Australian Research Data Commons

Virtual Laboratories and the Nectar Research Cloud

The Nectar Research Cloud is part of the Australian Research Data Commons, which is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy to establish eResearch infrastructure in partnership with Australian research institutions, organisations and research communities.

PRESENTED BY

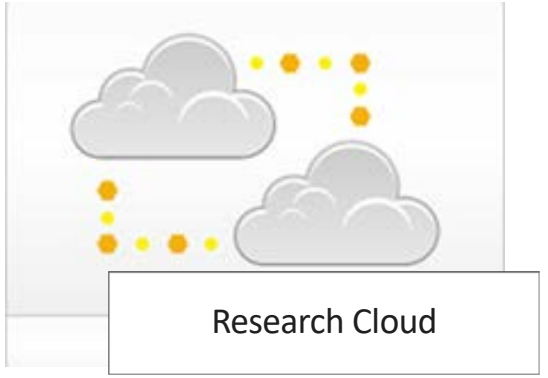
Paul Coddington, Associate Director Research Cloud and Storage
Australian Research Data Commons

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

NeCTAR

National eResearch Collaboration Tools and Resources

- Funded by Commonwealth Govt since 2010
- Cloud based infrastructure to enable research collaboration across institutional boundaries
- Enable researchers to store, access, share, and analyse data



National Research Cloud

- Computing infrastructure, software and services - at scale
- Self-service capability to quickly provision resources

Virtual Laboratories

- Domain-oriented online environments
- Combine research data, models, analysis tools and workflows
- Support national and international collaborative research

From NeCTAR to ARDC

- The NeCTAR program was funded through the federal government's National Collaborative Research Infrastructure Strategy (NCRIS)
- The 2016 NCRIS Roadmap called for a merging of 3 NCRIS facilities - NeCTAR, Research Data Storage, Australian National Data Service
  - "Establish an integrated data-intensive infrastructure system, incorporating physical infrastructure, policies, data, software, tools and support for researchers"
- The Australian National Data Commons (ARDC) was formed in 2018
- ARDC now supports the Nectar Research Cloud and Virtual Laboratories (through its Platforms program)
- ARDC is part of a broader Digital Data and eResearch Platforms theme including national HPC centres, AARNet NREN, and AAF
  - "Nationally coordinated eResearch infrastructure that builds on existing capabilities and leverages institutional investments"

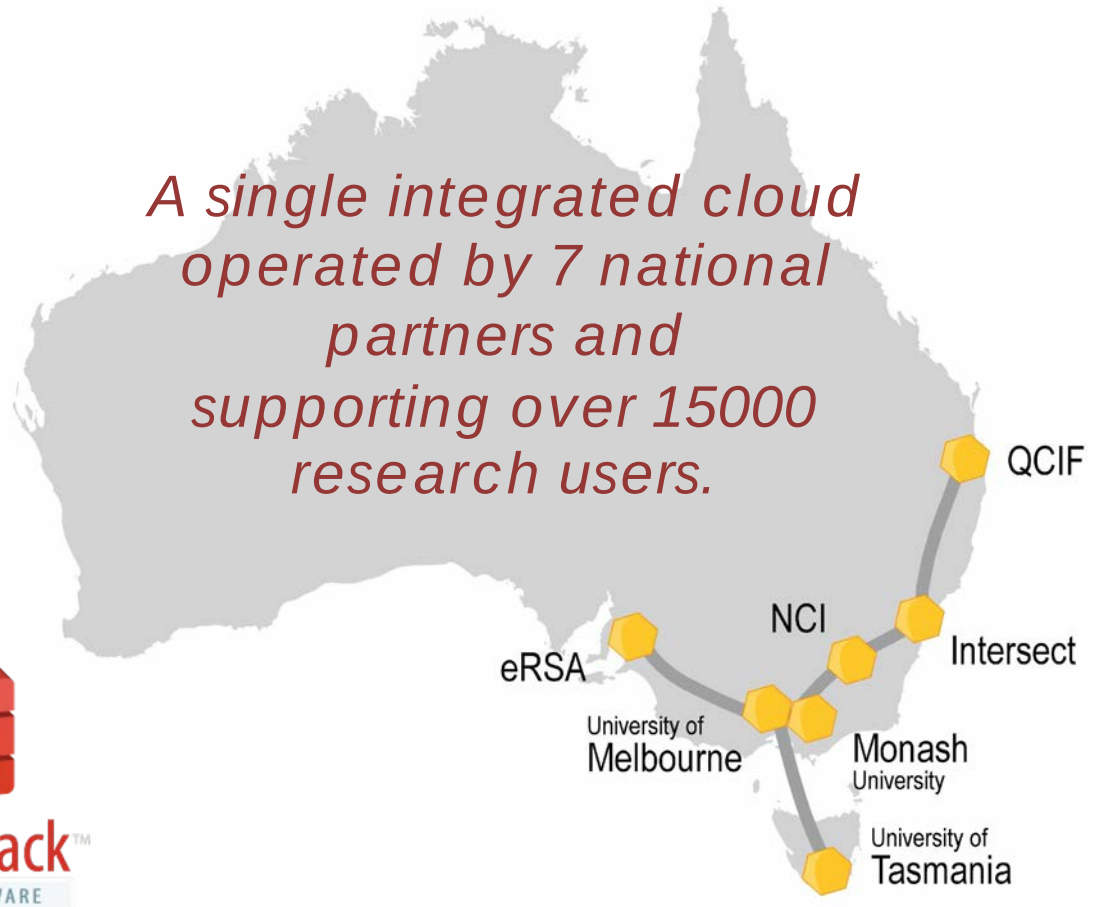
The Nectar Research Cloud

A world first…

The Nectar Research Cloud is a partnership between 7 institutions and research organisations who are operating a federated Openstack research cloud.

- University of Melbourne
- National Computation Infrastructure (NCI)
- Monash University
- Queensland CyberInfrastructure Foundation (QCIF)
- eResearch SA (eRSA)
- University of Tasmania
- Intersect, NSW

A single integrated cloud operated by 7 national partners and supporting over 15000 research users.



openstack™ CLOUD SOFTWARE

Australian Research Data Commons

THE UNIVERSITY OF MELBOURNE

MONASH University

qcif

eRESEARCHSA

NCI NATIONAL COMPUTATIONAL INFRASTRUCTURE

UTAS

INTERSECT

Nectar Research Cloud Federation

A single national cloud interface

- OpenStack cells to support 7 regional sites
- Other institutions have joined with their own resources (Swinburne, Auckland, RMIT)
- Users can request a specific site – or *deploy anywhere.*

National services

- Cloud dashboard (Horizon) and API
- Authentication (Keystone, AAF)
- Image repository (Glance)
- App store (Murano)

Federated services

- Object store (Swift)
- Compute (Nova) and volume storage (Cinder)
- User support, help desk, user guides, documentation

Australian Research Data Commons

Project Trials and Allocations

- User logs in to the Nectar Research Cloud dashboard with AAF credentials.
- Agree to Terms of Use on initial login.
- Project Trial (2 VCPUs for 3 months of usage) automatically allocated.
- Users can apply for a project allocation before or on expiry of project trial by completing and submitting an allocation request form from the dashboard.
- Allocations are approved for provisioning under the Research Cloud National Allocation Scheme by the Nectar Allocation Committee
- Allocation Requests usually assessed within 2-3 weeks.

Project Allocations

- The Nectar Research Cloud has two categories of cloud infrastructure:
  - National - ARDC funded, accessible through a national allocation scheme
  - Local - Node funded, for Node-prioritised (local) allocations
- Projects are eligible for a national allocation if they meet specific criteria:
  - National competitive research grant (e.g. ARC, NHMRC)
  - Supports (or is funded by) an NCRIS capability – including ARDC Platforms projects
  - Is approved by the Nectar Allocation Committee as meeting other merit criteria
- Local allocations require some arrangement with a Node

Nectar Cloud User Support

National Distributed Helpdesk (DHD)

Operates 7am to 6pm AEST/AEDT.

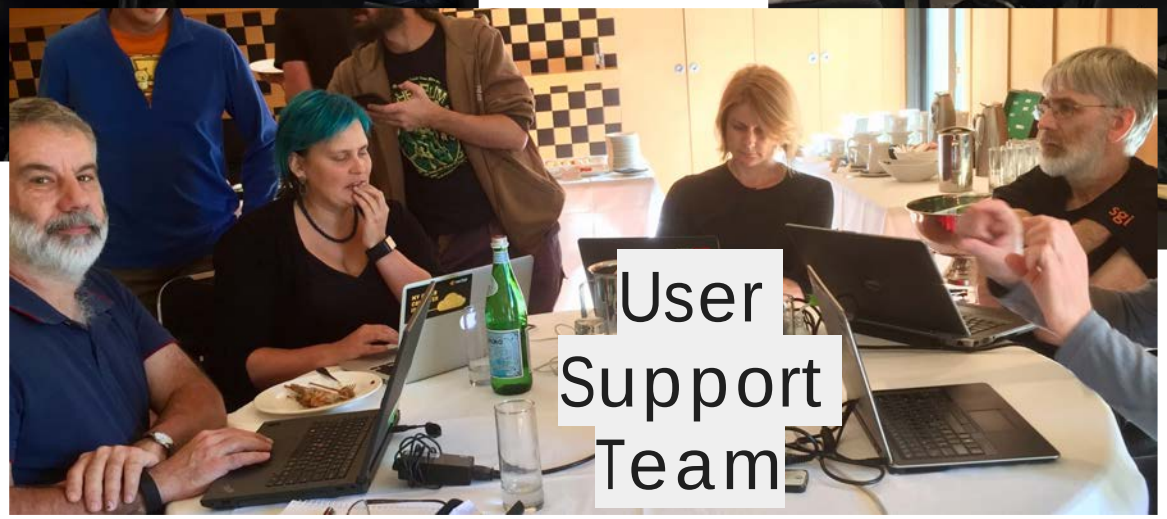
Users can contact the national helpdesk via:

- Email support@ehelp.edu.au
- Web interface from the Support Portal or the Helpdesk
- Live chat from the Helpdesk

SLA User Support Targets in place - respond to support tickets within 2hrs and resolve within 8 hours

Nectar Research Cloud Ops Team

…a federation of 8 Node operators



Australian Research Data Commons

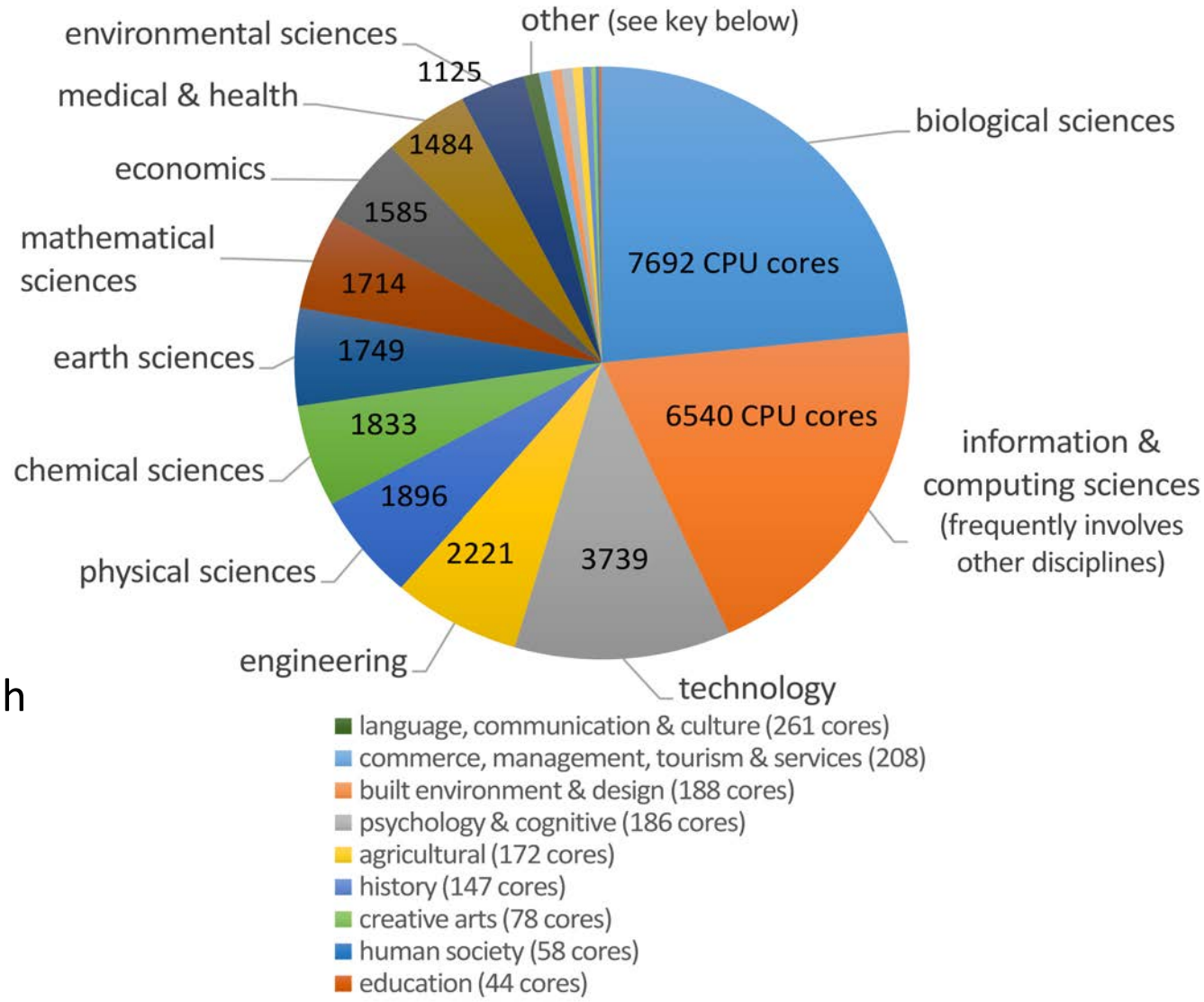
Usage

- ~8000 VMs in use at any time
- ~50,000 cores and 2 Pbytes of storage
- ~2000 users running VMs
- ~3500 users ran VMs last year
- >17,000 users since 2012
- 200 new registered users per month
- >3900 project allocations
- >400 projects with national research grants last year
- 16 Virtual Labs, 11 CRCs, 11 ARC CoEs, 13 NCRIS capabilities
- Tens of thousands of users of VMs and services that run on the cloud

Allocation request word cloud

Australian Research Data Commons

Who is using the Nectar Research Cloud?



Project allocations

2-digit Field of Research codes

Australian Research Data Commons

Cancer Therapeutics CRC

Access to cancer research data, tools and visualisation on the NeCTAR Cloud

Providing access to analysis and visualisation tools, and over 30TB of cancer research data **on the Research Cloud.**
The Nectar choice was easy, and the migration process seamless.

"The service, support and responsiveness that we have received from the Nectar team has been first class, and feels like an extension to our own internal support services."

Paul Reeve, Director of Operations, Cancer Therapeutics CRC.

Plant Energy Biology Centre of Excellence

Building collaboration on the Research Cloud.

Researchers study how plants capture energy from sunlight and how they use that energy to grow and develop.
Hosting collaborations with the Max Planck Institute and the Beijing Genomics Institute – ***on the NeCTAR Research Cloud.***

"NeCTAR makes it much easier, much faster. It means more collaborations — projects that would have just been too hard to go ahead."

Professor Ian Small, Laureate Fellow, West Australian Scientist of the Year 2015.

NeCTAR Research Cloud

*Supporting innovation and collaboration in the **business** of research.*

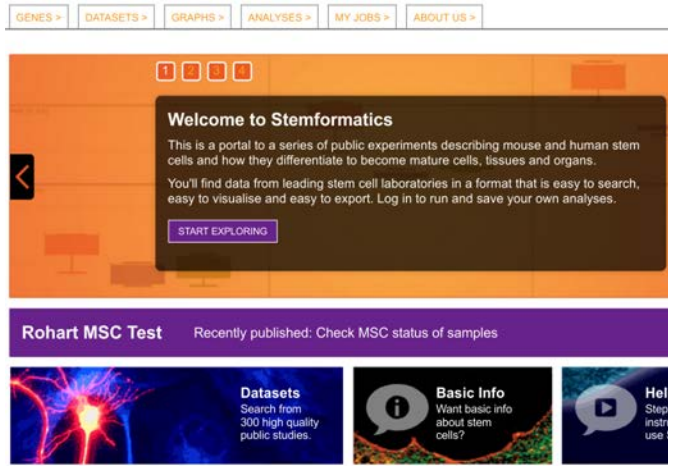
Australian Research Data Commons

Stemformatics

Stem Cell data visualisation on the Cloud.

Find and visualise interesting genes in datasets from leading stem cell laboratories on the **Research Cloud.**

- Over **400 users** nationally
- **100 cores**, multi-site
- NCRIS supported.



Future of the Nectar Research Cloud

- $4.1M of ARDC funding to refresh existing Node infrastructure this year
- Plus $2-3M per year for next 3 years
- 2021 focus on supporting needs of ARDC Platforms
- More high-end VMs – GPUs, big memory
- Make it easier to use
- More online training material, webinars, hands-on training courses
- Add to the app store
- New services – Jupyter/R, Virtual Desktop, high-throughput computing, …
- Better alignment with international cloud initiatives e.g. EOSC
- How to best take advantage of public cloud and multi-cloud

Aims of Virtual Laboratories program

- Accelerating access to research data, tools and models
  - Bringing together access to ***modelling and observation***; platforms, tools and applications to derive knowledge from data.
- Removing barriers to collaboration
  - Supporting cross-institutional and international research collaboration
- Creating partnerships/collaborations at precinct and national scale
  - Multi-institutional – highly networked – supporting and building research communities
- Researcher led governance
  - Addressing real needs identified by national research communities
- Partnering with eResearch capability providers

Nectar Virtual Laboratories (2012-17)

Climate and Weather Science Laboratory – Lead: Bureau of Meteorology – 6 Partners

- *Integrated environment for climate and weather science modelling and data*

Genomics Virtual Lab – Lead: University of Queensland/University of Melbourne – 9 Partners

- *Easy access to Genomics tools and resources for Australian biologists.*

Endocrine Genomics Virtual Lab – Lead: University of Melbourne – 7 Partners

- *Statistical power for clinical research*

Marine Virtual Lab – Lead: University of Tasmania – 8 Partners

- *Ocean observations and modelling to improve planning for marine and coastal environments.*

All Sky Virtual Observatory – Lead: Astronomy Australia Limited – 4 Partners

- *Theoretical and observational astronomy data, simulations and tools accessible from your desktop*

Biodiversity and Climate Change Virtual Lab – Lead: Griffith University – 18 Partners

- *Simplifies biodiversity-climate change modelling.*

Nectar Virtual Laboratories

Humanities Network Infrastructure – HuNI – Lead: Deakin University – 13 Partners

- *Integrating 28 of Australia's most important cultural datasets*

Characterisation Virtual Lab – Lead: Monash University – 11 Partners

- *Integrating Australia's key research imaging instruments with data and analysis tools on the cloud.*

Geophysics Virtual Lab – Lead: CSIRO – 7 Partners

- *Easy access to geophysics workflows, simulations and datasets.*

Alveo – Human Communications Sciences – Lead: Western Sydney University – 16 Partners

- *Studying speech, language, text, and music on a larger scale*

Industrial Ecology Virtual Laboratory – Lead: Sydney University – 9 Partners

- *Supporting comprehensive environmental carbon footprinting and sustainability assessments*

Virtual Laboratory Case Studies

Genomics VL

"This is the best exemplar of this kind of platform in the world… Genomics capability for the masses."

Associate Professor Andrew Lonie, Director, EMBL-ABR.

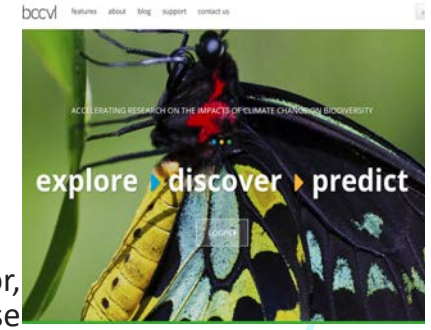
The **Peter MacCallum Cancer Centre** is using the GVL in the NeCTAR Research Cloud, providing instant access to Genomics tools and data for Australian biologists

BPA

Biodiversity and Climate Change VL

"..decreases the time to complete biodiversity analysis from 2 months to 5 minutes, supporting new applications in research, government and industry."

Professor Brendan Mackay, Director, Griffith Climate Change Response Program



Accelerating biodiversity–climate change modelling across large disparate datasets quickly and easily on the Research Cloud.

TERN

Virtual Laboratories are:

- *Accelerating* research
- Bringing together **observation** and **modelling**
- Removing barriers to collaboration
- Leveraging the Research Cloud for wide access

Marine VL

"MARVL enables researchers to start thinking about their problem sooner."

Dr Roger Proctor, Director e-Marine Information Infrastructure Facility.

Ocean observations and modelling for marine and coastal environments

Ian Coghlan is studying coastal erosion. **MARVL saves him 3 months effort** to access local data, wave model simulations and computing resources.

IMOS

Australian Research Data Commons

A complete bioinformatics analysis platform with:

1.34M bioinformatics jobs run

1178 bioinformatics tools

220 reference genomes

... and on average 2 new tools or references added weekly on user request

Supporting published research across many areas:

- Cardiac
- Toxicology and Drug Discovery
- Cancer
- Vision
- Agriculture and land management
- Obesity
- Ecology
- Infection and immunity
- Evolution

62 peer reviewed papers

14 higher degree theses

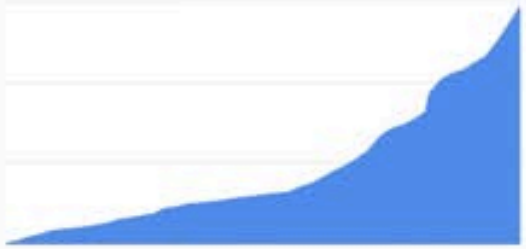
have used Galaxy Australia to support their bioinformatics analyses

An active and engaged user community

8111 registered users.

1950 active users (last 90 days)

User growth 2016 - 2020



Registered users in Australia from:

30 Australian Universities

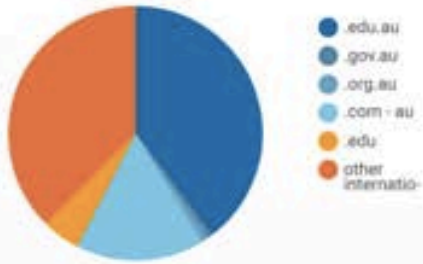
22 Medical Research Institutes or Organisations

19 Other Research Organisations

4 teaching courses utilising Galaxy Australia annually

2171 people trained in hands-on workshops

Users per domain



Users represented across **420** organisations

Users represented across **49** countries

"We have sequenced well over 500 transcriptomes/genomes, and routinely use Galaxy Australia, powered by the GVL, for many bioinformatics processes. It is easy to use, has high computational power, a sophisticated support structure and enables global collaboration through straightforward data sharing. We greatly appreciate the service."

- Dr Fabio Cortesi and Prof Justin Marshall, Queensland Brain Institute
Science Advances 2017 3(11), eaao4709

All statistics are Jan 2016 - May 2020

Galaxy Australia is supported by Bioplatforms Australia and ARDC, both of which are enabled by NCRIS. Co-investments have also been made by project partners.

ARDC Australian Research Data Commons

NCRIS National Research Infrastructure for Australia

BIOPLATFORMS AUSTRALIA

DeVLs (2017-2019)

- All Sky Virtual Observatory – Astronomy
- AuScope Virtual Research Environment – Geophysics
- C-DeVL - Characterisation
- Climate and Weather Data Science
- EcoCloud – Ecology
- Galaxy Australia – Genomics
- Humanities and Social Sciences (HASS) - Tinker
- MARVL – Marine VL

ARDC Platforms projects (2019-23)

- Australian Imaging Service – XNAT for microscopy and imaging
- Australian Characterisation Commons at Scale – CVL++
- Australian Transport Research Cloud – from AURIN
- Australia's Scalable Drone Cloud
- BioCommons Platform - Galaxy Australia, Genomics
- Coordinated Access for Data, Researchers and Environments (CADRE) - HASS
- EcoCommons – EcoCloud, BCCVL, etc
- Environments to Accelerate Machine Learning Based Discovery
- E-Research Institutional Cloud Architecture (ERICA) – sensitive data analysis
- FAIMS 3.0 Electronic Field Notebooks

Increase the number of researchers with access to platforms, and the diversity of disciplines