LEADING AUSTRALIA TO DATA-DRIVEN RESEARCH IMPACT

ARDC

Australian Research Data Commons

A MESSAGE
FROM OUR CEO

2020 began with national-scale bushfire and flood disasters and then confronted us with the COVID-19 pandemic. These events highlighted the critical role of data in addressing unpredictable challenges.

The COVID-19 pandemic continues to demonstrate the value of making data findable, accessible, interoperable and reusable (FAIR) for identifying potential vaccines, tracking the spread of the virus and uncovering ways to minimise exposure. FAIR data empowers researchers, industry and government to collaborate across organisational, sector, and geographical borders to rapidly respond in times of emergency.

Improving the ways we collect, store, share and use data within and across research domains builds resilience for Australia—empowering our society to prepare for, manage and emerge from global-scale challenges as a stronger and more agile nation.

The Australian Research Data Commons (ARDC) is well placed to address these data challenges as the only national organisation dedicated to providing Australian researchers with a competitive advantage through data. This competitive advantage accelerates research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets that lead to high-impact benefits locally and internationally.

The ARDC has a history of delivering successful digital infrastructure solutions such as analytical platforms and national data assets that support the Australian research community to solve real-world problems. Success stories like the digital fieldwork app, smart irrigation, Macquarie Perch conservation and the EcoCommons research platform featured in this booklet highlight how we deliver real impact.

Our portfolio of programs also drives broader transformations across the digital research ecosystem by leading culture change on how data is described and stored long term.
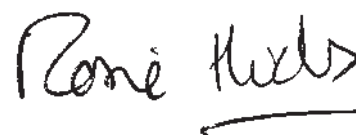
WHERE TO FROM HERE?

The ARDC will become the hub of expertise in how to build integrated national digital research infrastructures that address significant national and international challenges, with a focus on Australia's science and research priorities.

We are applying this new approach to create a suite of national 'Thematic Research Data Commons' for Australia. Thematic Research Data Commons will bring together data and related resources (storage, models and computing infrastructure) to enable researchers from across disciplines and industries to deliver world-class data-intensive research outcomes in strategic thematic areas.

In this booklet, we feature two initiatives launched last year, Bushfire Data Challenges and Health Studies Australian National Data Asset (HeSANDA), both of which demonstrate the way we plan to approach our future activities. We are progressively expanding this work to encompass sectors including Humanities, Arts and Social Sciences, and industries such as agriculture.

We look forward to you joining us on our journey to create a national research data commons for Australia that will accelerate the translation of research into real benefits for Australian society as well as contribute to the future global open science commons.

Rosie Hicks
Chief Executive Officer,
Australian Research Data Commons
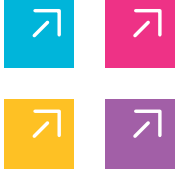
ARDC
SNAPSHOT

ARDC was established under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) in 2019. We have over 70 staff based at host institutions around Australia, and 20 member institutions and growing.

MEMBERS

Australian National University — HOST —

CSIRO — HOST —

Curtin University — HOST —

ECU Edith Cowan University

Federation University

Flinders University

Griffith University Queensland, Australia

Macquarie University Sydney·Australia

Monash University — HOST —

QUT

Swinburne University of Technology

The University of Adelaide — HOST —

The University of Melbourne — HOST —

The University of Queensland Australia — HOST —

The University of Sydney

The University of Western Australia

UNSW Sydney — HOST —

University of Tasmania

UTS — HOST —

University of Wollongong Australia

ARDC
STRATEGY

PURPOSE
To provide Australian researchers with competitive advantage through data.

MISSION
To accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

PEOPLE & POLICY

Connecting the ARDC

Connecting the ARDC to researchers, research institutions, industry and government to enhance knowledge exchange and drive an effective national data skills ecosystem.

COMMUNICATIONS
Communicate our aims, impact and activities.

ENGAGEMENTS
Manage the ARDC's interactions with stakeholders.

SKILLS & WORKFORCE DEVELOPMENT
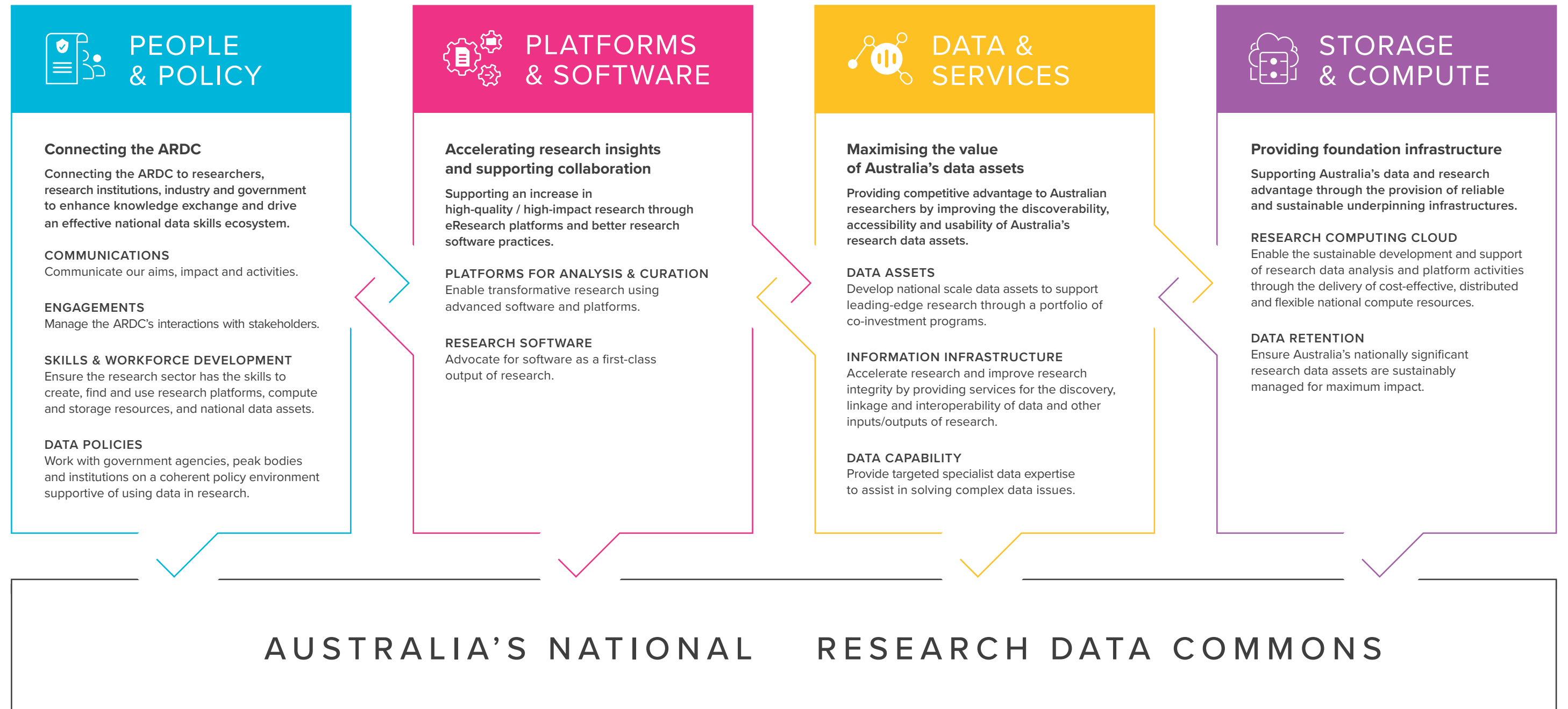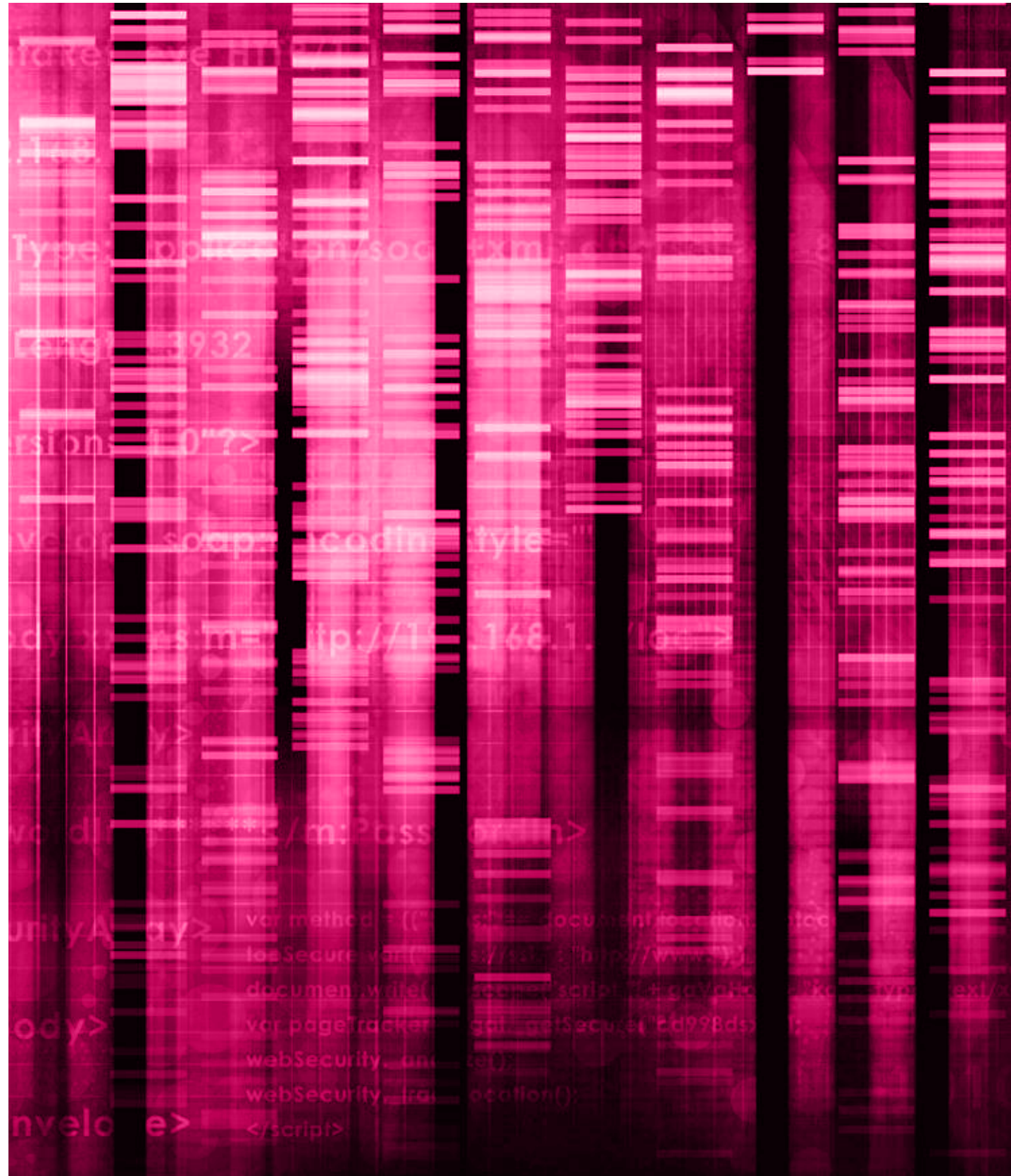Ensure the research sector has the skills to create, find and use research platforms, compute and storage resources, and national data assets.

DATA POLICIES
Work with government agencies, peak bodies and institutions on a coherent policy environment supportive of using data in research.

PLATFORMS & SOFTWARE

Accelerating research insights and supporting collaboration

Supporting an increase in high-quality / high-impact research through eResearch platforms and better research software practices.

PLATFORMS FOR ANALYSIS & CURATION
Enable transformative research using advanced software and platforms.

RESEARCH SOFTWARE
Advocate for software as a first-class output of research.

DATA & SERVICES

Maximising the value of Australia's data assets

Providing competitive advantage to Australian researchers by improving the discoverability, accessibility and usability of Australia's research data assets.

DATA ASSETS
Develop national scale data assets to support leading-edge research through a portfolio of co-investment programs.

INFORMATION INFRASTRUCTURE
Accelerate research and improve research integrity by providing services for the discovery, linkage and interoperability of data and other inputs/outputs of research.

DATA CAPABILITY
Provide targeted specialist data expertise to assist in solving complex data issues.

STORAGE & COMPUTE

Providing foundation infrastructure

Supporting Australia's data and research advantage through the provision of reliable and sustainable underpinning infrastructures.

RESEARCH COMPUTING CLOUD
Enable the sustainable development and support of research data analysis and platform activities through the delivery of cost-effective, distributed and flexible national compute resources.

DATA RETENTION
Ensure Australia's nationally significant research data assets are sustainably managed for maximum impact.

AUSTRALIA'S NATIONAL RESEARCH DATA COMMONS

ARDC
IN NUMBERS

The ARDC delivers data projects and digital research platforms that support the Australian research community in providing tangible outcomes for Australian society.

The following pages showcase the breadth and depth of the services that the ARDC provides the research community, the majority of which are either available to all or to merit-based researchers at no cost to them.
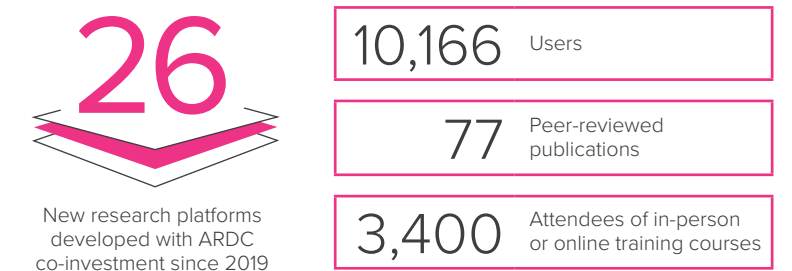
In the period to 30 April 2021, the ARDC attracted greater than 1:1 partner and industry co-investment of Australian Government funding.

PLATFORMS & SOFTWARE

RESEARCH PLATFORMS

Digital research platforms enable and promote world-class collaborative research across institutional and discipline boundaries using diverse data and analytical tools. The ARDC pioneered digital research platforms in Australia and continues to support their development:

26

New research platforms developed with ARDC co-investment since 2019

10,166	Users
77	Peer-reviewed publications
3,400	Attendees of in-person or online training courses

RESEARCH SOFTWARE

The ARDC is at the forefront of developing a national research software agenda to recognise research software as a first-class research output and a research enabler:

400 VIEWS

200 DOWNLOADS

of the draft agenda within a week of launching in June 2021

| 97 | Registrations of interest for the pre-release consultation |
| 51 | Institutions |
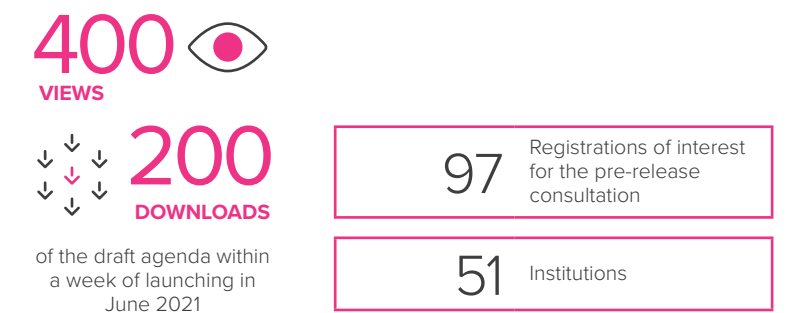
DATA & SERVICES

NATIONAL DATA ASSETS

The ARDC develops national-scale data assets through strategic partnerships with research communities and institutions to accelerate leading-edge research:
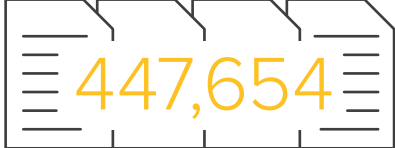
209

Partnerships with universities, medical research institutes, national facilities, and research networks, as well as state and federal government agencies

26 National-scale data assets that aggregate, harmonise and integrate complementary data

IDENTIFIER SERVICES

The ARDC provides persistent identifier services that support Australian researchers with citation of their research outcomes and attribution of the resources and services used during that research:

447,654

Digital Object Identifiers (DOIs) minted since 2011 - uniquely identifying and facilitating access to papers, publications, data collections, and research outputs

394,391 Handles minted since 2010 - uniquely identifying and facilitating access to data collections and other research outputs

RESEARCH DATA AUSTRALIA (RDA)

RDA is Australia's online portal for researchers to find research data and associated projects, researchers, and data services:

ARDC Research Data Australia

49,477

Unique searches per year (average)

179,145 Collections (since 2011)

RESEARCH VOCABULARIES AUSTRALIA (RVA)

The ability to read datasets relies on vocabularies; RVA helps researchers find, access, and reuse vocabularies for their research as well as facilitating collaborations:
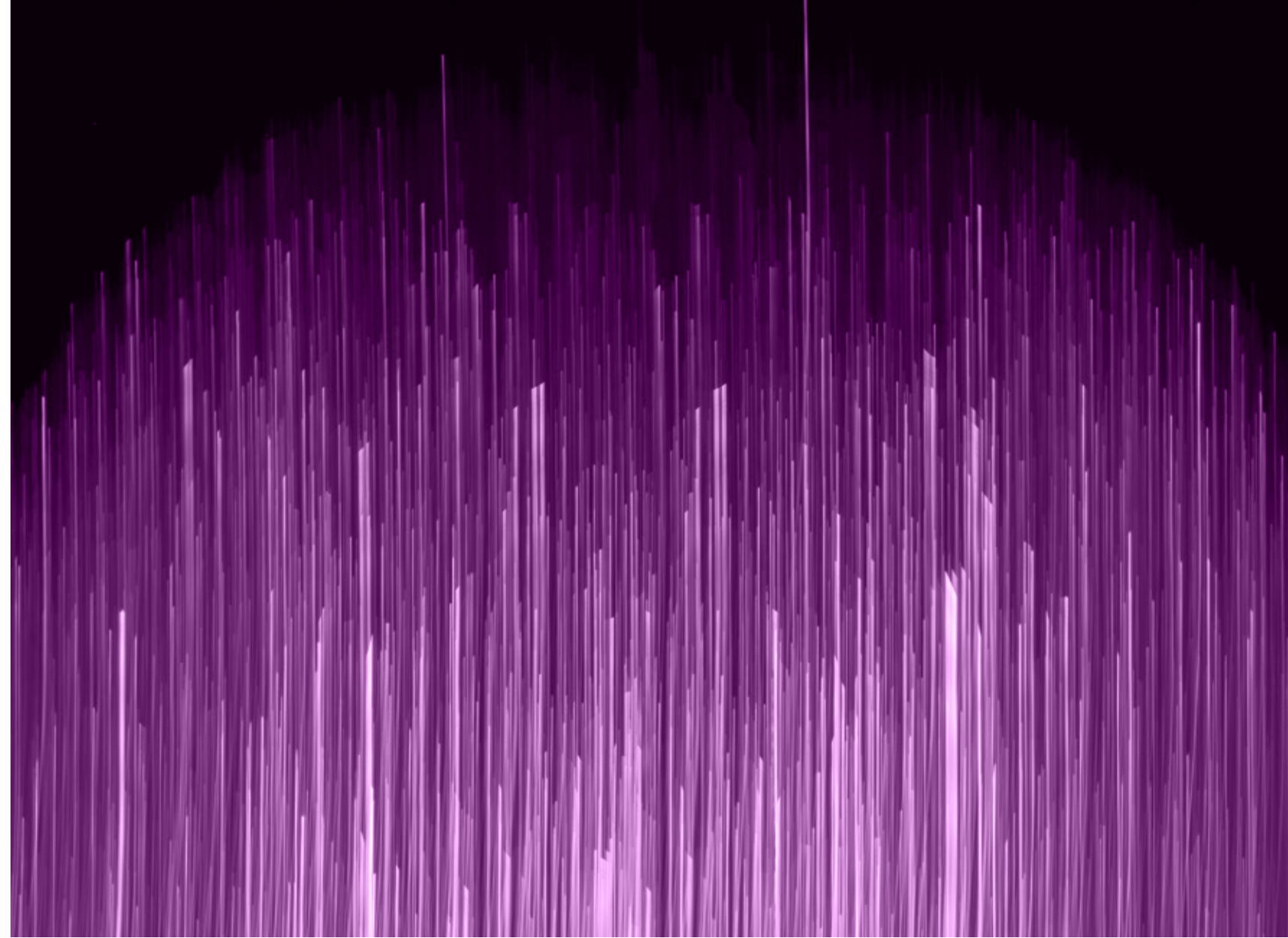
ARDC Research Vocabularies Australia

391

Vocabularies (since 2015)

10,596 Filtered page views per year (average)

STORAGE & COMPUTE

DATA RETENTION

The ARDC partners with the research sector to increase the impact of investment in underpinning infrastructure that store important data collections:

1300

Nationally significant data collections supported
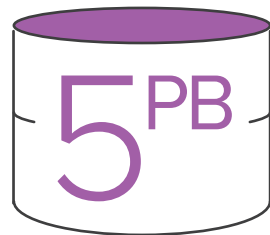
52,000TB of storage capacity across 13 national partners

ARDC NECTAR RESEARCH CLOUD

The ARDC Nectar Research Cloud is Australia's federated national research cloud providing the research community with on-demand computing infrastructure and software:

AUSTRALIA'S NATIONAL RESEARCH CLOUD

5 PB

File Storage

33,000	Physical CPU Cores	8,000	Virtual Machines
150,000+	Virtual CPU Cores	2 PB	Object Storage

SUPPORTING AUSTRALIAN RESEARCH SINCE 2012

37

Australian Universities

11	ARC Centres of Excellence	13	NCRIS Capabilities
4	Publicly Funded Research Agencies	16	Virtual Laboratories
4	Medical Research Institutes	4,000	Projects
13	Cooperative Research Centres	50,000+	Researchers supported

IN 2019/2020 ARDC NECTAR RESEARCH CLOUD SUPPORTED:

1788

Projects

|			
2,921	Users running virtual machines	430	Collaborative multi-institutional projects
344	ARC & NHMRC grants	ALL	Research fields

PEOPLE & POLICY

SKILLS, WORKFORCE DEVELOPMENT AND ENGAGEMENTS

The ARDC helps build the data management capacity of Australian researchers by inspiring and supporting researchers to communicate, collaborate and build knowledge and skills, as well as engaging with institutions and organisations to implement best practice data governance and stewardship policies, tools, and processes:

70
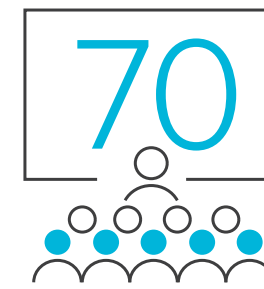
National and international ARDC-run events in 2020

3,300 Attendees

COMMUNITIES OF PRACTICE

The ARDC leads and supports communities of practice and interest groups dedicated to supporting the improvement of digital research infrastructure, skills, and capacity for Australian researchers:

15

Local and international communities of practice and interest groups

2,000 Participants

## INPUT INTO AUSTRALIAN GOVERNMENT CALLS FOR CONSULTATION

The ARDC provides specialist input into state and federal government consultations regarding the treatment and use of data, including:

- Data Access and Transparency Bill
- Department of Prime Minister and Cabinet's Technical Advisory Committee
- OECD Council Recommendation on Access to Research Data from Public Funding
- OECD Council Recommendation on Enhancing Access to and Sharing of Data

ARDC
IMPACT
STORIES

DIRT TO DESKTOP:
DIGITAL TOOLS IN THE FIELD
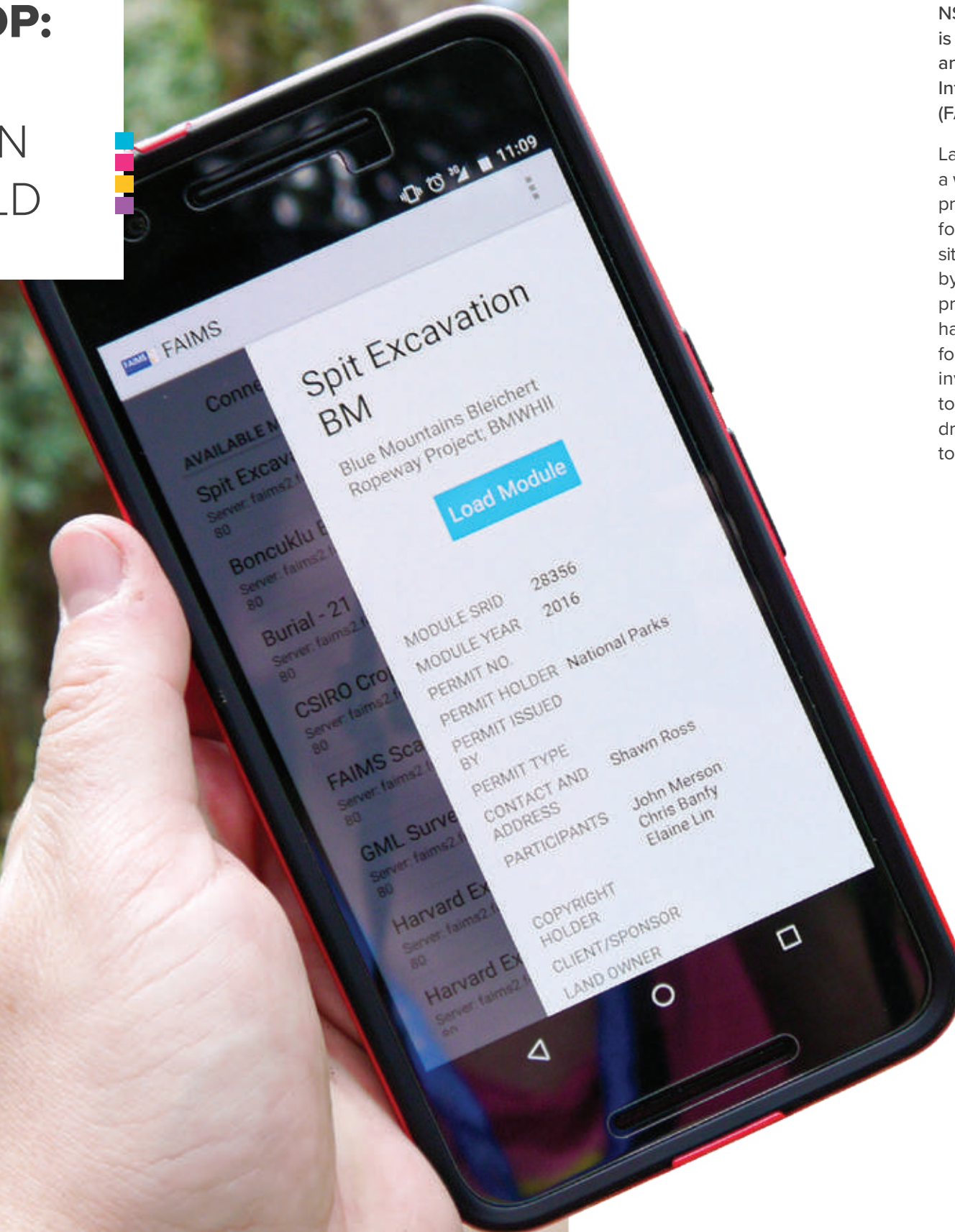
Field research data around the globe, from the Blue Mountains in NSW to remote villages in Cameroon, is being digitally captured, organised and analysed thanks to the Field Acquired Information Management Systems (FAIMS) Mobile platform.
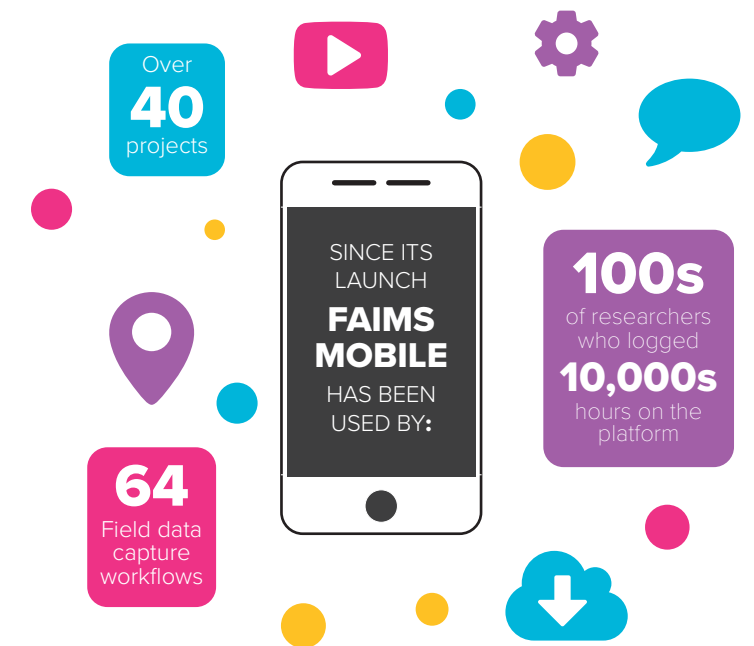
Launched in 2012, FAIMS Mobile is a world-class, open-source platform for producing custom electronic notebooks for data collection in diverse field situations. FAIMS was originally funded by Nectar, one of the three NCRIS projects that became the ARDC, and has seen continuous improvements for almost 10 years. In 2020, the ARDC invested in a major rework of FAIMS to modernise its technology base, and dramatically expand its ability to adapt to many varied field-work settings.

UPGRADING FROM PAPER TO MOBILE

There are a myriad of ways that manual data collection in the field, particularly in remote locations, can go wrong and cause inconsistencies — leading to hours of additional work 'cleaning' the data. FAIMS was developed to address these challenges and simplify field work by creating custom electronic notebooks for data collection.

"In the past you had to transcribe handwritten notes into a spreadsheet, then make the association between the image file and location — it was time consuming and manual," said Dr Jens Klump, Team Leader Geoscience Analytics at CSIRO.

FAIMS has completely changed the way researchers like Dr Klump collect samples, ditching paper for the mobile app to produce richer data sets.

Over **40** projects

64 Field data capture workflows

SINCE ITS LAUNCH **FAIMS MOBILE** HAS BEEN USED BY:

100s of researchers who logged **10,000s** hours on the platform

The platform has been taken up around the globe, including in Australia, Bulgaria, Cameroon, China, Israel, Malawi, Peru and Turkey, by researchers at Yale, Harvard, Brown, Aarhus, La Trobe, Macquarie, UNSW, New England and Queensland universities. It has supported research in geoscience, ecology, archaeology, oral history, ethnography, linguistics and citizen science applications.

Since its initial development in 2012, FAIMS has received support from: Nectar, the Australian Research Council, the NSW Government, several Australian universities and small firms, and the ARDC. ▶

SEARCHING FOR HIDDEN MINERAL DEPOSITS WITH FAIMS

In Australia, FAIMS is speeding up the mineral resources industry's hunt for traces of high-value mineral deposits by halving the time required for geochemical sampling.

"80% of Australia is not well explored for mineral resources," said Dr Jens Klump, Team Leader Geoscience Analytics at CSIRO. "Everything that's easy to find has already been found."

There could be large high-value mineral deposits waiting to be discovered deeper underground.

However, the chance of finding economic mineral deposits is very small, explained Dr Klump: "Finding them is even more difficult than finding a needle in a haystack."

As part of the CSIRO Mineral Resources discovery program, Dr Klump and his colleagues work with independent contractors, state geological surveys and the minerals resources industry to sample thousands of boreholes and soils, looking for anomalies in the geochemistry that might point to economic deposits of high-value mineral resources that warrant further investigation.

With so much of Australia's mineral resources still unknown, FAIMS is reducing the time and money spent on water, vegetation and soil sampling to speed up the quest to find valuable mineral deposits.

"FAIMS has halved the time needed for sampling," said Dr Klump.

> "
> ────────────

FAIMS has halved the time needed for sampling.

Dr Jens Klump
Team Leader
Geoscience Analytics at CSIRO

ACCURATE AND EFFICIENT REMOTE DATA COLLECTING

FAIMS allows data collection offline, automatically synchronising data collected by multiple teams in the field when they connect to a network, eliminating double-entry and automating exports, which saves time and reduces errors.

FAIMS was used in fieldwork for the cross-disciplinary Malawi Earlier-Middle Stone Age Project, which aims to understand changes in human technology, subsistence, and demography across the Middle Stone Age (~315,000-30,000 years ago).

"FAIMS Mobile has been such an incredible advantage in terms of workload, data quality, and other data management issues," said Jessica Thompson, Assistant Professor of Anthropology at Yale University, who used FAIMS on the Malawi project.

"It readily links disparate data types that are otherwise stored separately, such as photographs, tabular logs, and context relationships."

FAIMS 3.0 — FLEXIBLE, CROSS-PLATFORM AND SELF-SERVE

Recognising the potential for the commercialisation of FAIMS the ARDC co-invested in the FAIMS 3.0 project to rebuild FAIMS Mobile to be cross-platform (Android, IOS and desktop), as well as to allow users to customise and rapidly deploy modules themselves through a graphic user interface. An open source FAIMS library will also enable researchers to reuse and adapt applications developed by others.

While FAIMS 3.0 will always have a no-cost pathway for research, there will also be commercial opportunities through the provision of value-added services around customisation. The project team participated in the CSIRO ON Prime pre-accelerator program and a commercialisation program at Macquarie University, which influenced the redesign of the platform. Archaeology, geoscience and environmental science companies, along with the university sector, have already shown interest in the commercial version of FAIMS 3.0. ■

FAIMS 3.0 beta will launch for testing at the beginning of 2022, with the final release in 2023.

PROJECT PARTNERS

MACQUARIE University SYDNEY-AUSTRALIA

ARDC Australian Research Data Commons

aarnet

CSIRO

UNSW SYDNEY

LA TROBE UNIVERSITY

AARHUS UNIVERSITY

12 other local and international institutions

SMART URBAN IRRIGATION:

SAVING WATER & PROTECTING THE GREAT BARRIER REEF

Researchers from Central Queensland University worked alongside Cairns Regional Council and irrigation equipment manufacturer Rain Bird Australia to optimise urban irrigation using artificial intelligence, the ARDC Nectar Research Cloud and the Internet of Things, resulting in significant environmental benefits.

Through the use of smart irrigation, water usage and chemical runoff were significantly reduced. Urban parks in Cairns are now healthier green spaces for the community to enjoy year round. The successful project marked the beginning of further research and commercial opportunities to improve irrigation for high-value sporting fields and agriculture.

IRRIGATING URBAN PARKS IS NOT SIMPLE

Managing water and irrigation in urban parks can be challenging.

Parklands are exposed to many external factors including rainfall, winds, public traffic, and temperature and humidity variations.

Unfortunately, traditional irrigation systems generally operate on pre-set timers and can't respond to actual weather conditions, soil moisture levels or the changing needs of plants.

In 2019, researchers from Central Queensland University (CQU) and members of Cairns Regional Council started the Smart Urban Irrigation Project to address this problem, with funding from the Australian Government Smart Cities Collaboration Platform and Cairns Regional Council.

The project had a number of goals: improve and optimise irrigation in two of the city's parks, reduce water consumption, adapt to the workload of park employees, and ensure parks are healthy and green year round for the community to enjoy. ▶

INTRODUCING PARKS TO THE INTERNET OF THINGS AND CLOUD COMPUTING

The solution involved creating an Internet of Things: with real-time monitoring sensors, software and cloud computing connected and exchanging data over the Internet. Core data is collected from the city's parks, including soil water content, soil properties and plant characteristics. The data is channelled into a computer model designed by the team, which incorporates these and other variables that impact parkland conditions, including weather conditions and park management practices.

The ARDC Nectar Research Cloud played a key role in the project.

"We used Nectar to collect nearly two years' of data from the sensors," said Dr Biplob Ray, the CQU lead researcher on the project.

"We also used it to develop the AI [artificial intelligence] system and to present the data on a dashboard."

The dashboard provides a visual representation of the data and sends alerts to park managers based on real-time events, including extremely low or high moisture content. The AI-powered 'brain' of the system allows it to make accurate automatic decisions for the whole irrigation system, based on actual parkland conditions.

"Nectar has enabled us to store and present this project to end users [park managers]. It's allowed us to create a controlled decision system to run a completely automated system of controlled irrigation," Dr Ray said. "The data we're storing on Nectar is invaluable."

LESS RUNOFF MEANS A HEALTHIER GREAT BARRIER REEF

By deploying the smart irrigation system, the Cairns Regional Council has seen their sprinklers, which usually operate every day, turn off for 9 days every month on average. The project team calculated that the council is saving 583 litres of water per year per square metre of parkland. The smart irrigation system has proved to be a success, and the council is still using the technology to this day.

Along with creating greener parks, saving water and easing the workload of park employees, the reduction in water use and runoff means fewer chemicals are being transported into the Great Barrier Reef.

PATENT FILED AND THE SEARCH IS ON FOR AN AUSTRALIAN MANUFACTURER

The Smart Urban Irrigation Project marked the beginning of more research with enormous potential to optimise irrigation not only in urban parks, but also for agriculture and high-value sporting fields such as cricket pitches and golf courses.

As an extension of the project, the CQU team has filed an Australian patent application (#2021900891) for a new sprinkler system with two-way communication for improved remote decision making for irrigation.

The new technology could help irrigate high-value sporting fields remotely and efficiently. For example, the intelligent sprinkler system would help a curator know remotely whether irrigation has made a cricket pitch better for spin or pace — a significant efficiency improvement compared to the current practice, which relies on a curators expert experience on the ground.

The CQU team is currently working with a Queensland Government manufacturing hub in Rockhampton to find a local manufacturer for the patented technology.

"The project wouldn't be this successful without the support from the ARDC," said Dr Ray. ∎

> "
>
> The project wouldn't be this successful without the support from the ARDC.
>
> **Dr Biplob Ray**
> Lead researcher
> Smart Urban Irrigation Project
> *Central Queensland University*

ARDC SUPPORTING THIS RESEARCH

ARDC Nectar Research Cloud

Easy to scale resources

Cost effective for researchers and their institutions

Local and national expertise and support

Easy to build and host platforms due to standard template

PROJECT PARTNERS

Cairns REGIONAL COUNCIL

CQUniversity AUSTRALIA

RAIN BIRD

Australian Government
Department of Infrastructure, Transport, Regional Development and Communications

THE GENOMIC FIGHT
TO SAVE THE
MACQUARIE PERCH

Researchers measured genetic diversity of endangered Macquarie perch populations in the Murray-Darling Basin to improve the long-term success of conservation programs for the species. The large computational effort required for the genomic analyses was made possible by the ARDC Nectar Research Cloud.

All of the recommendations from the research have been incorporated into the species' National Recovery Plan, and the genetic management approach is now being used to help conserve more of Australia's unique biodiversity.

THE PLIGHT OF THE ENDANGERED MACQUARIE PERCH

The endangered Macquarie perch was once widespread across the Murray-Darling Basin, but human-built river barriers and other human-induced threats have caused it to become locally extinct. Only a few isolated populations remain.

The fish need to swim upstream to breed, laying their eggs onto rocks where they are aerated by riffles in the creeks — an essential process for the eggs to develop. The lowland populations trapped downstream of the barriers have died out.

MANAGING FOR GENETIC DIVERSITY, NOT JUST ABUNDANCE

In the past, the recovery plan might have been to stock a shrinking population with hatchery-bred offspring from a single population, after mitigating the environmental threats.

Now, genomic testing conducted by a team of researchers working on an ARC Linkage project — which included the sequencing of the entire Macquarie perch genome — has found that isolated populations have a low chance of survival on their own and fish managers should be mixing them up.

"We don't know yet whether they have adapted in different ways," said molecular ecologist Dr Alexandra Pavlova, Monash University, who co-leads the project with Professor Paul Sunnucks. "But we know that the genetic diversity of most populations is so low that they might be suffering from inbreeding and loss of fitness as a result. Mixing them up would reverse this decline."

Inbreeding can lead to an unhealthy population, which can lead to extinction. And while managing for genetic diversity as well as for abundance of the fish is a relatively new concept, all of the project's recommendations have been incorporated into the species' National Recovery Plan. ▶
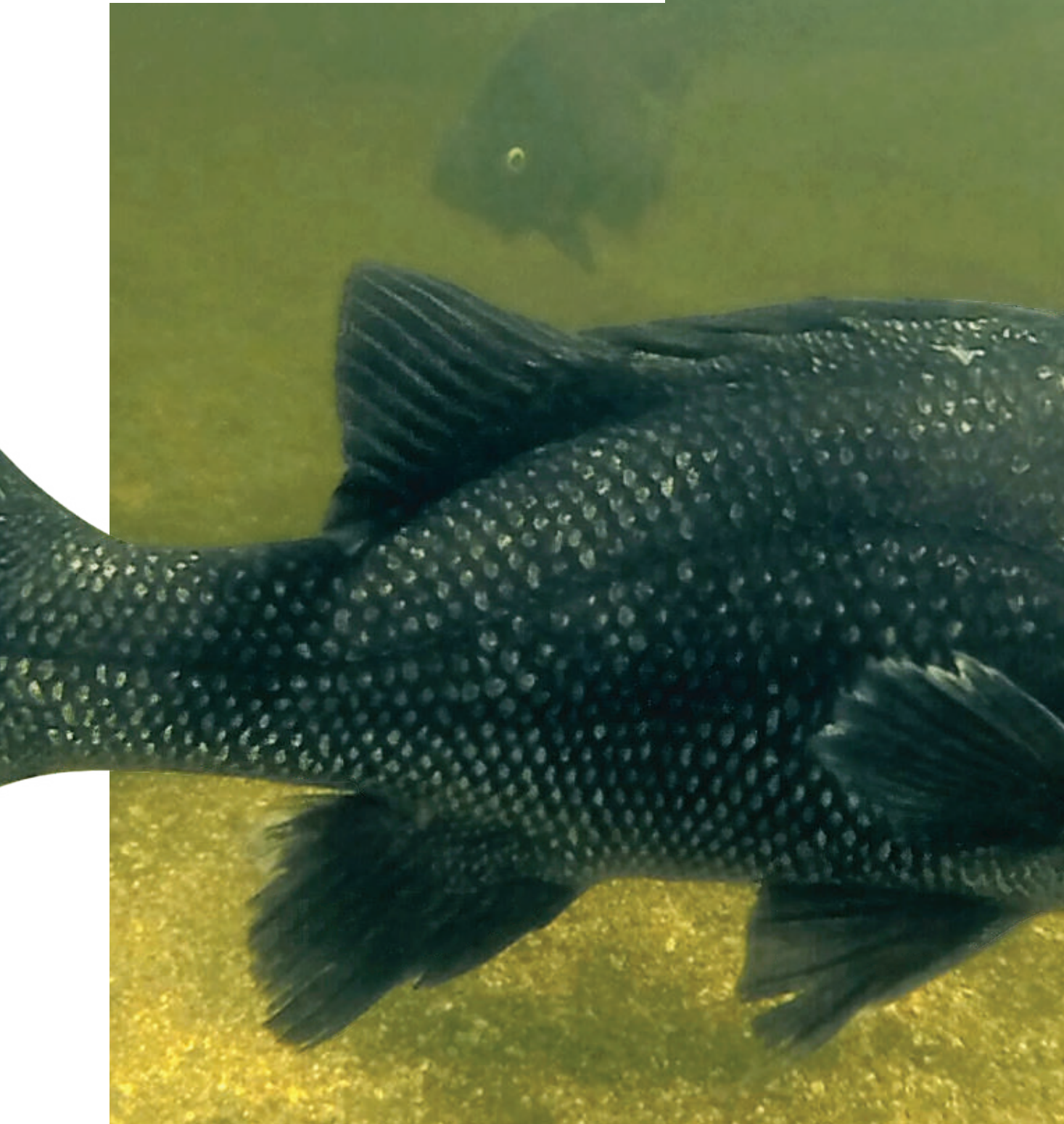
IMAGE — ZEB TONKIN

"

I think that Nectar is brilliant. We are grateful to those people who invented it and made it available to our lab.

Dr Alexandra Pavlova
Molecular Ecologist
Monash University

WORKING REMOTELY ON NECTAR BEFORE AND DURING THE COVID-19 CRISIS

Since the advent of COVID-19, being able to work remotely and internationally on the Nectar Research Cloud has become more important than ever.

"The servers are accessible from anywhere, so it is great for collaboration. And our PhD students, including one in Mexico and one in Singapore, are able to work from their homes. The big machine runs the jobs, but the interface is actually on your laptop, so we can easily communicate, which is really convenient."

"I think that Nectar is brilliant. We are grateful to those people who invented it and made it available to our lab," said Dr Pavlova.

FISH ARE NOT THE ONLY BENEFICIARIES

Fish managers in the Murray-Darling Basin have embraced the findings of the genomic testing, incorporating all of the lab's recommendations for Macquarie perch into the species' National Recovery Plan.

Other species also stand to benefit. The team has already sequenced the genome of the Golden perch, as well as the iconic Murray cod. Genomic testing is underway for the endangered Trout cod, the critically endangered Helmeted Honeyeater and the critically endangered Leadbeater's Possum. ▪

THE ARDC NECTAR RESEARCH CLOUD PROVIDES THE POWER

Even for a fish, genome sequencing needs serious compute grunt. Dr Pavlova was able to secure an allocation of computational resources on the ARDC Nectar Research Cloud.

"We purchased our own hardware in the beginning but it started to break down and needed a lot of maintenance," explained Dr Pavlova.

"The allocation on Nectar was a huge help for our lab. Next-generation sequencing yields huge data files, even when compressed. You have to use Linux machines and command line interface to process the data. And that's what Nectar provided — Linux machines with huge computing power and storage capacity. Most of our analysis was done on these machines, which is invaluable."

Monash eResearch Centre assists with managing the virtual servers on Nectar. Most of the storage capacity for the project is supported by the ARDC's Data Retention project.

GENOMIC TESTING — IT'S A LONG PROCESS

Genomes are measured by the number of 'bases', the chemical components of DNA and RNA. The Macquarie perch came in at 675 million bases. Even with the power of the ARDC Nectar Research Cloud, analysing genomic data takes a staggering amount of time.

"The latest analysis we are running is genotyping whole-genome resequencing data for 75 individuals. It's already taken two months and it's not even half finished. That kind of gives you an idea," said Dr Pavlova.

Genome assembly (the computational process) and annotation (listing the genes) takes weeks. "It's a long process. Basically, you write a piece of code, submit it on the computer; if the computer doesn't flick an error, you wait and wait and in a few weeks you may get the happy answer. Sometimes you see that you've made an error or your search parameters were not good enough, so you change them and start over."

To get a good quality genome for one individual fish, they sequenced the genome more than 350 times. Then, to understand variation in the Macquarie perch populations, they sequenced 100 fish more than 15 times.

ARDC SERVICES SUPPORTING THIS RESEARCH

Cost effective for researchers and their institutions

Easy to scale

Simple to collaborate across institutions

ARDC Research Data Australia
The dataset produced by this research can be searched for and accessed by other researchers through ARDC Research Data Australia.

Data Retention Project
The datasets produced by this research are being stored with the support of the ARDC Data Retention Project.

Persistent Identifiers Service
This dataset has a Digital Object Identifier (DOI) used to identify, cite and track research impact. The DOI was minted by Monash University, which use the ARDC DOI service. We provide this service free of charge to the Australian research sector.

IMPROVING ENVIRONMENTAL DECISION MAKING
WITH RESEARCH PLATFORMS

Over 7,000 users based at 400 different institutions in 35 countries have accessed ARDC-supported digital research platforms for ecological modelling to inform environmental research and decision making.

AUSTRALIA'S ENVIRONMENT IS IN A STATE OF DECLINE

In a recent independent review of Australia's environmental legislation, the Environment Protection and Biodiversity Conservation Act 1999, Professor Graeme Samuel AC wrote:

> "Australia's natural environment and iconic places are in an overall state of decline and are under increasing threat. The pressures on the environment are significant – including land-use change, habitat loss and degradation, and feral animals and invasive plant species. The impact of climate change on the environment will exacerbate pressures and contribute to further decline." [1]

The ability to make good decisions about our environment to reverse this decline relies on sound ecological modelling and good biodiversity data. It is a powerful tool, but ecological modelling is not simple to do.

Dr Elisa Bayraktarov, EcoCommons Program Manager and ecologist, explains the dilemmas faced by ecosystem modellers, using an example of predicting the future distribution of a species:

"Imagine there's a conservation problem you need to solve. First, you have to access all the available data on the species occurrences.

"Now, you have to prepare your data. You have to filter it for the species you need and the area of interest, and you have to get rid of data that might be biased or that has simple errors, and ensure all the metadata is consistent.

"Then you need to choose the right Species Distribution Model Algorithm out of the 25+ that are published in the scientific literature.

"You then have to evaluate your model and make sure that your data is being visualised properly."

And finding the current species distribution is only the first step. Then it's onto modelling how species will distribute in the future considering the impact of climate change, which requires even more datasets on emissions scenarios, climate and more.

In the end, ecological model results require careful application of the latest science, but these modelling steps can take days, weeks or even months for the most experienced coder depending on the model and data.

That's where a digital research platform steps in.

"The EcoCommons Australia program is building a research platform to bring all the fantastic data that already exists into one place, connect it with published methods, tools and analytical workflows, and back this all up with high performance computation and cloud storage.

"And by doing this in one connected platform, running scientific workflows becomes much more rapid, transparent, reproducible and streamlined," explained Dr Bayraktarov.

▶

1   (Samuel, G 2020)

BUILDING ON 6 YEARS OF EXPERIENCE

The ARDC is supporting the development of the EcoCommons platform — a $5 million investment to create a world-leading ecological and environmental modelling tool to support environmental problem solving. Building on 6 years of research and experience, the platform will become a superhighway for users; shortening the time from question, to answer, to environmental decision.

The design of the underlying technology will make the platform more easily deployable in different disciplines. For instance, EcoCommons is being used as the basis for a BioSecurity Commons, in development with the Queensland Government.

EcoCommons is a partnership of nine organisations:

It also involves investment from the Queensland Government's Research Infrastructure Co-investment Fund (RICF).

EcoCommons will give Australian decision makers, practitioners, researchers and students access to trusted ecological and environmental data sets and modelling tools. It will also provide ready-to-use educational materials.

For example, EcoCommons will enable a user to investigate the habitat available for the critically endangered Southern Corroboree Frog in Australia in a scenario where the world is 1 degree warmer. The platform will use data already collected and stored in its database to plug into a trusted ecological distribution model and provide actionable insights into protecting Australia's biodiversity.

> **"**
>
> The BCCVL takes out nearly all the technical drudgery users are commonly faced with when running species distribution models. With its easy to use interface, accessible from anywhere, it opens the field to a whole new array of researchers who understand the systems they are working on, but do not have the technical skill-sets or hardware to properly answer their questions.
>
> Professor Shawn Laffan
> *University of NSW*

EcoCommons is building upon and integrating services that are still popular with researchers and students today, including:

- **Biodiversity and Climate Change Virtual Laboratory (BCCVL)** — species distribution modelling and climate change impact

- **EcoCloud** — cloud based computing for researchers to address any ecological question with their own code

- **EcoEd** — an ecoscience training and skills development initiative

- **Collaborative Species Distribution Model (CSDM) Program** — a pilot program associated with BCCVL that produces species distribution models to support decision making by state and federal government.

IMPACT IN AUSTRALIA AND AROUND THE GLOBE

Two of EcoCommons' offerings, BCCVL and EcoCloud, received significant start-up support from the ARDC's predecessors (ANDS, Nectar and RDS) and are currently hosted on the ARDC Nectar Research Cloud. They are used by 1,000s of decision makers and researchers, catering to a range of experience levels from undergraduates and PhD candidates through to experienced researchers.

When Casey Kirchhoff, PhD candidate at the University of New South Wales, wanted to study the expected impact of global warming on a selected species of alpine wildflowers, she turned to the BCCVL. Ms Kirchhoff used the BCCVL to assess the future prospects of these species under different climate change scenarios, finding that the climatically viable habitat of all 19 plant species is significantly shrinking, severely limiting survival potential by 2055.

"The BCCVL was not only useful for conducting my research, it is also a great learning tool for teaching second year students," Ms Kirchhoff said.

Internationally, researchers conducted ecological niche modelling on the BCCVL to help understand the factors leading to Yellow Fever outbreaks in Brazil.[2] The BCCVL was used to model where humans and monkeys, the hosts and vectors of the disease, would live in the future due to climate change. These models predicted at-risk areas for future outbreaks of Yellow Fever and informed public health authorities on where to focus vaccination efforts.

"The BCCVL takes out nearly all the technical drudgery users are commonly faced with when running species distribution models. With its easy to use interface, accessible from anywhere, it opens the field to a whole new array of researchers who understand the systems they are working on, but do not have the technical skill-sets or hardware to properly answer their questions." Professor Shawn Laffan, University of NSW.

The ARDC is proud to support the evolution of EcoCommons, taking the learnings and successes of initiatives such as BCCVL and EcoCloud to develop a trusted national data infrastructure for the ecosciences.

EcoCommons will be launched at the Ecological Society of Australia conference in late 2022, opening for user testing as early as the end of September 2021. ■

2   (de Thoisy et al. 2020)

TACKLING
AUSTRALIA'S BIG
CHALLENGES

HEALTH STUDIES AUSTRALIAN NATIONAL DATA ASSET (HeSANDA)

In partnership with the health research community, the ARDC is establishing national infrastructure to enable researchers to access and share data from health studies, with an initial focus on clinical trials.

The initiative aims to stimulate new data-driven research ideas, increase the impact of health research and ultimately improve the health and wellbeing of Australians.

THE POTENTIAL OF SHARING HEALTH RESEARCH STUDY DATA

A wealth of data is created through health research studies, including information about the people taking part in the research, their health and their response to interventions being studied.

The data collected in one study is extremely valuable to other studies. However, the challenges of patient privacy and the naturally siloed approaches of research groups and state jurisdictions have created a confusing array of inadequate options and cottage-industry approaches to data sharing. The ARDC is playing a critical role in partnership with the health research community to synchronise efforts, align approaches and build national data capability.

This untapped trove of health research data will bring immense value to health research, maximising the return on investment of past research and allowing future research to build upon it to improve health outcomes for Australians.

Similar capability has already been established in European health research infrastructure (ECRIN and EOSC Life) and data sharing platforms are emerging in the UK and USA. Australia risks being left behind without a corresponding capability.

COLLABORATING TO MAXIMISE THE VALUE OF DATA

The Advisory Committee for the initiative includes key national health research organisations: ▶

Australian Clinical Trials Alliance

AHRA
AUSTRALIAN HEALTH RESEARCH ALLIANCE

ANZCTR
Australian New Zealand Clinical Trials Registry

CHF
Consumers Health Forum OF Australia

Cochrane Australia

Australian Government
National Health and Medical Research Council

NHMRC

PHRN
Population Health Research Network

RESEARCH AUSTRALIA
CONNECTING ● ENGAGING ● INFLUENCING

> ARDC has taken a highly collaborative approach to the HeSANDA initiative and we are glad to both represent and help capture the voice of the investigator-led clinical trials sector. ACTA supports the ethical, feasible, and valid sharing of clinical trials data for use in further research, toward better health through best evidence.

Professor Steve Webb
Chair *of the Australian Clinical Trials Alliance (ACTA)*

SYSTEMATIC CONSULTATION TO BUILD CONSENSUS

Informed by the NCRIS facilitation process, the HeSANDA initiative implemented an 18-month systematic consultation and co-design process with assistance from the Australian Institute of Health and Welfare (AIHW), clinical trialists, researchers, health consumers and research institutions, as well as infrastructure providers and policy makers involved in clinical trials research. This established consensus around the purpose, content and requirements for a national health data asset to be created through the HeSANDA initiative, summarised in Figure 1:
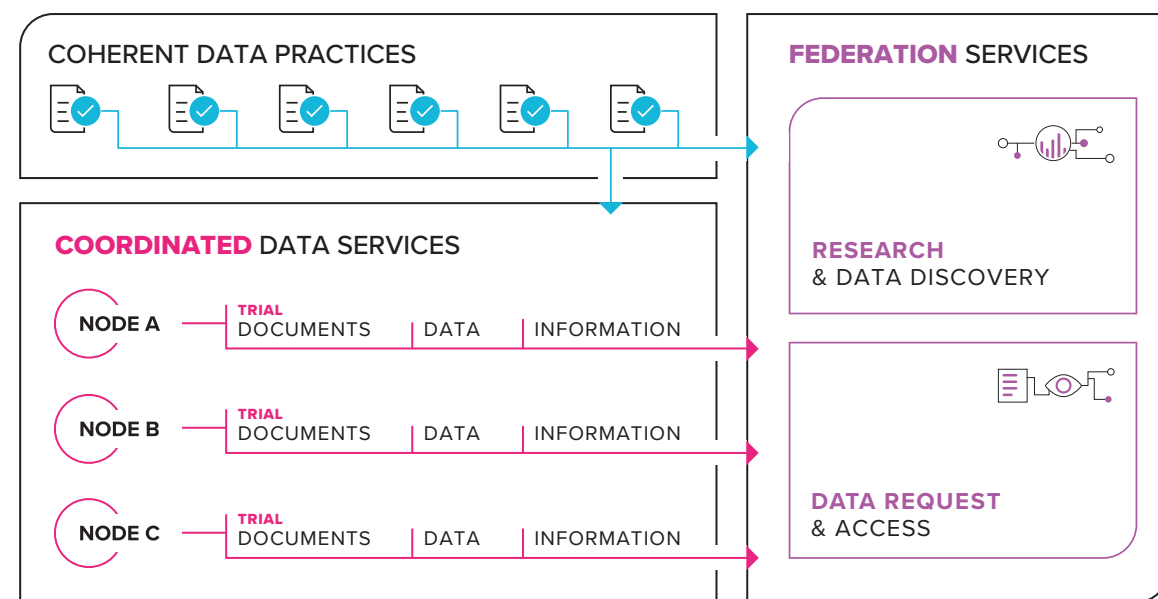
HESANDA INFRASTRUCTURE MODEL

Figure 1. HeSANDA infrastructure model

> The Australian Health Research Alliance (AHRA) across our ten NHMRC accredited Centres, integrates healthcare, research and education to deliver better health. Our priorities include improving data standards, governance, access and use for research and healthcare improvement. We strongly support, and are partnering to deliver, the aims of HeSANDA, with nodes across Australia to enhance collaboration, knowledge and technology exchange, capability building and delivery of impact through better use of data.

Professor Helena Teede
MBBS PhD FRACP FAAHMS FRANZCOG
Chair *of AHRA's Data-Driven Healthcare Improvement Committee* | Executive Director *Monash Partners Academic Health Science Centre* | Professor of Women's Health and Director *Monash Centre for Health Research and Implementation, School Public Health, Monash University* | Chair *of the National Women's Health Research, Translation and Impact Network* | Lead NHMRC Centre's Research Excellence on Women's Health in Reproductive Life and on PCOS, co-lead on CRE in Healthy Preconception, Pregnancy and Postpartum | Endocrinologist *Monash Health*

CREATING THE NATIONAL DATA ASSET

In June 2021, the HeSANDA initiative hit a significant milestone with confirmation of 9 initial nodes of a nationally distributed network of health data sharing infrastructure.

These 9 initial nodes confirm a broad national participation covering 72 health research organisations including:

- 18 universities, 10 medical research institutes, 19 health service operators, 16 clinical trial networks and 9 other organisations
- All Australian states and territories with the exception of Tasmania and the ACT
- Considerable depth in several specialised fields of health research:
    - The National Cooperative Cancer Trials Group node represents the 14 major cancer clinical trials networks in Australia
    - The Mental Health node includes nascent clinical trials networks spanning major universities, medical research institutes and health services conducting research into adult, child and youth mental health
    - The Northern Australian node provides national leadership in Aboriginal and Torres Strait Islander health research and international leadership in tropical medicine and disease research.

The infrastructure build phase of this 9-node network commences in Q3 2021 with coordinated design, implementation and testing. The initial rollout of the data asset will contain clinical trials data generated from >$56 million of public investment from the NHMRC and MRFF.

TRANSLATING INFRASTRUCTURE INTO BENEFIT

The broad alliance of partners in the HeSANDA initiative are committed to translational research. Together they have a demonstrated track record of translating health research into positive changes to:

- Patient health apps (e.g. for diabetes management)
- Australian health guidelines (e.g. pregnancy guidelines to reduce the risk of prematurity)
- World Health Organisation guidelines (e.g. WHO malaria, cancer and essential medicines guidelines)
- Decision-support on the clinical floor (e.g. evaluating AI support for clinicians).

The new HeSANDA data sharing capability put in place by these node partners is designed to accelerate and transform such activities by providing a robust and coordinated national infrastructure.

FORWARD ROADMAP

Phase 1 of the HeSANDA initiative has 9 nodes focused on clinical trials; future phases of the initiative are designed to include more nodes and support high value data from other health study types such as clinical registries or cohort studies.

A broader health-themed data commons would also tackle health data standardisation and access to public and private sector data sources. Connections to other life science capabilities such as genomics or bio-banks would build more extensive capacity, similar to the European EOSC Life cluster.

BUSHFIRE
DATA CHALLENGES

The ARDC is supporting the development of national digital infrastructure to overcome the data challenges associated with bushfire research and response in Australia.

SHARING DATA IS CRITICAL FOR PREPARING FOR AND MANAGING BUSHFIRES

The devastating Australian bushfires in 2019/2020 left 1,000s homeless, burnt over 24 million hectares of land, claimed the lives of 33 people and over a billion animals, incurred financial impacts of over $10 billion, and significantly impacted health, infrastructure, industry and small businesses.

It will take a long time to recover from such an enormous disaster. The Royal Commission into National Natural Disaster Arrangements (2020) recommended supporting better decision making through the national sharing of data.

Data sharing is a key challenge for the numerous Australian research agencies, federal, state and local governments, and emergency management organisations involved in preparing for and managing bushfires. Data is being captured in different formats, with different metadata, and by different levels of government in the states and territories. The ability to harness the full power of data by sharing it across organisations and borders will have a significant impact on how Australia prepares for the next fire season and responds to it effectively.

The ARDC has unique breadth and depth of expertise in facilitating collaboration across organisations to create national data capabilities. In partnership with the bushfire community, the ARDC is investing in developing innovative digital infrastructure solutions with the aim of improving Australia's bushfire resilience, response and recovery.

▶

IMAGE — MATT-PALMER - BMDR2JJXHC / UNSPLASH

CONSULTATION AND COLLABORATION — THE KEY TO A NATIONAL APPROACH

For the Bushfire Data Challenges program, the ARDC conducted 9 months of targeted consultations, including a series of facilitation meetings and participatory design workshops with 92 stakeholders. This process captured priority areas for national digital research infrastructure.

The ARDC has convened key national stakeholders from the research and public sectors to design a national data capability that spans research, development, operational planning, response and recovery, as shown in Figure 1.
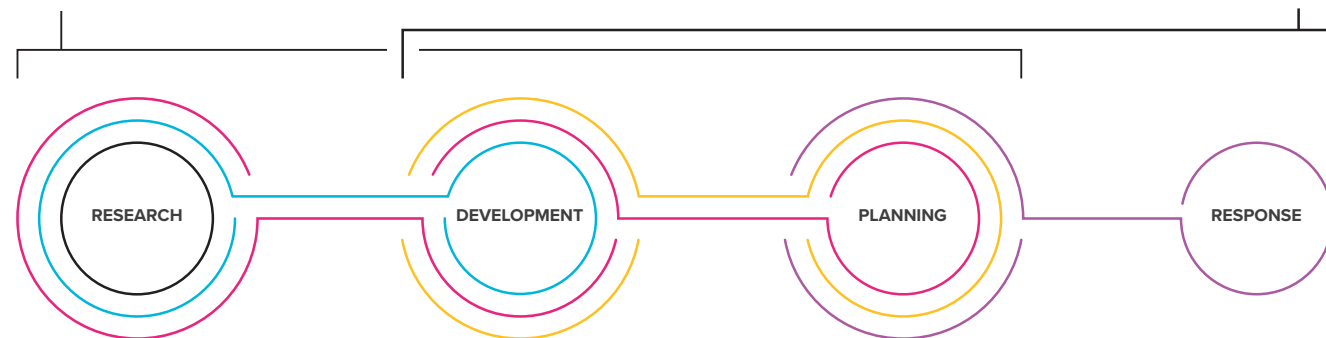
"

The ARDC and the project teams have done an excellent job of understanding the fire industry's needs and these projects could solve some long-standing data sharing and integration problems. The new data sets and capability created will set Australia up for improved and more consistent fire management and prediction in the future.

Stuart Matthews
Senior Project Officer, *NSW Rural Fire Service*

OPERATIONS & INFRASTRUCTURE SCOPE

RESEARCH INFRASTRUCTURE
EG. ARDC BUSHFIRE DATA COMMONS

OPERATIONAL INFRASTRUCTURE
EG. NATIONAL BUSHFIRE INTELLIGENCE CAPABILITY (NBIC)

RESEARCH — DEVELOPMENT — PLANNING — RESPONSE

NHRA | Universities

NHRA: National Hazards Research Australia
Universities

NESP 2 | NCRIS | CSIRO

NESP 2: National Environmental Science Program 2
NCRIS: National Collaborative Research Infrastructure Strategy
CSIRO: Commonwealth Scientific and Industrial Research Organisation

GA | BOM

GA: Geoscience Australia
BOM: Bureau of Meteorology

AFAC

AFAC: Australasian Fire and Emergency Service Authorities Council

State & Australian Government Departments and Agencies

EMA: Emergency Management Australia
NRRA: National Recovery and Resilience Agency

Figure 1. Bushfire data challenges operations and infrastructure scope

PARTNERS FOR PHASE 1: MODELLING BUSHFIRE BEHAVIOUR

Australian Government Geoscience Australia | EMSINA | tern Ecosystem Research Infrastructure | Australian Government Department of Agriculture, Water and the Environment | Australian National University | CSIRO | afac

---

BUSHFIRE MODELLING INFRASTRUCTURE FRAMEWORK

FUEL · ENVIRONMENT · HEALTH · AIR QUALITY · FIRE HISTORY · BUILT INFRASTRUCTURE · DATA FRAMEWORK · PLATFORM · IMPACT MODELLING · FIRE BEHAVIOUR MODELLING

BUSHFIRE DATA COMMONS
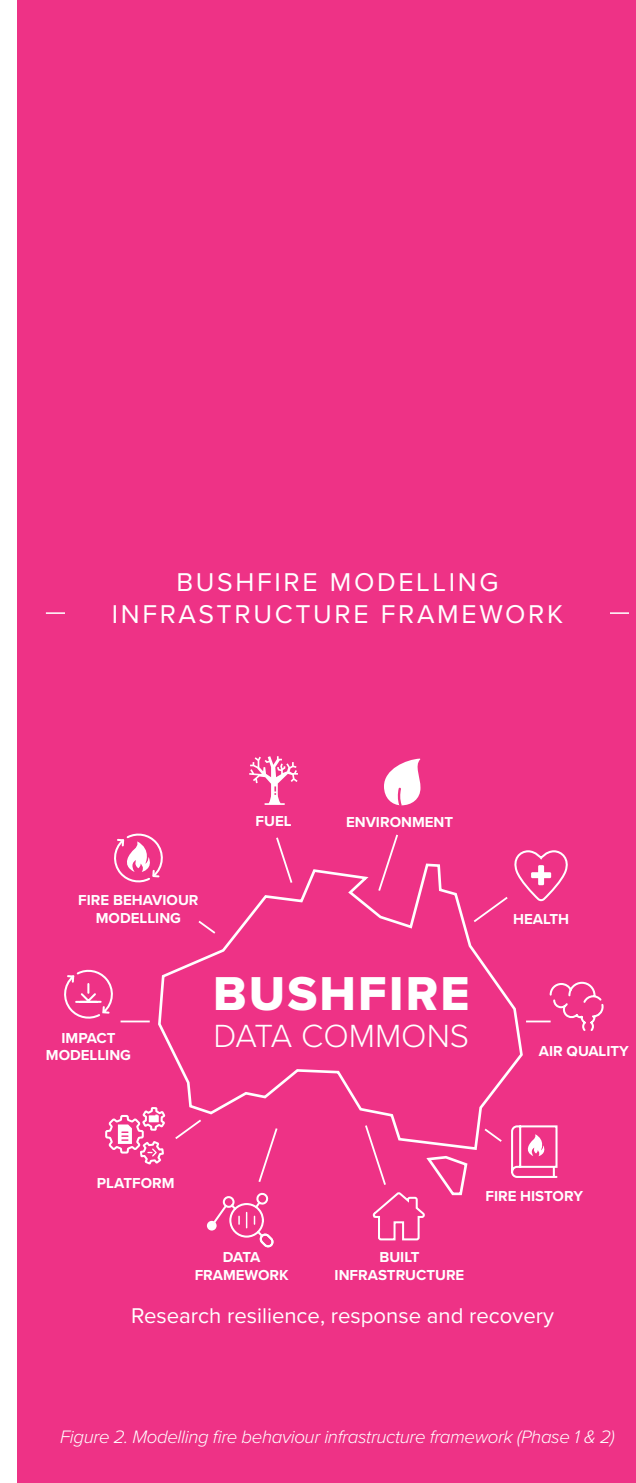
Research resilience, response and recovery

Figure 2. Modelling fire behaviour infrastructure framework (Phase 1 & 2)

UNDERSTANDING FIRE BEHAVIOUR AND IMPACT

The consultation process identified two key areas of national digital research infrastructure focus, which will happen in two phases:

Phase 1: Modelling bushfire behaviour: nationally aggregating data and models on bushfire history and fuel.

Phase 2: Modelling bushfire impact: nationally aggregating data and models on the impact of bushfires on physical & mental health and the built & natural environment.

Data governance will also be addressed to ensure data is collected, used and communicated according to an agreed framework and structure.

DIGITAL INFRASTRUCTURE SOLUTIONS

The ARDC Bushfire Data Challenges national digital research infrastructure will benefit:

Researchers
Researchers will have access to high-quality data assets, which will lead to better models for fire behaviour, enabling risk reduction and more accurate predictions for fire spread.

Operational Organisations
Improved models empower operational and planning organisations to prepare and manage fires more effectively, with the right data and predictions at their fingertips.

Flora and Fauna
Integrating bushfire data improves our understanding of the impact fires have on flora and fauna. It enables researchers to pinpoint species at risk of extinction due to habitat loss from extreme fires and recommend appropriate action.

Health
Improved data will help us to understand the impact of bushfires on physical and mental health to ensure Australia is best prepared to support the community during and after natural disasters.

Infrastructure and Urban Planning
A wide variety of data inputs, when synthesised, provide insights into the impact of bushfires on infrastructure, urban planning and livability.

Phase 1 and phase 2 of the Bushfire Data Challenges program will be completed by June 2023. ■

The Bushfire Data Challenges program is part of the ARDC's Translational Research Data Challenges initiative, providing innovative and high-impact digital infrastructure solutions for real-world problems. It focuses on translational research, operating at the intersection of research and broader society, to provide tangible and enduring economic, environmental and social benefits to Australia.

DESIGNED FOR THE FUTURE —
THEMATIC RESEARCH DATA COMMONS

The ARDC is developing a suite of Thematic Research Data Commons that scale up digital research infrastructure to meet Australia's future research needs.
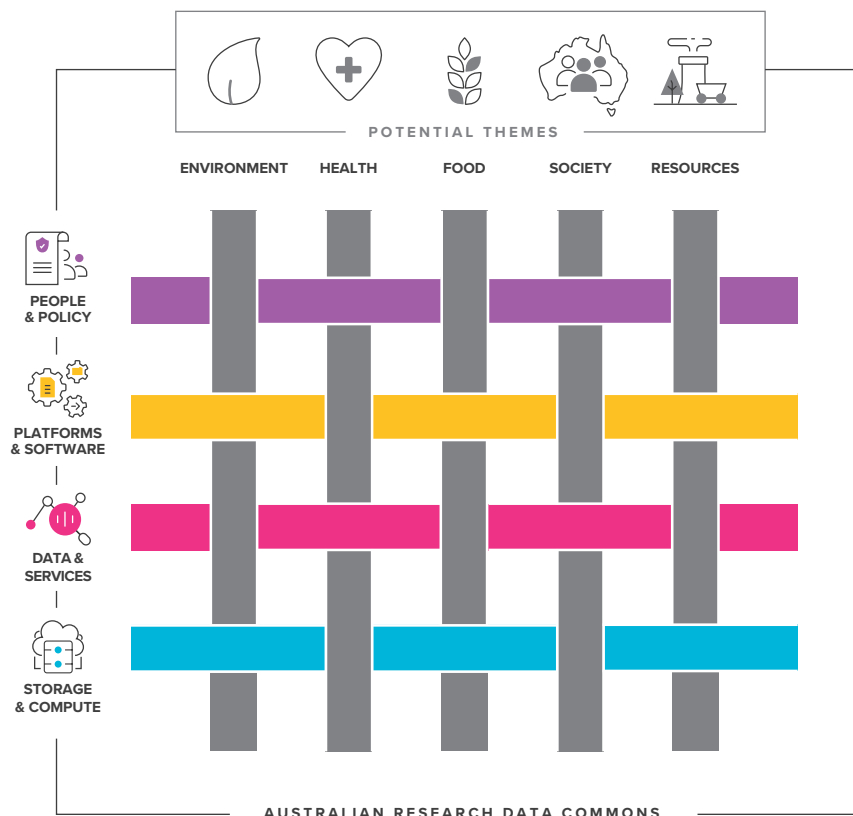
THE OPPORTUNITY

Over the past few decades we have seen an explosion of new technologies to support all aspects of research, from producing and analysing data, through to integrating datasets to solve new questions.

As we adopt new technologies, what was once innovative digital research becomes mainstream, yet the need to continually innovate is undiminished. New digital research infrastructure is being driven by the need to collaborate on big global challenges, the exponential increase in data production, new instruments such as the Square Kilometre Array that pose entirely new challenges in data management and visualisation, and the rise of open science that reaches out to include the citizens who will ultimately benefit from the discoveries being made.

We now have the opportunity to evolve Australia's digital research infrastructure and consolidate around national scale initiatives to meet escalating demand, address data challenges, and enhance collaboration to maximise the benefit for Australians and our return on investment. ▶

THEMATIC RESEARCH DATA COMMONS

POTENTIAL THEMES

ENVIRONMENT HEALTH FOOD SOCIETY RESOURCES

PEOPLE & POLICY

PLATFORMS & SOFTWARE

DATA & SERVICES

STORAGE & COMPUTE

AUSTRALIAN RESEARCH DATA COMMONS

THEMATIC RESEARCH DATA COMMONS

The ARDC is expanding on its experience creating digital research infrastructure over the past decade, along with its hub of expertise and established relationships with research institutions, to create a suite of Thematic Research Data Commons for Australia.

A data commons can be defined as the bringing together of "data with cloud computing infrastructure and commonly used software services, tools and applications for managing, analysing and sharing data to create an interoperable resource for a research community" (Grossman, 2018).

The new ARDC Thematic Research Data Commons (RDC) will develop national-scale data assets, digital tools and platforms within a thematic area to **address Australia's science and research priorities**. The Thematic RDCs will be supported by existing ARDC underpinning infrastructure: cloud compute, data retention, our hub of expertise, and training and outreach activities. Our capabilities will be applied many times over to the Thematic RDCs, resulting in significant optimisation and driving maximum return on investment.

Strategically selecting thematic areas in which to build larger, coordinated and optimised initiatives will enable

the ARDC to meet the needs of the greatest number of researchers and ensure Australian researchers are best positioned to collaborate both domestically and internationally. The Thematic RDC model will also enhance interoperability within and across domains, as well as **enable better collaboration** with international partners, such as the European Open Science Cloud.

The suite of Thematic RDCs will **build on existing digital research platforms and incubate new emerging areas**. For example, building on the HeSANDA, Bushfires and EcoCommons initiatives, mentioned in this booklet, will deliver a broader health and environment themed RDC that will tackle data standardisation and access to public and private sector data sources.

A core goal of thematic RDCs is to **support research translation and commercialisation** by coordinating research activities within a strategic priority area, providing the necessary foundational digital infrastructure to facilitate industry partnerships.

While the resources in a Thematic RDC will evolve to meet the changing needs of researchers, the ongoing commitment to sustainability obtained by embedding the RDC in a national research infrastructure capability will prevent the loss of valuable resources. The enduring nature of an RDC provides organisational resilience to enable a rapid response to crises such as floods, bushfires and pandemics, as well as a meaningful commitment to data retention. ∎

HASS & INDIGENOUS RESEARCH DATA COMMONS

The ARDC is leading the development of a national Humanities, Arts and Social Sciences (HASS) Research Data Commons, which will act as a pilot for the Thematic Research Data Commons model and could contribute to a future Thematic RDC.

The program is supported over the next 2 years by an $8.9m investment by the Australian Government to build national HASS digital research infrastructure that will create new tools and platforms to extend researcher capacity. The investment represents the initial stage of a proposed long-term program to address the challenges and opportunities presented by a diverse grouping of research fields.

THE 3 STREAMS OF THE HASS RESEARCH DATA COMMONS

The HASS RDC comprises 3 streams of activity, which complement the Indigenous Research Capability program.
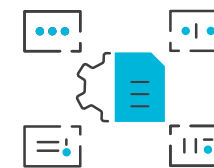

1 Developing the Linguistics Data Commons of Australia

This platform will capitalise on existing infrastructure, rescue vulnerable and dispersed collections, and link with improved analysis environments for new research outcomes.

2 Developing a Trove researcher platform for advanced research

Augmenting existing National Library of Australia resources, this platform will enable a focus on the delivery of researcher portals accessible through Trove, Australia's unique public heritage site. The platform will create tools for visualisation, entity recognition, transcription and geocoding across Trove content and other corpora.
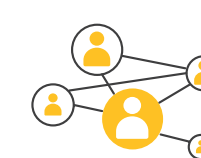
3 Integrated social sciences research infrastructure

This platform will expand existing social sciences initiatives and provide a coordinated governance model for access to data. It will improve the capacity of researchers to access, preserve and disseminate quantitative and qualitative social sciences data sources, and will drive the development of systems and tools for capturing new and emerging real time, or near real time, data.

Supporting Indigenous Research Capability

As part of the Australian Government investment, the ARDC will also work in collaboration with the Indigenous Data Network (IDN) at the University of Melbourne to support the consolidation and expansion of its technological, training and governance initiatives.

The HASS RDC will help institutions share data more freely, ethically and cooperatively, following the FAIR data principles and indigenous data governance protocols maintained by the Indigenous Data Network.

The HASS RDC brings together existing and ongoing investments in text analysis, cultural collections, linguistics, social sciences and indigenous data.

It will leverage existing NCRIS capabilities, including the Population Health Research Network (PHRN) and the Australian Urban Research Infrastructure Network (AURIN).

The infrastructure will enhance research in a broad range of fields including education, Australian cultural studies, history, economics, commerce, tourism, law and legal studies, and the creative arts. ∎

ARDC'S MOST VALUABLE RESOURCE:
OUR STAFF

ARDC

Australian Research Data Commons

MEMBERS

Australian National University

CSIRO

Curtin University

ECU Edith Cowan University Australia

Federation University

Flinders University

Griffith University Queensland, Australia

Macquarie University Sydney·Australia

Monash University

QUT

Swinburne University of Technology

The University of Adelaide

The University of Melbourne

The University of Queensland Australia

The University of Sydney

The University of Western Australia

UNSW Sydney

University of Tasmania

UTS

University of Wollongong Australia

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

The ARDC is enabled by NCRIS