

NCRIS AGILITY FUND

Australian Research Community Clouds Overview

NeCTAR

The University of Melbourne

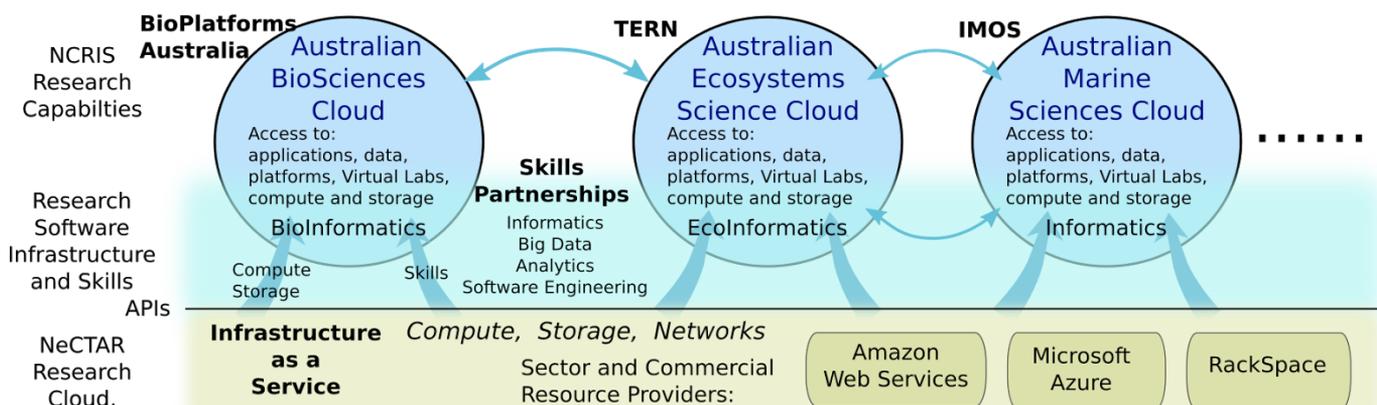


NCRIS
National Research
Infrastructure for Australia
An Australian Government Initiative

Overview

This proposal will establish for the first time:

- An **Australian BioSciences Cloud** to be operated by NeCTAR, RDS and NCI in partnership with BioPlatforms Australia (BPA);
- An **Australian Ecosystems Science Cloud** to be operated in partnership with the Terrestrial Ecosystem Research Network (TERN); and
- An **Australian Marine Sciences Cloud** to be operated in partnership with the Integrated Marine Observing System (IMOS).



The proposal will leverage significant co-investment of more than three times the requested NCRIS Agility funds to support the establishment and initial operations of these *Australian Science Clouds* (or *Research Community Clouds*).

These *Science Clouds* will leverage the existing NeCTAR Research Cloud, Virtual Laboratories, Research Data Storage Infrastructure (RDSI) collections, NCI capabilities and RDS Data Services infrastructure.

Links to existing research infrastructure:

These three science clouds will operate on the NeCTAR Research Cloud and provide support and access for BPA, IMOS and TERN related initiatives, including EMBL-ARB, QFAB, AODN, MaRVL, GVL, CoESRA, BCCVL and so on. Establishment of these Research Community Clouds will elevate international recognition and enhance linkages with international science platforms, including between the *Australian BioSciences Cloud* and Europe's ELIXIR program and the UK's CLIMB microbial bioinformatics cloud. The *Australian Ecosystems Science Cloud* will enable easy access to data and analytical tools in an "as a Service" model so that the cloud infrastructure will be used more effectively. The *Australian Marine Sciences Cloud* will establish the first stages of a fully functional cloud service for marine scientists nationally and internationally.

This initiative will ensure that the Research Cloud and Virtual Laboratories are responsive to the national priority needs of BPA, TERN, IMOS and their stakeholders.

The proposed infrastructure will be established by extending, upgrading and integrating existing NeCTAR, RDS, BPA, TERN and IMOS infrastructure, including through enhanced integration and access. The proposal will also support further scoping/planning processes to direct further

development and planning against future need, including international access and alignment.

Benefits

The NeCTAR Virtual Laboratories and Research Cloud programs were established as novel investments in the development of innovative research infrastructure. These innovative investments have supported over 7,500 users of the Research Cloud infrastructure service and over 10,000 research users of the Virtual Laboratory platforms.

The proposal will establish a new model for ***sustaining improved strategic relationships and alignment between NCRIS research-domain capabilities and eResearch infrastructure providers (NeCTAR)***; addressing key issues raised during the consultation phase of the eResearch Framework process.

The proposal will provide a vehicle for:

- NCRIS domain capabilities to partner with eResearch infrastructure providers to:
 - influence the design and delivery of eResearch infrastructure to better support their strategic planning and changing infrastructure needs;
 - to guide the prioritisation of resource allocations and services to better address national and research community priorities; and
- The NeCTAR Research Cloud to deliver greater value by supporting higher-level services in partnership with research communities:
 - Specific research computing needs vary significantly between, and even within, research disciplines. Partnering with NCRIS domain capabilities provides a mechanism to support higher-value services on the Research Cloud.

This strategy is proposed as a key element enabling future multi-year planning with NCRIS domain capabilities, and ensuring that e-infrastructure is fully integrated into, and supporting, the national research infrastructure landscape.

Additional benefits provided by such an approach are expected to include:

- Deeper engagement and uptake with Australian researchers:
 - Leading to more attractive platforms to attract institutional and sectoral co-investment;
- Provision of a mechanism for drawing together the cloud program and ongoing support for Virtual Laboratories (and other hosted platforms in the domain clouds);
- Opportunity for engagement and interoperability with emerging international research-domain oriented cloud platforms (eg. ELIXIR EU, CLIMB UK, NSF Clouds - Jetstream, Chameleon);
- Provision of a mechanism to jointly support higher-value research applications and services deployed on the NeCTAR Research Cloud;
- Expected to be more attractive vehicle for delivering sectoral co-investment; and
- Provision of a mechanism for addressing issues identified in eResearch Framework consultations.

It is proposed that *Australian BioScience Cloud*, *Australian EcoSystems Science Cloud* and the *Australian Marine Science Cloud* will be associated (“branded”) with BPA, TERN and IMOS; with NeCTAR, RDS and NCI as supporting partners).

The proposal will also support initial planning for leveraging multi-cloud support (including commercial and strategic international scientific clouds).

Australian BioSciences Cloud

Biosciences research in Australia involves a diverse set of activities from local to national, across a range of life sciences areas (e.g. genomics, clinical, translational, population studies, etc.) and consequently has a diverse set of complex digital infrastructure needs. Identifying how informatics capabilities and infrastructure can strategically support national priorities in this research community is central to the value of this platform. By providing a connected environment, linking existing bioinformatics skills, infrastructure, software and data, the platform will support a number of prioritised research activities.

National initiatives such as BPA are already producing reference datasets of international relevance and interest (e.g. Sepsis multi-omics data resource). Bioinformatics and expertise is existent through key international efforts such as EMBL-ABR, and key research projects such as ASPREE are generating data which requires integrated environments from which to derive most value. The *Australian BioSciences Cloud* will play host to tools, data and managed services which support the informatics needs of these already-prioritised initiatives. Primary users of the platform will come from medical research precincts (Clayton and Parkville) and their medical research institutions, as well as biosciences-rich institutions including UQ, Monash, Melbourne and UNSW.

Australian EcoSystems Science Cloud

The *Australian Ecosystems Science Cloud* will significantly lower the barriers for research and industry sectors to access relevant data and analysis tools and accelerate research impact and innovation. The *Australian Ecosystems Science Cloud* combines advanced software and cloud infrastructure, scientific and information technology expertise, and training to meet emerging needs of the ecosystem sciences research community.

There are approximately 6,000 ecosystem science researchers and students within the Australian research and higher education sector. Most of these researchers use wide varieties of datasets in their research including climate, ecology, biodiversity and ecosystem related remote sensing datasets. These datasets are available from different and often from multiple sources. The NCRIS investment has delivered unprecedented access and development of eResearch infrastructure involving TERN, Atlas of Living Australia (ALA), NeCTAR, RDS and the Australian National Data Service (ANDS). These eResearch infrastructures have offered researchers access to cloud computing widely as an infrastructure as a Service, open access data services and customised application specific virtual laboratories. However, the significant impact of the previous investment will be realised with the more co-ordinated approach for the service delivery tailored to the needs of researchers with minimal impediments to access to ecosystem data, tools and compute resource for their research completely abstract from where the infrastructure is hosted.

The *Australian Ecosystems Science Cloud* will lower the barriers to access ecosystem science data collections, analytical and synthesis tools and cloud compute resource. The *Australian Ecosystem Sciences Cloud* will be a customised cloud infrastructure under NeCTAR to provide collaborative cloud resource and infrastructure to enable high-impact science in ecosystem science discipline. This will be accomplished by providing an open and publicly accessible centralised repository hosted on NeCTAR OpenStack swift object store service. The repository will be built based on the community engagement and input. The repository will host valuable data collections from TERN, ALA and other ecosystem science data providers and contributors. This will bring siloed, fragmented and structured datasets under single platform that can be accessible by a simple API. The data will be made seamlessly accessible from the NeCTAR cloud computing infrastructure including existing virtual laboratories and analytical tools. For researchers to build reusable tools and applications, a scalable cloud computing resource will be offered. The *Australian Ecosystem Sciences Cloud* will streamline the accessibility of ecosystems science large data collections for analysis, synthesis and providing customised services over the data. The *Australian Ecosystems Science Cloud* will also enable researchers and informaticians to write custom tools and services to access, analyse and visualise datasets without worrying about underlying platform hurdles.

Australian Marine Sciences Cloud

Marine science is driving the development of Australia's blue economy. The *National Marine Science Plan* (2015) states that by 2025, Australia's marine industries will contribute around \$100 billion each year to our economy. There is a clear need to facilitate coordinated national studies on a range of marine system processes, and eResearch infrastructure is critical for achieving this goal.

National eResearch infrastructure initiatives led by IMOS, TPAC and NeCTAR, amongst others, have created data repositories and toolsets to facilitate advancement of collaborative research in marine science. This proposal will enable hosting of existing platforms in one place, enabling deeper engagement with the Australian and international marine science communities on those issues identified as priorities.

The IMOS Ocean Portal (<http://imos.aodn.org.au>) allows marine and climate scientists and other users to discover and explore data streams coming from all of these Facilities, and the Marine Virtual Laboratory (supported by IMOS, TPAC and NeCTAR) provides a research environment within which to explore BioSciences the science questions around seamless integration of the complex multi-disciplinary marine environment. These initiatives integrate with major Australian data collections, including the RDSI/RDS Marine Science Data Collection at TPAC, Pawsey Centre and JCU, and data collections in the National Environmental Data Collection at NCI, and feed into the National Environmental Information Infrastructure through the Australian Ocean Data Network.

Partners

BioPlatforms Australia

BPA has funded the creation of a number of internationally significant reference data sets within their Framework dataset program, one of which the Sepsis dataset (the focus of an RDS Flagship and the NeCTAR microbial-genomics-VL platform).

University of Melbourne

Research Platform Services at the University of Melbourne operates the Melbourne node of the research cloud and is engaged in a number of data and cloud initiatives in the biosciences space. It will focus on technical integration, tool deployment and publishing relevant data. It also leads the Research Bazaar, a national-scale digital training initiative for researchers.

VLSCI/EMBL-ABR

EMBL Australian Bioinformatics Resource (EMBL-ABR) is a distributed national research infrastructure providing bioinformatics support to life science researchers in Australia, with in-kind focussing on national training and reference data resources. VLSCI is strongly engaged in a number of national and international data and tool efforts, and will provide in-kind co-investment and significant expertise in workflow and analysis tools.

Monash University

Monash's in-kind will focus on training development and support, as well as the development of the medical genomics data safe-haven.

QCIF/QFAB

QCIF delivers data-intensive digital infrastructure and services to accelerate Queensland research, innovation and application. It will be the primary cloud node for the *Australian Ecosystem Sciences Cloud*. TERN and QCIF have a long history of collaboration on initiatives including the Terrestrial Data Discovery Project, the Collaborative Environment for Ecosystem Science Research and Analysis, and the Biodiversity and Climate Change Virtual Laboratory, a NeCTAR Virtual Laboratory that facilitates access by over 300 users to 4,400 datasets and tools. QCIF delivers data-intensive digital infrastructure and services to accelerate Queensland research, innovation and application. QCIF supports bioscience research across the six leading universities in Queensland by operating the QRIScloud node of the national research cloud and providing expert bioinformatics consulting services through QFAB. It will become a partner node in the biosciences cloud tool deployment, data access, data publication, and training for the biosciences research community.

TERN

Key partners in the development of the Australian Ecosystem Sciences Cloud include TERN and QCIF, who are already involved in facilitating the transformational change to create a national, multi-disciplinary, networked, and collaborative community that for delivering Australia's future. TERN delivers critical research infrastructure and supporting national and international networks of scientists, environmental managers and stakeholders, needed to improve understanding and management of Australia's ecosystems. They are ideally placed to lead engagement with the research community on the development of the *Australian Ecosystem Sciences Cloud*, provide

research expertise and offer technical support for the development of the *Australian Ecosystem Sciences Cloud*.

IMOS

IMOS and TPAC are ideally situated to represent the marine science community in the development of the *Australian Marine Sciences Cloud*. Since 2006, IMOS has been routinely operating a wide range of observing equipment throughout Australia's coastal and open oceans, making all of its data accessible to the marine and climate science community, other stakeholders and users, and international collaborators. IMOS operates as a multi-institutional collaboration. IMOS is led by the University of Tasmania in partnership with the CSIRO, Australian Institute of Marine Science, Bureau of Meteorology, Sydney Institute of Marine Science, University of Western Australia, Curtin University and the South Australian Research and Development Institute. IMOS is also instrumental in developing the Australian Ocean Data Network.

TPAC

The Tasmanian Partnership for Advanced Computing (TPAC) is partnered by the University of Tasmania, CSIRO Marine & Atmospheric Research, the Australian Government Antarctic Division, Antarctic Climate & Ecosystems Cooperative Research Centre (ACE CRC), and the Australian Maritime College. TPAC provides expertise in storage and hosting of Research collections (2.7 Pbytes disk, 2Pbytes or tape silo), operate Australia's third largest cloud deployment and high performance computing facilities for the Australian and international research community. The University of Tasmania is investing \$3million in a new HPC facility and data centre (in 2016 and 2017) calendar years. This data centre has e-Research as its primary reason for existence.