



# eResearch at VUW: the eScience Consultant's Tale

---

**Kevin M. Buckley**

School of Engineering and Computer Science

Victoria University of Wellington  
New Zealand

[Kevin.Buckley@ecs.vuw.ac.nz](mailto:Kevin.Buckley@ecs.vuw.ac.nz)



eRA14: eResearch at VUW: The eScience Consultant's Tale

## The Abstract

In 2008, Victoria University of Wellington finally filled the post of eScience Consultant, although back then the post had been drawn up with the title of eResearch Programmer, not that either name really gives much insight into the range of facilitation that the role has provided.

This tale will highlight recent eResearch activity at the capital city university, from the viewpoint of its eScience Consultant and in light of the experiences of the last six years.

## Outline

- Cycle-stealing Grids vs Power-saving Initiatives
- How VUW's Science Faculty got its HPC facility - and what it's been used for
- SKA/MWA Activity at VUW - and how it aided 10GB network research at the Univ. of Auckland
- Distributed computing - why don't computer science students get it?
- The effects of MATLAB on researcher programming skills
- The Death of the Institution ?

HTML and PDF renditions of the slides should always be available here:

<http://www.ecs.vuw.ac.nz/~kevin/Conferences/eRA14>



eRA09: Kevin Buckley: A Grid-Based Facility for Large-Scale Cross-Correlation of Continuous Seismic Data

And finally ... who is supposed to do this "no-boundary" science ?

Domain scientists ?

- Might need to brush up on their BEPL, SCUFL, WSDL, MPI and DRMAA

Workflow/Grid computer scientists ?

- Might need to read up on SEED, SAC, QuakeML

Phew !

Might still be a few jobs for people who can straddle the boundaries

and, indeed, that "theme" has been picked up in a few eRes Aus talks since then, including GA's Lesley Wyborn's talk from 2102 that explicitly mentioned a

- Comparison of eResearch Scientific vs Technical Teams



eRA14: eResearch at VUW: The eScience Consultant's Tale

## Cycle-stealing Grids vs Power-saving Initiatives

- ECS Grid: 300-ish Eng and Comp Sci UNIX workstations - SGE
- SCS Grid: 900/600/49 VUW public lab windows machines - Condor

These grids are operated by the School of Engineering and Computer Science and not VUW's central IT facilitator, although there's not that much "operating" needed, given that they make use of existing resources.

Early SCS usage was UNIX/Cygwin: some recent projects have been windows-native:

- The Zoological Society of London's Colony program (genotype sibship and parentage)
- Compiled MATLAB simulations of historic stock pricing strategies

VUW's central IT facilitator were recently asked to look at power saving across the "enterprise"

Possibly because they don't operate the SCS Grid, the "solutions" considered by VUW's enterprise IT facilitator didn't consider any effects on the operation of the Grid but identified a commercial product that they could purchase, so as to do "what was required"



eRA14: eResearch at VUW: The eScience Consultant's Tale

## Cycle-stealing Grids vs Power-saving Initiatives 2

Somewhat belatedly, a fuller investigation, as to how the free software scheduler and the commercial product play together, was started, and here's what was found:

- Condor could already do everything that's needed in terms of powering machines down

Hardly a surprise - once/if you stop to think about it - because:

in order to schedule jobs into idle cycles, the scheduler needs to know how resources are being used and so, by running such a grid, you already know machines are, and/or have been, idle and so, instead of accepting jobs, you could power them down - however,

- Some infrastructural changes are needed to allow the free software to power machines up

Bad! The grid uses spare cycles within existing infrastructure: can't dictate it - however,

- Some infrastructural changes are needed if the commercial product is to power machines up

Good! VUW's IT purchasers also operate the infrastructure, so can just change it to fit.

Better still ? Wouldn't you know it, these are the same changes Condor would need!

It's possible to save power as a by-product of deploying free cycle-stealing grid software and yet VUW already was, before it started looking around for something to purchase



eRA14: eResearch at VUW: The eScience Consultant's Tale

## MWA-related activity

The Murchison Widefield Array (MWA) is one of the precursor projects to the Square Kilometer Array.

VUW is very fortunate to have Melanie Johnston-Hollitt within in the SKA project, as well as having her current research group, based within VUW's Science Faculty, working on data from the MWA programme, a key component of which, a 24-node IBM iDataplex system, is Melanie's

Melanie even "rescued" New Zealand's initial tranche of MWA-related hardware, after it moved from AUT and ended up over in WA, about the same time as an AUT researcher moved to UWA.

That computational resource, based around NVIDIA Tesla C2070 cards (a RADON Duo 2U chassis) was provisioned by Xenon Systems, one of the regular sponsors of eResearch Australasia.

Not content with just working on the MWA data, Melanie has also initiated the deployment of a New Zealand "data node", running the same NGAS platform as MIT, and RRI in India, to mirror data from the main SKA/MWA archives in WA

New Zealand's "data node" was provisioned by SGI, who put MWA/SKA-related kit into iVEC, and currently comprises an SGI IS5000 "tray", housing 96TB. (Melanie's data ~160TB)



eRA14: eResearch at VUW: The eScience Consultant's Tale

## MWA-related activity 2

That 2U chassis gave us the ability to have two separate, from a networking viewpoint, nodes.

In order to avoid data bottlenecks at VUW's edge (think: deep packet inspection), REANNZ (think: AARNET) helped Melanie facilitate what's now, in effect, a 10GbE "Science DMZ"

Even before moving things to 10GbE (a story in itself), we, VUW, and folk at UoA/CeR/NeSI had "maxed out" this REANNZ private 1GbE circuit in some GridFTP testing.

Roll forwards to a dinner, alongside the river, one night at last year's eRes Aus in Brisbane. I, and a few people from UoA/CeR/NeSI were together and they were bemoaning the lack of any 10GbE capability for a UoA researcher who wanted to do some network profiling tests.

Of course, I was then able (VUW isn't part of NeSI to the extent that this kind of knowledge flows easily within NZ anymore) to tell of the previous testing that'd been done, and suggest that, when they got back to NZ, they get Nevil Brownlee's group access to their own resource.

Sadly, the Auckland folk found that their central IT people had been allowed to take away their old "research DMZ" capability, so Nevil had to wait a while, whilst UoA caught up with VUW.

Everyone's aware that VUW is currently rated #1 for research in New Zealand ?



eRA14: eResearch at VUW: The eScience Consultant's Tale

## Science Faculty HPC Facility

VUW's Science Faculty only got an HPC Facility after

a lecturer moved from Massey but, because of the then existing BeSTGRID community, was then able to contact VUW's eScience Consultant to ask what resources VUW had and, on discovering that VUW had next to nothing, bar a grid of desktop PCs running windows, decided to not keep his head down at his new home, but to ask for some HPC kit, leading to (Doppler effect studies using the sirens of the paramedics heading to/from VUW) the Dean of the Science Faculty showing some vision, and seeing his Faculty obtain

25 computers, 52 CPUs, 624 cores, 1920 GB RAM, IB interconnect

- 6-off, SGI C2112, 2x12-core AMD Opteron 6174, 64GB RAM (4-node units)
- 1-off, SGI H2106, 4x12-core AMD Opteron 6174, 512GB RAM
- RHEL5 OS hosting an SGE local resource manager (note: no vendor appl. stack)

Since added:

- 2-off, SGI C2112 2x16-core AMD Opteron 6174, 64GB RAM (4-node units)
- 1-off, SGI ISS3500, which houses around 30TB storage
- OS upgrade to CentOS6
- So, currently at: 784 cores with access to 64GB, and the 0.5TB node.



eRA14: eResearch at VUW: The eScience Consultant's Tale

## Science Faculty HPC Facility: 2

Main users of the facility have been computational chemistry research groups within VUW's School of Chemistry and Physical Sciences (SCPS), one led by that lecturer who moved from Massey, Matthias Lein, and more recently, Nicola Gaston, who moved into SCPS from IRL, but brought her MacDiarmid Institute PI funding with her.

Their research focuses around `Gaussian` and `VASP`, respectively, although Nicola has been looking at deploying `Crystal` but found the code's Italian authors operate a "code of silence"

The large memory node, originally slated for use by VUW's School of Biological Sciences (SBS), gets, as a result of the researcher driving its acquisition leaving for ANU, sporadic use by SBS researchers for `BLAST` searches, with a recently completed PhD project around protein-docking studies using `RosettaLigand` although without a large memory footprint.

In terms of "really testing the beast" we've had to rely on an School of Engineering and Computer Science project which used `COMSOL` to study far-field superlens effects, which touched 360 GB.

Large usage has also come from our Faculty of Architecture and Design, where a combination of `GenOpt` and `EnergyPlus` are being used for optimisation studies of the energy performance of New Zealand's commercial building stock.



eRA14: eResearch at VUW: The eScience Consultant's Tale

## Science Faculty HPC Facility: 3

So, the Facility has seen quite a range of projects and disciplines and so here are some of the issues that people, across those disciplines, seem to have when using it

### ■ Transition from PC to HPC

From as simple as not really "getting" directories or job submission scripts to treating the shared resource as their own and writing job-scheduling "daemons" (read: script-kiddy self-spawning python scripts)

### ■ Knowledge transfer within research groups/communities

New users in a research group look to (or are steered to?) the resource facilitator, not to their group, for initial help (recall that I only claimed to straddle the boundaries!)

Equally, code authors happy to share the code, but not to correspond about it

So, unless you can match the development environment, you may not be able to run supposedly "open source" codes





eRA14: eResearch at VUW: The eScience Consultant's Tale

## Here's one way to unroll loops for distributed processing

Code is from learning classifier system research, and needed to be run "embarassingly parallel":

```
... main(int argc, char **argv){ ...  
    //for(int j = 0; j < MAX_NUM_RUNS; j++) {  
    for(int j = 49; j < 50; j++){  
    ...
```

What the researcher is doing here is clear: but why do they think editing a loop to run once for a specific value, thus requiring a recompilation of each job's binary, was "the way to go" ? This:

```
if ( 1 <= argc ) {  
    int j ;  
    sscanf(argv[1], "%d", &j ) ;  
    ...  
} else {  
    for(int j = 0; j < MAX_NUM_RUNS; j++) {
```

also allows for a specific value to be provided at invocation, or see the default, "loop over all", be run, with no recompilation of the binary - once the traditional run-debug-edit cycle is finished.

And I blame MATLAB !



eRA14: eResearch at VUW: The eScience Consultant's Tale

## The effects of MATLAB on researcher programming skills

Here's a piece of a MATLAB program, from the rainfall-runoff modelling arena, that does something similar to the AI program we saw before

```
runs=10000

for i=1:runs
    modrunoff=[]; Vsoil=[]; of=[]; ae=[]; ir=[];vi=[];vf=[];vb=[];
    for j=1:ceil(totalmodelsteps/20000)
        ...
    end
end
```

Here's one that can read from the command line and so can be compiled to run standalone

```
function [] = runmodel_ECS(catchno,runs,commontag,uniquetag)
...
if ischar(runs)==1
    runs=str2num(runs);
end
...
for i=1:runs
    ...
end
```

but MATLAB users, cosy and warm inside their GUI, have no command line, so won't do that ...



eRA14: eResearch at VUW: The eScience Consultant's Tale

## The effects of MATLAB on researcher programming skills: 2

... but nor do many MATLAB users, unaware of simple, pipeline, programming practice, akin to

```
+-----+           +-----+           +-----+  
| input_file | =is read by=> | binary | =writes to=> | output_file |  
+-----+           +-----+           +-----+
```

write something akin to the following, based on an example from MATLAB's own help resources, so as to get their input values from a file, and not hard-code values into the program:

```
infile = 'input.txt'  
fid = fopen(infile);  
tline = fgetl(fid); catchno=str2num(tline);  
...
```

It's a great shame our researchers are not encouraged to: they could still edit values inside the GUI, as they're used to, but also compile the very same M-files to standalone programs. Same shameful story with plotting images on-screen before saving the image to a file: it's unnecessary but not well documented or promoted!

What if "MATLAB programmers" were being taught programming basics: not just MATLAB?



eRA14: eResearch at VUW: The eScience Consultant's Tale

## The Death of the Institution ?

Question prompted by some comments from last year's eRes Aus suggesting that tertiary education institutions needed to replicate their traditional content inside facebook - really?

- Institutional web presence controlled by its "marketing silo", rather than the content providers, the academics and researchers themselves
- Students, but increasingly academics, use `@gmail.com` addresses, not `@vuw.ac.nz`  
Sam Searle's talk, pointed you to her `samsearle.net` domain, not an institutional one.
- Institution's central IT facilitators often just "purchase products" rather than "provision protocols"  
Latest trend perhaps? Your institution's hardware will have been purchased from "the cloud" ?

Problem for institutions is that their IT facilitators, in operating what they like to call "enterprise IT" are making it hard to do "simple stuff": and researchers simply go off and do it themselves!  
And it gets worse, because the supra-institutional support bodies, understandably, want to go direct to the researchers, and so further bypass any institutional-level insight into research

Institution's IT facilitator becomes even less able to offer "solutions" that are in anyway relevant



# Colophon

---

Slides for this talk have been created and delivered using MagicPoint

<http://member.wide.ad.jp/wg/mgp/>

An X11 based presentation tool which has the slide sources in plain text and which also provides for creation of an HTML slide-set.



eRA14: eResearch at VUW: The eScience Consultant's Tale

## References

<http://www.eresearch.edu.au/buckley2009>

<http://homepages.ecs.vuw.ac.nz/~kevin/Conferences/eRA09/>

[http://eresearchau.files.wordpress.com/2012/11/01\\_lesley\\_wyborn.pdf](http://eresearchau.files.wordpress.com/2012/11/01_lesley_wyborn.pdf)

<https://ecs.victoria.ac.nz/EResearch/ScienceFacultyHPCFacility>

[http://technologyinterface.nmsu.edu/5\\_1/5\\_1f/5\\_1f.html](http://technologyinterface.nmsu.edu/5_1/5_1f/5_1f.html)

(Converting M-Files to Stand-Alone Applications - from 2003!)

[http://www.mathworks.com.au/help/matlab/import\\_export/import-text-data-files-with-low-level-io.html](http://www.mathworks.com.au/help/matlab/import_export/import-text-data-files-with-low-level-io.html)

[http://www.mathworks.com.au/help/matlab/import\\_export/import-text-data-files-with-low-level-io.html](http://www.mathworks.com.au/help/matlab/import_export/import-text-data-files-with-low-level-io.html)

(As one line to aid cut and paste)

<http://www.dsprelated.com/showcode/223.php>

(Mentions the un-documented "hgexport" for figures)



eRA14: eResearch at VUW: The eScience Consultant's Tale

## The Author Bio

Kevin Buckley is currently employed as the eScience Consultant at the Victoria University of Wellington, NZ, based in the School of Engineering and Computer Science, a role which he entered after following the England cricket team's tour of New Zealand in 2008, during which his resume ended up on the desk of Prof. John Hine, currently VUW's Director of eResearch. (A post likely to disappear!?)

His biography for eResearch Australasia 2009 stated that he was "attempting, within frameworks provided by NZ's MoRST-funded BeSTGRID organisation and BeSTGRID's federation into the trans-Tasman ARCS community, to evangelise and enable e-Research across a institution which is just taking its first steps towards facilitating academic research IT needs above and beyond an existing, windows-desktop-computing-focused, support provision. Kevin finds the role an interesting, if sometimes frustrating, challenge."

Although the last six years have seen both New Zealand's BeSTGRID and MoRST, and Australia's ARCS, long since superceded and/or replaced, Kevin's institution still has the windows-desktop-computing-focused, support provision.

Kevin thus still finds, as he did in 2009, the role an interesting, if sometimes - make that constantly - frustrating, challenge.