Australian Research Data Commons

PARTNERING FOR SUCCESS

The ARDC is enabled by NCRIS

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

We acknowledge and celebrate the First Australians on whose traditional lands we live and work, and we pay our respects to Elders past, present and emerging.

Aboriginal and Torres Strait Islander people should be aware that this booklet may contain images and names of deceased persons.

Thank you to ARDC staff and ARDC-supported project and program team members for their contributions to this publication.

Approved by:
Keith Russell,
Director of Outreach, ARDC

Published:
July 2023

DOI: 10.5281/zenodo.8015861

Suggested citation:

ARDC Ltd. Partnering for Success. 2023.
Available from: doi.org/10.5281/zenodo.8015861.

Australian Research Data Commons

CONTENTS

A MESSAGE FROM OUR CHAIR 1

A MESSAGE FROM OUR CEO 3

ABOUT THE ARDC 5

PARTNERING FOR SUCCESS WITH THEMATIC RESEARCH DATA COMMONS 9

PEOPLE RESEARCH DATA COMMONS 11

Unlocking the Potential of Clinical Trial Data 13

Secure Data Analytics Platform, ERICA, Brings Better Health Outcomes 19

Transforming Clinical Image Management and Analysis 23

Blood Transfusions and Patient Outcomes: A New National Dataset 27

PLANET RESEARCH DATA COMMONS 29

Plant Trait Data Guides Australia's Bushfire Recovery 31

Unlocking Australia's Restricted Access Species Data 35

Finding Fast Answers to Urgent Ecological Questions 39

Streamlined Data for National Environmental Reporting 43

Reducing Deaths from Air Pollution 45

HUMANITIES, ARTS AND SOCIAL SCIENCES (HASS) AND INDIGENOUS RESEARCH DATA COMMONS 49

Empowering HASS and Indigenous Researchers with Essential Computational Skills 51

Introducing the ARDC's First Indigenous Intern, Ms Lisa Rigney 55

Telling Human Stories with Time-Layered Cultural Maps 58

Tweets Illuminate the Impact of COVID-19 on Society 62

Minding Our Language Data 66

TRANSLATIONAL RESEARCH DATA CHALLENGES 70

Bringing Together Bushfire History Data to Plan for the Future 71

BUILDING ON PARTNERSHIP SUCCESS 75

Co-Investment Partnerships 76

IMPACT SHOWCASE A Research Data Management Framework for Institutions 79

Research Software 83

IMPACT SHOWCASE The ARDC Awards Software Innovation 84

Data Retention 85

Digital Research Skills 85

IMPACT SHOWCASE Digital Research Skills Trainers Convene at the 2023 ARDC Skills Summit 86

IMPACT SHOWCASE Ocean Data Repository Certified as Trustworthy 87

Communities of Practice 88

Engagements, Expertise & Support 88

Our Input to Policy 88

IMPACT SHOWCASE Mapping the Vision for Data-Driven Research in Australia 89

OUR SERVICES 91

ARDC Nectar Research Cloud 92

IMPACT SHOWCASE Understanding the Impact of Our Future Climate with the ARDC Nectar Research Cloud 95

ARDC Research Data Australia 99

Persistent Identifiers 100

IMPACT SHOWCASE Research Efficiency Through Persistent Identifiers 101

ARDC Research Vocabularies Australia 103

IMPACT SHOWCASE FAIR Vocabularies in the Spotlight 104

Resources for Researchers 105

IMPACT SHOWCASE Our New Website to Meet the Needs of the Research Community 106

ACKNOWLEDGEMENTS 107

Our Members and Memberships 108

Our Partners 109

GOVERNANCE 111

Board of Directors 112

Research and Technology Advisory Committee 113

Senior Leadership Team 114

Our Team 115

A MESSAGE FROM OUR CHAIR

Craig Roy FAICD MBA MSc

As we witness the power of AI using big data to transform the way the world works, it is clear that providing Australian researchers with competitive advantage through data — the ARDC's mission — is more relevant than ever.

Over the past year, ARDC co-investment projects and expertise have contributed to the creation of national datasets and platforms, and excellence in research data management. By making research data more findable, accessible, interoperable and reusable, we ensure the greatest benefit for society.

ARDC co-investment projects are feeding directly into government reporting and decision making. EcoAssets is a great example, integrating biodiversity data from Australia's 3 national infrastructure facilities — the Atlas of Living Australia, Integrated Marine Observing System (IMOS) and Terrestrial Ecosystem Research Network (TERN) — to inform the 2021 State of the Environment report. And a new database of plant traits, AusTraits, has informed recovery action following the catastrophic bushfires of 2019/2020, through investment from the Threatened Species Commissioner. ▶

Australian Research Data Commons

Over the past year, ARDC co-investment projects and expertise have contributed to the creation of national datasets and platforms, and excellence in research data management. By making research data more findable, accessible, interoperable and reusable, we ensure the greatest benefit for society.

Craig Roy FAICD MBA MSc
Chair
Australian Research Data Commons

The ARDC Leadership Series this year was a highlight for the ARDC board. The series brought together leaders from research, industry and government to discuss the issues of managing sensitive and identifiable human data, translating research to industry, and closing the digital research skills gap. The series demonstrates the ARDC's unique role as a leader of collaboration to support data-driven research and innovation.

The future of the ARDC is secure for the next 5 years thanks to the acceptance of our phase 1 submission request for $119 million in operational funding by the Australian Government Department of Education Research Infrastructure Investment Plan (RIIP). As this booklet goes to press, we expect to learn whether our major initiatives in digital research infrastructure will benefit from the RIIP 2 investment.

The ARDC's success has been possible due to the commitment, dedication, expertise and work ethic of the ARDC team, led by our CEO, Rosie Hicks. I take this opportunity to express my gratitude to them. I also thank our partners for continuing to support, collaborate, and engage with us. We remain focused on our vision of an ecosystem where Australia's researchers can discover, access and reuse data, enabling them to collaborate, innovate and deliver research outcomes that make a difference to society.

We will continue in our unique role of bringing together diverse organisations — research institutions, research infrastructure providers, industry, state and federal governments — to collaboratively attain our shared goal of making research data more accessible for the benefit of all Australians.

On behalf of the ARDC board, I thank our team, our partners, and the Australian research community. We look forward to continuing to work with you on advancing research data management in Australia. ■

Image — Anthony McKee / ARDC

A MESSAGE FROM OUR CEO

Rosie Hicks

Our vision is for world leading knowledge infrastructure that enables Australian researchers to transform our lives. At the ARDC, we are committed to delivering enduring digital research infrastructure that will support researchers to tackle the grand challenges facing our society.

As digital research infrastructure matures and evolves, we must optimise the delivery of digital research services to researchers, reduce duplication of effort across institutions and research domains, and build enduring capability in the research community.

Further, we must prioritise, resource and govern the integration of data from diverse sources to create nationally significant collections.

Focused on Australian researchers' needs

The ARDC is leading the development of targeted, enduring digital research infrastructure to meet Australian researchers' needs and realise our vision.

Over the past 12 months, we've undertaken extensive nationwide consultations to gain a deep understanding of the research community's needs. As a result, we now have clear priorities for extending and optimising existing infrastructure to deliver a seamless research infrastructure ecosystem.

Building on the success of our mission-driven Translational Research Data Challenges, we are delivering a suite of Thematic Research Data Commons (RDCs) to meet the digital research infrastructure needs of about 70 per cent of the research sector. **The People RDC** is for health and medical research, **the Planet RDC** for earth and environmental sciences, and **the HASS and Indigenous RDC** for humanities, arts, social sciences and Indigenous research.

The Thematic RDCs draw on our extensive expertise in data management and the provision of digital research infrastructure and associated services to give researchers a coherent and scaled-up suite of data and tools, enabling deeper research insights, better policy, and enhanced services and products.

I hope that you enjoy reading about these initiatives in this impact booklet: *Partnering for success*. ▶

> As digital research infrastructure matures and evolves, we must optimise the delivery of digital research services to researchers, reduce duplication of effort across institutions and research domains, and build enduring capability in the research community.

Rosie Hicks,
Chief Executive Officer,
Australian Research Data Commons

On the global stage

We are delighted to have won the bid to host International Data Week in October 2025 in Brisbane with the support of leading institutional and industry partners in the Asia Pacific region.

International Data Week is the pre-eminent global research data conference, organised by the Committee on Data (CODATA), the World Data System of the International Science Council, and the Research Data Alliance. The conference will illuminate the power of data-driven research to solve the grand challenges facing the world, including climate change, food security and energy security.

This year, we're also celebrating 10 years of the Research Data Alliance, which was established by the Australian Government, the European Commission and the US Government. The Research Data Alliance builds the social and technical bridges that enable open sharing and reuse of data. This international network now numbers 13,500 academics, researchers and infrastructure providers.

A small but mighty sector

When compared with our European and American counterparts, the Australian research sector is small — but being small has its benefits. Australia's national research infrastructures are well connected and actively collaborate to give the research sector access to national data assets, research platforms, tools and skills.

At the ARDC, we've been able to harness this unique positioning to advance Australia's national digital research infrastructure, creating a seamless research ecosystem. Highlights include:

- In February 2023, in a world-first collaboration with 25 Australian universities, we launched the Research Data Management Framework for Institutions. The culmination of almost 2 years of collaboration by hundreds of experts, this landmark framework provides Australian universities a national approach for managing research data.
- To highlight the vital but hidden role of research software in almost all contemporary research, we are sponsoring the inaugural Australian Museum Eureka Prize for Excellence in Research Software, which will be awarded in August 2023. We are excited to see applications for the award from domains as diverse as linguistics, bioinformatics and ecology.
- The release of the Health Studies Australian National Data Asset (HeSANDA) in mid-2023 will see a new national data-sharing capability for health studies, with an initial focus on clinical trials. Created in partnership with the health research community, it will provide a secure and ethical way for clinical trialists to fulfil the data-sharing requirements of publicly funded research, and for researchers to discover and request access to clinical trial data.

As we transform the delivery of ARDC projects and services, I look forward to continuing to support researchers to maximise the value and impact of their data through the provision of innovative and sustainable digital research infrastructure.

I'd like to take this opportunity to thank our dedicated team for their hard work and commitment to our mission to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

Thank you also to our partners for your ongoing support and collaboration. ■

Clockwise:

Liz Stokes, ARDC, presenting at the ARDC Skills Summit

Grant Sarra, facilitator, and **Uncle Michael Williams,** former head of the Aboriginal and Torres Strait Islander Studies unit at The University of Queensland, at the event, Bringing Data to Life: Co-Designing a Language Data Commons

Keith Russell, ARDC, presenting at the ARDC Skills Summit

Associate Professor Sandra Philips, Western Sydney University, at the event, Bringing Data to Life: Co-Designing a Language Data Commons

Dr Paul Coddington and Darcelle Malby, ARDC, with partners from the ARDC Nectar Research Cloud nodes

Images — Renee Nowytarger | Anthony Mckee | Marc Grimwade / ARDC

Australian Research Data Commons

ABOUT THE ARDC

At the ARDC, we drive development of world-class national digital research infrastructure that gives Australian researchers competitive advantage through data and supports research impact.

The ARDC is Australia's leading facility for research data infrastructure. We facilitate access to research datasets and tools from academia, industry and government for all Australian researchers.

We run programs and form partnerships that ensure Australian researchers are internationally competitive through having access to high-quality data assets, platforms, infrastructure, policies, people and training to transform our lives.

Solving society's greatest challenges takes the collective efforts of society. Through our collaborations and partnerships — national and international — we are ensuring that valuable data and software assets are developed, made accessible and sustained for everyone.

Our 80+ staff are based at host institutions around Australia. We have 25 member institutions and growing.

The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). ■

OUR PURPOSE

To provide Australian researchers with competitive advantage through data.

OUR MISSION

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

OUR MEMBERS

Australian National University

CSIRO

Curtin University

Australia ECU Edith Cowan University

Federation University

Flinders University

Garvan Institute of Medical Research

Griffith University Queensland, Australia

Macquarie University Sydney · Australia

Monash University

QUT

RMIT University

SAHMRI South Australian Health & Medical Research Institute

Swinburne University of Technology

The University of Adelaide

The University of Melbourne

The University of Notre Dame Australia

The University Of Queensland Australia

UNSW Sydney

University of South Australia

The University of Sydney

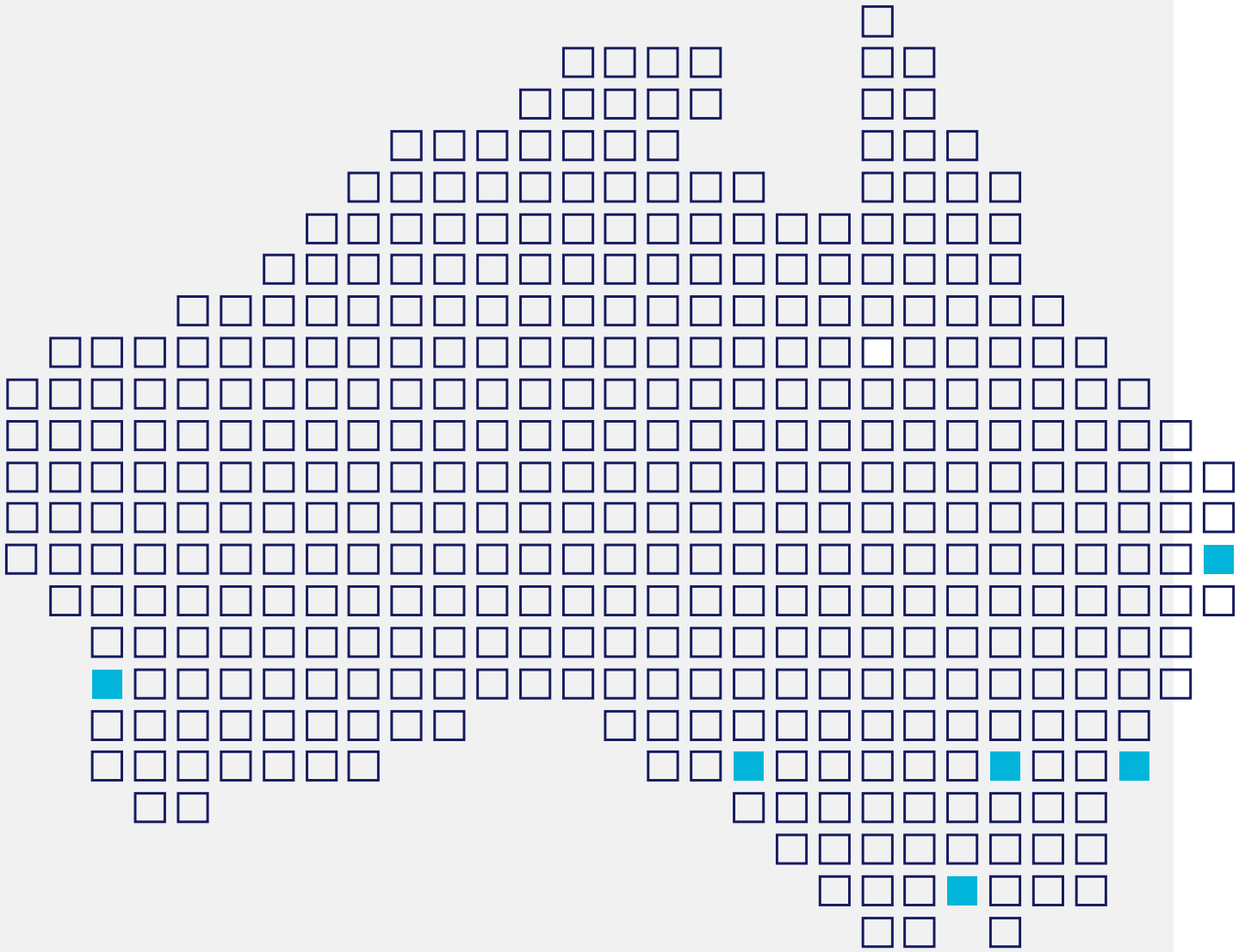
The University of Western Australia

University of Tasmania

UTS

University of Wollongong Australia

Australian Research Data Commons



ARDC HOST INSTITUTIONS

Australian Research Data Commons

PARTNERING FOR SUCCESS WITH THEMATIC RESEARCH DATA COMMONS

In partnership with the research community, we're building 3 national-scale Thematic Research Data Commons to meet Australia's future research needs with long-term, enduring digital infrastructure.

Each Thematic Research Data Commons integrates the ARDC's underpinning compute, storage infrastructure and services with analysis platforms and tools. Each is supported by our expertise, and our work on developing community-agreed standards and best practices.

These coordinated, structured, and complementary activities are building data assets, tools, and skills that will constitute a national 'knowledge infrastructure' that enables Australian researchers to transform our lives.

Co-designed with the research community through extensive consultations and broad partnerships, they will enable us to achieve our goal of supporting the maximum number of researchers in strategic priority areas of research through a new approach to participation and organisation.

Read on to learn more about our Thematic Research Data Commons, along with impact case studies that show how digital research infrastructure is meeting researcher needs in these research areas. ■

PEOPLE RESEARCH DATA COMMONS
for health and medical research
PAGES 11 – 28

PLANET RESEARCH DATA COMMONS
for earth and environmental science research
PAGES 29 – 48

HASS AND INDIGENOUS RESEARCH DATA COMMONS
PAGES 49 –69

TRANSLATIONAL RESEARCH DATA CHALLENGES
PAGE 70 – 74

Image — Anthony Mckee / ARDC

Left to right:
Rosie Hicks, CEO, **Jenny Fewster,** Director, HASS and Indigenous Research Data Commons, **Dr Sheida Hadavi,** Director, Translational Research Data Challenges, **Dr Adrian Burton,** Deputy CEO, Director, People Research Data Commons, **Hamish Holewa,** Director, Planet Research Data Commons

> Building on the success of our mission-driven Translational Research Data Challenges, we are delivering a suite of Thematic RDCs to meet the digital research infrastructure needs of about 70 per cent of the research sector.
>
> **Rosie Hicks,** Chief Executive Officer, *Australian Research Data Commons*

Australian Research Data Commons

PEOPLE RESEARCH DATA COMMONS

National-scale data infrastructure for health research and research translation

Australia's health industry is world-leading due to significant investment in both research and the translation of health research to benefit the community. To enable this translation, evidence-based healthcare improvements are required, which in turn are driven by data. The national data landscape for health research is rich, diverse and complex, spanning multiple sectors, jurisdictions and data types. The sensitive nature of health data, combined with privacy and regulatory requirements, demands that digital research infrastructure can support secure and seamless national-scale research.

That's why we're partnering with the research community, industry and government to co-design and deliver the People Research Data Commons (People RDC), which will support health and biomedical researchers to develop and sustain cross-sector and multidisciplinary data collaborations at a national scale.

To support new research and translation for better health outcomes, the People RDC is delivering national digital health capability in 4 key areas where digital health research is a challenge:

SECURE DATA ACCESS

Data custodians typically operate bespoke secure environments for data analysis; however, these environments can create data silos that hinder national-scale multidisciplinary collaborations. To protect data and maintain control across secure data analysis platforms, it is crucial to improve interoperability and incorporate privacy-preserving technical measures and seamless data governance.

ADVANCED ANALYTICS

Applying advanced analytics techniques such as machine learning to national sensitive data collections across multiple secure data repositories raises data challenges that require innovative solutions.

DATA INTEGRATION

Australia has world-leading data linkage capabilities supported by NCRIS, enabling health data to be linked within and across jurisdictions. To solve the grand challenges facing society, it is essential to broaden and connect national data integration capabilities with a wide range of other data assets while ensuring the highest privacy standards to meet research demand.

DATA STRATEGY AND DISCOVERY

Health research has a complex data ecosystem with national data assets distributed across government, research and health service sectors. For researchers and industry, identifying the location, availability and access of this data is a significant challenge.

Through building these fundamental digital capabilities, the ARDC can meet the needs of diverse health and medical research communities throughout Australia in a strategic and comprehensive way.

The People RDC builds on our experience creating national digital research infrastructure for health and medical researchers. For example, HeSANDA, our Health Studies National Data Asset, is a national research infrastructure that is enabling researchers to share and access health studies data.

To learn more about how our partnership approach, expertise, and digital research services have helped create digital research infrastructure for researchers in this domain, read our impact case studies on the coming pages. ■

Be part of the People Research Data Commons. ***Visit - bit.ly/ARDCPeopleRDC***

CASE STUDIES

Unlocking the Potential of Clinical Trial Data

Researchers and clinicians can now share and find clinical trial data via a new national platform.

PAGE 13

Secure Data Analytics Platform, ERICA, Brings Better Health Outcomes

The ERICA platform is a fully scalable and secure online environment, created to enable people to analyse health-related and other highly sensitive data.

PAGE 19

Transforming Clinical Image Management and Analysis

The Australian Imaging Service is enhancing imaging research on a global scale and making it more secure and easier.

PAGE 23

Blood Transfusions and Patient Outcomes: A New National Dataset

The new National Transfusion Dataset links the use of life-saving blood products with patient outcomes to reveal new insights.

PAGE 27

Image — Gorodenkoff - 236237514 / AdobeStock.com

Image — Arnela Ramba/peopleimages.com - 528850160 / AdobeStock.com

> HeSANDA... is creating new opportunities in health research in Australia and internationally. For example, sharing data enables the pooling of data from multiple trials, which can lead to better understanding of the effect of a treatment in certain cohorts of participants that may have been previously unexplored.

Dr Melina Willson,
Infrastructure Manager, NHMRC Clinical Trials Centre,
The University of Sydney

Unlocking the Potential of Clinical Trial Data

Sharing and requesting access to data from health studies just became easier, with the launch of a new national platform.

Initially focusing on data from clinical trials, the Health Studies National Data Asset (HeSANDA) allows clinical trialists and data custodians to share clinical trial data according to standards designed by the Australian clinical trials community and fulfil data-sharing requirements that come with public funding. The platform can subsequently be used by researchers who were not involved in the clinical trials to find and request access to this data.

The HeSANDA program aims to stimulate new data-driven research ideas, increase the impact of health research and, ultimately, improve the health and wellbeing of Australians. Researchers stand to unlock the massive potential of the data in understanding health and medical problems. Sharing data also avoids duplicating research, saving time and money. ▶

Health Research Data Without Borders

Health research studies in Australia produce huge amounts of data that contains information about research participants, their health, and their response to the interventions being studied.

Data collected in one study can be extremely valuable to other studies. However, patient privacy requirements and the naturally siloed approaches of research groups and state jurisdictions have made it hard to share data.

In partnership with the health research community, the ARDC played a critical role in synchronising efforts, aligning approaches and building national data capability through the HeSANDA program.

Now, this untapped trove of health research data is set to bring immense value to health research, maximising the return on investment of past research, and laying a foundation for future research to improve health outcomes for Australians.

In her role as the infrastructure manager at the NHMRC Clinical Trials Centre, The University of Sydney, Dr Melina Willson manages the Australian New Zealand Clinical Trials Registry (ANZCTR).

“HeSANDA will have an impact at 3 levels,” said Dr Willson. “Firstly, it is creating new opportunities in health research in Australia and internationally. For example, sharing data enables the pooling of data from multiple trials, which can lead to better understanding of the effect of a treatment in certain cohorts of participants that may have been previously unexplored.

“Secondly, it is creating a network of Australian researchers that will be well equipped with the know-how and processes that encourage sharing of data. And lastly, HeSANDA can spur on other countries to develop a similar initiative. This is significant because it would enable better international data sharing and collaboration.” ▶

> While there is still a way to go, it is evident that the HeSANDA team has ensured that relevant individuals and organisations across Australia are engaged in this journey together and that the momentum for change is not only building, but the key infrastructure will be in place to enable these changes to be adopted. This project has the potential to accelerate the translation of research into practice, reduce research waste and ultimately improve patient outcomes.
>
> **Wendy Keech,** Executive Director, *Health Translation SA (HTSA)*

Australian Research Data Commons

A Co-Design Approach

The framework for sharing clinical trial data was co-designed by the ARDC and experts and representatives for the initial 9 nodes of the distributed HeSANDA network, which cover 72 health research organisations, health service operators and clinical trial networks from across Australia. More than 90 people took part in working groups, incorporating feedback from research trial participants, consumers, researchers and trial organisers.

“We’re pleased to be establishing consensus amongst a very large and diverse group for the design and approach of HeSANDA,” said the ARDC’s Dr Kristan Kang, Program Manager for HeSANDA.

The co-designed HeSANDA framework would not have been created without the valuable contributions of a wide range of organisations, including:

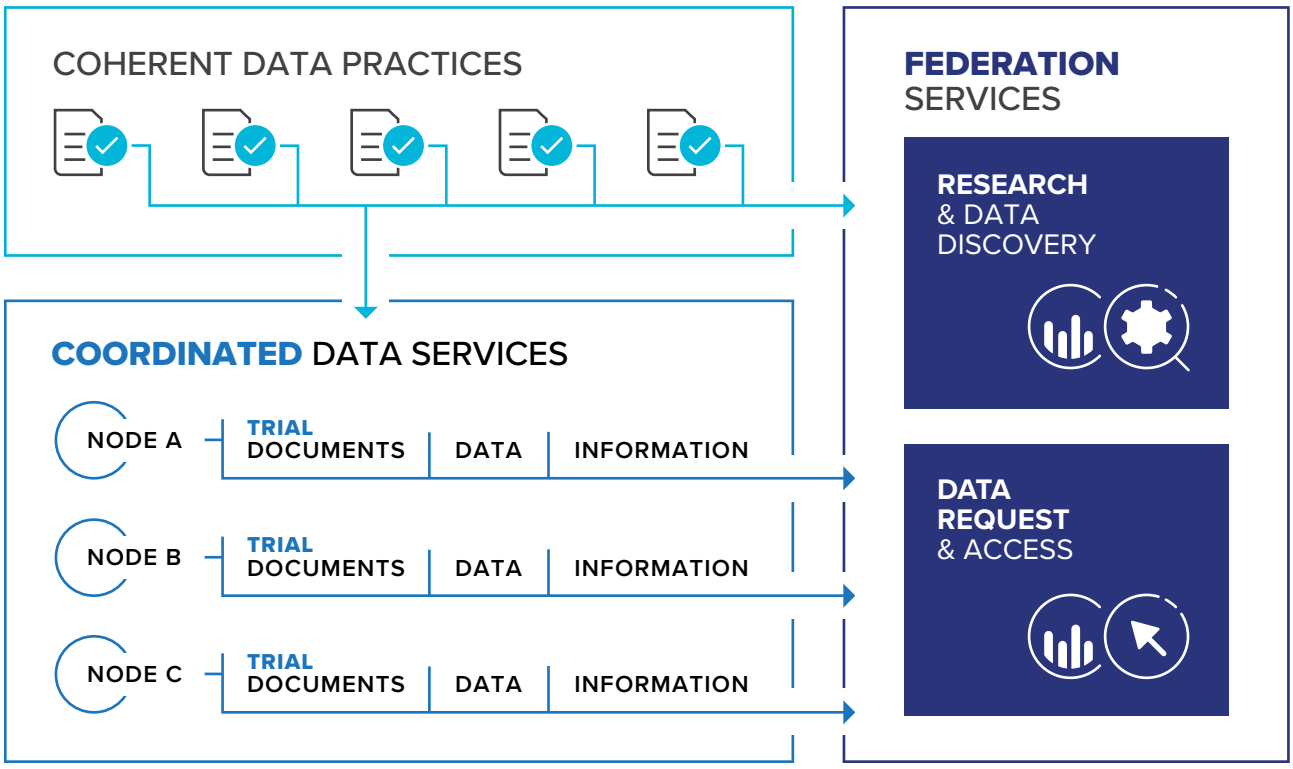
- Australian Clinical Trials Alliance (ACTA)
- Australian Genomics
- Australian Health Research Alliance
- Australian Institute of Health and Welfare
- Australian New Zealand Clinical Trials Registry
- Australian Research Management Society
- Cochrane Australia
- Consumers Health Forum of Australia
- CT:IQ
- National Health and Medical Research Council (NHMRC)
- Population Health Research Network
- Research Australia.

Image — DC Studio - 574566575 / AdobeStock.com

Wendy Keech is the Executive Director of Health Translation SA (HTSA), a node of HeSANDA.

“The HTSA Board of Partners is committed to enabling the secondary use of clinical trial data across Australia. HTSA is very pleased to be one of the 9 HeSANDA nodes that are working to implement systems and engage clinical trialists to achieve this vision,” said Ms Keech.

“While there is still a way to go, it is evident that the HeSANDA team has ensured that relevant individuals and organisations across Australia are engaged in this journey together and that the momentum for change is not only building, but the key infrastructure will be in place to enable these changes to be adopted. This project has the potential to accelerate the translation of research into practice, reduce research waste and ultimately improve patient outcomes.” ▶



HeSANDA infrastructure model

Standard Practices for Sharing Data

During the design phase, a set of coherent data practices was identified to complement the data management infrastructure (see diagram above). These data practices give the clinical trial data community a common framework for sharing data via HeSANDA.

The data practices cover:

- ethics and consent for data sharing
- data-sharing policies and procedures
- standard metadata to describe clinical trials and the availability of their data.

Keeping Data with Its Custodians

The HeSANDA infrastructure is a nationally distributed network of 9 nodes, which launched in mid-2023. The nodes, which comprise 72 organisations, deliver coordinated data services federated at the national level.

The participating organisations have agreed on a common high-level framework for curating and sharing clinical trial data. To adhere to this framework, they have built or modified their policies, systems and procedures.

Resources have been developed for health researchers to support their data management practices, as well as addressing the culture and policy aspects that support this valuable data infrastructure.

While HeSANDA is focusing on sharing clinical trial data, the ARDC is growing its digital research services for the medical and health research community. Our aim is to catalogue and share other types of health and medical data and provide secure data access for clinical trials via the People Research Data Commons. ▶

HeSANDA Partners

There are 9 HeSANDA nodes that represent 72 health research organisations across Australia. They are:

Melbourne Academic Centre for Health (MACH) Clinical Trials Consortium Node

Led by

The University of Melbourne

Partners

- Austin Health
- BioGrid Australia
- Centre for Eye Research Australia
- Murdoch Children's Research Institute & Royal Children's Hospital
- Northern Health
- PeterMac Cancer Centre
- St Vincent's Hospital Melbourne
- The Royal Melbourne Hospital
- Western Health

Mental Health Node

Led by

Deakin University

Partners

- Australian Early Psychosis Collaborative Consortium
- Barwon Health
- Orygen
- The University of Melbourne

Monash and Partners Node

Led by

Monash University

Partners

Monash partners

Northern Australia Node

Led by

Menzies School of Health Research

Partners

Charles Darwin University

National Cancer Cooperative Trials Groups Node

Led by

Australasian Leukaemia and Lymphoma Group

Partners

- Australian & New Zealand Childrens Haematology/Oncology Group
- Australasian Gastro-Intestinal Trials Group
- Australia and New Zealand Sarcoma Association
- Australia New Zealand Gynaecological Oncology Group
- Australian & New Zealand Urogenital and Prostate Cancer Trials Group
- Breast Cancer Trials
- Cancer Symptom Trials, Palliative Care Clinical Studies Collaborative
- Cooperative Trials Group for Neuro-Oncology
- Melanoma and Skin Cancer Trials
- Primary Care Collaborative Cancer Clinical Trial Group
- Psycho-oncology Co-operative Research Group
- Thoracic Oncology Group of Australasia
- Trans-Tasman Radiation Oncology Group (TROG)

Image — Sharne T/peopleimages.com - 577006515 / AdobeStock.com

Australian Research Data Commons

Image — Kay A/peopleimages.com - 566595863 / AdobeStock.com

Queensland Node

Led by

Health Translation Queensland

Partners

- AARNet
- Australasian Kidney Trials Network
- Brisbane Diamantina Health Partners
- Cairns and Hinterland Hospital and Health Service
- Central Queensland University
- Children's Health Queensland Hospital and Health Service
- CSIRO's Australian e-Health Research Centre
- Griffith University
- James Cook University
- Mackay Hospital and Health Service
- Mater Misericordiae Ltd
- QIMR Berghofer Medical Research Institute
- QCIF
- Queensland Government Department of Health
- Queensland Government Metro North Hospital and Health Services
- QUT
- The University of Queensland
- Townsville Hospital and Health Service
- Tropical Australian Academic Health Centre Ltd
- University of the Sunshine Coast
- University of Southern Queensland

South Australia Node

Led by

South Australian Health and Medical Research Institute (SAHMRI)

Partners

- Flinders University
- Health Translation SA

Sydney Health Partners Node

Led by

NHMRC Clinical Trials Centre at The University of Sydney

Partners

- Cardiovascular Centre of Excellence
- Digital Health CRC
- Institute for Musculoskeletal Health
- Institute of Bone and Joint Research
- Sydney Children's Hospitals Network
- Sydney Health Partners
- Sydney Local Health District
- The Brain and Mind Centre
- The University of Sydney
- Westmead Applied Research Centre
- Westmead Institute for Medical Research
- Woolcock Institute for Medical Research

West Australian Health Translation Network Node

Led by

Curtin University

Partners

- Child and Adolescent Health Service
- Ear Science Institute Australia
- Edith Cowan University
- Government of Western Australia East Metropolitan Health Service
- Government of Western Australia North Metropolitan Health Service
- Government of Western Australia South Metropolitan Health Service
- Harry Perkins Institute of Medical Research
- Institute of Respiratory Health
- Lions Eye Institute
- Murdoch University
- The University of Western Australia
- WA Country Health Service ■

Australian Research Data Commons

Secure Data Analytics Platform, ERICA, Brings Better Health Outcomes

When you're in an ambulance after a suspected stroke, every minute matters. You want a quick diagnosis and the right treatment.

Health professionals have observed that after a stroke women were not recovering as well as men — they had higher rates of disability and death.

To get to the bottom of this puzzle, health researchers at The George Institute for Global Health in Sydney used the new, secure and powerful cloud-based platform, ERICA, which stands for E-Research Institutional Cloud Architecture. ERICA enables researchers to analyse huge datasets while adhering to strict ethics guidelines.

With ERICA, the researchers were able to pull together and analyse detailed and sensitive datasets related to more than 200,000 patients in a safe, secure environment. The data included hospital records from all hospitals and day procedure centres in New South Wales, and NSW ambulance records.

Their analysis revealed that women were less likely to receive stroke-specific management in the ambulance. This valuable finding is now informing paramedic training to help them better spot stroke symptoms in women.

ERICA Has Transformed How Researchers Solve Problems

According to Professor Louisa Jorm, Director of the Centre for Big Data Research in Health, who is leading the national expansion of ERICA through the ARDC, the platform has transformed how researchers and data analysts can solve big, complex problems that have life-or-death consequences.

"Prior to now the secure data infrastructure didn't exist that could link together records from ambulances, hospitals and historical health data," said Prof Jorm.

"In the case of the research into stroke, we've been able to compare the diagnosis patients were given when they arrived in hospital with what the ambulance paramedics had assessed and the treatment provided in the ambulance.

"As a result, paramedics and clinicians are being upskilled in what they can do to increase the accuracy of early identification of stroke, particularly in female patients." ▶

Image — arliftato22205 - 473945102 / AdobeStock.com

In the case of the research into stroke, we've been able to compare the diagnosis patients were given when they arrived in hospital with what the ambulance paramedics had assessed and the treatment provided in the ambulance.

As a result, paramedics and clinicians are being upskilled in what they can do to increase the accuracy of early identification of stroke, particularly in female patients.

Professor Louisa Jorm, Director, *Centre for Big Data Research in Health*

Secure, Ethical Research at Scale

ERICA is a fully scalable and secure online environment, created to enable people to analyse health-related and other highly sensitive data. The platform is approved by eHealth NSW under the Privacy and Security Assurance Framework, and Commonwealth Infosec Registered Assessor Program assessment is underway.

ERICA provides a secure environment, like a virtual data laboratory, for projects to conduct analyses safely and securely.

Researchers can access the computational power of ERICA from anywhere and, because it uses Amazon Web Services, there's no limit to the physical infrastructure; for example, there's no need to buy, install and configure larger servers to hold data.

ERICA Project Manager Phi Nguyen explains: "We don't have the limitations of physical infrastructure to spin up, or buy and house. And when we finish a project, the capability is transferable as well. So, there's a lot of flexibility for us to be able to move projects, create new environments.

"For example, if someone needs a graphical processing unit or a lot more compute power, then we can create it really quickly rather than having to go out and procure something. There's nothing similar in the research space, as far as secure environments are concerned," he said.

Without secure data laboratories like ERICA, research into critical health issues isn't possible. Conservative estimates put the recent impact of ERICA as enabling over $8 million worth of grants through the Medical Research Future Fund. The future fund is a $20 billion long-term investment in Australian health and medical research. ▶

ERICA Gets to the Heart of Shared Data

A team of Australian doctors and university researchers with a focus on heart health are also using ERICA. They are creating an Australian cardiovascular data resource that allows researchers to access ongoing linked data, with appropriate privacy and security protections.

> ERICA has been a key enabler to the establishment of the CardiacAI project, a unique electronic medical record data resource for cardiovascular research... (it) has allowed us to gain the trust of local health districts and ethics committees in our ability to keep patient data secure.
>
> **Vicky Blake,** Data Manager/Analyst, *Eastern Heart Clinic*

Image — C Malambo/peopleimages.com - 566596170 / AdobeStock.com

Uptake by Universities and Government Agencies

ERICA has hundreds of users in organisations such as the Australian Institute of Health and Welfare, the New South Wales Data and Analytics Centre, and the Western Australia Department of Health. At UNSW Sydney, where ERICA began, there are now 57 enthusiastic users.

Medical Records Next Up for ERICA

While sensitive data is analysed in all sectors, the health sector is the most significant area for innovation, according to Phi Nguyen. His focus now for ERICA is electronic medical records (eMR) in hospitals, which track patients' care while at hospital.

"These medical records hold really complex and large data. They contain things like free-text data from clinical notes, images, ECGs, physiological readouts from machines," said Mr Nguyen.

"These multimodal datasets, made of many different data types, are massive in size, and it requires lots of computing power to bring to life what data analysts want to create, such as predictive analytics using AI techniques, machine learning.

"To realise the potential benefits from analysing this data, you need ERICA because of its combination of computing power and security, which makes it unique," he said. ■

> These multimodal datasets, made of many different data types, are massive in size, and it requires lots of computing power to bring to life what data analysts want to create, such as predictive analytics using AI techniques, machine learning.
>
> **Phi Nguyen,**
> Project Manager,
> *ERICA*

ERICA

ERICA received co-investment (doi.org/10.47486/PL109) from the ARDC. It is led by UNSW Sydney in partnership with the ARDC, The University of Melbourne, Intersect Australia, NSW Data Analytics Centre, The George Institute Australia, Australian Institute of Health and Welfare, SA-NT Datalink, The University of Adelaide, University of South Australia, University of Western Australia, and the Population Health Research Network.

Australian Research Data Commons

Transforming Clinical Image Management and Analysis

Three million people around the world – more than 33,000 of whom are in Australia – are living with multiple sclerosis (MS). It is a disease that damages nerves and disrupts communication between the brain and the body. This can result in difficulty walking, vision changes and various other ailments. ▶

Image — Maksym Povozniuk - 425342467 / AdobeStock.com

There is currently no cure for MS, but neurologists around the world are collaborating to better understand it and improve healthcare for patients.

The MSBase Registry is a global online registry for MS researchers, dedicated to sharing, tracking and evaluating data on outcomes in MS and other neuro-immunological diseases. Since its launch in 2004, it has become the largest organised repository of longitudinal MS patient data, with over 85,000 patient records from health facilities in 45 countries.

Now a new global platform – the MSBase Imaging Repository (MSBIR) – is further facilitating international collaboration in MS research by enabling secure sharing and analysis of de-identified clinical images of MS, such as MRI scans. Integrated with the MSBase Registry, it was developed in a partnership between The University of Sydney's Brain and Mind Centre, MSBase and the Sydney Neuroimaging Analysis Centre (SNAC), and one of its building blocks is the ARDC-supported Australian Imaging Service (AIS).

> Our mission is to increase research reproducibility and drive the adoption of innovative but trusted analysis techniques. We aim to create a unified service underpinning all imaging research conducted by Australians, both nationally and abroad, on which more specific research and development programs can be built.

Dr Ryan Sullivan, Head, *Australian Imaging Service*

Image — zinkevych - 144209929 / AdobeStock.com

Introducing the AIS

The AIS is a national platform for secure management, analysis and informatics of mainly biomedical and veterinary imaging data. It is led by The University of Sydney in partnership with and with co-investment from the ARDC, the National Imaging Facility and 10 other universities and research organisations.

"Our mission is to increase research reproducibility and drive the adoption of innovative but trusted analysis techniques. We aim to create a unified service underpinning all imaging research conducted by Australians, both nationally and abroad, on which more specific research and development programs can be built," said Dr Ryan Sullivan, who leads the AIS.

The AIS operates as a federation. It co-maintains a central set of software repositories with each partner institution operating their own node according to their local governance, infrastructure and cost structures.

The AIS integrates with imaging devices in a hub-and-spoke model. Each node integrates their local academic and clinical equipment, and data can be transferred between nodes to facilitate multi-site studies. ▶

Easy, Secure and Powerful

By adopting and standardising user authentication, instrument integration, data ontologies and mature software tools, the AIS lets researchers and facilities spend more time on innovation. It also allows reuse of national datasets by building a provenance trail from image capture to manipulation.

Taking a data-centric computing approach, the AIS integrates analysis and informatics directly into data management. This is broken into 4 key capability areas, the first 2 of which are live now:

- secure data management and collaboration built around XNAT, a popular open-source imaging data informatics platform developed by Washington University, and the Open Health Imaging Foundation (OHIF) DICOM medical image viewer
- non-interactive pipelines built using Arcana, a Python framework for repository-centric analysis
- interactive analysis integrating Jupyter and Neurodesk to provide secure virtual desktops, allowing direct access of the data without leaving the platform
- machine learning, starting with MONAI Label, a tool for AI-assisted image annotation and segmentation.

Semiautomated MS lesion analysis using Jim, an imaging analysis software toolkit for MRI, CT and other types of medical image. The AIS is providing secure, scalable and audited infrastructure across multiple international jurisdictions for the long-term management of these images to enable the large-scale analytics needed for new insights.

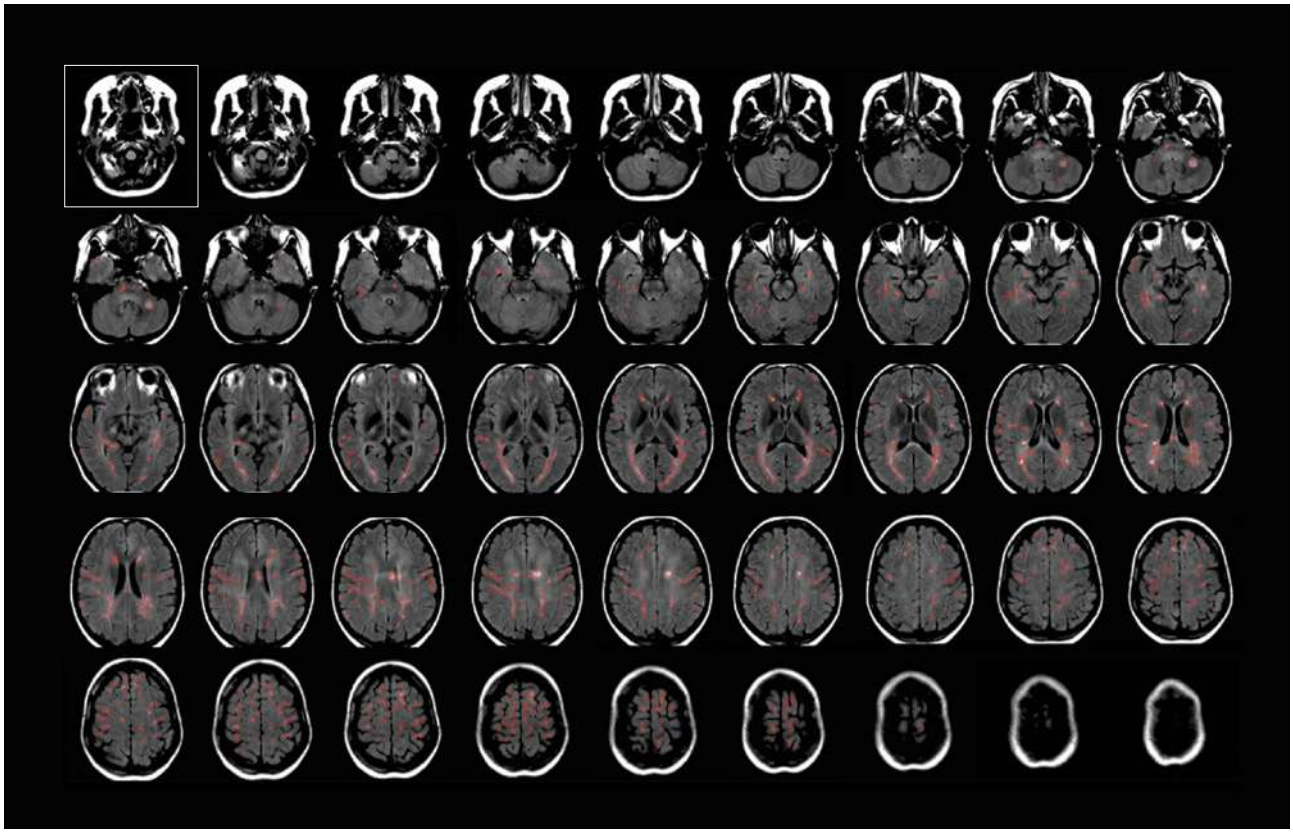


Image - Sydney Neuroimaging Analysis Centre

> The conclusion of the ARDC AIS project is the start of the AIS platform. We currently support 398 users across 279 projects and have a number of large programs that we are partnering with to support, such as the ACRF Australian Centre of Excellence in Melanoma & Diagnosis and recently announced Medical Research Future Fund projects.
>
> **Dr Ryan Sullivan,** Head, *Australian Imaging Service*

A Global Platform

The AIS team is keenly aware of the need for international collaboration in imaging research. This is precisely the case with research into MS, a global challenge requiring global solutions like the MSBIR.

For the MSBIR project, the AIS collaborated with Radiologics, which designed a federated system that would allow patient data to be domiciled in its home country while coordinating access for specialists from a single point that dynamically aggregates data for analysis.

Thanks to the cloud-native design of the AIS, the Australian, US, EU and lead MSBIR AIS nodes are in full production. Each node is accompanied by the Sydney Neuroimaging Analysis Centre's TORANA software, which handles seamless data ingestion and custom integration with the MSBIR analysis engine to provide AI-based, quantitative imaging metrics. By providing secure, scalable and audited infrastructure across multiple jurisdictions around the world for long-term clinical image management, the AIS enables the large-scale analytics needed by global MS researchers for new insights into the disease.

"Going forward, our partnership with the AIS will further optimise the successful workflow and ongoing global implementation of the MSBIR," said Professor Michael Barnett and Dr Heidi Beadnall, leaders of the MSBIR Project Control Board. It is also opening the way for global, more widely accessible AIS nodes to facilitate partnerships between Australia and the rest of the world.

Reaching Far and Wide

The AIS team is now working with other ARDC-supported projects, such as the Australian Electrophysiology Data Analytics Platform, the Veterinary and Animal Research Data Commons, and the Australian Characterisation Commons at Scale.

"The conclusion of the ARDC AIS project is the start of the AIS platform. We currently support 398 users across 279 projects and have a number of large programs that we are partnering with to support, such as the ACRF Australian Centre of Excellence in Melanoma & Diagnosis and recently announced Medical Research Future Fund projects," said Dr Sullivan. ■

The AIS (doi.org/10.47486/PL102) was a partnership of the ARDC, the National Imaging Facility, The University of Sydney, Macquarie University, Queensland Cyber Infrastructure Foundation, Queensland University of Technology, The University of Queensland, UNSW Sydney, Neuroscience Research Australia, The University of Western Australia, and SAHMRI.

Swinburne University, Monash University and The Florey advised on the project.

Blood Transfusions and Patient Outcomes: A New National Dataset

Blood is a precious resource. Every day, Australians receive transfusions for critical bleeding, whether from trauma, surgery, ulcers, obstetric bleeding or transplants, and for other medical conditions such as cancer, iron deficiency and bleeding disorders. Australia spends over $1.2 billion each year on blood products, and even more on the systems to administer them. Transfusions are done in ambulances, hospitals, community facilities and even in people's homes.

Now, for the first time, data on how, where and when the various blood products are used has been aggregated nationally and linked to the outcomes for patients receiving blood transfusions.

The new National Transfusion Dataset builds on the existing Massive Transfusion Registry for Australia and New Zealand and is integrated with clinical registries for blood disorders and cancers.

Developed with ARDC co-investment, the National Transfusion Dataset is an invaluable resource for understanding:

- how and why the various blood products are used
- the characteristics of patients receiving transfusions
- clinical outcomes, such as mortality rates, for patients following transfusion.

The dataset is available for use by researchers, clinicians, governments, industry and others to inform patient care, develop clinical guidelines, and generate hypothesis-driven research in the area.

Professor Erica Wood, Head of the Transfusion Research Unit at Monash University, leads the program to develop the National Transfusion Dataset.

"We've been very pleased to work with the ARDC on this important research. Blood products can save lives, but they carry risks and costs. They are donated by volunteers, so it's essential that we use them wisely."

"We've been very pleased to work with the ARDC on this important research. Blood products can save lives, but they carry risks and costs. They are donated by volunteers, so it's essential that we use them wisely.

Professor Erica Wood,
Head of the Transfusion Research Unit,
Monash University

Pilot Shows the Benefits of Linking Transfusion Data With Patient Outcomes

Prof Wood and colleagues ran a pilot to make the first linkages of transfusion data with national clinical registry data. The initial data came from Ambulance Victoria and pilot hospitals The Alfred and the Flinders Medical Centre.

"While transfusions are currently captured in some registries, data are often incomplete as the registries rely on on-site staff to enter clinical and laboratory data for participating patients," said Prof Wood.

"The linkages enabled the team to examine blood use and clinical outcomes in patients with blood disorders such as aplastic anaemia (who often require high-volume, medium to long-term support with red cells and platelets), and blood cancers such as myeloma (who may require immunoglobulin replacement prepared from donated plasma to prevent or treat infections).

"The National Transfusion Dataset was able to supplement transfusion data in these registries to provide a clearer picture at our pilot sites of how blood products were being used in these patient populations." ▶

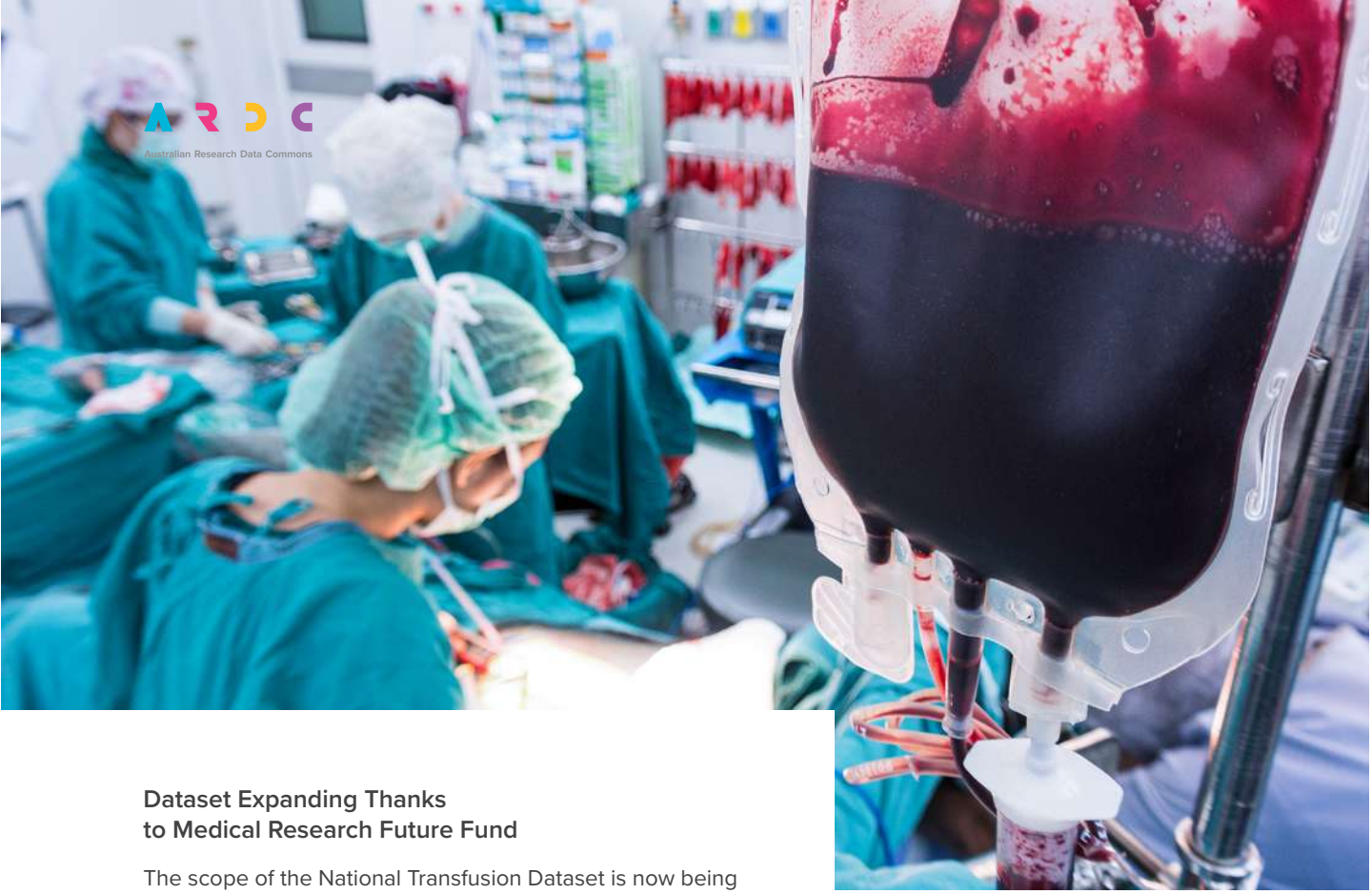


Image — chanawit - 71654301 / AdobeStock.com

Dataset Expanding Thanks to Medical Research Future Fund

The scope of the National Transfusion Dataset is now being expanded as part of the National Transfusion Research Data Infrastructure Initiative, which is receiving $2.9 million from the Australian Government's Medical Research Future Fund (MRFF).

"The success of the National Transfusion Dataset pilot laid the groundwork for the successful MRFF application," said Prof Wood.

New linkages are being made to more "pre-hospital" datasets (from ambulance and retrieval services), more clinical registries (such as critical care), and data from Lifeblood, Australia's national blood service.

"We need better national data to improve our practice and outcomes, and securing this additional MRFF funding and bringing new partners on board will really help develop the project further."

The MRFF-funded project, led by the Transfusion Research Unit at Monash University, also includes a cutting-edge project using AI algorithms to analyse unstructured data in the form of hospital electronic medical records.

The expansion will facilitate research into national priority areas, including:

- areas of high use of blood products, such as the support of critically ill patients and those with major haemorrhage or bone marrow failure
- the development of haemovigilance datasets to monitor blood product safety
- economic analyses of blood product use for health
- Australia's first registry-based transfusion clinical trial. ■

> We need better national data to improve our practice and outcomes, and securing this additional MRFF funding and bringing new partners on board will really help develop the project further.
>
> **Professor Erica Wood,**
> Head of the Transfusion Research Unit,
> *Monash University*

The National Transfusion Dataset received co-investment (doi.org/10.47486/DP708) from the ARDC. It is led by Monash University in partnership with the ARDC, Ambulance Victoria, Blood Synergy, Myeloma and Related Diseases Registry, Lymphoma and Related Diseases Registry, Australian and New Zealand Intensive Care Society, Pre-hospital Emergency Care Australia and New Zealand, and Aplastic Anaemia and Other Bone Marrow Failure Syndromes Registry.

Australian Research Data Commons

PLANET RESEARCH DATA COMMONS

National-scale data infrastructure for earth and environmental science research and decision making

To meet the needs of earth and environmental science researchers, the ARDC is co-designing a Planet Research Data Commons (Planet RDC) with partners in academia, industry and government and with other national stakeholders.

The Planet RDC will enable seamless access to data and digital infrastructure to meet the current and emerging research needs across multiple domains, including biodiversity, ecology, earth sciences, marine sciences, agriculture, biosecurity and climate adaptation.

Through supporting cross-sectoral, multidisciplinary and nationwide data collaborations, it will help to accelerate research.

The Planet RDC will provide:

LEADERSHIP

National leadership to co-design, identify and deliver connected infrastructure; and to develop and support world-class digital skills and networked capability in environmental and earth sciences data and modelling.

FOUNDATIONAL INFRASTRUCTURE

Sustainable and secure digital research infrastructure to support the development of analysis, modelling and decision-support infrastructure across multiple sectors.

NATIONAL DATA PROGRAMS

The national capability to catalyse and enable data integrations, connectivity and access across research, government and industry.

PARTNERSHIPS

Cross-sectoral engagement to promote and develop research translation pathways.

The Planet RDC builds on the ARDC's experience creating national digital research infrastructure for earth and environmental science researchers. To learn more about how our partnership approach, expertise, and digital research services have helped create digital research infrastructure for researchers in this domain, read our impact case studies below. ■

Be part of the Planet Research Data Commons. *Visit - bit.ly/ARDCPlanetRDC*

CASE STUDIES

Plant Trait Data Guides Australia's Bushfire Recovery

A plant trait database has informed the prioritisation of recovery actions and national-scale conservation planning following the 2019–2020 megafires.

PAGE 31

Unlocking Australia's Restricted Access Species Data

In a first-of-its-kind collaborative project supported by the ARDC, a nationally consistent approach to restricted access species data has been developed to increase the discoverability, access and sharing of restricted access species data.

PAGE 35

Finding Fast Answers to Urgent Ecological Questions

Building on existing econinformatics services, EcoCommons is making it faster and easier to tackle pressing challenges for our planet.

PAGE 39

Streamlined Data for National Environmental Reporting

EcoAssets provides freely accessible, integrated data for Australian environmental reporting and assessment.

PAGE 43

Reducing Deaths from Air Pollution

A new research platform standardises and links environment and health observations, enabling researchers to quantify population exposure and health burden due to air pollution, which can inform policies that support sustainable cities and environments.

PAGE 45

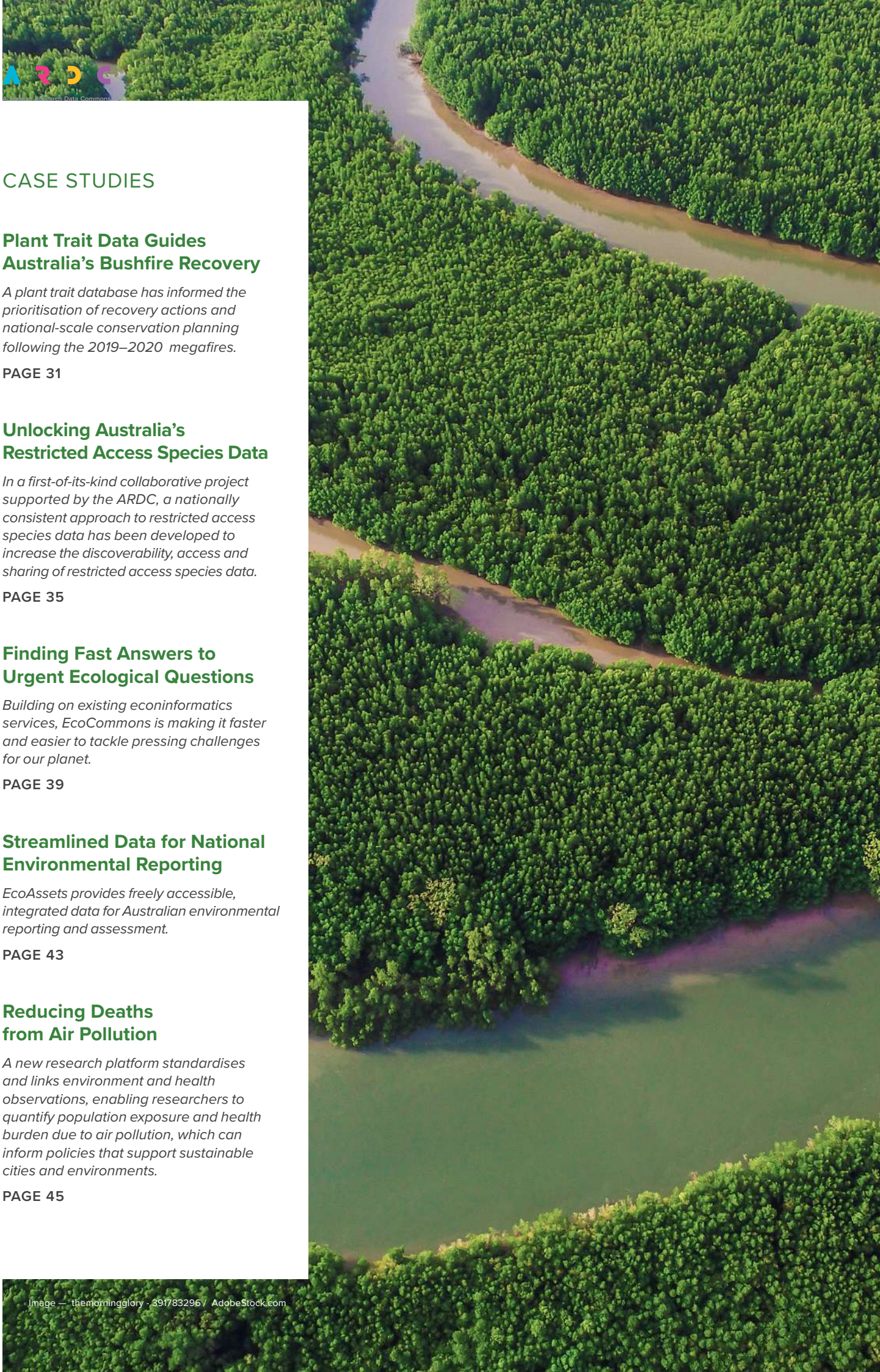


Image — themorningglory - 391783296 / AdobeStock.com

Image — jamenpercy - 318878058 / AdobeStock.com

Plant Trait Data Guides Australia's Bushfire Recovery

The catastrophic 2019–2020 Australian bushfires burnt more than 10 million hectares of the Australian landscape, a scale unprecedented in recent history. The fires impacted the habitat of 17,197 species of plants, which is 69 per cent of all Australian plant species[1]. Working with the Threatened Species Commissioner, scientists immediately began prioritising Australian Government funding to help plant and animal populations recover.

Plants hold the key to ecosystem recovery. Aiding the recovery of Australia's plant species ensures its unique biodiversity is sustained into the future. With more than 87 per cent of Australian plant species occurring nowhere else in the world, the impact of the 2019 and 2020 bushfires is of global significance for plant conservation[1].

But how do we quickly find out, on a national scale, which plant species recover well from fire? ▶

1 *Gallagher, RV, Allen, S, Mackenzie, BDE, et al. High fire frequency and the impact of the 2019–2020 megafires on Australian plant diversity. Divers Distrib. 2021; 27: 1166–1179. doi.org/10.1111/ddi.13265*

Image — Renee Nowytarger / ARDC

Dr Rachael Gallagher used AusTraits to strategically prioritise plant species for recovery actions and conservation planning at a national scale following the 2019 and 2020 bushfires.

> If we hadn't had AusTraits, which harmonises hard-won data on fire response from hundreds of Australian scientists who are our contributors, it would have been doubly difficult to be able to do the work for the Threatened Species Commissioner.

Dr Rachael Gallagher,
Associate Professor of plant conservation and ecology,
Western Sydney University

Understanding How Plants Respond to Fire is Vital

The first stop for Dr Rachael Gallagher, Associate Professor of plant conservation and ecology at Western Sydney University, and her colleagues was AusTraits[2].

Developed in partnership with the ARDC and 19 institutions, AusTraits is a national database synthesising data on 448 plant traits across more than 21,000 Australian plant species.

Using AusTraits to access data on plant fire response traits, Dr Gallagher and her colleagues were able to combine data on a plant's geographic range and fire response traits, fire extent and fire history to estimate the proportion of the species' extent that burnt in the past as well as in the 2019–2020 fires[1]. They then used this information to strategically prioritise the species for recovery actions and conservation planning at a national scale.

Two species that grow near Mount Imlay National Park in southeast New South Wales are now high priority species for recovery actions, thanks to the assessment. These are *Boronia imlayensis*, the Mount Imlay boronia, and *Eucalyptus imlayensis*, the Imlay Mallee.

"Synthesised national information about plant response to fire was vital. If we hadn't had AusTraits, which harmonises hard-won data on fire response from hundreds of Australian scientists who are our contributors, it would have been doubly difficult to be able to do the work for the Threatened Species Commissioner," said Dr Gallagher.

The research has led to plant species and ecosystems being newly listed as threatened under the *Environment Protection and Biodiversity Conservation Act 1999*. This will help ensure Australia's unique plant species recover, reproduce and thrive. ▶

Image — Mikeybear - 20110903 Boronia imlayensis 1.jpg / commons.wikimedia.org

2 *Falster DS, Gallagher RV, et al. AusTraits, a curated plant trait database for the Australian flora. Scientific Data. 2021;8:254. doi:10.1038/s41597-021-01006-6.*

Australian Research Data Commons

> The co-investment from the ARDC has helped make the international definitions of traits possible in two ways. One, it forces us to pay attention to vocabularies, which we probably would have overlooked otherwise, like most ecologists would. Two, the ARDC provided the skills to navigate the very complex landscape of formal vocabularies and definitions.

Dr Daniel Falster,
ARC Future Fellow, *UNSW Sydney*
Project co-leader, *AusTraits*

Image — Renee Nowytarger / ARDC

Dr Daniel Falster, UNSW Sydney, and Dr Hervé Sauquet, Royal Botanic Gardens and Domain Trust, from the AusTraits project team

A National Data Asset for Cutting-Edge Science and Decision Making

The AusTraits team worked with ecologists and herbariums across Australia to capture, harmonise and rigorously review data from over 300 sources. But understanding how a plant responds to fire is just one application of the database.

To date, over 50 scientific papers have cited AusTraits, and many more projects using the data are underway. AusTraits is facilitating studies in patterns of flowering, changes in traits with the environment, and patterns of richness across the landscape. It is also helping researchers devise next-generation models for simulating vegetation dynamics. ▶

Australian Research Data Commons

AusTraits – the International Gold Standard

AusTraits has also garnered attention outside Australia, being lauded at a recent international Open Traits Network meeting as the new gold standard on how to make a reproducible trait database. Key to its global success is the ARDC's co-investment model, said Dr Daniel Falster, ARC Future Fellow at UNSW Sydney and AusTraits project co-leader:

"The co-investment from the ARDC has helped make the international definitions of traits possible in two ways. One, it forces us to pay attention to vocabularies, which we probably would have overlooked otherwise, like most ecologists would. Two, the ARDC provided the skills to navigate the very complex landscape of formal vocabularies and definitions."

Following workshops with researchers to clearly define traits, the AusTraits team worked with experts from the ARDC Research Vocabularies Australia to establish a plant trait vocabulary.

AusTraits will soon be the first large traits database in the world to connect its trait concepts with those in all other major plant trait databases internationally. According to AusTraits Project Manager Dr Elizabeth Wenk, groups working on huge trait databases in Europe are already looking to AusTraits for transparent, defined, publishable and connected definitions.

AusTraits data is also being incorporated into the Atlas of Living Australia and the ARDC co-investment platform EcoCommons.

The AusTraits team is now working to apply the workflow building their database to organisms other than plants – and that is just the beginning.

"We envisage AusTraits as a framework for anyone who wants to build a database by putting lots of fragments of data together," said Dr Falster. ■

Members of the AusTraits project team. **Back:** Lily Dun, David Coleman, Sophie Yang, Dr Rachael Gallagher, Dr Elizabeth Wenk **Front:** Dr Daniel Falster, Dr Hervé Sauquet

Image — Renee Nowytarger / ARDC

AusTraits

AusTraits received co-investment (doi.org/10.47486/DP720) from the ARDC. The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).

AusTraits is a partnership between UNSW Sydney, Macquarie University, Royal Botanic Gardens and Domain Trust, Centre for Australian National Biodiversity Research, The University of Melbourne, Western Sydney University, EcoCommons, CSIRO, Landcare Research New Zealand, University of New England, Murdoch University, University of Arizona, Australian Biological Resources Study (Department of Agriculture, Water and the Environment), Parks Australia, Atlas of Living Australia, Arthur Rylah Institute for Environmental Research, NSW Department of Planning, Industry and Environment, Western Australia Department of Biodiversity, Conservation and Attractions, and Greening Australia.

Australian Research Data Commons

Unlocking Australia's Restricted Access Species Data

National-scale biodiversity data, including species location, quantity, behaviour and habitat interactions, are critical for researchers and decision-makers to understand, manage and conserve sensitive species. ▶

The princess parrot (*Polytelis alexandrae*) is a listed sensitive species. It is protected by law from capture.

> Collaborating with such a large group of partners was both challenging and rewarding. We couldn't have developed a successful national approach without engagement from the jurisdictions and across the biodiversity sector.
>
> **Cameron Slatyer,**
> Project Manager, National Biodiversity Initiatives (Atlas of Living Australia),
> *CSIRO*

Image — Nick Cairns - 146282857 / inaturalist.org

Sandalwood (*Santalum spicatum*), found in Australia's western and southern rangelands, is threatened by commercial harvesting for its fragrant, high-value timber.

Sometimes data needs to be withheld or restricted to reduce the risk of harm to species, protect personal information, ensure Indigenous data sovereignty, avoid breaching legal or confidentiality agreements or manage biosecurity risks. For example, data on the location of the princess parrot may be restricted as it is a target for illegal collectors.

Inaccessible Data Impedes Research and Decision Making

Biodiversity data has proliferated with the growth of big data and citizen science, as has the demand for easy access to the data.

The inability to access species data is a significant impediment to biodiversity analysis. About 45 per cent of Australia's biodiversity data is withheld or restricted. This makes tasks such as developing a national plan for threatened species or assessing cross-jurisdictional bushfire vegetation recovery challenging.

States, territories and other data custodians such as Birdlife Australia, private conservation organisations, researchers and companies hold restricted access species data. Currently, species data must be requested directly from custodians, making it difficult and time-consuming to discover and access.

Collaboration Is Key

The Atlas of Living Australia (ALA), with co-investment from the ARDC, together with 13 other partners — including all state and commonwealth jurisdictions, research infrastructure and other stakeholder groups — collaborated widely to develop a national framework for managing and sharing restricted access species data.

The first-of-its-kind project also consulted with the community, government organisations, research agencies and non-government and industry bodies to streamline processes and systems, better define restricted access species data, agree on guiding principles, and improve data sharing processes and tracking.

"Collaborating with such a large group of partners was both challenging and rewarding," said Cameron Slatyer, Project Manager - National Biodiversity Initiatives (Atlas of Living Australia), CSIRO. "We couldn't have developed a successful national approach without engagement from the jurisdictions and across the biodiversity sector." ▶

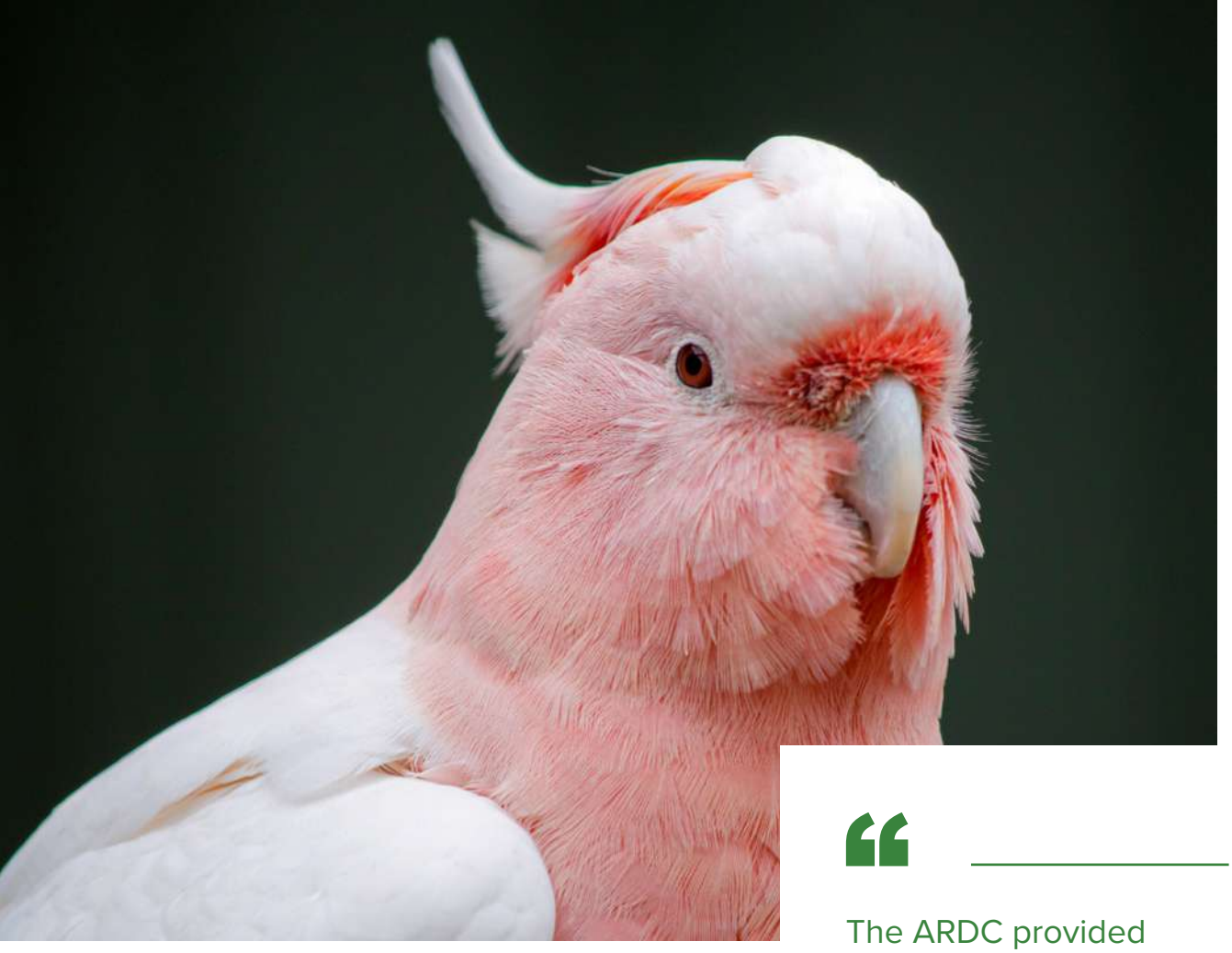


Image — susan flashman -495034070 / AdobeStock.com

The illegal trapping and the collection of nests for avian trade may threaten the Major Mitchell's cockatoo (*Lophochroa leadbeateri*).

A National Framework for Sharing Restricted Access Species Data

The national framework is a non-legally binding document providing approaches and principles to safeguard sensitive species data, enable secure sharing, offer clauses to legal agreements, and ensure nationally consistent methods for modifying restricted access data for public release.

Project partners developed the framework through ongoing and broad consultation. It has undergone a FAIR audit (Findability, Accessibility, Interoperability, and Reusability of digital assets) and includes CARE principles (Collective benefit, Authority to control, Responsibility and Ethics) for Indigenous data governance.

"BirdLife Australia has long upheld a restricted access policy for sensitive species data as a basis for our citizen science work. This gives contributors the confidence that sharing their data will not put species at risk, while still allowing for its use in conservation. We are excited to now have a nationally consistent framework for sensitive species data sharing, which will massively increase the use and impact of bird and other species data in science and conservation", said Dr Golo Maurer, BirdLife Citizen Science Program Leader.

> The ARDC provided valuable input on FAIR principles, access to communities of practice to mine ideas, and advice on cross-cutting issues needed to develop the framework. This was critical to the success of the project.
>
> **Cameron Slatyer,**
> Project Manager, National Biodiversity Initiatives (Atlas of Living Australia),
> *CSIRO*

A Secure Way of Using Restricted Access Species Data

To ensure the security of restricted access species data, the project developed a password-secured environment where data can be displayed and interrogated without being downloaded. Led by ARDC-supported project partner EcoCommons, this functionality has now been built into the new EcoCommons platform. ▶

Australian Research Data Commons

A Data Service for Accessing Restricted Access Species Data

A data-sharing service was also developed by some of the project partners to streamline requests for restricted access species data. The data service does not hold or aggregate data. Instead, it connects users with data custodians using the service, provides an audit pathway to track request progress, captures metadata and provides a DOI for each request to ensure it is discoverable and reproducible.

"We developed an agreed nationally consistent set of data fields to enable the processing of requests at a national scale", said ALA policy officer Tania Laity. "The data service also allows users to request data from multiple custodians simultaneously."

The project has developed approaches that suit government and non-government data users and custodians. It avoids duplicating existing processes and supports the data sovereignty of custodians.

Image — Ken Griffiths - 495034070 / AdobeStock.com

The southern corroboree frog (*Pseudophryne corroboree*) is threatened by habitat damage by feral species, climate change and the chytrid fungus.

> We developed an agreed nationally consistent set of data fields to enable the processing of requests at a national scale. The data service also allows users to request data from multiple custodians simultaneously.

Tania Laity,
Policy officer,
Atlas of Living Australia

Leading the Way

Improved access, discoverability and sharing of restricted access species data benefits researchers, governments and industry. The result is better decision making about the sustainable management, conservation and recovery of Australia's sensitive and threatened species.

The collaborative and practical approach used to develop the national framework and service has not gone unnoticed. The Global Biodiversity Information Facility — an international network and data infrastructure — has recognised the value of the approach taken in this project, using learnings to improve the discoverability of international restricted access species data.

The ARDC's Planet Research Data Commons will also champion these approaches to make government and industry environmental data more accessible for research. ■

The project received co-investment (doi.org/10.47486/PS027) from the ARDC, the Atlas of Living Australia and CSIRO. The project was led by the Atlas of Living Australia, in partnership with the ARDC, CSIRO, NSW Department of Planning, Industry and Environment, NT Department of Environment and Natural Resources, the ACT Environment, Planning and Sustainable Development Directorate, WA Department of Biodiversity, Conservation and Attractions, SA Department of Environment and Water, Victorian Department of Environment, Land, Water and Planning, Tasmanian Department of Primary Industries, Parks, Water and Environment, QLD Department of Environment and Science, Australian Government Department of Agriculture, Water and the Environment, EcoCommons Australia, Pawsey Supercomputing Centre, The Western Australian Biodiversity Science Institute, and the Council of Heads of Australian Faunal Collections.

Image — Nautilus Creative - 551247362 / AdobeStock.com

Predictions about the migratory patterns of the yellowtail kingfish (*Seriola lalandi*) have been made using a machine-learning-based workflow available on EcoCommons.

Finding Fast Answers to Urgent Ecological Questions

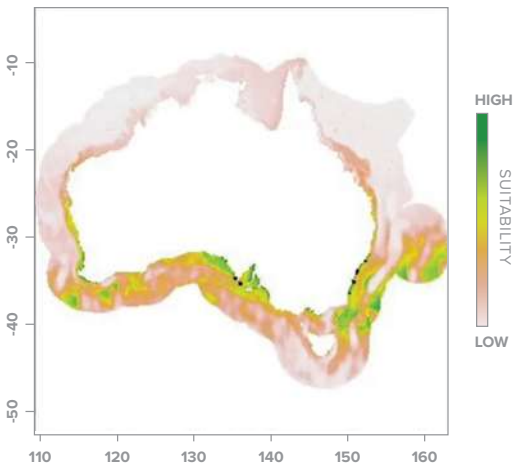
In the depths of Australia's coastal waters, hundreds of acoustic receivers are monitoring the movement of more than 100 fish, reptile and mammal species. This network of trackers was first built in 2007 by the Integrated Marine Observing System (IMOS), and to date it has made over 100 million detections. This wealth of data promises insights into many important questions, such as how climate change affects the animals' migration patterns and how they should be protected. The challenge is to gain these insights quickly and easily.

A solution came on 29 November 2022, when EcoCommons Australia, an ARDC-supported platform for analysing and modelling ecological data, was launched at the joint conference of the Ecological Society of Australia (ESA) and the Society for Conservation Biology Oceania (SCBO) in Wollongong. ▶

We were thrilled to release EcoCommons as a successor to BCCVL but with workflows and visualisations that are so much faster and smoother. And we've taken great care to update and expand the curated data collections which are now available at the fingertips of the EcoCommons modeller.

Dr Elisa Bayraktarov,
Program Manager,
EcoCommons

YELLOWTAIL KINGFISH (*SERIOLA LALANDI*)
PREDICTED DISTRIBUTION IN OCTOBER



Source — EcoCommons

The predicted distribution of the yellowtail kingfish in October based on IMOS acoustic detection data and averaged IMOS environmental data (2011–2021). This was produced using models on EcoCommons, and improvements are coming soon to the way users can model species migration or distribution over time.

Making the Good Better

Developed with co-investment from the ARDC and partners, and hosted on the ARDC Nectar Research Cloud, EcoCommons is the result of consolidating and re-engineering several existing ecoinformatics services. These include the Biodiversity and Climate Change Virtual Laboratory (BCCVL) and Ecocloud, which have powered environmental science for almost a decade. Using BCCVL, international as well as Australian researchers and policymakers could easily model biodiversity in different climate scenarios to understand, say, the changing distribution of alpine wildflowers in the Snowy Mountains[3] or of yellow-fever-carrying monkeys in Brazil[4] as the climate warms.

With EcoCommons, all this and more can be done on a single platform. It offers access to more high-quality ecological, environmental and climate data through its Data Explorer, expanded data visualisation and analysis capabilities through the new BCCVL Modelling Wizard, and greater computational power through the Coding Cloud. It is now faster and easier than ever to answer pressing questions about the environment.

Two of the species tracked by IMOS were the yellowtail kingfish (*Seriola lalandi*) and the bull shark (*Carcharhinus leucas*). Based on some 20,000 movement records for each species, predictions have been made about their migratory patterns using a machine-learning-based workflow available on EcoCommons. Predictions like these are exactly what is needed to design conservation measures such as fishing regulations, and to understand how the species may fare in the ongoing climate crisis.

"We were thrilled to release EcoCommons as a successor to BCCVL but with workflows and visualisations that are so much faster and smoother," said Dr Elisa Bayraktarov, EcoCommons Program Manager. "And we've taken great care to update and expand the curated data collections which are now available at the fingertips of the EcoCommons modeller." ▶

3 *Kirchhoff CM. Rapid assessment of future habitat suitability: A case-study of the Snowy Mountains endemic alpine flora using the Biodiversity and Climate Change Virtual Library (BCCVL). Australasian Plant Conservation: Journal of the Australian Network for Plant Conservation. 2020;28(3):5-8. Available from: https://search.informit.org/doi/10.3316/informit.093463002457583.*

4 *de Thoisy B, Silva NIO, Sacchetto L, de Souza Trindade G, Drumond BP. Spatial epidemiology of yellow fever: Identification of determinants of the 2016-2018 epidemics and at-risk areas in Brazil. PLOS Neglected Tropical Diseases. 2020;14(10):e0008691. Available from: https://doi.org/10.1371/journal.pntd.0008691.*

Australian Research Data Commons

GREATER SOOTY OWL (*TYTO TENEBRICOSA*)

ESTIMATED EXTENT OF SUITABLE HABITAT AFFECTED BY BUSHFIRE IN 2019/2020

Source — EcoCommons

On the Biodiversity and Climate Change Virtual Laboratory (BCCVL) – now part of EcoCommons and more powerful than ever – various algorithms are available for researchers to estimate the full extent of habitat suitable for the greater sooty owl (in yellow). The results can then be overlaid with external fire data to yield an estimate of the extent of the habitat affected by the 2019/2020 bushfires.

Image — Clarissa Human - 514123668 / AdobeStock.com

Greater Than the Sum of Its Parts

EcoCommons is also helping to maximise the potential of other ecoinformatics platforms and data sources. With EcoCommons, researchers can, for instance, make better use of the growing body of audio recordings of animal calls on Open Ecoacoustics, also an ARDC co-investment platform, to model species distribution and to understand how species compositions change over time and space.

Meanwhile, the EcoCommons team is working to integrate the platform with the open-source, ARDC-supported AusTraits, which holds plant traits data for more than 21,000 Australian species. This data can be plugged into models on EcoCommons that aim to predict how plant species ranges may shift under climate change, the understanding of which can be crucial for conservation planning.

Speaking at the EcoCommons launch, Professor Hugh Possingham, the former Queensland chief scientist, shared his vision for EcoCommons:

"I dream of a day when Indigenous people and fishermen and mining companies are sitting around a laptop in Mullumbimby deciding how to plan land-sea reserves; equitably discussing the data and the information and the cultural knowledge that needs to go into those land-sea reserves. And that, I think, is what EcoCommons will hopefully deliver."

Building Digital Research Skills in Environmental Sciences

EcoCommons is also building on the ecoscience training and skills development initiative EcoEd, providing users with peer-reviewed educational materials, real-world use cases of its tools, and guides to using the platform. The EcoCommons team also delivered in-person workshops at the 2022 ESA-SCBO conference and the 2022 Queensland Research Bazaar (ResBaz).

"There is a lot of pressure on researchers and policymakers to come up with fast solutions for the environment," said Dr Rob Clemens, former EcoCommons change and communications manager. "We want to empower our users to become the next generation ecological modellers who can utilise the whole suite of data, models and cloud compute." ■

> There is a lot of pressure on researchers and policymakers to come up with fast solutions for the environment. We want to empower our users to become the next generation ecological modellers who can utilise the whole suite of data, models and cloud compute.
>
> **Dr Rob Clemens,**
> former change and communications manager,
> *EcoCommons*

EcoCommons

EcoCommons Australia (doi.org/10.47486/PL108) is a partnership of the ARDC, the Atlas of Living Australia, CEBRA at The University of Melbourne, CSIRO, Griffith University, Macquarie University, the Queensland Cyber Infrastructure Foundation, TERN, and UNSW Sydney. It has also received investment from the Queensland Government's Research Infrastructure Co-investment Fund.

Australian Research Data Commons

Streamlined Data for National Environmental Reporting

Each year, large amounts of data on all aspects of the Australian environment, from our oceans to our deserts, are collected using national environmental research infrastructure.

For the data to be suitable for environmental reporting and assessment at national or state/territory levels, they must be integrated and be at a certain scale and in a certain format.

To achieve this, 4 of Australia's national research infrastructure teams worked together to develop EcoAssets, a collection of open, integrated data from 3 National Collaborative Research Infrastructure Strategy (NCRIS) facilities — the Atlas of Living Australia (ALA), Integrated Marine Observing System (IMOS) and Terrestrial Ecosystem Research Network (TERN) — with co-investment and expertise from the ARDC.

By making it simpler and faster to obtain integrated data, EcoAssets is enhancing government reporting and supporting better decision making about the environment. ▶

Images — WeedZard - 583034380 / AdobeStock.com | Ian R McCann - species/8372 / museumsvictoria.com.au | Peripitus - Themeda triandra - kangaroo grass.jpg / commons.wikimedia.org

Breaking New Ground for State of the Environment Reporting

Data and maps from EcoAssets were used in the 2021 State of the Environment report, which is an independent, comprehensive and evidence-based assessment of the state of Australia's environment.

Dr Kristen Williams, a CSIRO principal research scientist and a lead author of the land chapter of the report, said that the EcoAssets project "broke new ground" in national-level reporting on land pressures related to introduced and invasive species.

"The aggregation of introduced species location data across time and space, combined with information on invasiveness, enabled reporting on status and trends with unprecedented granularity," she said.

"When further combined with contextual data, such as land-use zones and bioregions, a wide range of interpretive products could be developed. These outputs contributed to the multiple lines of evidence used to inform assessment summaries in both the land and biodiversity chapters, and key findings in the overview chapter."

EcoAssets biodiversity data is also informing the 2023 State of the Forests, to be published by the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). The authors are using the data to explore species' dependence on forests at national and state scales.

Freely Available Data

All 7 datasets from EcoAssets are now publicly and freely available for biodiversity reporting and assessment. The primary biodiversity data asset is the aggregated Australian species occurrence data asset, which enables users to explore the data by threatened species status, bioregion, and protected area category, among others, for both terrestrial and marine systems. A future release will include an environmental monitoring and observation summary dataset.

More than 120 users have registered to access EcoAssets, most of them coming from government (all levels), industry, the university and education sector, community-based organisations and not-for-profit organisations. In the 9 months following the launch of EcoAssets in mid-2022, its datasets were downloaded over 400 times.

> The aggregation of introduced species location data across time and space, combined with information on invasiveness, enabled reporting on status and trends with unprecedented granularity.

Dr Kristen Williams,
Principal Research Scientist,
CSIRO

Data Pipelines Are Ready for New Data

By better aligning the data, the research infrastructure teams have created not only new datasets, but also data pipelines to ensure that each data asset can be updated at regular intervals. This should speed up future reporting and assessment.

EcoAssets data is also available via ARDC Research Data Australia, data.gov.au, and EcoCommons, another ARDC co-investment project. ■

EcoAssets

EcoAssets received co-investment (doi.org/10.47486/XN005) from the ARDC. It is led by the Atlas of Living Australia in partnership with the ARDC, TERN, IMOS and the Australian Government Department of Climate Change, Energy, the Environment and Water.

Image — timalenphoto - 213805942 / stock.adobe.com

Reducing Deaths from Air Pollution

Policy decisions that aim to reduce human-caused air pollution, and the resulting deaths, need integrated data about the environment and mortality. A new platform to be launched in mid-2023 brings together air pollution and health data for Australia for the first time.

A 2020 study by Curtin University's Dr Ivan Hanigan and colleagues[5] found that every year about 2,600 Australians die from human-caused air pollution — including from particulate matter, carbon monoxide, ozone and ultrafine particles. Air pollution also reduces life expectancy, with costs of about $6.2 billion a year.

To reduce the impact of these emissions on people's health, policymakers need to make evidence-based decisions to reduce air pollution in areas such as urban planning and planning prescribed burns for bushfire management. Recent ARDC-supported research has improved access to air pollution and health data and tools to better quantify how environmental changes affect people's health. Armed with this knowledge, environment and health agencies can better understand the likely impact of proposed interventions to address the problem. ▶

5 *Hanigan IC, et al. Avoidable mortality attributable to anthropogenic fine particulate matter ($PM_{2.5}$) in Australia. Int J Environ Res Public Health. 2021;18(1):254. doi:10.3390/ijerph18010254.*

A major hurdle to more accurate quantification of the health burden due to air pollution is obtaining appropriate mortality data, which can be difficult to access even for governments. The other challenge is making those data available so that it becomes knowledge, which can in turn be used practically by policymakers to benefit society.

Dr Aditya Vyas,
Deputy Director, WHO Collaborating Centre for Climate Change and Health Impact Assessment,
Curtin University

Getting an Accurate Picture of Mortality

Part of the challenge is the restricted access to health data at a local level. For ethical reasons, people's health records are protected by several layers of data security and confidentiality requirements.

Also, air pollution reporting standards vary by state and territory. And it's difficult to consolidate and integrate air pollution and health datasets for use in computer models that quantify air pollution exposures — which is what is needed to inform policy changes and reduce human health impacts.

"A major hurdle to more accurate quantification of the health burden due to air pollution is obtaining appropriate mortality data, which can be difficult to access even for governments," said Dr Aditya Vyas, Deputy Director of the WHO Collaborating Centre for Climate Change and Health Impact Assessment at Curtin University. "The other challenge is making those data available so that it becomes knowledge, which can in turn be used practically by policymakers to benefit society."

Establishing a National Platform

With support from the ARDC, Dr Hanigan from Curtin University and Associate Professor Geoff Morgan from The University of Sydney coordinated a project to develop the Integrated National Air Pollution and Health Data platform.

Launching in mid-2023, the platform brings together data about the environment, pollution, health and mortality across time, place, type and source. It gives researchers access to restricted but de-identified health data to estimate risk and numbers of deaths attributable to human-caused air pollution. Importantly, researchers and decision-makers can see the likely impact on mortality of reducing these emissions. This information can inform changes in economic and environmental policies to create sustainable cities and environments.

The data platform builds on successful collaborations between the Centre for Air pollution, energy and health Research (CAR) and the Healthy Environments And Lives (HEAL) national research network, which both received funding from the NHMRC. The data platform extends the CAR Data and Analysis Technology (CARDAT) project. Partners include Curtin University, the NSW Department of Planning and Environment, EPA Victoria, the Australian Institute of Health and Welfare, Monash University, The University of Queensland and The University of Sydney.

The ARDC-supported project team worked closely with researchers and the Australian Institute of Health and Welfare to aggregate and summarise confidential mortality data for the new platform. The summarised data enables more accurate quantification of the health burden due to air pollution while maintaining the confidentiality of people's data.

The platform standardises past and current air pollution observations from around Australia, including for particulate matter, carbon monoxide, ozone and nitrogen dioxide. These data feed into epidemiological modelling of air pollution exposure and mortality.

"The ARDC encouraged everyone to be more systematic," said Dr Hanigan. "With their support, the research and policy translation community can reduce the time taken to acquire relevant health data for standardised processes such as air pollution health impact assessment, while maintaining health data confidentiality and security. The creation and storage of data on a platform will reduce time spent on data grunt work." ▶

City Heat and Death

As an example, the new platform has been used to find ways to reduce heat-related deaths in cities by reducing the so-called urban heat-island effect.[6]

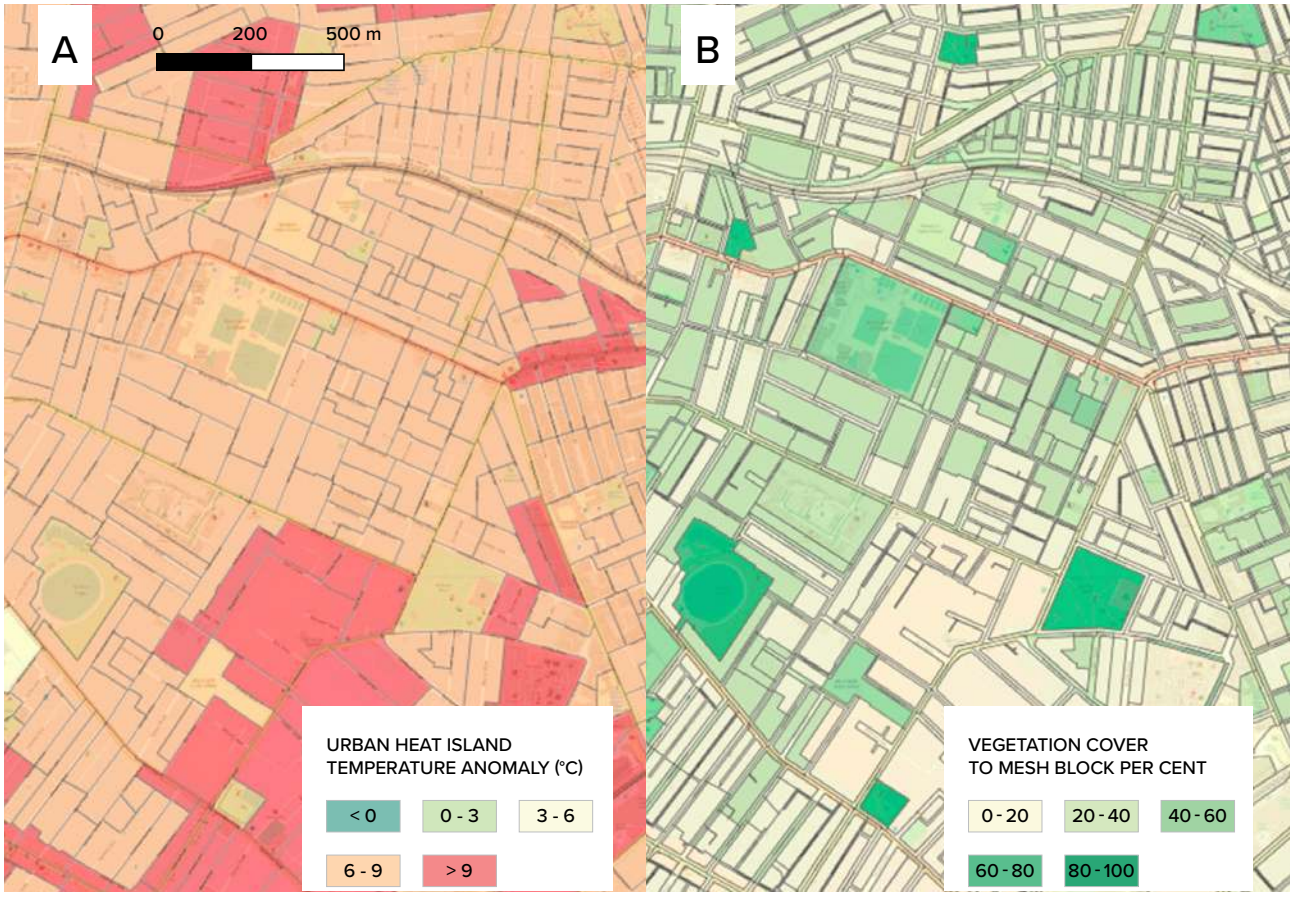
Air temperature observations aggregated from a range of agencies, combined with modelled results of the excess heat from cities, can be linked with downscaled health data and seasonal trends in death rates. From this, the number of deaths attributable to the urban heat-island effect can be estimated. This number can be projected forward to show the likely future mortality from heat in cities, under continued global warming.

An added bonus, said Dr Hanigan, is that "through the platform, we can share the results, heat maps, and other analyses without requiring IT support or expensive software." ▶

> The platform's standardisation and improved accessibility of air pollution and health data is a huge step forward.

Dr Ivan Hanigan,
Senior Lecturer, Curtin School of Population Health,
Curtin University

Urban heat island temperature anomaly (A) and % vegetation cover (B)



Source — Adapted from Chaston et al. [6]

6 *Chaston TB, Broome RA, Cooper N, Duck G, Geromboux C, Guo Y, et al. Mortality burden of heatwaves in Sydney, Australia is exacerbated by the urban heat island and climate change: Can tree cover help mitigate the health impacts? Atmosphere [Internet]. 2022 Apr 30;13(5):714. Available from: http://dx.doi.org/10.3390/atmos13050714*

Image — VMJones - 947621710 / istockphoto.com

Rural Drought and Suicide

Using the new platform to analyse data about drought exposure combined with the number of rural suicides has already revealed an increase in rural suicides during drought.

"We found, on average each year, 1.8 per cent of suicides among men in Australia's rural communities can be attributed to drought," said Dr Vyas.

"Under climate change, we will see more drought in Australia, with models suggesting drought-related suicides will increase to 3.3 per cent."

These results suggest an urgent need for actions to adapt to climate change and drought to avoid impacts on vulnerable communities.[7]

> Under climate change, we will see more drought in Australia, with models suggesting drought-related suicides will increase to 3.3 per cent.
>
> **Dr Aditya Vyas,**
> Deputy Director, WHO Collaborating Centre for Climate Change and Health Impact Assessment,
> *Curtin University*

Near-Real-Time Health Warnings

Dr Hanigan envisages the platform will enable a more resilient society.

"This technology will enable people to rapidly map the impacts of climate hazards such as bushfires, heatwaves and floods, and potentially pre-empt them to give near-real-time health warnings," he said.

"And the platform's ability to show the impact of data on decisions may lead to more proactive evidence-based policies that help mitigate future climate change and justify investment in mitigation activities."

Dr Vyas agrees. "Outputs from this will be considered by policymakers responding and adapting to climate change," he said. "In addition, the international focus of our work means we can translate the infrastructure, methods, tools and technology for other countries, to have even more impact." ■

This project received co-investment (doi.org/10.47486/PS022) from the ARDC. It is led by Curtin University in partnership with The University of Sydney, Monash University, The University of Queensland, the Australian Institute of Health and Welfare, the NSW Department of Planning and Environment, the Centre for Air pollution, energy and health Research, and EPA Victoria.

7 *Hanigan IC, Chaston TB. Climate change, drought and rural suicide in New South Wales, Australia: Future impact scenario projections to 2099. International Journal of Environmental Research and Public Health 2022;19:7855. DOI: 10.3390/ijerph19137855.*

Australian Research Data Commons

HUMANITIES, ARTS AND SOCIAL SCIENCES (HASS) AND INDIGENOUS RESEARCH DATA COMMONS

National-scale data infrastructure for HASS and Indigenous research

In collaboration with Indigenous Australians, the research community, industry and government, the HASS and Indigenous Research Data Commons (HASS and Indigenous RDC) is harnessing research data to enhance Australian social and cultural wellbeing, and help Australia understand and preserve our culture, history and heritage.

New digital platforms and data directories will improve how researchers discover and access Australia's rich HASS and Indigenous data and innovative analysis tools. The program is also upskilling researchers to use data-driven approaches to HASS and Indigenous research, ensuring Australian researchers gain a competitive advantage through data.

Projects of the HASS and Indigenous RDC include:

- Language Data Commons of Australia
- Trove Researcher Platform
- Integrated Research Infrastructure for Social Sciences
- Improving Indigenous Research Capabilities
- HASS Community Data Lab
- Gazetteer of Historical Places Names.

As an engine for research translation, the HASS and Indigenous RDC will enable researchers to develop and sustain cross-sectoral and multidisciplinary data collaborations at a national scale through federated models. It integrates the ARDC's services for compute, storage infrastructure, persistent identifiers and data discovery with analysis platforms and tools that are supported by expertise, standards and best practices.

Learn more about the impact and community engagement of the HASS and Indigenous RDC on the coming pages.

Be part of the HASS and Indigenous RDC. *Visit - bit.ly/HASSRDC*

CASE STUDIES

Empowering HASS and Indigenous Researchers with Essential Computational Skills

In February, close to 100 early-career researchers, librarians and students came together on Gadigal Country (Sydney) to upskill in digital research methods and tools.

PAGE 51

Introducing the ARDC's First Indigenous Intern, Ms Lisa Rigney

Lisa is assigned to the Indigenous Data Network at The University of Melbourne.

PAGE 55

Telling Human Stories with Time-Layered Cultural Maps

Patterns in HASS and Indigenous research data are being revealed and visualised using the Time-Layered Cultural Map and the Gazetteer of Historical Australian Places.

PAGE 58

Tweets Illuminate the Impact of COVID-19 on Society

The Australian Text Analytics Platform is helping researchers analyse large volumes of text to enhance our understanding of the impact of major events, such as the COVID-19 pandemic, on society.

PAGE 62

Minding Our Language Data

Scattered language data can now be found and accessed in a single place — the Language Data Commons of Australia.

PAGE 66

Image — Mario Purisic - jG1z5o7NCq4 / unsplash.com

Image — Renee Nowytarger / ARDC

Almost 100 early-career researchers, librarians and students participated in the ARDC HASS and Indigenous RDC Computational Summer School in February 2023.

Empowering HASS and Indigenous Researchers with Essential Computational Skills

An ARDC-run computational skills summer school was held on Gadigal Country (Sydney) in February 2023 for researchers from academia and from Aboriginal community-controlled organisations. The aim was to uplift their digital research skills in data-intensive research and decision making.

Almost 100 early-career researchers, librarians and students participated. More than 20 experts conducted training on how to use the research infrastructure that is being created in the HASS and Indigenous Research Data Commons for participants' research.

A stream specifically for Aboriginal community-controlled organisations was presented by Lisa Rigney from the ARDC and Levi-Craig Murray from the Indigenous Data Network at The University of Melbourne.

Popular skills development workshops included:

- data and governance
- preparing and analysing unstructured data
- managing sensitive data
- analysing spatial and temporal data
- analysing social and cultural data.

Sessions were also run on the relationship between data management and the importance of language and awareness to Country and cultural beliefs, ensuring respect and inclusion in decisions about collecting and managing data. ▶

Summer School Participant Experiences

The ARDC sponsored 10 travel bursaries for Aboriginal and Torres Strait Islander researchers and students, early-career researchers and higher-degree research students to attend the summer school. Here are some of their insights and takeaways. ▶

> In my experience, there are little-to-no opportunities for HASS researchers to build their computational skills in a general, collaborative setting.
>
> The session on Jupyter Notebooks was most valuable for my research. The introduction to the ARDC Nectar Research Cloud was very informative and will certainly steer my future research. Tully Barnett's overview of the digital humanities landscape was also very encouraging.
>
> The range of participants, in terms of discipline, research/industry and career, was refreshing and made for rich and diverse conversation.
>
> **Charbel El-Khaissi,**
> PhD student in linguistics,
> *Australian National University*

> I took great interest in our discussion around FAIR and CARE, and how we can work with our Indigenous communities to ethically, safely and respectfully collect data for purposeful research.
>
> **Annie Sowter**

> The introduction to the ARDC Nectar Research Cloud was very informative and will certainly steer my future research. Tully Barnett's overview of the digital humanities landscape was also very encouraging.
>
> **Charbel El-Khaissi**

> I was fortunate to attend the Indigenous stream, hosted by Levi-Craig Murray. I found his sessions incredibly informative – from explaining types of data, contextualising information and metadata, and beginning to explore research strategies.
>
> I took great interest in our discussion around FAIR and CARE, and how we can work with our Indigenous communities to ethically, safely and respectfully collect data for purposeful research.
>
> Cultural safety is such an important part of my role, and it was interesting to see how much this presents in research strategies. I particularly enjoyed our conversations around implicit bias in research, and the impact this has on overall data analysis and outcomes for research studies.
>
> "It was fascinating to learn how research data works its way into each individual's role, regardless of their profession.
>
> **Annie Sowter,**
> Aboriginal pathway worker,
> Koolin Balit Babaneek Booboop Early Years Project,
> *Western Health*

Robert McLellan (left) and Christie Mancktelow (right) at the Computational Skills Summer School

> I attended to learn more about the Indigenous Data Catalogue and how to use it, and to connect with others working in Indigenous research.
>
> I found both sessions with the Indigenous Data Network team to be valuable. The discussions generated were enriching.
>
> **Christie Mancktelow,**
> Higher degree by research student,
> *CQUniversity*

Yolanté Jones, Community Coordinator for the ARDC-supported CADRE project at the Australian National University presents the Five Safes framework.

Uncle Michael West gives the Welcome to Country.

All Images — Renee Nowytarger / ARDC

Levi-Craig Murray, Indigenous Data Network at The University of Melbourne, co-led the stream for Aboriginal community-controlled organisations.

> I joined Levi-Craig Murray in the Indigenous data stream, where we workshopped a research question for an issue raised by 2 women from Cherbourg about the lack of resources for treating rheumatic heart disease. This was powerful for many reasons. Through the discussion, we all gained insight into the enormity of the issues in this community.
>
> The summer school was a great way to meet people who have experience working across many sectors, sharing first-hand stories of the fraught nature of data and data sovereignty.
>
> **Frances Silberstein,**
> **Master of Research candidate,**
> *Curtin University*

Hana Hamilton (left) and Frances Silberstein (right) at the Computational Skills Summer School

> Listening to Levi-Craig Murray during the Indigenous Data Network group deepened my understanding of Aboriginal health issues and, more broadly, the perspectives of Aboriginal community members and academics/ researchers. I will be keeping these views close to me in my future work.
>
> I really feel that I made some valuable and lasting connections.
>
> **Hana Hamilton,**
> **PhD candidate, School of Humanities, Languages and Social Science,**
> *Griffith University*

Introducing the ARDC's First Indigenous Intern, Ms Lisa Rigney

In August 2022, Lisa Rigney joined the ARDC as our first Indigenous intern, assigned to the Indigenous Data Network at The University of Melbourne.

Image — Anthony McKee / ARDC

A proud Ngarrindjeri, Talkindjeri and Ramindjeri woman from Raukkan (formerly known as Point McLeay), she is also a Kaurna woman from the Adelaide Plains area and a Boandik woman from Mount Gambier.

Lisa moved from Kaurna Country (Adelaide) to take up the internship as Project Support Officer, Indigenous Data. When it ends in August 2023, she will remain with the Indigenous Data Network as an education coordinator, facilitating culturally safe and appropriate ongoing training and development opportunities in Indigenous data and governance.

Levi-Craig Murray, who works alongside Lisa at the Indigenous Data Network, said, "The commitment of the ARDC to create and foster opportunities such as the Indigenous Internship is what changes the trajectory for individuals and their communities.

"For Lisa, who is already a senior and accomplished educator, the internship has provided a platform for two-way learning, not only within institutional settings, but also, critically, into the Aboriginal and Torres Strait Islander communities that we come from and serve.

"As an intern, Lisa's insights have ensured that the work undertaken within the Improving Indigenous Research Capabilities project positions and elevates Indigenous voices."

Jenny Fewster, Director of the ARDC HASS and Indigenous RDC, said, "The ARDC's Indigenous Internship program is a vital component of our program to uplift digital research skills in the Indigenous research sector. At the same time as we are creating powerful digital research tools and platforms for the HASS and Indigenous research sectors, we are also ensuring we upskill researchers to use the technology. Lisa's contributions to the ARDC this past year have been considerable, and we thank her for her dedication to elevating Indigenous research capability. We are committed to continuing our internship program and look forward to welcoming our next Indigenous intern in August 2023." ▶

Lisa's contributions to the ARDC this past year have been considerable, and we thank her for her dedication to elevating Indigenous research capability. We are committed to continuing our internship program and look forward to welcoming our next Indigenous intern in August 2023.

Jenny Fewster, Director, HASS and Indigenous Research Data Commons, *ARDC*

Australian Research Data Commons

We spoke with Lisa about her experience at the ARDC.

How has your internship impacted your professional growth?

I have a dream and goal to one day become a CEO of an Aboriginal community-controlled organisation. There are so many things I need to consider before becoming a CEO, such as both qualifications and on-the-job experiences and opportunities.

Last year I completed a Diploma of Business Administration / Diploma of Leadership and Management, and I am now studying for my Master of Business Administration at the Australian Institute of Management. I also completed the Introduction to Aboriginal and Torres Strait Islander Research through the Lowitija Institute, which has given me skills and knowledge to become a researcher, but most importantly an Aboriginal researcher.

[Alongside my goal to become a CEO], I am interested in pursuing a career in Indigenous research, maybe doing my PhD in Aboriginal genealogies in South Australia or Indigenous education.

My ARDC Indigenous internship has prepared me to work within Indigenous research and governance, leadership and management, equipping me with project management skills and knowledge. ▶

Image — Renee Nowytarger / ARDC

Lisa Rigney (left) co-led a stream at the ARDC HASS and Indigenous RDC Computational Summer School.

> [Alongside my goal to become a CEO], I am interested in pursuing a career in Indigenous research, maybe doing my PhD in Aboriginal genealogies in South Australia or Indigenous education.
>
> **Lisa Rigney,**
> Project Support Officer (Indigenous Data), *ARDC*

Image — Lisa Rigney / ARDC

Lisa Rigney (centre) with her colleagues Jo Savill (ARDC), Levi-Craig Murray (Indigenous Data Network), Jenny Fewster (ARDC), Kristen Smith (Indigenous Data Network)

What have been the most memorable experiences during your internship?

The ARDC workshops have been a great experience for me, including the ARDC HASS and Indigenous RDC Symposium, the Computational Summer School and the ARDC Digital Research Skills Summit. I was able to learn new things to improve my practice as a project support officer.

One of the most special memorable experiences in my internship was working under and alongside the country's most influential and strong Black woman, Professor Marcia Langton AO (The University of Melbourne's Associate Provost, Foundation Chair of Australian Indigenous Studies, Redmond Barry Distinguished Professor, Director of the Indigenous Studies Unit). I see her dedication and commitment to getting a voice for our people. I am in awe of her work and how passionate she is, like me, about her people and culture.

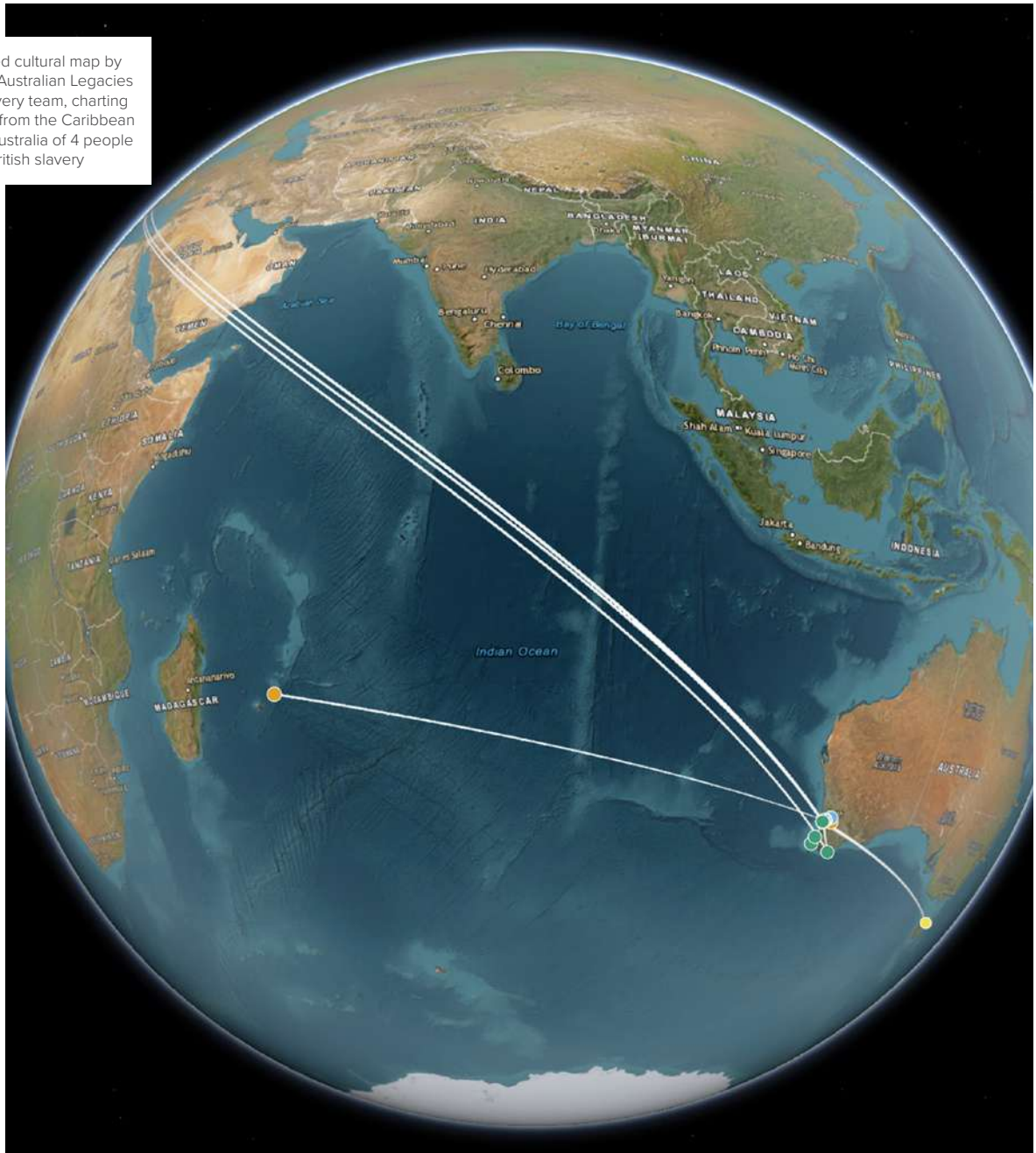
Another memorable experience has been working with Levi-Craig Murray and the team at The University of Melbourne. I have thoroughly enjoyed working in an inclusive and culturally safe work environment. We have a deadly team. [I am] forever grateful, thankful and appreciative.

I am particularly proud of my contribution to the ARDC HASS and Indigenous RDC Computational Summer School. Both Levi and I prepared a comprehensive, culturally appropriate program on Indigenous data, research and governance. We got a lot of positive feedback from everyone who attended, and we have been invited to present the same workshop with CSIRO. ■

Australian Research Data Commons

Telling Human Stories with Time-Layered Cultural Maps

On 14 August 1829, James Walcott and Charles Dawson Ridley set sail from Portsmouth, England. Walcott had had a financial interest in several slave plantations in Demerara (now part of Guyana), and Ridley was the administrator of several others. But this time they were not travelling to the Caribbean. Bringing their money, servants and equipment, they were bound for Western Australia. ▶



A time-layered cultural map by the Western Australian Legacies of British Slavery team, charting the journeys from the Caribbean to Western Australia of 4 people with ties to British slavery

Source — Western Australian Legacies of British Slavery

Combining other historical records with the location data from GHAP [the Gazetteer of Historical Australian Places] is helping the team identify which Victorian pastoral estates were indeed connected to Caribbean slave plantations and, conversely, when shared names are just coincidental.

Professor Zoë Laidlaw,
School of Historical and Philosophical Studies,
The University of Melbourne

Walcott and Ridley were part of a wave of British colonisers divesting from slavery, which was becoming economically and morally unsustainable, and turning their attention to settling in Britain's new colonies in the early 19th century. Most likely using funds drawn from their estates in Demerara, they would claim and reshape what is now the suburb of Caversham in Perth and parts of the Avon Valley.[8]

This is just one chapter of a larger story of how the people, capital and culture of British slavery were channelled into and left their mark on Australia – a story that the Western Australian Legacies of British Slavery project is reconstructing using the Time-Layered Cultural Map (TLCMap), an ARDC-supported suite of no-code or low-code digital mapping tools.

Locating Scars of Slavery in Australia

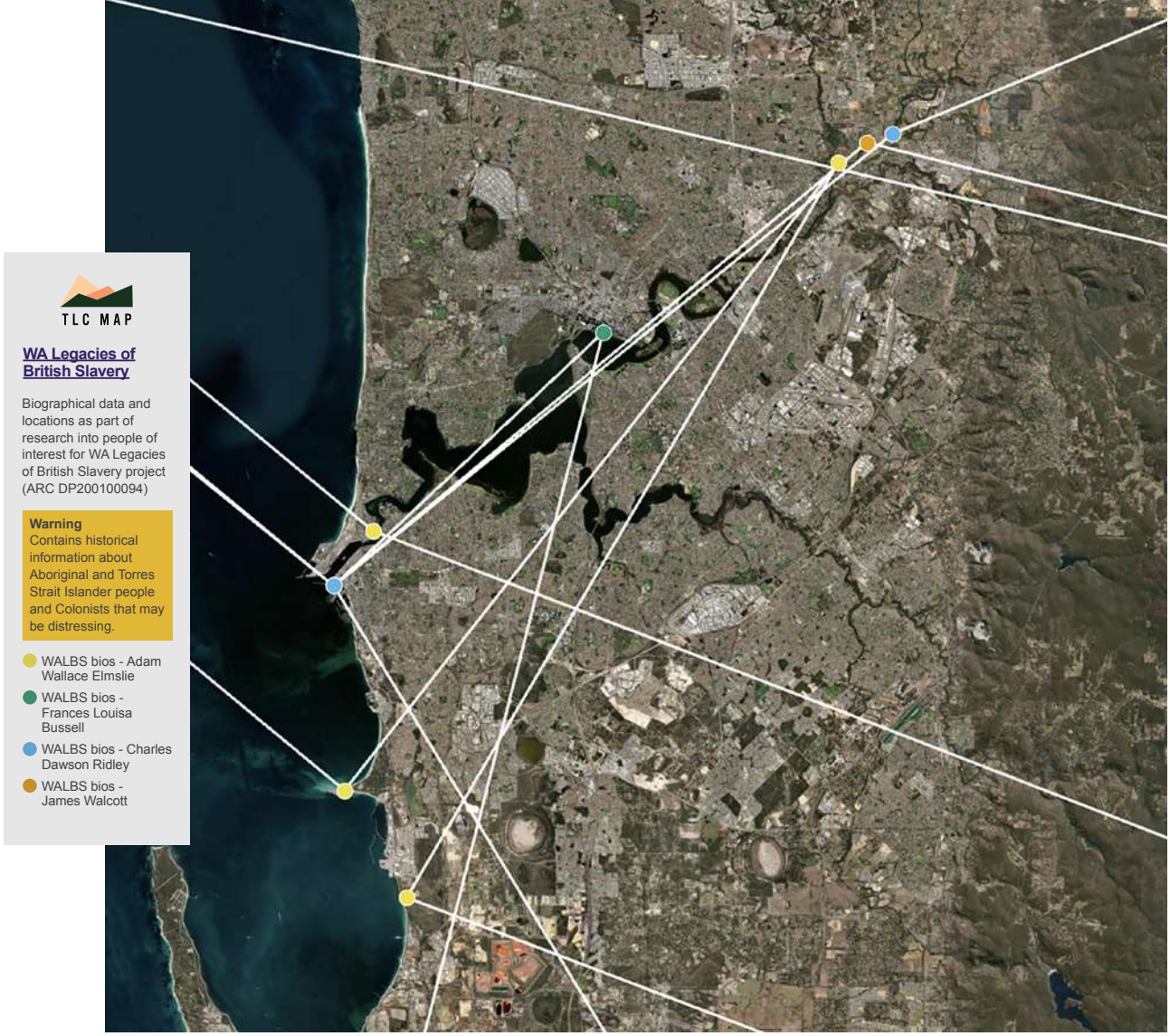
Developed with an ARC Linkage, Infrastructure, Equipment and Facilities (LIEF) grant and later co-investment from the ARDC, TLCMap was launched in 2020 for HASS and Indigenous researchers.

One component of TLCMap was the Gazetteer of Historical Australian Places (GHAP), which has evolved into a standalone service with further co-investment from the ARDC. Based on the Australian National Placenames Survey (ANPS), GHAP provides access to over 330,000 historical and contemporary Australian placenames and their coordinates. It also allows the community to contribute Indigenous and historical names that are not in the ANPS database.

Professor Zoë Laidlaw, a historian at The University of Melbourne, found GHAP to be crucial to her team's work on the WA Legacies project. As part of foregrounding the lasting impact of British slavery on Australia, 2 interns from the university's Digital Studio identified colonial estates in Victoria that shared their names with British Caribbean slave plantations. Then, using GHAP, they located the estates.

"Combining other historical records with the location data from GHAP is helping the team identify which Victorian pastoral estates were indeed connected to Caribbean slave plantations and, conversely, when shared names are just coincidental," said Prof Laidlaw.

8 *Lydon, Jane. From Demerara to Swan River: Charles Dawson Ridley and James Walcott in Western Australia. Australian Journal of Biography and History, 2022;(6): 23–49. https://search.informit.org/doi/10.3316/informit.489240258762875.*



Source — Western Australian Legacies of British Slavery

A close-up of the time-layered cultural map by the WA Legacies team, showing the movements of 4 people with ties to British Caribbean slavery as they settled in Perth

Visualising British Colonisers' Journeys to Australia

TLCMap also allows people, objects and events to be plotted on maps. This enabled the WA Legacies team to visualise the flow of British colonisers from the Caribbean to Australia.

Using TLCMap, Professor Paul Arthur and research associate Isabel Smith from Edith Cowan University charted the journeys of Walcott, Ridley and several other British colonisers who left the Caribbean for Western Australia. Drawing on the WA Legacies team's biographical research, they were able to plot and timestamp every stop the colonisers made with annotations detailing their ties to slavery and their activities at each location. They then stacked the maps together using TLCMap's multilayer function to compare the colonisers' movement.

"We originally set out to develop our own visualisation tools that would help reveal aspects of this hidden history," said Prof Arthur. "TLCMap is now fulfilling the need for spatio-temporal mapping at the national scale for researchers across disciplines." ▶

> We originally set out to develop our own visualisation tools that would help reveal aspects of this hidden history. TLCMap is now fulfilling the need for spatio-temporal mapping at the national scale for researchers across disciplines.
>
> **Professor Paul Arthur,**
> Vice-Chancellor's Professorial Research Fellow and Chair in Digital Humanities and Social Sciences,
> *Edith Cowan University*

> Overall, the aim is to bring new digital tools and resources to bear on some of the central concerns of the humanities. We hope researchers will use these tools and resources for important new perspectives on Australian history and culture.

Emeritus Professor Hugh Craig,
Director, Centre for Linguistic and Literary Computing, *University of Newcastle*
Project Leader, *TLCMap and GHAP*

Mapping HASS and Indigenous Research Data

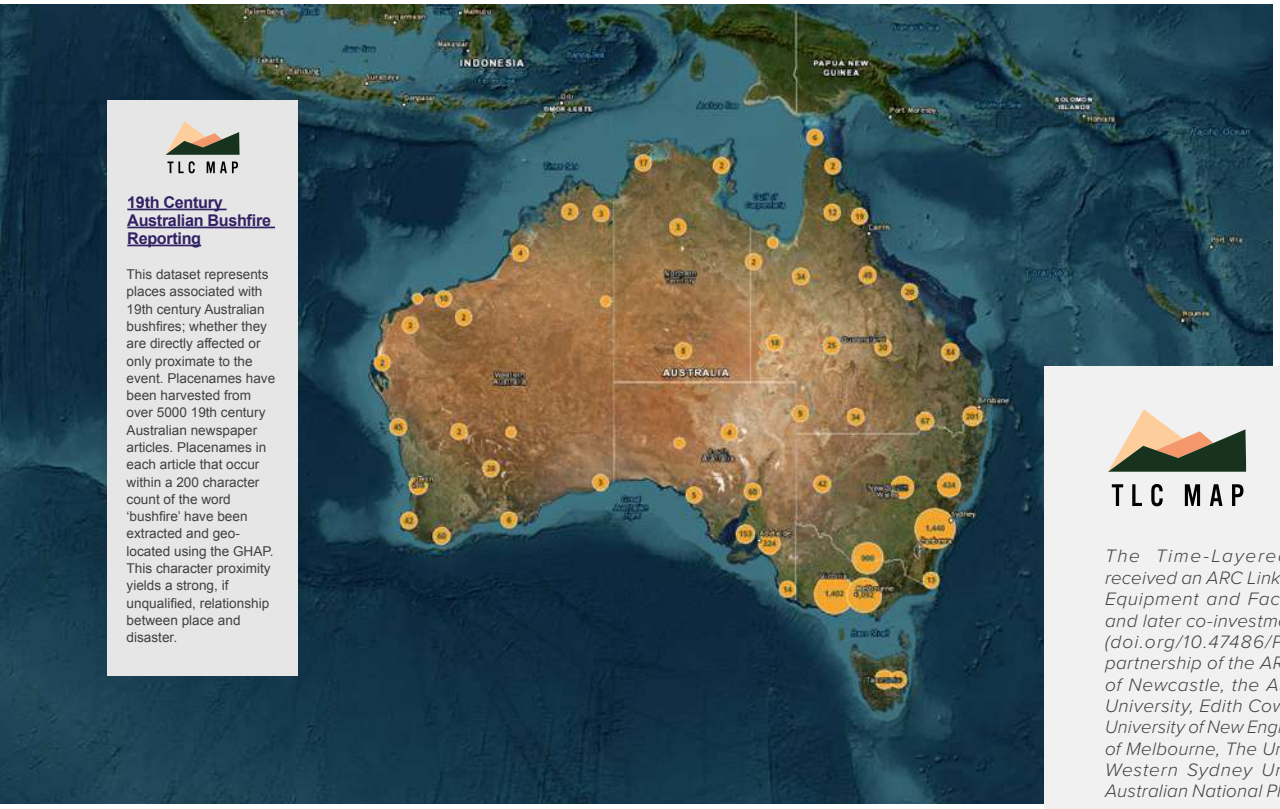
In 2021, TLCMap won gold for education services at the Good Design Awards. With more HASS and Indigenous researchers expected to take advantage of TLCMap and GHAP — both now hosted on the ARDC Nectar Research Cloud — the possibilities are endless for revealing patterns in our culture, society and heritage that tell important stories.

"We are embarking on a new LIEF-funded project this year to detect placenames in documents programmatically and link them to places in our gazetteer for instant mapping," said Emeritus Professor Hugh Craig, Director of the University of Newcastle's Centre for Linguistic and Literary Computing and leader of the TLCMap and GHAP projects.

"We are also developing ways of finding hotspots in spatio-temporal datasets and working on historical census data to add population numbers by year as a foundation map layer in the platform.

"Overall, the aim is to bring new digital tools and resources to bear on some of the central concerns of the humanities. We hope researchers will use these tools and resources for important new perspectives on Australian history and culture." ■

A time-layered cultural map showing places mentioned in 19th century Australian news reports about bushfires



Source — Fiannuala Morgan, Australian National University

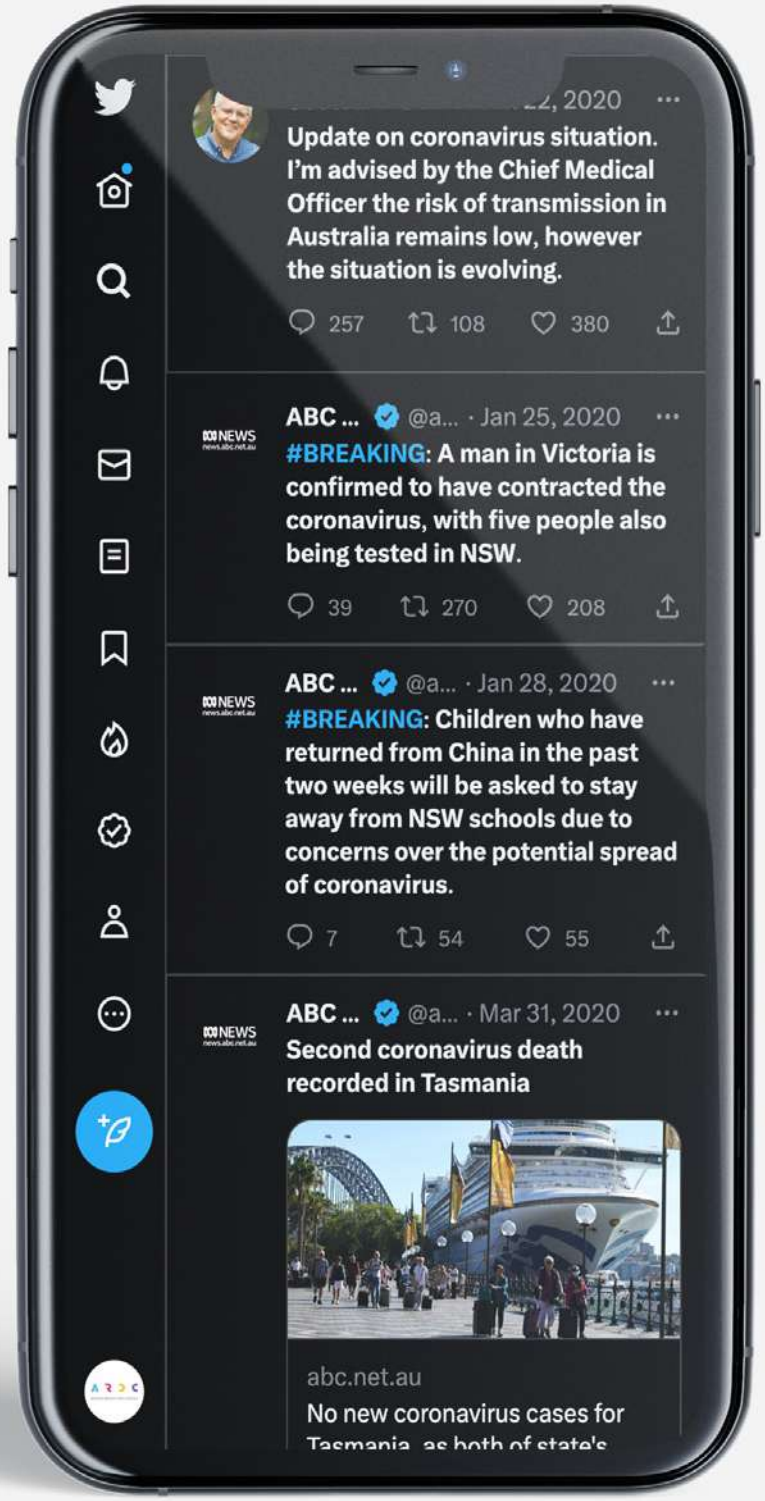
TLC MAP GHAP

The Time-Layered Cultural Map received an ARC Linkage, Infrastructure, Equipment and Facilities (LIEF) grant and later co-investment from the ARDC (doi.org/10.47486/PL069). It was a partnership of the ARDC, the University of Newcastle, the Australian National University, Edith Cowan University, the University of New England, The University of Melbourne, The University of Sydney, Western Sydney University, and the Australian National Placenames Survey.

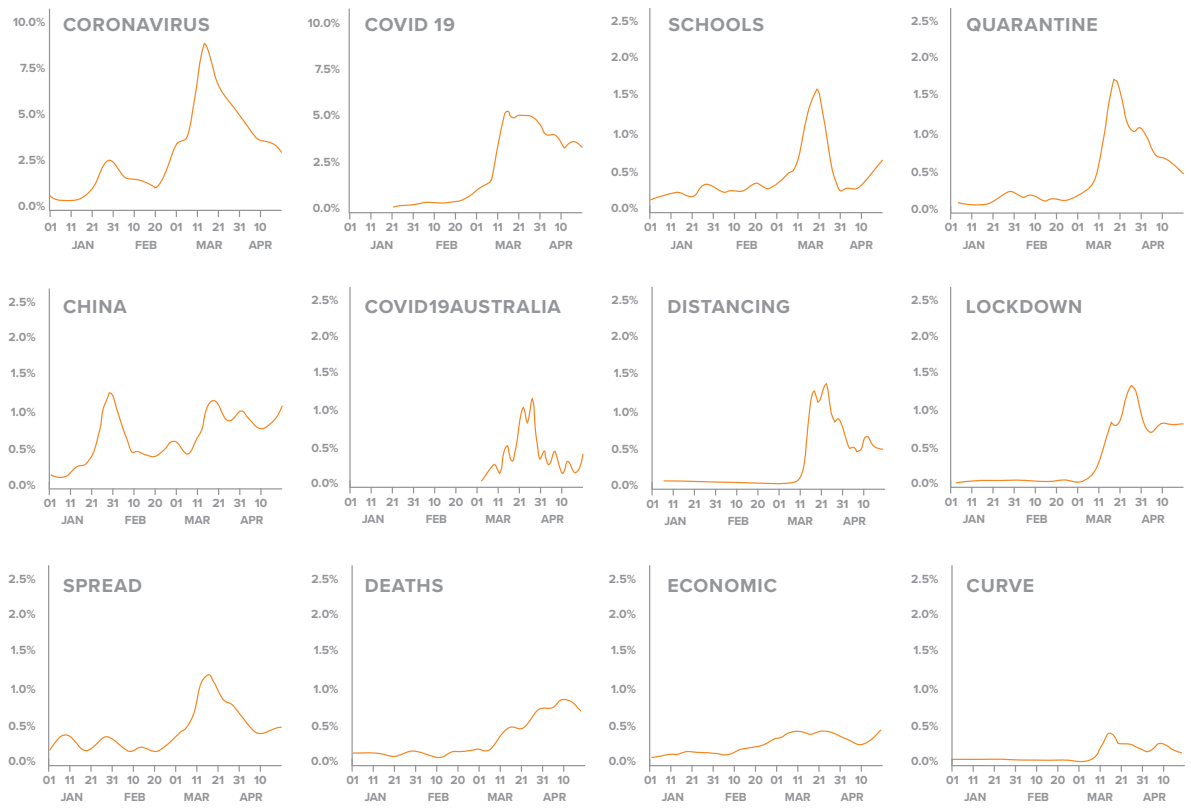
The Gazetteer of Historical Australian Places (doi.org/10.47486/HIR007]) is led by the University of Newcastle in partnership with the ARDC.

Tweets Illuminate the Impact of COVID-19 on Society

China, curve, deaths, distancing, economic, lockdown, quarantine, schools, spread... These were the top words mentioned on Twitter in Australia about the pandemic from January to April 2020.[9] ▶



PERCENTAGES OF TWEETS CONTAINING SELECTED KEYWORDS PER DAY



Source: Adapted from Schweinberger et al.[9]

Note the Y axis is not consistent.

Percentages of tweets about the pandemic containing selected keywords per day from January to April 2020

Following the declaration of the COVID-19 pandemic, researchers began to study what people were sharing on Twitter to gain insight into society's response to a global disaster, one that is still unfolding today.

Software engineers were mining the text to look for trends on social media, but their analyses lacked nuance in determining what conversations were dominant at different stages of the pandemic.

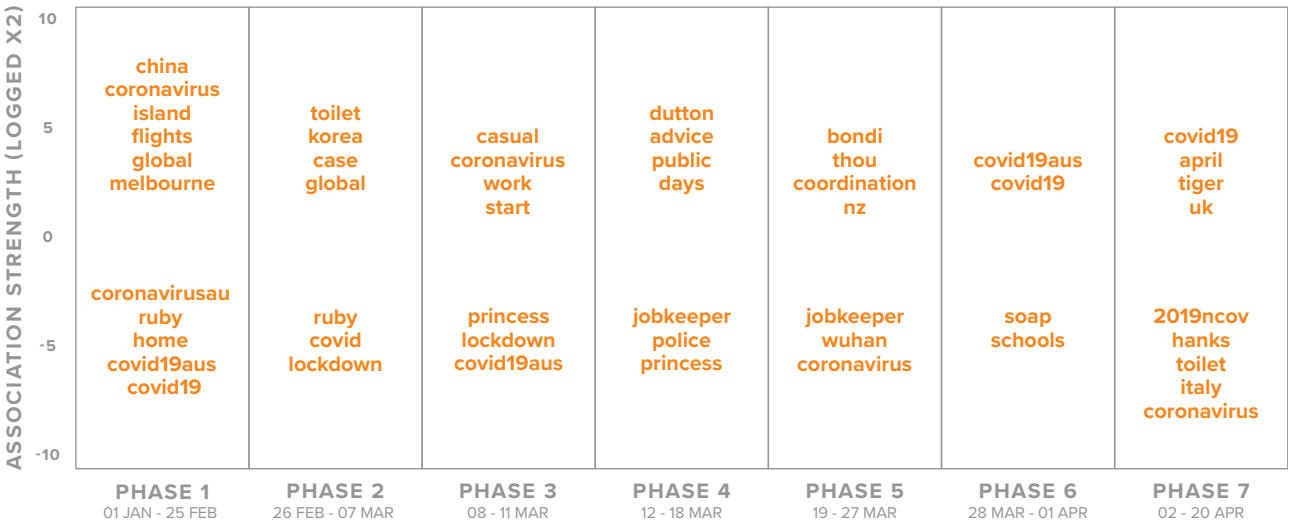
Enter the linguists. Dr Martin Schweinberger is a lecturer in applied linguistics at The University of Queensland and Director of the Language Technology and Data Analysis Laboratory (LADAL).

"As linguists, we know that you cannot view language as one big lump of words," he said. "The discourse on COVID-19 in Australia evolved over time through different topics and different layers of discussion. And when you want to have something meaningful come out of the analysis, you need to separate these layers." ▶

> As linguists, we know that you cannot view language as one big lump of words. The discourse on COVID-19 in Australia evolved over time through different topics and different layers of discussion.
>
> **Dr Martin Schweinberger,**
> Lecturer in Applied Linguistics and Director of the Language Technology and Data Analysis Laboratory,
> *The University of Queensland*

9 *Schweinberger M, Haugh M, Hames S. Analysing discourse around COVID-19 in the Australian Twittersphere: A real-time corpus-based analysis. Big Data & Society. 2021;8(1). DOI: 10.1177/20539517211021437*

TERMS SIGNIFICANTLY OVER-USED (POSITIVE VALUES) AND UNDER-USED (NEGATIVE VALUES) IN COVID-19-RELATED TWEETS FROM THE AUSTRALIAN TWITTERSPHERE, 1 JANUARY – 20 APRIL



Source: Adapted from Schweinberger et al.[9]

> Discourse began by focusing on China and the coronavirus, and later Australians were more concerned about toilet paper, lockdown, casual contacts, jobkeeper, and school closures.

Dr Martin Schweinberger,
Lecturer in Applied Linguistics and Director of the Language Technology and Data Analysis Laboratory,
The University of Queensland

> By combining linguistics and text mining on a large dataset, we enhanced understanding of the public's response to social events in a way that is not possible when relying on only engineering approaches.

Dr Martin Schweinberger,
Lecturer in applied linguistics and Director of the Language Technology and Data Analysis Laboratory,
The University of Queensland

Analysing Unstructured Text

To unravel the early discourse, Dr Schweinberger and Dr Sam Hames, a postdoctoral research fellow in computational humanities at The University of Queensland, used the Australian Text Analytics Platform (ATAP), which received ARDC co-investment, to analyse over 41,000 COVID-related tweets posted between January and April 2020. The Twitter dataset was obtained from another ARDC co-investment project, the Australian Digital Observatory.

The linguistically informed text analysis showed the dominant words shared on Twitter in the early stages of the pandemic. "Discourse began by focusing on China and the coronavirus, and later Australians were more concerned about toilet paper, lockdown, casual contacts, jobkeeper, and school closures," said Dr Schweinberger.

"By combining linguistics and text mining on a large dataset, we enhanced understanding of the public's response to social events in a way that is not possible when relying on only engineering approaches," said Dr Schweinberger.

It's not only linguistics that's harnessing large datasets to gain insights on Australian society. More and more, researchers from the humanities, arts and social sciences are using computational approaches to understand society and culture. The ARDC's HASS and Indigenous Research Data Commons is creating the infrastructure needed to support them in taking data-driven approaches.

Tools and Training in Text Analysis for Australian Researchers

The Australian Text Analytics Platform, ATAP, which is part of the Language Data Commons of Australia, is also one of the projects within the HASS and Indigenous Research Data Commons. It's an open-source platform with tools and training for researchers to analyse, process and explore text. Australian researchers can use it to access an ecosystem of data and code repositories, online workspaces, scripts, and training in text analytics.

Using text analytics, researchers can extract and analyse information from unstructured text, enabling data-driven research. Due to the ever-increasing availability of large amounts of unstructured text, not only from Twitter but other digital media platforms, such techniques are becoming more and more important across diverse research disciplines.

ATAP can also work with existing archives, such as the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) and the Australian National Corpus, making it easier for researchers to access the content in their collections.

A powerful toolbox within ATAP — the Language Technology and Data Analysis Laboratory (LADAL) — offers a multitude of resources on methods used to analyse textual data. It includes basic research showcases, self-paced tutorials for everyone from novices to experts, and readymade interactive notebooks that allow users to try out a method, also with their own data.

ATAP is a technical platform but training is an important component — digital research methods are vital skills for early-career researchers in HASS and Indigenous studies. In 2022, the ATAP team trained over 400 researchers through hands-on workshops, online training modules and online office hours, and advised and collaborated with partners. ■

The Australian Text Analytics Platform project received co-investment (doi.org/10.47486/PL074) and expertise from the ARDC. It is led by The University of Queensland, with support from AARNet and The University of Sydney.

Minding Our Language Data

Next time you speak with a person in their 80s, take a moment to consider how they speak. What words do they use? Is their grammar different to yours? Do you notice anything about their pronunciation? ▶

Images — rastilly - 172088064 / AdobeStock.com | shane-rounce - I4dZ4h6vL6o / unsplash.com
clarinta.e - cyFNGgeUI / unsplash.com | sheila-regita-pasaribu - eyOJ6RIwb-c / unsplash.com
Kenya Chan / WikiCommons

Professor Catherine Travis from the Australian National University analyses changes in the way English is spoken in Australia over time. For example, she compares how people of different ages speak, from teenagers to octogenarians, as well as comparing how Australians speak today to how we spoke 40 years ago.

To conduct this research, Prof Travis and colleagues compiled the Sydney Speaks Corpus — a collection of recordings of Sydneysiders telling their stories. However, to understand what changes had occurred over several decades, they had to find old recordings.

“There was a set of recordings of Sydney residents speaking English in the 1970s and 80s created by a very famous linguist, Dr Barbara Horvath, which is cited in introductory linguistics textbooks around the world,” said Prof Travis. “I was communicating with Dr Horvath when she said ‘By the way, I’ve got all my cassettes of the recordings sitting in my garage. Do you know anybody who’s interested in that?’

“This is like a gold mine from a linguistic perspective. I was able to digitise them, and incorporate them into the Sydney Speaks Corpus.”

Many more such recordings are waiting to be discovered and documented. And because recordings like these often contain stories about people’s lives and experiences, they are useful not just to linguists, but also to historians, sociologists, and anthropologists, among others.

The benefits of being able to find and access language data are immeasurable. Not only can language data help researchers from many disciplines answer a multitude of questions without having to collect new data, it is also crucial for Aboriginal and Torres Strait Islander peoples, who are revitalising hundreds of languages in Australia and the region. ▶

There was a set of recordings of Sydney residents speaking English in the 1970s and 80s created by a very famous linguist, Dr Barbara Horvath...

This is like a gold mine from a linguistic perspective. I was able to digitise them, and incorporate them into the Sydney Speaks Corpus.

Professor Catherine Travis,
Chair of Modern European Languages,
Australian National University

Access to Language Data May Be Restricted

The Language Data Commons of Australia (LDaCA) was created, with ARDC co-investment and expertise, to make language data easier to find. It is a portal that points users to where relevant data is held. Making the “gold mines” of language data scattered across Australia findable is a crucial step towards making this data FAIR — findable, accessible, interoperable and reusable.

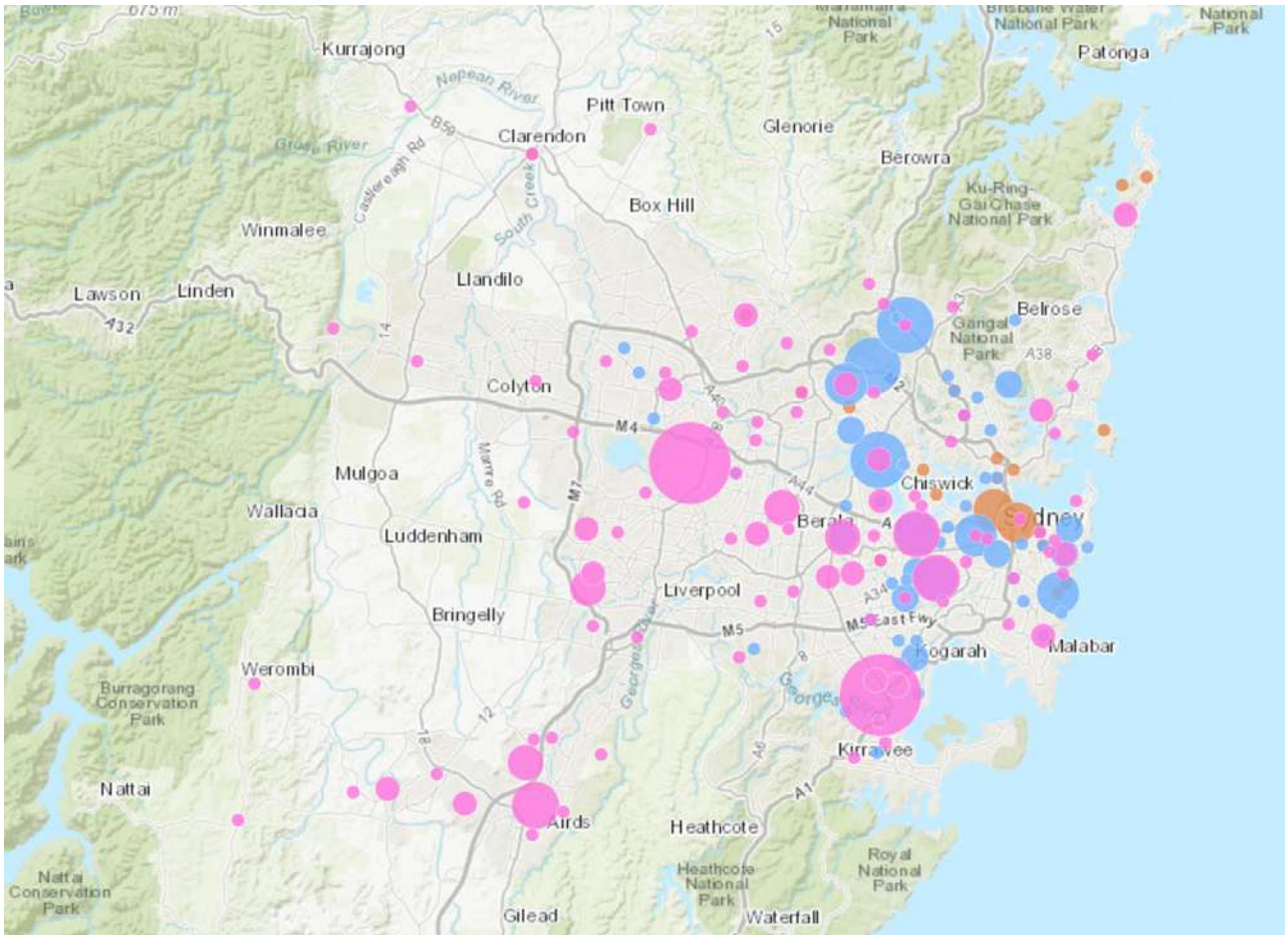
However, providing access to language data is not always straightforward. Language data is inherently identifiable and may contain sensitive information — a recording of a conversation between friends will contain identifiable information, or may include personal topics, for example, and in this era of deep fakes, could be used for malicious purposes. To make the data accessible, LDaCA provides a way for users to request access to the data through its online portal.

“All personal data is identifiable, so it can’t always be made open,” said Dr Peter Sefton, the technical lead for LDaCA at The University of Queensland. “Research is conducted with many different access licences, meaning that some data can be reused, but a lot has specific restrictions.”

The Sydney Speaks Corpus, for example, contains data that was collected under a few different licences. It includes the NSW Bicentennial Oral History Project recordings from 1987–1988, which are freely accessible in their entirety through the National Library of Australia, as well as recordings from 2016 onwards that are more restricted due to the agreements with the participants about this data collection. LDaCA helps data stewards like Prof Travis manage the access restrictions in accordance with ethical, moral and legal obligations.

“The access conditions for each item in LDaCA are determined by the data steward, and are managed using an authorisation system,” said Dr Sefton. “The access can be a simple click-through licence, where you agree to licence terms, through to a detailed multi-step workflow where applicants are vetted based on criteria assigned by the rights-holder, such as qualifications or membership of a cultural group. In some cases, there is a manual approval process.” ▶

The Sydney Speaks project brings together recordings of 250 speakers from 3 collections of spoken language. The size of each bubble on this map corresponds to the number of speakers.



Source — Sydney Speaks

Image — ake1150 - 220246410 / AdobeStock.com

Giving Access to Communities, Not Just Researchers

In Australia, the Australian Access Federation (AAF) mediates access to hundreds of national and international research platforms and resources using the researcher's institutional ID and password.

However, since access is based on institutional IDs, a researcher who changes institutions or retires loses access because their email changes. Also, language data is not only useful for researchers working in research institutions; communities might be searching for language data for revitalisation projects or as a cultural record. A different identity authentication system was needed for recordings in LDaCA.

The LDaCA team worked with AAF to integrate CILogon, a system that enables people outside traditional research institutions to create authenticated identities using social logins such as Google and GitHub and access the data they need.

> It would be such a waste for those stories to stop at the analysis of the vowels. That's what gets us linguists excited, but there's much, much more to offer.
>
> **Professor Catherine Travis,**
> Chair of Modern European Languages,
> *Australian National University*

Taking the Stories Beyond Linguistics

Information about the Sydney Speaks Corpus and other language collections are now in LDaCA, and linguists, historians and community groups can request access to the hundreds of hours of language recordings.

"It would be such a waste for those stories to stop at the analysis of the vowels," said Prof Travis. "That's what gets us linguists excited, but there's much, much more to offer." ■

This project received co-investment (doi.org/10.47486/DP768 and doi.org/10.47486/HIR001) and expertise from the ARDC. It is led by The University of Queensland in partnership with AARNet, ARC Centre of Excellence for the Dynamics of Language, Australian National University, First Languages Australia, Monash University, The University of Melbourne and The University of Sydney.

Australian Research Data Commons

TRANSLATIONAL RESEARCH DATA CHALLENGES

Partnering to Overcome Translational Research Data Challenges

Our Translational Research Data Challenges program, which kicked off in 2020, is a national-scale initiative bringing innovative and high-impact digital infrastructure solutions to real-world problems. In its current program, we have brought researchers together with the public and private sector to solve the pressing societal challenges of bushfires and food security.

To establish a strong network of partners and identify the strategic gaps in digital infrastructure, we have taken a co-design approach. For both the bushfires and food security projects, we ran targeted consultations over 11 months, including a series of facilitation meetings and participatory design workshops with over 150 stakeholders to identify key data challenges. Through these processes, we identified priority areas where national digital research infrastructure could improve Australia's research capability to address these societal challenges.

Both projects are establishing data infrastructure, such as platforms for visualising or analysing data; data assets; data catalogues; and data-sharing frameworks and policies.

Bushfire Data Challenges

The unprecedented 2019/2020 Australian bushfires displayed in full force the devastating impact bushfires can have on the natural landscape, the built environment and human health.

To protect our environment and communities, respond effectively to bushfires, and recover from them in the future, the Royal Commission into National Natural Disaster Arrangements (2020) recommended the sharing of data to support better decision making.

In response to the Royal Commission, the ARDC co-invested $4.77 million in 13 national-scale projects that tackle data challenges in bushfire research to improve Australia's bushfire resilience, response and recovery. To improve understanding of bushfire behaviour, these projects are aggregating data on fire history and fuel at the national scale, and developing a platform for modelling bushfire behaviour. Understanding of bushfire impacts is also improving through the aggregation at a national scale of data on air quality; human health; and invertebrate species' traits, genomics and distribution. Data management is also being addressed to ensure data is collected, used and communicated according to an agreed framework and structure.

As a result, scientists can tackle the operational and planning challenges of Australia's fire agencies using national-scale data, modelling and advanced analytics.

Food Security Data Challenges

The ARDC is partnering with stakeholders in food security to develop innovative digital infrastructure solutions to improve research into Australia's production, consumption and distribution of safe and high-quality food.

Our co-investment of $3.6 million boosts the $5.7 million co-investment from our partners. Projects are addressing areas such as agriculture, antimicrobial resistance, aquaculture and fisheries, biosecurity, food provenance and traceability, and food equity and relief.

As a result, researchers can tackle food security problems using national-scale data from government agencies, non-government organisations, and private-sector food producers. ■

Image — Stuart+ 507395677 / AdobeStock.com

Bringing Together Bushfire History Data to Plan for the Future

Bushfire analysts and planners now have access to a national dataset of past fires, which will allow them to prepare a faster, more effective response.

Until recently, Australians could almost be forgiven for thinking bushfires stopped at state borders. That's because the bushfire authorities of each state and territory gather their own data separately and have management responsibility only within their own boundaries. The media reports reflected that. Even the fire danger rating systems differed between states.

But the Black Summer fires of 2019–2020, which burned more than 240,000 square kilometres across multiple states and territories, clearly demonstrated the limitations of such an approach. A high level of information sharing and management between jurisdictions became a necessity. ▶

Looking at that response and discussing needs with bushfire analysts, modellers and planners, the ARDC recognised the value of pulling together all the state/territory-based data into a harmonised, accessible national dataset documenting Australia's bushfire history.

Dr Sheida Hadavi,
Director, Translational Research Data Challenges,
ARDC

Looking at that response and discussing needs with bushfire analysts, modellers and planners, the ARDC recognised the value of pulling together all the state/territory-based data into a harmonised, accessible national dataset documenting Australia's bushfire history, said director of the ARDC Bushfire Data Challenges program Dr Sheida Hadavi. And the expertise to do so was available in the national bodies of Geoscience Australia and the Emergency Management Spatial Information Network Australia (EMSINA).

So, in 2021 the ARDC, with a contribution from Minderoo Foundation, initiated, facilitated and invested in a $6 million program to deliver such a dataset, as well as other data tools and platforms for bushfire data. The bushfire history dataset was created using data from trusted state and territory agencies, who each contributed their own government database to Geoscience Australia through EMSINA. Geoscience Australia specialists then combined and standardised the data using advanced geospatial software.

The historical bushfire dataset documents fires from 1899 to 2022. It is available for anyone to access and use through the Digital Atlas of Australia, a new online geospatial platform being developed by Geoscience Australia. Currently in its beta phase, the Digital Atlas allows anyone to layer the historical bushfire data with other curated trusted national datasets on Australia's geography, population, economy and environment. This integration will help unlock the full potential of the historical bushfire dataset, enabling a better understanding of collective challenges across borders and empowering proactive and informed decision making. ▶

With how everything was set up before, you couldn't really look beyond your own border. But as ACT fire planners, we can't just look at our own patch. We are a small jurisdiction with a large border with NSW. The fire history data allows us to adjust fuel accumulation rates for different vegetation types so we can put the best fuel inputs into the different fire behaviour models we use.

Ailish Milner, Assistant Director, Strategic Planning Bushfire, *ACT Rural Fire Service*

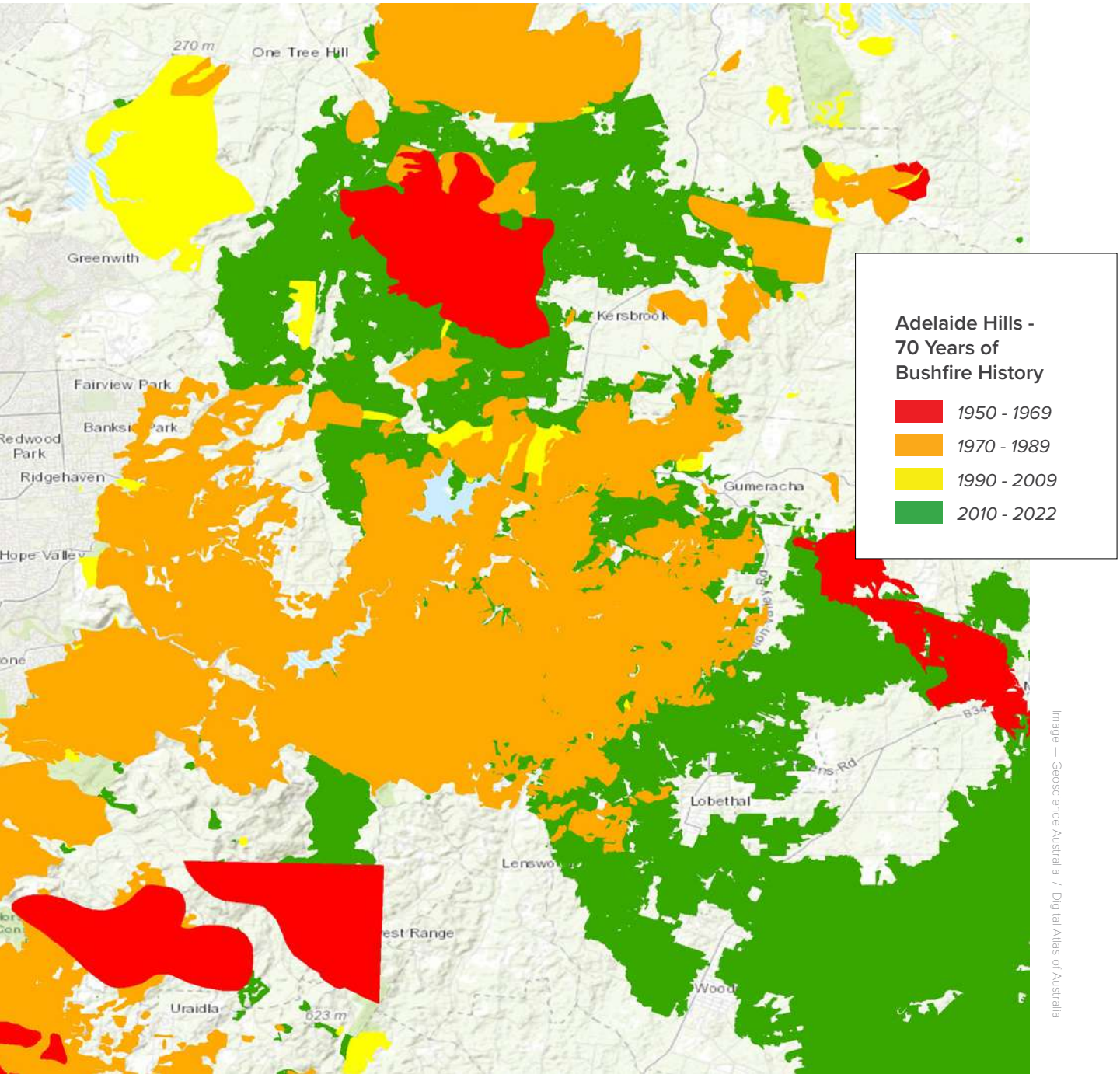


Image — Geoscience Australia / Digital Atlas of Australia

Bushfire Data That Goes Beyond Borders

The creation of the bushfire history dataset means that for the first time, researchers and emergency management professionals can look beyond borders to understand bushfire behaviour and plan accordingly.

"With how everything was set up before, you couldn't really look beyond your own border," said ACT Rural Fire Service fire behaviour analyst and modeller Ailish Milner. "But as ACT fire planners, we can't just look at our own patch. We are a small jurisdiction with a large border with New South Wales. The fire history data allows us to adjust fuel accumulation rates for different vegetation types so we can put the best fuel inputs into the different fire behaviour models we use."

The impact of the dataset will extend far beyond bushfire modelling. Energy Queensland has already been using it to assess when their electricity poles were last exposed to bushfire.

Knowing where, when and how frequently fire has struck is important in determining Australian Fire Danger Ratings and assisting disaster recovery, as well as for carbon accounting, planning controlled burning, Indigenous land management, water utility management, urban planning, determining insurance risk, conserving nature and protecting species.

Creating the bushfire history dataset was only one of the 5 work packages on bushfire behaviour included in the Bushfire Data Challenges program. Another group worked on determining what information needed to be included and in what form. Yet other groups identified gaps in the data and secured resources to help plug them. And, a team has been looking at how remote sensing can best contribute to the assembly and assessment of bushfire data.

The bushfire work will be further developed now be undertaken by the CSIRO-based National Bushfire Intelligence Capability (NBIC).

Data and Platforms to Address Bushfire Data Challenges

More projects from the Bushfire Data Challenges program are launching in 2023, including:

- a fire behaviour modelling platform
- a framework for sharing bushfire data between jurisdictional agencies
- bushfire research data management plans and catalogue
- national-scale fuel data from state agencies, ground observations and remote sensing
- curated species data, invertebrates traits data and the Australian Reference Genome Atlas
- a federated air quality data system and a bushfire-specific air pollution exposure dataset
- aggregated and integrated data on health outcomes associated with bushfires at a national scale. ■

This project received co-investment from the ARDC (doi.org/10.47486/DC001) and a contribution from Minderoo Foundation. It was led by Geoscience Australia and Emergency Management Spatial Information Network Australia was a partner.

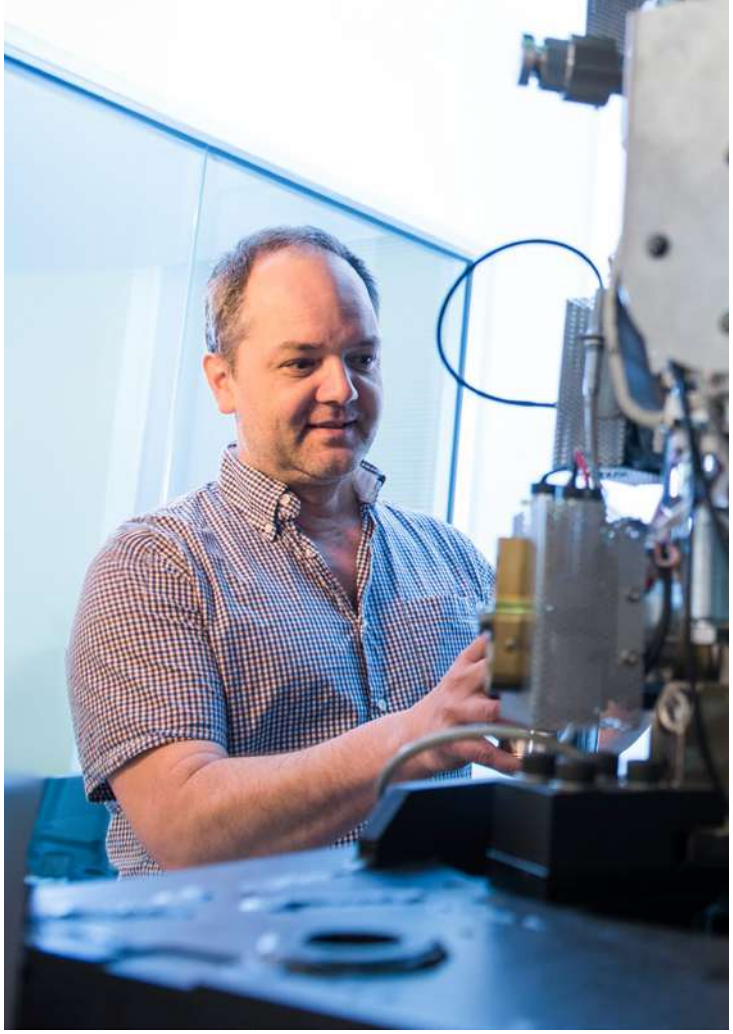


Image — Monash University

The Australian Characterisation Commons at Scale is solving big data challenges faced by researchers using imaging. This photo shows Associate Professor Georg Ramm, Head, Monash Ramaciotti Centre for Cryo-EM, Senior Research Fellow & Head, Advanced Cellular Imaging Laboratory, using the Titan microscope.

BUILDING ON PARTNERSHIP SUCCESS

Our success in delivering digital research infrastructure that meets the needs of researchers is built on our deep knowledge of the Australian research community combined with our strong partnerships.

Partnering is the bedrock of our work to develop our Thematic Research Data Commons and Translational Research Data Challenges. And it is the consistent strand running through our work in developing research platforms and national data assets, recognising research software as a first-class research output, building digital research skills, supporting data retention, facilitating communities of practice, advising the research community, and contributing to policy. ■

Co-Investment Partnerships
PAGE 76

IMPACT SHOWCASE
A Research Data Management Framework for Institutions
PAGE 79

Research Software
PAGE 83

IMPACT SHOWCASE
The ARDC Awards Software Innovation
PAGE 84

Data Retention
PAGE 85

Digital Research Skills
PAGE 85

IMPACT SHOWCASE
Digital Research Skills Trainers Convene at the 2023 ARDC Skills Summit
PAGE 86

IMPACT SHOWCASE
Ocean Data Repository Certified as Trustworthy
PAGE 87

Communities of Practice
PAGE 88

Engagements, Expertise & Support
PAGE 88

Our Input to Policy
PAGE 88

IMPACT SHOWCASE
Mapping the Vision for Data-Driven Research in Australia
PAGE 89

Co-Investment Partnerships

This year alone, 52 ARDC-supported research platforms and national data assets are launching to give Australian researchers a competitive advantage through data. These projects received expertise from ARDC and were created by bringing together co-investment from partners and the ARDC.

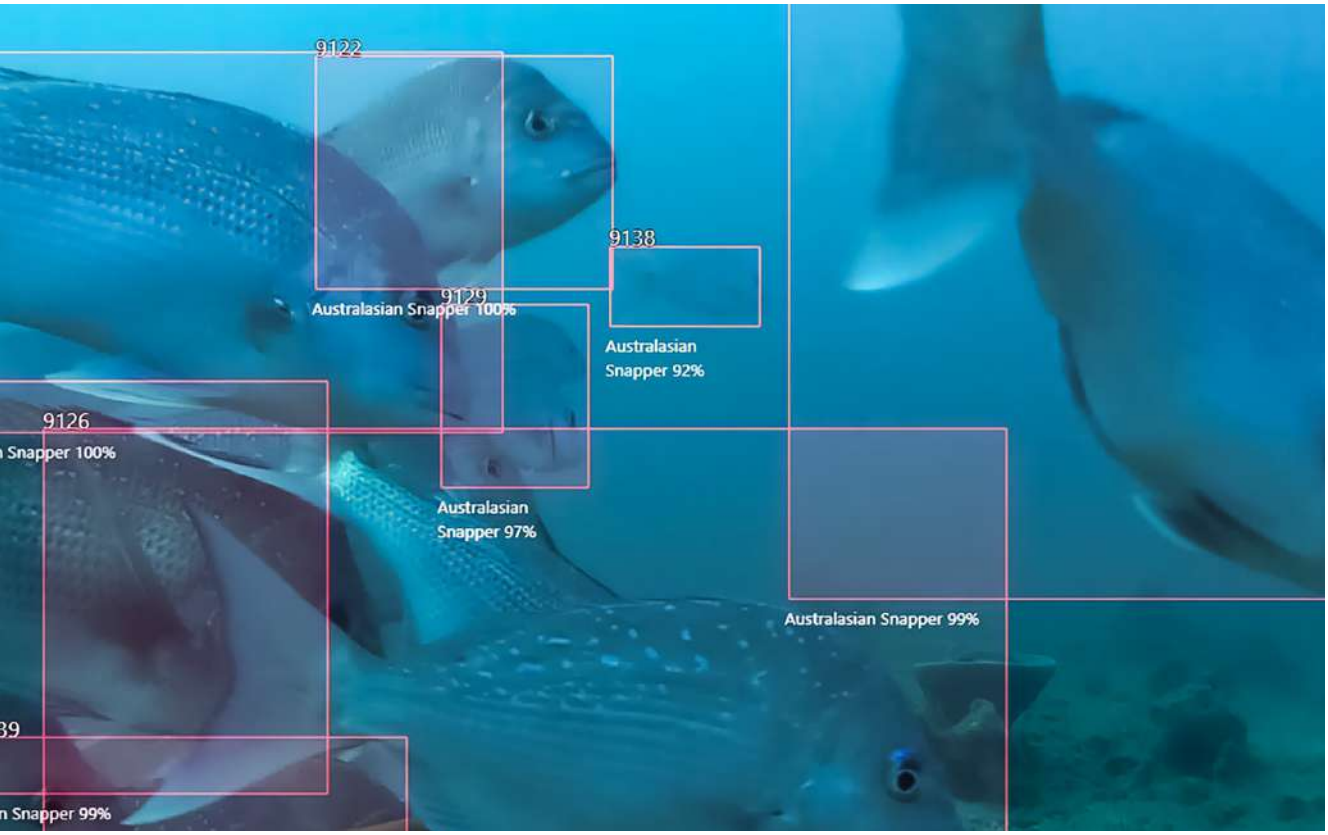
Through our research platforms program, we have co-invested $21.7 million in 26 research platforms that enable researchers to collect or generate data, analyse those data, and produce outputs that can be made findable, accessible, interoperable and reusable (FAIR). The platforms often integrate or connect to specific data resources.

With our national data assets program, we partnered with more than 200 organisations across the research sector to develop a portfolio of national-scale data assets that support leading-edge research. 26 national data assets are now available to Australian researchers.

The research platforms and national data assets listed on the following pages will be available in 2023.

Health and Medical Research
Platforms and Data Assets Launching in 2023

- Air Health Data
- Australian Cancer Data Network
- Australian Characterisation Commons at Scale
- Australian Companion Animal Registry of Cancers
- Australasian Computational and Simulation Commons
- Australian Electrophysiology Data Analytics Platform
- Australian Imaging Service
- Australian Urban Health Indicators
- BioCommons Bring Your Own Data BYOD Expansion Project
- Biomedical Discovery Asset
- Child Health and Development Atlas
- Health Studies National Data Asset (HeSANDA) (ARDC-led)
- Hospital EMR Data as a National Data Asset for Research
- Human Genomes Platform
- Integrated Microscopy and Proteomics
- Linked Data Asset for Australian Health Research
- National Poisons Information Centre Dataset
- National Transfusion Dataset
- Scalable Governance, Control and Management of FAIR Sensitive Research Data
- Secure Cloud Computing for Sensitive Data (ERICA)
- Veterinary and Animal Research Data Commons



Source — FishID/The Global Wetlands Project (GLOW) / WA Department of Primary Industries and Regional Development

FishID uses artificial intelligence to automate the analysis of videos and images of fish and other sea creatures, making large volumes of data faster and easier to process.

Earth and Environmental Science

Platforms and Data Assets Launching in 2023

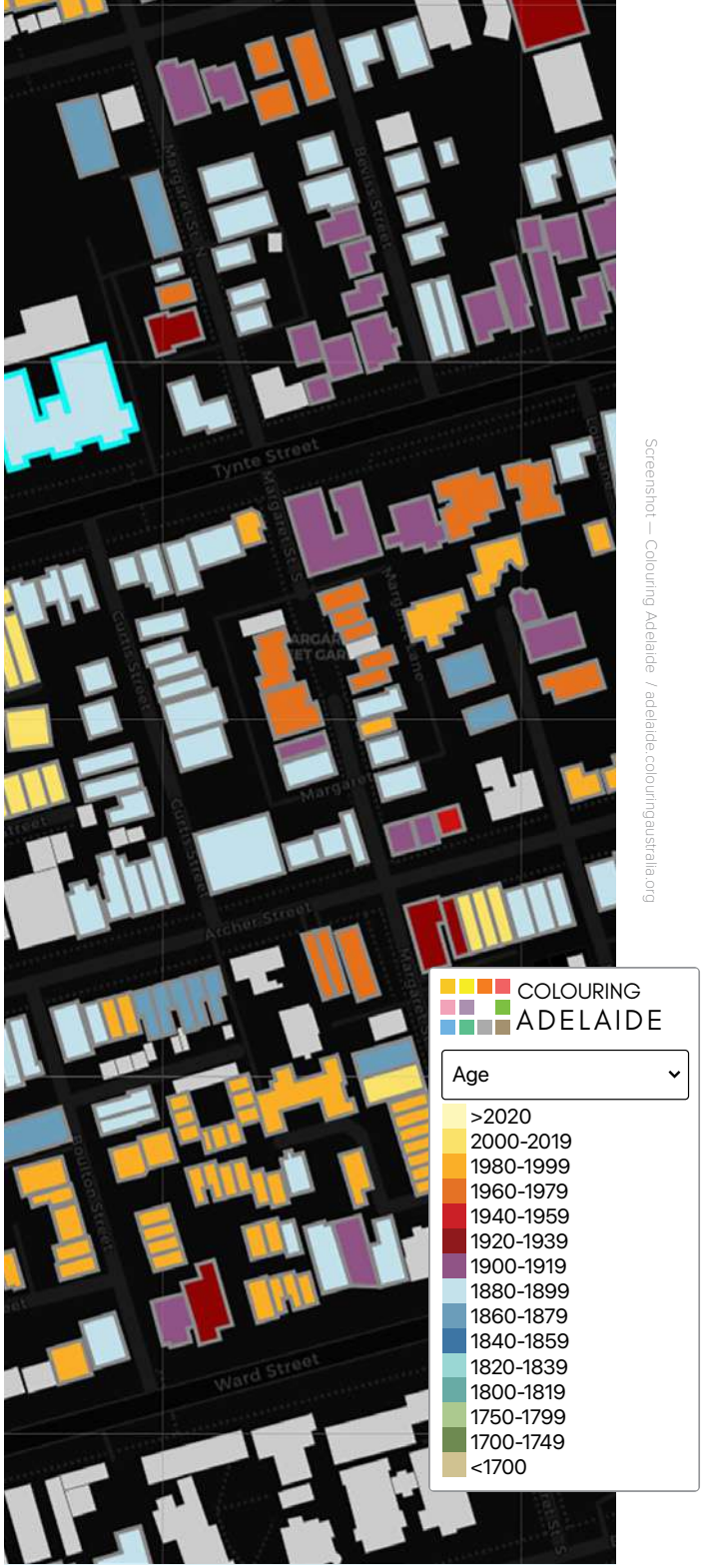
- 2030 Geophysics Collections
- AgReFed: A platform for the transformation of agricultural research
- Air-Health Scientific Workflow System
- AU Pest Genomes
- AusTraits
- Australia's Scalable Drone Cloud
- Australian Transport Research Cloud
- Biosecurity Commons Australia
- Catching Oz Waves
- EcoAssets
- EcoCommons Australia
- Fish and Shark Data
- FishID
- Geodynamic Adjoint Optimization Platform
- Global Multi-Resolution Topography for Australian Coastal and Ocean Models (GMRT-AusSeabed)
- Microbial Ocean Atlas
- Mobility Australia Dataset
- Open Ecoacoustics
- OzBarley Data Asset
- Restricted Access Species Data Pathways From Decision Making to Research

Housing data across Australia is available through many agencies, across governments and jurisdictions, but the access is fragmented – there isn't a central place to springboard off and find those datasets.

The Australian Housing Data Analytics Platform will help direct users to the best available housing data that exists, whether it's in government or industry.

Professor Christopher Pettit leads the ARDC-supported project for housing data analytics.

Christopher Pettit,
Professor of Urban Science and
Director of the City Futures Research Centre,
UNSW Sydney



Screenshot — Colouring Adelaide / adelaide.colouringaustralia.org

Part of the Australian Housing Data Analytics Platform, Colouring Australia is the new Australian contribution to an international collaboration that uses open source software to collect, collate and visualise data about buildings.

Humanities, Arts and Social Sciences (HASS) and Indigenous Research
Platforms and Data Assets Launching in 2023

Australian and New Zealand Leaders, Elections and Democracy Data Asset

Australian Digital Observatory

Australian Housing Data Analytics Platform

Australian Text Analytics Platform

Coordinated Access for Data, Researchers and Environments (CADRE)

FAIMS 3.0 Electronic Field Notebooks

Language Data Commons of Australia

Leveraging Data to Support Young People's Education and Wellbeing Data Asset

National Free Access Coronial Findings, Recommendations & Responses Data Asset

Time-Layered Cultural Map of Australia 2.0

Natasha Simons from the ARDC speaks at the launch of the Research Data Management Framework for Institutions.

IMPACT SHOWCASE

A Research Data Management Framework for Institutions

On 21 February 2023, the ARDC launched the Research Data Management Framework for Institutions in collaboration with 25 universities.

The framework brings together knowledge from hundreds of experts from these universities to create a national approach for institutional research data management (RDM). The approaches and ideas documented in the framework have been tested by institutions across the country through dedicated projects and a passionate community of practice.

An In-Person and Online Launch

The framework was officially launched at in-person events in Canberra, Brisbane, Toowoomba, Sydney, Melbourne and Adelaide, and more than 500 people registered for the live broadcast.

Speakers at the launch reflected on the goodwill and enthusiasm of those who created and tested the framework, describing the process as “open, collaborative and fun”. They also showcased the ways the framework has been (and can be) applied by institutions.

The Growing Complexity of RDM in Institutions

RDM is a complex challenge for institutions. It requires many departments or business units to work together. This challenge is becoming more pressing as:

- the quantity of research data generated within institutions is increasing
- institutions are moving towards increasingly open research practices, with a focus on data availability for both integrity assurance and reuse
- institutions are facing changing data security needs and policies
- institutions are also facing financial pressure to improve the efficiency of their RDM as storage costs are no longer decreasing fast enough to allow all research data to be stored indefinitely. ▶

All Images — Ben Appleton / ARDC

How the Framework Was Developed

The ARDC's Institutional Underpinnings program began in February 2021 with the goal of coordinating research data management across Australian universities. All Australian universities were invited to participate and 25 Australian universities signed up to jointly develop the first draft of the framework.

Through a co-design process, the participants identified 19 elements that the framework should cover, 9 of which were prioritised for development in detail and assigned to working groups.

While the framework is intended specifically for Australian universities, it may also be useful to other research institutions. It provides an institutional-level perspective and includes recommendations and calls to action. ▶

Roxanne Missingham OAM,
University Librarian
(Chief Scholarly Information Officer),
Australian National University

"The Institutional Underpinnings program has been an amazing achievement. We've had 25 universities with multiple people from all over the universities genuinely come together and share ideas, all treating each other with respect, all exploring issues that are complex.

There was never a time when we were looking at easy issues. The universities also operate in a very complex environment, across the states and territories and nationally. We brought all of that together, in an amazing way.

Thanks very much to the staff of the ARDC for supporting us and keeping us on track. I still can't believe we had months and months of fortnightly meetings and everyone came — it was a real sign that improving research data support is something from our hearts, not just from our positions in the universities.

The end result is just remarkable and will live on as we meet all the changes in the infrastructure around us and think of new opportunities to continue to work together. It's been a real privilege for me to be involved.

Launch Round-Up – What People Think of the Framework

We asked a few people at the Canberra launch at the Australian National University what they think about the framework and the broader approach of the Institutional Underpinnings program. Here's what they said:

> The Institutional Underpinnings program is very important, because it's all about open data and sharing of knowledge. So we all can progressively get better. In terms of collaboration, everyone's been really open to share their challenges and what worked for them or didn't. There's a really good community spirit.

Anesh Nair,
Manager of Research Ethics, Integrity and eResearch,
University of Canberra

Dr Adrian Burton from the ARDC speaking at the launch of the Research Data Management Framework for Institutions

All Images — Ben Appleton / ARDC

> It's really exciting seeing the way that the RDM space has evolved and the collaborative effort that's gone in to produce this report. It gives a lot of guidance and information and advice.
>
> I think that's going to be really valuable in terms of advocacy, not only for the importance of research data management but the practical steps to implement to move [RDM] forward.

Jo Croucher,
Research Data Management Specialist,
NCI Australia

The Next Phase – The RDM Network

The framework may look like a final product but it is already evolving and will continue to do so. To leverage the community and guidance generated by the Institutional Underpinnings program, the ARDC has established the RDM Network, which 31 universities have joined.

Representatives from these universities take part in monthly workshops to share information and discuss best practice in RDM. They have also co-designed 4 collaborative projects that fill gaps identified by the framework, and these projects are now underway. Throughout 2023 they will continue to identify opportunities to add to the framework guidance. ■

Dr Jac Charlesworth from the University of Tasmania speaking at the launch

Institutional Underpinnings received co-investment from the ARDC. It was led by the ARDC in partnership with 25 Australian universities:

Australian National University
Bond University
Charles Darwin University
Curtin University
Edith Cowan University
Federation University Australia
Griffith University
Macquarie University
Monash University
QUT
Swinburne University of Technology
The University of Queensland
The University of Adelaide
University of Canberra
The University of Melbourne
University of New England
University of New South Wales
University of Southern Queensland
The University of Sydney
University of Tasmania
University of Technology Sydney
The University of Western Australia
University of Wollongong
Victoria University
Western Sydney University.

More than 6,000 people in Australia are developing and maintaining vital research software. The ARDC is working with the research community to ensure these people are recognised and valued, and that they can develop their skills so that the research sector can meet future demands and remain world class.

Research Software

Research software — the software that is created for or through research — is an integral part of modern research. In a 2022 ARDC survey, 92 per cent of researchers said that research software is essential or very important for their research.[10] Our National Agenda for Research Software positions research software as a first-class output of research, alongside publications and research data. We are making this a reality by leading coordinated activities with the Australian research community to see, shape and sustain research software, including driving the uptake of FAIR principles for research software.

More than 6,000 people in Australia are developing and maintaining vital research software.[11] The ARDC is working with the research community to ensure these people are recognised and valued, and that they can develop their skills so that the research sector can meet future demands and remain world class.

Sponsoring awards and prizes that elevate the research software community is one example of this work. The ARDC is encouraging change by sponsoring awards for new, innovative software in research fields where developing such software is a common but unrewarded practice. We are also thrilled to be awarding the inaugural 2023 ARDC Eureka Prize for Excellence in Research Software, recognising the software creators and maintainers that have had a broad impact on research. ■

Image — Josh Brightman / ARDC

10 *Stevens F. Understanding how researchers find research software for research practice. Zenodo: ARDC; 2022. doi.org/10.5281/zenodo.7340034*

11 *Barker M, Chue Hong NP, Katz DS et al. Introducing the FAIR principles for research software. Sci Data 2022;9:622. doi.org/10.1038/s41597-022-01710-x*

ARDC CEO Rosie Hicks presents the ARDC-sponsored Ecological Society of Australia Award for New Developers of Open Source Software in Ecology to Dr Fonti Kar.

IMPACT SHOWCASE

THE ARDC Awards Software Innovation

Inaugural Award for Open Source Software in Ecology

An R package for the Australian plant trait database AusTraits was the recipient of the inaugural ARDC Award for New Developers of Open Source Software in Ecology from the Ecological Society of Australia.

"It is rare to come across awards and funding opportunities to support ECRs foster new professional interests and career transitions into areas such as research software development," said Dr Fonti Kar, a postdoctoral research associate at UNSW Sydney, who led the application.

"Receiving the award was a massive confidence boost. It will allow me to upskill certain aspects of package development and provide me opportunities to connect with other R developers to share ideas and experiences."

Bioinformatics Software Developer Award

Dr Michael Roach of Flinders University was "blown away" to learn that he had won the 'Torsten Seemann' Outstanding Bioinformatics Software Developer Award from the Australian Bioinformatics and Computational Biology Society. Sponsored by the ARDC, the award recognises an outstanding early-to-mid-career bioinformatic software developer.

"It was Torsten's talk at BioInfoSummer 2016 on developing bioinformatics software that motivated me to publish my first pipeline and I've been hooked on writing software ever since," said Dr Roach.

"I would love to pursue a career in academia as a group leader. There is, unfortunately, a lot of attrition at my career level, especially given the current state of research funding, but this award will go a long way in helping me to realise this dream." ■

Image — David Ma

Dr Michael Roach receives his ARDC-sponsored award from the Australian Bioinformatics and Computational Biology Society.

> QCIF has been pleased to participate in ARDC's Data Retention program. With hundreds of data collections residing on QCIF storage, we recognise the importance of metadata and for having a nationally consistent approach to metadata by program participants. That consistency has ensured a minimum level of metadata for each collection is available through DataCite and has made collections more readily findable for re-use, and more easily citable when used.
>
> **Stephen Bird,** Manager of Shared Infrastructure Services, *QCIF*

Data Retention

National data collections are expensive to create and maintain. Through our Data Retention program, we have partnered with the research sector to increase the impact of significant and important national data collections by investing in data storage infrastructure at scale.

Our partnerships leverage contemporary research data management practices to enrich data collections with controlled and consistent metadata, which is critical to further and future use of the collections as part of the scholarly record. Enriching data collections with metadata also means we can plan and maximise storage investments to underpin a sustainable and inclusive national digital research infrastructure.

Having a coherent, national view of significant data collections greatly increases their findability, accessibility and reusability and enables wider adoption of the FAIR and CARE principles.

ARDC investment has supported 1300 data collections, with a total storage capacity of 50 petabytes across 14 infrastructure partners and all research disciplines. ■

Digital Research Skills

The ARDC National Digital Research Skills Agenda brings together research institutions, peak bodies and other stakeholders in a coordinated approach to develop training infrastructure, encourage the sharing of training materials, and support key communities. With our partners, we have developed the tools and resources that identify and develop skills, roles, job profiles and learning pathways needed for an effective digital research sector. Increasingly, these capabilities and skills frameworks will ensure Australian researchers have access to the experts and expertise they need to use national digital research infrastructure effectively. ■

IMPACT SHOWCASE

Digital Research Skills Trainers Convene at the 2023 ARDC Skills Summit

The 2023 ARDC Skills Summit saw digital research trainers and skills professionals from across Australia work together within a culture of respect, collegiality, and open sharing. The trainer community came together to address the significant skills and workforce development challenges now faced by the research sector.

Our keynote speaker Jason Williams, Assistant Director at Cold Spring Harbor Laboratory's DNA Learning Center in New York, centred his talk on what makes a good researcher. He spoke about career progression and the role that effective, inclusive and targeted short-format training can play in ensuring that researchers maintain their edge across their entire career.

Day 1 of the summit focused on exploring the success factors for skills training. On Day 2, the conversation expanded to national approaches that narrow the digital research skills gap. ■

> I recommend this annual event for all involved in planning, coordinating, delivering or funding training for researchers to understand more about the importance and role of these skills in academia and industry. It's a great way to work collaboratively to solve common challenges we face.

Lyndal Gunton
Coordinator, Information Research Skills
QUT Library

Participants at the 2023 ARDC Skills Summit

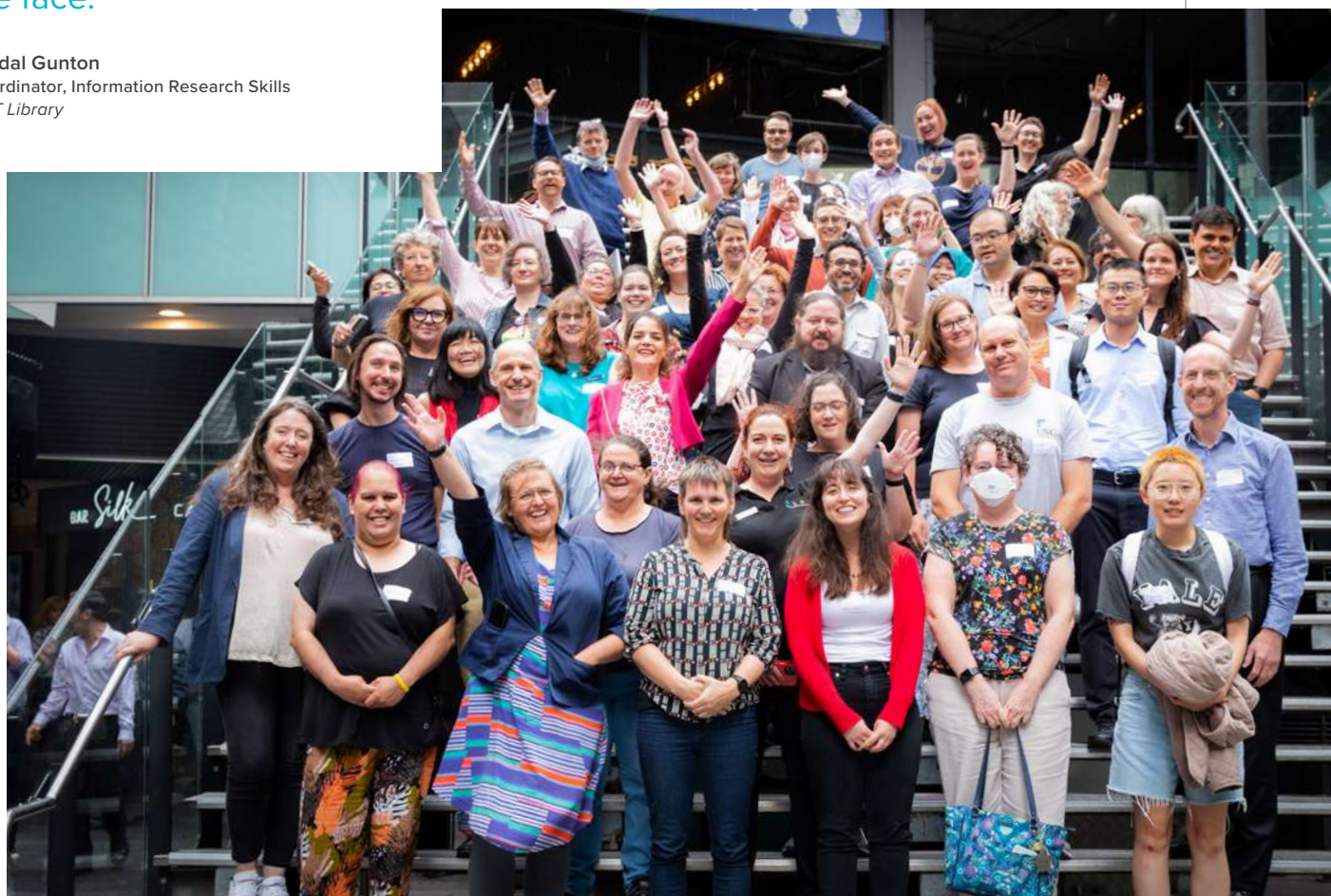


Image — Renee Nowytarger / ARDC

Image — Fabrice Jaine / Sydney Institute of Marine Science (CC BY-NC-ND)

> IMOS is proud to have gained certification from CoreTrustSeal for our Australian Ocean Data Network.
>
> **Dr Natalia Atkins,**
> Project Officer, Australian Ocean Data Network, *IMOS*

Collecting data for the Australian Ocean Data Network

IMPACT SHOWCASE

Ocean Data Repository Certified as Trustworthy

In late 2022, the Australian Ocean Data Network, which is managed by the Integrated Marine Observing System (IMOS), was recognised as a Trustworthy Data Repository with the internationally recognised CoreTrustSeal certification.

IMOS worked with the ARDC Trusted Data Repository Community of Practice to gain the certification. The community meets regularly to share, review and progress applications for certification; participants come from NCRIS facilities, federal government agencies and universities. Gaining CoreTrustSeal certification shows researchers and users that the repository provides a reliable, long term home for the data, it is well governed, and will meet the discipline needs.

IMOS was the first organisation to submit their application to CoreTrustSeal as part of the ARDC community, and several more are in the pipeline.

“IMOS is proud to have gained certification from CoreTrustSeal for our Australian Ocean Data Network,” said Dr Natalia Atkins, Project Officer for the data repository at IMOS.

“The process of gaining certification was invaluable for reviewing our documentation and processes [...] to ensure that we are a trustworthy data repository. In particular, the regular meetings with the Trusted Data Repository community and the writing sessions within our writing cohort were key to us progressing and finalising our application.” ■

Communities of Practice

The ARDC facilitates 16 communities and groups to strengthen bonds between people working in research data. Anyone with an interest in this area can join. These communities are a platform for sharing information and best practice, developing standards and approaches, solving problems and supporting peers. Some communities are established to respond to urgent issues, like the national community to support universities to become accredited data users to access Australian Government data under the *Data Availability and Transparency Act 2022*. ■

Engagements, Expertise & Support

The ARDC Engagements team helps the Australian research community tackle data and digital research challenges. We offer information and expertise on the FAIR principles, metadata and persistent identifiers, and on managing sensitive data and its governance; we can advise on data description and the publishing and sharing of data; and we identify appropriate national infrastructure available for use, which helps researchers conduct data-intensive research more efficiently and effectively.

We offer a range of workshops on self-evaluating data maturity, both at the project and the institutional level, such as the Research Infrastructure Self Evaluation (RISE) workshops. And we run the ARDC Leadership Series to engage decision-makers on topics such as sensitive data, digital research skills, and industry–researcher collaboration. ■

Our Input to Policy

The ARDC's Policy Program advocates for a coherent data policy environment so that Australia's research data infrastructure can accelerate innovation and foster engagement among Australian and international researchers, governments and industry.

Though we are neither a policy nor lobbyist organisation, we contribute to public discourse and influence policy owners whose rules affect the access, use and publishing of data by and for researchers.

Input to Australian Government Calls for Consultation

The ARDC provides specialist input to consultations about the treatment and use of data. In the past year, we have provided input to 22 consultations, including:

- Accreditation, Cost Recovery and Data Code (Office of the National Data Commissioner)
- Australian Data Strategy (Department of Prime Minister and Cabinet)
- Australian Research Council Act and Universities Accord (Department of Education)
- Copyright and Privacy Acts (Attorney General's Department)
- National Data Security Action Plan Discussion Paper (Department of Home Affairs)
- National Science and Research Priorities (Office of the Chief Scientist). ■

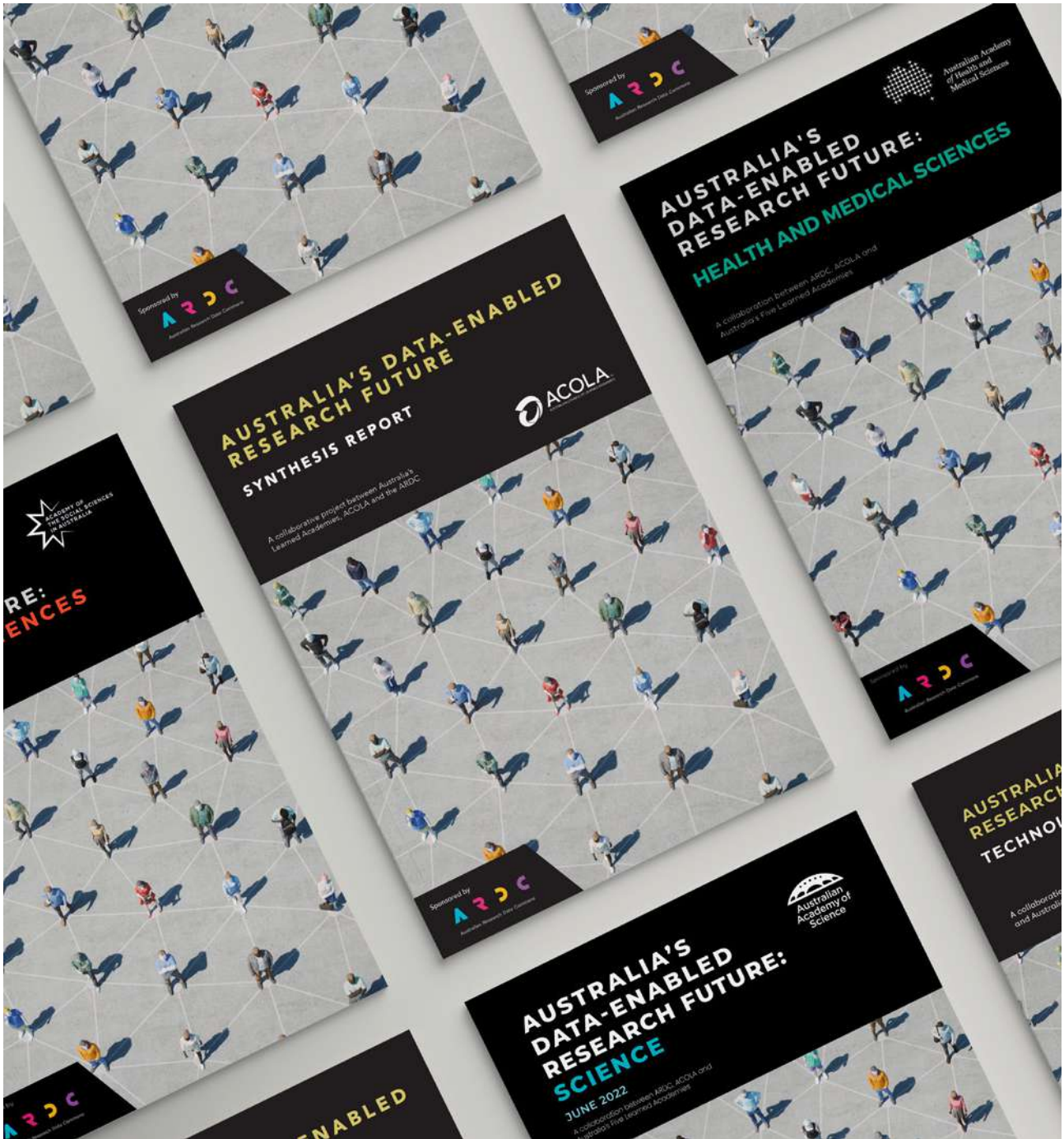
Australian Research Data Commons

IMPACT SHOWCASE

Mapping the Vision for Data-Driven Research in Australia

The pandemic and recent natural disasters have reinforced the importance of researchers having ready access to data to inform timely decision making. Coordinated and integrated data infrastructure is vital for making data available when it's needed for emergencies, and for answering research questions about the world around us.

To uplift data capabilities for researchers in Australia, the ARDC partnered with the Australian Council of Learned Academies (ACOLA) and the 5 Learned Academies to enable more data-driven research and decision making across all fields of research. ▶



> This work will help progress our vision for data-driven research to thrive in Australia. The academies and ACOLA have clear roles in uplifting data-driven research, alongside governments, the ARDC, universities and industry. Together we can ensure a fit-for-purpose and future-proof approach to Australian research data.
>
> **Professor Richard Holden,** Board Chair, *ACOLA;* President, *Academy of the Social Sciences in Australia*

Australia's 5 Learned Academies provide independent, authoritative and influential research-based advice across research domains, build public awareness and understanding of research, and champion, celebrate and support excellence in Australian research and innovation. The academies are:

- Academy of the Social Sciences in Australia
- Australian Academy of Health and Medical Sciences
- Australian Academy of the Humanities
- Australian Academy of Science
- Australian Academy of Technology and Engineering.

ACOLA is the forum in which these Learned Academies come together to contribute expert advice to inform national policy; and to develop innovative solutions to complex global problems and emerging national needs.

In June 2022, a suite of 6 reports was published arising from this partnership offering insights into opportunities for further driving Australian research and innovation through data, and guidance on how governments, industry and research institutions can achieve this.

Each of the 5 academies explored the current state, barriers and opportunities for realising an effective, data-enabled research sector. In the sixth report, ACOLA synthesises the insights from the academies, highlighting common themes, opportunities and the needs of multidisciplinary research.

The reports leverage the expertise and capabilities of some of Australia's greatest minds. They describe the potential of a data-enabled research future, where researchers have the skills, resources, tools and technologies to leverage data to its full extent.

Professor Richard Holden, ACOLA Board Chair and President of the Academy of the Social Sciences in Australia, said, "This work will help progress our vision for data-driven research to thrive in Australia. The academies and ACOLA have clear roles in uplifting data-driven research, alongside governments, the ARDC, universities and industry. Together we can ensure a fit-for-purpose and future-proof approach to Australian research data."

Broadly, the reports highlight 3 key ways to improve digital research in Australia:

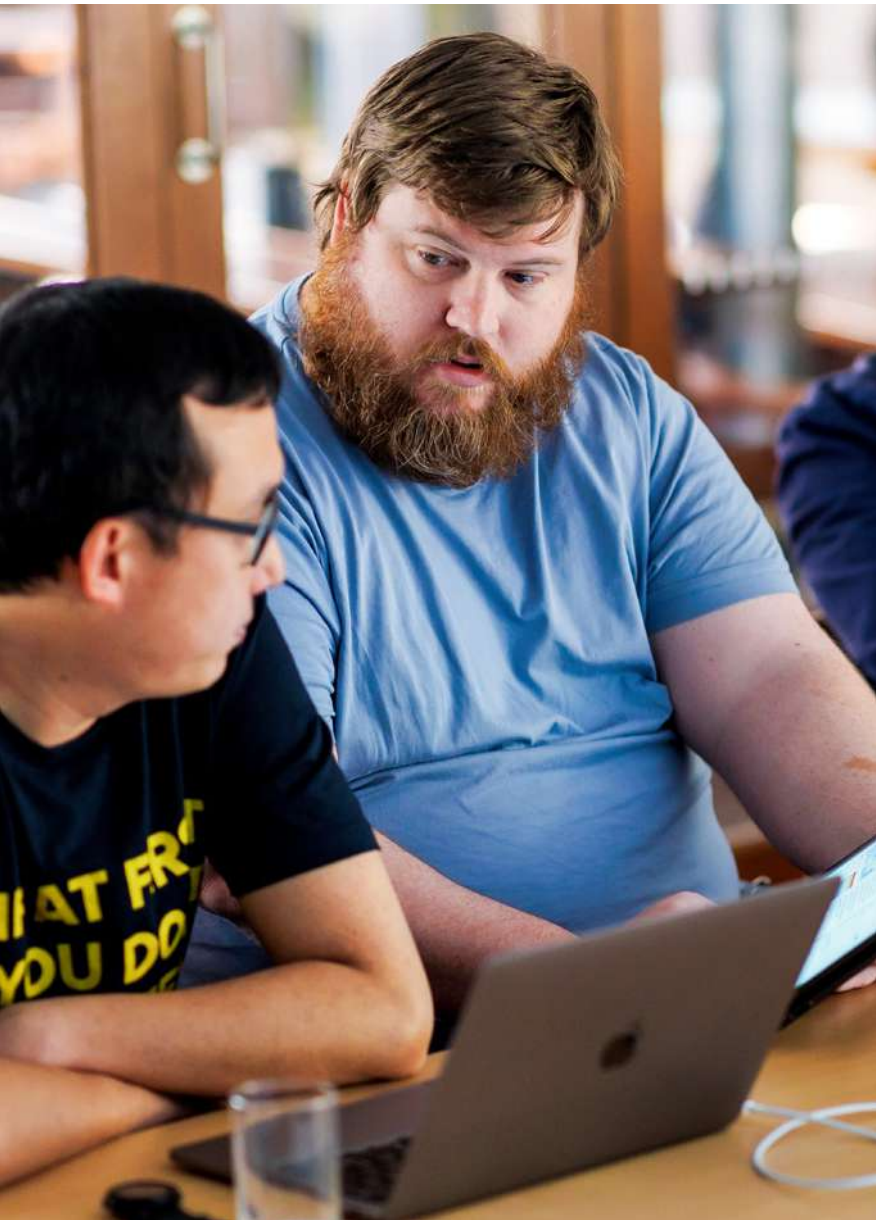
- Advocate for investment in data as a public good, but especially for research and innovation.
- Drive real action to implement the FAIR and CARE principles.
- Lead change and build capability to realise Australia's data-driven future.

OUR SERVICES

We provide the infrastructure that underpins national digital research capabilities and platforms, including our Thematic Research Data Commons. Our services operate at a national scale to support every aspect of the research data lifecycle, accelerating research and innovation.

We offer cloud computing for researchers through the ARDC Nectar Research Cloud, we make data discoverable through ARDC Research Data Australia, we help research institutes create and manage persistent identifiers, and much more. ■

Jake Yip from the ARDC with Michael Mallon from QCIF, a node of the ARDC Nectar Research Cloud, at a technical workshop in 2022



Image—Marc Grimwade / ARDC

ARDC Nectar Research Cloud
PAGE 92

IMPACT SHOWCASE
Understanding the Impact of Our Future Climate with the ARDC Nectar Research Cloud
PAGE 95

ARDC Research Data Australia
PAGE 99

Persistent Identifiers
PAGE 100

IMPACT SHOWCASE
Research Efficiency Through Persistent aIdentifiers
PAGE 101

ARDC Research Vocabularies Australia
PAGE 103

IMPACT SHOWCASE
FAIR Vocabularies in the Spotlight
PAGE 104

Resources for Researchers
PAGE 105

IMPACT SHOWCASE
Our New Website to Meet the Needs of the Research Community
PAGE 106

Image — Marc Grimwade / ARDC

The team behind the ARDC Nectar Research Cloud – team members from the nodes and the ARDC at a technical workshop in 2022

Nectar provides the underpinning cloud computing infrastructure for the ARDC's Thematic Research Data Commons and for over 200 services, including databases, data repositories, web portals and tools, and large-scale research platforms run by NCRIS facilities, institutions and research organisations.

ARDC Nectar Research Cloud

The ARDC Nectar Research Cloud (Nectar) is Australia's national research cloud, specifically designed for research computing.

Launched in 2012, Nectar provides Australia's research community with fast, interactive, self-service access to large-scale computing infrastructure, software and data. A powerful platform for collaboration, it allows researchers and research support staff to access compute resources, software and data from their office and home and easily share them with collaborators at other institutions.

Nectar is a co-designed federation, receiving co-investment from universities across Australia. The federation enables research collaboration across institutional boundaries and gives research collaborations access to research computing services at a national scale. ▶

AARNet is committed to developing and delivering technical solutions and infrastructure that meet the unique and specialised needs of researchers. Collaborating with the ARDC to develop a BinderHub service and build digital skills is a great example of this and, together, we can deliver many benefits to the research community.

Ryan Fraser, Director, Digital Research, *AARNet*

Nectar provides the underpinning cloud computing infrastructure for the ARDC's Thematic Research Data Commons and for over 200 services, including databases, data repositories, web portals and tools, and large-scale research platforms run by NCRIS facilities, institutions and research organisations.

In the past year, we launched 4 new services on Nectar, which we describe overleaf.

In addition to those new services, we are developing a national machine learning service, and we will soon launch our national BinderHub service, built in collaboration with AARNet. BinderHub allows researchers to share custom interactive computing environments from code repositories, enabling reproducible research. It takes considerable technical skill to set up BinderHub, which prevents many researchers from using it. Our national BinderHub service will give all Australian researchers easier access to BinderHub.

"AARNet is committed to developing and delivering technical solutions and infrastructure that meet the unique and specialised needs of researchers," said Ryan Fraser, Director of Digital Research at AARNet. "Collaborating with the ARDC to develop a BinderHub service and build digital skills is a great example of this and, together, we can deliver many benefits to the research community."

What Can Researchers Do With Nectar?

Nectar can be used in many ways to support research, such as a virtual desktop for a single researcher, or as a powerful computational server that can be shared by researchers in Australia and internationally.

Researchers can use Nectar to:

- access advanced research computing resources directly from home or in the office
- access large-scale computing including large memory machines and GPUs
- run Jupyter Notebooks or virtual desktops
- deploy and run software, workflows or platforms using containers and container orchestration
- customise their computing infrastructure to meet their needs
- collaborate globally to ensure their research meets the highest standards, with easy shared access to compute, software and data
- access and rapidly deploy and share research software tools and data to easily collaborate with peers
- benefit from centralised support, expert knowledge and continual development of cloud-based infrastructure designed for Australian researchers. ▶

Our Nectar Partners

Nectar is a federated cloud hosted at our partner institutions:

- The University of Melbourne
- Monash University
- University of Tasmania
- Queensland Cyber Infrastructure Foundation
- Intersect Australia Ltd
- Swinburne University of Technology
- University of Auckland

New ARDC Services Powered by Nectar

The services we host on Nectar can save researchers time, boost productivity and add the power they need for ground-breaking research.

Virtual Desktop Service

This service gives researchers an extra personal computer in the cloud. It's a simple way to access Nectar compute resources.

National GPU Service

Researchers can now reserve GPUs and large-memory virtual machines (16 flavours) through the Nectar dashboard.

Jupyter Notebook Service

This service makes it easier than ever to launch, develop and serve Jupyter Notebooks — a great way to develop and share code and computational output, with formatted explanatory text and multimedia.

Preemptible Instances

Researchers can easily access extra computational resources for a short time, on demand. The service is ideal for running automated workflows that need a lot of compute resources for a limited time. ■

Darcelle Malby and Sonia Ramza from the ARDC at an ARDC Nectar Research Cloud technical workshop in 2022

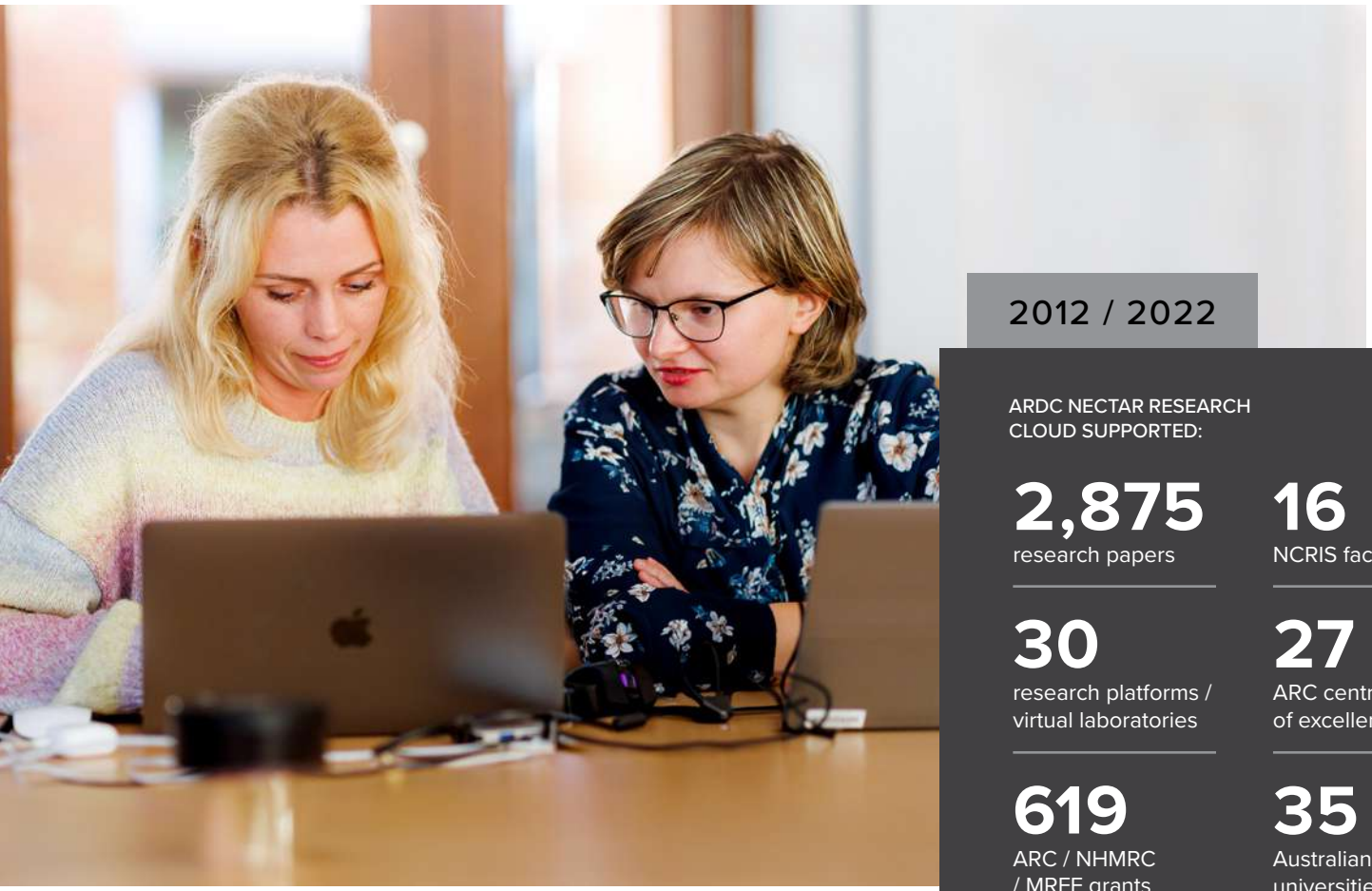


Image — Marc Grimwade / ARDC

2012 / 2022

ARDC NECTAR RESEARCH CLOUD SUPPORTED:

2,875 research papers

16 NCRIS facilities

30 research platforms / virtual laboratories

27 ARC centres of excellence

619 ARC / NHMRC / MRFF grants

35 Australian universities

Australian Research Data Commons

IMPACT SHOWCASE

Understanding the Impact of Our Future Climate with the ARDC Nectar Research Cloud

For almost 10 years, the ARDC Nectar Research Cloud (Nectar) has been vital for Dr Tom Remenyi's research to understand the future impacts of climate change. It has allowed him to analyse big climate datasets and model future climate scenarios.

A senior climate research fellow at the University of Tasmania's Climate Futures Group, Dr Remenyi was one of the first researchers in Tasmania to use Nectar when it was launched at the Tasmanian Partnership for Advanced Computing node in 2014. Since then, he has completed 30 projects using Nectar with partners from industry, state and federal governments, and academia. ▶

The outbreak of fires in the Tasmanian Wilderness World Heritage Area in 2015 and 2016 underscored the need to understand the risk of fire in the area under future climate scenarios.

Dr Tom Remenyi

> The bushfire research industry has a lot of technically minded people that are doing a lot of modelling, so they want to be able to work with the code.

Dr Tom Remenyi
Senior Climate Research Fellow,
Climate Futures Group,
University of Tasmania

Government and Academia Collaborate to Analyse Climate and Bushfire Data for World Heritage Area

The outbreak of fires in the Tasmanian Wilderness World Heritage Area in 2015 and 2016 underscored the need to understand the risk of fire in the area under future climate scenarios. Understanding the risk would help the government and authorities to better prepare for, and respond to, bushfires in the area, which spans almost 20 per cent of Tasmania's land surface.

Collaborating with natural disaster experts in the Tasmanian state government, Dr Remenyi and colleagues worked on the Tasmanian Wilderness World Heritage Area Climate Change and Bushfire Research Initiative.

The project assessed bushfire risk across Tasmania, which has influenced bushfire policy and state emergency management policy.

To aid the development of the planned burning strategy for the area, Dr Remenyi and colleagues developed new and updated decision-making tools.

Nectar made it easy to collaboratively test data analyses for projecting future bushfires in the area, said Dr Remenyi. Instead of his government partners needing to laboriously download data and software to undertake analysis, he gave them access to Nectar, where they could easily access what they needed to run transformations and test fire network systems.

"The bushfire research industry has a lot of technically minded people that are doing a lot of modelling, so they want to be able to work with the code," said Dr Remenyi. Access to Nectar allowed them to do that. ▶

Image — Janelle - 202711780 / AdobeStock.com

> One thing I've loved about Nectar is it allows people to learn things that can match their capacity.

Dr Tom Remenyi
Senior Climate Research Fellow,
Climate Futures Group,
University of Tasmania

Climate Atlas Reveals Australia's Wine Future

Working with Wine Australia, an Australian government statutory authority, Dr Remenyi and colleagues published *Australia's Wine Future – A Climate Atlas*.

A world-first, the atlas details how Australia's famous wine-growing regions will be forced to adapt to climate change over the next 80 years. For each region, it describes trends for growing conditions under future climate conditions, including changes in temperature, rainfall, evaporation, and heat and cold extremes.

Throughout the 4-year project developing the atlas, the data analysis and modelling were run on Nectar.

Winegrowers are using the atlas to understand the future climate of their region and change how they manage their vines. They are seeking new techniques for managing soil, and for managing and irrigating vines; some may even change varieties to suit their changing climate. ▶

A Valuable Service for Upskilling the Next Generation of Researchers

Nectar has been the ideal platform for PhD and early-career researchers to learn the digital research skills they need for cutting-edge research, said Dr Remenyi.

“One thing I’ve loved about Nectar is it allows people to learn things that can match their capacity,” he said.

“There’s so much to learn, and there’s such a big gap between working on your laptop or home computer and using a supercomputer – and Nectar fills that gap perfectly. They need something that can work as if you’re on a laptop but has the grunt of a bigger machine. Nectar allows for this transition.”

Dr Remenyi and his colleagues have trained dozens of higher-degree and early-career researchers using Nectar as a learning space where researchers can develop their computational skills.

“By the time you start getting to the top end of what Nectar offers, you’re already doing multi-threading, and you are then set to ask NCI [the National Computational Infrastructure] for time on a supercomputer. You have built someone up to a point where they at least know why they have to do it this way.” ■

> There’s so much to learn, and there’s such a big gap between working on your laptop or home computer and using a supercomputer – and Nectar fills that gap perfectly.

Dr Tom Remenyi
Senior Climate Research Fellow, Climate Futures Group, *University of Tasmania*

Image — brooke-cagle - gKt4Ozfoac / unsplash.com

Australian Research Data Commons

ARDC Research Data Australia

Operating since 2011, Research Data Australia is an online portal for finding research data and associated projects, researchers and data services. More than 100 Australian data sources — including research organisations, government agencies and cultural institutions — contribute to the portal.

Research Data Australia covers a broad spectrum of research fields in science, technology, engineering, mathematics, social sciences, arts and humanities.

Visitors to the portal can also search for Australian research grants and projects, including grant information from Australia's principal research funders, and for project descriptions from various institutions and agencies. ■

2021 / 2022

ARDC RESEARCH DATA AUSTRALIA:

202,585
data collections

891
data services and tools listed

784,142
page views

59,449
searches conducted

Alice Springs Mulga Flux Data Collection
Provider: Terrestrial Ecosystem Research Network (TERN)

This dataset consists of measurements of the exchanges of energy, carbon and water between the ecosystem and the atmosphere in semi-arid mulga woodlands, using eddy covariance techniques. The TERN Alice Mulga SuperSite is located on Pine Hill cattle station, near Alice Springs in the Northern Territory. It is managed by TERN in collaboration with James Cook University and the Northern Territory Government.

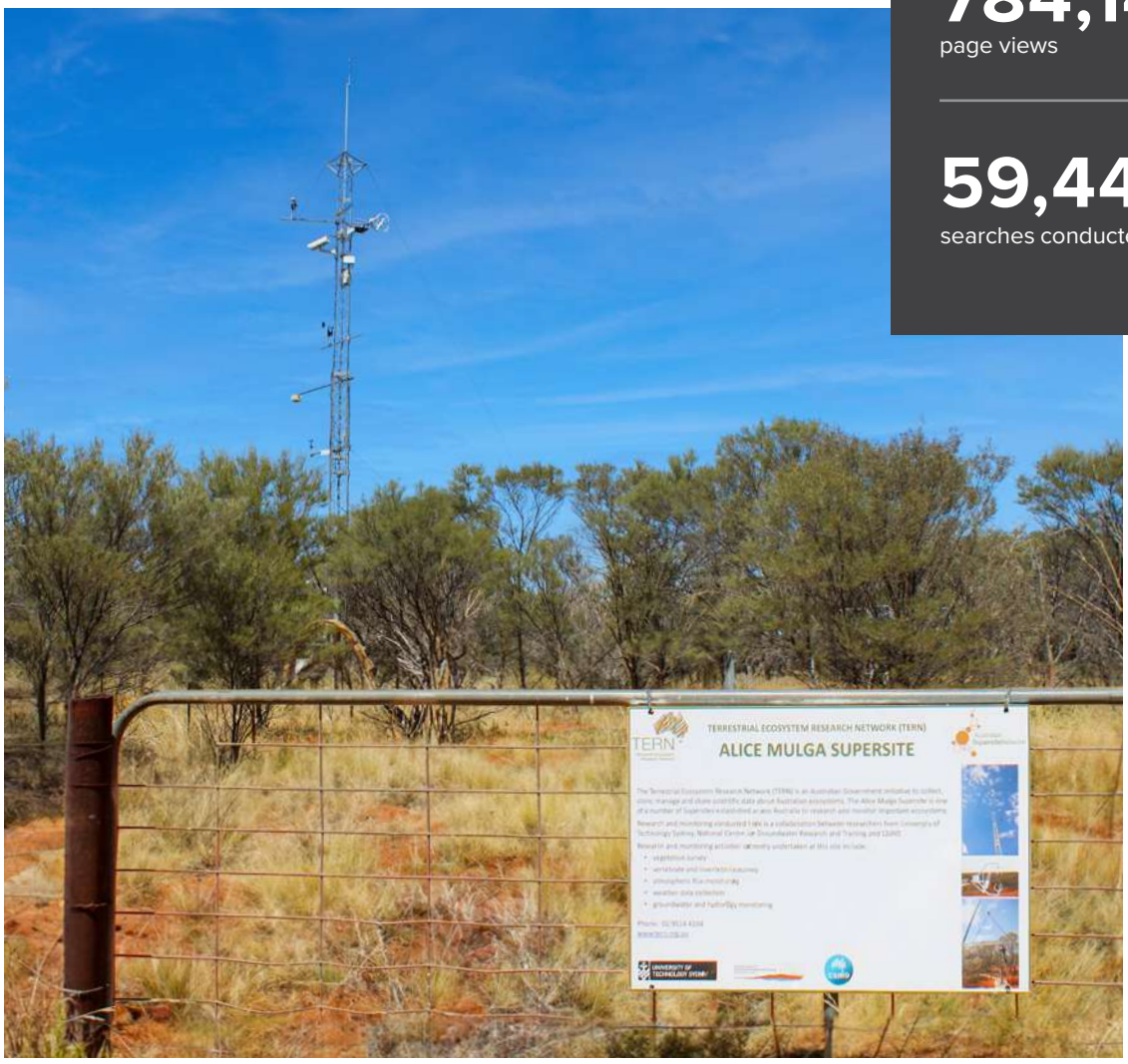


Image — Jamie Cleverly / TERN, James Cook University

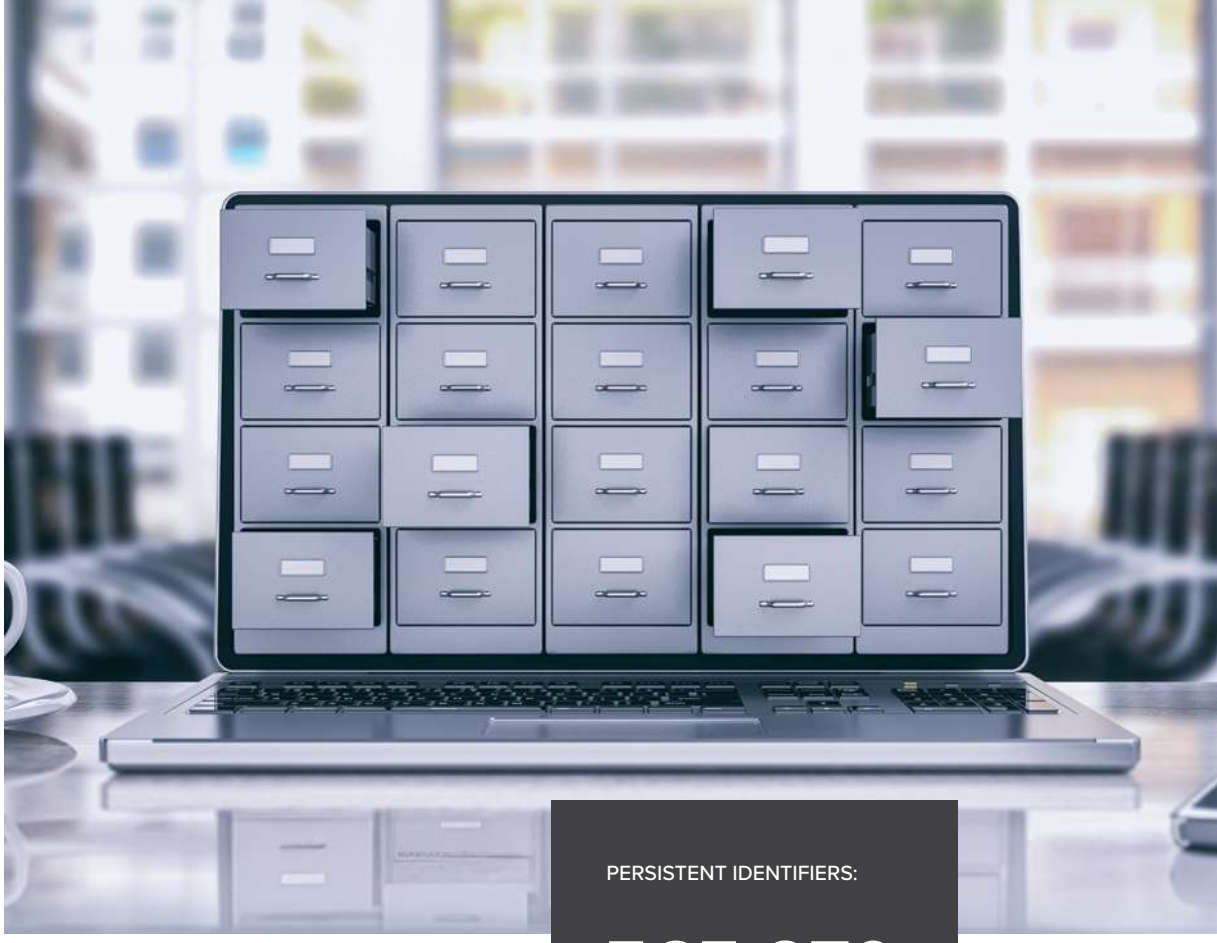


Image — Rawf8 - 162076001 / AdobeStock.com

Persistent Identifiers

The ARDC offers several services that enable researchers at Australian institutions to assign globally unique persistent identifiers to their data and to related materials such as samples, files, documents, tools and services. This supports researchers in discovering, citing and attributing these resources.

In partnership with providers of international persistent identifier systems — and at no cost to Australian research organisations and institutions — the ARDC provides the following 4 types of persistent identifier, each of which may be linked to a researcher's ORCID:

2021 / 2022

PERSISTENT IDENTIFIERS:

565,970
DOIs minted

435,652
handles minted

3,574
RAiDs minted

Digital Object Identifier (DOI)

A DOI is used internationally by researchers to identify research data and related objects such as software. It provides a persistent link to the object's location on the internet and a description of it. This supports research citation and tracking of citation metrics.

International Generic Sample Number (IGSN)

An IGSN is a globally unique persistent identifier for physical samples. The service facilitates the location, identification, and citation of physical samples used in research.

Research Activity Identifier (RAiD)

A RAiD is a persistent identifier for research projects and activities. It connects existing persistent identifiers for researchers, institutions, outputs and tools to create a timeline of research activity, making the tracking of research projects and their outputs clear and easy.

Handle

A handle is a reference to datasets, collections, papers and other research objects, which persists even if the location of the objects changes. ■

IMPACT SHOWCASE

Research Efficiency Through Persistent Identifiers

In 2022, the ARDC and the Australian Access Federation (AAF) commissioned MoreBrains Cooperative to analyse the incentives for the Australian research sector to adopt persistent identifiers (PIDs) — the globally unique digital codes that identify entities such as people, places and things.

The results[12], published on 1 October 2022, found that:

- researchers spend **30–40 per cent** of their time on administrative tasks
- retyping metadata about grants, publications and projects takes nearly **38,000 person days per year**
- the direct financial cost of this duplicated effort is nearly **$24 million per year**. When the opportunity cost associated with technology transfer and innovation-led growth is factored in, the cost expands to **$84 million per year**.

In the Australian Research Council's (ARC) 2020 review *Excellence in research for Australia* and its *Engagement and impact assessment*, reducing the administrative burden on universities and researchers was seen as a priority. Less administration would enable them to focus on conducting meaningful and impactful research in Australia's national interest, which has become a key policy goal.

There is growing interest in using PIDs to address the administrative burden. PIDs offer a way to embed metadata in entity descriptions and records at the point of creation or publication of the entity. They also provide a way to openly store metadata in standardised formats that are both human- and machine-readable, making information exchange easier and eliminating the need to tediously retype information into multiple systems.

The MoreBrains report includes case studies from the ARC and Australia's Terrestrial Ecosystem Research Network (TERN) that demonstrate the value that PIDs can bring. Professor Joe Shapter, former pro-vice-chancellor for research infrastructure at The University of Queensland, gives a personal account of his experience submitting grant applications. By connecting his ARC profile to his ORCID record, his publication record is automatically added to his grant applications, saving **3–4 days of effort per submission**—"a mountainous saving of work".

The report also illustrates how the Australian research sector benefits from:

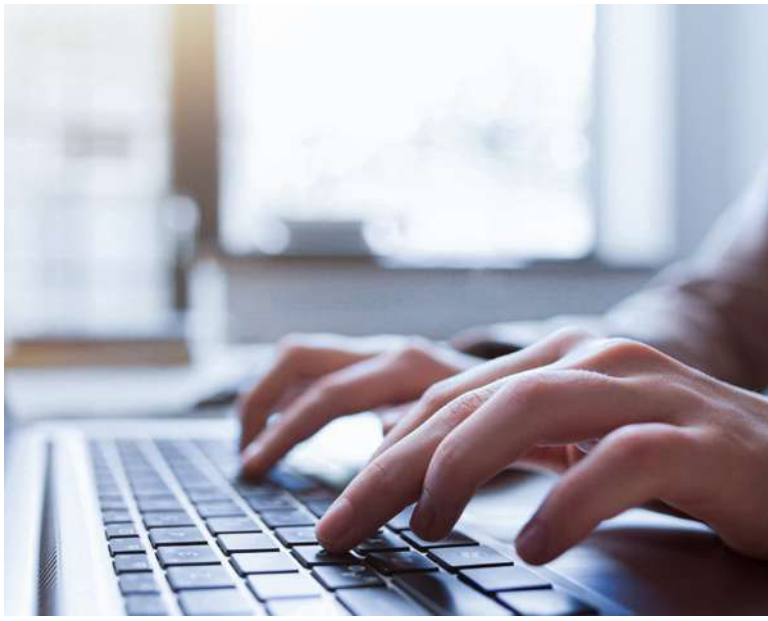
- the AAF-led ORCID Consortium, which has saved its 43 member institutions $4.5 million since 2016 and provides valuable local support for ORCID system integrations
- the ARDC's suite of PID services, provided free to Australian research institutions, including services for identifying and linking research data (DOIs), grants (PURL), physical samples (IGSNs) and projects (RAiDs).

A National PID Strategy for Australia

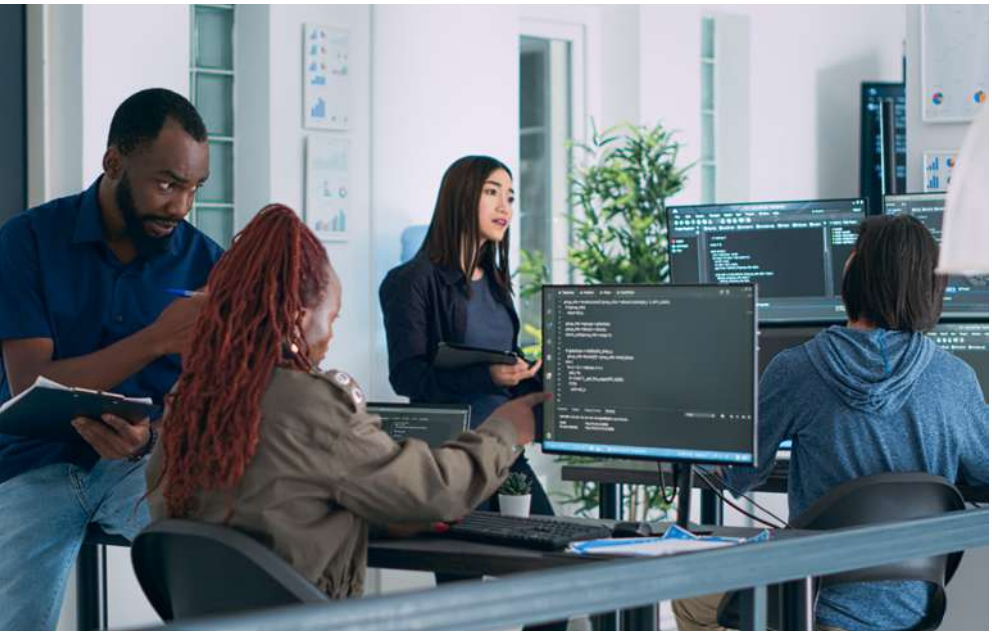
The ARDC is now coordinating the collaborative development of a national PID strategy and roadmap for Australia — a key recommendation of the MoreBrains report. This involves defining the aspirations of the research sector for PIDs over the next 5 years and agreeing on the steps needed to realise these. A National PID Strategy Taskforce has been formed to advise on the development of the strategy, in which working groups will also provide input. An open public consultation commenced in February 2023 with a national workshop. ■

12 *Brown J, Jones P, Meadows A, Murphy F. Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems. 2022. Zenodo. doi.org/10.5281/zenodo.7100578*

Researchers spend **30–40 per cent** of their time on administrative tasks.



Retyping metadata about grants, publications and projects takes nearly **38,000 person days per year.**



The direct financial cost of this duplicated effort is nearly **$24 million per year**. When the opportunity cost associated with technology transfer and innovation-led growth is factored in, the cost expands to **$84 million per year.**

ARDC Research Vocabularies Australia

Since 2015, Research Vocabularies Australia has been used to create, maintain, find, access, and reuse vocabularies for research. It caters for researchers and those who support, describe and discover research, including vocabulary managers, ontologists, data managers and librarians.

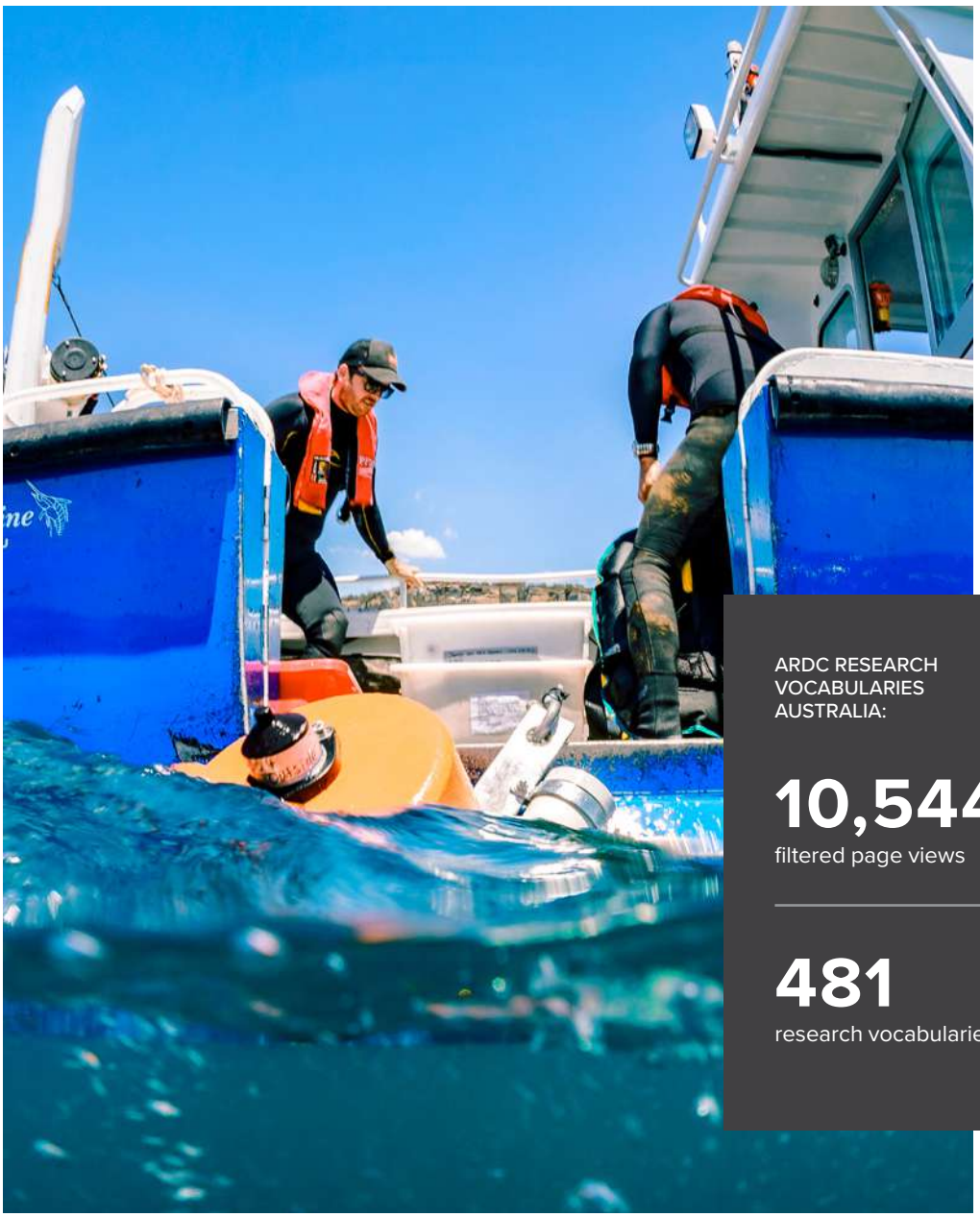


Image — Fabrice Jaine / Sydney Institute of Marine Science (CC BY-NC-ND)

ARDC RESEARCH VOCABULARIES AUSTRALIA:

10,544
filtered page views

481
research vocabularies

2021 / 2022

Australian Ocean Data Network Instrument Vocabulary
Publisher: Integrated Marine Observing System (IMOS)

This is an Australian Ocean Data Network (AODN) controlled vocabulary describing instruments. The AODN is the data management component of the NCRIS-funded IMOS.

Image — kasto - 547892205 / AdobeStock.com

IMPACT SHOWCASE

FAIR Vocabularies in the Spotlight

The 2022 Vocabulary Symposium was held on 14–15 November online and in person at the Australian National University. Co-produced by the ARDC, the Australian Data Archive and CODATA, it brought together more than 200 users, creators and publishers of vocabularies across multiple domains and sectors in Australia and internationally.

Participants shared experiences and identified requirements for FAIR (findable, accessible, interoperable and reusable) vocabularies that underpin cross-domain data. Controlled vocabularies are an important part of research and scholarly communication. They provide a consistent way to describe data, so research can be replicated and verified. Vocabularies are also a key part of making data FAIR to support data discovery and reuse.

> Australia's been playing a leading role globally in data stewardship and good data practices. It started off with the Australian National Data Service and now that's being continued by the ARDC. It has been an effective and influential model for the rest of the world.
>
> **Dr Simon Hodson,**
> Executive Director,
> *CODATA*

The symposium was followed by an invitation-only workshop where a spectrum of domains was represented, including government; humanities, arts and social sciences; health and medicine; Indigenous studies; and geospatial, earth and environmental sciences. The diverse participants worked through hands-on exercises and scenarios to map out the current vocabulary landscape and identify future requirements. ■

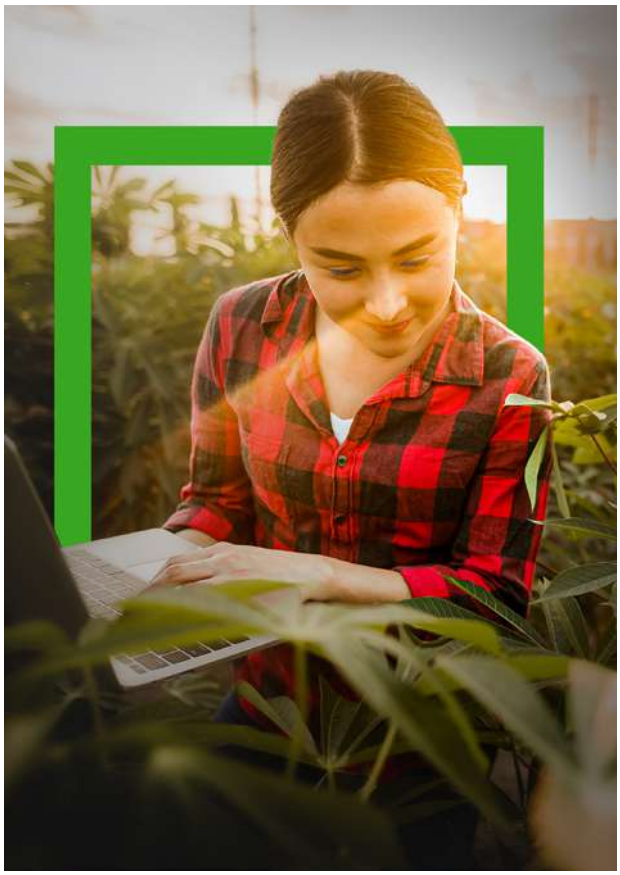
Resources for Researchers

The ARDC provides curated resources for digital research on our website. We share datasets, platforms, free tools and upskilling materials for data-intensive research in the areas of our 3 Thematic Research Data Commons:

HASS and Indigenous

People

Planet



Australian Research Data Commons

IMPACT SHOWCASE

Our New Website to Meet the Needs of the Research Community

In October 2022, we launched a refreshed website, **ardc.edu.au**, for the Australian research community. It shares our rapidly expanding ecosystem of research infrastructure tools, services and resources that can accelerate data-driven research in Australia and beyond.

The refreshed website provides researchers, research institutions, industry, digital research infrastructure providers, and our national and international partners with intuitive navigation and improved search functionality.

We collaborated with Australian researchers across all career stages and fields of research and key stakeholders in the sector to ensure it meets their needs.

Visit the website to learn more about our strategy, programs, projects, news, resources, events and how to get involved in our activities. Researchers can find resources on data management, research software, sensitive data, FAIR data guidance, cloud computing, and more. ■

> We are excited to launch our new ARDC website, which showcases the breadth of digital research infrastructure tools, communities and guides supported by the ARDC that help accelerate research.

Keith Russell,
Director of Outreach,
ARDC

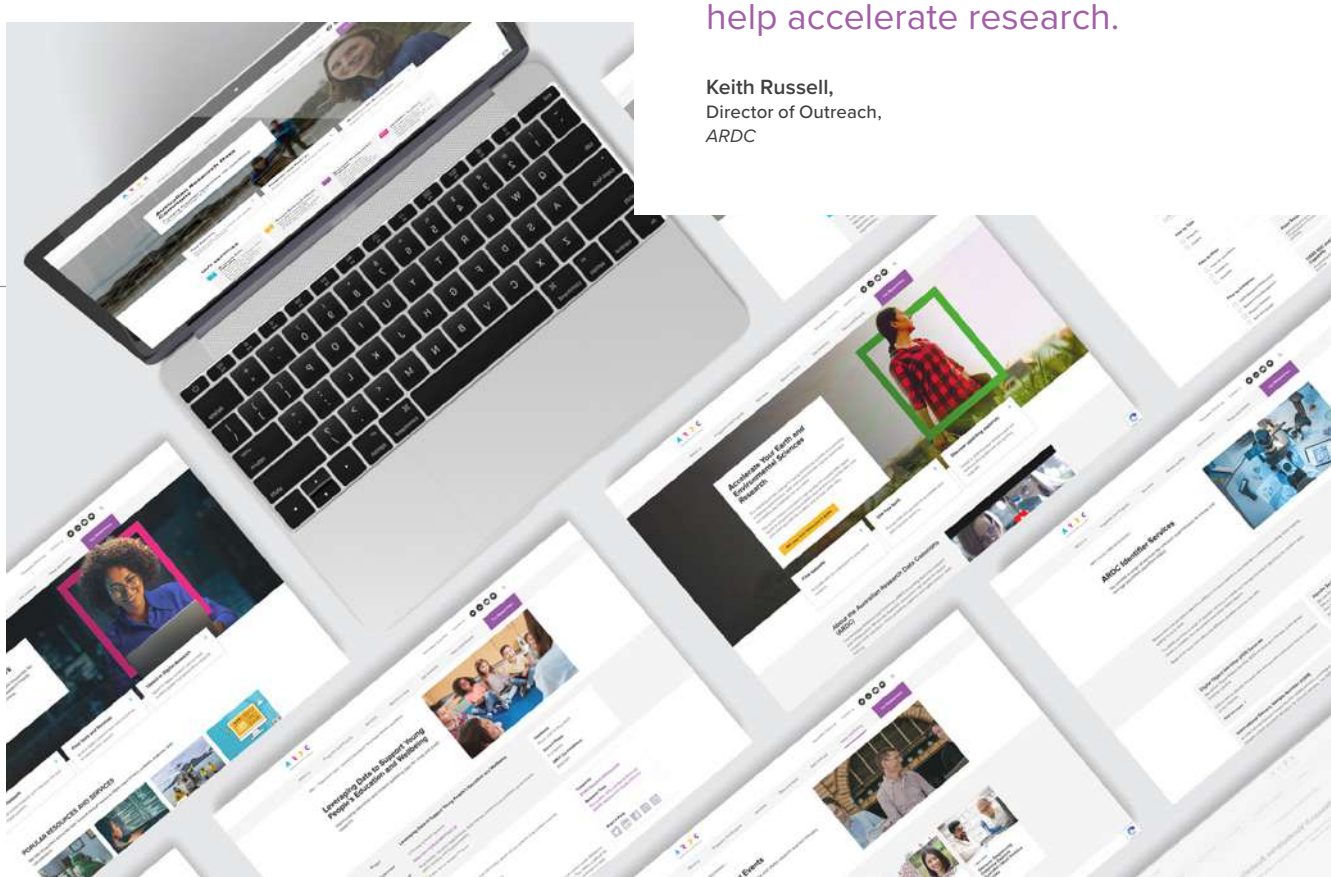


Image — Anton Balazh - 209276103 / AdobeStock.com

ACKNOWLEDGEMENTS

We'd like to acknowledge all of our members, partners, collaborators and community for their contribution and support. The ARDC would not be possible without you.

Partnerships are essential for developing a national data and computational research infrastructure environment, and a partnership approach is a core principle for the ARDC. Our national and international collaborations and partnerships ensure that valuable data and software assets are developed, made available and sustained for everyone.

Our Members and Memberships
PAGE 108

Our Partners
PAGE 109

Australian Research Data Commons

OUR MEMBERS AND MEMBERSHIPS

Our Members

Australian National University

CSIRO

Curtin University

Edith Cowan University

Federation University

Flinders University

Garvan Institute of Medical Research

Griffith University

Macquarie University

Monash University

Queensland University of Technology

RMIT University

SAHMRI

Swinburne University of Technology

The University of Adelaide

The University of Melbourne

The University of Notre Dame Australia

The University of Queensland

The University of Sydney

The University of Western Australia

University of South Australia

University of Tasmania

University of Technology Sydney

University of Wollongong

UNSW Sydney

Our National Memberships

Australian Data Science Network

Australasian eResearch Organisation (AeRO)

Science & Technology Australia

Our International Memberships

DataCite (member and Consortium Lead)

International Data Space Association (member)

International Generic Sample Number (IGSN e.V., member)

OpenInfra Foundation (associate member)

Research Data Alliance (member)

Research Software Alliance (member)

Image — Adam Calaitzis - 163234683 / AdobeStock.com

Australian Research Data Commons

Australia's national research infrastructures are well connected and actively collaborate to provide the research sector with access to national data assets, research platforms, tools and skills. The ARDC has been able to harness this unique positioning to advance Australia's national digital research infrastructure creating a seamless research ecosystem.

Rosie Hicks, Chief Executive Officer, *Australian Research Data Commons*

OUR PARTNERS

Peak Bodies

- Academy of the Social Sciences in Australia
- Australian Academy of Health and Medical Sciences
- Australian Academy of Science
- Australian Academy of the Humanities
- Australian Academy of Technological Sciences & Engineering
- Australian Council of Learned Academies
- Council of Australasian University Directors of Information Technology
- Council of Australian University Librarians
- Research funders (ARC, NHMRC, MRFF)

Research Organisations

- ARC centres of excellence
- Cooperative research centres
- Medical research institutes
- Publicly funded research agencies
- Australian universities

Research Infrastructure Providers

- Commercial digital infrastructure providers
- Digital data and research platform (DDeRP) members: AARNet, Australian Access Federation, ARDC, National Computational Facility, Pawsey Supercomputing Centre
- Galleries, libraries, archives, museums and repositories
- National Collaborative Research Infrastructure Strategy (NCRIS) facilities
- Other national and state infrastructure providers (including the ARDC Nectar Research Cloud node partners)

International Partnerships

- American Geophysical Union
- Committee on Data of the International Science Council (CODATA)
- DataCite
- Earth Science Information Partners (ESIP)
- FORCE11
- Global Open Science Cloud
- Open Researcher and Contributor ID (ORCID)
- OpenStack community
- Organisation for Economic Co-operation and Development (OECD)

Key Government Departments We Work With

- Australian Government Department of Climate Change, Environment, Energy and Water
- Australian Government Department of Prime Minister and Cabinet
- NSW Department of Health
- NSW Department of Planning, Industry and Environment
- NSW Department of Primary Industries
- NT Department of Environment, Parks and Water Security
- NT Department of the Attorney-General and Justice
- QLD Department of Agriculture and Fisheries
- QLD Department of Environment and Science
- QLD Department of Health
- SA Department for Education
- SA Department for Environment and Water
- TAS Department for Education, Children and Young People
- TAS Department of Natural Resources and Environment
- TAS Department of Justice
- VIC Department of Energy, Environment and Climate Action
- WA Department of Biodiversity, Conservation and Attractions
- WA Department of Fire and Emergency Services
- WA Department of Health
- WA Department of Justice
- WA Department of Primary Industries and Regional Development
- WA Department of Transport

Image — baspley - 500422288 / AdobeStock.com

Key Australian Government Agencies and Offices We Work With

- Australian Bureau of Statistics
- Australian Centre for International Agricultural Research
- Australian Institute of Aboriginal and Torres Strait Islander Studies
- Australian Institute of Health and Welfare
- Australian Institute of Marine Science
- Bureau of Meteorology
- CSIRO
- CSIRO's Data61
- Defence Science and Technology Group
- Geoscience Australia
- National Emergency Management Agency
- Office of the National Data Commissioner

Australian Research Data Commons

GOVERNANCE

The ARDC governance structure includes the ARDC Board, Research and Technology Advisory Committee and Senior Leadership Team.

The ARDC was formed in 2018, enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). As part of our governance activities, we report annually to the Australian Government Department of Education, which administers NCRIS. In May 2019, the ARDC became a company limited by guarantee and is registered with the Australian Charities and Not-for-profits Commission.

Our Board of Directors sets the strategic objectives of the ARDC and ensures the company is best positioned to deliver on its commitments and obligations under the NCRIS Funding Agreement. The ARDC Research and Technology Advisory Committee provides independent advice to the CEO relating to longer-term national digital research infrastructure requirements and the efficacy of current ARDC programs. ■

Board of Directors
PAGE 112

Research and Technology Advisory Committee
PAGE 113

Senior Leadership Team
PAGE 114

Our Team
PAGE 115

Image — Who is Danny - 291529560 / AdobeStock.com

BOARD OF DIRECTORS

Mr Craig Roy

FAICD MBA MSc (Chair)

Chair, *Silex Systems Limited (ASX: SLX), Global Laser Enrichment (USA, LLC)*

Non-Executive Director, *Sydney Water and Sovereign Manufacturing Automation for Composites Cooperative Research Centre (SoMAC)*

Honorary Fellow,
University of Technology Sydney

Dr Sue Barrell

AO FTSE

Chair, *Australia's Marine National Facility Steering Committee, Stawell Underground Physics Laboratory Ltd and Reef Restoration and Adaptation Program Intervention Risk Review Group*

Member, *UTS Council and Australian Mathematical Sciences Institute Board*

Advisory Board Member, *ARC Centre of Excellence for Dark Matter Particle Physics*

Ms Anne-Marie Lansdown

Independent Board Member,
Cooperative Research Australia

Advisory Board Member, *ARC Centre of Excellence in Synthetic Biology*

Ms Toni Moate

PSM FCPA GAICD

Director, *National Collections and Marine Infrastructure, CSIRO*

Member, *National Marine Science Committee*

Ex-Officio member, *Atlas of Living Australia Advisory Board*

Professor Joe Shapter

Adjunct Professor, *Flinders University and The University of Queensland*

Member, *Advisory Group for National Research Infrastructure, Department of Education*

Independent Chair, *The Australian ORCID Consortium Governance Committee*

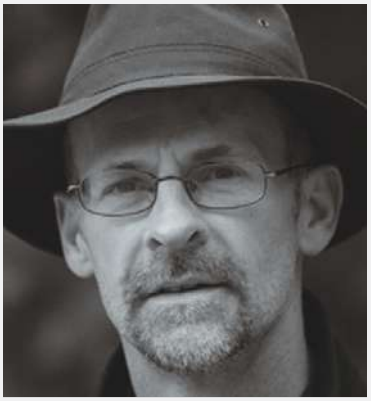
Mr Russell Yardley

Non-Executive Director,
Alannah & Madeline Foundation

RESEARCH AND TECHNOLOGY ADVISORY COMMITTEE

Dr Ian Oppermann
(Committee Chair)
Chief Data Scientist / CEO, *NSW Data Analytics Centre at NSW Government*

Professor Joanna Batstone
Director, *Monash Data Futures Institute, Monash University*



Emeritus Professor Stephen Dovers FASSA
Fenner School of Environment and Society, Australian National University

Dr Stephanie Duce
Lecturer in Spatial Science
James Cook University

Mrs Rosie Hicks
(ex-officio)
CEO, *ARDC Ltd*

SENIOR LEADERSHIP TEAM

Mrs Rosie Hicks
Chief Executive Officer

Dr Adrian Burton
Deputy CEO
Director, *People Research Data Commons*

Mr Matt Dailey
Chief Operations Officer

Dr Emma Crott
Chief of Staff
Company Secretary

Ms Jenny Fewster
Director, *HASS and Indigenous Research Data Commons*

Dr Sheida Hadavi
Director, *Translational Research Data Challenges*

Mr Hamish Holewa
Director, *Planet Research Data Commons*

Mr Keith Russell
Director, *Outreach*

Ms Natasha Simons
Director, *National Coordination*

Dr Andrew Treloar
Director, *International Strategy*

Australian Research Data Commons

OUR TEAM

Dr Muhammad Ali
Madhu Bahl
Melanie Barlow
Joel Benn
Dr Gnana Bharathy
Kylie Black
Andy Botting
Catherine Brady
John Brown
Rowan Brownlee
Dr Robin Burgess
Dr Adrian Burton
Dr Nichola Burton
Shannon Callaghan
Ping Chen
Myra Cheng
Dr Paul Coddington
Adelle Coote
Stephen Crawley
Dr Emma Crott
Matt Dailey
Shubhra Dargar
Horatio Davis
Dr Richard Ferrers
Jenny Fewster
Mary Filsell
Dr Sheida Hadavi
Doris Harrison
Rosie Hicks
Dr Tom Honeyman
Tamara Hooper
Tika Kandel
Dr Kristan Kang
Dr Stefanie Kethers
Tran Le
Kerry Levett
Matthias Liffers
Liz Lopes
Darcelle Malby
Julia Martin
Dr Paula Andrea Martinez
Siobhann McCafferty
Dylan McCulloch
Julie McCusker
Bernard Meade
Leo Monus
Jo Morris
Sam Morrison
Paul Murphy
Fiona Nelson Campbell
Kelli O'Hern
Cel Pilapil
Sonia Ramza
Timothy Rice
Lisa Rigney
Prof Shawn Ross
Keith Russell
Jo Savill
Olivia Shanahan
Natasha Simons
Jonathan Smillie
Adrian Smith
Mihail Staicu
Liz Stokes
Amanda Tombolato
Dr Andrew Treloar
Dr Aleem Uddin
Kathryn Unsworth
Asher Vennell
David Vitkin
Richard Walker
Grant Ward
Jaye Weatherburn
Andrew White
Dr Max Wilkinson
Dr Rhys Williams
Dr Lyle Winton
Adeline Wong
Liz Woods
Dr Mingfang Wu
Dr Lesley Wyborn
Rocky Yan
Jake Yip
Jason Yuen
Enming Zhang

Australian Research Data Commons

Image — Anthony McKee / ARDC

Australian Research Data Commons

HEAD OFFICE

Monash University, Building T,
Ground Floor, 100 Sir John Monash Drive
Caulfield East, VIC 3145, Australia

CONTACT

ardc.edu.au
+61 3 9902 0585
contact@ardc.edu.au

FOLLOW

@ardc_au
australian-research-data-commons
ARDC_AU
Subscribe to the ARDC Connect Newsletter
https://ardc.edu.au/subscribe/

MEMBERS

Australian National University
CSIRO
Curtin University
ECU Edith Cowan University
Federation University
Flinders University
Garvan Institute of Medical Research
Griffith University Queensland, Australia
Macquarie University Sydney · Australia
Monash University
QUT
RMIT University
SAHMRI South Australian Health & Medical Research Institute
Swinburne University of Technology
The University of Adelaide
The University of Melbourne
The University of Notre Dame Australia
The University Of Queensland Australia
UNSW Sydney
University of South Australia
The University of Sydney
The University of Western Australia
University of Tasmania
UTS
University of Wollongong Australia

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The ARDC is enabled by NCRIS