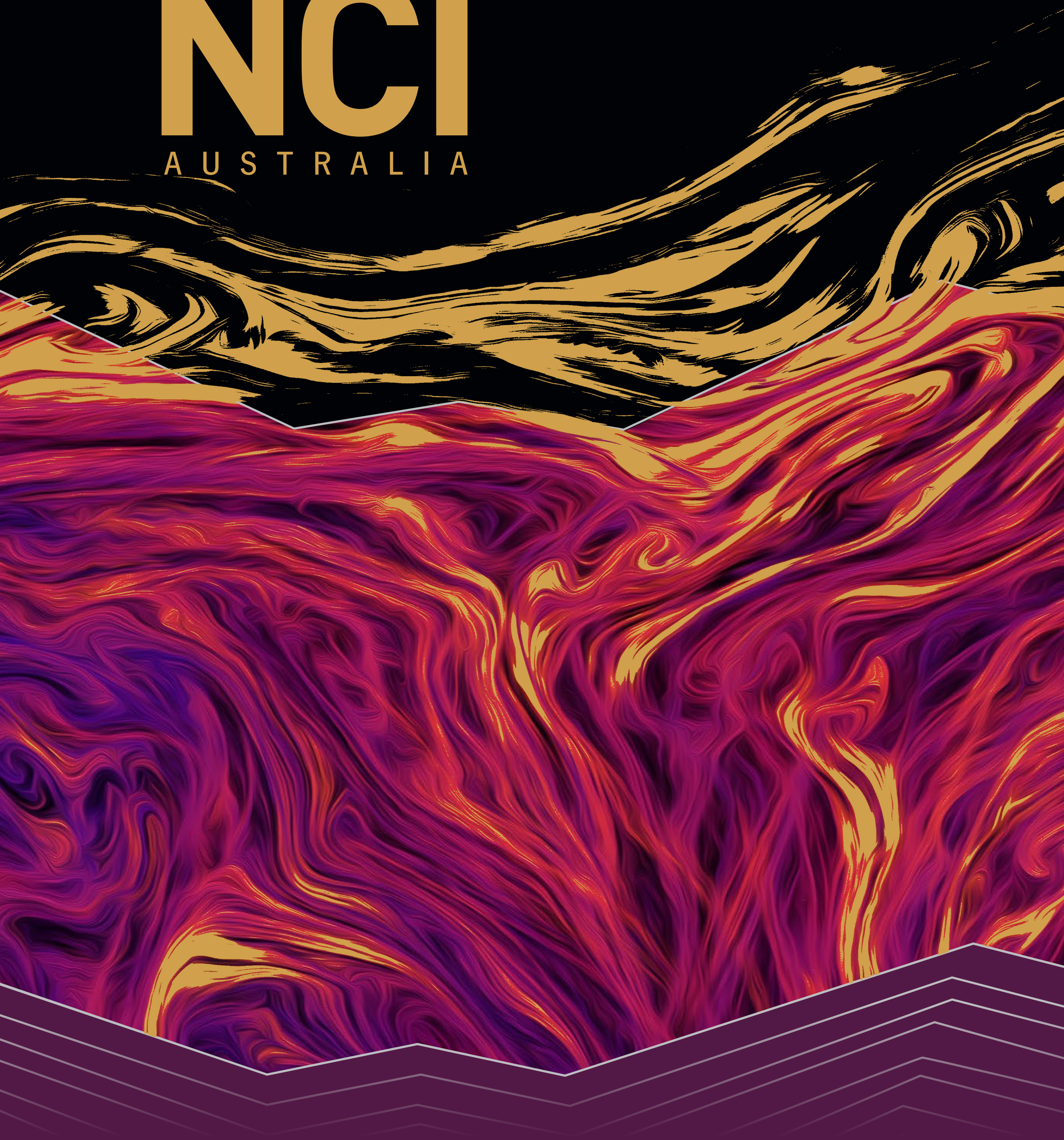
NCI
AUSTRALIA



ANNUAL REPORT
2020-2021

HOME OF GADI – POWERING LEADERSHIP-CLASS COMPUTING AND DATA SCIENCE FOR AUSTRALIA

This report and additional data are available from the NCI website at nci.org.au/about-us/annual-reports

Requests for authorisation and enquiries concerning the contents of the report should be directed to enquiries@nci.org.au

Produced by
National Computational Infrastructure
143 Ward Road, The Australian National University, Acton, ACT 2601

Designed by Result Design.

Cover Image:
A zoom-in of supersonic, turbulent particle motions that trace the gas dynamics around a strong, coherent magnetic field, coloured by the density and velocity of the particles. Understanding the motions of the particles helps us understand how energy is transported through flows of gas and dust in molecular clouds, the nurseries of stars in the interstellar medium. Image by James Beattie, The Australian National University.

We acknowledge the Traditional Custodians of the ACT, the Ngunnawal people. We acknowledge and respect their continuing culture and the contribution they make to the life of this city and this region.

CONTENTS

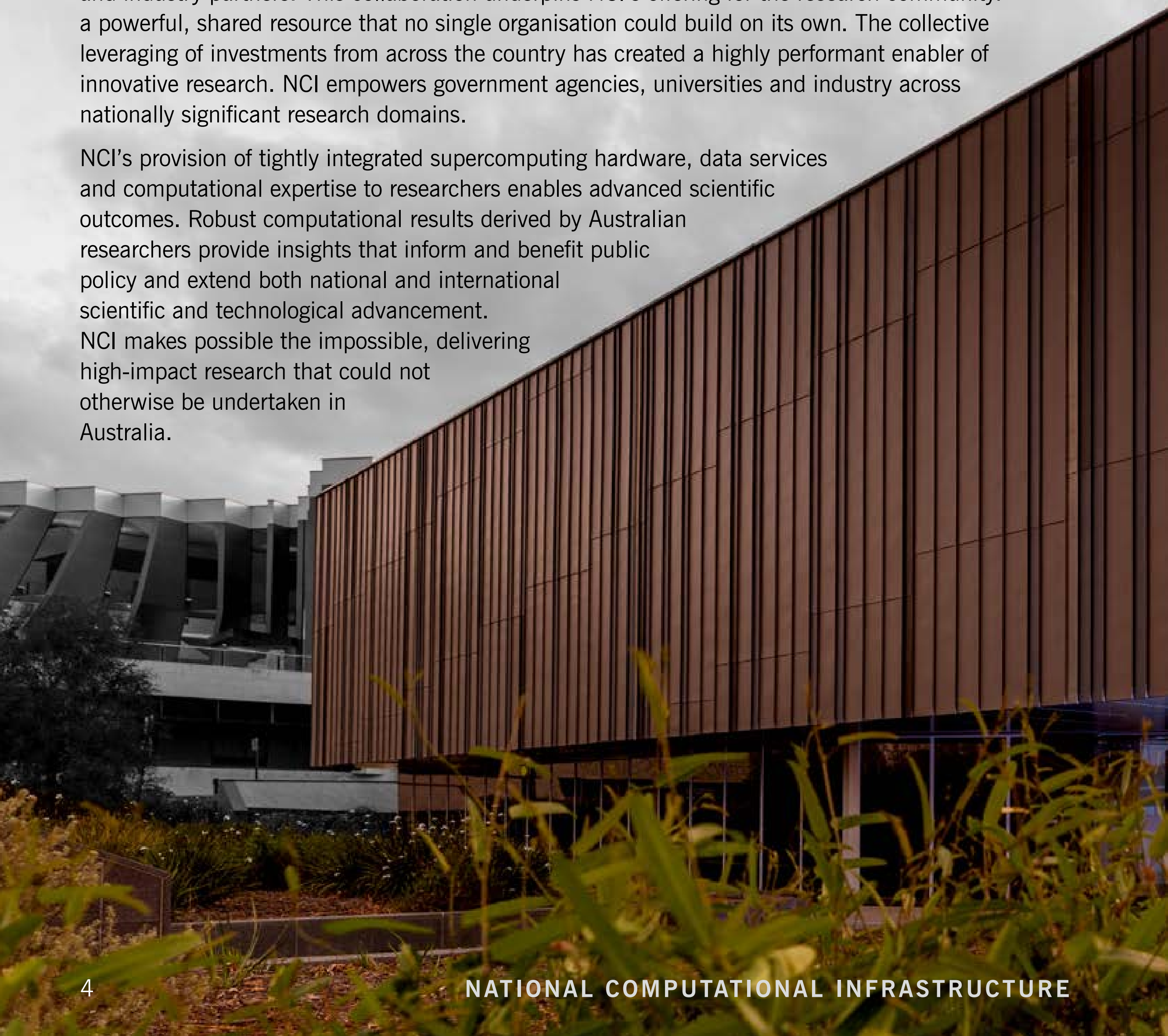
INTRODUCTION

About NCI

Australia's National Computational Infrastructure, NCI, is a unique supercomputing and big data facility supporting a vast network of scientists, research infrastructure providers and national science agencies. Comprising one of the country's most powerful supercomputers alongside vast data repositories, NCI provides the data management, infrastructure and services to deliver on the critical national need for high-performance data, storage and computing. NCI's service offerings are critical to the nation, supporting the high-impact work of research, policy development, government, industry and innovation.

The Australian Government and the Australian research sector come together through NCI in a broad collaboration involving the largest national science agencies (CSIRO, the Bureau of Meteorology, Geoscience Australia), the National Collaborative Research Infrastructure Strategy (NCRIS) network of research infrastructure capabilities, research universities, the Australian Research Council and industry partners. This collaboration underpins NCI's offering for the research community: a powerful, shared resource that no single organisation could build on its own. The collective leveraging of investments from across the country has created a highly performant enabler of innovative research. NCI empowers government agencies, universities and industry across nationally significant research domains.

NCI's provision of tightly integrated supercomputing hardware, data services and computational expertise to researchers enables advanced scientific outcomes. Robust computational results derived by Australian researchers provide insights that inform and benefit public policy and extend both national and international scientific and technological advancement. NCI makes possible the impossible, delivering high-impact research that could not otherwise be undertaken in Australia.



Our Mission

NCI's mission is to radically enhance the high-performance computational methods and capabilities available to Australian researchers. By doing this, we add value to Australian society, fulfil the needs of our collaborators and enable transformative ideas. Thus, the research sector provides a direct benefit to Australia's industry, environment, policy and society.

We expand the range, scope and ambition of Australian research. Our world-class computing and data systems, services and practices enable thousands of computational and data scientists each year to produce breathtaking science that benefits and impacts the country. As data-intensive methods become ever more central to Australian science, NCI's reliable and innovative high-performance platform is essential. NCI is pushing the boundaries of what such a facility can offer: tightly integrated high-performance computing (HPC), high-throughput computing (HTC) and high-performance data (HPD) open up previously unimagined possibilities for Australian research.

NCI enables Australia to have a globally competitive HPC, HTC and HPD capability. Through us, Australian researchers continue to experience a significant shift in what they can do computationally. We are expanding the number of researchers using high-end computing in Australia, and are enabling Australian researchers to produce more high-impact discoveries and innovations.

Chair's Report

It is a great pleasure to introduce to you the National Computational Infrastructure's 2020–2021 Annual Report, my first as Chair of the Advisory Board. Since joining the Board in December 2020, I have been fascinated to learn more about the great work that the NCI team and the whole user community are doing with high-performance computing (HPC) and data.

My years working in and leading research in CSIRO and the Bureau of Meteorology gave me a front row seat to the evolution of supercomputing over the past 3 decades. In 1990, CSIRO consolidated all the Organisation's mainframe computing to a CRAY Y-MP2/216 located at Port Melbourne, and established a 2 megabit per second link to my Division, Atmospheric Research at Aspendale, that gave us access to AARNet and the Internet. The CRAY had 2 cores and a peak performance of 0.0006 teraflops. None of us then could have conceived of a Gadi with 180,000 cores and a peak performance of 9,264 teraflops only thirty years later. You don't need a supercomputer to work out the magnitude of that increase in computational power, or the potential for more research that it provides.

As a result, computing is now ubiquitous in our economic, social and scientific systems, with supercomputers and the accompanying big data a staple for nationally significant science advances.

With the pandemic still affecting us all in 2020–2021, NCI has continued to provide crucial support to the Australian research community, cementing its place as vital infrastructure for the nation. Over the past 12 months, Australian scientists have conducted significant research in areas including COVID-19 molecular modelling, short- and long-term climate modelling, gas turbine combustion simulations and much more.

NCI's range of computational and data services has been particularly effective in providing a platform for COVID-19 related research and community resilience efforts. On the HPC front, biochemists have been simulating virus binding and potential treatment pathways in detail. NCI has also provided the support for rapid response and reliable virtual computing environments underpinning data sharing dashboards and healthcare modelling.

Supported by the Department of Education, Skills and Employment (DESE)'s National Collaborative Research Infrastructure Strategy (NCRIS), the Gadi supercomputer and the flagship Australasian Leadership Computing Grants scheme are underpinning some of the most exciting computational research in Australia. Detailed research into the fundamentals of future battery technologies, pandemic prevention and next-generation ocean circulation models all help lead us towards a cleaner, safer, healthier future.

A big project touching NCI and many in the Australia science community this year is the National Research Infrastructure Roadmap currently being developed by DESE. This Roadmap, produced every five years, outlines the infrastructure needs of Australian research, the potential benefits of world-class infrastructure and sets out recommendations for the Federal Government. The 2016 Roadmap, which identified HPC as a vital underpinning capability for advanced analysis and simulation, led directly to the funding for the current Gadi supercomputer.

NCI's world-class HPC and HPD services continue also to support research commercialisation and industrial modelling from across the private sector. Six companies are using NCI's systems to support their business development including commercial weather modellers, environmental consultants and award-winning long-time NCI users and water modellers DHI.

After taking on the Chair role, I was fortunate earlier in the year – before interstate travel became impossible – to spend two periods of a few days each at NCI familiarising myself not only with Gadi but also with a wide range of NCI staff. Their technical competence, professionalism, and enthusiasm for providing researchers with high-performance computational services at the leading edge was infectious. It is the combination of hard infrastructure in Gadi with the soft infrastructure of the NCI staff that, to use the language of management consultants, makes NCI as a facility "best of breed".

I look forward over the coming year to seeing much more great research and powerful new computational technologies come out of NCI.

Dr Greg Ayers NCI Advisory Board Chair

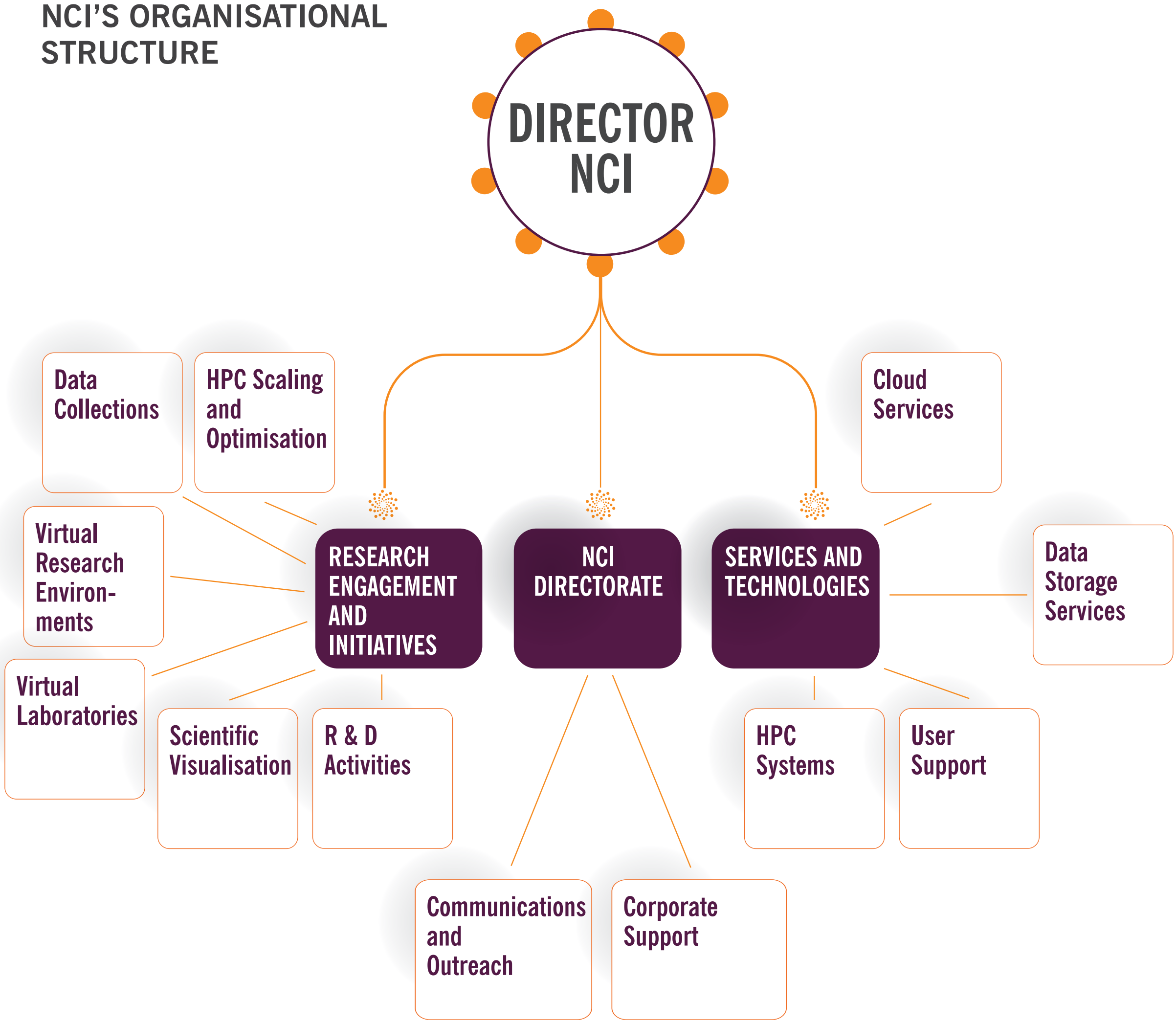
NCI Governance

THE NCI ADVISORY BOARD

NCI is governed by The Australian National University on the advice of the NCI Advisory Board, which comprises:

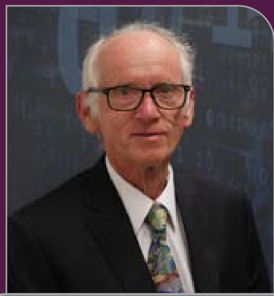
> an independent chair appointed by the Advisory Board
> the Director of NCI
> one member appointed by each of the Major Collaborators
> additional independent board members appointed by the NCI Advisory Board on the basis of their expertise.

NCI'S ORGANISATIONAL STRUCTURE



ADVISORY BOARD MEMBERS

Dr Greg Ayers
Chair



Emeritus Professor Robin Stanton
Deputy Chair

Professor Keith Nugent
FAA
The Australian National University, Deputy Vice-Chancellor (Research and Innovation)

Professor Sean Smith
FAAAS
NCI, Director

Dr Gilbert Brunet
FCMOS
Bureau of Meteorology, Chief Scientist and Group Executive (Science and Innovation)

Dr James Johnson
Geoscience Australia, Chief Executive Officer.

Dr David Williams
CSIRO, Executive Director, National Facilities and Collections

Dr Simone Richter
ANSTO, Group Executive (Nuclear Science & Technology and Landmark Infrastructure)

Accurate as at 30 June 2021.

Director’s Report

Colleagues and friends,

To start this report, let me offer a warm welcome to NCI Advisory Board Chair Dr Greg Ayers, who took up the role at the beginning of 2021. I’d like to offer my sincere thanks to Emeritus Professor Michael Barber for his contributions to the governance of NCI during his tenure as Chair over the past five years.

2020–2021 has been a year like no other in most of our memories, with the enormous challenges posed by the global pandemic impacting us all. Indeed, this report is penned while the ACT is – like large parts of the country – in lock-down. I’d like to pay special tribute to the deep commitment of the NCI staff in maintaining excellence in the delivery of high-performance computing and data services for the country throughout the past year.

Last year’s Annual Report marked the introduction of the Gadi supercomputer – debuting at #24 in the world – highlighted by the fabulous artwork of local artist Lynnice Church. This year, the NCI Annual Report recognises the Gadi-enabled world-class science of some of NCI’s leading users. It is appropriate, therefore, to feature on the cover a visualisation from one of Gadi’s largest single computing jobs, a 99,000-core astrophysics simulation.

Gadi, with its heterogeneous architecture designed to carry a broad range of workflows, has been a standout performer. It has delivered scientific simulations of unprecedented magnitude, facilitated through the meritorious Australasian Leadership Computing Grants on topics as diverse as COVID virus studies, weather simulations, astronomical simulations, combustion research and drug design. At the same time, it has delivered leading high-throughput computing capacity in topic areas such as bioinformatics and materials science.

A recurring theme in many of our strategic activities over this past year has been facilitating access to the unique capabilities and resources at NCI for a broad range of Australia’s research communities.

> NCI has built and launched a cloud platform that acts as a convenient and user-friendly interface into our systems, drawing together in a seamless fashion numerous tools that previously were known – and in part implemented – but not integrated.

> NCI has collaborated with NCRIS partners Australia Astronomy Limited and BioPlatforms Australia in two separate co-investment projects to facilitate flexible community access to our platforms from targeted astronomy and life-sciences communities.

> NCI is collaborating with the Australian BioCommons to build tools and services to support a robust and performant platform for the computational research of the bioinformatics community.

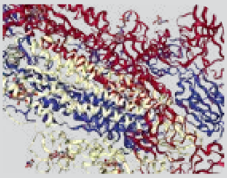
> Following extensive consultation, NCI is mounting a major thrust in user skills training to assist our user community in their diverse pathways up the high-performance computing and data skills ladder – from novices to mature and expert users.

Through its decade-long collaborations with CSIRO, the Bureau of Meteorology and Geoscience Australia, NCI is positioned at the vanguard of trusted hosting and provisioning of ultra-large national data collections in climate, weather and geospatial imaging. This is also increasingly the case with genomics datasets through our collaboration with the Garvan Institute. Tight integration of supercomputing, data analytics and petascale data collections will increasingly be a powerful national vehicle for the advancement of research and industry. This and related topics have been prominent in the consultations and dialogues that have accompanied the current National Research Infrastructure Roadmap development process, in which NCI has been an enthusiastic and active participant. We look forward to the outcomes of the Roadmap and assisting to advance this important national agenda.

Professor Sean Smith, FAAS Director, NCI

Highlights of the Year

NCI and the Pawsey Supercomputing Centre join an international White House-led COVID-19 research consortium.



Comprehensive collection of ERA5 climate model data from Europe available at NCI.

NCI announces grants to Australia's leading computational researchers totalling 150 million units of computing time.

J A S O N D J F M A M J

2020 2021

NCI joins leaders of Australasian HPC and eResearch to launch new Women in HPC chapter.

Gadi completes 724 million hours of computing time in its first full year of operations.

Gadi remains Australia's fastest supercomputer, ranked 44th in the world.

1

A CENTRAL ENABLER IN A NATIONAL RESEARCH ECOSYSTEM

Research Overview – Supporting COVID-19 Projects

ENABLING THE SCIENCE WE NEED WHEN WE NEED IT THE MOST

The work of computational researchers can sometimes take place years or decades before the products or technologies they help develop make their way into people's lives. New manufacturing or environmental management technologies developed on the NCI supercomputer can take years to develop and implement at large scales.

In addition to supporting long-term fundamental research, NCI also provides the infrastructure that researchers need to respond to urgent national and global priorities as they arise. Streamlined access and the increased computational capacity of Gadi allow scientists to respond to issues impacting all Australians. Never has this been more apparent than with the plethora of COVID-19 research projects that NCI has been able to support this year.

Understanding a novel virus requires many kinds of expertise: clinical, medical, chemical, biological. Thankfully, much of that capability already existed in the Australian computational science community before the pandemic took hold in 2020. Australia's researchers used their decades of experience, and enabling infrastructure platforms like NCI, to help Australia and the world progress our understanding of the virus

For NCI's users, high-performance computing and easily accessible resources made it possible for them to quickly focus their research efforts on this new direction. High-resolution simulations of the molecular interactions on the surface of human cells require thousands of processors and millions of hours of computing time. Dashboards with up-to-date case numbers, maps of transmission locations and localised alerts require reliable and robust cloud environments to run around the clock.

Long-time NCI user Professor Irene Yarovsky, Professor Mibel Aguilar and their collaborators from RMIT University used their existing knowledge of the human ACE2 protein that the COVID virus binds with to model how they could slightly interfere with the protein's physical structure. By modifying the structure a small amount, they could potentially reduce the ability of the virus to bind to human cells.

Through a dedicated merit allocation process in 2020, NCI awarded four research groups significant resources enabling them to advance our knowledge of COVID-19. This included learning about how the virus works, whether existing drugs could be used to treat the disease, and how vaccines could be designed to work more effectively. This scheme, offering rapid-response resources at both NCI and Pawsey, linked Australia to the US-led COVID-19 High-Performance Computing Consortium. The Consortium brought together supercomputer facilities, research institutes and industry leaders from around the world, including NCI and Pawsey in Australia, the Partnership for Advanced Computing in Europe and NASA in the USA, a truly international effort in a time of global crisis.

Associate Professor Megan O'Mara from The Australian National University pivoted her molecular biology simulations towards understanding the virus. She says, "This research uses simulations of the around 800,000 atoms that make up a key receptor of the human body to understand exactly how the coronavirus uses it to invade human cells. It is only with high-resolution modelling that accurately replicates the true behaviours of these receptors that we can figure out where vulnerabilities in the virus' binding process are."

NCI's resourcing in response to the COVID-19 pandemic extends beyond simply providing computing time. We also rapidly provisioned a cloud dashboard capability that has run for more than a year to provide graphical updates on the virus spread and transmission hotspots. CRISPER – the COVID-19 Real-time Information System for Preparedness and Epidemic Response – is an initiative of The Australian National University Research School of Population Health, the University of Queensland, the National Health and Medical Research Council's Australian Partnership for Preparedness Research on Infectious Disease Emergencies and state health departments. CRISPER continues to share invaluable real-time case numbers, testing numbers and exposure site data that benefit public safety and lead to a more effective clinical response.

Early in the pandemic, ACT Health worked with ANU epidemiologists, public health researchers, visualisation experts and data specialists to predict COVID-19's impact on the territory's medical system. Project engineer Dr Kathy Reid says, "This necessitated rapid stand-up of computing infrastructure. We were delighted not only with the responsiveness, personal assistance and customer service ethic exhibited by NCI, but also the ability of the technical infrastructure to scale rapidly to demand. The team was able to begin modelling on NCI infrastructure within hours of an initial request – due both to the technically skilled NCI team, and the maturity of NCI infrastructure automation."

The response to the COVID-19 pandemic is drawing on the expertise and capabilities of Australian researchers and public health professionals. As a key element of national research infrastructure, NCI continues to support leading research efforts. Whether the scientists are notifying clinical colleagues, simulating molecular interactions or modelling medical system capacity, NCI's robust and rapid-response computing platform is there to support nationally significant research.

"The team was able to begin modelling on NCI infrastructure within hours of an initial request, due both to the technically skilled NCI team, and the maturity of NCI infrastructure automation."

– Dr Kathy Reid, ANU

National Benefits

Computational and data-intensive science in Australia underpins some of the most important areas of national progress for the coming years and decades. Advances in weather forecast accuracy, bushfire model resolution, industrial material design and genetic illness diagnosis benefit all Australians. These steps forward for health, environment, transport, tourism, industry and more require the high-performance infrastructure that NCI provides.

In 2020–21, the Australasian Leadership Computing Grants, designed to use the full scale and capability of the Gadi supercomputer, have provided a showcase for Australian computational research. They have highlighted the leadership ambition of the research community and the need for this capability among our foremost researchers. The incredibly competitive process led to four successful projects sharing in a total allocation of more than 180 million units of computing time on the nation-leading, world-class supercomputer Gadi.

Led by some of the most experienced computational scientists in Australia, these projects were some of the largest and most ambitious ever attempted in the country, allowing our researchers to compete at an international scale. Researching topics of national significance, they looked into critical questions of Australian climate change impacts in coming decades, of the formation of black holes and galaxies, and of efficient designs for gas turbines to drive future energy generation.

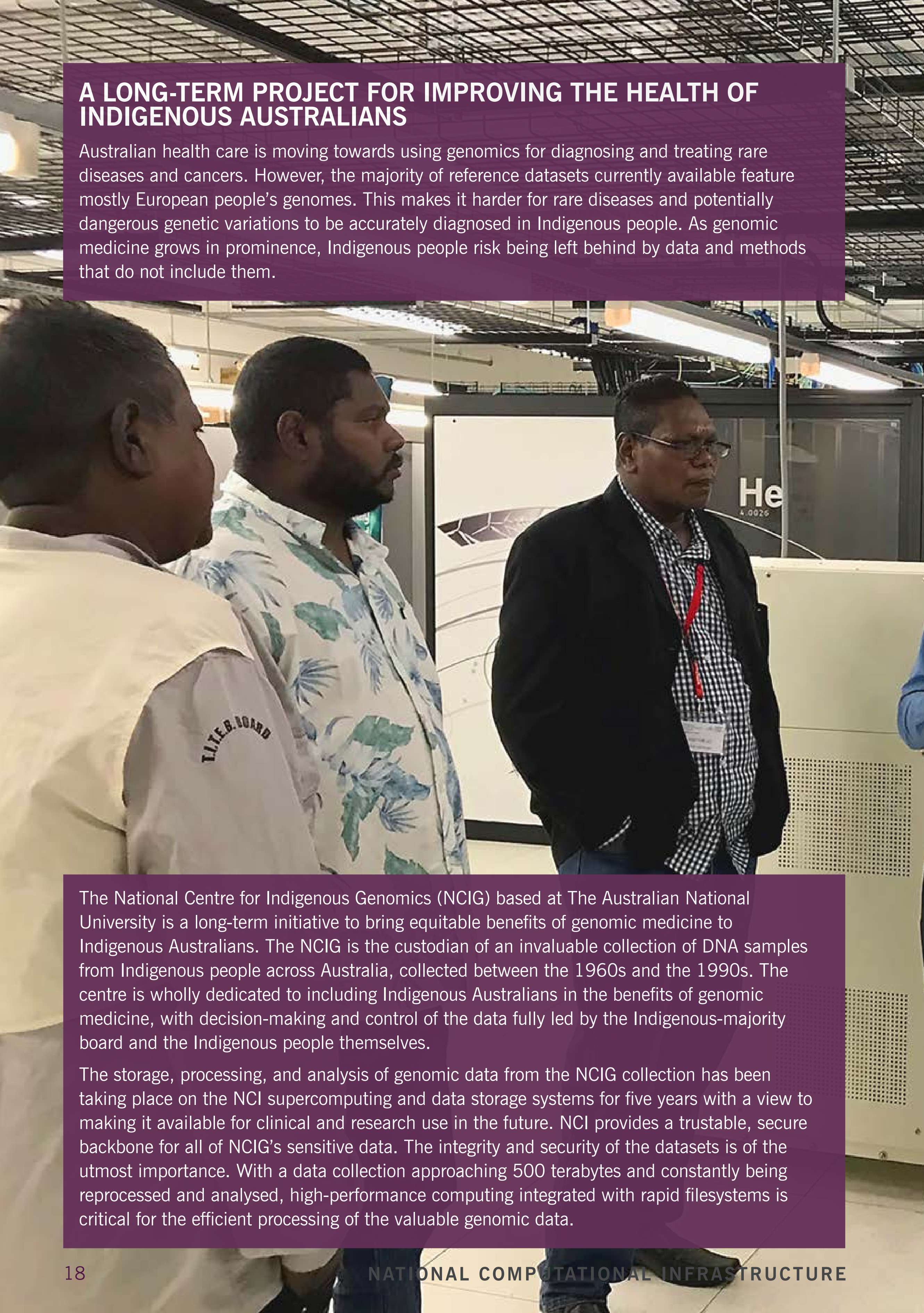
This year, NCI also augmented its ever-growing catalogue of nationally significant datasets with several collections of international climate observations, model outputs and reanalysis data. The CMIP6 (Coupled Model Intercomparison Project Phase 6), ERA5 (European Reanalysis 5) and Australian Gridded Climate Data (from the Bureau of Meteorology) collections provide new inputs for scientists studying regional weather, long-term climate, and environmental change. Key components of innovative weather and climate prediction systems, these large and expertly curated datasets are centrally located at NCI for thousands of researchers around the country to efficiently access and analyse.

As the COVID-19 pandemic continues to impact communities around Australia and the world, Australian researchers are using NCI to share outbreak data, understand the virus and search for drugs to treat COVID-related illnesses. Researchers across the fields of genomics, molecular modelling, population health and biochemistry are all pursuing their own lines of inquiry to understand both the current pandemic and the potential for further pandemics in the future (See Research Overview *Enabling the science we need when we need it the most* on page 14 and the Case Study *Predicting the next coronavirus outbreak* on page 26 for more information).

Tightly integrated supercomputing and big data systems, powerful analysis environments and expert support: NCI provides the key components that drive discovery and high-impact Australian science efficiently and effectively.

A LONG-TERM PROJECT FOR IMPROVING THE HEALTH OF INDIGENOUS AUSTRALIANS

Australian health care is moving towards using genomics for diagnosing and treating rare diseases and cancers. However, the majority of reference datasets currently available feature mostly European people's genomes. This makes it harder for rare diseases and potentially dangerous genetic variations to be accurately diagnosed in Indigenous people. As genomic medicine grows in prominence, Indigenous people risk being left behind by data and methods that do not include them.



The National Centre for Indigenous Genomics (NCIG) based at The Australian National University is a long-term initiative to bring equitable benefits of genomic medicine to Indigenous Australians. The NCIG is the custodian of an invaluable collection of DNA samples from Indigenous people across Australia, collected between the 1960s and the 1990s. The centre is wholly dedicated to including Indigenous Australians in the benefits of genomic medicine, with decision-making and control of the data fully led by the Indigenous-majority board and the Indigenous people themselves.

The storage, processing, and analysis of genomic data from the NCIG collection has been taking place on the NCI supercomputing and data storage systems for five years with a view to making it available for clinical and research use in the future. NCI provides a trustable, secure backbone for all of NCIG's sensitive data. The integrity and security of the datasets is of the utmost importance. With a data collection approaching 500 terabytes and constantly being reprocessed and analysed, high-performance computing integrated with rapid filesystems is critical for the efficient processing of the valuable genomic data.

NCIG Bioinformatics Lead Dr Hardip Patel says, “NCIG is based around the principle that Indigenous people are in charge of their data. We build a relationship with those communities and individuals whose DNA was collected many decades ago, and ask them how they would like us to proceed with the samples we do have. Some samples have been recently repatriated to Galiwin’ku after many years away from their country with the permission to use the data for research.”

“We are using NCI at every step of the process from the initial quality assessment and alignments to complex genetic analysis, research, reprocessing, storage and curation. The data collection we are building will provide a reference baseline of genomic data for Australian Indigenous people. This resource will be used for carefully controlled research and clinical settings to benefit the whole of Australia.”

NCI regularly hosts representatives from Indigenous communities who are working with the National Centre for Indigenous Genomics to develop appropriate protocols for the storage and use of Indigenous genomic data.

Since 2017, the NCIG has used 17 million units of computing time at NCI. In collaboration with research groups around the country, they are using Gadi’s Graphics Processing Units to run Machine Learning algorithms on the DNA sequences. They are also developing an Indigenous Australian variant of the reference human genome that more fully captures specific genetic details that are missing in the current generic reference genome to improve accuracy of genomic analyses and interpretations.

The aims of the NCIG are long-term and carefully considered. Over time, its work will lead to improved health outcomes for Indigenous people. NCI is proud to support their nationally significant mission through reliable and secure data processing and analysis infrastructure.

“This resource will be used for carefully controlled research and clinical settings to benefit the whole of Australia.” – Dr Hardip Patel, ANU

Impactful Science

Some of the most rapidly evolving elements of Australian society rely on knowledge generated through computational and data-intensive science at NCI: battery and hydrogen technologies to support a decarbonised electricity grid and future industries, satellite imagery to support urban planners and farmers, and human genomics to support fast, effective and efficient disease diagnosis and treatment.

Australia's leading supercomputer and research data repositories at NCI are central to scientific research into weather forecasting, climate change, new materials, health, resources and future technologies. With the research tools that NCI makes available, scientists are producing results that create an impact on our everyday lives. Over the past two years, nothing has had a bigger impact on our lives than the coronavirus pandemic.

Throughout 2020 and 2021, computational chemists, biologists and technologists have been using NCI to investigate every aspect of the COVID-19 virus, pandemic and outbreak distribution. Through the provision of dedicated resources, as well as ongoing

support for impactful projects from around the country, NCI has enabled critically important national research. Scientists are applying their decades of expertise and all the computational and data analysis tools at their disposal to assist with vaccine development, medical treatment discovery and future pandemic preparations.

Associate Professor Megan O'Mara from The Australian National University has been using Gadi to simulate the spike protein of the COVID-19 virus and how it binds to human cells (See Research Overview *Enabling the science we need when we need it the most* on page 14 for more detail). Understanding the exact behaviours of large molecules such as coronaviruses requires accurate models that can duplicate the forces and interactions affecting each atom. Computational methods are the only way to properly grasp these intricate effects.

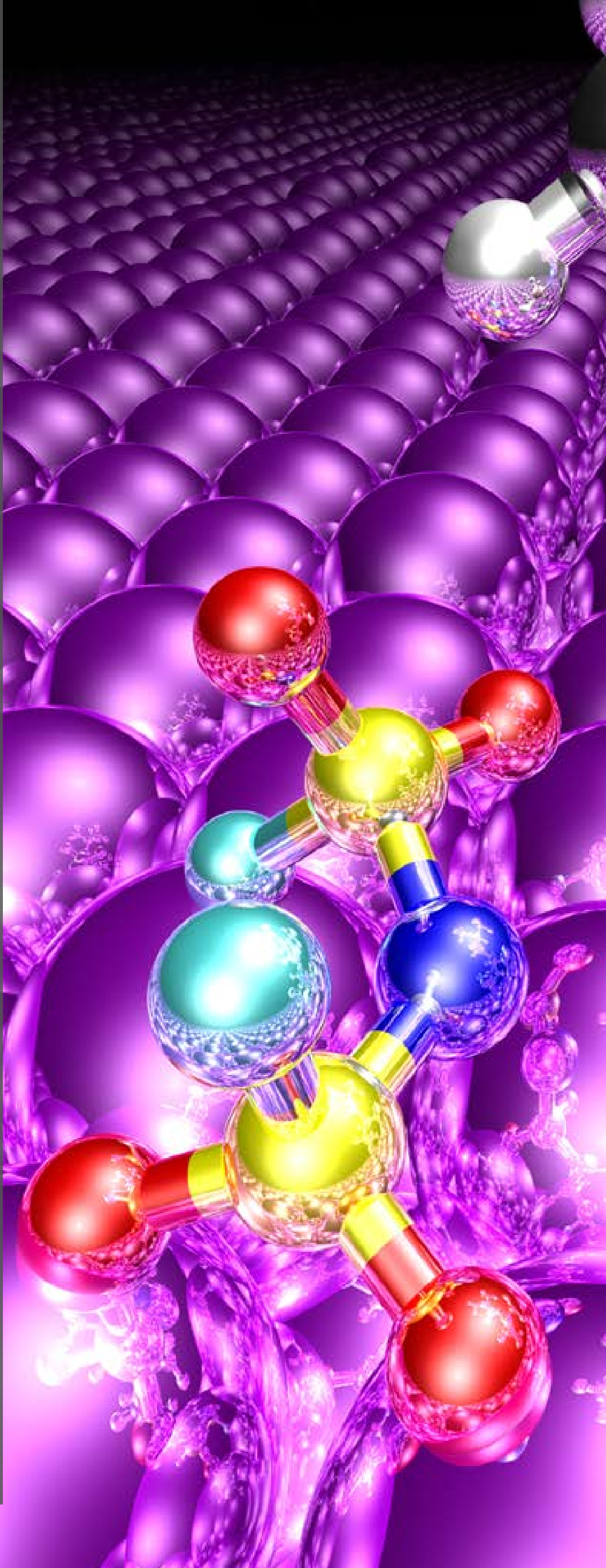
The science taking place at NCI – in the data collections, virtual analysis environments and supercomputer – benefits Australians now and into the future. The impact of the discoveries computational scientists are making will grow over the coming years. Some of the central technologies of the future, whether they be in pain treatments, energy storage, land management or advanced manufacturing, are making their start in the nation's leading supercomputing and big data facility.

BATTERY TECHNOLOGIES OF THE PRESENT AND FUTURE

Every item and process in our lives is made up of technologies, discoveries and inventions that have been combined into a useful form. Every advance in creating a new material or method helps us gradually move society forward. In our electrified and energy-efficient future world, we will have high-capacity batteries that are lighter and safer than current ones, and energy-efficient computers made from silicon variants that have much lower energy consumption for each calculation. Professor Michelle Spencer from RMIT University and the Australian Research Council Centre of Excellence in Future Low-Energy Electronics Technologies (FLEET) is conducting research in these critical areas.

Professor Spencer uses the Gadi supercomputer to understand what is going on inside battery materials at the molecular level. She says, “We can’t see the molecules react under a microscope, but we can predict what will happen with very powerful computers and then test only the most promising reactions in the laboratory.” Predicting molecular interactions is incredibly hard to do. The number of possible movements and interactions for every proton and electron in the system increases exponentially as molecules get bigger and more complex. Modelling even a simple system accurately for a few seconds is already hugely computationally intensive.

The Gadi supercomputer allows the research team to run hundreds of simulations looking at complex, previously unresearched combinations of molecules. Using specialised scientific modelling software running on thousands of processors at the same time, Professor Spencer can avoid wasting experimental time looking into unsuitable candidate materials. Furthermore, simulations allow the research team to learn about why certain behaviours occur in the materials of interest. This provides much greater insight and lets researchers everywhere build on and share their understanding of the fundamentals of the molecular interactions.



Visualisation of molecules from Professor Michelle Spencer’s research.

Professor Spencer says, “Access to high-performance computers is essential if we are going to design and test new materials on a reasonable timescale. If we are to see exciting new technology like electric passenger planes in our lifetime, then computational chemists working with supercomputers are as essential as the engineers and material scientists at the bench.”

Efficiently searching for new materials that we can use to support a cleaner world requires the processing power and capabilities of supercomputers. NCI and Gadi give Australian researchers the performance they need to understand the fundamental characteristics of some of the materials that will soon power our planes, ships and computers. Molecular interactions control everything about the world around us and how it works. Researchers like Professor Spencer spend their days looking into that fascinating world to develop the advanced materials of our future.

> *“Access to high-performance computers is essential if we are going to design and test new materials on a reasonable timescale.”*
>
> – Professor Michelle Spencer, RMIT

Merit-based Access to NCI

NCI provides computing resources to the most impressive Australian researchers. Across the merit-based portion of NCI's allocations in 2020–21, NCI allocated more than 586 million units of computing time to researchers from across the scientific spectrum. Peer-reviewed and allocated through a stringent process, the successful proposals are richly deserving for both their scientific excellence and national significance. The three major merit allocation streams are supported each year by investments from the Commonwealth Department of Education, Skills and Employment's National Collaborative Research Infrastructure Strategy (NCRIS), the Australian Research Council (ARC) and the NCI Collaboration.

The three allocation streams are the longstanding National Computational Merit Allocation Scheme, the Australasian Leadership Computing Grants (ALCG, now in its second year), and the continuing Flagship Scheme supporting the work of 5 ARC Centres of Excellence. Together, these support some of the most experienced and important Australian research teams looking into battery technology, climate variability, high-efficiency turbines and much more.

Professor Ben Corry from The Australian National University, recipient of more than 26 million units of computing time through the ALCG, is using high-resolution molecular modelling to investigate the potential for currently circulating coronaviruses to cause future pandemics like the COVID-19 pandemic (See Case Study *Predicting the next coronavirus outbreak* on page 26 for more details). The ALCG is a scheme catering to the biggest and most impactful computational research projects, allocating 150 million units of computing time in 2021. Enabled by the Gadi system, NCI is able to offer unprecedented computational resources to highly

meritorious, incredibly valuable science: fundamental semiconductor research for improved device manufacturing, fluid flow simulations for high-resolution bushfire physics research, energy storage and energy efficient materials development for low-emissions housing, and more.

The National Computational Merit Allocation Scheme, together with partner centres the Pawsey Supercomputing Research Centre, the Monash University eResearch Centre and the University of Queensland Research Computing Centre, allocated 286 million units of computing time in 2021 to 151 projects on Gadi. These projects cover all domains of science, from turbine simulations to particle physics, human genomics, astronomy, molecular modelling and weather model development.

The Flagship Scheme, providing no-cost computational resources to the ARC Centres of Excellence, underpins their nationally significant missions. The current Flagship Scheme recipients specialise in astrophysics, climate extremes, materials science and fundamental physics. Further support goes to research infrastructures in NCRIS and from across the scientific community. These highly meritorious centres and research organisations share in 150 million units of computing time each year with the support of both NCRIS and the ARC.

PREDICTING THE NEXT CORONAVIRUS OUTBREAK

Professor Ben Corry from The Australian National University received one of five prestigious Australasian Leadership Computing Grants (ALCG) for the 2021–2022 period to support his research into currently circulating viruses with the potential to generate future outbreaks like COVID-19. This highly meritorious research received an allocation of 26.5 million units of computing time through the ALCG scheme, enabled by funding from the Department of Education, Skills and Employment's National Collaborative Research Infrastructure Strategy (NCRIS).

The only way of preventing future global pandemics is to understand where they might come from and prepare ourselves for when and how they might break out into the human population. Professor Corry says, "Access to the enormous resources of Gadi enables us to take the methods we have honed in studying the COVID-19 coronavirus and apply them to the huge range of coronaviruses circulating in animals to determine which could seed the next pandemic."

Accurate understanding of the specific virus structures that bind to human cells, the spike proteins, requires high-resolution simulations. Predicting the flexibility of a coronavirus spike protein and how tightly it binds to its receptor requires simulating the motion of all the protein atoms in their environment for an extended period of time. Getting to the required modelling precision means simulating each atom and all the forces acting on each one, and then running the simulation forward through billions of minuscule time-steps.

Such a computationally intense activity is only possible using a supercomputer. Ordinary PCs would not be able to handle the number of calculations nor the data produced and processed throughout. Only powerful, modern supercomputers can provide the necessary computing power. At their peak, these simulations will use hundreds of state-of-the-art Graphics Processing Units and thousands of processors simultaneously, working through every atom and time-step of the calculation in parallel.

Maintaining a close eye on potentially dangerous viruses we are already aware of is a critical step in preparing for future pandemics. If needed, the knowledge gained through research such as this could be used for pre-development of vaccines or medicine. Learning about possible threats ahead of time means that we can be much more prepared with suitable medical and clinical responses if need be.

Computational simulations run on supercomputers such as NCI's Gadi are a key element in the race to understand the functioning of these dangerous viruses. NCI provides fast and efficient access to the supercomputing resources Australian researchers need and supports the important work of the Australian science community.

"Access to the enormous resources of Gadi enables us to determine which currently circulating coronavirus could seed the next pandemic." – Professor Ben Corry, ANU

2

A HUB FOR THE RESEARCH COMMUNITY

Supporting National Research Infrastructure

NCI is a research infrastructure platform, an investment for the community to underpin nationally significant science. Researchers from across the scientific disciplines access NCI for the analysis-ready datasets, data storage, computational performance, scientific visualisation and expert support we provide. By leveraging the investment of our Collaborators from across the spectrum of the national science agencies, research universities and medical institutes, along with the Australian Government, we support a larger and more capable set of computing and data services than any single institution could build individually.

The Australian Government, through the Department of Education, Skills and Employment's NCRIS program, funds many other research infrastructure facilities, collectively making up the NCRIS network. Through strategic collaborations, NCI compute and data services underpin many of the research tools and data portals that other NCRIS facilities provide their users. By supporting the rest of the national research infrastructure network, we help researchers get easy access to the data collections and robust analysis tools they need.

NCI's variety of services covers the requirements of different scientific disciplines, workflows, code bases and data sources. From the data processing of optical astronomy to the virtual analysis of environmental sensors, NCI helps other capabilities provide their expert scientific platforms in the ways that work best for their users. Offering a national platform, NCI makes it easier for the entire research community to share and access the resources they need.

Whether supporting biodiversity researchers, urban planners, geophysicists or ocean modellers, NCI is the hub for the research of some of Australia's major scientific communities. The Gadi supercomputer's finely tuned performance, the capacity and accessibility of its filesystems, and the availability of virtual analysis environments make NCI a key part of the research community's toolset for Australia's most intensive computational science.

The SkyMapper telescope used by astronomers around the country. Credit: James Gilbert, ANU

National Research Infrastructure Vignettes

TERRESTRIAL ECOSYSTEM RESEARCH NETWORK:

The Terrestrial Ecosystem Research Network (TERN)'s central goal, to measure and share observation and model-ready data of key terrestrial ecosystem attributes, requires robust data access and discovery capability. NCI helps to bring the data from hundreds of locations around Australia to the research community through computational and data storage infrastructure, including NCI's FAIR data repository, which underpins TERN's Data Discovery Portal. Combining the strengths of each NCRIS capability highlights the power of the NCRIS network and its role in supporting national science endeavours.

AUSCOPE:

Valuable geophysical data about the structure of the ground beneath our feet has many applications in the resources sector and in minimising risk from natural hazards and human activities. The Geophysics 2030 project, funded in part by the Australian Research Data Commons and AuScope, that NCI is undertaking with AuScope, TERN, Geoscience Australia and CSIRO will focus on processing rawer forms of data from around the country to common, high-performance standards and combining them with existing geophysical data assets at NCI. Integrating and securing this data in this way will enable more rapid processing and scalable, data-intensive computation.

PAWSEY:

NCI works closely with the Pawsey Supercomputing Research Centre to support the Australian computational research community. During a time of global crisis, NCI and Pawsey worked together to provide rapid-response computational resources for highly deserving COVID-19 research. As a result of this collaboration, Australia's two leading HPC centres joined the COVID-19 High-Performance Computing Consortium – an initiative led by the White House Office of Science and Technology, IBM, and the US Department of Energy – supporting the fight against the pandemic. As part of the COVID-19 HPC Consortium, NCI and Pawsey enabled international research into potential treatments and assisting vaccine development studies.

ASTRONOMY AUSTRALIA LIMITED:

NCI is supporting Astronomy Australia's Optical Data Centre project as part of a collaboration with Macquarie University and The Australian National University. The Optical Data Centre, a service for holding and serving optical and infrared astronomy data, also supports research teams to carry out their research through simplified data management, team management and data service provisioning. NCI provides robust data storage and management for this new and exciting astronomy research space.

BIOPLATFORMS AUSTRALIA:

Bioplatforms Australia (BPA) provides a wide range of biology and health-related research services. Through BPA's Australian BioCommons capability, NCI hosts and maintains a node of the Galaxy Australia web-platform. Galaxy Australia lets researchers conduct accessible, reproducible and transparent computational biology research with more than 1,000 tools, more than 220 reference datasets and extensive online training materials. NCI also provides a vast software repository of codes and processing pipelines available across the NCI computational platforms and accessible to BPA researchers.

ACCESS:

Announced in October 2020, the ACCESS-NRI is becoming a National Research Infrastructure dedicated to developing and supporting the Australian Community Climate and Earth System Simulator – Australia's national climate and weather modelling suite. ACCESS-NRI will provide a software engineering facility for the community into the future. NCI will be an important partner in ACCESS-NRI's operations and is already playing a key role in the development and formation of the centre. Over the coming years, NCI and the ACCESS-NRI will work closely to develop, optimise and run the next generation of our climate prediction and weather forecasting models.

AARNET AND PAWSEY:

NCI, Pawsey and AARNet, Australia's national research and education network, are constantly working together on strategy development and looking to the future to prepare for the further development of high-performance data and computing across Australia and the region. Our close collaboration enables innovative data analysis tools, access to international data repositories and preparation for future upgrades and service developments.

ATLAS OF LIVING AUSTRALIA:

NCI hosts and makes available high-quality, analysis-ready datasets for the research community, including users of the Atlas of Living Australia (ALA). Through the facility's data portal, ALA users can browse, analyse and download biodiversity data, including species occurrence records, eDNA records, and species information, as well as spatial and environmental layers. Under the surface, NCI provides storage and support for the data portals to run smoothly and efficiently.

Credit: Kim Calders, Terrestrial Ecosystem Research Network

AUSTRALIAN RESEARCH DATA COMMONS:

NCI and the Australian Research Data Commons (ARDC) work together to support Australian eResearch activities from across the scientific disciplines. Together, NCI and the ARDC partner to provide enhanced capabilities and future-focused technological developments to research communities, including for climate prediction, weather forecasting and geophysics. Enhanced capabilities include the creation of dedicated Data-enhanced Virtual Laboratories, and the rescue, management and sharing of valuable geophysical data collections.

TERN
Terrestrial Ecosystem Research Network

AuScope

Pawsey

Astronomy Australia Ltd.

BIOPLATFORMS AUSTRALIA

aarnet
Australia's Academic and Research Network

Atlas of Living Australia
ala.org.au

ARDC
Australian Research Data Commons

Supporting the NCI Collaboration

Central to NCI's operations is the NCI Collaboration, a grouping of Australia's most prominent national science agencies, research universities and medical research institutes, which provides the funding and leadership that underpin NCI's work and supports that of scientists around the country. The collaborative model allows our Collaborators to each contribute to and share in NCI's systems and services, thereby benefiting from a more capable computing and data ecosystem than any individual facility could have built on their own.

The NCI Collaboration consists of the four Foundation Collaborators, a growing group of Minor Collaborators and a selected set of Merit Flagships. The Foundation Collaborators, who have supported NCI since it took its current form in 2011, are CSIRO, the Bureau of Meteorology, Geoscience Australia and The Australian National University. The Minor Collaborators include the University of New South Wales, the Garvan Institute of Medical Research, RMIT University, The Queensland Cyber Infrastructure Foundation and The University of Melbourne. Supported also by the Department of Education, Skills and Employment's NCRIS program and the Australian Research Council, NCI provides no-cost allocations of computing time to select ARC Centres of Excellence with large computational requirements. All of NCI's Collaborators are listed on page 84.

The NCI Collaborators conduct some of Australia's most important science. From the Bureau of Meteorology's regular improvements to regional weather forecasts to CSIRO's critical climate modelling activities and Geoscience Australia's mapping of Australia's underground resources, the Collaborators produce high-impact, otherwise-impossible science at NCI. Making use of the Gadi supercomputer, high-performance data storage systems, virtual analysis environments and expert support, the NCI Collaboration has access to Australia's most integrated and capable computational science platform.

NCI Collaboration Vignettes

BUREAU OF METEOROLOGY:

The Bureau of Meteorology invests in NCI as part of its research and development process. Gadi serves as an ideal system on which to develop and test improvements to the national weather forecasting software suite, ACCESS. Recent achievements include significant efficiency gains enabling better resolution and accuracy further into the future, and major improvements to both capital city forecasts and tropical weather forecasts in particular. The Bureau is also a key contributor to the ACCESS National Research Infrastructure facility with which NCI is closely engaged.

Australian Government
Bureau of Meteorology

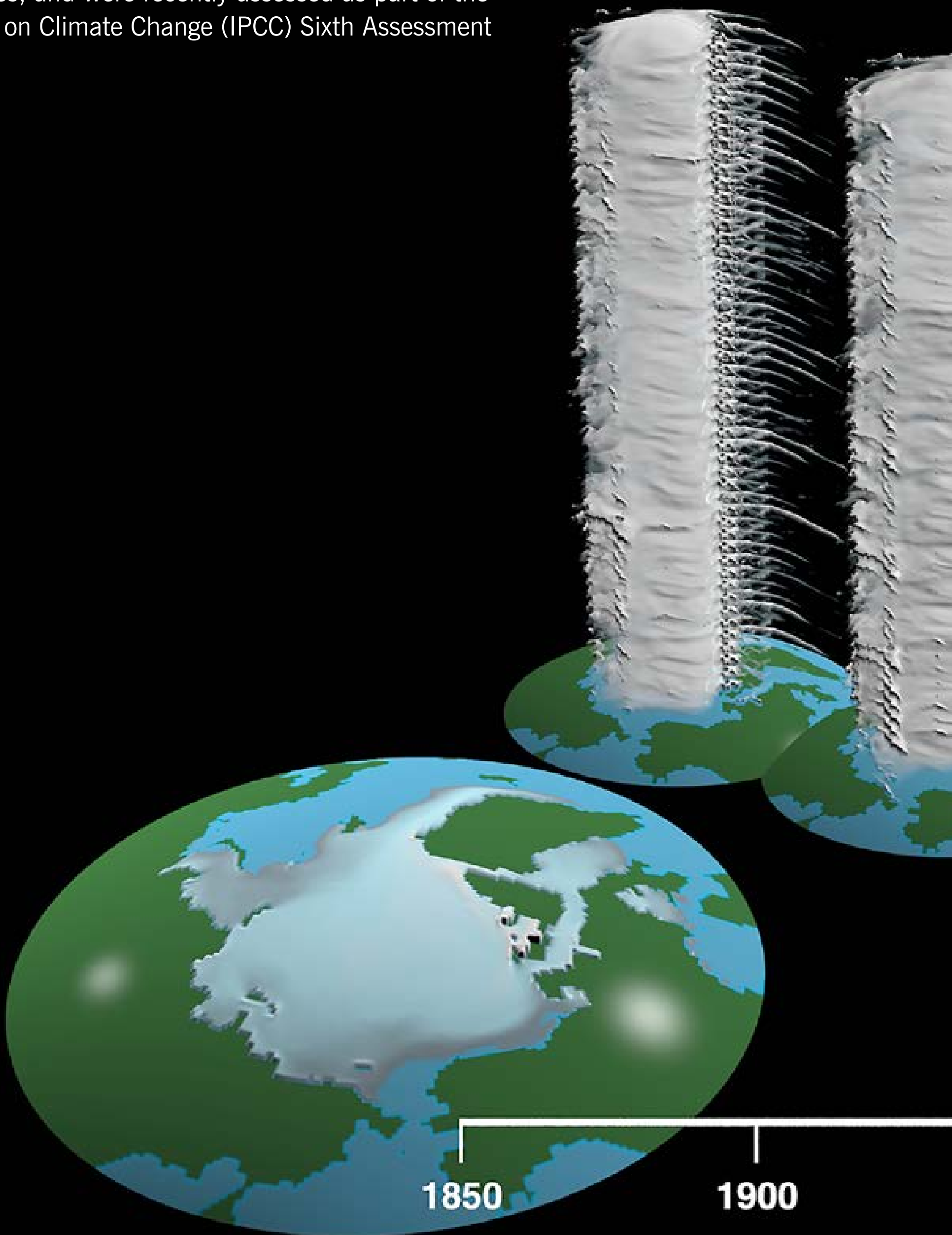
THE AUSTRALIAN NATIONAL UNIVERSITY:

Researchers from ANU, experienced in molecular modelling, turned their attention to COVID-19 over the past 18 months in an attempt to understand how we might create medicines to treat the disease, and how we might prevent future pandemics. Making use of Gadi's performance and large allocations of compute time, the researchers can perform high-resolution, scientifically accurate simulations at the atomic scale. Researchers from across the university have access to Gadi and all of NCI's data services for their high-impact science.

NATURAM PRIMUM COGNOSCERE RERUM
Australian National University

CSIRO:

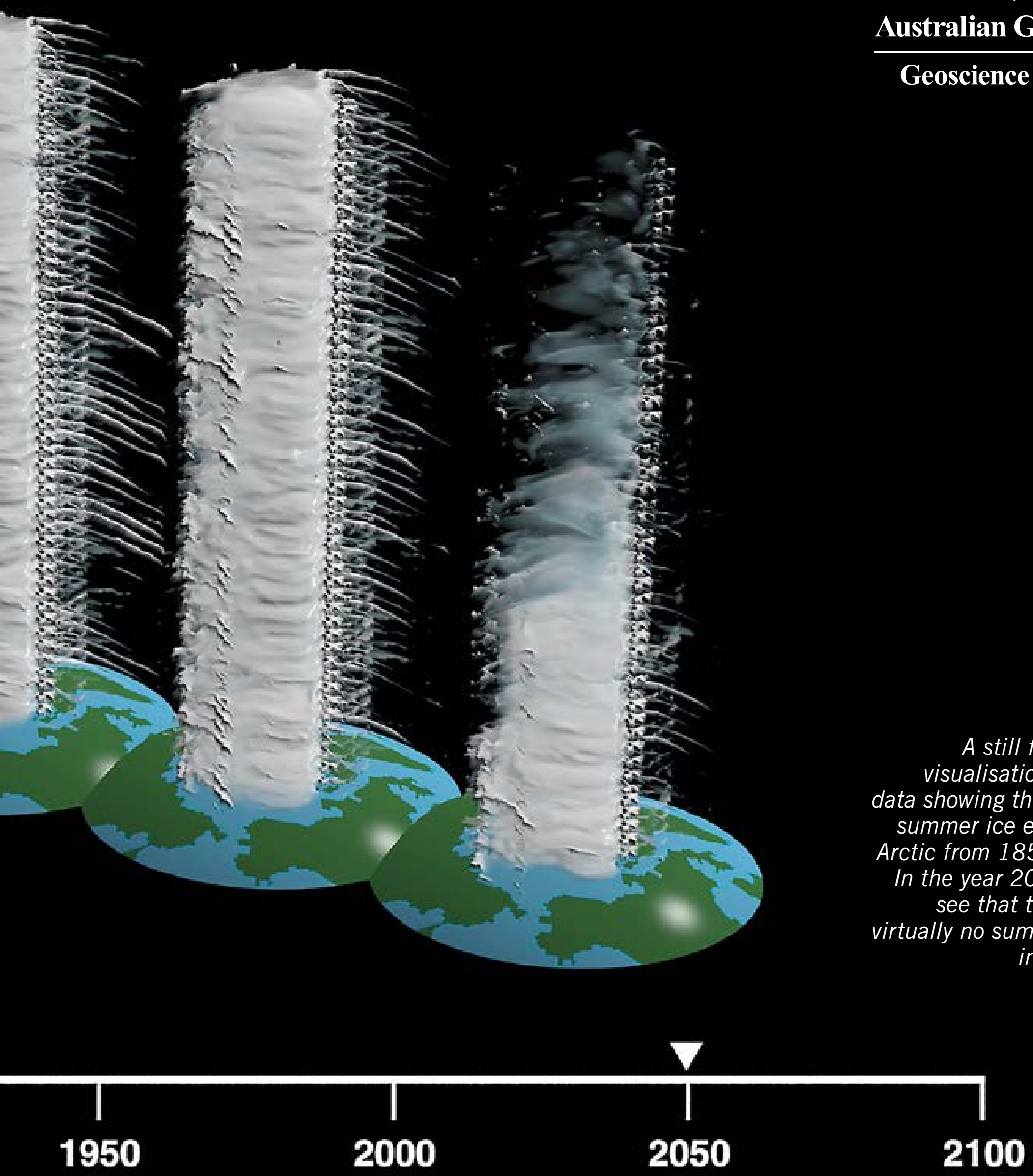
Australia's national science agency, CSIRO, performs climate and weather modelling over many timescales and resolutions. Over the past year, scientists Richard Matear, Simon Marsland and their research teams have been conducting climate studies at both long-term and decadal timescales. Incredibly computationally intensive and producing huge amounts of data, high-resolution simulations such as these rely on access to a high-performance supercomputer and filesystem, an ultra-fast network and optimised software environment, all of which Gadi provides researchers. Some of this work was delivered through the National Environmental Science Program Earth Systems and Climate Change Hub, through delivery of global simulations that explore future possible climates. This was part of a much larger international project coordinated by the World Climate Research Program. These model simulations contribute to our understanding of possible future climates, and were recently assessed as part of the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report.



GEOSCIENCE AUSTRALIA:

Learning about Australia's deserts, forests, oceans, rivers, cities and underground resources requires large datasets and rigorous processing. Geoscience Australia (GA) uses NCI as a trusted repository and analysis platform for many of its research data collections. Comprising daily updated collections of images of the Australian surface taken from a range of satellites in space, derived data products, direct measurements of ocean temperature and colour and much more, the earth-observation data collections that GA uses are stored, quality controlled and made easily available at NCI. At the same time, NCI has developed tools such as the GSKY and On Demand data services to enable smoother processing of these valuable data collections (read more on page 54).

Australian Government
Geoscience Australia



A still from an NCI visualisation of CMIP6 data showing the maximum summer ice extent in the Arctic from 1850 to 2100. In the year 2050, we can see that there will be virtually no summer ice left in the Arctic.

A HUB FOR THE RESEARCH COMMUNITY

3

BUILDING NEW TECHNOLOGIES FOR AUSTRALIAN SCIENCE

Research Overview – Pushing the Supercomputing Boundaries

Replacing a supercomputer after seven years opens up many new possibilities to researchers. Suddenly, they have a lot more computing power at their fingertips, and better underlying technologies. New worlds of calculations are possible. For NCI to understand the limits of a singular system such as Gadi, we first had to test them.

When NCI's Gadi supercomputer came fully online in early 2020, some of Australia's leading research groups were tasked with testing the new capabilities with technically and scientifically ambitious computations. From astronomy to atmospheric modelling, and from genomic science to fluid dynamics, the groups extended their previous models to test what Gadi could do.

The Gadi supercomputer, with 180,000 processor cores and 640 Graphics Processing Units (GPUs), has double the number of processors, double the memory, three times as much active project storage and five times the number of GPUs as its predecessor Raijin. The new generation of processors from Intel, code-named Cascade Lake, can run existing codes faster, especially for some of the more memory-intensive tasks common to large atmospheric and molecular modelling software.

The increased capacity of Gadi allows researchers to undertake larger, more realistic simulations than were previously possible. The advance in technology and the bigger machine mean that truly ground-breaking computations are possible. Having that capacity on hand encourages the researchers to aim higher and extend their conceptions of what might be possible. More researchers can perform more calculations faster, with less waiting and greater performance. As an additional benefit, a new baseline has been set. These scientific outcomes will be used for years to come to show what researchers can achieve with supercomputers and big data.

Professor Christian Jakob and his team from The University of Melbourne and the Australian Research Council Centre of Excellence in Climate Extremes ran a weather simulation of the whole Australian continent at an incredible 400-metre resolution. At the time, this was the largest simulation ever performed with the ACCESS model, the model currently used by the Bureau of Meteorology for producing Australia's daily weather forecasts. The model run used 42 terabytes of memory, split across 12,000 processor cores. It used almost 10% of the entire Gadi supercomputer and more memory than was available in the previous machine.

Professor Jakob says, "One of the prime considerations for the run being feasible at all was the amount of memory available to each process. Getting output from the model written to disk was also a large concern. With guidance from NCI's User Support Team, we managed to output the entire 2.4 terabytes of data produced from an hour of model run within 30 minutes, down from six hours previously."

Professor Richard Sandberg, a computational fluid dynamics researcher at the University of Melbourne investigating ways of making jet engines more fuel efficient, says "The significant increase in computing power, mainly enabled by having access to significant numbers of GPUs, will allow fundamental studies in ever more realistic environments. It will allow us to include more complexity, such as consideration of roughness on surfaces rather than having to assume they are perfectly smooth."

In another growing area of computational science, NCI experts worked with bioinformaticians from the Garvan Institute of Medical Research to improve their genome processing software. This led to major increases in efficiency, allowing faster processing of multiple genomes at the same time.

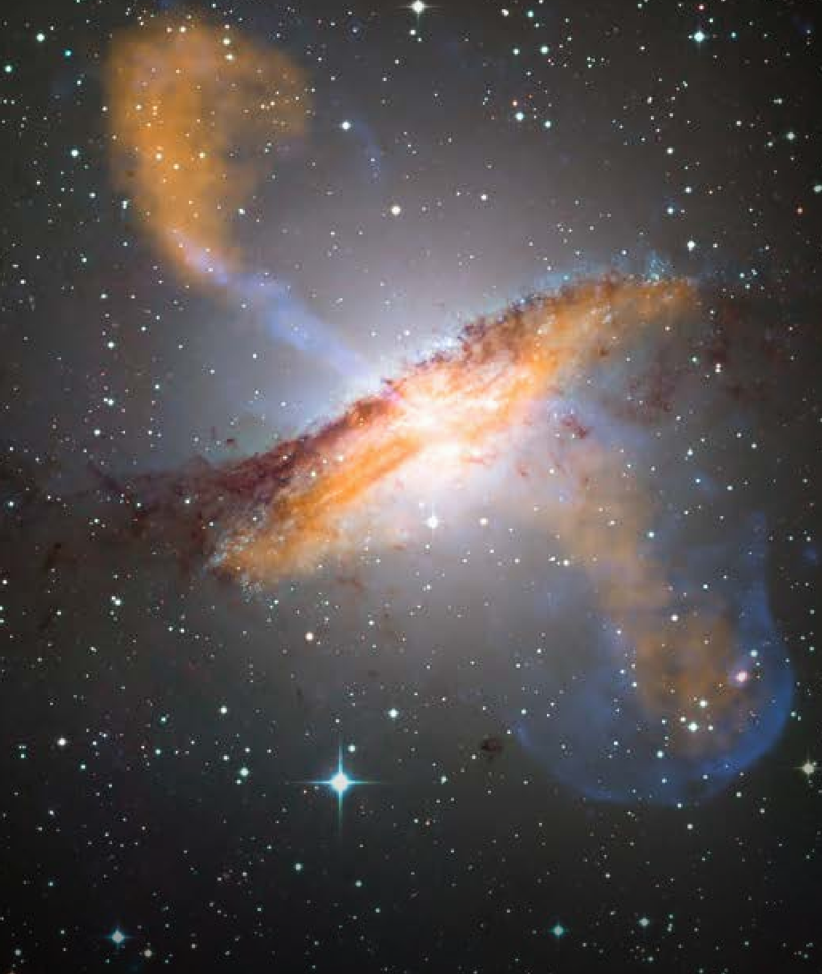
Making use of Gadi's new technologies – and supplementing them with improved software taking advantage of the added functionality – ensures that scientific disciplines keep advancing alongside the new computer systems.

When it came to testing the limits of Gadi's capabilities, the research teams did not hold back. The scale of the data and compute-intensive tasks went up a level – the teams produced so much data through their models that the challenge became finding a way to handle it all. This is exactly the kind of result that NCI was hoping to find. A collaborative process led to software optimisations and code improvements that take advantage of what Gadi's new filesystems can do.

Professor Chris Power, an astronomer from Curtin University and the International Centre for Radio Astronomy Research, says, "If we saved snapshots of the state of the simulation and did the structure finding once the simulation had finished, each run would generate petabytes worth of data.

"By doing the structure finding 'on the fly', we are able to reduce the volume of data produced and track galaxy growth on finer timescales. We were able to get this framework working properly through this stress testing period."

NCI's Gadi supercomputer makes available to researchers more than seven hundred million hours of computing each year. Its technology, partnered with the expertise of the NCI staff, make it a valuable resource for nation-leading science endeavours. The early test projects showed just what is possible with a performance jump of this size, laying the groundwork for the next few years of Australian computational science. Across all disciplines and around the country, researchers are leveraging the national investment in Gadi to advance ambitious research goals.



Colour composite image of Centaurus A, revealing the lobes and jets emanating from the active galaxy's central black hole. Image from ESO/WFI (Optical); MPIfR/ESO/APEX/A.Weiss et al. (Submillimetre); NASA/CXC/CfA/R. Kraft et al. (X-ray)]

The Supercomputer and its Filesystems

GADI – THE SUPERCOMPUTER

3,074 nodes each with **2 24-core Intel Cascade Lake Central Processing Units** (CPUs)

155,000 processor cores

9.26 Petaflops

160 nodes each with 4 **Nvidia V100 Graphics Processing Units** (GPUs)

50 CPU nodes also offer **1.5 TERABYTES RAM**

Data transfers inside the **supercomputer** at up to **200 GIGABITS PER SECOND**

After its first year and a half of operations, the Gadi supercomputer has already demonstrated its value to Australian researchers. It remains the most powerful supercomputer in the Southern Hemisphere, supporting more than 6,000 researchers and underpinning the vast array of NCI's data and compute-intensive services. Powerful CPUs augmented with an increased proportion of GPU accelerators enable next-generation data-intensive research and software to run efficiently.

Powered both by new Intel Cascade Lake processors and by existing Broadwell-generation processors now integrated into the Gadi system, researchers have access to powerful and

efficient hardware for running the biggest computing jobs. High-resolution weather and climate models, simulations of air flow inside gas turbines, and detailed investigations of virus molecule behaviour in the human body: all these intricate and scientifically important topics require software packages that can run across tens of thousands of processors at once.

Similarly, data-intensive research using tens of thousands of files and huge amounts of memory also requires the capabilities of the Gadi supercomputer. NCI enables the most data-heavy applications to run smoothly through the provision of more than 560 terabytes of memory available across the entire system and the ultra-fast network connection to more than 60 petabytes of active storage. Large genome sequencing tasks and hyper-resolution simulations both produce and process so much data that the whole of NCI's integrated storage and compute systems are required.

The NCI data storage systems provide Australian researchers with more than 60 petabytes of research data including more than 10 petabytes of nationally significant managed data collections, and these continue to grow to support Australia's research needs. With performance up to 500 gigabytes per second on the newest filesystems, NCI's data storage systems provide the performance needed for in-depth analysis of research datasets.

In 2020–2021

Total number of projects

1,000+

Total number of users

6,000+

Total computing time

960 MILLION hours

For a full list of all compute allocations on NCI in 2020–21, visit bit.ly/NCIAR20-21CompAllocations

EXTREME EFFICIENCY ASTROPHYSICAL TURBULENCE SIMULATIONS

Astronomy researchers are using Gadi's scale and performance to test the next generation of their star formation simulations. They are simulating the random motions found in magnetised clouds of interstellar gas. These are complex flows of hydrogen gas spread thinly throughout our own and other galaxies in the Universe. The clouds are undergoing supersonic, turbulent motions, interacting through gravity and magnetic fields, and collapsing in some of the ultra-dense regions to form powerful stars, like the Sun.

Computational simulations that run on the world's biggest supercomputers aim to replicate the real turbulent phenomena as closely as possible. To do this, researchers need to use and create software specially designed to run for tens of millions of hours of computing time. Only then can the simulation start to cover the range of length, time and density scales that are essential to understand interstellar turbulence. This type of turbulence is so complex because it couples large and small scales, is a mixture of slow and ultra-fast motions, and has a gigantic range in density that spans many orders of magnitude. Getting an accurate computational representation of this turbulence therefore requires extremely high spatial and temporal simulation resolutions, and extremely efficient computational and processing tools.

Associate Professor Christoph Federrath from The Australian National University uses supercomputers around the world to run turbulence and star formation simulations. Using Gadi, Professor Federrath and PhD student, James Beattie tested their high-performance FLASH simulation package on 99,840 processors at the same time. By efficiently distributing large calculations across thousands of processors, researchers speed-up and transform the quality of research they can produce.

James Beattie, says, "We wanted to push Gadi to its limits and run this code on more processors than we ever have before in Australia. This proof of concept shows that our version of FLASH works extremely well at large scales and that Gadi is highly capable of running jobs that utilise a significant fraction of the whole machine. We are moving to ever more realistic simulations that take into account all of the complexity and chaos of real interstellar gas. These help inform astrophysical gas dynamics and star formation researchers all over the world. Gadi, and the next-generation of Australian supercomputers, are going to be vital for helping us understand the nature of the turbulent interstellar medium."

Testing high-performance software at extreme scales helps researchers develop their codes to run as efficiently as possible for the maximum scientific benefit. The mathematics and simulations of interstellar turbulence in galaxies are remarkably similar to fluid flow simulations used for understanding the movement of air and water on Earth. Thus, scientific and technological advances from one field benefit many others. Computational disciplines at all points on the research spectrum are working with Gadi and other leading supercomputers to optimise, improve and develop their codes for their ever-increasing scientific ambitions.

> *"We wanted to push Gadi to its limits and run this code on more processors than we ever have before in Australia."*
> – Mr James Beattie, ANU

A zoom-in of supersonic, turbulent particle motions that trace the gas dynamics around a strong, coherent magnetic field, coloured by the density and velocity of the particles. Understanding the motions of the particles helps us understand how energy is transported through flows of gas and dust in molecular clouds — the nurseries of stars in the interstellar medium.

Computational Science Enhancements

NCI has a dedicated team working to enhance the capabilities, performance, efficiency and reliability of scientific software. Improvements to the codes that underpin the biggest computational science projects extend the potential outcomes and impact of the research. NCI's expertise in code optimisation starts with large climate and ocean models, and extends all the way to genomics, geophysics, fluid dynamics and astrophysics codes, as well as underlying tools beneficial to all disciplines.

Taking advantage of NCI's world-class supercomputing hardware requires expertise to optimise codes on the new hardware in order to deliver the maximum performance improvements. NCI's HPC Simulation, Scaling and Data Analysis Optimisation Team helps research groups understand the bottlenecks and inefficiencies in their code and plan for future growth in resolution or scale. The team can provide solutions for researchers to implement that save them time and computational resources and lead to better scientific outcomes.

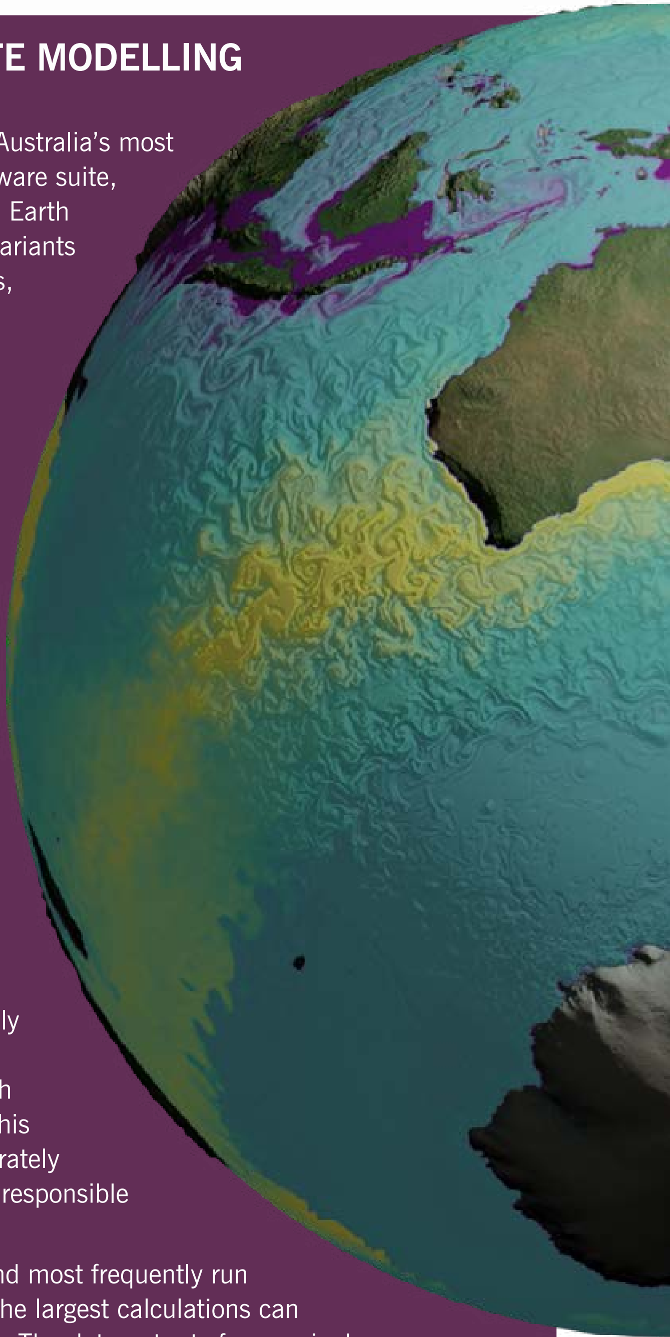
For more than a decade, NCI has been supporting the Consortium for Ocean-Sea Ice Modelling in Australia (COSIMA) with the development and optimisation of some of Australia's most important scientific codes: the various forms of the Australian Community Climate and Earth-Simulation Suite (ACCESS). From regional to global scales, and with a particular focus on Southern Ocean simulation, the ACCESS model plays a central role in national daily weather forecasts, seasonal planning and long-term climate projections.

Over the years, NCI has helped COSIMA implement improvements to the ACCESS models that have led to higher resolutions, more accurate handling of the biogeochemistry of ocean circulation and more efficient operations (see Case Study *Critical support for climate modelling suite* on page 48 for more information about the COSIMA work). The large data collections produced by and used for running software such as the global ACCESS models on tens of thousands of processors at once also requires specialised development work from NCI's team. Processing, storing, labelling and making the data available for open-access sharing is an important part of the computational science enhancements that NCI offers researchers.

NCI continues to build our support for genomics software development, in particular with our long-term collaborator the Garvan Institute of Medical Research. As the field of bioinformatics develops, efficient workflows for the processing and analysis of genomes are becoming ever more critical. NCI and Garvan are already working on next-generation technologies for scaling up the capabilities of existing genomics technologies.

NCI's code optimisation efforts help create the high-performance computing workflows that lead to national impact and significant scientific advancement. Code optimisation produces more efficient jobs, enabling increased resolution and decreased turn-around times on models that impact the everyday lives of Australians. Getting the most from massive supercomputer calculations requires carefully optimised codes, and NCI provides the necessary expertise.

CRITICAL SUPPORT FOR CLIMATE MODELLING SUITE

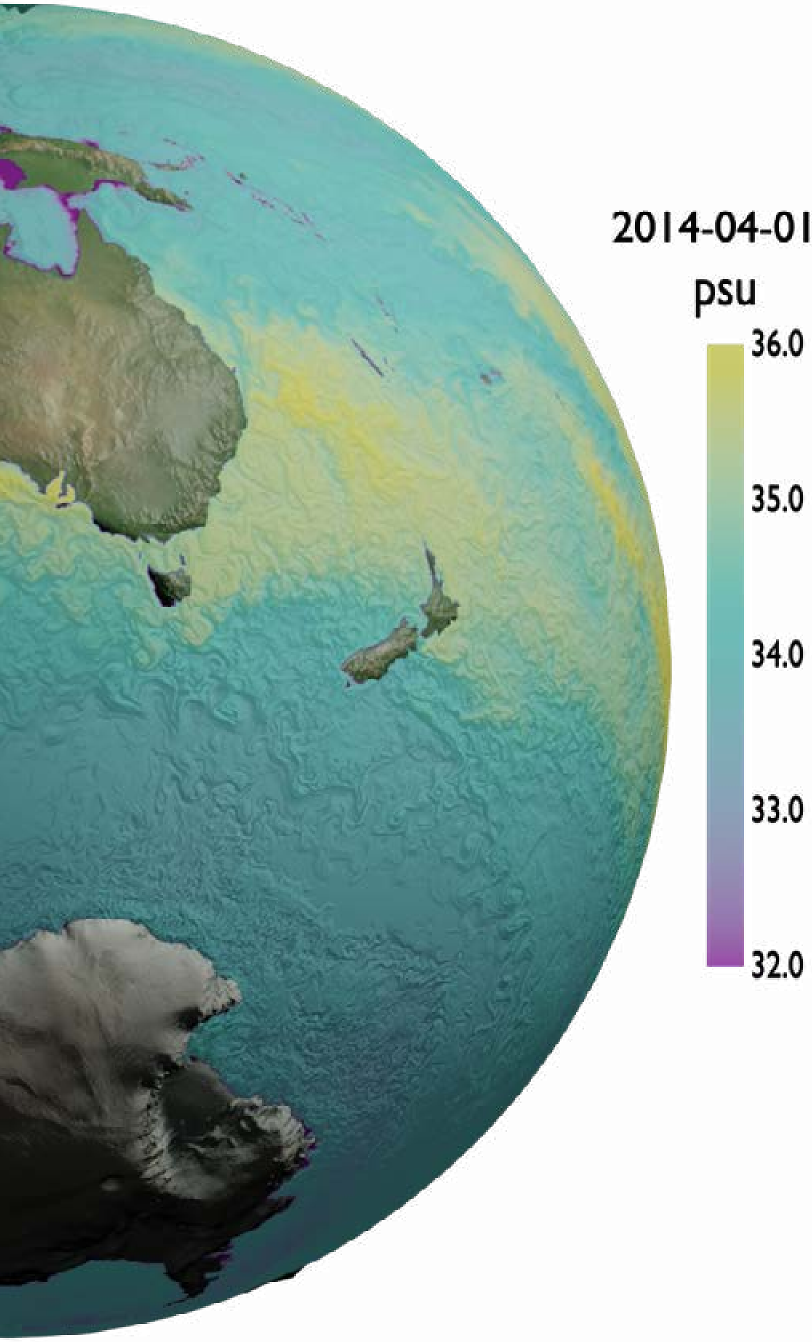


Ocean and sea-ice models are a key element of Australia's most widely used weather and climate prediction software suite, ACCESS (the Australian Community Climate and Earth System Simulator). Comprising multiple model variants suited to different areas of interest and use cases, the ACCESS suite is used daily by the Bureau of Meteorology as part of the weather forecasts all Australians rely on, and by weather and climate scientists all over the country. These high-resolution models allow us to understand how ocean currents and sea-ice affect our weather systems. Since 2012, the Consortium for Ocean-Sea Ice Modelling in Australia (COSIMA) has been building a collaborative community of modellers and scientists working together to improve Australia's ocean models. Supported by NCI for its significant computing needs and critical code optimisation, COSIMA has become a central part of climate model development in Australia.

NCI has played an important role in the COSIMA group since its beginnings, most notably through the provision of expert assistance with optimising the code to run efficiently at higher resolutions. Focusing primarily on the ocean and sea-ice version of the ACCESS model, ACCESS-OM2, NCI helped COSIMA reach the current 0.1 degree resolution milestone. At this resolution, the flagship model is capable of accurately simulating the turbulent swirling eddies that are responsible for most of the ocean's variability.

The ACCESS model suite is one of the biggest and most frequently run pieces of software on the Gadi supercomputer. The largest calculations can run for months on more than 12,000 processors. The data outputs from a single simulation can reach more than 10 terabytes, and over recent years the COSIMA group has produced more than 500 terabytes of valuable data. Making sure that the software runs as efficiently as possible is critical for future resolution improvements, overall model performance and energy savings.

COSIMA researcher Dr Andrew Kiss from The Australian National University says, "Our high-resolution models require enormous computational resources that are only available from supercomputing facilities like NCI. NCI's expert team have provided very valuable expertise which has helped COSIMA make efficient use of NCI's hardware in running ACCESS-OM2."



Surface salinity in the oceans around Australia from a run of the ACCESS-OM2 0.1 degree model. Credit: Josué Martínez Moreno, ANU

The improvements and developments that COSIMA has produced over the past 9 years directly benefit the broader Australian climate and weather community. The improvements have made their way through to the national long-term climate model ACCESS-CM2, and to the forthcoming Bluelink ocean forecasting system from the Bureau of Meteorology. Dr Kiss says, "COSIMA has made state-of-the-art models readily accessible, facilitating groundbreaking research and developing postgraduate students into experts who can contribute to solving some of the world's most important problems in the decades to come."

NCI has been the key supporting infrastructure providing the necessary expertise in code optimisation, data management, software testing and high-performance computing that has enabled COSIMA and many other research groups to produce world-class science.

"NCI's expert team have provided very valuable expertise which has helped COSIMA make efficient use of NCI's hardware in running ACCESS-OM2."

– Dr Andrew Kiss, ANU

Data Science Technologies

To compete and collaborate internationally, scientists in all disciplines are increasingly reliant on large, distributed data collections for use in their highly specialised workflows. NCI provides the research environment that they need through collating and curating this data into high-quality, analysis-ready datasets. This opens up new research opportunities for scientists, who can focus on their work while NCI maintains the collections in a research-ready state.

We have constantly been adding to our data collections' functionality to help users better access, analyse and share the data they need every day. A core aim is making sure that the data stored at NCI is safe, secure and accessible. As a trusted repository for some of Australia's biggest datasets, we play a key role in the national data science community.

All the significant national data collections at NCI, in total making up more than 10 petabytes of data, meet the international FAIR data standard: data should be Findable, Accessible, Interoperable and Reusable for the research community. Maintaining the datasets entrusted to us by the research community and the national science agencies for broad scientific use is a central facet of our data collections work. We are constantly working with our data providers to make sure that their data collections are up-to-date and accurate.

We provide new tools and technologies for scientists to conduct their data science research. Over the past year, we have developed high-performance capabilities for efficient and scalable data processing and computing. Using software such as Dask and Xarray, we can engineer ways for researchers to inhabit a middle ground between data analysis and parallel processing. Effective data science relies on frictionless access to both the data and the means of understanding it. NCI's data science technologies help researchers extract the relevant pieces of information from datasets and find creative connections between them.

The Shoemaker Crater, the circular shape at the bottom right, is visible in this false-colour satellite image of the landscape approximately 100 kilometres north-west of the town of Wiluna in the Little Sandy Desert in Western Australia. Image from the European Space Agency's Copernicus satellite program.

BUILDING POWERFUL DATA SCIENCE ENVIRONMENTS

The growth of computational science has led to more diverse and varied workflows across the scientific spectrum, and much more complex data collections. New computational use cases, and the analysis-ready datasets that NCI has systematically organised through its National Reference Data Collections, call for fast, programmatic access tools for analysis across multiple domains. NCI's solution is a suite of simplified environments called Pangeo that speeds up data analysis using the powerful Gadi supercomputer and a range of carefully optimised software tools and libraries.

Researchers around Australia in areas such as ocean modelling and earth observation regularly handle data collections hundreds of terabytes in size. The size of the datasets involved is such that specialised tools are required simply to view and analyse the data for scientific purposes. Visualisation, interpretation and compilation of datasets uses parallel computing methods implemented through software such as Dask and Xarray. NCI has provided researchers a suite of software that make possible much quicker and more scientifically useful workflows. The ultimate aim is to redirect the time that researchers would normally spend on data collecting, data pre-processing and managing software installations towards research and analysis.

Like all the services operating in NCI's tightly integrated computing and data ecosystem, these environments provide researchers access to petabytes of nationally significant research data collections, as well as their own research datasets, all stored on site for fast and simple access. These tools come together to form an interactive distributed parallel computing environment. NCI is uniquely placed to offer environments such as this that make use of the integrated data, software and computing infrastructure.

NCI set up this novel data analysis environment by integrating the Python software stack and NCI computing resources. This gives researchers the best of both worlds: powerful parallel computing at scale and capable data analysis tools, all working together.

The Pangeo environment is an acceleration tool encompassing technologies from high-performance computing, data science, big data analysis and more to create a powerful space for interacting with and interrogating analysis-ready data. With petabytes of data collections available on site at NCI, the Pangeo solution brings the analysis to where the data is. NCI's storage, computing, data management and expertise work together to enable major advances in data science for Australian researchers.

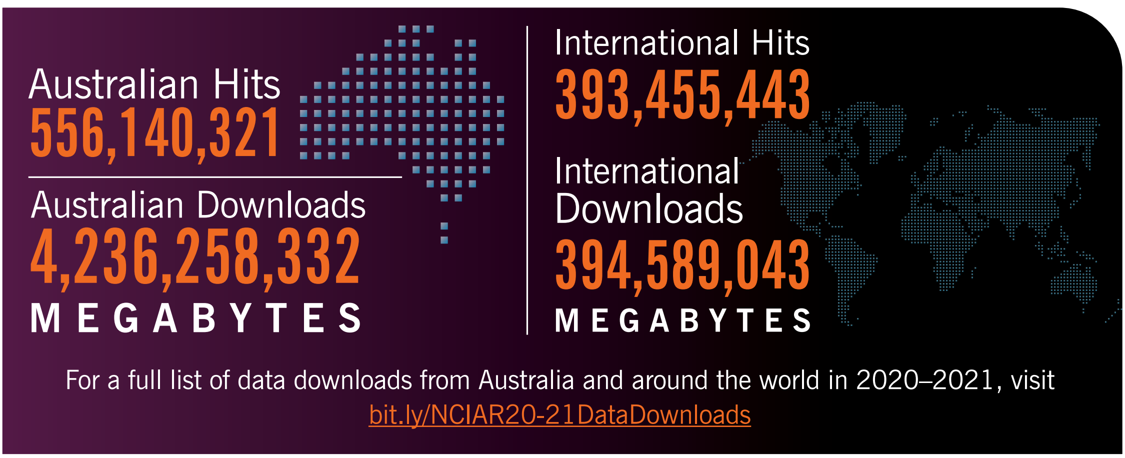
Data-Intensive Services

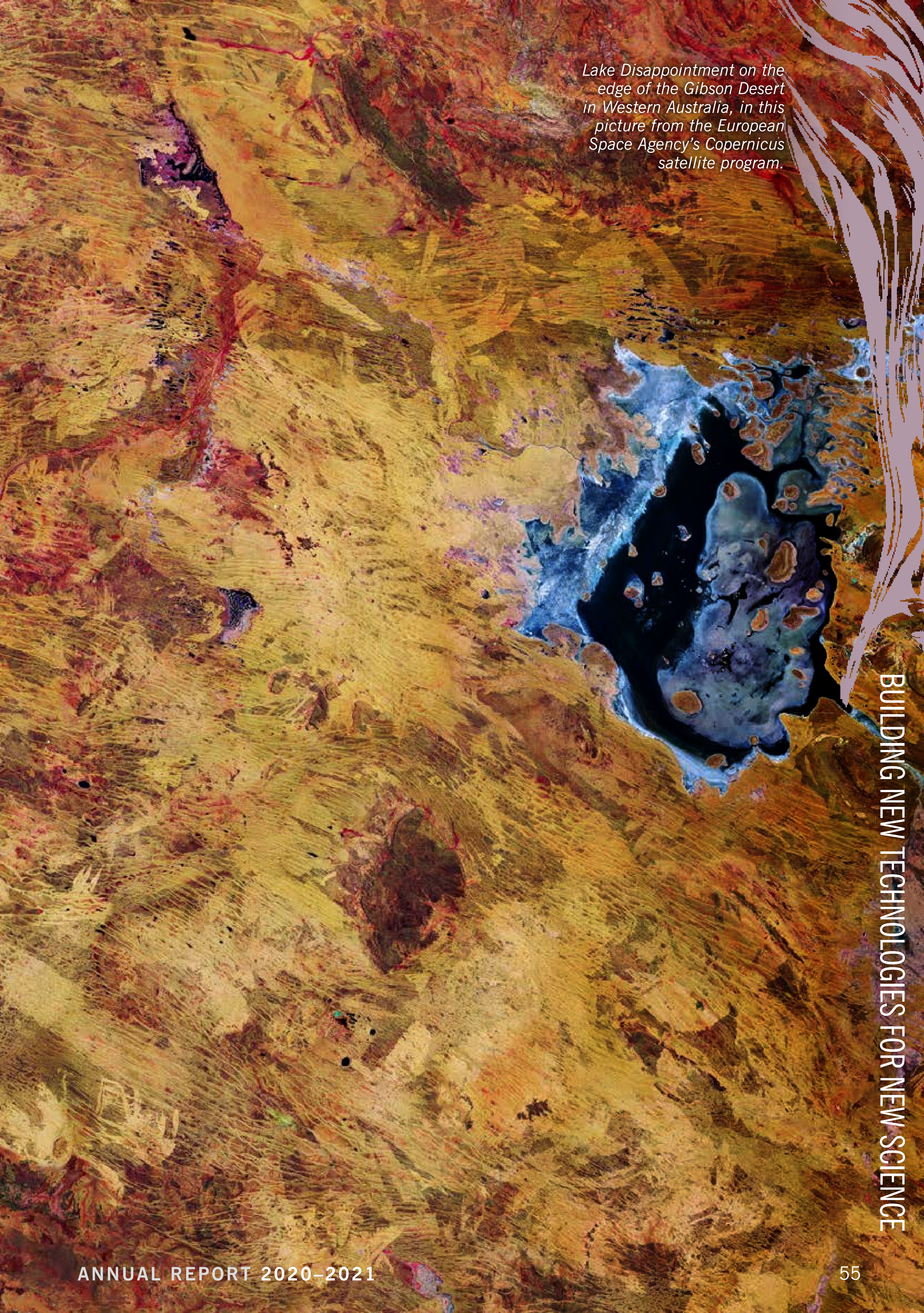
Data-intensive services open up opportunities for researchers to take existing datasets and approach them in new, compelling ways. We develop and implement new computational technologies that take advantage of our integrated big data, high-performance computing and scientific software ecosystem. This allows us to offer powerful analysis and processing services for our users conducting data-intensive science. NCI's combination of computational performance, data storage capacity and data management expertise offers dynamic ways of interrogating data that extend the capabilities and ambition of the research community.

NCI has built a new service providing access to virtual computing resources and tools for data analysis activities. Using the Open OnDemand (OOD) software developed under an American National Science Foundation grant at the Ohio Supercomputing Centre, this new service allows NCI users to perform data post-processing and analysis in an interactive, Python-based environment called Jupyter. This benefits a wide range of scientific disciplines, such as ocean modelling, in which the data modelling and the data analysis can be decoupled. The initial modelling takes place on the Gadi supercomputer across tens of thousands of processors at the same time, and then the post-processing, analysis and data visualisation takes place on the OOD running its own dedicated hardware.

We continue to build our GSKY (pronounced ji-skee) data service to support the needs of data-intensive research taking place at NCI. A technology to dynamically process and distribute earth-observation data from the NCI Data Collection, it streamlines data discovery and automated processing. A high-performance service that makes accessible otherwise obscure and complex datasets, GSKY is now playing an even bigger role as the compute engine for environmental science data cubes.

Science in all disciplines is now producing – and requires storage for – large amounts of research data. The increased volume of data collections provides an opportunity to develop new ways of analysing data going beyond the methods of the past. Computational methods integrated with large data collections extend the science that the data enables in the first place. The combination of supercomputing performance and innovative data services connects and enriches the valuable data stored on site at NCI.





Lake Disappointment on the edge of the Gibson Desert in Western Australia, in this picture from the European Space Agency's Copernicus satellite program.

DEVELOPING A HIGH-PERFORMANCE DATA ANALYSIS PLATFORM AT NCI

Over the past year, NCI has been developing a high-performance online data analysis platform, Open OnDemand (OOD), that streamlines and speeds up big data workflows from across the scientific disciplines. OOD allows NCI to provide users different tools for use in data analysis, all running on NCI's cloud infrastructure. Through OOD, users have simple and efficient access to the Gadi supercomputer, virtual desktop environments, dedicated data analysis spaces and more.

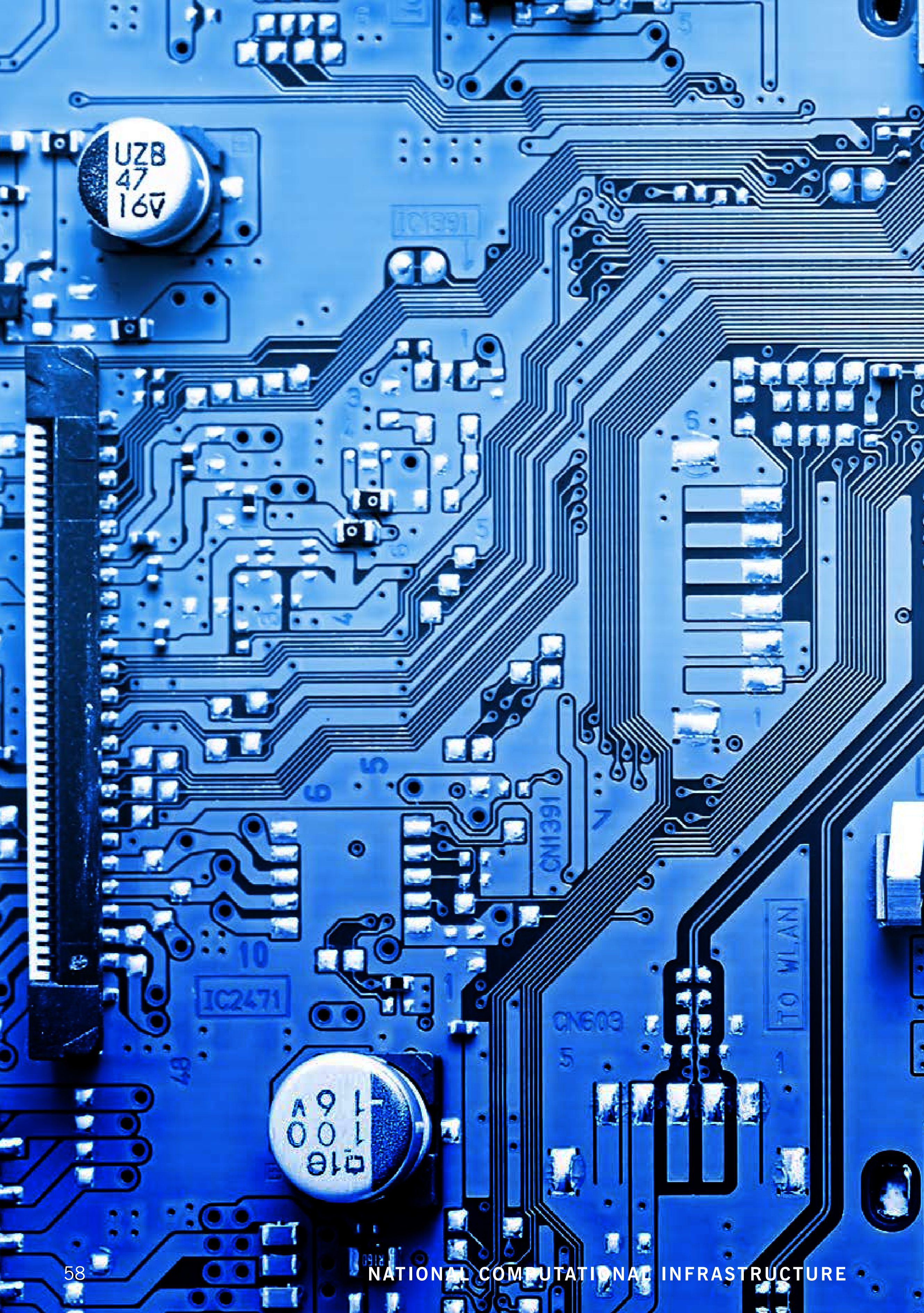
The OOD environment, easily accessible through a web browser, offers several different specially designed apps, custom-built by NCI. Over time, the available apps will grow, with additional functionalities currently being planned. The apps allow users to take advantage of NCI's high-performance computing infrastructure, access research data collections stored at NCI, and perform their analysis in an interactive way. Leveraging a user's NCI credentials, OOD easily lets users log in and request the processors, memory, and software packages they need for their data analysis tasks.

This new platform allows users to launch JupyterLab and Virtual Desktop Infrastructure instances running on dedicated cloud hardware. Early-access users in the fields of genomics and ocean modelling have already found that the OOD system makes powerful, large-scale data analysis much more seamless.

OOD is a prime example of the power of connecting vast data collections to powerful computing systems and to optimised software solutions. NCI continues to streamline the researcher experience by taking on the management, administration and building workload that such a tightly integrated system requires. Additional apps with deep connections to the Gadi supercomputer, specialised hardware, dedicated software environments and more are all in the pipeline.

As a national hub for data and compute-intensive research, NCI increases the productivity of the research community by making available robust, high-performance tools with access to some of Australia's most significant data collections. NCI's OOD allows researchers to easily access far more powerful compute resources through far more capable virtual data analysis environments than they ever have before. Previously, researchers would have had to spend days or weeks setting up and managing their own virtual environments, with less functionality, security and availability. Now, they can focus on their science.

BUILDING NEW TECHNOLOGIES FOR NEW SCIENCE



4

A BIG DATA AND BIG COMPUTING LEADER

Research Overview – A Central Climate Data Hub

Climate modelling taking place around the world underpins global agreements and transformative decision-making processes. Those climate models, and the decisions they help inform, impact all people through policy changes and the life-changing environmental shifts the models signal are upon us. As the global climate continues to change and affect us, our way out from the threat of climate change requires robust scientific data and processes to underpin it. NCI tackles the unique data challenges that global climate action requires, to make important science possible for Australian researchers.

Historically, one of the most challenging aspects of dealing with climate data has been the storage, transfer and accessibility of the various global datasets. NCI has stepped up to the challenge and provides the storage and data quality assurance that the community needs to pursue its nationally significant work. By removing the data handling step from the scientific workflows, and by providing research-ready data to the Australian researchers at CSIRO, the Bureau of Meteorology, the ARC Centre of Excellence for Climate Extremes and more, NCI streamlines and simplifies the scientific process.

In the climate world, all data is big data. Routinely surpassing hundreds of terabytes produced by large climate model runs, and with reference datasets now totalling in the tens of petabytes, robust and reliable solutions for sharing, accessing and validating the data are required. In the early 2000s, when climate datasets were relatively small and less complex, research groups were still sending each other physical hard drives and monitoring distribution informally. As datasets have expanded and the number of researchers involved has grown, dedicated solutions for the distribution of climate data have become necessary.

Now, NCI is the Australian home for climate and environmental data from research centres around the world, allowing scientists here to share their modelling outputs and interrogate those of colleagues overseas. The datasets include:

- CMIP6, the Climate Model Intercomparison Project Phase 6, used in the 2021 Sixth Assessment Report of the Intergovernmental Panel on Climate Change;
- ERA5, the 5th version of the European Centre for Medium-Range Weather Forecasting Reanalysis, providing more than 40 years of global climate observations; and
- BARRA, the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia, providing more than 30 years of high-resolution observations of the Australian region.

NCI's role as the Australian hub for global climate data is built on many years of experience and close partnerships with international data peers. As a Tier-1 node of the Earth Systems Grid Federation (ESGF) alongside centres in the USA, Europe and the UK, NCI is the only home of CMIP6 data in Australia and the region. Our advanced big data management processes create a powerful and reliable environment for researchers to find the data they need to perform their critical science. The data repositories at NCI provide advanced search, citation and sharing functionality, central to any modern collaborative scientific endeavour.

The future of high-performance data will be linked, carefully managed datasets and integrated processing tools. When the size of the datasets and the complexity of the analysis grow like they have in the climate space, connecting the storage platform to the computational platform is a core improvement to the process. Instead of spending ineffective time moving data between repositories and computing centres, researchers can perform their analysis, reach conclusions, share their findings and collaborate based on the stable foundations of the data hub.

NCI's data management expertise, computational capabilities and central position within the national research ecosystem makes it the ideal hub for nationally significant climate research. Australian contributions to the CMIP6 datasets have been incorporated into the 6th IPCC Assessment Report, and scientists all over the country are finding new and powerful ways of using the international and national datasets now available to them to improve our weather and climate modelling processes, increase model accuracy, and add extra detail to our scientific predictions.

National Engagements

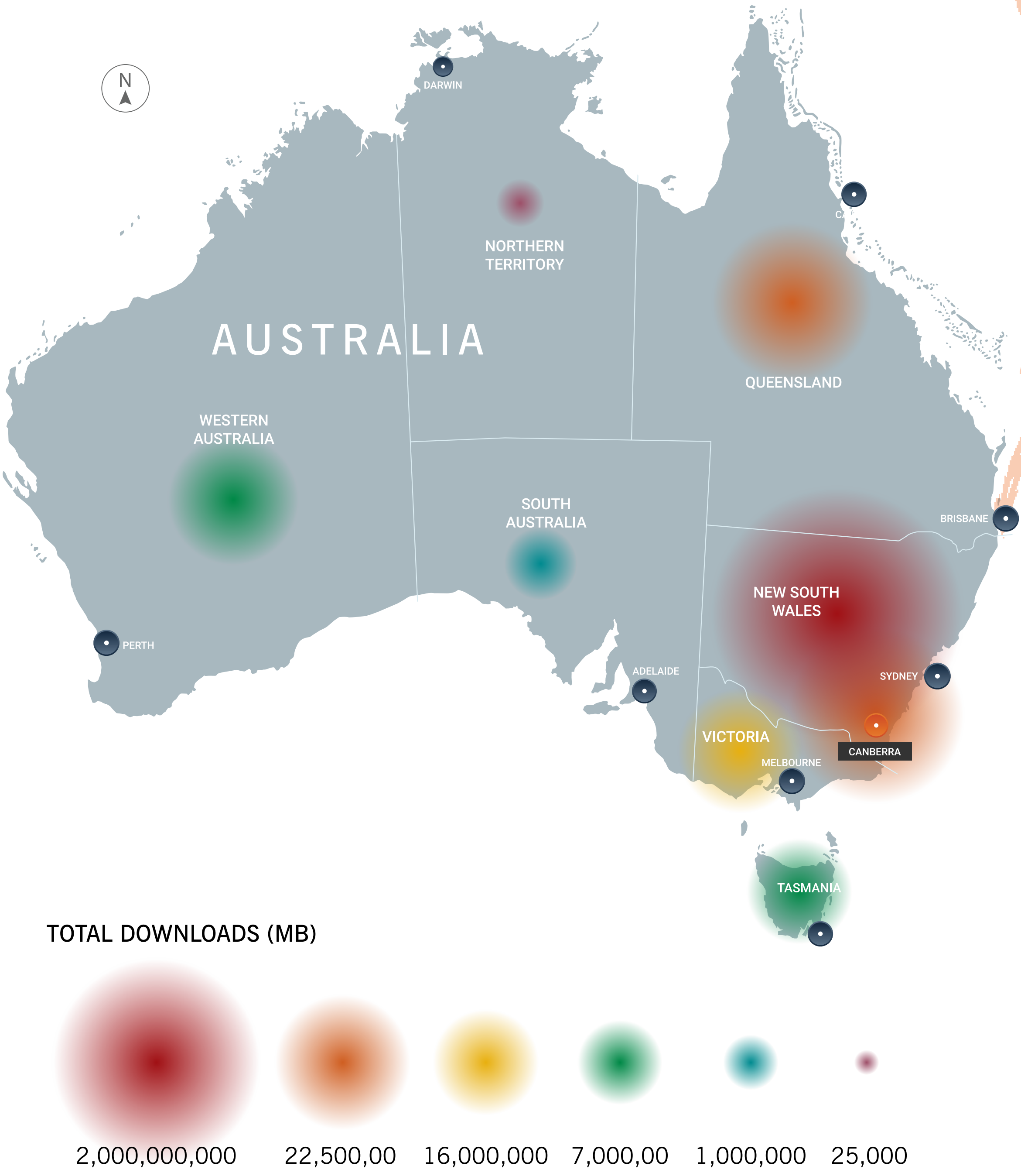
NCI is a central pillar of the Australian computational and data science community. As a critical component of national science infrastructure, we provide the tools that researchers and research institutions all over the country rely on for nationally significant data access and sharing, simulations and modelling, and collaboration. We are constantly working with our stakeholders to further our shared ambitions and anticipate the research needs of our user communities.

Our primary technical partner for HPC and HPD is the Pawsey Supercomputing Research Centre, the Western Australian supercomputing facility providing computational and data support for the Square Kilometre Array radio astronomy project and many other research groups around the country. We work closely with Pawsey on user training, technical development and events.

We similarly engage closely with the major national science agencies: CSIRO, the Bureau of Meteorology and Geoscience Australia. Together, they are responsible for the majority of Australia's peak scientific research. Foundation Collaborators within the NCI Collaboration support, and in turn rely on, NCI's vast array of services. These ongoing relationships set the stage for advances in their scientific programs and scientific ambition. The other members of the NCI Collaboration include The Australian National University as the fourth Foundation Collaborator, and a host of other universities and research organisations as Major and Minor Collaborators. The NCI Collaboration brings together the country's leading research bodies, all contributing funds to support and access the computational infrastructure underpinning their science.

Region	Total Users	Total Hits	Download (MB)
New South Wales	21,128	440,821,519	2,220,778,219
Australian Capital Territory	1,723	48,711,603	35,644,892
Queensland	3,525	25,747,475	22,093,161
Victoria	4,305	6,458,161	17,305,156
Western Australia	1,881	2,101,433	15,136,844
Tasmania	470	675,212	7,645,941
South Australia	1,198	2,509,090	1,807,333
Northern Territory	69	135,400	32,912
Unknown	2,190	28,980,428	1,915,813,873
Grand Total	**36,489**	**556,140,321**	**4,236,258,332**

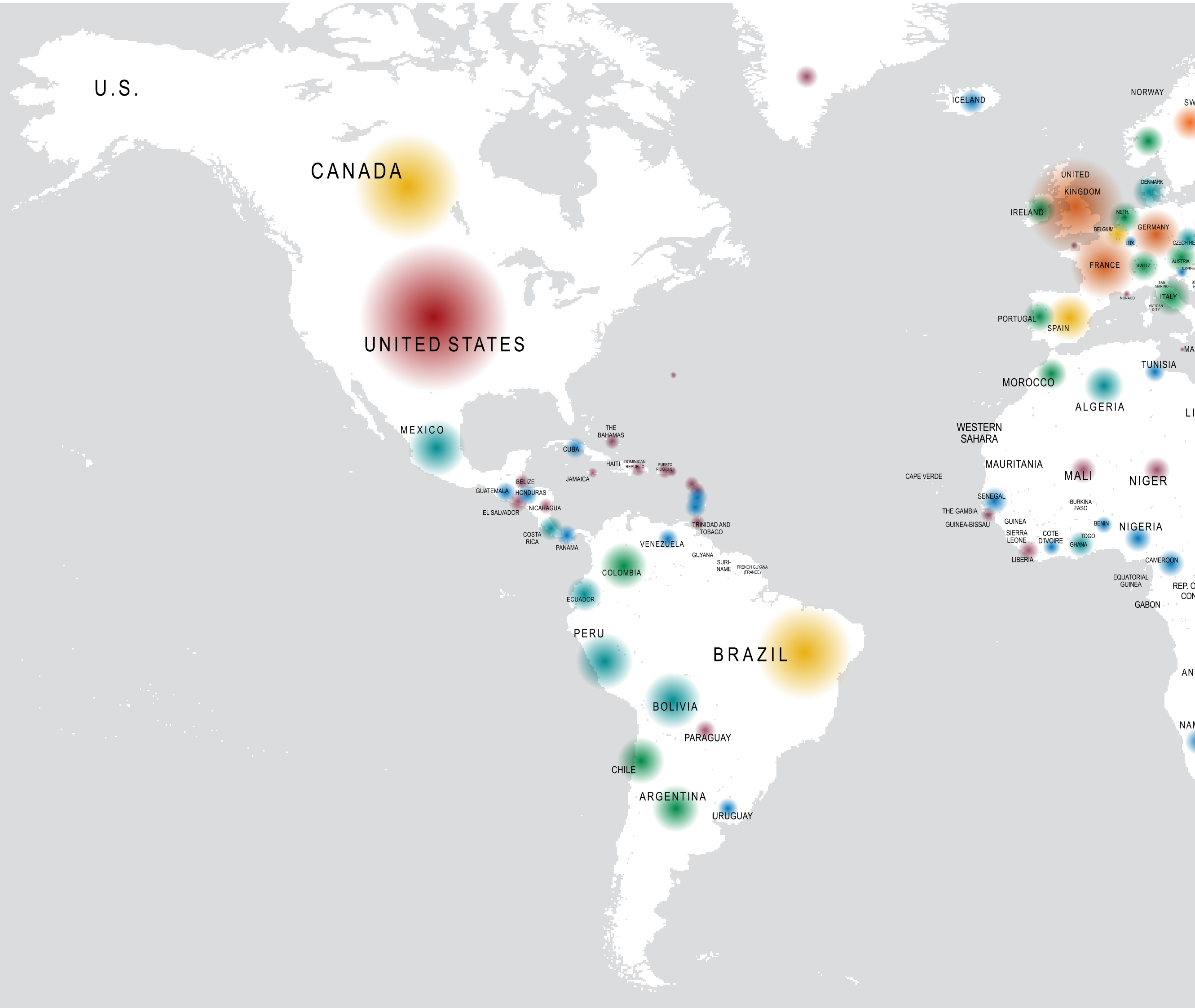
Funded under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), NCI also works closely with other NCRIS facilities including Pawsey, the Australian Research Data Commons, Bioplatforms Australia, AuScope, the Terrestrial Ecosystem Research Network, the Australian Community Climate and Earth-System Simulator Research Infrastructure and more. These national science infrastructure facilities support each other, building on their strengths to leverage even more value and impactful research capabilities for the nation (see section A Hub for the Research Community on page 30 for more about these national engagements).



International Engagements

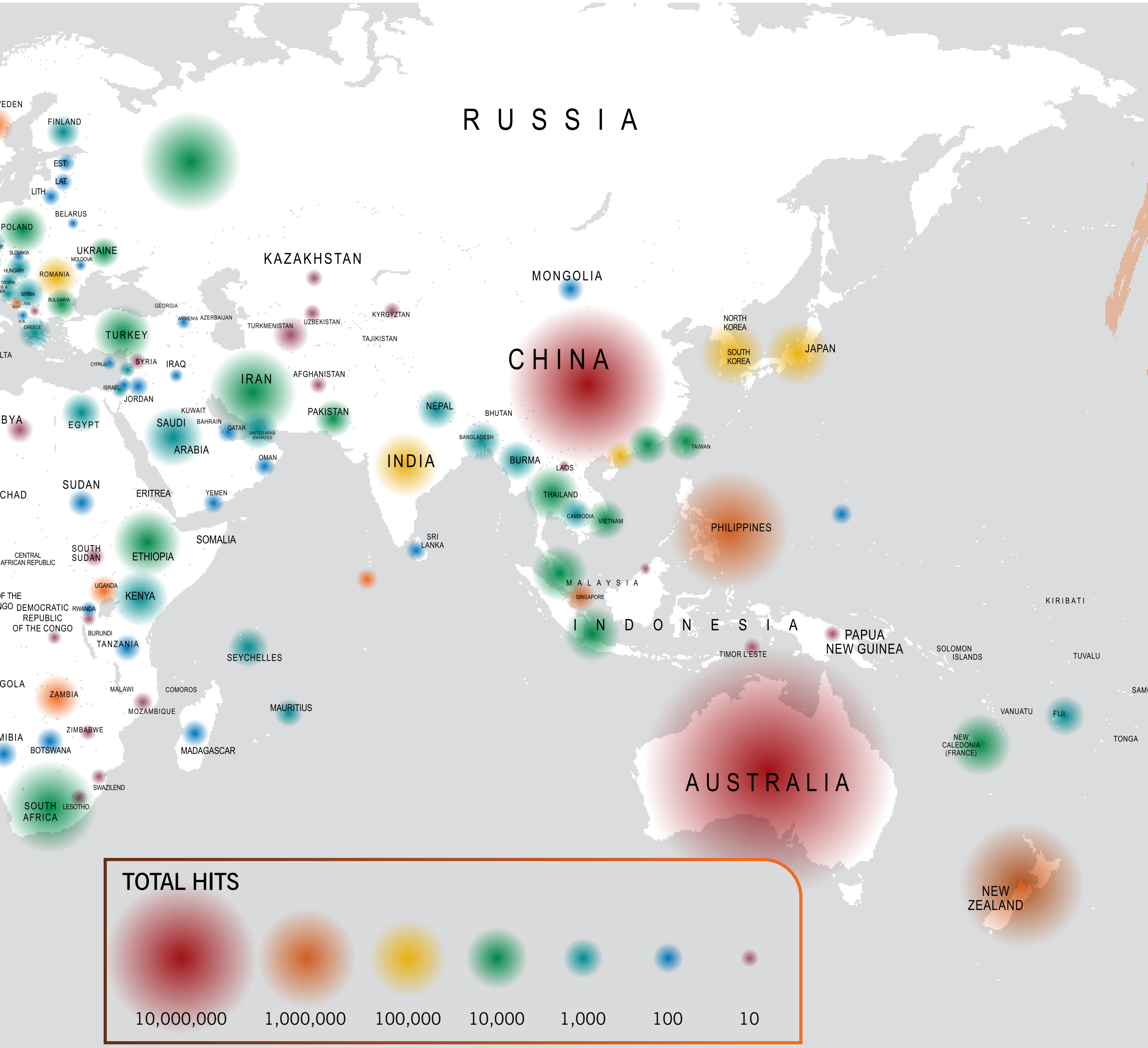
As a global leader in high-performance data and computing, NCI works closely with numerous international research organisations. We are in constant contact with our peer facilities in Japan, South Korea, Singapore, New Zealand, the European Union, the United Kingdom and the United States of America. We have much to learn from each other, from networking developments to training methods, user interaction and software development.

NCI plays a central role in the climate research community, as one of the six nodes worldwide collaborating to share and make available climate data through the Earth Systems Grid Federation. Alongside partners including the US Department of Energy, NASA, the German Climate Research Centre, the UK Centre for Environmental Data Analysis and others, NCI provides access to petabytes of climate model data for the Australian and global research community. This data, critical for thousands of researchers worldwide and for the Intergovernmental Panel on Climate Change's latest Sixth Assessment Report, is a striking example of NCI's deep and continuing engagement with the most central scientific research activities.



On the high-performance computing front, NCI is a member of ADAC, the Augmented Data and Accelerated Computing Consortium. Other members include the Swiss Supercomputing Centre, Oak Ridge National Laboratory in the USA and RIKEN in Japan, all organisations at the pinnacle of extreme-scale supercomputing innovation. Aiming to develop technologies and skills between collaborating organisations, ADAC highlights that the future of supercomputing and big data technologies will rely on global cooperation. NCI continues its involvement with the US-led COVID-19 HPC Consortium of research institutes and major industry partners working together to support computational scientists around the world with COVID-19 research projects.

NCI also engages closely with the American Geophysical Union and European Geosciences Union, helping guide their annual meetings and contributing to their busy programs and skills-sharing networks. This fits into a broad range of data-related international engagement activities featuring the European Space Agency, the United States Geological Survey, the United Kingdom Meteorological Office and others who all work with NCI to share data further and improve scientific outcomes.



Diversity and Inclusion

NCI organises the Women in High-Performance Computing (WHPC) Australasian Chapter with colleagues from Pawsey, Monash University, Australian eResearch Organisations and the New Zealand eScience Infrastructure. First conceived in 2019, and formally launched in October 2020 as part of the global WHPC movement, the group has rapidly become an important way for HPC and HPD organisations to embed principles of Diversity and Inclusion throughout their work.

In 2020–21, NCI facilitated fortnightly WHPC catch-up meetings to help women in our community find their peers and helped lead initiatives to improve diversity and inclusion in the HPC workforce. Networking and mentoring sessions at prominent data science and computing conferences have provided valuable spaces for researchers to meet and discuss ways to support their communities. The Directors of all the WHPC facilities have committed to work together to implement concrete steps for improving Diversity and Inclusion in the workplace and the broader research community.

NCI, as Secretariat for the National Computational Merit Allocation Scheme, is also working closely with the Office of the Women in STEM Ambassador on the National Trial of Anonymising Grant Proposals. This trial, now in its second year, will provide important data to a University of New South Wale study on the effectiveness of anonymising proposals to reduce bias. The study is based on international grant application models which demonstrated that removing identifying information from processes can level the playing field. Improving equity and diversity practices can support women and other marginalised groups, and early careers researchers, in competitive merit grant programs.

A fortnightly gathering of the WHPC+ Australasian Chapter.

Training

NCI continues to expand on the breadth and frequency of training for our cohort of Australian researchers. Recognising the need to develop an increasingly comprehensive training curriculum, NCI established a dedicated user training team in January 2021. The first role of this team was to further refine the scope of user training supported by NCI.

A strategy based around this scope was realised using internal and external stakeholder communication, as well as extensive user polling. With the support of the NCI executive, the user community feedback that we received assisted the training team in creating a comprehensive approach to the delivery of user training for the 2021 calendar year, with further development already planned for 2022 and beyond.

Ably assisted by the NCI User Support Team, NCI has been delivering a regular series of training webinars for beginners, with an emphasis on the fundamentals of using the Gadi supercomputer. In addition, the User Support Team has been holding virtual 'office hours' for NCI users each month. These sessions are a great opportunity for NCI users to get to know the Team, ask questions and highlight any challenges they have come across.

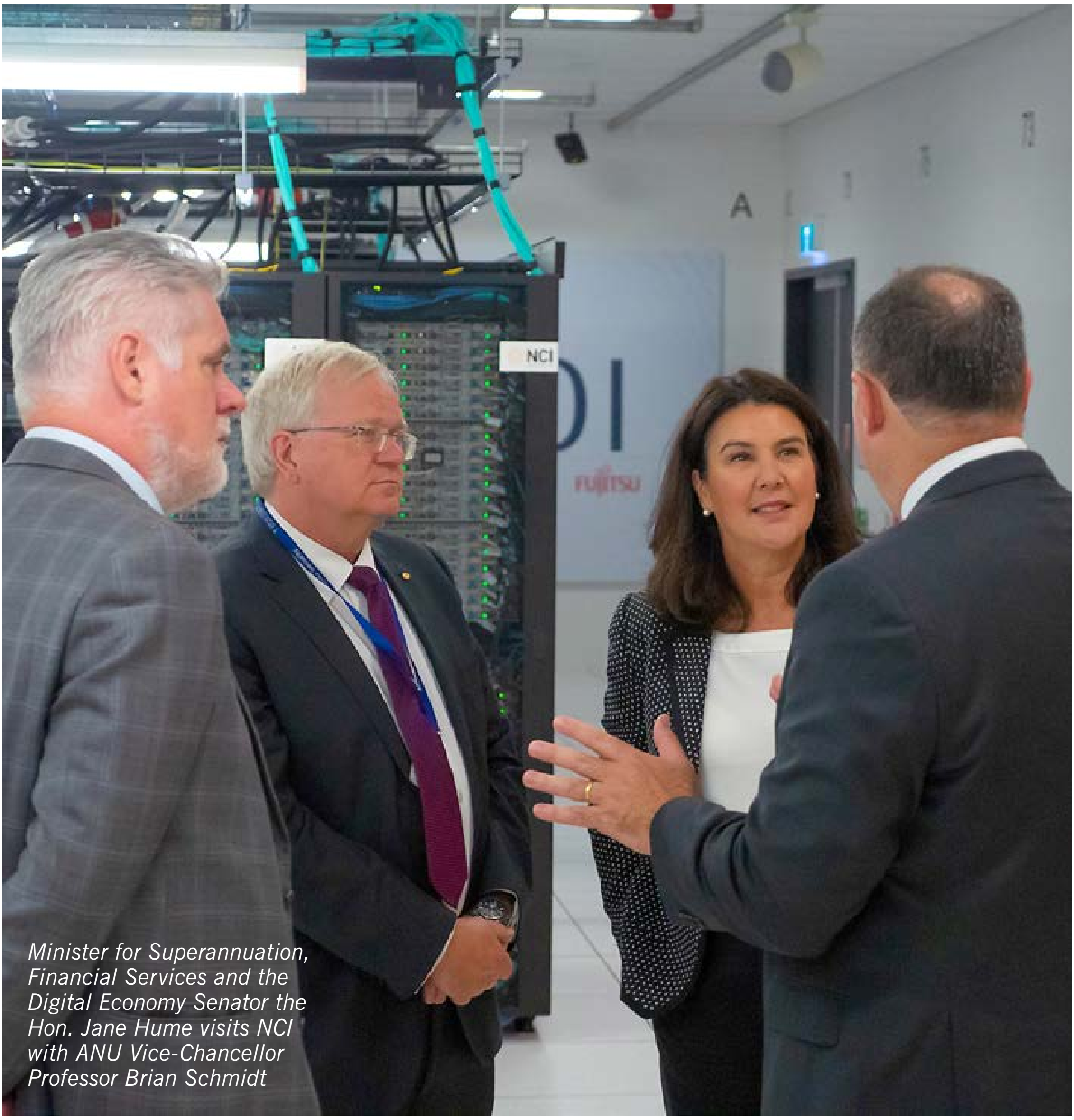
Aiming to build connections and knowledge across the user community, NCI's new TechTake series features science-focused presentations from renowned national and international computational scientists. These events have featured topics including the machine learning future of weather forecast models, GPU-accelerated computational fluid dynamics and a look at new NVIDIA GPU programming tools.

In collaboration with institutional partners around Australia, we continue to offer engaging and educational events for our user community. Of particular interest to newer users will be a comprehensive set of training courses to be delivered for the NCI user cohort in collaboration with Intersect Australia starting in Q4 of 2021.

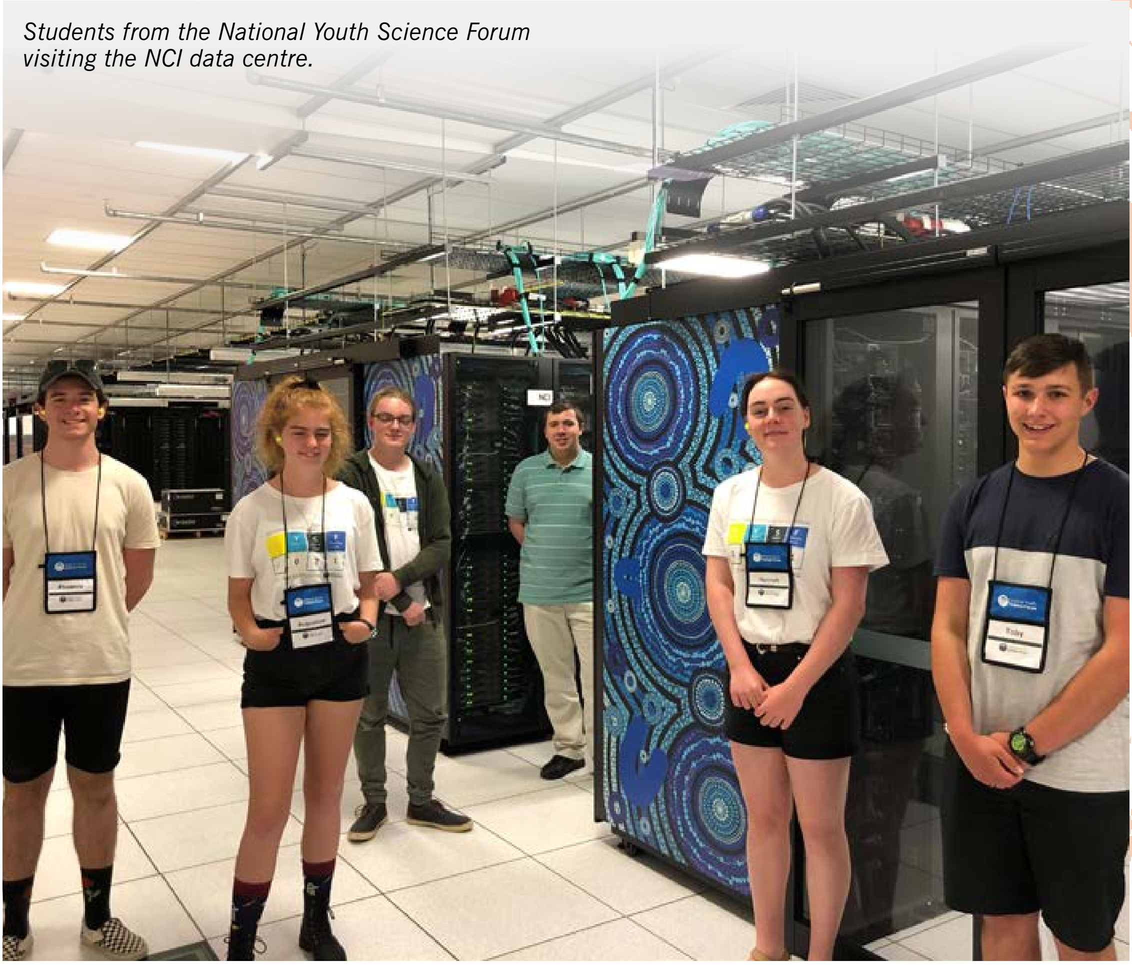
Outreach

NCI continues to offer high school students, university students, parliamentarians and the general public an inside look at the workings of supercomputers and the big data world of modern computational science. NCI reaches hundreds of dedicated learners and future scientists every year through in-person and virtual visits, exhibitions at computational science conferences, stalls at school career events and public engagement on social networks. NCI also runs outreach activities for our user community, finding ways of making connections between users and furthering research collaborations.

This year's National Youth Science Forum students had this to say about their in-person and online visits to NCI: "The NCI was absolutely inspiring. The amount of detail that we got to see in the facility was truly a gift. The presenters were very enthusiastic and executed a wonderful visit. I loved NCI so much! It was amazing getting to see the technologies implemented in the supercomputer and the application of such technologies."



Minister for Superannuation, Financial Services and the Digital Economy Senator the Hon. Jane Hume visits NCI with ANU Vice-Chancellor Professor Brian Schmidt



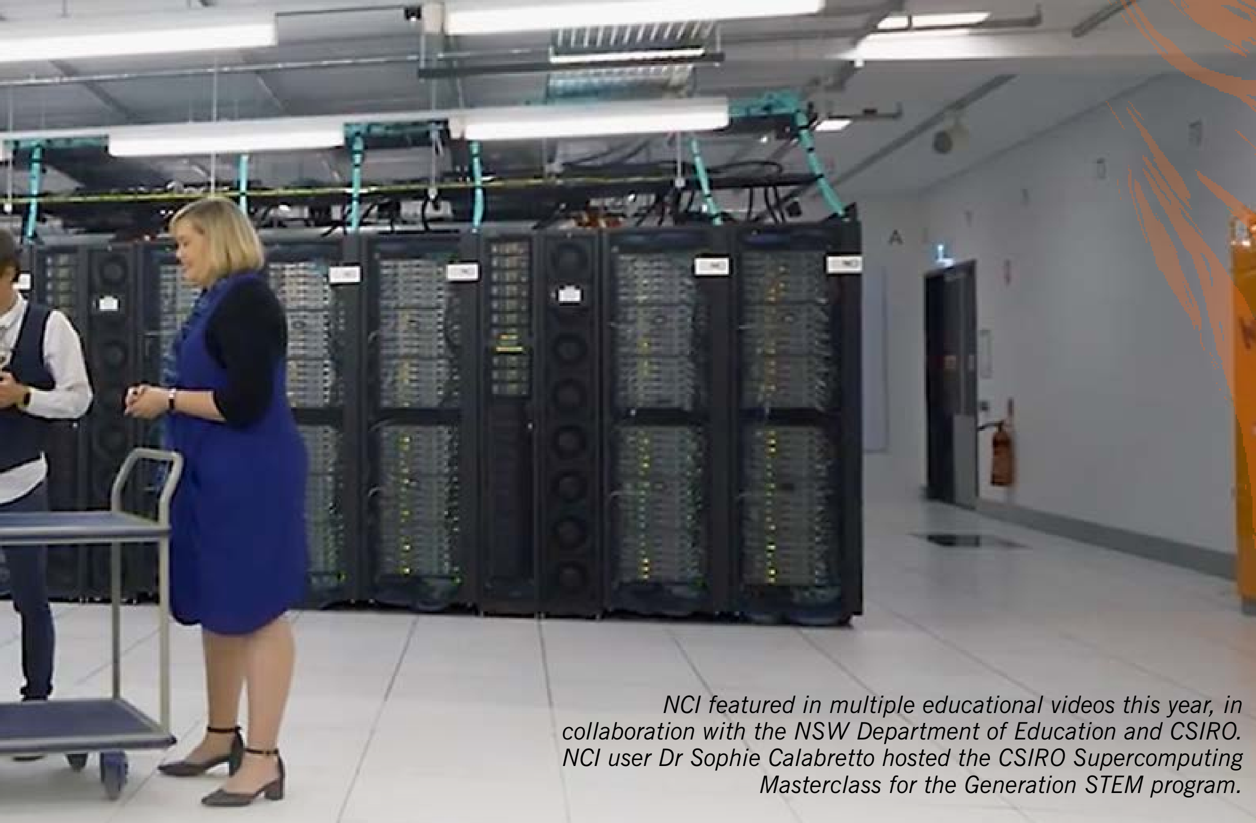
Students from the National Youth Science Forum visiting the NCI data centre.

ACT Chief Minister Andrew Barr (right) visited NCI and met with NCI Director Professor Sean Smith.

In 2020–21, NCI collaborated with CSIRO's Generation STEM and the New South Wales Department of Education's Technology 4 Learning student programs on two video projects. The two educational videos filmed at NCI showcase the role of supercomputing and big data in our everyday lives, illustrate who NCI's users are, and highlight the kinds of science careers available to students who follow that same path. The completed videos and worksheets are now being used to inspire and teach students across New South Wales.

NCI continues to highlight the range of careers and career paths available to budding computational scientists and high-performance computing and data experts. Through our Humans of NCI video interview series, we bring to the fore the multidisciplinary lives and interests of our staff and user community. Additionally, as part of Education Services Australia's MyFuture career profiles, NCI staff with the varied careers of Visualisation Programmer, Systems Administrator and Cloud Architect shared the stories of their education and career paths towards working on Australia's most powerful supercomputer. Their stories add to the existing collection of STEM career options presented to school students, teachers and parents through the MyFutures dedicated online portal.

Associate Professor Hanna Suominen from The Australian National University introduces her Natural Language Processing research in a Humans of NCI video.



NCI featured in multiple educational videos this year, in collaboration with the NSW Department of Education and CSIRO. NCI user Dr Sophie Calabretto hosted the CSIRO Supercomputing Masterclass for the Generation STEM program.

"The NCI was absolutely inspiring. The amount of detail that we got to see in the facility was truly a gift."
– NYSF student, ANU

Financial report

PREAMBLE

NCI is an organisational unit of The Australian National University (ANU). The ANU, as represented by NCI, administers numerous funding contracts that support the operations of NCI. In the interests of providing a comprehensive picture of the NCI operation, a financial report consolidating these funding contracts is presented.

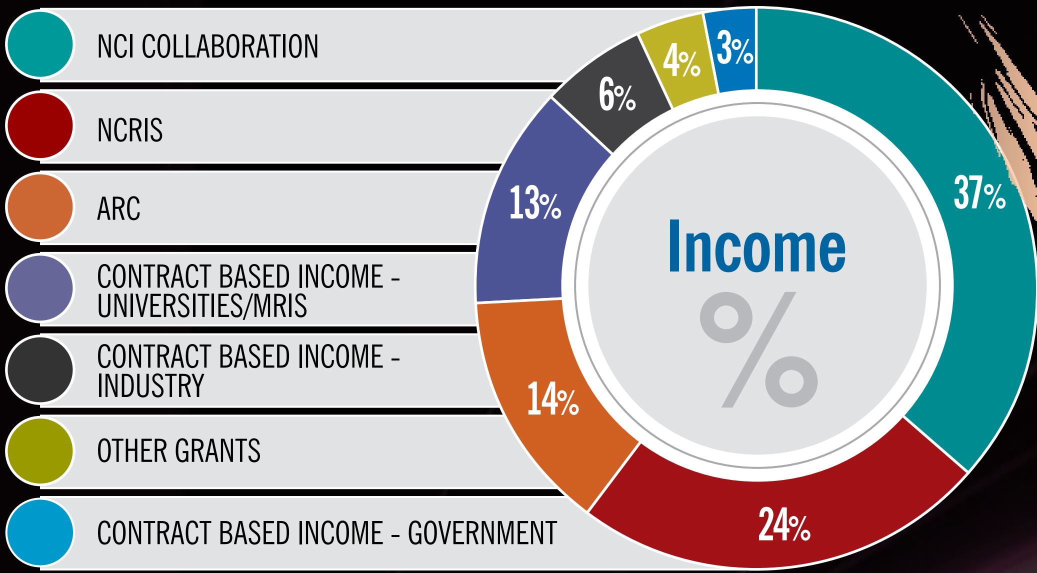
Each funding contract is accounted for in a distinct account within the University ledger, and the University facilitates, and where appropriate, acts on, the NCI Advisory Board's directions and resolutions on NCI matters insofar as they are consistent with the relevant funding contract and not contrary to University statutes and policies.

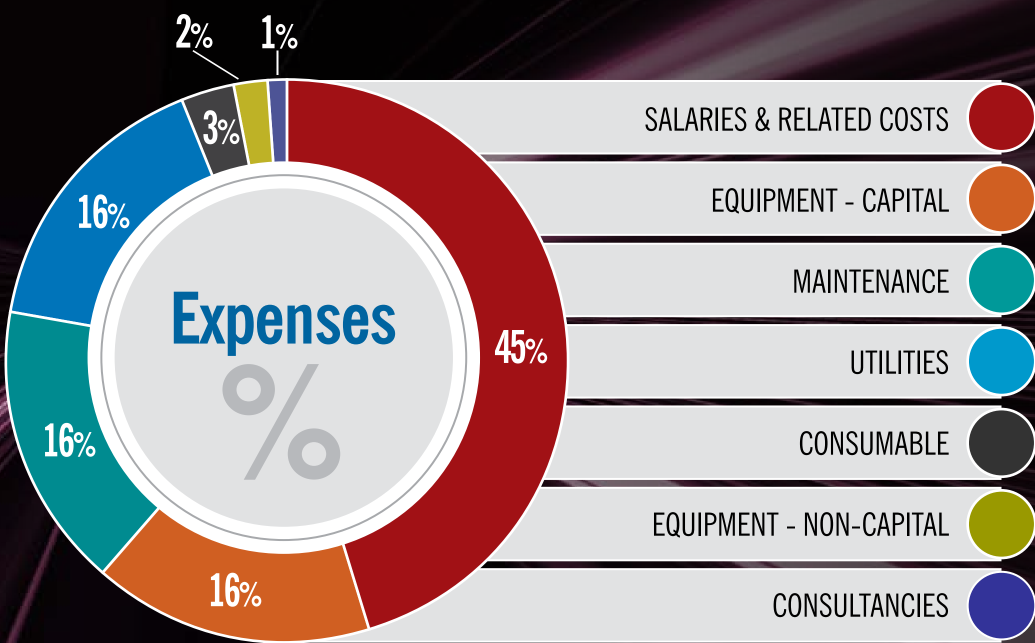
STATEMENT OF INCOME AND EXPENDITURE

For the period 01 July 2020 to 30 June 2021

For the NCI collaboration and associated project accounts

	NCI $
Income	
NCI collaboration income	9,522,596
Other income	16,088,175
Total income	25,610,771
Expenditure	
Salaries and related costs	8,010,200
Equipment - capital	2,923,811
Equipment - non-capital	398,595
Utilities and maintenance	5,703,808
Travel, field and survey expenses	(1,798)
Expendable research materials	414
Consultancies	149,770
Consumables	508,826
Internal purchases	12,117
Other expenses	81,723
Total expenditure	17,787,466
Surplus / (Deficit)	7,823,306


NCI COLLABORATION
NCRIS
ARC
CONTRACT BASED INCOME - UNIVERSITIES/MRIS
CONTRACT BASED INCOME - INDUSTRY
OTHER GRANTS
CONTRACT BASED INCOME - GOVERNMENT
Income
%
37%
24%
14%
13%
6%
4%
3%


SALARIES & RELATED COSTS
EQUIPMENT - CAPITAL
MAINTENANCE
UTILITIES
CONSUMABLE
EQUIPMENT - NON-CAPITAL
CONSULTANCIES
Expenses
%
45%
16%
16%
16%
3%
2%
1%

NCI COLLABORATION INCOME

The NCI Collaboration Agreement enables many of Australia's leading research-intensive universities and science agencies to collectively fund a capability beyond the capacity of any single institution. The Australian Government, through the Department of Education, Skills and Employment (DESE), together with these institutions (including ANU, CSIRO, Bureau of Meteorology, Geoscience Australia, the Australian Research Council, and a range of other research-intensive universities and consortia) fund a significant proportion of NCI's operating costs. Operational funds from DESE are received under the National Collaborative Research Infrastructure Strategy.

A small but growing proportion of NCI Collaboration income comes from the commercial sector which has a need for NCI services to support innovation and their growing research needs in driving the economy. Under the strategic guidance of the NCI executive and NCI Advisory Board, income that has not been expended in the current year is strategically used in subsequent years to meet contractual obligations and expenditure on emerging infrastructure requirements.

NCI administers a number of grants and contracts outside of the NCI Collaboration and NCRIS accounts. These special purpose arrangements fund clearly defined projects, infrastructure and services that provide synergistic benefits not only to the NCI Collaboration, but also to the wider Australian research community and, through them, the nation.

EXPENSES

NCI, as Australia's national research computing service, provides world-class, high-end services to Australia's researchers. In order to do this, NCI invests significant amounts of money in its expert team of staff and high-performance computing and data storage infrastructure – with these two factors driving NCI's expenditure profile.

REVIEW/AUDIT

Each funding contract held by the ANU as represented by NCI has specific financial reporting and auditing requirements, and NCI, in conjunction with the University's Finance and Business Services Division and Corporate Governance and Risk Office, acquits individual project funds in accordance with these requirements.

This consolidated statement has been reviewed by ANU's Finance and Business Services Division. The Chief Financial Officer certifies that:

The statement accurately summarises the financial records of these grants and that these records have been properly maintained so as to accurately record the Income and Expenditure of these grants.

A BIG DATA AND BIG COMPUTING LEADER

The NCI Collaboration

NCI AUSTRALIA

SUPPORTED BY

Australian Government
Department of Education, Skills and Employment

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

MAJOR COLLABORATORS

Australian National University

CSIRO

Australian Government
Bureau of Meteorology

Australian Government
Geoscience Australia

COLLABORATORS

AAPP Australian Antarctic Program Partnership

DEAKIN UNIVERSITY AUSTRALIA Worldly

Garvan Institute of Medical Research

QCIF

RMIT UNIVERSITY

UTS UNIVERSITY OF TECHNOLOGY, SYDNEY

UNIVERSITY OF WOLLONGONG AUSTRALIA

INTERSECT RESEARCH FASTER

OTHER CONTRACTS

Australian Government
Australian Research Council

MACQUARIE University

MONASH University

NSW GOVERNMENT Planning, Industry & Environment

THE UNIVERSITY of ADELAIDE

THE UNIVERSITY OF MELBOURNE

UNSW AUSTRALIA

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

THE UNIVERSITY OF SYDNEY

Victor Chang Cardiac Research Institute

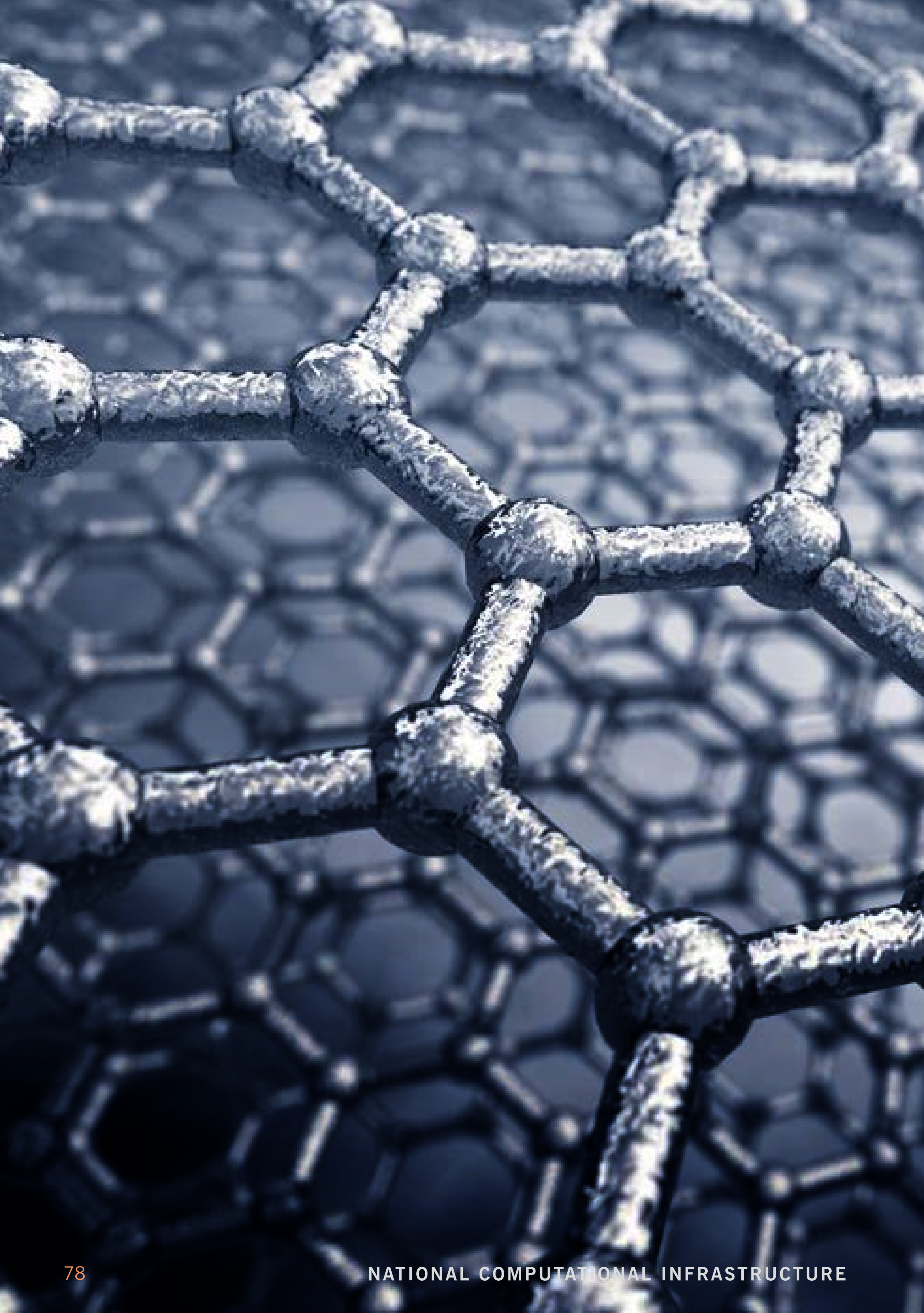
UNIVERSITY of TASMANIA

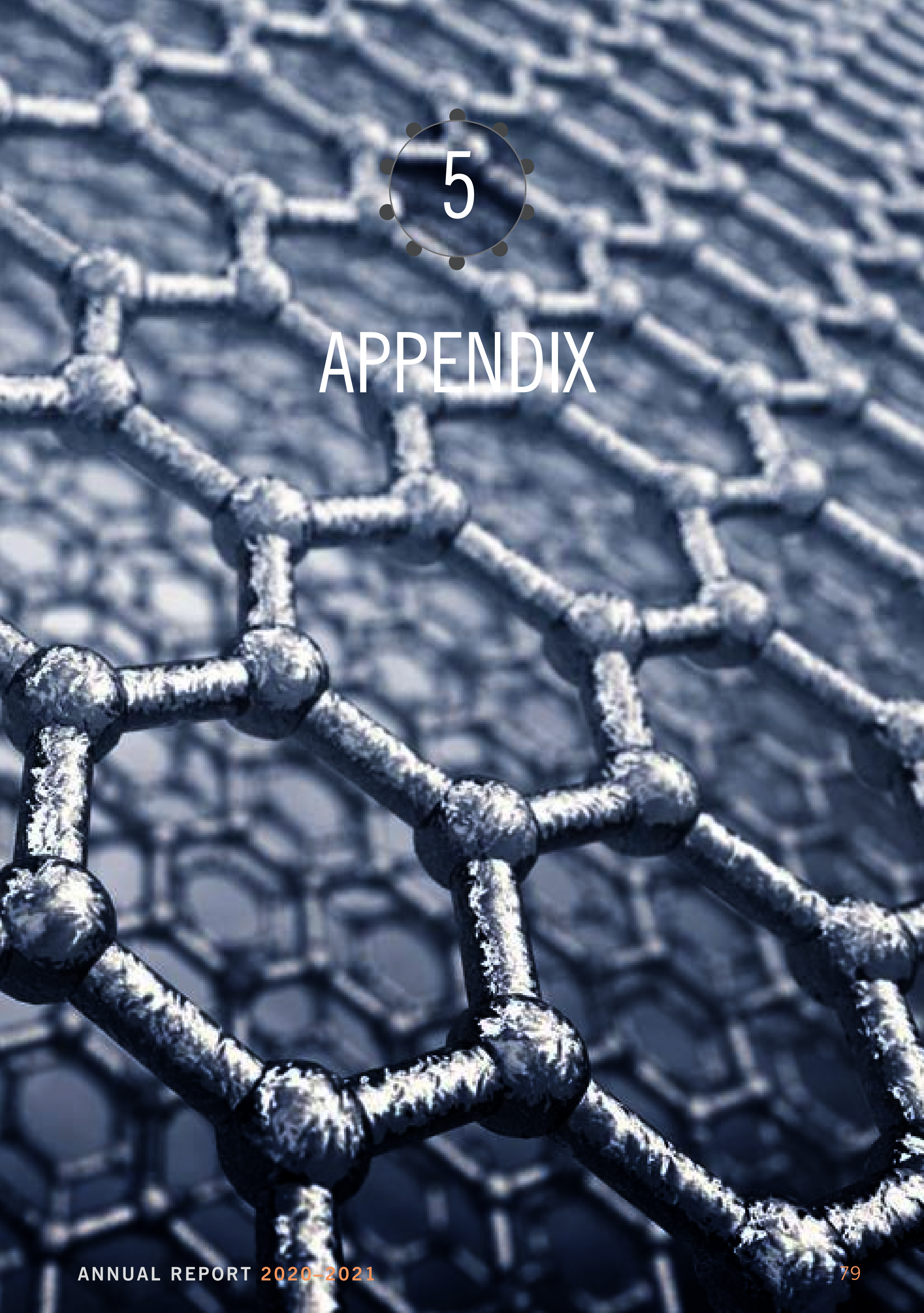
Australian Government
Australian Antarctic Program

Queensland Government

MERIT FLAGSHIPS

Our Vendors





5

APPENDIX

NCI Links to Government Portfolios

Government Department Impacted	Programme/Agency Impacted	Activities/Projects supported by NCI
Education, Skills and Employment	Australian Research Council (ARC)	Dependencies from more than 530 projects funded by ARC's National Competitive Grant Programs (NCGP) underpinning in excess of $606m in research investment
	National Collaborative Research Infrastructure Strategy	Support for services provided by numerous other NCRIS Capabilities (AuScope, TERN, ACCESS-NRI, ALA, AARNet,…)
Industry, Science Energy and Resources	CSIRO	Australian Community Climate and Earth System Simulator (ACCESS)
		Earth Systems and Climate Science (ESCC) Hub of the National Environmental Science Program
		Climate and Weather Science Data-Enhanced Virtual Laboratory
		Marine Virtual Laboratory
		AuScope Virtual Research Environments Geoscience Data Enhanced Virtual Laboratory
		CMIP6 Climate Dataset
	CSIRO and the Australian Institute of Marine Science (AIMS)	eReefs
	Geoscience Australia (GA)	Digital Earth Australia
		Copernicus Data Hub
		National Reference Grid
		Exploring For The Future Initiative
		Water Observations from Space (WOfS)
		Australian Natural Hazards Data Archive
		AuScope Virtual Research Environments Geoscience Data Enhanced Virtual Laboratory
Foreign Affairs and Trade	Policy development for, and by, the tourism sector	eReefs (through CSIRO)
		International Collaboration development (ESGF, ADAC, NeSI)

Government Department Impacted	Programme/Agency Impacted	Activities/Projects supported by NCI
Agriculture, Water and the Environment	National Environmental Science Programme (NESP)	Earth Systems and Climate Science Hub
	Environmental policy development	eReefs
		Coupled Model Intercomparison Project (CMIP)
	Bureau of Meteorology	Australian Community Climate and Earth System Simulator (ACCESS)
		ESCC Hub of the NESP
		Climate and Weather Science Data Enhanced Virtual Laboratory
		Marine Virtual Laboratory
		BARRA Reanalysis
		Bushfire Model Development
	Australian Antarctic Division	Antarctic Climate and Environment CRC (ACE-CRC) ACCESS Southern Ocean and cryosphere models
	Policy development for the agricultural industry and water resources	ACCESS development with BoM/CSIRO
		Development and hosting of Digital Earth Australia Collaboration with GA and CSIRO
		Hosting and curation of WOfS data with GA
	Murray-Darling Basin Authority	Exploitation of Digital Earth Australia data
		Development of water-related Digital Earth Australia capabilities in collaboration with GA
Health	National Health and Medical Research Council	Dependencies from more than 64 NHMRC funded projects and fellowships
		Australian Genomics Health Alliance (AGHA)
		Support for COVID-19 Research projects
Defence	Australian Geospatial Intelligence Organisation (AGIO)	Onshore topographic data and products provided by GA
	Australian Hydrographic Service (AHS)	Raw and processed bathymetric data collections provided by GA
Infrastructure, Transport, Regional Development and Communications	Australian Marine Safety Authority (AMSA)	Managing risks to marine vessels in Australian waters (undertaken with consultant DHI)
	Aviation Programs	Development of specialist weather reporting products for the aviation industry (with BoM)
	Transport Infrastructure Programs	National Reference Grid

Outreach Activities

Event	Date
NCI Pawsey COVID-19 Australian Response Showcase	31 July 2020
National Science Week NCI Virtual Tour	19 August 2020
ANU Digital Open Week	22-28 August 2020
HPC-AI Advisory Council Australasia Conference 2020	1-2 September 2020
eResearch Australasia	19-23 October 2020
WHPC Launch Morning Tea	11 November 2020
NCI Supercomputer Virtual Tour	12 November 2020
Super Computing 2020 Conference (Virtual)	17-19 November 2020
National Youth Science Forum Digital Tour	13 January 2021
National Youth Science Forum On-Site Visit	14 January 2021
Australian Meteorological and Oceanographic Society Data Workshop	5 February 2021
WHPC Birds of a Feather Session – eResearchNZ	10 February 2021
Senator the Hon. Jane Hume, Minister for the Digital Economy Visit	23 February 2021
The Hon. Nola Marino MP, Assistant Minister for Regional Development and Territories Meeting	25 February 2021
Ms Madeleine King MP Visit	26 February 2021
Supercomputing Asia (Virtual)	2-4 March 2021
ACT Chief Minister Mr Andrew Barr MLA Visit	22 March 2021
ACT Health Minister Ms Rachel Stephen-Smith MLA Visit	26 March 2021
Mr Peter Cain ACT MLA Visit	8 April 2021
Mr Julian Hill MP Visit	8 April 2021
COSMOS Supercomputer Public Briefing	15 April 2021
Department of Education, Skills and Employment Visit	19 April 2021
Hawker College Women and Girls in Science and Engineering Day Exhibition	6 May 2021
Geoscience Australia Graduates Visit	13 May 2021
AI and ML in Materials Design and Discovery Showcase	13 May 2021
Data Science Week Supercomputer Virtual Tour	14 May 2021
ANU College of Engineering and Computer Science Visit	17 May 2021

Training Activities

Training	Date
GPU Bootcamp	August 2020
CUDA Bootcamp	October 2020
eResearch and Data Skills Summit	October 2020
VTune Seminar	November 2020
AMOS 2021 (NCI Data Workshop)	February 2021
Sydney Informatics Hub training series (Introduction to Gadi)	February 2021
Data Science Week (Introduction to HPC, Big Data, Scientific Visualisation)	May 2021
AI/ML in Materials Design and Discovery Showcase	May 2021
ResBaz Perth (Introduction to HPC, Big Data, Scientific Visualisation)	June 2021
Introduction to Gadi – Parts 1 and 2 (monthly beginner webinars)	2021 (ongoing)
NCI Presents: TechTake (monthly)	2021 (ongoing)
Support Session (office hours, monthly)	2021 (ongoing)
/ML4AU ML showcasing events	2021 (multiple)

APPENDIX