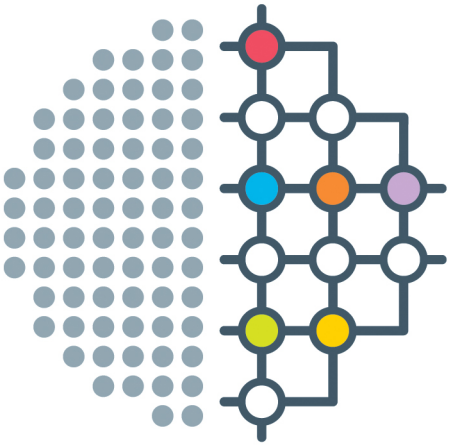


NeSI

New Zealand eScience Infrastructure

New Zealand Research Data Transfer Services

Kieron Mottley
Business Manager

GlobusWorld 2014
Chicago

Thanks to:

- Vladimir Mencl, NeSI eResearch Services and Systems Consultant
- Michael Keller, NeSI eResearch Services and Systems Consultant
- Markus Binsteiner, NeSI Software Developer
- Sat Mandri, Tuakiri Service Manager
- Andrew Farrell, NeSI Technical Programme Manager

Overview

1. New Zealand eScience Infrastructure and the national research landscape
2. New Zealand research is increasingly collaborative
3. Developing a national data transfer service

#1:

New Zealand eScience Infrastructure and the national research landscape


agresearch


SCION
forests · products · innovation


AUT
UNIVERSITY
AUCKLAND, NEW ZEALAND


VICTORIA
UNIVERSITY OF WELLINGTON


Plant & Food
RESEARCH
RANGAHAU AHUMĀRA KAI


ESR


UC
UNIVERSITY OF
CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND


MASSEY
UNIVERSITY


Landcare Research
Manaaki Whenua


GNS
SCIENCE
TE PŪ AO


Lincoln
University
Te Whare Wānaka o Aoraki
CHRISTCHURCH-NEW ZEALAND
New Zealand's specialist land-based university


UNIVERSITY
OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND


NIWA
Taihoro Nukurangi

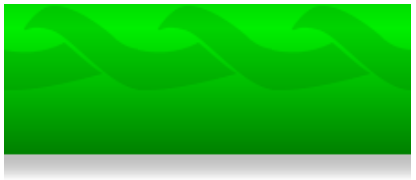

CallaghanInnovation

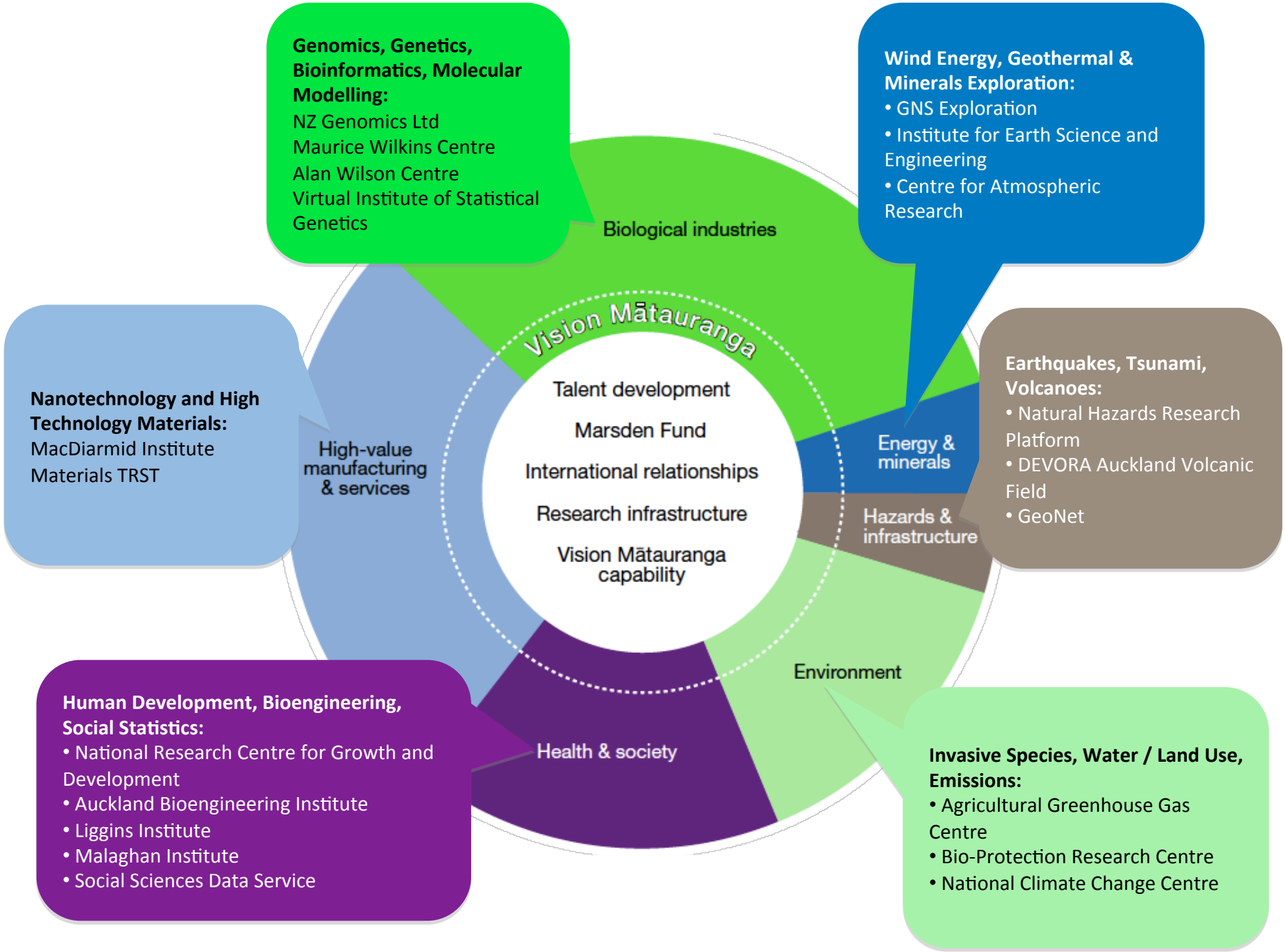

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato


THE UNIVERSITY
OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

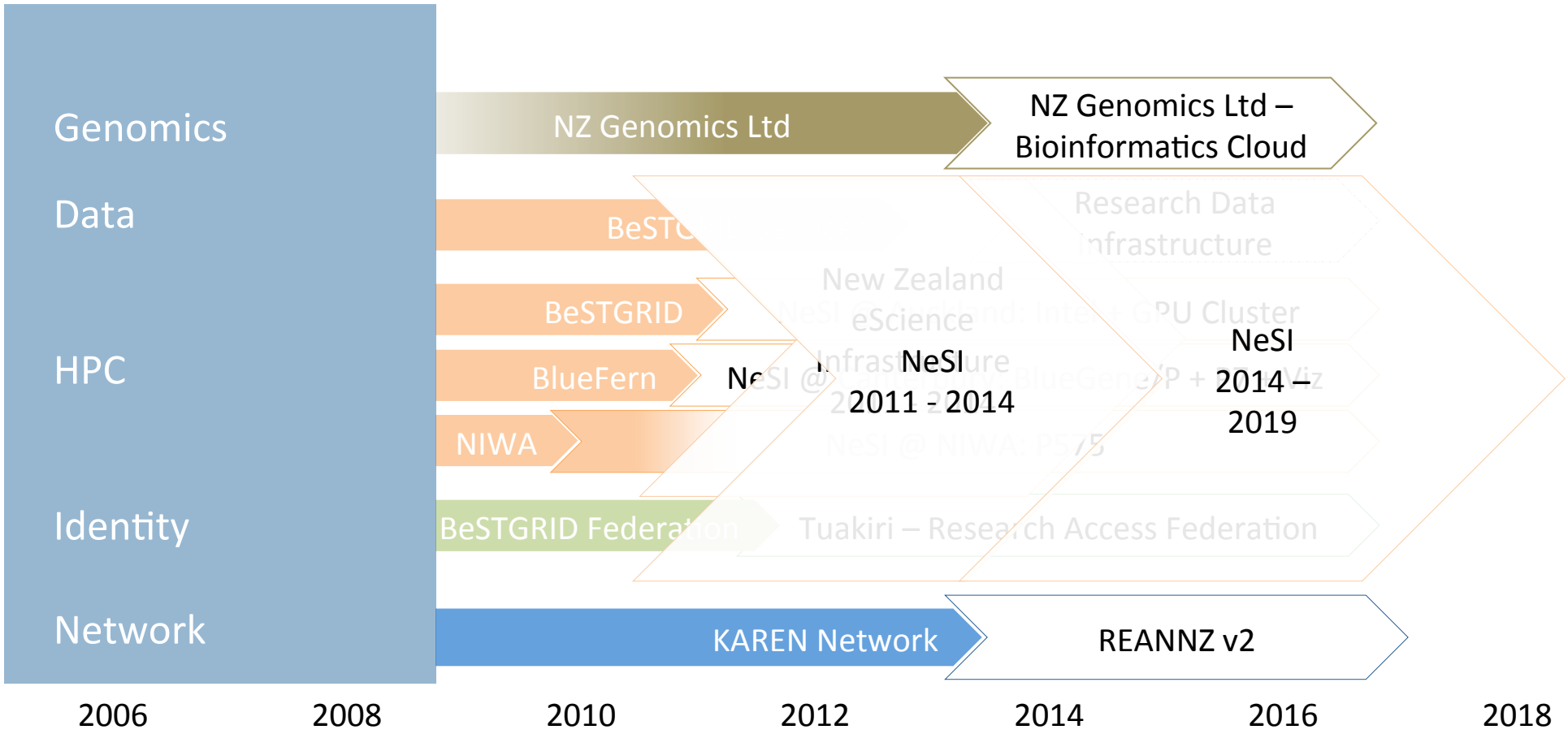
… with significant research data investments

Social Statistics
Humanities
Climate
Ecosystems
Marine
Geology
Environment
Physiology




Genomics, Genetics, Bioinformatics, Molecular Modelling:
NZ Genomics Ltd
Maurice Wilkins Centre
Alan Wilson Centre
Virtual Institute of Statistical Genetics
Wind Energy, Geothermal & Minerals Exploration:
• GNS Exploration
• Institute for Earth Science and Engineering
• Centre for Atmospheric Research
Nanotechnology and High Technology Materials:
MacDiarmid Institute
Materials TRST
Earthquakes, Tsunami, Volcanoes:
• Natural Hazards Research Platform
• DEVORA Auckland Volcanic Field
• GeoNet
Biological industries
Energy & minerals
Hazards & infrastructure
Environment
Health & society
High-value manufacturing & services
Vision Mātauranga
Talent development
Marsden Fund
International relationships
Research infrastructure
Vision Mātauranga capability
Human Development, Bioengineering, Social Statistics:
• National Research Centre for Growth and Development
• Auckland Bioengineering Institute
• Liggins Institute
• Malaghan Institute
• Social Sciences Data Service
Invasive Species, Water / Land Use, Emissions:
• Agricultural Greenhouse Gas Centre
• Bio-Protection Research Centre
• National Climate Change Centre

NZ Research e-Infrastructure Roadmap



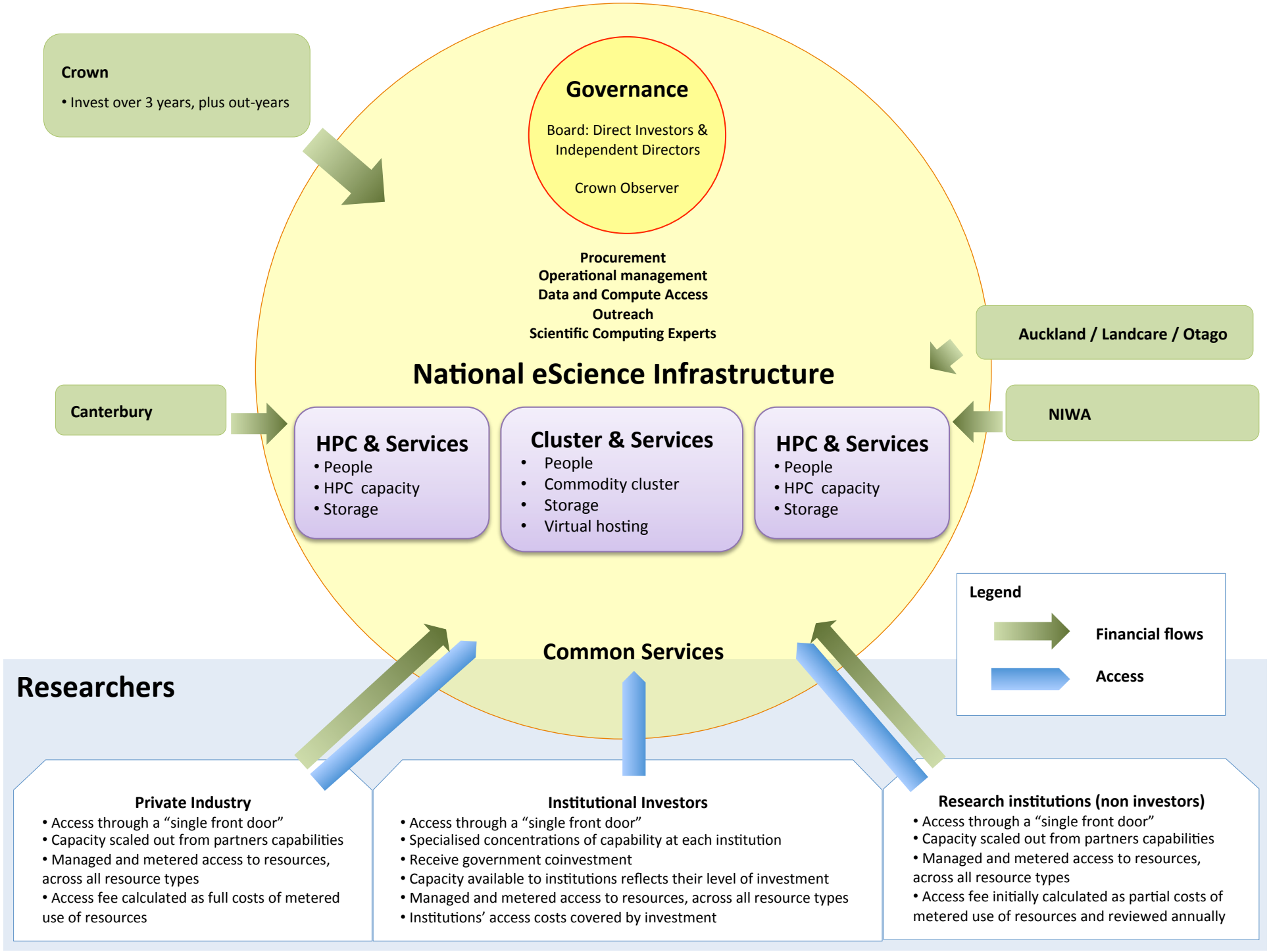
Building a national infrastructure

- Collaborative model
- Capabilities embedded within institutions

Teams are typically **distributed** and **multi institutional**, primarily as they're built from existing capability within historical research computing centres and groups

Operating in this environment creates two early barriers:

1. Culture & identity need to integrate and mature
2. Support needs to scale up from local to national

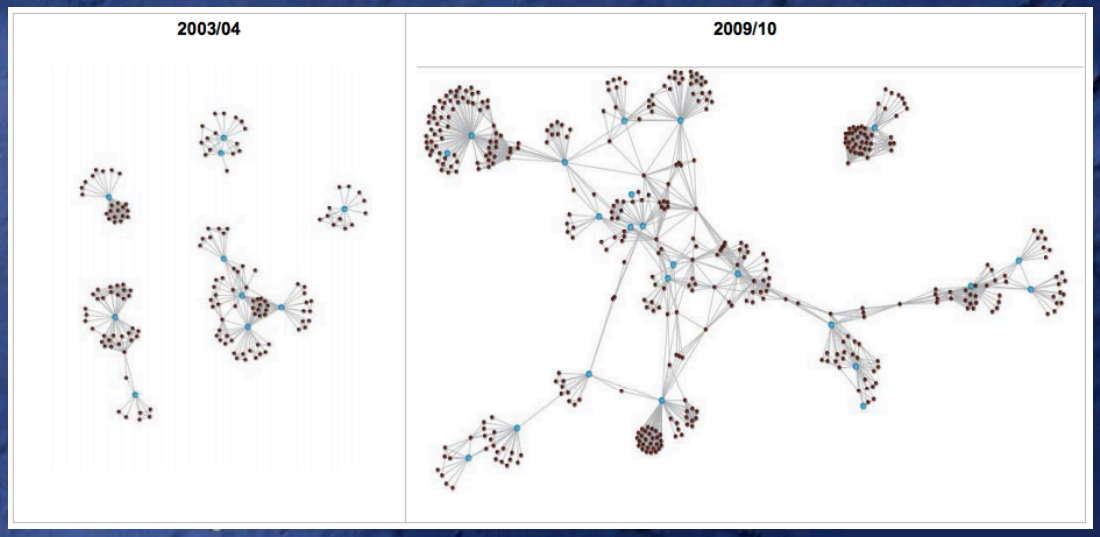

Crown
• Invest over 3 years, plus out-years
Governance
Board: Direct Investors & Independent Directors
Crown Observer
Procurement
Operational management
Data and Compute Access
Outreach
Scientific Computing Experts
National eScience Infrastructure
Auckland / Landcare / Otago
Canterbury
NIWA
HPC & Services
• People
• HPC capacity
• Storage
Cluster & Services
• People
• Commodity cluster
• Storage
• Virtual hosting
HPC & Services
• People
• HPC capacity
• Storage
Common Services
Legend
Financial flows
Access
Researchers
Private Industry
• Access through a "single front door"
• Capacity scaled out from partners capabilities
• Managed and metered access to resources, across all resource types
• Access fee calculated as full costs of metered use of resources
Institutional Investors
• Access through a "single front door"
• Specialised concentrations of capability at each institution
• Receive government coinvestment
• Capacity available to institutions reflects their level of investment
• Managed and metered access to resources, across all resource types
• Institutions' access costs covered by investment
Research institutions (non investors)
• Access through a "single front door"
• Capacity scaled out from partners capabilities
• Managed and metered access to resources, across all resource types
• Access fee initially calculated as partial costs of metered use of resources and reviewed annually

#2:

New Zealand research is increasingly collaborative

Research collaborations are growing, becoming more highly connected

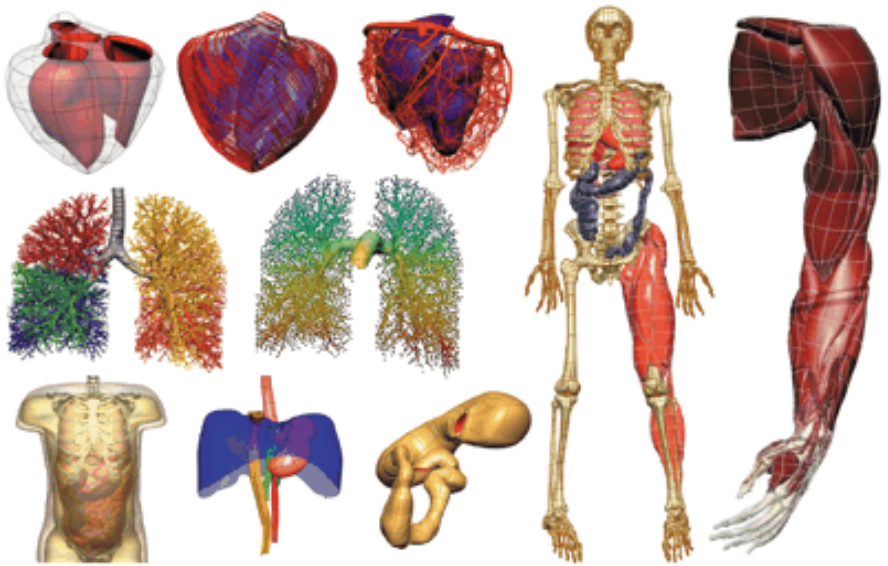
Research networks and collaboration at the Allan Wilson Centre



http://sciblogs.co.nz/a-measure-of-science/2011/04/14/the-new-zealand-innovation-ecosystem-map/

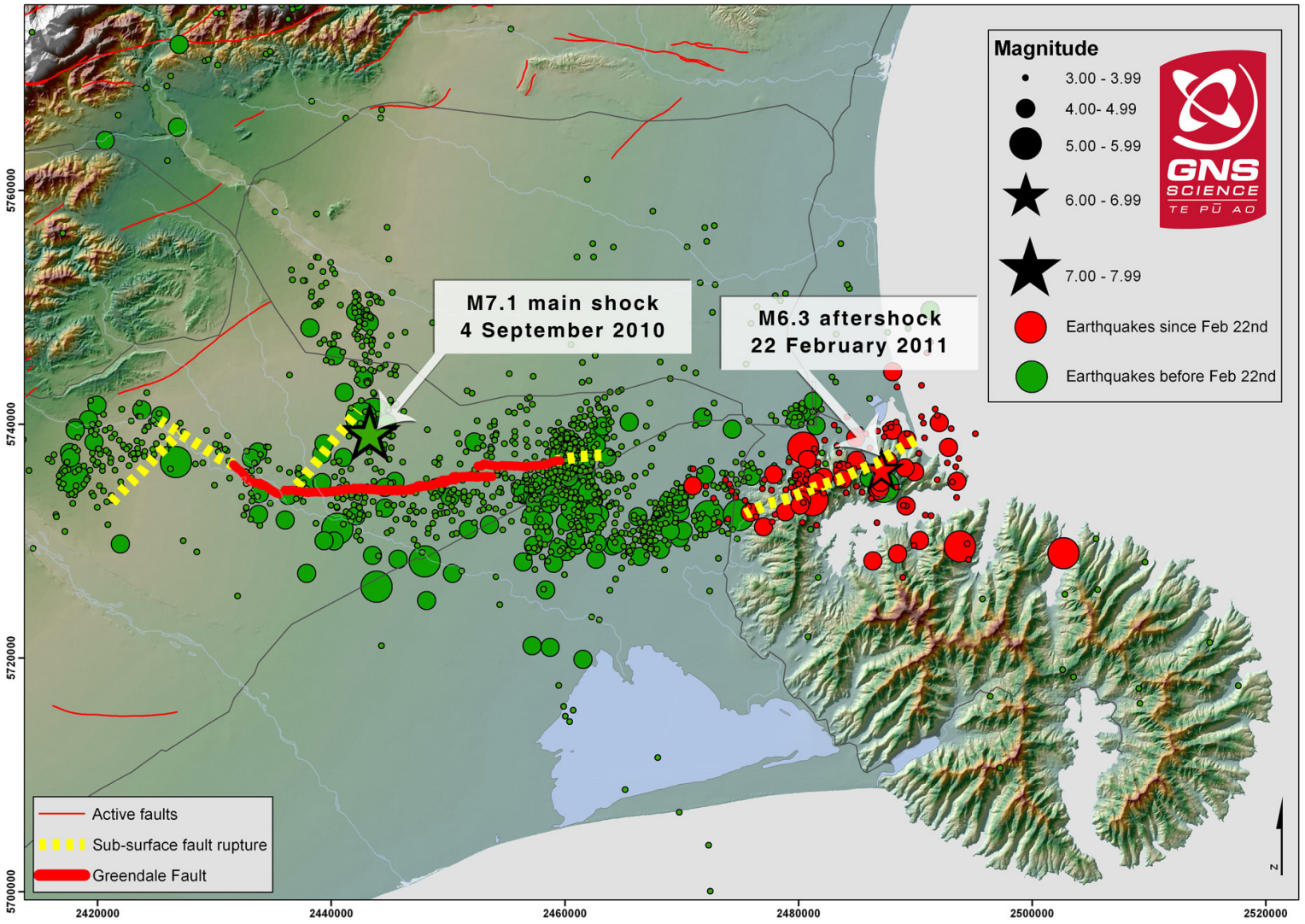
+

international collaborations



Finite-element organ models used for computational physiology in the IUPS Physiome Project. Hunter and Nielsen, Physiology, 316, October 1, 2005

Managing Big Data

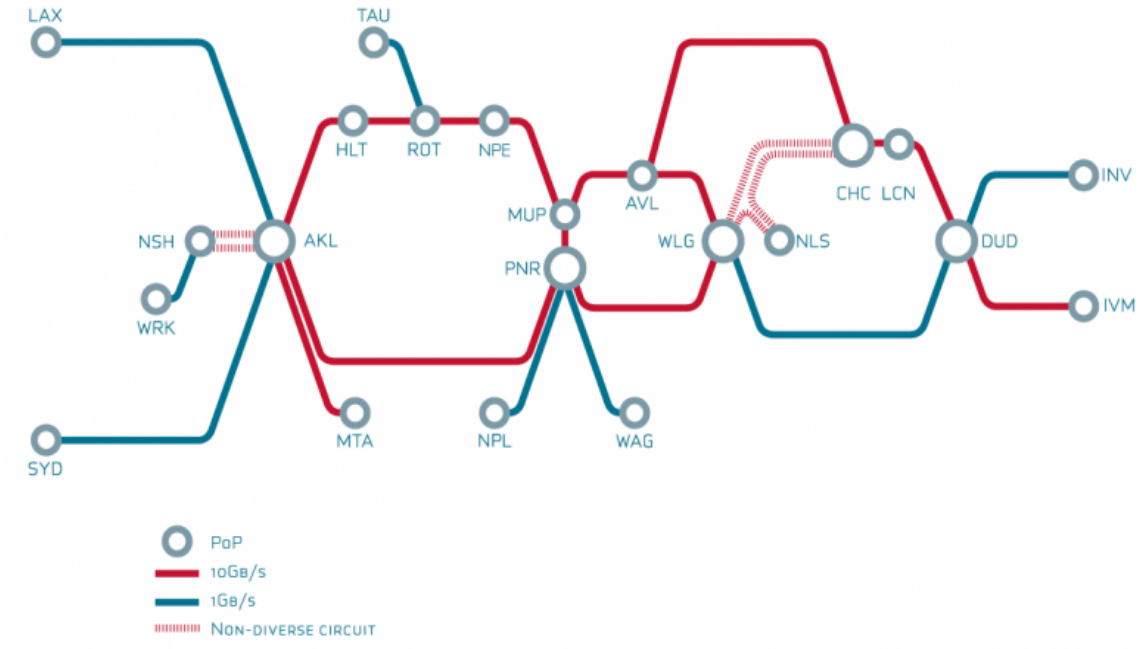


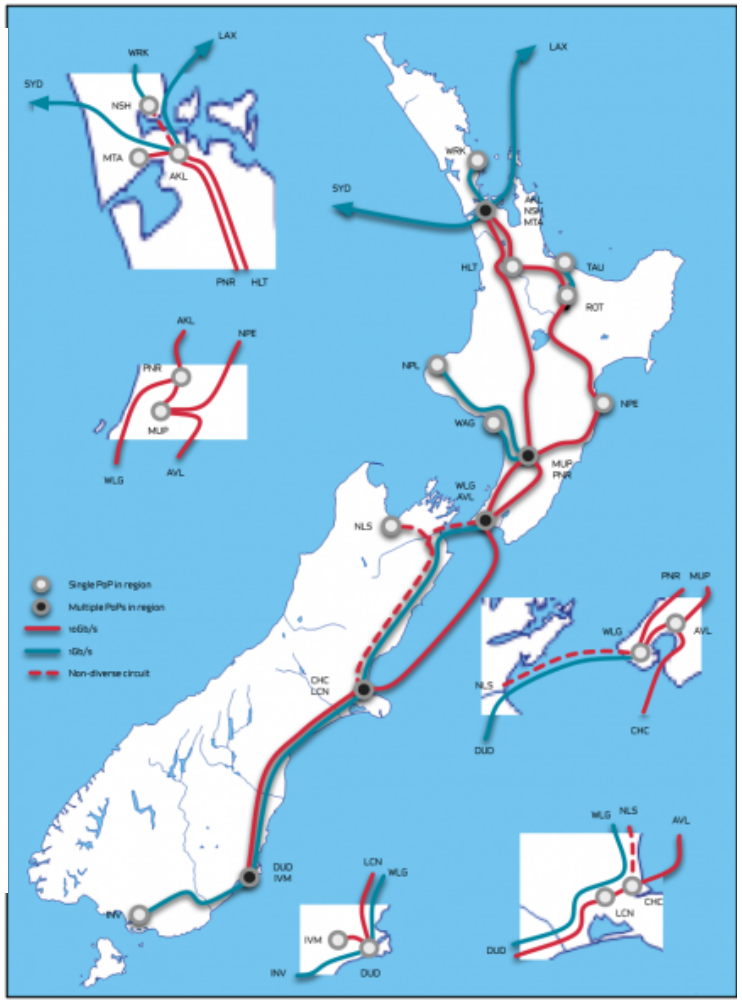
This map from GNS Science in New Zealand shows the earthquakes of 4 September 2010 and 22 February 2011 along with their aftershocks.

#3:

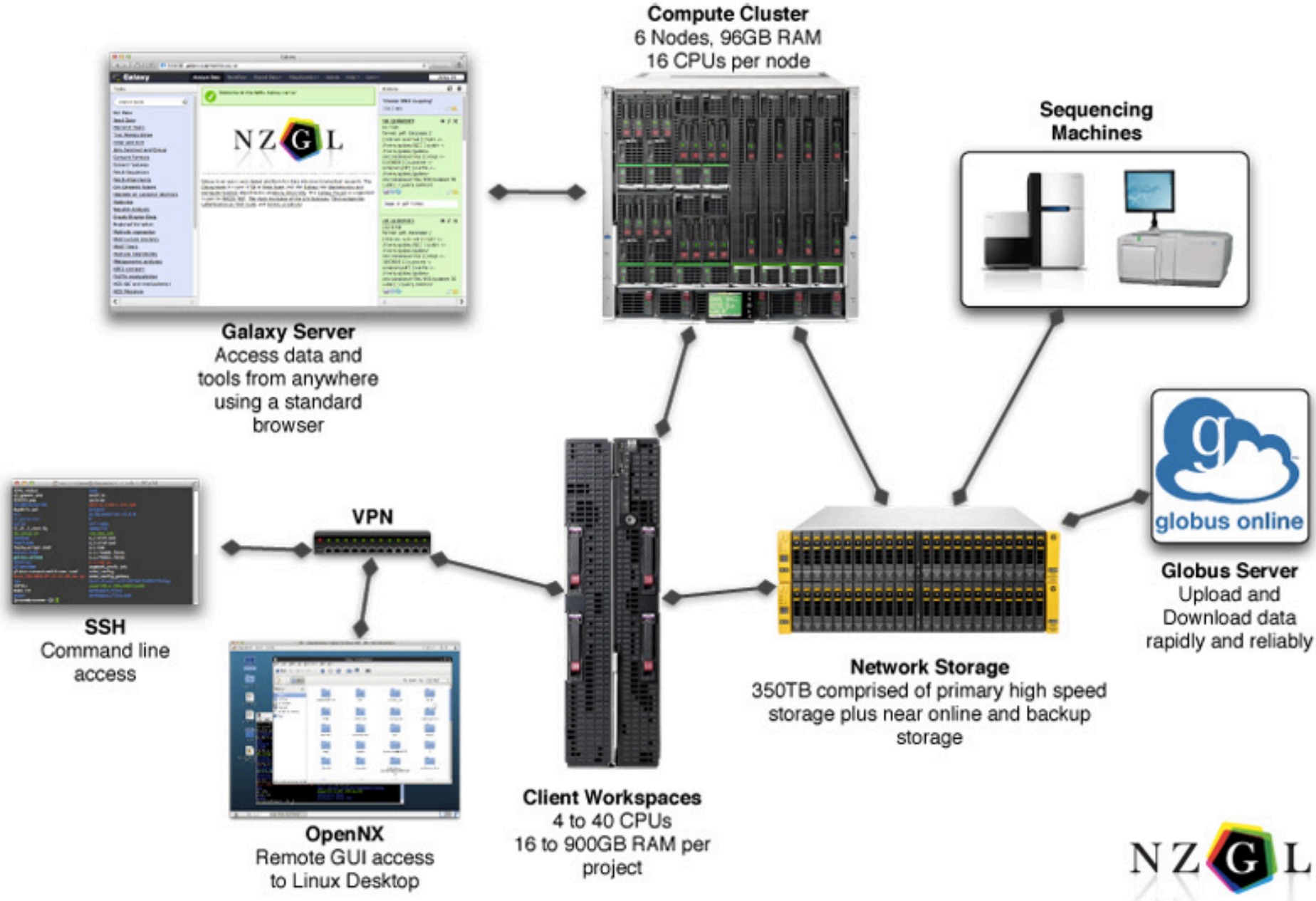
Developing a national data transfer service

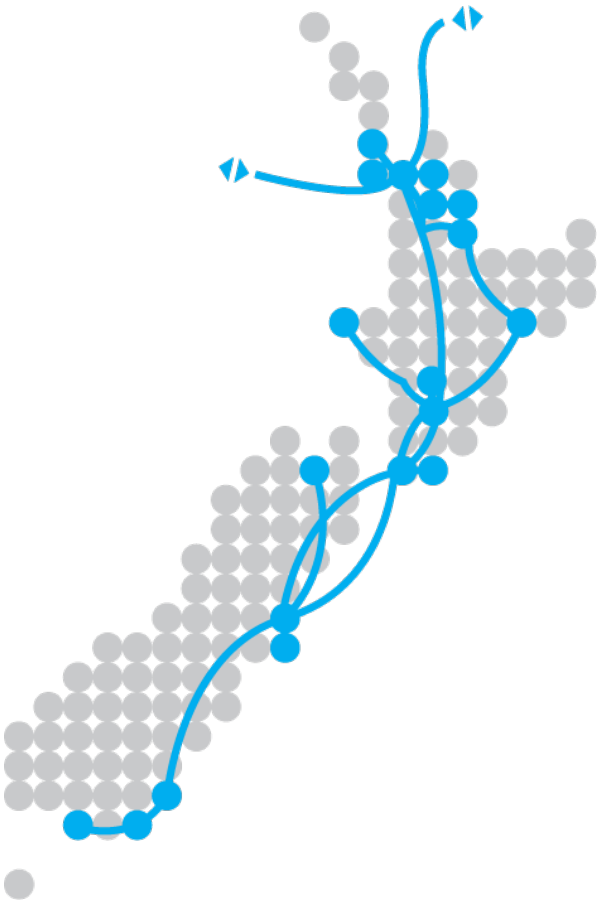
Transfer links





Goal: Facilitate high-performance transfer, making good use of available b/w, particularly on red, 10Gbps links


Compute Cluster
6 Nodes, 96GB RAM
16 CPUs per node
Sequencing Machines
Galaxy Server
Access data and tools from anywhere using a standard browser
globus online
Globus Server
Upload and Download data rapidly and reliably
VPN
SSH
Command line access
Network Storage
350TB comprised of primary high speed storage plus near online and backup storage
OpenNX
Remote GUI access to Linux Desktop
Client Workspaces
4 to 40 CPUs
16 to 900GB RAM per project



REANNZ

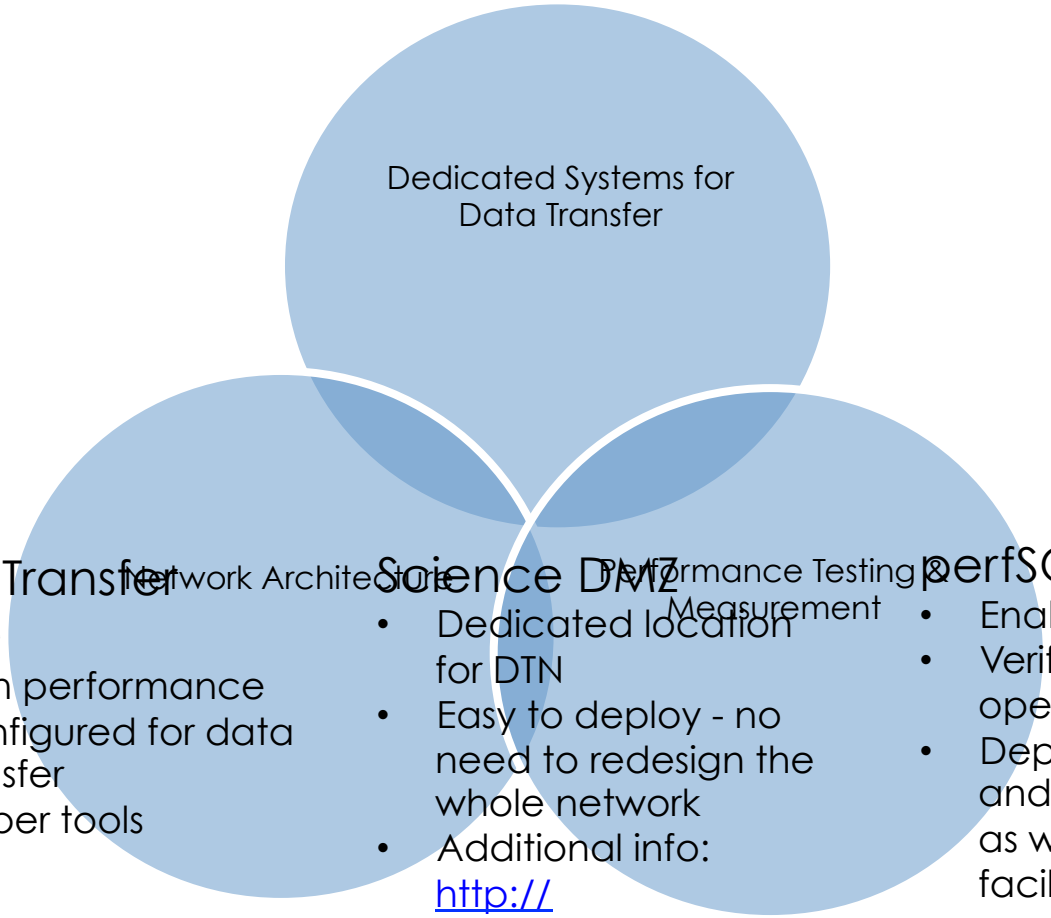
Current: 10 Gb/s national
1 Gb/s international

2014: 40 Gb/s national
40 Gb/s international
(shared AARNet science wave)

23 connection points nationwide

Low latency, low jitter

The “Science DMZ” Model



Data Transfer Node

- High performance
- Configured for data transfer
- Proper tools

Science DMZ

- Dedicated location for DTN
- Easy to deploy - no need to redesign the whole network
- Additional info: http://fasterdata.es.net/

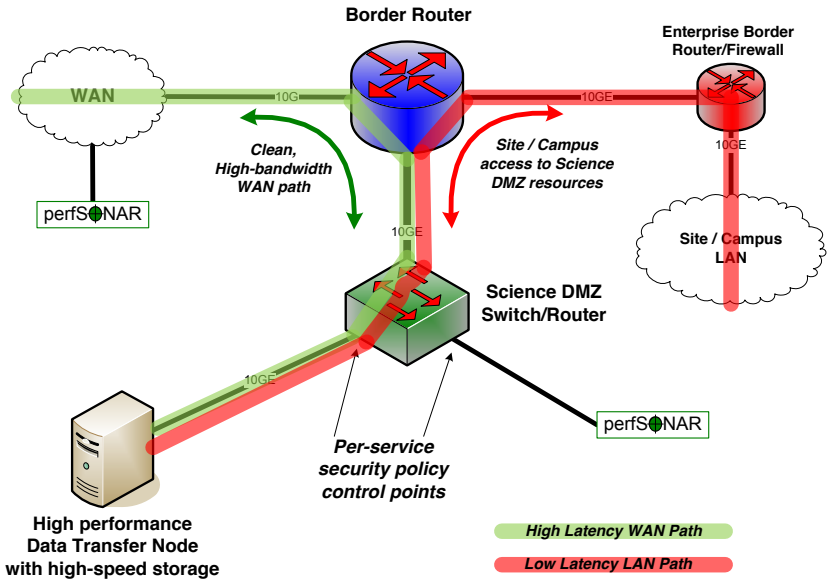
perfSONAR

- Enables fault isolation
- Verify correct operation
- Deployed in REANNZ and other networks, as well as sites and facilities

REANNZ

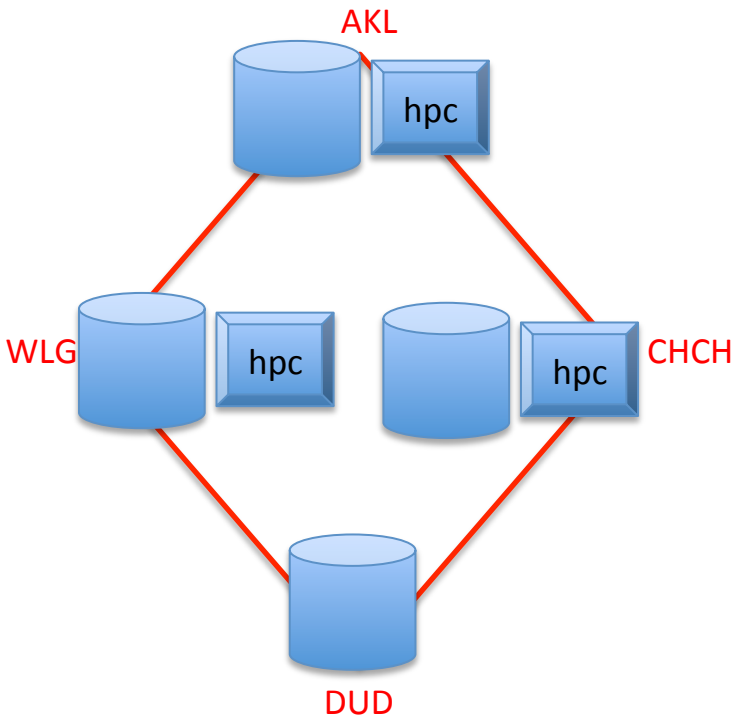
Science DMZ in NZ

- Widely deployed architectural concept connecting the science instrument directly to the high-performance network
  - Fully instrumented with perfSONAR
  - Data transfer node
  - High performance switch
- REANNZ is working with NeSI, NZ universities and Crown Research Institute customers to deploy Science DMZs



What do we want to do with data transfer?

- Transfer large datasets to advanced computing resources, e.g. from DUD to AKL
- Support collaborative sharing of data sets where access is often latency-sensitive
  - Local copies – of at least subsets of shared data
  - Replicate data between local stores using network backbone, at high-speeds where possible



Why enable delegation of Globus authentication to NZ IdPs?

- Make some fundamental improvements to UX, for NZ users
  - Enable login to Globus website using institutional credentials, rather than Globus-specific credentials.
  - Enable automatic activation of end-points using same credentials – single sign-on.

- Tuakiri is New Zealand Access Federation Service for the NZ Higher Education and Research Sector, established in 2011.

- Tuakiri is a formal federation of member institutions focused on creating a common framework for collaborative trust in support of research and education.

- Tuakiri makes sharing protected online resources easier, safer, and more scalable in our age of digital resources and services.

New Zealand eScience Infrastructure "NeSI", is the federation operator, providing Tuakiri with:

- Service hosting
- Service support
- Future development and innovation potential

New Zealand members

Tuakiri
NEW ZEALAND ACCESS FEDERATION

Production IdPs IdP deployment in pipeline

AUT UNIVERSITY

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

SCION
forests · products · innovation

GNS SCIENCE
TE PŪ AO

Lincoln University
Te Whare Wānaka o Aoraki

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Plant & Food RESEARCH
RANGAHAU AHUMĀRA KAI

ESR

NIWA
Taihoro Nukurangi

CallaghanInnovation

Massey University

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

KO TE TANGATA
THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha

Landcare Research
Manaaki Whenua

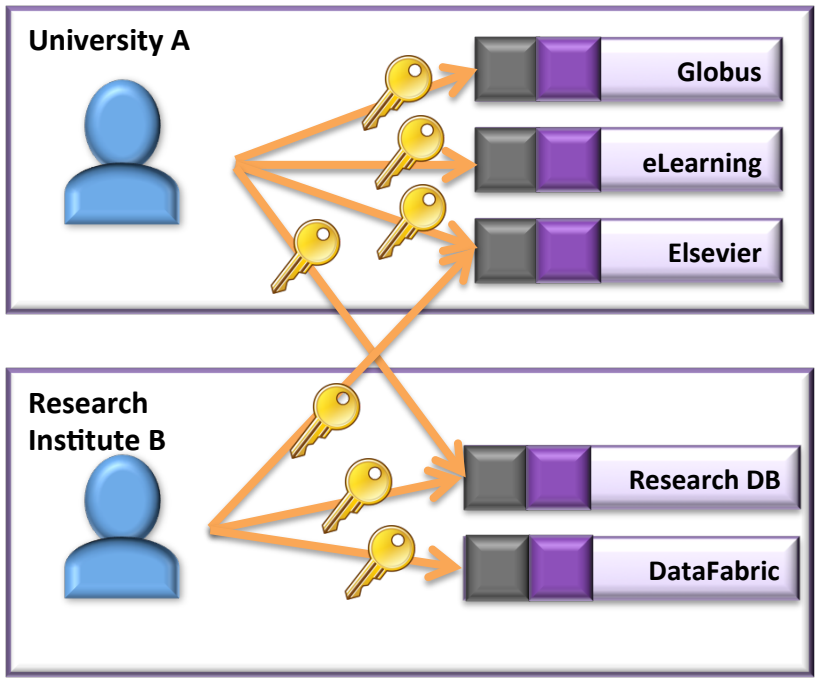
agresearch

CPIT

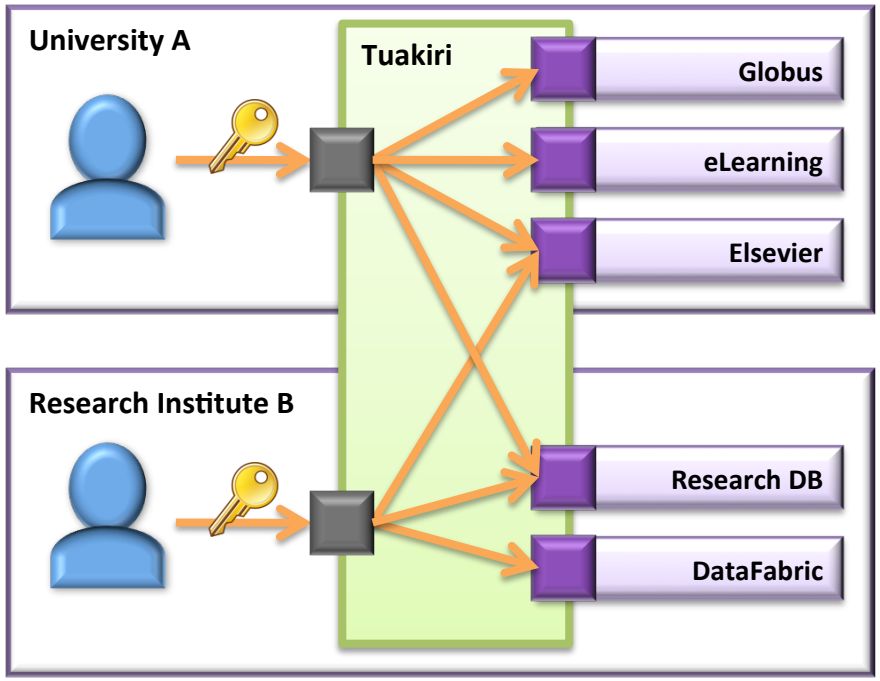
What it does

Tuakiri
NEW ZEALAND ACCESS FEDERATION

Without Tuakiri



With Tuakiri



User Administration & Authentication
Authorisation
Resource
Credentials

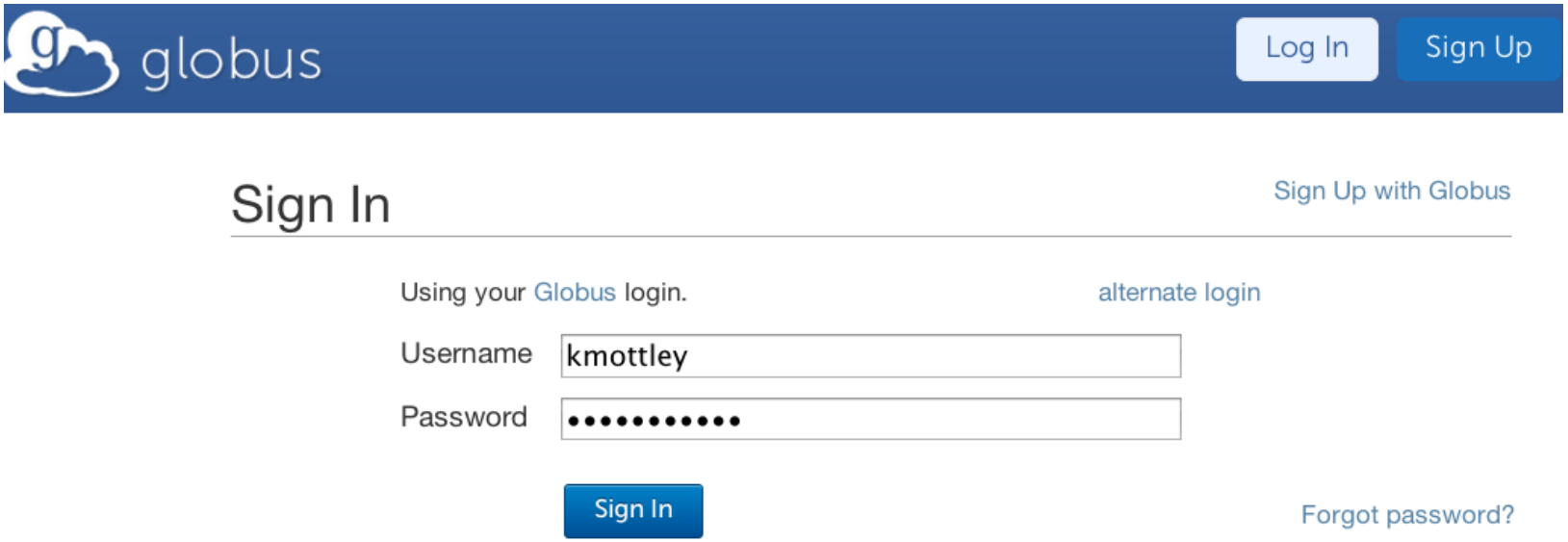
User Administration & Authentication
Authorisation
Resource
Credentials

Securing online resources with shibboleth

Tuakiri
NEW ZEALAND ACCESS FEDERATION

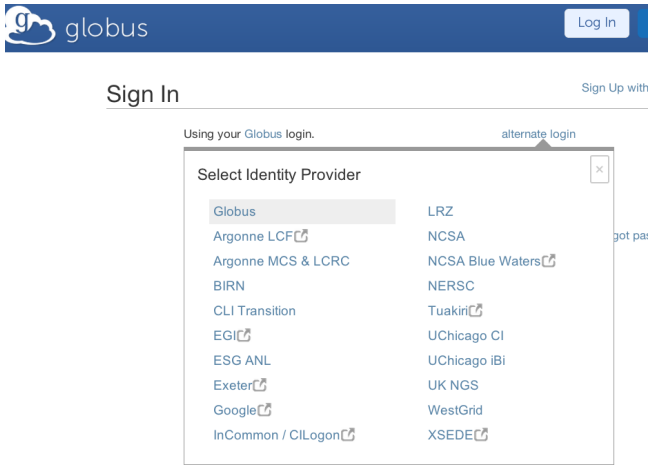
- Using Shibboleth you can secure an online resource like Globus by implementing authenticated and authorised access.
- Following Service Provider enablement, an online resource will request a Shibboleth Session **"single-sign-on federated access"** instead of the traditional local user account login.
- The online resource will consume relevant and required user information from the attributes supplied by the IdP and make authorisation decisions based on that information provided.

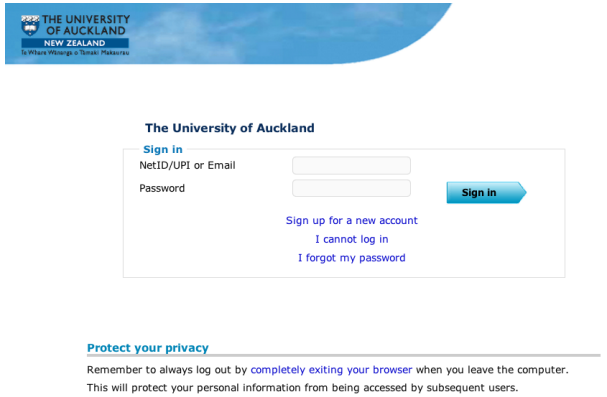
Without NZ IdP integration – Globus website login



I need to remember yet another set of credentials!

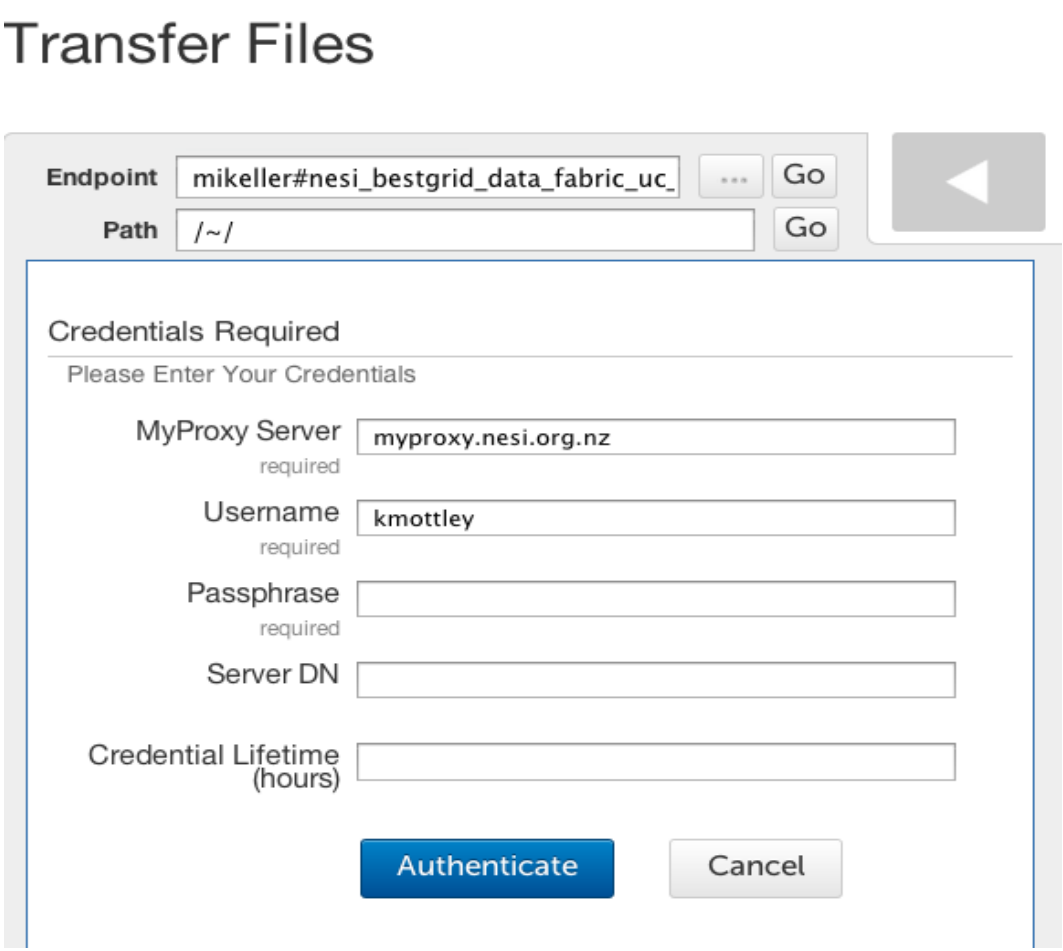
With integration – Globus website login





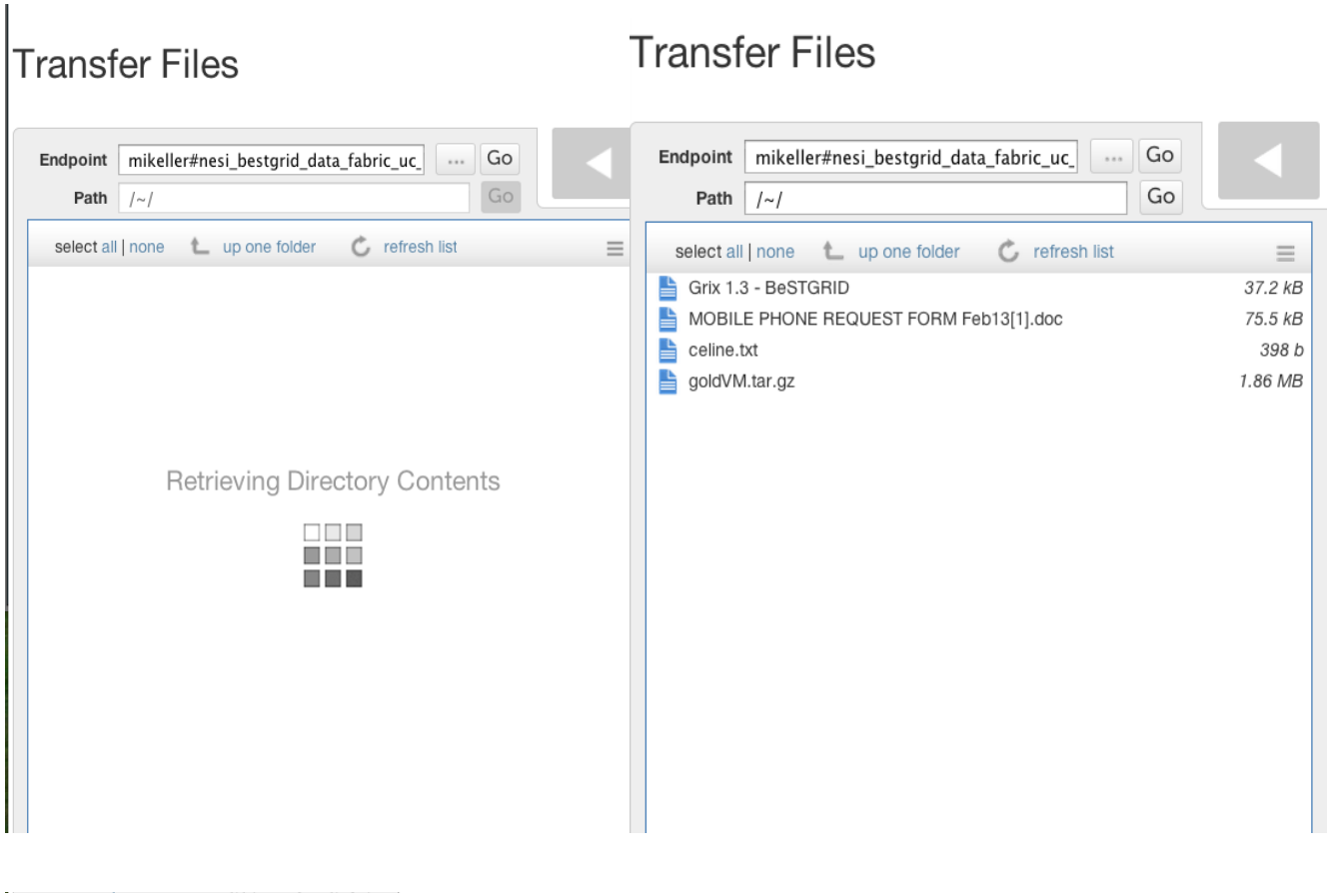
I can use my home university or research institution's credentials!

Without NZ IdP integration – Transfer end-point authentication



I now have to authenticate with different credentials than I used for logging into the website, e.g. using some stored in a myproxy repository

With integration – Transfer end-point authentication



Single sign-on epitomized – automatic authentication using cached credentials from website login

Enabling Globus SSO to access NeSI resources

Challenge 1

Globus supports myproxy-oauth for Single Sign-On (SSO), but there exists no ready made solution to integrate shibboleth (Tuakiri) with myproxy-oauth.

Action

NeSI developed patch for myproxy-oauth source code allowing injection of user attributes from shibboleth, and construction of customized Distinguished Name (DN) used in certificate issued by myproxy-oauth.

Next step

Discussion with authors of myproxy-oauth to get our patch integrated into official version of myproxy-oauth, as a contribution back to the community.

Enabling Globus SSO to access NeSI resources

Challenge 2

Globus supports version of myproxy-oauth integrated in their 'Globus Connect Server' (GCS) (formerly 'Globus Connect Multi User') product – NeSI is using a non-standard implementation of a GridFTP server, so using GCS is not an option.

Action

Reverse engineering of GCS source code to enable replication of the registration of a myproxy-oauth server for authentication with Globus

- Globus support's reaction: 'How on earth did you get this registered'. Request lodged with Globus support to have them develop official documentation of the process.

Enabling Globus SSO to access NeSI resources

Challenge 3

Our solution should be interoperable with existing legacy NeSI services (SLCS server, myproxy.nesi.org.nz), to enable Globus SSO sessions to access HPC storage.

Action

As stated, a patch was developed to enable a custom DN to be included in myproxy-oauth certificates. The custom DN was chosen in a way to be identical to the custom DN that is included in certificates issued by the NeSI SLCS server.

Enabling Globus SSO to access NeSI resources

Challenge 4

NeSI is not able to provide hosting for highly available service - the myproxy-oauth server (myproxyplus.nesi.org.nz) currently runs on ‘best effort’ hardware.

Next Step

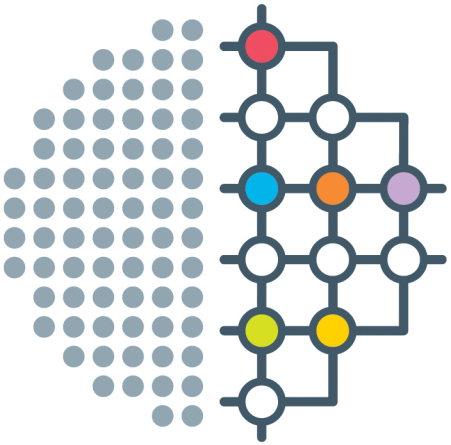
Finding a new, highly available home for NeSI’s myproxy-oauth server.

Current activities

- Discussion with Globus to get official documentation on the registration process for myproxy-oauth servers, to allow replication of our process based on official documentation.
- Establishing maturity in simplifying the setting up of other transfer end-points within NZ.
- Get operators of NZ HPC platforms to accept Tuakiri based authentication for access to their local data platforms.

Possibilities for NeSI/Globus

- As a customer, taking advantage of features offered by the subscription-based services.
- As a collaborator, working on select initiatives such as optimizing application-layer transfer protocols to maximize transfer performance.
- As a collaborator, continue to work with Globus to help support research data management use cases in NZ.



NeSI

New Zealand eScience Infrastructure

Thank you

www.nesi.org.nz