e-Research and New Zealand

Nick Jones

BeSTGRID and

Center for e-Research, University of Auckland

Topics

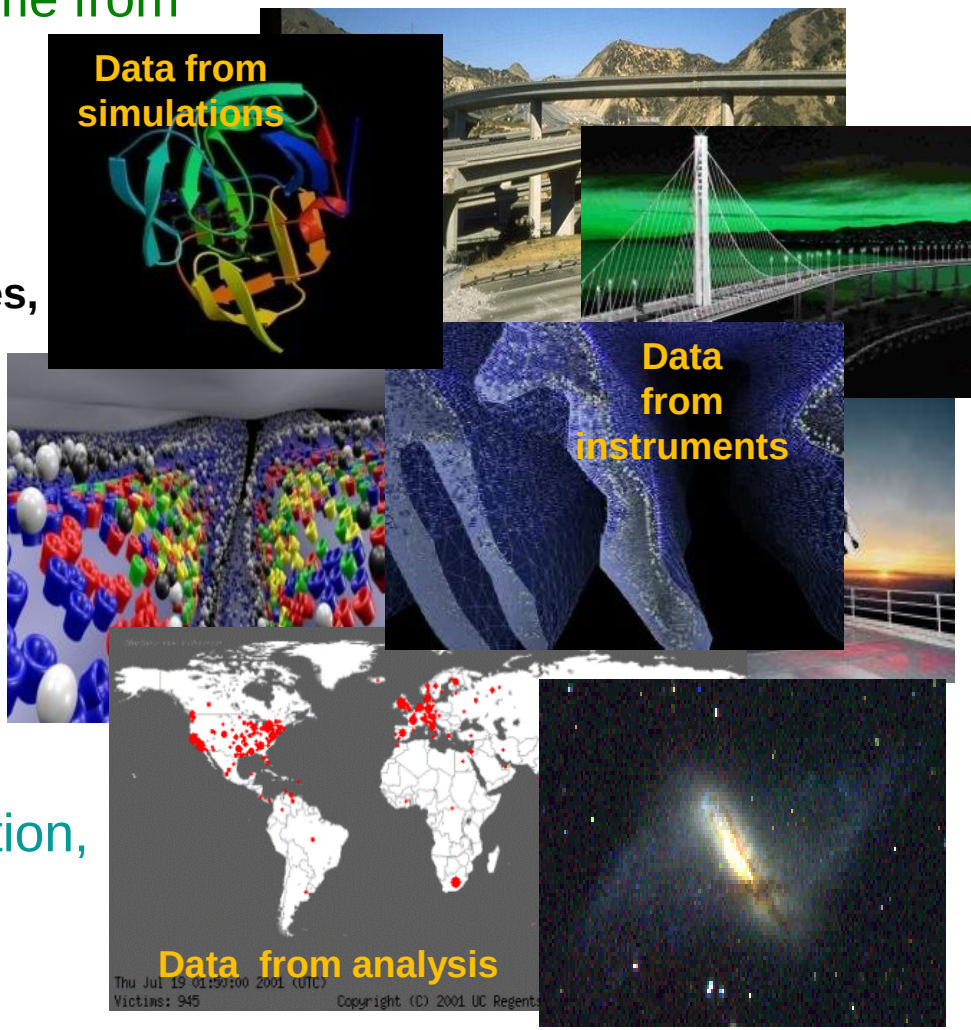
1. The nature of modern science
2. e-Research
3. e-Research in the New Zealand context
4. Summary

A knowledge economy

A knowledge economy requires a new and innovative infrastructure for data management, data exploration, information analysis, visualization, and knowledge sharing.

Coping with (anticipated) data torrents

- Over the next decade, data will come from everywhere
  - **Scientific instruments**
  - **Experiments**
  - **Sensors and sensor-nets**
  - **New devices (personal digital devices, computer-enabled clothing, cars, …)**
- And be used by everyone
  - **Scientists**
  - **Consumers**
  - **Educators**
  - **General public**
- Systems will need to support unprecedented diversity, globalization, integration, scale, and use

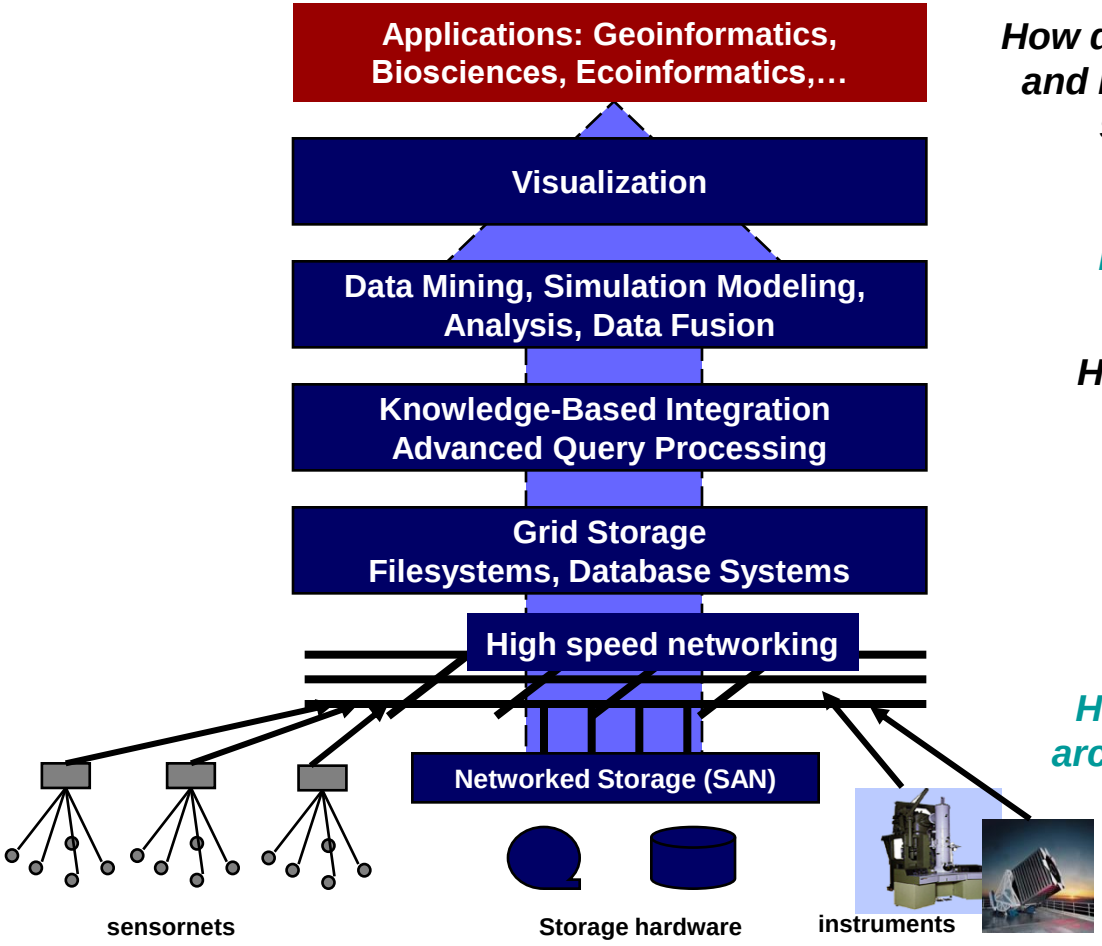


The data explosion

(from Wired 'Big Data, July 2008)

Terabytes	What it stores
1	2,600 songs Large hard disk ($200)
2	Daily data produced by a gene sequencer
20	Photos uploaded to FaceBook every month
120	All the data collected by the Hubble telescope
330	Weekly data produced by the Large Hadron Collider (est.)
440	All the international climate / weather data compiled by the National Climatic Data Center in the USA
530	All the videos in YouTube
1000 (1 petabyte)	Data processed by Google servers every 72 minutes

From Data to Information to Knowledge



How do we combine data, knowledge and information management with simulation and modeling?

How do we represent data, information and knowledge to the user?

How do we detect trends and relationships in data?

How do we obtain usable information from data?

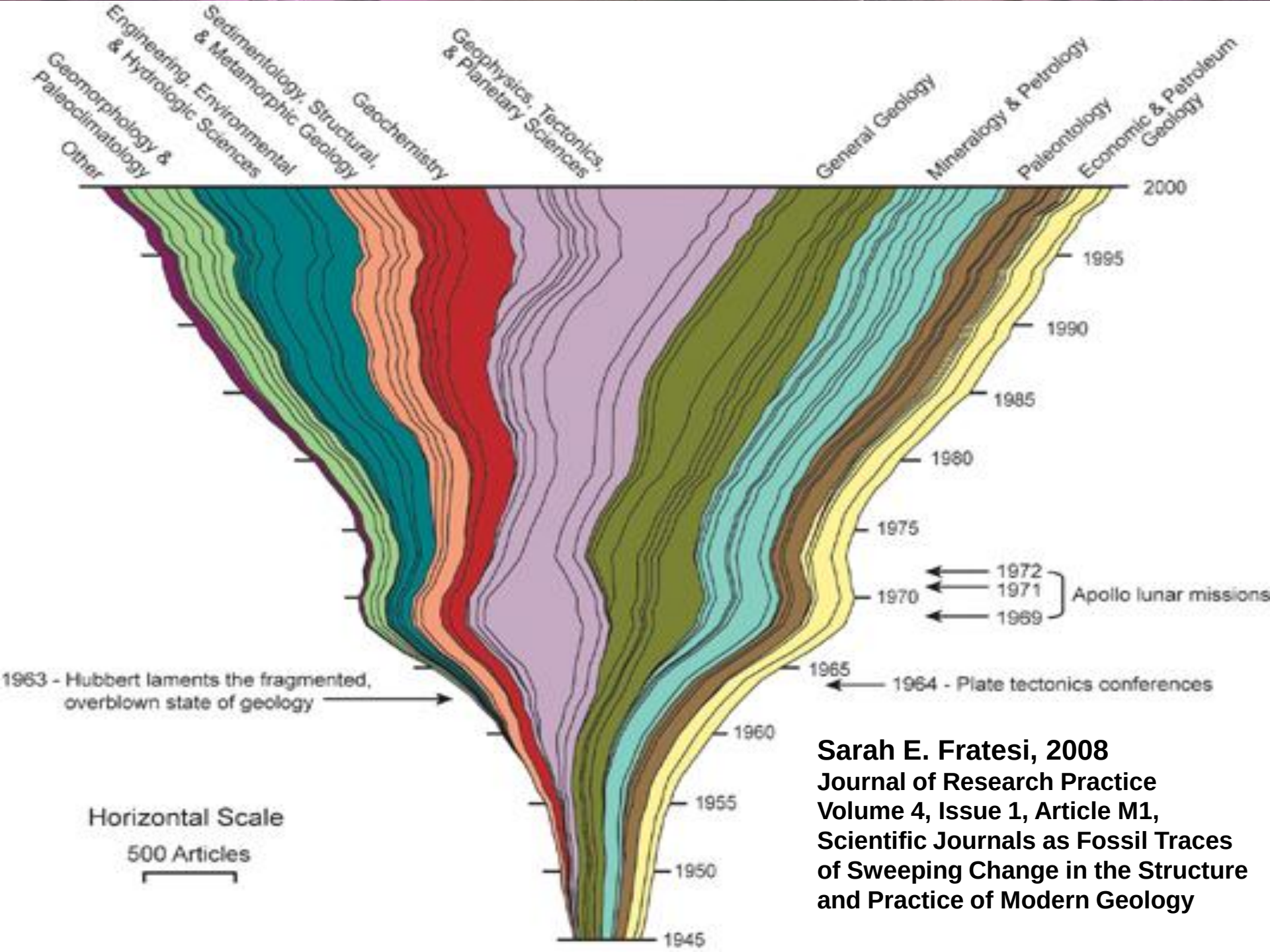
How do we collect, access and organize data?

How do we configure computer architectures to optimally support data-oriented computing?

The knowledge explosion

Amount of published research is expanding at a geometric rate

- has doubled in the last 24 months.
- Difficult to search, to find what we need
- Or to see where the gaps are
- Same is true of datasets, methods



Sarah E. Fratesi, 2008
Journal of Research Practice
Volume 4, Issue 1, Article M1,
Scientific Journals as Fossil Traces of Sweeping Change in the Structure and Practice of Modern Geology

Vannevar Bush, *As We May Think* (1945)

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers - conclusions which he cannot find time to grasp, much less to remember, as they appear.

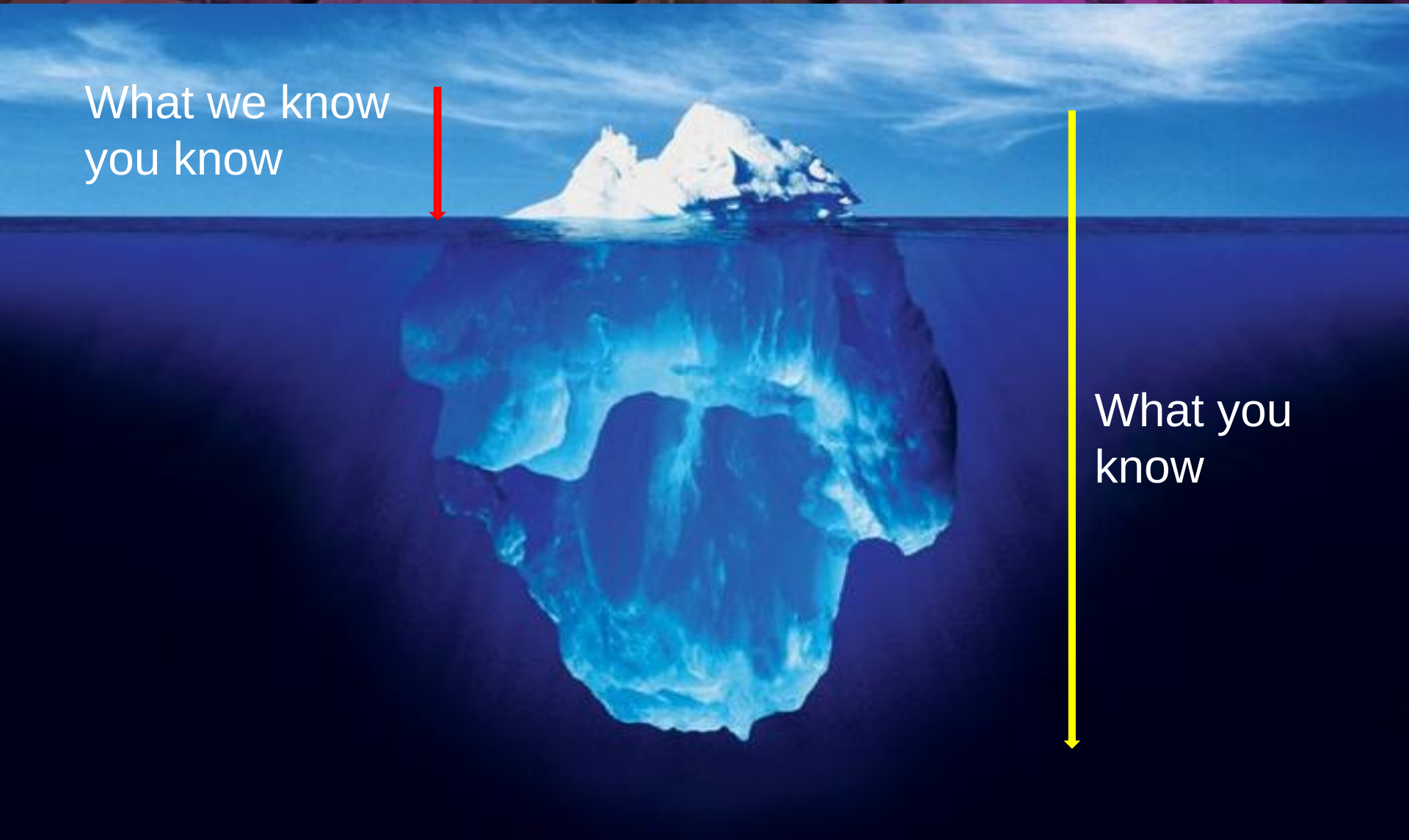
Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. If the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time might well be startling.

…A record, if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted. (Bush, 1945)

"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"

The Rock, T. S. Eliot

How much of our knowledge is actually published?



Research is Global

- Research collaborations are essential to modern research
- MoRST figures show that no New Zealand discipline produces more than 2% of citations globally
- or, at least 98% of publications are made outside of New Zealand

Problems with Research

- Accessibility to expensive, remote instrumentation & computation
- Repeatability and associated efficiencies of sharing data and: results, models, code, (outcomes)
- Complexity, especially across scales and between disciplines
- Collaboration: the notion of a virtual co-laboratory (leading to negotiation, synthesis)

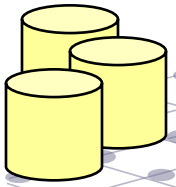
eResearch

What changed?

- Powerful simulation computers
- Metadata and data standards
- Grid Middleware support services
- Scientific workflow tools
- Fast and effective indexing technology
- Computational semantics & graph theory, description logics

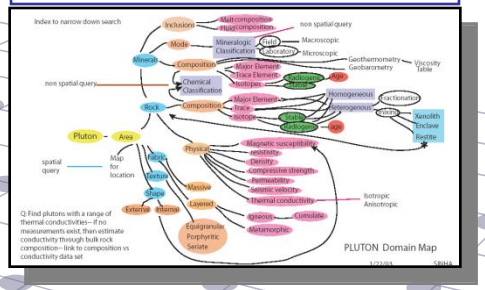
Societal context
Science drivers
Global issues

e-Research

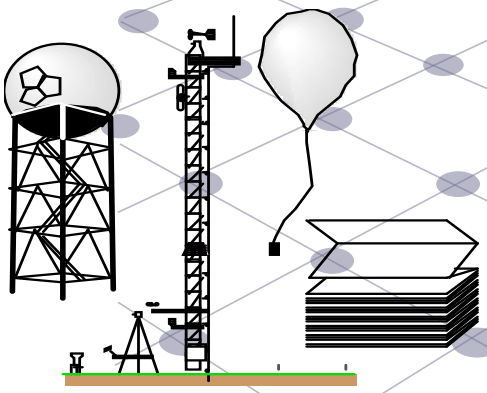


Connectivity resources

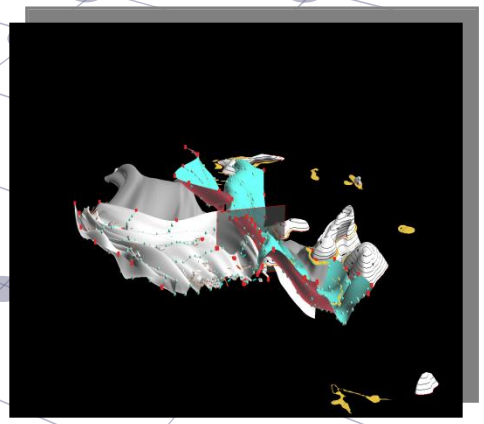
Theories, concepts
Knowledge representation



Information: real-time, archives, analyses
Informatics resources



People
Collaboration, visualization, education resources
Awareness / Outreach
Education
Support / Enabling



Data: Observations, measurements, experiments
Instrumentation

Models, simulations
Supercomputing

eResearch in the New Zealand Context

KAREN: *Kiwi Advanced Research and Education Network*

- Went live Dec 2006
- 10Gb/s NZ Backbone (“Squished” Ring)
- NZ$40million, Government Funding
- NZ$5million Capability Build Programme
- Linking all 8 New Zealand Universities and all 9 Crown Research Institutes, and National Library
- Additionally: ~622Mb/s link to US (and onto Europe)
- ~133Mb/s link to Australia

BeSTGRID www.bestgrid.org

... was a Pilot

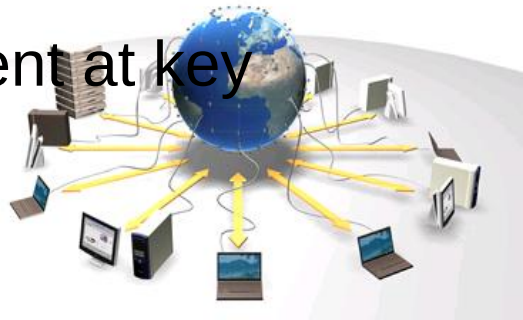
The University of Auckland New Zealand — Te Whare Wānanga o Tāmaki Makaurau
University of Canterbury — Te Whare Wānanga o Waitaha Christchurch New Zealand
Massey University

$2.5million: Sep 2006 – March 2008

Tertiary Education Commission — Te Amorangi Mātauranga Matua

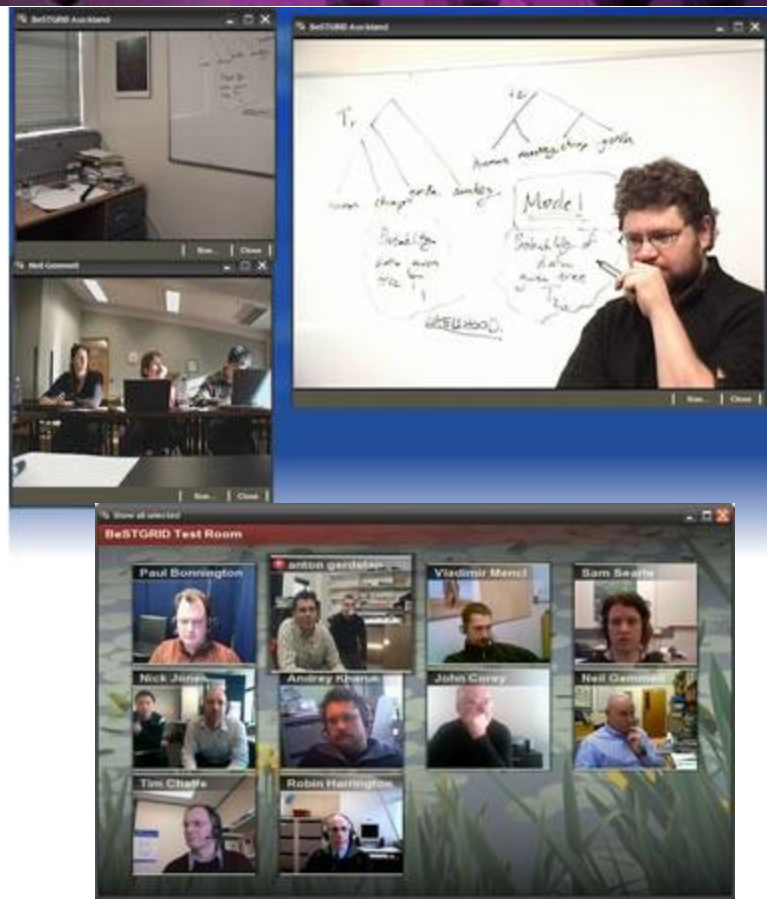
"**BeSTGRID** aims to enhance e-Research capability in New Zealand by providing the **skill base** to help the various research disciplines engage with new eResearch services"

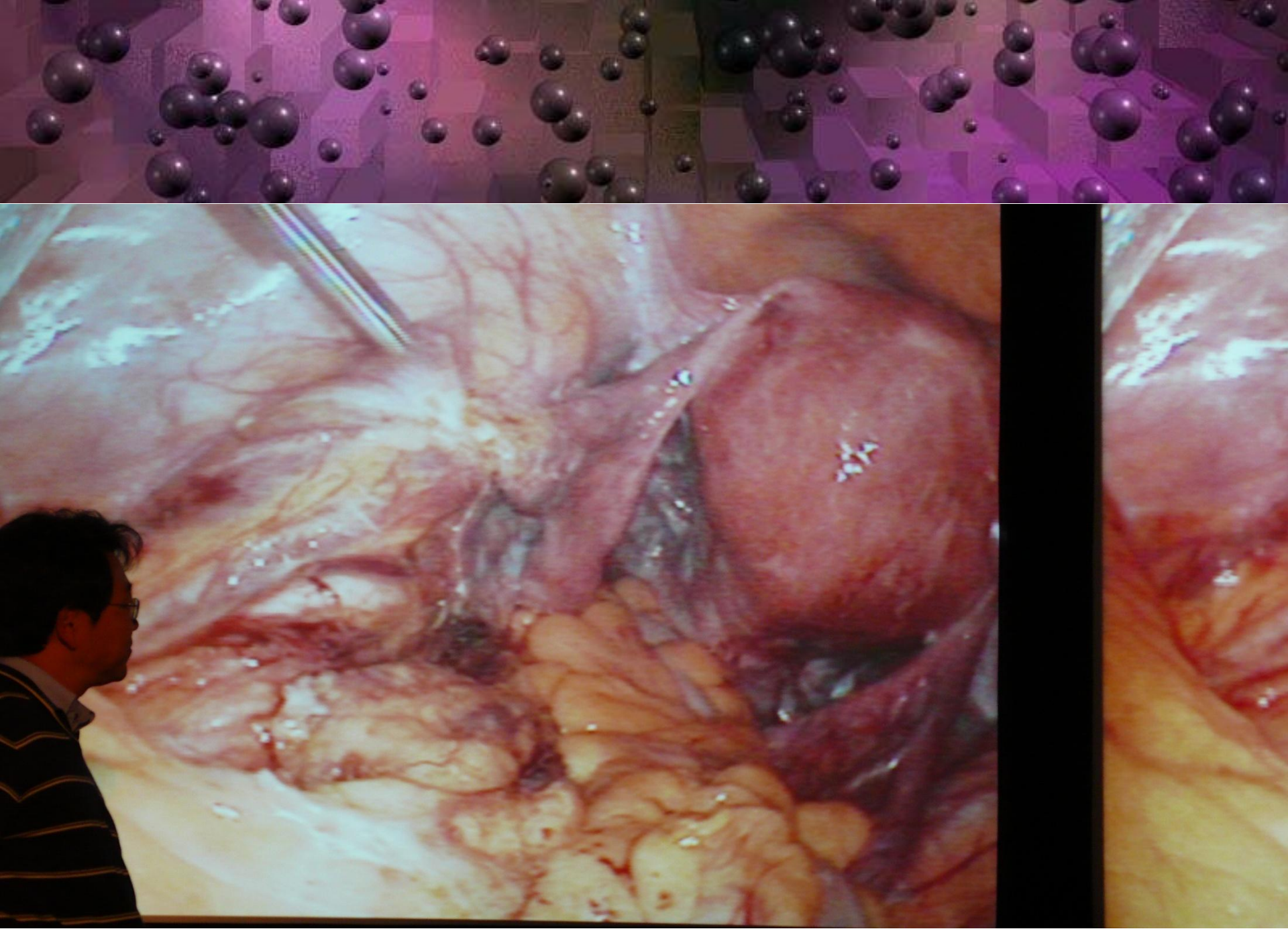
"**BeSTGRID** aims to kick start **centralised infrastructure** with some capital investment at key institutions"



Collaboration GRID

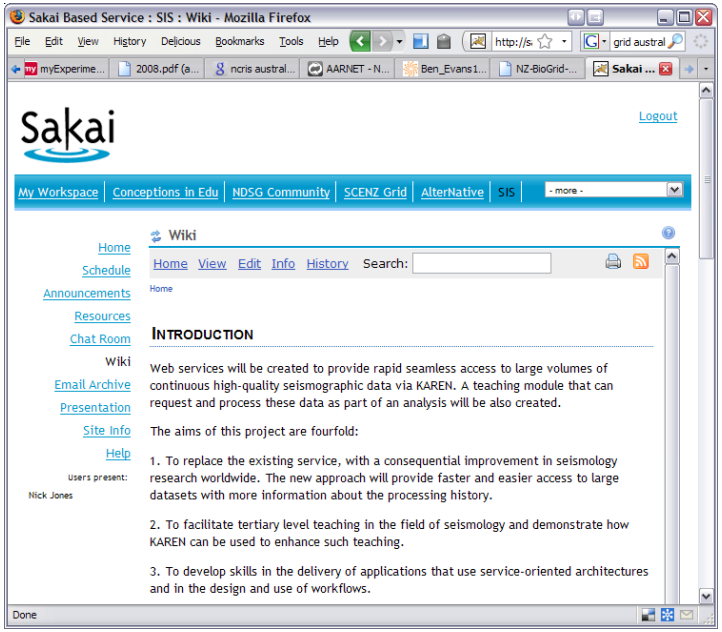
- Advanced Video Conferencing (Access GRID, EVO - Enabling Virtual Organisations, HD, H.323)





Collaboration GRID

- Sakai VRE (Virtual Research Environment) – collaboration space
  (wikis, mailing lists, event calendars, file sharing)



Computation and Data GRID

Bringing site specific HPC Resources to NZ Researchers

BeSTGRID streamlines access

Connect the resources and transparently present them to the community

To match users and their jobs to the best computational resources for their needs

Current
- The University of Auckland
- Massey University, Albany
- Canterbury University

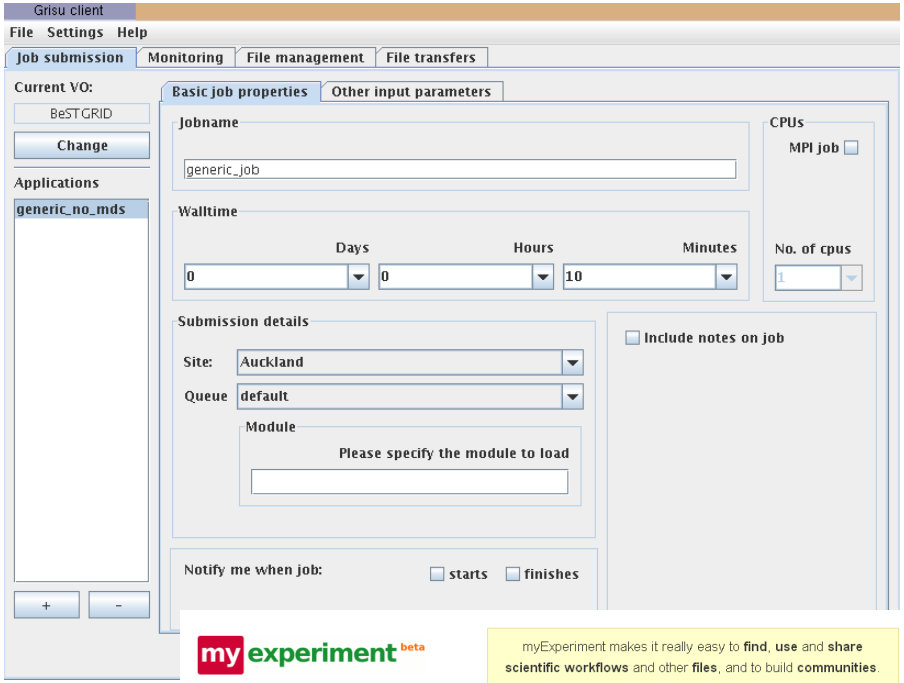
In development
- *Otago University*
- *Victoria University, Wellington*
- *Landcare Research*

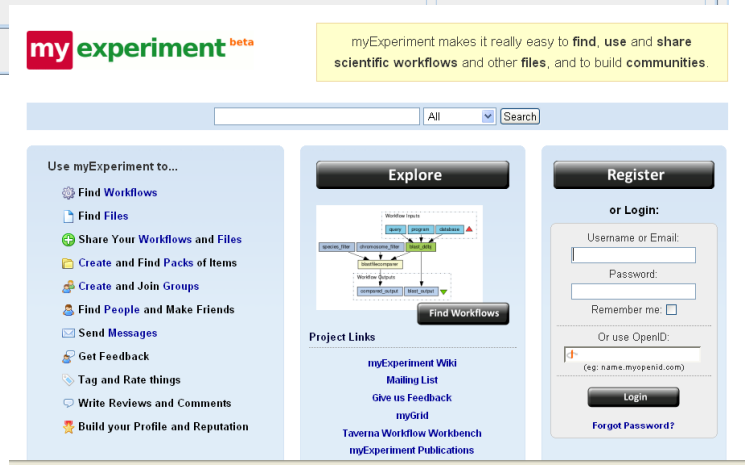
Computation Resources

Resource name	Location	Architecture	OS	#CPU cores	Freq	Notes
BeSTGRID Auckland Cluster	The University of Auckland	x86 Xeon	Linux	80	2.8GHz	Operational, since June 2008. Current Status
BeSTGRID Auckland Test Cluster	The University of Auckland	AMD	Linux	6	2.8GHz	for testing purposes only
BeSTGRID Prototype Cluster	University of Canterbury	x86 Xeon	Linux	4	3.0GHz	Prototype cluster, serving the community since June 2007.
BlueFern p575	University of Canterbury	Power5+	AIX/Linux	128	1.9GHz	8 nodes, 16 CPU cores each.
BlueFern BlueGene	University of Canterbury	PowerPC	Linux	4096	770MHz	Available only together with a local account.
Maggie	University of Otago	x86 Pentium 4	Linux	10	3.0GHz	No MPI available
Massey Cluster	Massey University, Albany	x86 Xeon	Linux	??	??	Gateway currently not accessible.

Grid Tools

- Grid Computing interfaces, such as GRISU and the Bioportal
- Workflow
  - Taverna + MyExperiment
- Data Services and Repositories
  - GridFTP file transfer tools





Successes: within NZ

- First grid infrastructure in NZ, focused on BioInformatics, with Industry collaboration with *BioMatters* and *New Zealand Supercomputing Centre*
- Implementation of first federated identity management framework in NZ
  - and contribution to a multi-sector working group to further evolve identity management across research and education (IMAGER, http://www.morst.govt.nz/current-work/science-infrastructure/imager/)
- Established a shared infrastructure service delivery model across institutions
- Recognition from MoRST of need for eResearch support, with $4.2M in ‘08/’09 budget for eResearch infrastructure

Issues: NZ Institutions

- Individual institutions struggle to fund dedicated staff to develop infrastructure and support research groups in changing their approach to use this infrastructure.
  - There is a medium term investment required with each research project to shift their approach without impacting negatively on their current programmes. To make significant and permanent changes of this nature requires long term sustained investment.

Issues: International collaborations

- Building international linkages to participate in the major grid programmes requires sophisticated capabilities at a larger scale than NZ groups are currently operating at

  (http://www.ogf.org/UnderstandingGrids/grid_projects.php)

- Need for deeper engagement with AeRIC (*Australian eResearch Infrastructure Committee* and NCRIS) to be funded from within NZ

Issues: Investment level

- Govt investment over 3 years is $4.2M:
  - http://www.morst.govt.nz/funding/budget/#knowledge
- In comparison to 4 year funding in Australia of $540M (in NCRIS).
  - NCRIS have been investing at that level since 2002 (12 science-focussed programmes)
  - (*In the USA, currently 18 full-scale research programmes connected to cyberinfrastructure)*

Issues: Education

- Who is training NZ's next generation of students to be e-aware?
- Will our students be able to conduct research & business within these emerging infrastructures?
- Will they be leaders or followers?

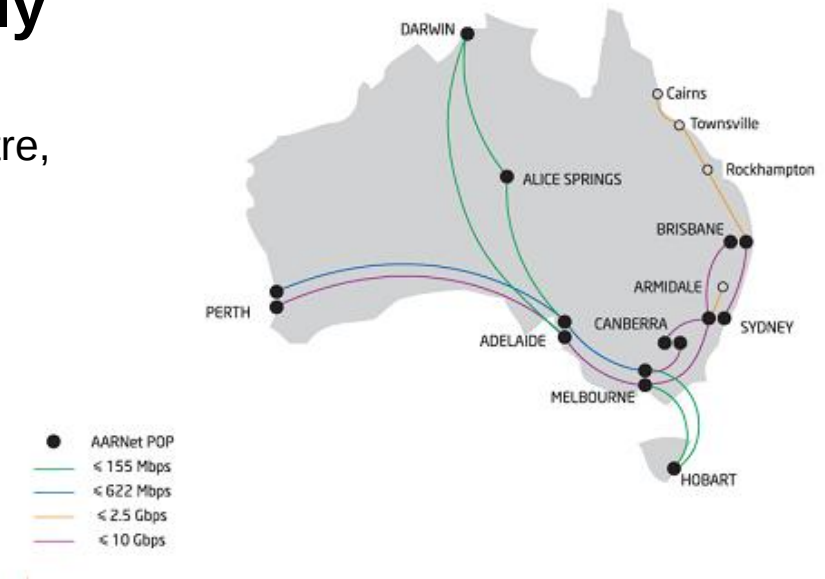
Strategic alignment with Australia

Grid Australia

AARNet National Network 2006

- BeSTGRID **aligns technically** with Grid Australia (Identity and Access Management, Grid Operations Centre, Grid Gateways, Policy Development)

- **Facilitates interoperability** and possible partnerships

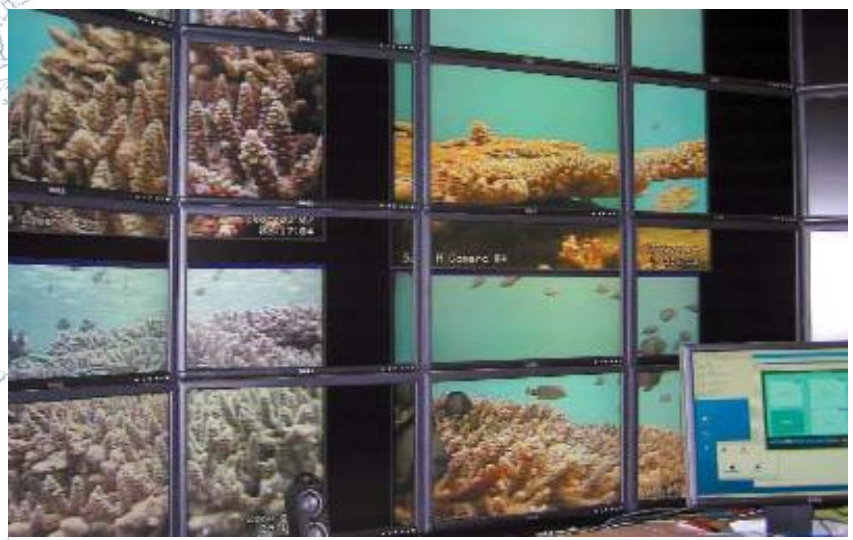


Successes: Trans-Tasman collaborations

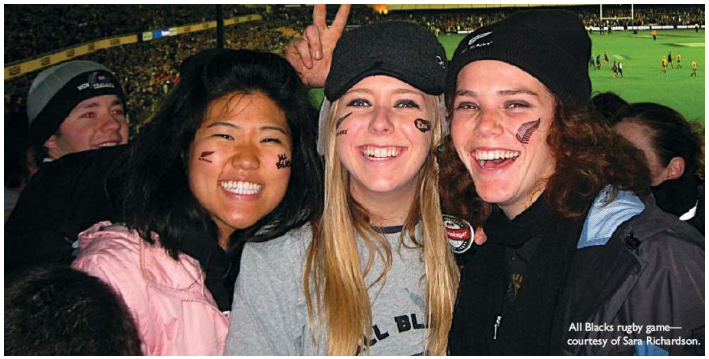
- Formative relationships in several technical programme areas with much more mature and larger scale community in Australia
- This conversation continued at APAN 26 and at eResearch Australasia, with the aim of developing insights into the next major issues to be resolved for the sector to seek Trans-Tasman collaborations
- Invitation to join the NCRIS, with support to be provided from NZ as Australia are unable to fund our engagement.
- Most opportunities still to be assessed (http://www.ncris.dest.gov.au/capabilities/).

PRAGMA


JLU
CNIC
GUCAS
LZU
KISTI
AIST
U. of Tsukuba
Titech
Osaka U.
HPC/HUT
AGCC
NCHC
CUHK
ASTI
NECTEC
ThaiGrid
HCMUT
HOIT-VN
MIMOS
IHPC
NGO
U. Sans Malaysia
NTU
IHPC/NGO
SKU
UI
QUT
Monash U.
BeSTGRID
NCSA
BU
U. UTAH
SDSC
CICESE
UNAM
UPRM
CeNAT-ITCR
U. Chile



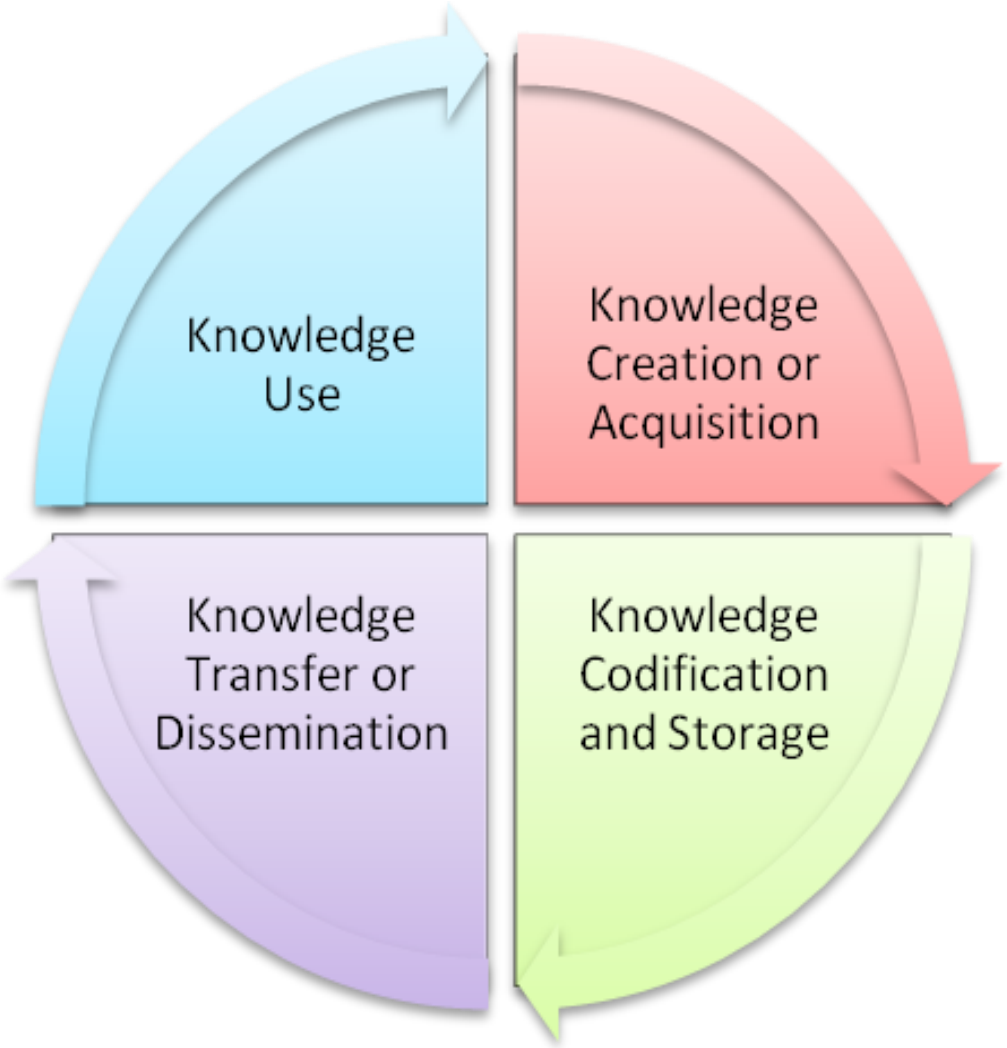

The New Science of Metagenomics
Calit2


All Blacks rugby game—
courtesy of Sara Richardson.

Summary

What do we have, what do we need and how to proceed

Components of knowledge computing (we need all 4)



What do we have?

- A good **high-speed network** with capacity for growth
- The **foundations of a sophisticated infrastructure**, some HPC and massive storage silos,
- **Authentication**, some **grid middleware** and **basic services** (but not nationally coordinated)
- Good **links to e-research development communities**: e.g. Australian National Collaborative Research Infrastructure Committee (NCRIS), San Diego Supercomputer Center (SDSC), eScience UK (Edinburgh, Manchester)

Science argument for e-research

- Help address previously intractable questions (plant pathology, epidemiology, plate tectonics, astrophysics, bioinformatics)
  - there is a need/obligation for NZ to share its science internationally
- Keep a better record of our data, our experiments, our knowledge, for future scientists.
  - Current science records are woefully inadequate for deep, long-term understanding.
- Enable more effective collaboration between experts, within NZ & internationally

Societal argument for e-Research

- Interoperable products and services between research agencies allow for anticipation, better planning, mitigation of pressing issues
  - such as extreme events, climate changes, invasive species, bio-security
- Better access by educators and general public to science products & government information (e-government)
- NZ researchers get good in emerging science practices & do not becotraining me isolated

Economic argument for e-Research

- Help keep NZ's top researchers in NZ (being there is everything)
- Promote efficiencies in the process of science and in research funding by providing access to expensive equipment
- Create a culture of sharing outcomes, data, tools early and often within the science community by providing a more coordinated infrastructure

What next? Aligning the NZ science community

We would like to hear from you if:

- You have compelling examples of the need for further investment in science infrastructure
- You have examples of successful research collaboration that were enabled by science infrastructure
- Please get in touch:

m.gahegan@auckland.ac.nz

n.jones@auckland.ac.nz

Thanks !!!

End

Questions, comments