Addressing the data movement issue across AU and NZ

eResearch Australasia Conference 2022

Claire Rye (NESI), Richard Tumaliuan (REANNZ)
Ryan Fraser & Chris Myers (AARNet)

The Challenge?

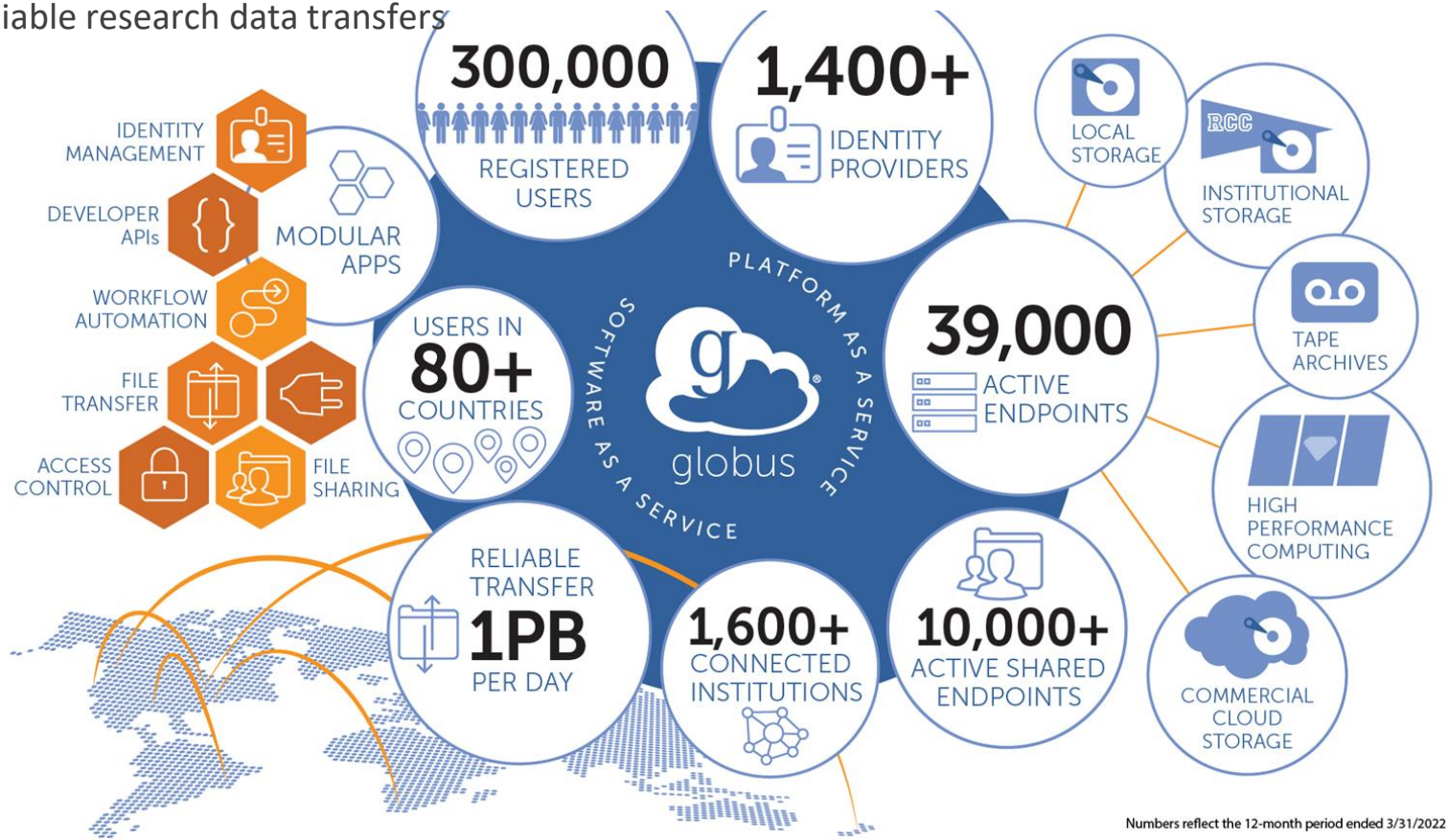
Can we transfer data

Faster
More reliably
More securely
More locations

Than UPS can?

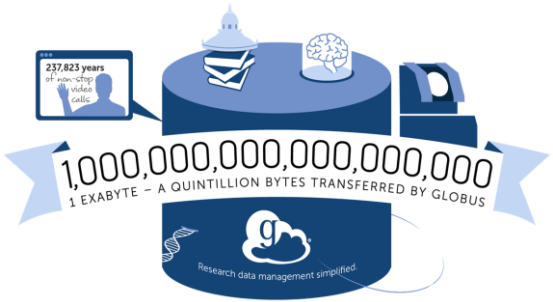
Globus

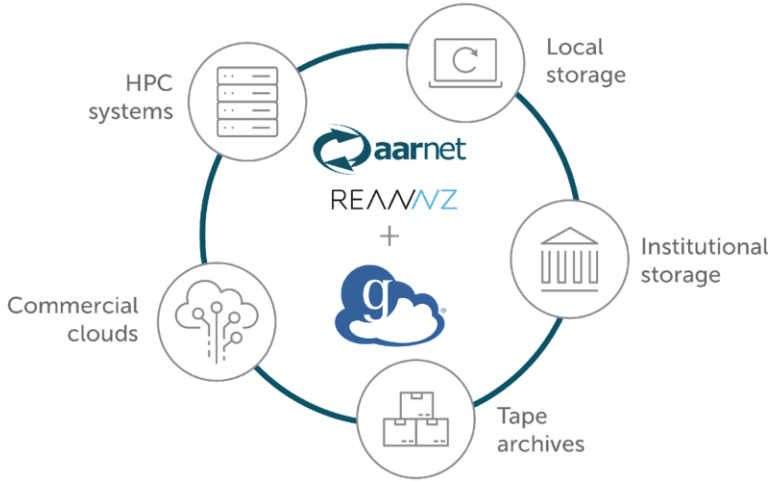
Fast, reliable research data transfers

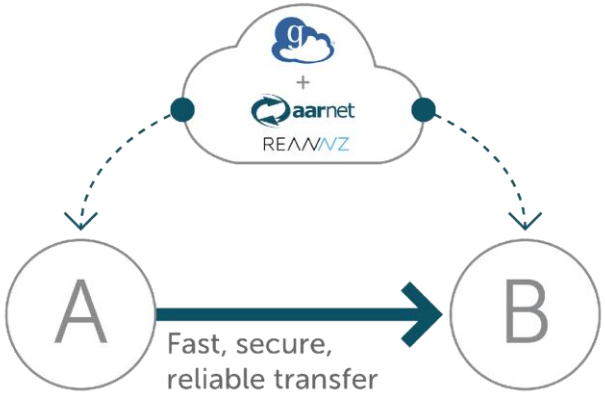


Globus

Fast, reliable research data transfers







Making it easier for researchers to collaborate and share large-scale data across organisational boundaries

NeSI
New Zealand eScience Infrastructure

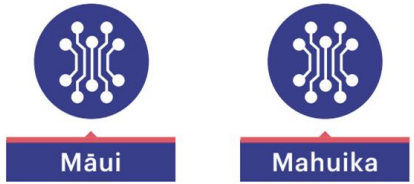
Disciplines Supported

- Biology
- Engineering
- Astronomy
- Physics
- Chemistry
- Computer Science
- Medical Science
- Earth Science
- Social Science
- Mathematics

Core Services

- High Performance Computing & Analytics
- Consultancy
- Training
- Data Transfer & Share

Shared infrastructure



>160 million	>1.3 petaflops	>130 GB/s
CPU core hours available per year	peak performance	IO bandwidth

Data replicated in Auckland. National research network provided by REANNZ.

University of Otago — Te Whare Wānanga o Otāgo — New Zealand

University of Auckland — Waipapa Taumata Rau — New Zealand

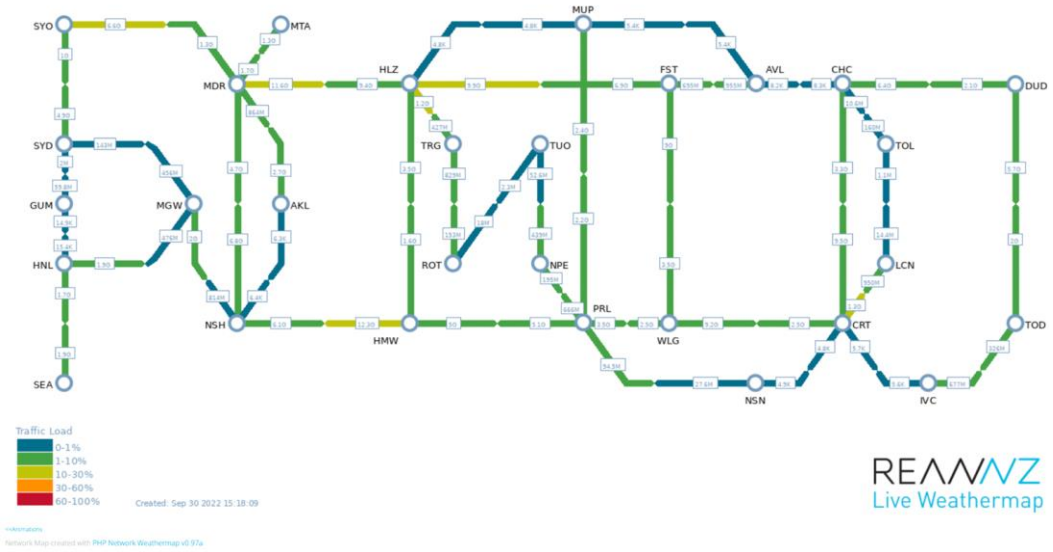
NIWA — Taihoro Nukurangi

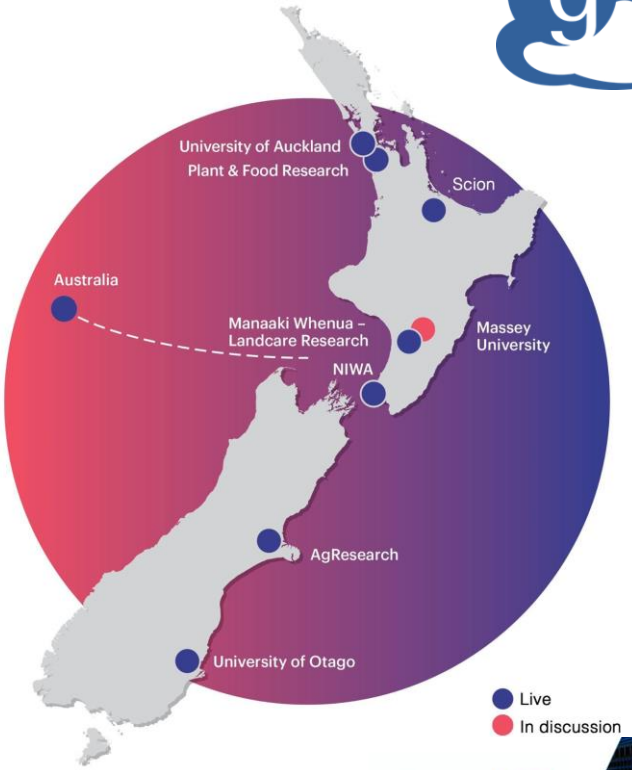
Manaaki Whenua Landcare Research

Ministry of Business, Innovation & Employment — Hīkina Whakatutuki

National data transfer platform activities in 2021:

799 TB	72 million	5,597
Amount of data transferred	Number of files transferred	Number of transfers made

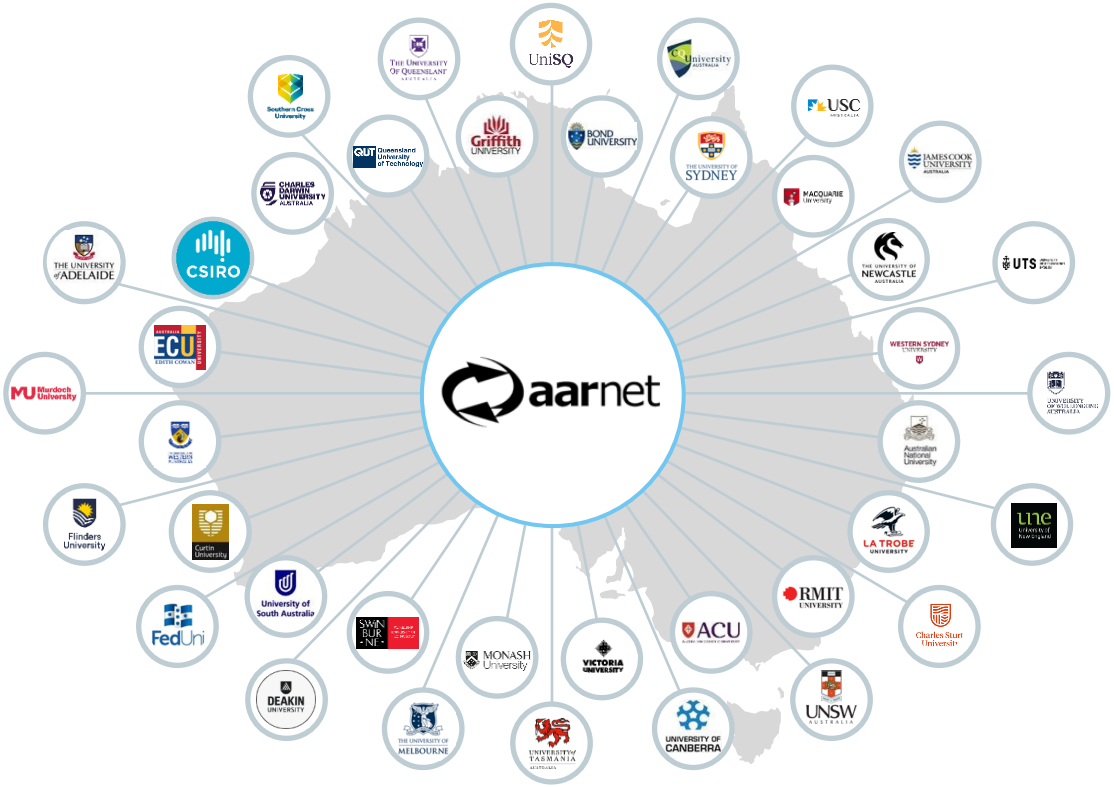




REANNZ
NeSI

Australian Approach: AARNet

- Australia's NREN
- Not for profit
- Owned by the Australian Universities and the CSIRO
- High-bandwidth, low-latency network
- Supports the research and education sector



AARNet: Globus

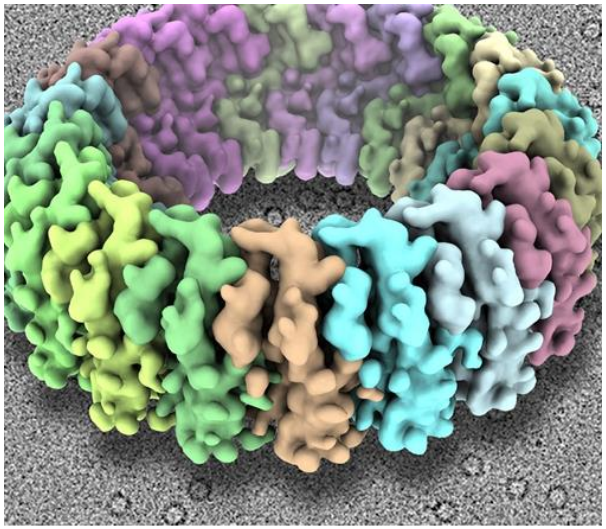
- Whole of Australia Research & Education Sector License, available through subscription
- Working with Institutes to configure and deploy (Institutes manage their own service endpoints)

Goal: Highly connected, accessible, reliable access to infrastructure and data within Australia and out to the World, so Researchers can focus on the science, rather than moving data

Use Case: BioCommons

- Research collaboration between Biomolecular Resource Facility based at the Australian National University and researchers at University of Otago
- Driven by need for researchers at BRF to easily and quickly move data out to collaborators
- NCI assisted BRF by establishing storage and globus endpoint for BRF. Leveraged existing University of Otago endpoints/storage

Use Case: Australian Characterisation (Microscopy)



- Focused on the measurement of material properties utilising Microscopy techniques including Cryo-Electron Microscopy (CryoEM)
- Huge data volumes generated and needed to move from labs to data storage and HPC (between long distances)
- Demonstrated between University of Wollongong and Monash University's MASSIVE HPC facility – now a production system

Key Point:

- **Focus on the Science rather than the data movement**

https://www.aarnet.edu.au/collaborating-to-solve-large-scale-research-data-transfer-challenges/

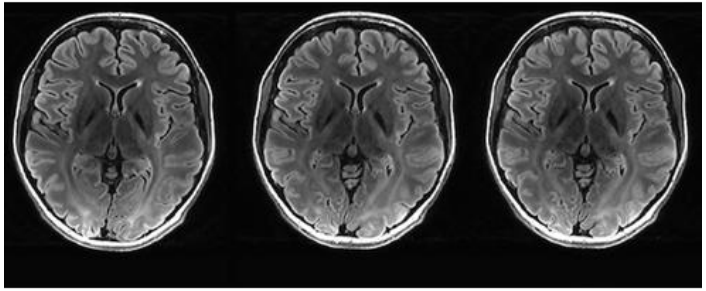
Use Case: Global Collaboration with ML on MRI Images

- Collaborating globally on the use of machine learning in effort to speed up reconstruction, pace and affordability of Magnetic Resonance Images (MRI)
- *"With the secure and fast data sharing enabled by Globus we can transfer large data sets both nationally and internationally, which made collaborations easier and faster"……"This frees up significant time to focus on data analysis rather than spending time on managing data transfers."*

Other Keys points:

Integrated into University of Queensland's data fabric – MeDiCi (data caching service) & supported by QCIF

https://rcc.uq.edu.au/article/2021/11/globus-web-app-makes-research-data-sharing-breeze

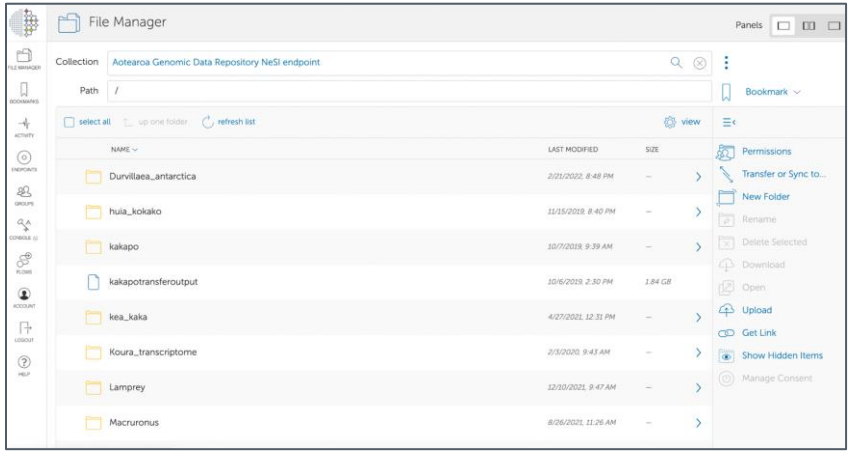


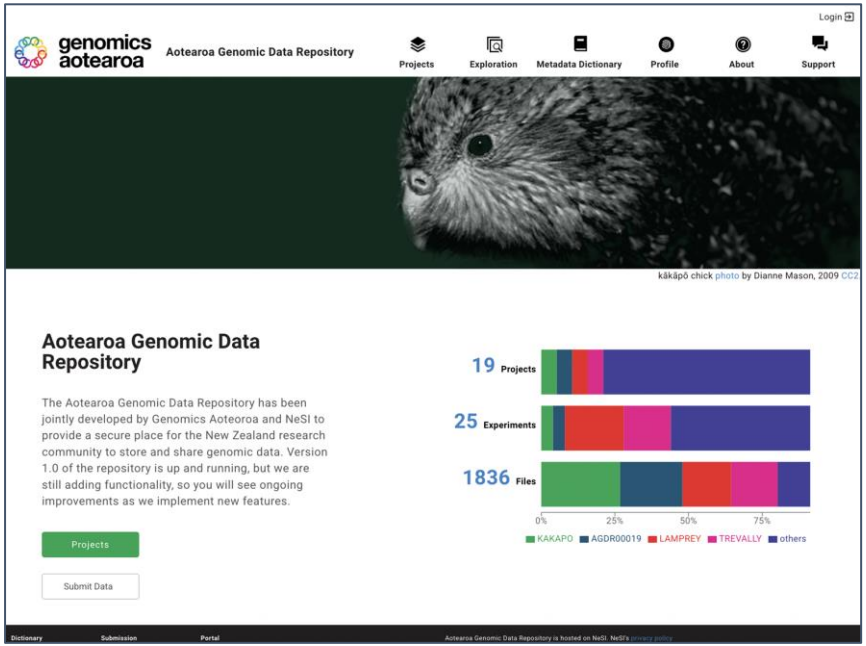
The brain scans show the reduction in acquisition time possible using deep learning without affecting image quality: Left—original scan 4min 49s; middle—original scan with calibration data 5min 18s; right—deep learning calibration 3min 18s. (Image courtesy of Dr Steffen Bollmann.)

Aotearoa Genomic Data Repository (AGDR)

Use case: Repository to Researcher

- The need: Māori data sovereignty over the taonga species repository, fine grain control on access to data.
- To move genomic datasets from AGDR to approved researchers wherever they are, once they had been approved for access.

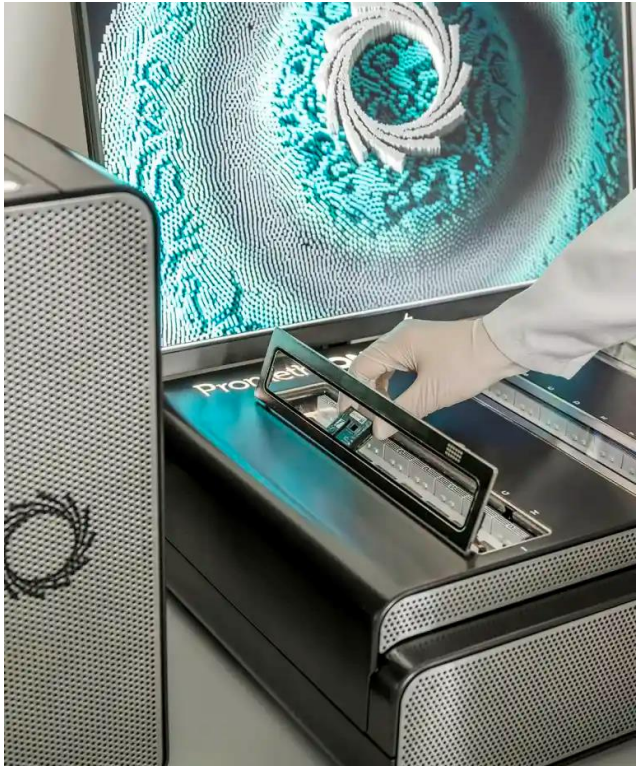




https://data.agdr.org.nz/

Bringing PromethION to Aotearoa

Use case: Instrument to Analysis compute



- The need: To move large volumes of sequence reads coming off the PromethION sequencer hosted at Lincoln University, and funded by Bragato Research Institute, to the compute facilities at NeSI to analyse and store.
- P24 for very high throughput. This provides up to 3.8Tb of data per full flow cell set in 72 hours.
- REANNZ created a Science DMZ using Lincoln University's infrastructure. This supported a personal globus endpoint to NeSI's Globus connect server, with a data transfer rate of ~400MB/s providing uninterrupted access to the compute process capability provided by NeSI.

https://bri.co.nz/2022/06/23/new-oxford-nanopore-promethion-sequencer/

BRAGATO RESEARCH INSTITUTE
RANGAHAU KAREPE, WĀINA O AOTEAROA

LINCOLN UNIVERSITY
TE WHARE WĀNAKA O AORAKI

REANNZ
NeSI

Accelerating New Zealand's Climate research

Use case: Sharing data with International collaborators

- The need: To move large volumes of climate simulation data between New Zealand and collaborators in the UK.
- The speed of data transfer across the network is enhanced by using New Zealand's National Data Transfer Platform, managed by NeSI and powered by Globus to provide fast, secure, and reliable transfers.

"It allows us to transfer data much faster than traditional methods. Because the data volumes that we need to transfer are so large, traditional methods are just simply not fast enough"

Dr Jonny Williams
NIWA climate scientist

https://www.reannz.co.nz/case-studies/accelerating-new-zealands-climate-research-global-collaboration-and-eresearch-infrastructure/

Get Moving your Data?

Summary

- Power of Globus is enabled through the physical network but strengthen by the network of endpoints
- Seeking researchers with local and “Trans-Tasman” data needs!

Contacts

Australia www.aarnet.edu.au/globus

- Ryan Fraser
  AARNet - Director, Digital Research
  ryan.fraser@aarnet.edu.au

New Zealand www.nesi.org.nz/services/data-services

- Claire Rye
  NeSI - Product Manager - Data Services
  claire.rye@nesi.org.nz
- John Graham
  REANNZ - Head of Engagement
  john.graham@reannz.co.nz

NeSI
New Zealand eScience Infrastructure

REANNZ

aarnet

Thank you

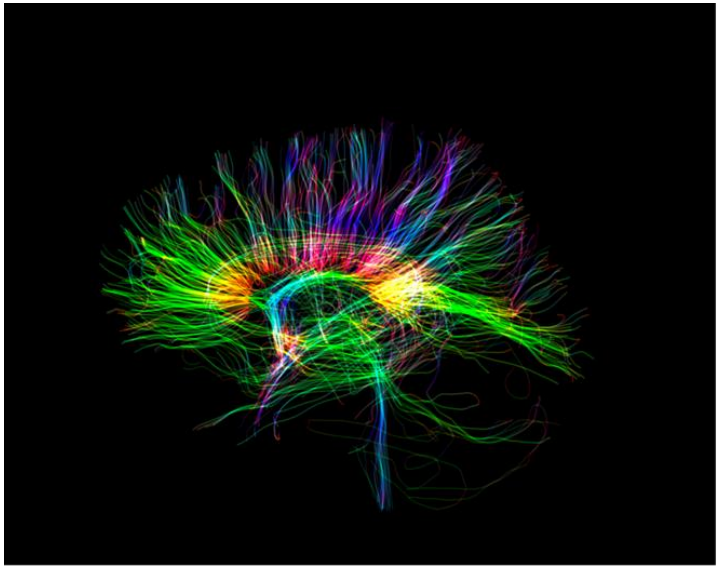
NeSI
New Zealand eScience Infrastructure

REANNZ

Extra slides

Use Case: MRI

Instrument to Analysis compute



- https://www.nesi.org.nz/case-studies/using-automation-enable-faster-collection-and-analysis-mri-data
  - Created a tool for automatically syncing a directory between two Globus endpoints
  - The tool can run on a schedule, such as overnight only, and resume transfers started previously but not completed
  - Option to delete the source once transferred to free up space at the source
  - Provides e-mail notifications on transfer progress and completion

"The automated data transfer with scheduling and destination setting functions allowed us to send large data out to multiple depositories, allowing streamlined image analysis."

Global National Research and Education Networks