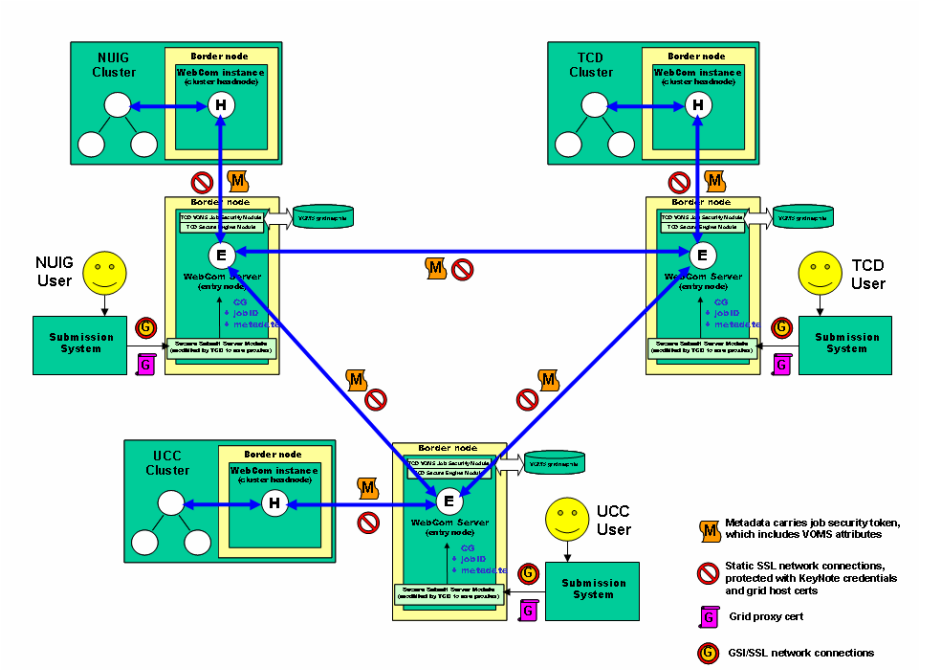

Figure 17: Webcom-G infrastructure allowing jobs between three campus sites in Ireland; National University of Ireland at Galway (NUIG), University College Cork (UCC), and Trinity College Dublin (TCD).

Although largely untested, Webcom-G should even be supportive of workflows distributed and running concurrently *between remote Grid sites*, for example, a workflow could be designed to make use of machines at two or three different campuses to make use of different computational abilities of machines at each site, combine the results, and store resultant data at a third location (see Figure 17 for how existing Webcom-G infrastructure in Ireland would support this capability).

Future works with Webcom-G aim to make the system interactive. Users can already visually monitor the execution of a Workflow from within an IDE, but developers would like to extend this to allow users to manually change the workflow path during execution (computational steering). The Centre for Unified Computing at University College Cork is currently undertaking a collaborative project with researchers in the University of Central Florida to implement a WebCom workflow for the three-dimensional reconstruction of virus structure. This application involves large datasets and execution time can sometimes run to several months. It is hoped that the addition of computational steering to this application will lead to higher quality results in less time [19].

The capabilities and future possibilities of middleware such as Webcom-G bring us closer to an *ideal* implementation of the Grid paradigm; truly distributed computing combined with powerful user tools. Webcom-G is a paramount example of the kind of capability that e-Research users and providers might look at moving towards in the next iteration of development in New Zealand.

5.2 eLearning

A considerable body of research in eLearning has been done by members of Grid-Ireland. During my visit in May I participated in a workshop on *PIC Simulations of Relativistic Collisionless Shocks* which featured eLearning tools developed by Kathryn Cassidy and Grid-Ireland researchers at Trinity College Dublin. As eLearning is a subject of research interest with BeSTGRID [20] the eLGrid project [21] [22] at Grid-Ireland should be of interest to eLearning researchers in New Zealand.

Kathryn describes the eLGrid project as "A Grid-Ireland initiative to develop an eLearning platform for delivery of Grid training and education. The project comprises two components, the eLGrid eLearning system and the eLGrid training infrastructure (t-Infrastructure). The eLearning system is a web-based application using the Adaptive Personalised eLearning Service (APeLS) developed by the Knowledge and Data Engineering group in Trinity College Dublin.

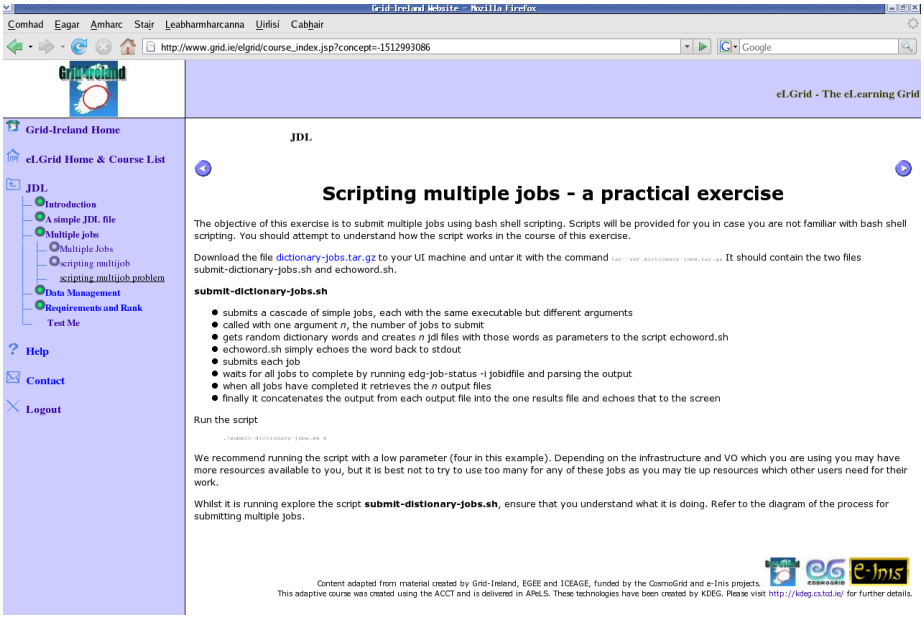

Figure 18: An eLGrid session teaches the user how to manage more complex Grid jobs.

The t-Infrastructure is a complete replica of the Grid-Ireland production infrastructure using virtual resources. By using adaptive eLearning technologies we hope to create personalised courses which will adapt to the learner's knowledge as measured by simple quizzes and tests and also to more complicated

practical assignments which must be carried out in the t-Infrastructure. Close integration of the eLearning system and the t-Infrastructure environment will allow us to capture the results and outputs of the learners' practical exercises and use these to personalise the courses."

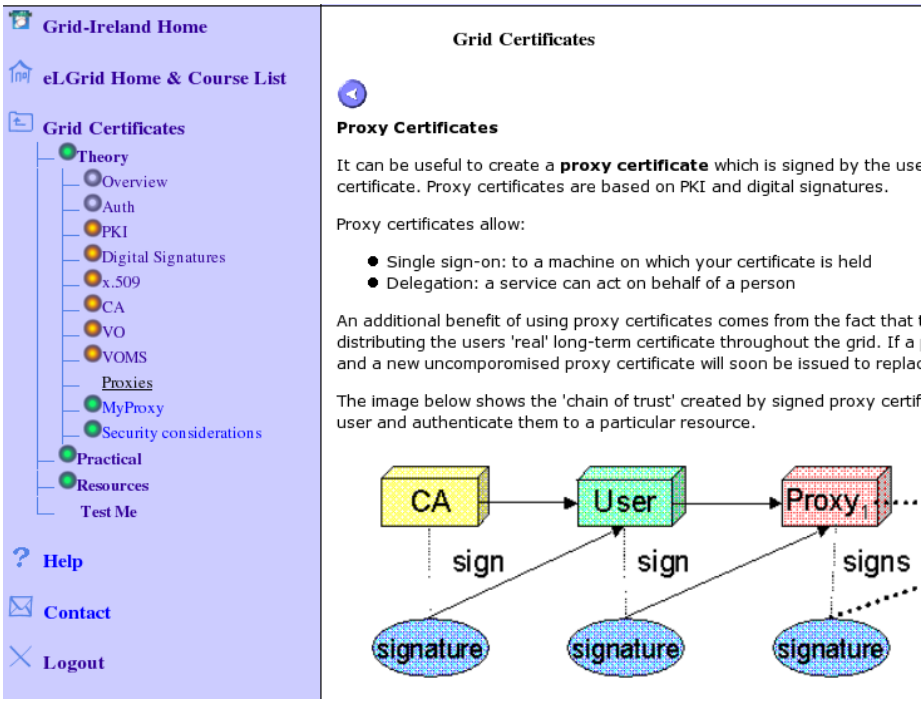

Figure 19: The front-end of the eLGrid system.

The eLGrid system makes use of Grid credentials to provide seamless access to the system. As a participant in the workshop I logged-in to the eLGrid system via a web browser using my Grid credentials (illustrated in Figure 19) and was taken to selection of interactive training courses (illustrated in Figure 20) which could be tutor-assisted or self-taught using a range of media and interactive examples and integrated tools. The courses taken included training sessions for Grid users, which would be particularly useful in New Zealand where staff available to train and support prospective Grid users are few and far between.

I personally found the eLGrid system to be a very efficient method for learning some of the more complex Grid and programming concepts; a sample course developed by Kathryn explained the Grid from the ground-up (no mean feat even for an expert), and was able to tackle and teach difficult concepts like MPI job construction to absolute beginners, using a whole range of learning tools - written, graphical, interactive, tutorial-like and practical 18. Figure 21 illustrates a tutorial guiding a user through an MPI workflow-based job construction using a tool called *Ganga* which is based on the gLite Grid middleware used by Grid-Ireland and allows users to assemble jobs using a graphical interface. As a user progresses through an eLGrid training session their progress through course sub-sections is visibly tracked, indicating modules that have been satisfactorily completed or mastered, and those which require further attention.

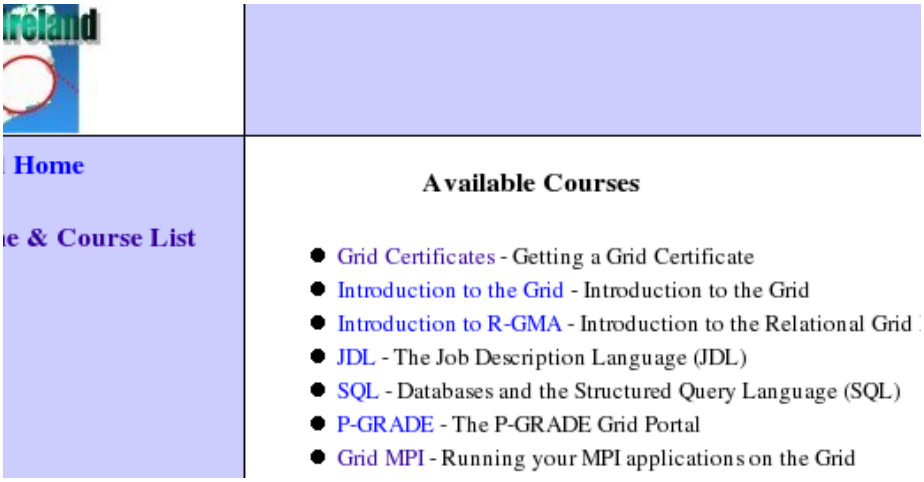

Figure 20: eLGrid presents the user with a selection of courses.

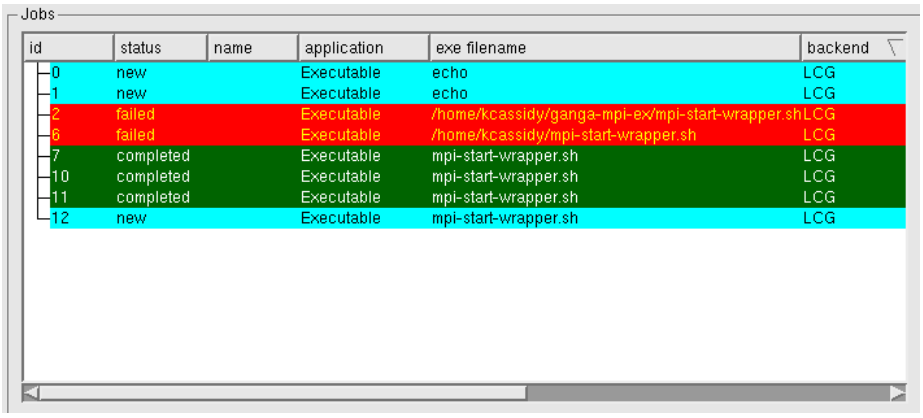

Figure 21: Monitoring submitted jobs via the Ganga programme.

5.3 Marine Grid

The Marine Grid project Grid-enables the current phase of Irish national seabed mapping in order to connect the more data-intensive and economically sensitive project with mass storage and security resources available over the Grid. The Irish National Sea-bed Survey has already mapped substantial areas of seabed surrounding Ireland, and has lead the world in this area; Ireland is only the third country to attempt a project of this type. The previous project generated 5.5 Terabytes of data, but the current iteration of the project, which will map sensitive and economically valuable priority bays around Ireland (see Figure 22) requires the managed and secure data and mass storage resources of the Grid. The Marine Grid in Ireland is funded by the Higher Education Authority and is a collaboration between the Geological Survey of Ireland, the Marine Institute and the four main Universities in Ireland.

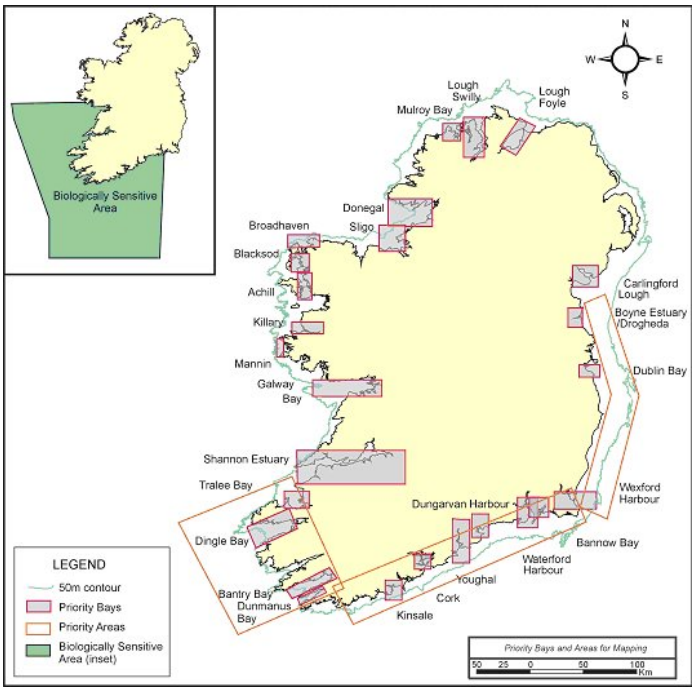

Figure 22: Priority bays to be mapped by using the Marine Grid.

I have included details of this project as I believe that there may be interest in a similar project in New Zealand, given the importance of marine resources to New Zealand and the availability of Grid and broadband resources. Further details about Marine Grid can be found at *http://www.scg.nuigalway.ie/marine-grid/* The Geological Survey of Ireland's Seabed mapping project can be found at *http://www.gsiseabed.ie/*.

6 Recommendations

I have been directly involved in the establishment and development of Grid and e-Research infrastructure in New Zealand over the last two years, and also had the opportunity to observe grid development in Australia, large-scale grid cluster deployment at the Texas Advanced Computing Center (TACC) as part of Teragrid in the United States of America, the latest grid research by INRIA at Lyon in France, studied within Grid-Ireland, and participated in many international Grid and e-research conferences, appreciating a sample of developments in these areas through much of the world. From these experiences it is my strongest recommendation that any further iteration of Grid development in New Zealand emulates first and foremost the structure and direction, and hopefully also the success of the Grid organisation in Ireland.

We have gained much experience in New Zealand from a close partnership with Grid groups across the Tasman, and the generosity of these groups has leveraged New Zealand to Grid-enabled status in a very short amount of time. However, as momentum is gained by e-Research infrastructure here, and New Zealand looks to become more autonomous, we need to look to role models of a similar national character; Grid-Ireland is an excellent example of how a relatively small nation can organise itself into a very effective provider to scientists nationally, stimulate greater collaboration between Universities or Colleges, and enable participation in high-level international scientific projects.

From my experience with Grid-Ireland, specific recommendations that I make for the next iteration of Grid or Grid-like infrastructural development in New Zealand are:

- Maintain a communicative and collaborative relationship with Grid-Ireland.
- Take on Quattor or a similar software-deployment and roll-out management system for rapid deployment and maintenance of Grid sites in New Zealand.
- Establish a Grid-dedicated cluster to avoid organisation-internal resource competition and have an available platform for international project participation.
- Consider building a National Data Centre for managed storage for researchers nation-wide
- Begin to participate in international Grid projects to shake-out any flaws or bottlenecks in e-infrastructure and prepare for large-scale participation
- Adopt workflow and MPI tools such as WebCom-G to provide our researchers with the grid tools that they will actually use
- Further publicise NZ as an e-research destination to attract experts from overseas and diversify local grid understanding
- Continue to follow progress and development of grid groups outside of NZ/Australia
- Continue the links with the Australian Grid in the spirit of a federation, just as Grid-Ireland maintains strong links with the UK Grid under the aegis of the EGEE UK/I grid federation.

Glossary

APAN26 The *Asia Pacific Advanced Network Consortium* meeting is Asia Pacific's foremost event for showcasing advanced broadband networking applications and technologies for research and education, 2

ATLAS Cloud The ATLAS Cloud is a Grid computing Virtual Organisation combining the national Grids of several European nations (led by the Netherlands) to compute data from the ATLAS Large Hadron Collider (LHC) particle detector experiment at CERN, 9

BeSTGRID *Broadband-enabled Science and Technology Grid*; New Zealand's grid organisation, ii

Celtic Tiger The Celtic Tiger (*Tíogar Ceilteach*) is a term used to describe the economy of the Republic of Ireland during a period resurgent booms from the 1990s to the current date, which transformed the Republic of Ireland from one of Europe's poorest economies to one of the richest, and has dramatically increased the standard of living in Ireland. The Celtic Tiger is usually attributed to strong state-driven economic policy, the long-term benefits of consistent investment in higher education, and Europen Union membership, 1

Cloud Computing Cloud Computing is the next level higher over Grid meta-computing. A *Cloud* is a collection of coorating national Grids, 9

DIAS The *Dublin Institute for Advanced Studies* is a publicly-funded research institution comprising three schools; The School of Celtic Studies, The School of Theoretical Physics, and The School of Cosmic Physics. DIAS is the leading member of e-Inis, 2

e-Inis Ireland's leading e-Research initiative, 2

e-Research e-Research (or *eResearch*) is a supportive research field, concerned with developing tools for collaboration, communication, education and distributed scientific research over networks and facilitating access to these resources for all scientists, ii

EGEE The *Enabling Grids for E-sciencE* is a European Commission-funded project to connect the grids of 27 countries into a large European grid using a Virtual Organisation hierarchy and custom grid middleware (gLite), and supporting and training a large range of users from science and industry, 4

Grid A system of distributed access to remote, loosely coupled resources analogous to a power grid, 1

Grid gatekeeper A Grid gatekeeper or Grid gateway is a model of Grid-resource access, where each site exposing resources to the Grid does so through a portal of one physical machine which has been loaded with several Xen virtual server machines which each provide a different service to connecting Grid-users., 1

Grid Operations Centre A *GOC* or *OpsCentre* is a central point of systems monitoring and administration for a Grid project, and is run by regular meetings at a central physical centre or *via* video conferencing where Grid site status is reviewed, 4

Grid-Ireland Ireland's grid organisation, covering 18 sites throughout the Republic of Ireland and Northern Ireland, ii

HEAnet *Higher Education Authority Network* is an advanced national broadband research network - Ireland's equivalent to KAREN, 1

HPC *High-performance computing* is a field that makes use of supercomputers and cluster computers (large inter-connected managed groups of smaller computers) to provide large amounts of processing power to solve complex computational problems, 2

ICHEC The *Irish Centre for High-End Computing* is a distributed national centre providing support for research in high-performance computing (HPC) and computational science in the Republic of Ireland. ICHEC has 14 support staff and 4 supercomputers including two new IBM Blue Gene machines. http://www.ichec.ie/, 2

ICT *Information and Communication Technology*; a blanket grouping of network-related technology that is often used by Government, business and media to refer to public telecommunications and Internet providers as well as loosely related academic or scientific activities, ii

IDE *Integrated Development Environments* are computer software suites that provide a user with a selection of design tools that help to simplify the construction of computer programmes and often provide graphical user-friendly interfaces that hide the more complex details that are being constructed behind them, 14

KAREN *Kiwi Advanced Research and Education Network* a broadband network connecting major universities and research institutions in New Zealand, ii

MPI *Message Passing Interface* is a method for computer programmes that have been designed to run in a distributed fashion over several machines in a cluster computer to communicate between the different parts of the programme being run on separate nodes or machines, 2

Virtual Machine A virtual machine is a running copy of a complete computer system that can reside inside another computer system. In this way a single physical computer can be running several systems (for example two different Linux servers and a Windows machine for a different task) simultaneously., 1

Virtual Organisation Virtual Organisations, or *VO*s, are groups under an e-Research organisation (such as the Grid) that have agreed to share resources linked over networks, 1

Webcom A complete middleware architecture supportive of Condensed Graph workflows in a parallel or distributed computing environment., 12

Webcom-G An adaptation of the Webcom distributed computing architecture to a Grid middleware, allowing Condensed Graph workflow execution over distributed Grid resources, 12

Xen One of several different technologies available to simultaneously host several *virtual machines* on a single computer's hardware., 1

References

[1] Irene Mia and Soumitra Dutta. The Global Information Technology Report 2007-2008. Technical report, World Economic Forum, 2008.

http://www.weforum.org/en/initiatives/gcp/Global Information Technology Report/index.htm.

[2] Stephen Childs, Brian Coghlan, David O'Callaghan, Geoffrey Quigley, and John Walsh. A single-computer grid gateway using virtual machines. In *19th International Conference on Advanced Information Networking and Applications (AINA'05)*, volume 1, pages 310–315, Taiwan, march 2005. IEEE.

[3] Antony P. Gerdelan, Martin J. Johnson, and Chris H. Messom. Performance analysis of virtualised head nodes utilising cost-effective network attached storage. In *Australian Partnership for Advanced Computing*, number ISBN 978-0-9579303-3-9, 2007. Winner of Best Paper Award and Best Presentation Award at APAC07.

[4] Hannah Fearn. After the Gold Rush. *Times Higher Education*, No. 1,855:pp. 36–39, 24-30 July 2008.

[5] John P. Morrison, David A. Power, Brian C. Clayton, Adarsh Patil, Philip D. Healy, and James J. Kennedy. *WebCom-G: Middleware to Hide the Grid*, chapter Grid Technologies Emerging from Distributed Architectures to Virtual Organizations, pages 271–303. WIT Press, 2006. ISBN 1845640551.

[6] Marie Boran. Trinity college researchers tackle big bang data. News article, SiliconRepublic.com, 09.09.2008 2008. http://www.siliconrepublic.com/news/article/11392/randd/tcd-tackles-big-bang-data.

[7] Stephen Childs, Brian Coghlan, and Jason McCandless. Gridbuilder: A tool for creating virtual grid testbeds. In *2nd IEEE International Conference on e-Science and Grid computing (eScience 2006)*, page 77, Amsterdam, The Netherlands, December 2006.

[8] Stephen Childs, Brian Coghlan, David O'Callaghan, Geoffrey Quigley, and John Walsh. The second-generation grid-ireland deployment architecture with quattor centralised fabric management. In *Cracow Grid Workshop 2005*, pages 114–123, Crakow, Poland, 2005.

[9] Eamonn Kenny, Brian Coghlan, John Walsh, Stephen Childs, David O'Callaghan, and Geoffrey Quigley. Autobuilding multiple ports of computing nodes for grid computing. In *Cracow Grid Workshop 2005*, number November, pages 232–239, 2005.

[10] Stephen Childs and Brian Coghlan. *Advances in Grid Computing*, volume 4854 of *Lecture Notes in Computer Science*, chapter Integrating xen with the quattor fabric management system, pages 21–223. Springer, Rennes, France, August 2008. High-Performance Cluster and Grid Computing (VHPC2007, part of Euro-Par 2007)IRISA, Rennes, France, 28-31 August, 2007.

[11] Brian A. Coghlan, John Walsh, and David O'Callaghan. *Advances in Grid Computing - EGC 2005*, volume 3470/2005 of *Lecture Notes in Computer*

Science, chapter Grid-Ireland Deployment Architecture, pages 354–363. Springer-Verlag, Berlin and Heidelberg, Deutschland, July 2005. ISBN: 978-3-540-26918-2.

[12] Eamonn Kenny, Brian Coghlan, John Walsh, Stephen Childs, David O'Callaghan, and Geoffrey Quigley. Heterogeneity of computing nodes for grid computing. In *European Grid Conference (EGC05)*, Amsterdam, The Netherlands, February 2005.

[13] Simon N. Foley, Thomas B. Quillinan, John P. Morrison, David A. Power, and James J. Kennedy. Exploiting keynote in webcom: Architecture neutral glue for trust management. In *Fifth Nordic Workshop on Secure IT Systems*, Reykjavik, Iceland, October 2001. Department of Computer Science, University College Cork, Ireland.

[14] Douglas Thain, Todd Tannenbaum, and Miron Livny. *Grid Computing: Making the Global Infrastructure a Reality*, chapter Condor and the Grid, page . Wiley Series in Communications Networking & Distributed Systems. Wiley, Department of Computer Sciences, University of Wisconsin-Madison, 1210 W. Dayton Street, Madison, Wisconsin 53706-1685, United States, May 2003. ISBN: 9780470853191.

[15] John P. Morrison. *Condensed Graphs: Unifying Availability-Driven, Coercion-Driven, and Control-Driven Computing*. PhD thesis, University of Eindhoven, The Netherlands, October 1996. ISBN: 90-386-0478-5.

[16] John P. Morrison, Philip D. Healy, David A. Power, and Keith J. Power. The role of XML within the WebCom metacomputing platform. *Scalable Computing: Practice and Experience*, 6(1):33–43, March 2005.

[17] Philip D. Healy. *Metacomputing on Clusters Augmented with Reconfigurable Hardware*. PhD thesis, National University of Ireland, Cork, May 2006.

[18] Neil Cafferkey, Philip D. Healy, David A. Power, and John P. Morrison. Job management in webcom. In *6th International Symposium on Parallel and Distributed Computing (ISPDC)*, Hagenberg, Austria, July 2007. IEEE Press.

[19] Chen Yu, Dan C. Marinescu, John P. Morrison, Brian C. Clayton, and David A. Power. An automated data processing pipeline for virus structure determination at high resolution. In *2007 IEEE International Parallel and Distributed Processing Symposium*, page 262. Computer Science Department, University College Cork, Cork, Ireland and School of EECS, University of Central Florida, Orlando, FL 32816 USA, 2007.

[20] C.H. Messom, A. Sarrafzadeh, A. P. Gerdelan, M.J. Johnson, and J. Shanbehzadeh. Operating system virtualization to support e-learning with affective intelligent tutoring systems. In *IEEE International Conference on Innovations in Information Technology*, number ISBN 978-1-4244-1841-1, pages pp. 143–147, 2007.

[21] Kathryn Cassidy, Stephen Childs, John Walsh, Brian Coghlan, and Declan Dagger. eLGrid - Adaptive eLearning in a Virtual Grid Environment. *Awaiting Publication*, .:., 2008.

[22] Kathryn Cassidy, Jason McCandless, Stephen Childs, John Walsh, Brian Coghlan, and Declan Dagger. Combining a virtual grid testbed and grid elearning courseware. In *Cracow Grid Workshop 2006 (CGW06)*, Cracow, Poland, 2006. Academic Computer Centre CYFRONET AGH.