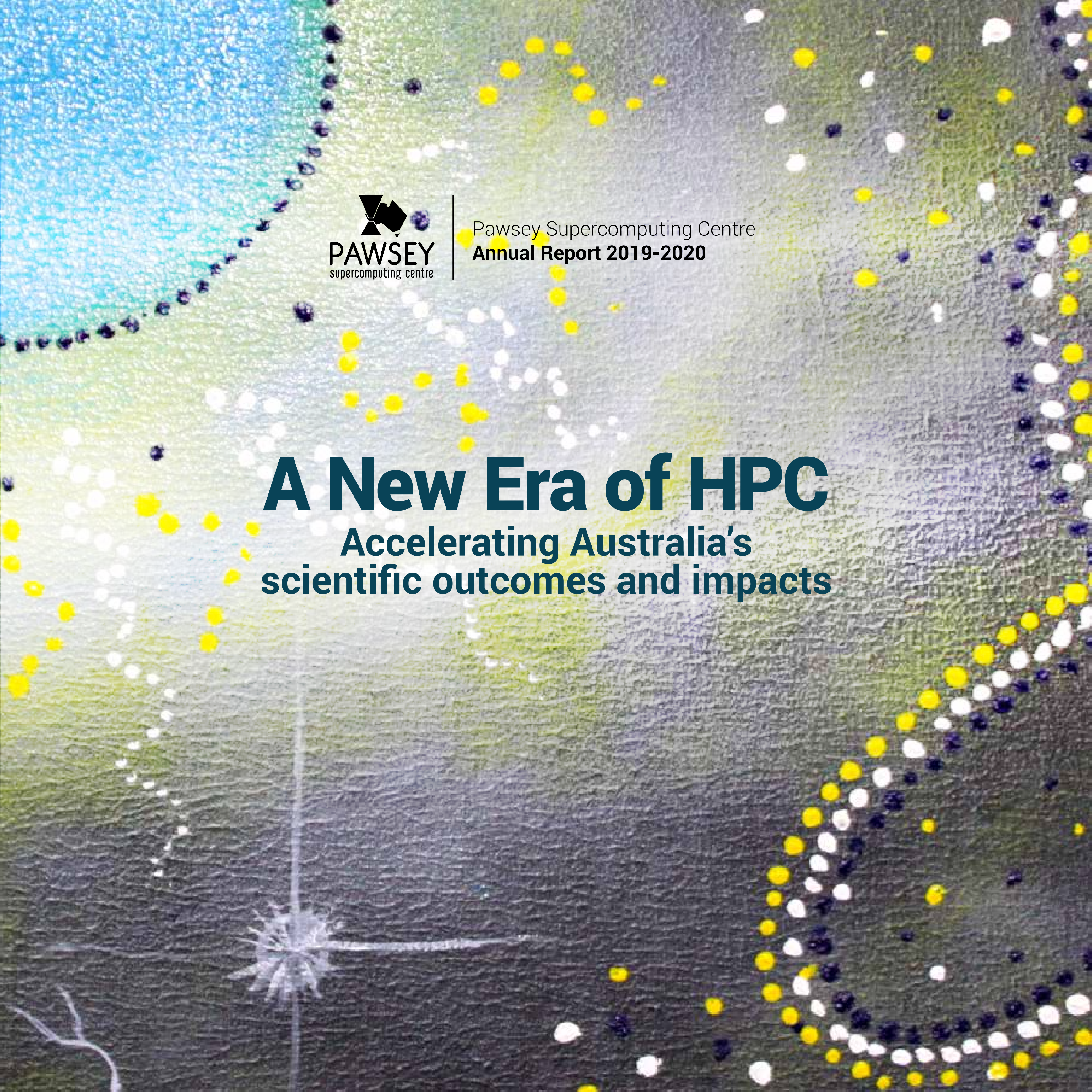


PAWSEY
supercomputing centre

Pawsey Supercomputing Centre
Annual Report 2019-2020

A New Era of HPC

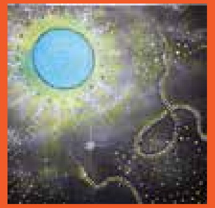
Accelerating Australia's scientific outcomes and impacts

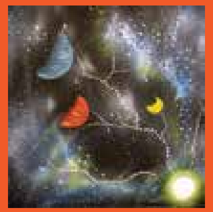
Artists' Acknowledgment

The Pawsey Supercomputing Centre selected these images to adorn its supercomputers in homage to our close connection to the north-west of Western Australia. The two areas of science that Pawsey supports most closely have been energy and resources, and radio astronomy, and we wanted to choose art that thematically reflected both the ground below and the sky above. In addition, the Centre has a close relationship with the Murchison Radio Astronomy through our handling of its radio astronomy data, and we wanted to work with indigenous artists who had a close personal relationship with that region. These two pieces capture the unique spirit of the north-west and pay homage to the knowledge held by the earliest Australians about the stars in their sky.

Margaret Whitehurst
SKA Satellites on the Murchison – 120 x 120cm
Yamaji Art

Margaret was born on Billabalong Station in the Murchison and is a Wajarri woman. She is the second of seven children. Her childhood was spent swimming and fishing in the Murchison River and she went to school at the Tardun Mission School. Her inspiration comes from the works of other Aboriginal artists and her Auntie Olive Boddington. She came to painting later in life and has only been working as an artist since joining Wila Gutharra Art Talk five years ago. Margaret is the mother of seven children herself and is an industrious and prolific artist. Her current works are reflections of Yamaji country. Margaret now lives in Geraldton Western Australia.





Jesse Pickett
1. *Rainbow Serpent and the Moon* – 56 x 55.5cm
2. *The Sun* – 56 x 55.5cm
Yamaji Art

Jesse was born in Quirading, East of Perth in Western Australia. He has been painting from an early age, taught by his father with both sides of his family painting and drawing. "I create as a way to express culture," Jesse said. His paintings, '*Rainbow Serpent and the Moon*' and '*The Sun*' are described by him as follows: "The Rainbow Serpent and the Moon are connected together due to the Wagu (Rainbow Serpent) creating water and the moon moving it to create life, while the Sun is giving life to every living thing on earth."

TABLE OF CONTENTS

MAKING TOMORROW HAPPEN,TODAY

AMAZING OUTCOMES

VOICES OF SCIENCE

A NEW ERA OF RESEARCH

A WORLD OF DIFFERENCE

FINANCIALS

A national research infrastructure and global player in science and innovation

Making Tomorrow Happen, Today

CHAIRMAN FOREWORD

Image courtesy of WA Business News.

> *Through removing supercomputing capacity constraints, local opportunities in health and medical research are also set to expand, and Western Australia's strong geoscience capabilities can only grow.*

WELCOME TO THE PAWSEY SUPERCOMPUTING CENTRE 2019–20 ANNUAL REPORT

This year Pawsey has taken its partnerships to new levels, driven by the opportunities that the Capital Refresh project is creating. The design and procurement of our new supercomputing system are now well advanced, and the massive increase in computational capacity Pawsey is creating represents a huge opportunity for research across Australia.

Local research through the Western Australian universities that support the Pawsey Centre has every opportunity to expand, consolidating capabilities in many sectors including resources and energy and providing opportunities for strong Western Australian industries to drive innovation and growth even further.

Research enabled through Pawsey's infrastructure upgrade is also underpinning the ever-expanding astronomy sector in Western Australia, and the State's continuing movement into the space sector through the establishment of the Australian Space Data Analytics Facility. This new capacity at Pawsey will unlock the value of space observational data and create a range of opportunities for Australian businesses from agriculture to emergency services and others.

Through removing supercomputing capacity constraints, local opportunities in health and medical research are also set to expand, and Western Australia's strong geoscience capabilities can only grow.

The challenge for both the university sector and industry is to engage and embrace the opportunities this Capital Refresh offers. This year has seen many sectors withdraw to their core businesses as they seek to weather the challenges brought by a global pandemic, but now is also the time to seize the opportunities that have come out of global change. For those looking to refocus on new activities, new ways of working, or new global problems, Pawsey is building the opportunity to do so.

The Board would like to acknowledge the foresight of the Commonwealth Government in maintaining Australia's High Performance Computing (HPC) capabilities through Pawsey and the National Computational Infrastructure (NCI) at an international standard, underpinning our ability to work together as equal partners with the global community on global problems, such as mitigating the effects of the COVID-19 pandemic.

Equally, the continuing support of the Western Australian Government has been critical to the ongoing operation of the Pawsey Supercomputing Centre, enabling it to meet both local and global research challenges. The State Government's support of Pawsey through its five-year funding agreements are essential to maintain our critical expertise. Without this funding, Pawsey could not continue.

Pawsey's Executive Director Mark Stickells and his team have brought significant stability and maturity to the management of Pawsey's operations this year. The Capital Refresh has been a business-changing undertaking that has been executed extremely well. From working with a diverse range of users to design systems that meet their research needs, to planning smooth transitions to new infrastructure, the team is making working with Pawsey easier and more productive than it has ever been before. This is seen in the strong working relationships with CSIRO and our joint venture partners.

I speak for the Board in congratulating Pawsey's team in navigating a complex year. We are extraordinarily impressed and delighted with the outcomes, evidence of which is throughout this report.

I would also like to thank the Board Directors for their support throughout the year. Next year is set to be a step change in success for our entire community.

John Langoulant AO
Chairman of the Board

EXECUTIVE DIRECTOR'S REPORT

For the Pawsey Supercomputing Centre, our staff and the research community we support, the past year has been one of significant challenges and achievements. We restructured our organisation to lead a new five-year strategic plan and we upgraded several key systems in addition to a major tender for our new Pawsey Supercomputing System. Pawsey's staff demonstrated outstanding commitment in supporting each other and our operations through what can only be described as an extraordinary year. It was a year in which we continued to deliver our services, now as critical as ever, and maintained the wellbeing of our own community.

The Pawsey team delivered an upgraded Nimbus cloud infrastructure, additional high-speed storage for both the Murchison Widefield Array (MWA) and the Australian Square Kilometre Array Pathfinder (ASKAP) radio astronomy telescopes and a new MWA compute cluster which also released additional compute power for ASKAP. Design and selection of the system architecture for Pawsey's new Supercomputing System was completed, tenders evaluated and the foundations have been laid for Pawsey's next-generation infrastructure to be announced later this year. Maintaining a state-of the-art research facility like Pawsey only happens with the support of both the Commonwealth and the Western Australian Governments and our partner institutions, and their funding commitments to both our physical facilities and continued operation.

Central to Pawsey's strategy is providing value to our customers, and we continue to invest in supercomputing infrastructure, and in our partnerships and research communities so they can take full advantage of our next generation of services.

Emerging research domains and areas of strategic importance to Pawsey include bioinformatics and space science. In partnership with the WA Data Science Innovation Hub, Pawsey received funding from the Australian Space Agency to establish the Australian Space Data Analysis Facility. The Facility will support researchers and small business with space data analytics expertise to make space observational data more readily available for the development of new products and services, and grow Australia's space industry sector.

As we grow and diversify our research community, we're also growing and diversifying our own workforce. Through summer internships, Pawsey HPC Research Fellowships, and relentlessly promoting careers in data intensive sciences, we are steadily becoming more representative of the community we serve. When I first joined Pawsey two years ago, we had more men named Mark in our workforce than we had women and I'm very pleased to say that is no longer the case. Our commitment to diversity and inclusion is supported by the stories in this report of the performance and achievement of our staff and our community. Genuine improvements in diversity and inclusion ensure Pawsey is not just a better place for all to work, but a better performing business.

Pawsey has been responsive to the COVID-19 pandemic and worked with the NCI to provide accelerated access to supercomputing resources for eight Australian research groups working to understand and stop the spread of the SARS-CoV-2 virus. Their work is summarised alongside other examples of Pawsey-supported research later in this report. We are also continuing to work with the international research and supercomputing community to facilitate the sharing and analysis of information to block the action of SARS-CoV-2 and treat COVID-19.

We've also reassessed how to deliver our services in a COVID-safe way, with widespread adoption of online and virtual tools for engagement, access, training and troubleshooting. Our staff have been incredibly proactive and innovative in expanding our ways of working this year, and some changes will become business-as-usual in the future as they are demonstrably improving our operations, engagement and services.

I'd like to wholeheartedly thank the Pawsey team and acknowledge their efforts through what has been a year of many challenges. Equally, our Board has provided unwavering support and direction through this period. As we together navigate this global period of uncertainty, upheaval, risk and loss, I am confident in Pawsey's future and its contribution to State, national and international efforts to support our social and economic recovery.

I commend this annual report, and throughout the report acknowledge our stakeholders and the research communities that benefit from Pawsey's expertise, infrastructure and services. Together we will continue to tackle the world's 'wicked' problems.

Mark Stickells
Executive Director

HIGHLIGHTS

PEOPLE

OVER 4000 HOURS
of training were delivered to Pawsey users and staff in joint training sessions.

3000 HOURS
of training was delivered to Pawsey users, with 2,000 of those hours delivered online between April and June, after COVID-19 restrictions were imposed. The shift online enabled Pawsey to reach a much broader and geographically-dispersed user group of over 1,550 researchers.

8 RESEARCH GROUPS
were granted prioritised and expedited access to computation and data resources as part of the Australian HPC COVID-19 accelerated access initiatives, a joint program delivered by NCI and Pawsey.

821 PEOPLE
participated in Pawsey events, including the third HPC-Artificial Intelligence (AI) Advisory Council Conference Perth, two virtual and two face-to-face Pawsey Fridays, and STEM activities.

37
people trained in the art of communicating science by Perth TED Talk CEO, Gavin Buckley and Science and Technology Australia.

THE HON KAREN ANDREWS
Federal Minister for Industry, Science and Technology, visited the Centre.

THE HON DAVE KELLY MLA
Minister for Water; Forestry; Innovation and ICT; Science; Youth visited the Centre to launch the Index of Marine Surveys for Assessments (IMSA) portal.

AT SC19, PAWSEY WAS AWARDED
the HPCWire Editor's Choice Award as the Best Use of HPC in Physical Sciences for a project undertaken with Professor Brett Harris at Curtin University to map and model Perth's aquifers.

SYSTEMS

1.8X OVERSUBSCRIPTION
of Pawsey supercomputers in 2020

3X MORE STORAGE
and 3x more performance in high-speed storage were procured for ASKAP and MWA researchers.

10X MORE STORAGE
and 3x more memory is now available in Pawsey's cloud for researchers.

3 MORE
procurements are underway, including the tender for the new supercomputer to replace the Magnus and Galaxy supercomputers and Long-Term Storage.

308 TERAFLOPS
of graphics processing Unit (GPU) power were delivered as part of Topaz, a new commodity Linux cluster which is an extension of Zeus.

75 DATA COLLECTIONS
of curated research were stored and managed at the Centre.

COLLABORATIONS

PAWSEY SUPERCOMPUTING CENTRE AND QUANTUM BRILLIANCE
are working together to develop Australia's first quantum-supercomputing hub for innovation– an Australian first commercial partnership between a supercomputing centre and quantum computing provider.

WA'S FIRST
Honours-level HPC course, developed by a team of computational physicists and chemists and supported by Pawsey staff, was delivered at Curtin University.

SPACE DATA ANALYSIS @PAWSEY
in partnership with the WA Data Science Innovation Hub, were awarded more than $2 million in Federal and State Government funding to establish a new national space data analysis facility in Western Australia.

ABOUT PAWSEY

THE PAST

THE PATH TO A WORLD-CLASS FACILITY

The Pawsey Supercomputing Centre, named in honour of the Australian radio astronomer, Dr Joseph Pawsey, is a federally funded purpose-built facility in Perth, Western Australia, constructed in 2012. It is a national HPC facility, delivering cutting edge computational support and services to key scientific areas such as radio astronomy, bioinformatics, energy and resources.

The Centre's compute capacity has grown by an order of magnitude since its inception, and it now provides world-class expertise and infrastructure in supercomputing, cloud services, and data-intensive analysis, storage and visualisation to over 1,600 researchers every year.

COLLABORATION

Pawsey is an unincorporated joint venture between CSIRO, Curtin University, Murdoch University, Edith Cowan University and The University of Western Australia. The Pawsey Supercomputing Centre is also funded by the State and Federal governments.

One of two Tier 1 HPC facilities within Australia's national research infrastructure network, Pawsey partners with the NCI in Canberra to unlock impactful scientific knowledge through research, critical infrastructure, expert staff, sector know-how, and customer-centric problem-solving.

THE PRESENT

REAL WORLD IMPACT

Through supporting Australian researchers and their collaborators around the globe, Pawsey enables high-impact research for the benefit of society across domains including energy and resources, engineering, bioinformatics, health sciences and agriculture. Pawsey-enabled research is tracking the spread and evolution of disease, designing drugs for pain management, building better batteries and improving our agricultural crops in a drying climate.

STAYING AT THE FOREFRONT OF HPC

Pawsey is halfway through a $70 million Capital Refresh project to secure its next generation of supercomputers, data and supporting infrastructure. This is keeping Pawsey at the forefront of global advances in supercomputing technology, and providing significantly increased computing power and speed, expanded data storage capacity, and more efficient high-performance computing services for Pawsey users.

OUT OF THIS WORLD

The Centre is a unique supercomputing facility in the world, directly ingesting and processing continuous real-time data from world-leading scientific research instruments. Pawsey supports the operation of two Square Kilometre Array (SKA) precursors, the MWA and ASKAP, and has been processing and storing their data since they began operations. These radio telescopes are making discoveries about the nature of our Universe while contributing to the design and development of the SKA, the world's largest mega-science project. Pawsey is equally contributing to the design, development and testing of the data processing and storage components of the SKA.

Australian Space Agency activities are now also being supported at Pawsey. Through a partnership with the WA Data Science Innovation Hub, and the support of the Commonwealth and the Western Australian Government, Pawsey is establishing the Australian Space Data Analysis Facility to support researchers and small to medium enterprises with space data analytics and expertise. This facility will support the growth of the Australian space industry, and unlock the value of space observational data for industries such as agriculture, mining, emergency services and maritime surveillance.

INNOVATION AND NEW TECHNOLOGIES

To remain at the forefront of digital innovation and potential disruption in computing technologies, Pawsey is collaborating with Quantum Brilliance, a leading Australian quantum computing company, to contribute to developing Australia's first quantum-supercomputing hub for innovation. This collaboration will see quantum expertise developed among Pawsey staff to then install and provide access to a quantum emulator at Pawsey, an important step in translating this technology into real quantum computing applications.

TACKLING COVID-19

Pawsey Supercomputing Centre and the NCI have again joined forces, offering computation and data resources to support the national and international research community to acquire, analyse and share information supporting COVID-19 research. The research undertaken through this special allocation is contributing to the global effort to understand and overcome the disease at the heart of our current global pandemic, from tracing the evolution and mutation of the virus to developing diagnostic tests and therapeutic drug targets.

COVID-19 has also forced Pawsey to re-evaluate how it delivers services, with a rapid transition to increased training, engagement and support for the research community through online and virtual tools. Pawsey's community is now united through digital technology as never before.

THE FUTURE

EXPONENTIAL GROWTH IN COMPUTE POWER

As a forward-looking, future-focused organisation, Pawsey is continuing to grow its supercomputing facilities and expertise, its research communities, and the next generation of HPC and data scientists for Australia. As the Capital Refresh progresses, Pawsey's main supercomputer procurement is moving forward, and expected to provide exponential growth in power and services to a wider research community. Innovations in computing technology continue to be explored to maintain our community's research advantage.

TOWARDS A NEW ERA

Pawsey is investing in staff expertise to support Australia's research community's transition to newer and more powerful supercomputing systems, and helping reduce entry barriers for researchers who need supercomputing resources to accelerate their discoveries.

Working together with the NCI, Pawsey is creating a cohesive supercomputing service for all Australian researchers. By driving innovation and accelerating discoveries across a multitude of different fields, the Centre is enabling Australian researchers and their international collaborators to tackle the world's 'wicked' problems.

PAWSEY
supercomputing centre

OUR VISION

Enable science and accelerate discovery

OUR PURPOSE

Unlock impactful scientific knowledge through research, critical infrastructure, expert staff, sector know-how, and customer-centric problem-solving

OUR VALUES

- Innovative
- Collaborative
- Impactful
- Responsive
- Honest

WHO WE SERVE

The Pawsey Supercomputing Research Centre was originally established to provide high-performance computing, data analytics and storage capability for the **Square Kilometre Array**.

We now serve many national and international research efforts.

Current research includes **climate change, health, medical, mining, resource minerals processing, agriculture systems and products, geosciences, renewable energy, engineering** and many more.

Demand for supercomputing services continues to rise from existing and new research disciplines and applications, including **emerging domains of space and bioinformatics**.

OUR STRATEGIC PILLARS & GOALS

RESEARCH & TECHNOLOGY

Prepare for and apply contemporary and future technologies that deliver peak performance for research insight and understanding

1. Identify and secure the best hardware and software options to support discovery, insights and investigation.
2. Pawsey's technology is a tool for flexible, expert and valued customer service.

PEOPLE

Attract, diversify, develop, and retain the best talent

1. Attract diverse talent who bring new perspectives from across the globe to challenge and enrich Pawsey's work.
2. Commit to staff development and retention, reflected in actions and opportunities that support career development, skills development and work satisfaction.
3. Lead by example with collaborative, can-do and solutions-focus attitude
4. Drive Pawsey's Culture by bringing our values to life.

INTEGRITY

Secure, trusted and safe services are protected by contemporary systems, policies, and highly skilled and security aware staff

1. Customers have confidence in and an understanding of the security of their data while it is at the Centre.
2. Staff are highly security aware and able to share security protocols with customers.

CUSTOMERS

Improve collaborative, responsive services to solve challenges, supported by strong communications skills and domain expertise

1. Continually improve internal and external meaningful customer engagement and responsiveness
2. Demonstrate elevated focus on customer service that validates a growing commercial mindset.
3. Apply strong communications skills with all customers.

SKILLS

Increase data and supercomputing literacy by delivering skills and knowledge frameworks that guide staff and diverse research groups

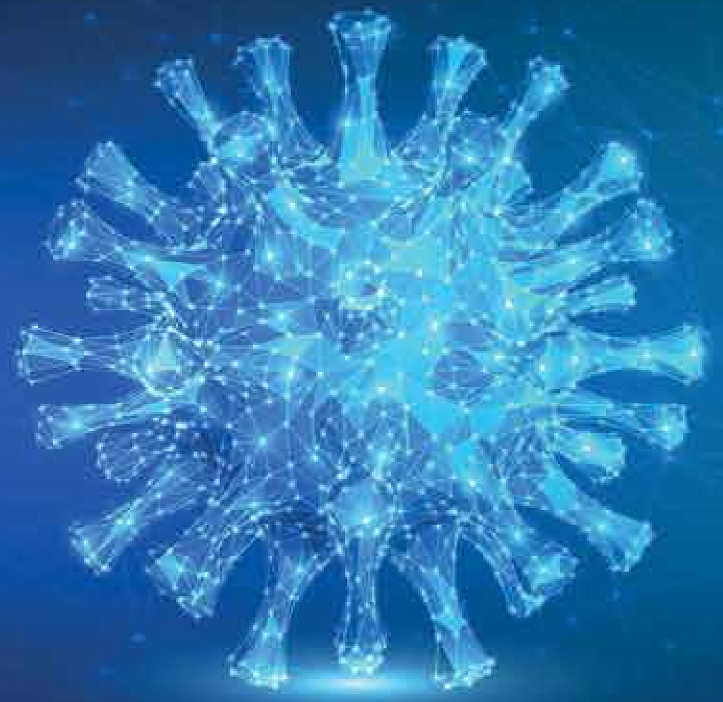
1. Establish a learning program that enables Pawsey staff to excel in their roles.
2. Build data, visualisation, cloud computing and supercomputing skills across a diverse range of national researchers.
3. Contribute strategically to a researcher pipeline and a data-informed Australia through partnered programs, activities, and events.


Enabling life-changing discoveries
Amazing Outcomes

COVID-19 ACCELERATED ACCESS INITIATIVES

In early April 2020, the Pawsey Supercomputing Centre and the National Computational Infrastructure (NCI) joined forces to offer computation and data resources to support the national and international research community to acquire, process, analyse, store and share data supporting COVID-19 research.

Through this initiative, NCI is supporting three targeted projects with more than 40 million units of compute time, and Pawsey is providing access for five projects to over 1,100 cores on the newly deployed Nimbus cloud and the Topaz compute cluster. The research supported through this special allocation is contributing to the global effort to understand and overcome the disease at the heart of the current global pandemic.



SPEEDING UP SARS-COV-2 GENOMICS ANALYSIS

Galaxy Australia (usegalaxy.org.au) is a web-accessible platform for computational biological research. It hosts almost a thousand tools supporting computational analysis of biological data against more than 200 reference data sets, and has access to over 7,000 tools from a global repository. Tools can be chained together to execute on remote computers to do complex series of analyses. Importantly, all aspects of the platform are peer-reviewable and transparent, allowing researchers to conduct accessible and reproducible biological research.

As part of the global research response to COVID-19, in conjunction with Galaxy Europe and Galaxy Main in America, Galaxy Australia has shared the workflows, tools, reference files and configuration styles to support rapid genomic analysis of the SARS-CoV-2 virus. Workflows specifically for analysing the genome of the virus were added to the platform in February 2020. These allow raw genetic data to be turned into some understanding of the evolution and variation of the SARS-CoV-2 strains and trace them as they spread through populations.

While Galaxy Australia is hosted at the Queensland Cyber Infrastructure Foundation (QCIF) Facility for Bioinformatics in Brisbane, the platform can send computational analytical jobs to a number of remote computing nodes around the country. Using Pawsey's special allocation, Galaxy Australia deployed a new node on Nimbus, setting up six virtual machines on the cloud infrastructure prioritised for COVID-19 research in just two business days.

Any analysis run through Galaxy Australia's platform identified as a COVID-19 workflow is now sent to this new node at Pawsey so computational time can be directly allocated without having to compete with other ongoing biological research on the other nodes, minimising the time for results to be returned to the researcher.

With Pawsey's support, Galaxy Australia has enabled the analysis of hundreds of SARS-CoV-2 genomes, amounting to over 9,000 Galaxy-mediated tool executions on Nimbus, since late February. This open and transparent analytics and data is enabling COVID-19 research around Australia.

TRACING THE EVOLUTION OF SARS-COV-2

The Severe Acute Respiratory Syndrome (SARS)-related coronavirus emerged in 2002, and is widely regarded as being in the same evolutionary group as the new SARS-CoV-2, sharing a common beta-coronavirus ancestor found in bats. Although the gene sequences of all beta-coronaviruses are known to be highly similar, the evolutionary grouping of SARS-CoV and SARS-CoV-2 was determined by comparing only three highly-conserved proteins of the 28 proteins coded by these single strand RNA viruses.

Associate Professor Michael Wise from the University of Western Australia has used a Pawsey special allocation to compare as many proteins as possible from SARS-CoV, SARS-CoV-2 and related bat-betacoronaviruses to more closely examine their evolutionary history. The comparison covered 25 proteins expressed by 58 strains of the new SARS-CoV-2, two old strains of SARS-CoV, 12 bat beta-coronaviruses and two related beta-coronaviruses found in civets, and looked for the emergence of variations at the protein level mapped by the dates of collection of the various virus strains.

The evolutionary relationships created from this phylogenetic analysis demonstrated that the protein expression of all of these viruses were highly conserved over time, with a resistance to change over evolutionary time similar to measles. This relative stability of protein output suggests that multiple possible targets exist for vaccines and other therapeutics, and a concerted vaccination campaign could potentially eliminate both SARS-CoV-2 and other SARS-CoV strains, much as has been done for the polio virus.

The most likely evolutionary model has a chain of bat beta-coronavirus strains linking SARS-CoV and SARS-CoV-2 in a single-species model, with a largely constant population circulating in bats since the 1980s and occasionally transferring to humans.

FAST, SENSITIVE DIAGNOSTIC TEST FOR SARS-COV-2

Scientists from DNA Zoo Australia and the University of Western Australia's Faculty of Science are part of a global team that have developed a fast, sensitive diagnostic test for SARS-CoV-2 using a Pawsey special allocation.

Existing Real-Time Polymerase Chain Reaction (RT-PCR) tests target and detect only around seven per cent of the SARS-CoV-2 genome, and can report a negative result if too little of the virus is present. Their new Pathogen-Oriented Low-cost Assembly and Resequencing (POLAR) test amplifies and sequences the entire SARS-CoV-2 genome. It can reliably detect 84 genome-equivalents per millilitre, more sensitive than nearly all diagnostic tests for SARS-CoV-2 currently approved by the US Food and Drug Administration (FDA).

The automated analysis pipeline developed at Pawsey provides a one-click diagnostic report containing the test result (positive or negative), and the specific SARS-CoV-2 genome identified. This genetic information can be used to identify and track how specific strains of the virus spread and mutate, aiding in tracing transmission in community and healthcare settings. The ability to rapidly generate detailed viral genome data will also underpin the creation of new diagnostic tests, vaccines and drug targets.

Associate Professor Parwinder Kaur, the Australian project lead for POLAR, said because the test was able to detect such low viral concentrations compared to other diagnostic tests, it could even be used to monitor COVID-19 through wastewater treatment plants to track its spread through communities.

The research team have demonstrated that using POLAR, a single person can process 192 patient samples in an 8-hour workday, enabling a 24-hour turnaround time for samples, even including sequencing and data analysis.

This preliminary SARS-CoV-2 testing approach has been published on bioRxiv, and the open-source analysis pipeline including documentation and test set is publicly available at https://github.com/aidenlab/Polar. The analysis pipeline developed at Pawsey can be run efficiently on a wide range of high-performance computing platforms.

POLAR is currently being used for COVID-19 research, and is consistent with US FDA guidelines for diagnostic testing for SARS-CoV-2.

MOLECULAR MODELLING OF COVID-19 THERAPEUTIC DRUG TARGETS

The SARS-CoV-2 virus, once it has infected a cell, needs to replicate to spread further. It makes copies of its genetic material and also makes long proteins that get cut down to size. These copies are repackaged using the cut proteins into new virions that can be released to infect other cells. The main protease that cuts proteins down to size is an obvious target for drug design and if it can be inhibited, the viral life cycle can be stopped.

Dr Tom Karagiannis heads the Epigenomic Medicine Laboratory at Monash University. With Dr Andrew Hung (Biomolecular Simulation Group), they have accessed the special allocation to use the Topaz advanced supercomputing cluster at Pawsey to model the main SARS-CoV-2 protease (around 133,000 atoms), map its active binding sites, and simulate binding with potential inhibitors including existing antiviral drugs. Initial studies were focussed on characterising the active site and other potential binding sites in the main protease. The interaction of the protease with three hundred potential inhibitors have so far been modelled to better identify which compounds bind tightly with the active site of the protease. From these, potential lead compounds have been identified which are now being explored in the laboratory to investigate antiviral effects in model systems. This project is building knowledge related to viral replication processes following infection of cells, and guiding the selection of small molecules for further investigation.

BUILDING DRUG DATABASES FOR COVID-19 TESTING

Professor Alan Mark and his team at the University of Queensland have over the last 10 years created the Automated Topology Builder (ATB, atb.uq.edu.au), a globally-recognised molecular modelling tool that lets researchers turn a molecular structure into an accurate 3D representation with high-quality atomic interaction parameters required for computational drug design.

These can then be used to predict which compounds have the potential to interact with the SARS-CoV-2 spikes or other viral proteins. The detailed molecular geometry and atomic interaction parameters provided by the ATB can also be used to understand how the viral protein dynamically adapts to the presence of a given compound. This cannot be observed directly at the atomic level and must be simulated using computers. These 'molecular movies' are used to work out how and where particular molecules may bind and if they can disrupt the action of the virus.

The ATB repository is one of the largest pre-calculated molecular structure and parameter databases in the world, and already contains over 430,000 compounds which are freely accessed up to 1,000 times daily by researchers worldwide. Its creation has required a significant computational effort over 10 years.

Professor Mark's team is now working with Pawsey and the NCI to ensure that all pharmaceutically active compounds that have already passed Phase II clinical trials for human safety are incorporated into the ATB, to support research investigating if existing drugs can be repurposed to treat COVID-19.

Using a special allocation at both Pawsey and the NCI, at least one structural example of all 7,300 compounds that have passed Phase II clinical trials have been processed and added to the database so they are now available within the ATB for computational drug studies. The extra computation allocation on the Nimbus cloud service has meant that even the largest compounds (50–150 atoms) could be processed.

Although all 7,300 Phase II therapeutic compounds are now available within the ATB, many of these compounds exist in multiple forms within the human body, and the different forms often have distinctly different properties and biological effects. By the end of the year-long project, Professor Mark's team anticipates having all of the biologically-relevant forms of all of the Phase II compounds – between 50,000 and 100,000 structures – both in the database and bundled for distribution to COVID-19 drug researchers.

CASE STUDY 1

Partner Institution: International Centre for Radio Astronomy Research (ICRAR)

How do galaxies evolve, and is ours 'normal'?

Project Leader: *Dr Claudia Lagos, ICRAR UWA*
Areas of science: *Astronomy*
System: *Magnus, Zeus*
Applications used: *Boost, GSL, Cmake, Python, Scipy, Matplotlib, HDF5, R*

Understanding how galaxies evolve is fundamental to understanding our place in the Universe, as our Milky Way is quite different to even our closest galactic neighbours such as Andromeda and the Magellanic Clouds. But spanning the reaches of both time and space to study the evolution of different galaxies is beyond the power of even our best telescopes.

Dr Claudia Lagos, an ASTRO 3D Senior Research Fellow at the International Centre for Radio Astronomy Research (ICRAR) at UWA is combining numerical modelling of the physical Universe with telescope observations to work out the details we can't see directly.

THE CHALLENGE

Within our Milky Way, we can directly observe individual stars, and study their masses, colours, ages and distributions to learn a lot about the structure of our galaxy. But we can only see what is still here, as many early stars have gone supernova and no longer exist. "We can only see the later portions of everything that has happened in our galaxy," explains Dr Lagos, "so to learn about galaxy evolution we have to look at progressively younger galaxies, which are progressively further away."

The problem is, our telescopes can only observe the generalised and cumulative signals coming from distant galaxies, and not the detail of their individual stars. To study the aggregated characteristics of galaxies 10 billion years ago, statistics and modelling is needed to make sense of what you're seeing.

Dr Lagos elaborates: "Because there are limitations on what we can see, we can model galaxy formation, and then compare the cumulative outputs of the model with the combined signals we get from our telescopes."

Unfortunately, galaxy formation can't be simulated just by solving all of the underlying physical laws from first principles. "We can simulate processes on a small scale, like how black holes accrete matter at the sub-parsec scale, but not at the same time as we simulate the scales at which galactic structure emerges, as that's eight or nine orders of magnitude larger."

"Since we can't model all of the physical processes at all of these scales, we make approximations at the smaller scales about how individual stars form, and then see how those assumptions affect the developing galactic structures within the model."

Working out which assumptions and approximations are most accurate then relies on comparing the simulation outputs with real telescope observations.

"Can we accurately predict the variety of galaxies that we can see, their colours, masses, sizes, and population frequency throughout the Universe? If we can, then the assumptions we have made in creating the model tell us a lot about the underlying structure of the Universe and how those galaxies must have evolved."

> *"We're building a model that will effectively let us run the Universe in reverse, which will give us unique insights into where we've come from."*

1100000
core hours

8
publications from this project

100000 EUR
(MERAC) Research Award

It shows on the left the predicted colours and shapes of galaxies in a high density region of the simulated universe (which we would consider similar to a local galaxy cluster), while the right panel shows the predicted dark matter distribution of the same region. You can see how galaxies are tracing the highest concentrations of dark matter but not the full structural details we would see if we were to directly image dark matter. This figure was produced with one of the models and dark-matter only simulations we use.

THE SOLUTION

Supercomputing is needed to run the model, because it needs to simulate a representative volume of the universe. "A characteristic scale of the universe is a cube with sides 300 megaparsecs long," says Dr Lagos. "In comparison, our local galaxy group containing the Milky Way and Andromeda is a cube of roughly 1 megaparsec."

Over 30 million Central Processing Unit (CPU) hours are needed to calculate the movement and interactions of a billion particles under gravity over time, and up to 40 million CPU hours if the hydrodynamics are considered, to evolve the skeleton structure where galaxies will form. Then thousands of simulations are run on this skeleton to see how different assumptions and physical parameters affect how the skeleton is populated with galaxies, for comparison with direct telescope observations.

OUTCOME

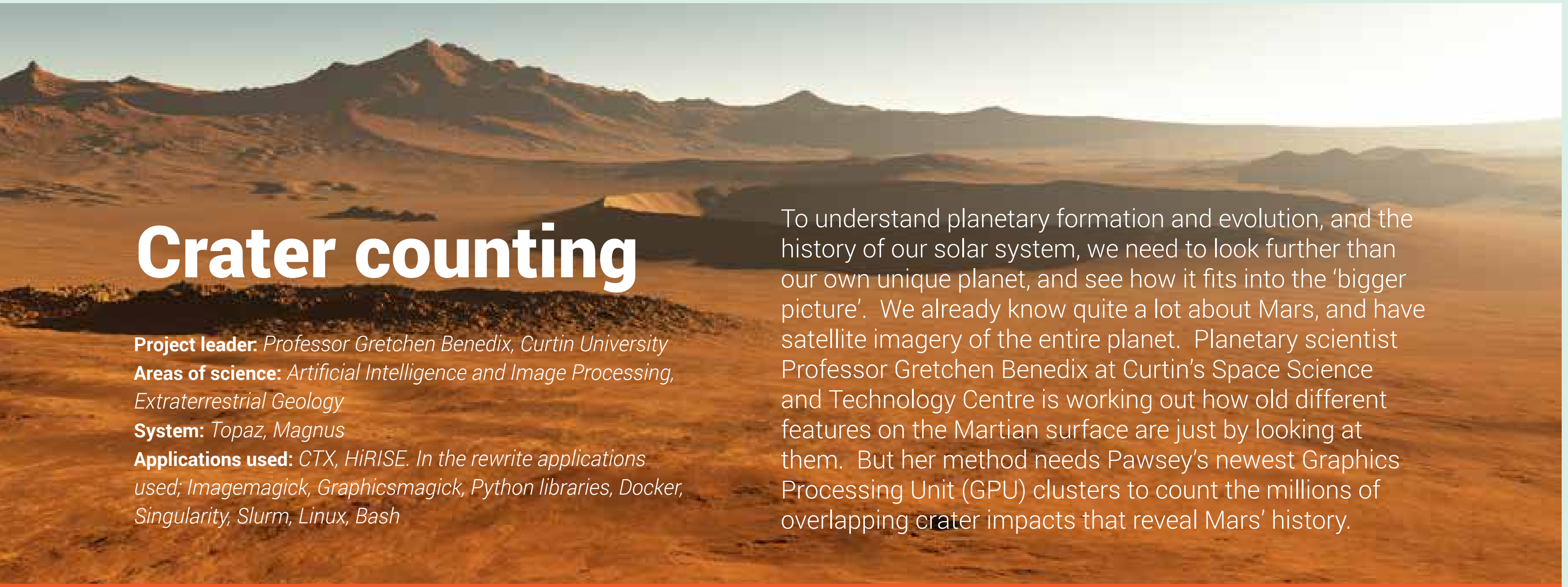
Comparing the model with the real Universe is painstaking as it is important to also reproduce the observation conditions, from the detection limits of the telescope to the effects of interstellar dust as it absorbs and re-emits starlight passing through. But Dr Lagos' model is already 'shedding light' on the formation of very bright, distant galaxies formed about 10 billion years ago, to fill in some of the early history of our own galaxy.

As the evidence mounts that the modelling is accurately representing our known Universe, it becomes more and more useful. Dr Lagos enthuses: "We're now starting to use the model to connect observations that represent different epochs in cosmic history – if I look at galaxies similar to the Milky Way, what did they look like five billion years ago? We're building a model that will effectively let us run the Universe in reverse, which will give us unique insights into where we've come from. And as telescopes become ever more powerful with the development of the Square Kilometre Array (SKA) and its precursors, we'll be able to check our understanding of galaxy evolution against what is really out there progressively further back in time."

CASE STUDY 2

Partner Institution:

Curtin University



Crater counting

Project leader: *Professor Gretchen Benedix, Curtin University*
Areas of science: *Artificial Intelligence and Image Processing, Extraterrestrial Geology*
System: *Topaz, Magnus*
Applications used: *CTX, HiRISE. In the rewrite applications used; Imagemagick, Graphicsmagick, Python libraries, Docker, Singularity, Slurm, Linux, Bash*

To understand planetary formation and evolution, and the history of our solar system, we need to look further than our own unique planet, and see how it fits into the 'bigger picture'. We already know quite a lot about Mars, and have satellite imagery of the entire planet. Planetary scientist Professor Gretchen Benedix at Curtin's Space Science and Technology Centre is working out how old different features on the Martian surface are just by looking at them. But her method needs Pawsey's newest Graphics Processing Unit (GPU) clusters to count the millions of overlapping crater impacts that reveal Mars' history.

200000 *core hours allocated in Topaz*

25000 *core hours allocated in Magnus*

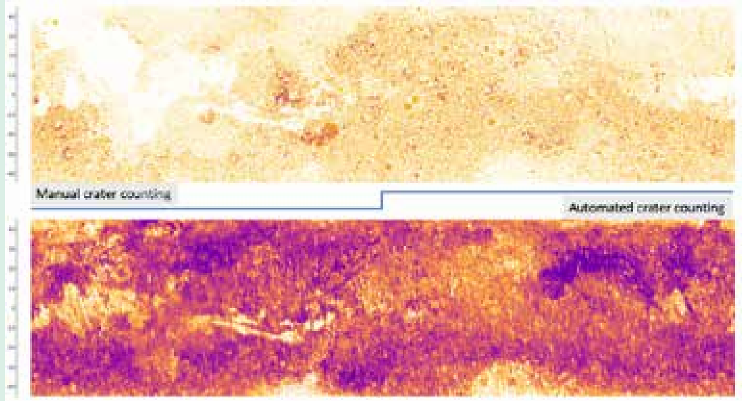
94000000 *craters found bigger than 25m in 24 hrs*

THE CHALLENGE

The craters on a planet's surface tell its history. The more craters, the older the surface since a volcanic or climactic event wiped it 'clean'. Older craters bear the scars of newer impacts on top of them. Crater counting is the principal tool used by astrogeologists to determine the surface ages of planets, our Moon and even large asteroids throughout the solar system. Until now, the technique has relied on scientists painstakingly identifying and counting craters by hand. The current database for Mars contains 385,000 identified craters with diameters of 1 km or larger. But it took at least six years to construct, before it was published in 2012.

"There is so much more information available now," enthuses Professor Benedix. "Using images sent by the Mars Reconnaissance Orbiter we have 5 m per pixel resolution, so we can see 25 m craters. But here's the catch – when asteroids collide in space, they break into a couple of really big pieces, a few more medium sized pieces, and lots of tiny pieces. This means there is a size frequency distribution for craters with smaller craters numbering exponentially higher than larger craters. The smaller craters we can see now amount to tens of millions."

Counting the smallest craters would let us create age maps with much higher accuracy and spatial resolution, and let us derive the surface ages for much smaller and younger features of the Martian surface.



The authors acknowledge the following links which include data sources for access to NASA mission data [THEMIS DayIR maps (https://tinyurl.com/y56c7542), CTX (http://murray-lab.caltech.edu/CTX/index.html), Mars Crater Database (https://tinyurl.com/y5c2qvfe)] as well as open source software [YOLOv3 (https://github.com/pjreddie/darknet and https://github.com/AlexeyAB/darknet), ImageMagick (https://imagemagick.org/index.php)]

THE SOLUTION

"We had to automate crater counting," says Professor Benedix. Combining the increasing surface detail provided by the Mars Orbiter with recent developments in object detection technology and machine learning, Benedix' team turned to specialists at the Curtin Institute of Computation (CIC) to help build a convolutional neural network to identify circles in an image. The algorithm was then put into a pipeline to use Pawsey GPUs to analyse the images from Mars.

Machine learning algorithms need a database of examples to learn from, so the algorithm was trained to recognise craters using images of 7,048 manually-identified craters from the Mars database. Once trained, when the algorithm analysed the remainder of the images making up the Mars crater database, it generated results comparable with the manual dataset, giving 91 per cent positive identification for craters 1 km and larger.

There is significant variability – around 85 per cent – between different manual crater data sets, arising from how subjectively people interpret an image, and their variations in attention and focus on repetitive tasks. Professor Benedix's algorithm identified craters just as well as a human observer, but more objectively, reproducibly, and much faster.

With the algorithm ready to identify even smaller craters from the 5 m per pixel high resolution imagery from the Mars Orbiter, Benedix requested Astronomy Data and Computing Services (ADACS) support to get a Pawsey developer embedded in the research project for six months.

"That made such a difference," says Professor Benedix. "Through improving the algorithm pipeline, the library training data set, and getting access to the newest GPU clusters at Pawsey, we went from being able to count craters in 100 m per pixel images fairly quickly to being able to analyse the global Mars dataset at 5 m per pixel, identifying 94 million craters bigger than 25 m in 24 hours. We're now analysing image sets of specific areas of Mars at a resolution of 30 cm per pixel. We can see craters the size of a car."

OUTCOME

Professor Benedix's team has created the largest Mars crater database in the world.

They've located and dated 20 of the youngest craters on the surface of Mars, and are now determining if any are related to the Martian meteorites that have been found here on Earth. They're working out when the last volcanoes were active and dating channel structures to see when water last flowed on the surface of Mars, along with any potential for life. It's information that may well help determine where the next NASA Mars Mission will explore, and all adds to our understanding of planetary formation and history.

Applications for the algorithm also extend well beyond Mars. "We've already applied it to the surface of the Moon and Mercury," says Professor Benedix. "We just needed to create a data training library specific to those different image sets and retrain the algorithm a bit."

It can even be used on Earth – the algorithm and pipeline at Pawsey, using the right library of training images, could equally see applications ranging from identifying cancerous spots on skin or unusual cells in pathology samples to recognising pipeline damage in images from remote underwater vehicles monitoring gas pipelines. Radar and sonar images could also be screened to automatically identify features of interest – you just need to teach the algorithm what to look for.

CASE STUDY 3

Partner Institution:

UNSW AUSTRALIA

Using fish to teach robots to swim

Project Leader: *Dr Fangbao Tian, University of New South Wales*
Areas of science: *AI/ Computational Fluid Dynamics*
System: *Nimbus, Magnus, Zeus, Data Portal*
Applications used: *ST FEM, SI IBM and IB LBM*

Underwater vehicles are used in environmental monitoring, resource exploration, offshore infrastructure construction and maintenance, maritime security, and search and rescue. Whether remotely-operated or truly autonomous, their movement through water cannot compare to the efficiency and precision of those that have adapted over millennia to their aquatic environment, such as fish.

Creating more efficient 'swimming' vehicles is one challenge, but optimising patterns of movement to suit changing marine conditions is an even bigger challenge. It takes a supercomputer to achieve what fish do naturally.

THE CHALLENGE

Fish have evolved a range of swimming modes to propel themselves efficiently through water, and adapt their swimming behaviour in different environments (steady currents, turbulent flow, vortices forming around obstacles, and the local flows generated within shoals of fish themselves) and to achieve different goals (obtaining food, escaping from predators, migrating). They can transition dynamically between swimming styles or gaits rapidly to increase speed, change direction or save energy.

Fish movement has been studied for a long time to improve the design concepts and locomotion control strategies for man-made submersible vehicles. Dr Fangbao Tian, from the School of Engineering and Information Technology at the University of New South Wales, Canberra, has been using computational fluid dynamics techniques to understand how fish can optimise their movement in different flow conditions. "Modelling how fish swim is very challenging," Dr Tian explains. "Understanding the flow–structure interactions, when the structure itself – the fish – is moving, requires novel Cartesian mesh-based methods to compute the continually changing boundary conditions at that moving interface."

It's like mapping how a robot 'dog' moves over an uneven surface. Except the shape of the surface is changing, and the shape of the dog is changing too.

But building a numerical simulation that can move like a fish in a defined flow is just the start of the problem. Dr Tian is now combining these models with AI learning techniques to see how they can adapt their movement to changing flow conditions, or changing goals.

THE SOLUTION

"We initially had to develop our own computational fluid dynamics solvers to simulate the flow conditions and the fish movement, because the computational efficiency of the existing solvers for this type of problem was very low. It requires supercomputing, and we need to obtain the fluid dynamics in a reasonable time if we want to combine multiple movements with learning methods to see if the simulation can learn and optimise its behaviour to the conditions," says Dr Tian.

Using Pawsey's supercomputing facilities, Dr Tian is now using reinforcement learning algorithms to train neural networks using the computational fluid dynamics simulations. Through sampling information about the flow field around the swimming model, using trial and error, and remembering the outcomes of historical actions, the model can be trained to achieve a specific goal.

"It takes a lot of computation to build up and explore a database of many different actions in many different flow conditions for the model to learn from," says Dr Tian. "Using Pawsey we can explore 4,000–5,000 episodes of a particular scenario, which takes about two weeks using around 300 cores."

"The ultimate aim is to use the understanding we're gaining of the flow physics and the movement control strategies to design underwater robots with improved locomotion, manoeuvring and holding capabilities in real marine environments."

250000
core hours allocated

14
team members

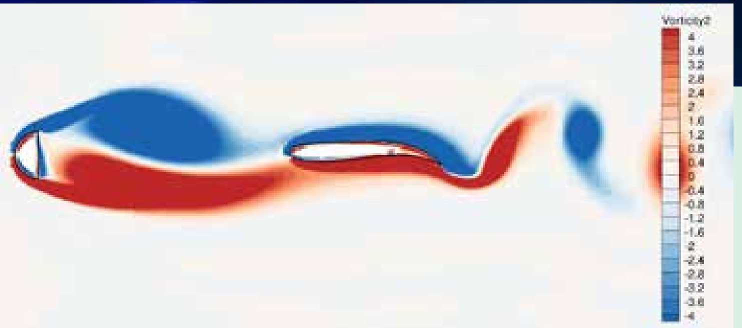
20
tail beat model optimatizacion achieved

8
journals published in a year

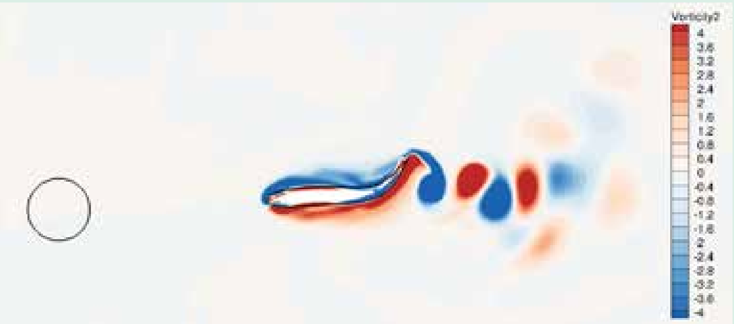
OUTCOME

Dr Tian's models have been learning how to swim point-to-point in a steady flow, hold their position in an on-coming current, and hold their position in the regular vortex 'street' that forms behind a cylindrical obstruction. Whereas real fish can adjust their swimming behaviour to these sorts of conditions within three or four beats of their tail, after training, the models can optimise their swimming motion to the conditions within about 20 tail beats.

"We can't match thousands of years of evolution yet, but we're getting closer," says Dr Tian. "The ultimate aim is to use the understanding we're gaining of the flow physics and the movement control strategies to design underwater robots with improved locomotion, manoeuvring and holding capabilities in real marine environments."



Visualisation fish point-to-point swimming in a steady flow



Karman gaiting into the vortex street forming behind a hemicylinder

CASE STUDY **4**

Partner Institution:

THE UNIVERSITY OF WESTERN AUSTRALIA

Tracking denizens of the deep to protect their future

Project Leader: *Dr Ana Sequeira, UWA*
Areas of science: *Aquaculture, Marine Science, Environmental Science*
System: *Nimbus*
Applications used: *R*

Scientists have tracked animals for decades, to learn more about their behaviour and also to identify the environmental niches they prefer to inhabit. Contrasting this knowledge with intersecting human activities can inform conservation and management practices for threatened species.

Dr Ana Sequeira from the UWA Oceans Institute and School of Biological Sciences is expanding this approach to the global scale. Combining the efforts of hundreds of researchers around the world, she has compiled tracking datasets for over 100 different species across the global oceans, and is using this information to map biodiversity hotspots and identify areas of potential conservation value for our most migratory animals.

THE CHALLENGE

Effective conservation and protection of highly migratory species like marine mammals is challenged by the fact that they move through different ecosystems and across national jurisdictions; two-thirds of the global oceans are effectively international waters. They can spend different seasons in entirely different environments, impacted by a range of human activities that may go unrecognised from a local or even national perspective.

Dr Sequeira is spearheading a global initiative to address this limitation by building an animal-tracking database and models that can combine information from many isolated tracking studies, to provide a global picture of animal activity and environmental preferences across our oceans.

"We've got a network of researchers around the globe that are providing their tracking data going back as far as 30 years", says Dr Sequeira. "Together, we have assembled many thousands of animal tracks containing millions of locations over time showing where individuals from over 100 species have travelled. The species range from polar bears, whales, seals and turtles to penguins and albatrosses, as well as a wide range of sharks."

"We've got a network of researchers around the globe that are providing their tracking data going back as far as 30 years. Together, we have assembled many thousands of animal tracks containing millions of locations over time showing where individuals from over 100 species have travelled. The species range from polar bears, whales, seals and turtles to penguins and albatrosses, as well as a wide range of sharks."

100
species tracked across global oceans

160000
different simulations

48
Nimbus cores across 2 virtual machines

300
researchers worldwide working on the project

Dr Sequeira is using this information to match animal activity with a range of environmental conditions measured at the same locations and times. Factors like temperature, chlorophyll concentration and current speed can be extracted from satellite data and oceanographic models, representing millions of additional measurements over time.

THE SOLUTION

"We're using Pawsey to prepare for this large analytical effort," explains Dr Sequeira. "As a starting point, we're running simulations to make sure we can cope with the variability and biases that can be found when doing these types of studies."

Different tracking datasets come with their own experimental biases. For example, animal locations measured using Global Positioning System (GPS) technology are inherently more accurate than datasets acquired using other technologies, such as light geolocation, which results in only two positions recorded per day. Data quality also depends on what animal is being tracked. For satellite tags to transmit positional data, the tag's antenna needs to breach the water surface – so the frequency of positions recorded for an air-breathing animal like a whale or turtle will be much higher than for a shark.

There are also gaps in the environmental data. "Because we're looking at a global scale, there are both temporal and spatial resolution limits to the data we can access. For example, with satellite data, it is often hard to get daily chlorophyll information for all areas we are interested in. For such cases, averaged weekly or monthly data might be more useful."

Using simulation datasets, results for models using specific combinations of known biases can be obtained in a controlled way. To prepare for the real dataset the team has run over 100,000 different simulations, each representing a set of tracks containing hundreds to thousands of positions over time, and identified local environmental information associated with each of those positions. The results from these simulations have been used to define the best strategy to cope with all of the real-world tracking data assembled.

OUTCOME

"We're now ready to run our models with the real animal tracking and environmental data, at a global scale," enthuses Dr Sequeira. "This will be the first global study incorporating so many different species from so many different taxa."

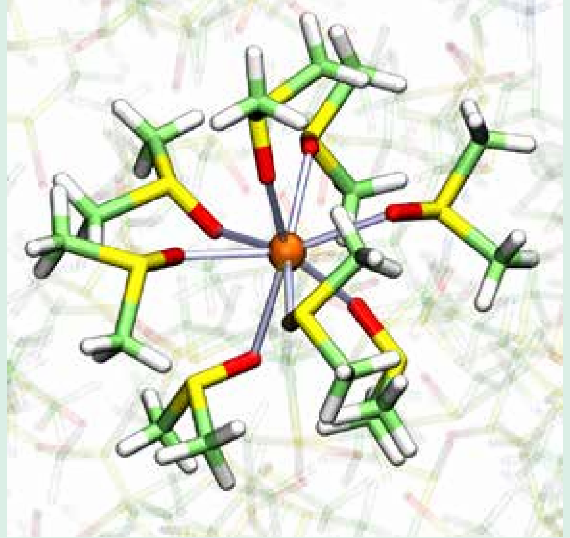
The aim is to be able to predict the areas of highest biodiversity and environmental value around the world for highly migratory species such as whales, sharks and turtles.

"Being able to overlay areas of ecological importance for a range of species with existing human impacts like shipping, fishing and oil and gas extraction will provide an evidence base to work from to assist understanding which areas might be critical to the survival of each species or the health of entire ecosystems."

CASE STUDY 5

Partner Institution:

THE UNIVERSITY OF SYDNEY



Simulation snapshots of perovskite precursors in solution.
Image courtesy Asaph Widmer-Cooper

Capturing sunlight with supercomputing

Project Leader: *Dr Asaph Widmer-Cooper, University of Sydney*
Areas of science: *Chemistry, Physics*
System: *Magnus, Topaz*
Applications used: *LAMMPS, Siesta, Quantum Espresso*

Converting sunlight into electricity is now commonplace, with efficient silicon-based solar cells steadily becoming more affordable. But manufacturing them from very pure silicon will always be a relatively slow and energy-intensive process. Alternatives to silicon solar cells are in development which promise to be just as efficient, simpler to manufacture, and cheaper – if they could be produced at commercial scale with improved stability. Dr Asaph Widmer-Cooper from the University of Sydney node of the ARC Centre of Excellence in Exciton Science is studying how these materials are created and break down at the atomic level, to guide efforts to make them commercially viable.

THE CHALLENGE

Metal halide perovskites are a class of minerals that could revolutionise renewable energy generation. Perovskite solar cells can capture sunlight as efficiently as silicon, but can be manufactured in a completely different way. Rather than melting, slowly crystallising and then slicing fine wafers from a large crystal, as is done with silicon to make solar cells and computer chips, metal halide perovskites can be formed into thin crystalline films by depositing a solution of their component parts onto an appropriate base and then removing the solvent, effectively 'printing' the solar cell on a surface.

A continuous printing process offers a low-cost route to solar cells on a range of surfaces, but while these can be produced at small scale, it is very difficult to form uniform thin films of perovskite at large scale. Trial-and-error printing is narrowing down the conditions needed to reproducibly produce thin crystalline layers, but very small changes in crystallisation conditions can lead to big differences in final product performance. Researchers still need a deeper understanding of how perovskite crystals form at the molecular level, and how to influence their crystallisation to make a more robust and reproducible commercial product. They also need a better understanding of how these thin films degrade in the presence of water, and how the insertion of other molecules into the crystalline film can inhibit that process.

Complementing practical printing experiments, Dr Widmer-Cooper is using computational modelling to study the crystallisation and dissolution of perovskites at the molecular level. The atomic structure of the crystal and its light-capturing properties are well known, but modelling its formation from solution requires calculating the behaviour of millions of atoms for long enough to observe crystal nucleation and growth. "The model is just a set of mathematical rules governing how the various ions and molecules in both the solution and at the surface interact with each other," explains Dr Widmer-Cooper, "but there's a trade-off between the accuracy of the model we can use and the amount of time we can run it for".

THE SOLUTION

Using Pawsey supercomputing, Dr Widmer-Cooper has tested a range of models to develop one that describes perovskites accurately both in the crystalline state and in solution.

"The first commercial outcome is likely to be printing perovskite on top of existing silicon solar cells. Perovskites can be tuned to absorb different parts of the solar spectrum than silicon, so a dual-layer cell will absorb and convert more sunlight to electricity than a silicon solar cell alone."

200000
core hours allocated in Topaz

1350000
allocated in Magnus

$31850000
in funding over six years

6
collaborating institutions

1_paper_fig4. Image courtesy Asaph Widmer-Cooper

"To see crystal growth we have to run the model for thousands of nanoseconds, recalculating atomic movements each quadrillionth of a second," says Dr Widmer-Cooper. "We can only do that using a very large CPU cluster like Magnus, or a GPU cluster like Topaz."

Dr Widmer-Cooper and his team are now using their model to 'watch' how the crystals both grow from solution and dissolve. "The next step is to explore those mechanisms, using the model to test how they are affected by variations in temperature, concentration, solvent and additives."

"Using this model, we're hoping to provide some molecular-level insight into the variations that occur in large-scale perovskite printing processes, and suggest approaches to test to make thin-layer crystallisation more reproducible and robust for commercial production."

OUTCOME

Once we really understand the mechanisms of crystallisation and dissolution and what affects them, we can create the most stable conditions to scale up perovskite printing for commercial production. It's a process worth pursuing, as Dr Widmer-Cooper has already seen what can be achieved at an experimental scale.

"The first commercial outcome is likely to be printing perovskite on top of existing silicon solar cells. Perovskites can be tuned to absorb different parts of the solar spectrum than silicon, so a dual-layer cell will absorb and convert more sunlight to electricity than a silicon solar cell alone."

From there, the aim will be to create fully printable solar cells, with the potential to create solar panels on a range of different surfaces – even curved or flexible ones. All at a fraction of the effort and cost of using silicon.

CASE STUDY 6

Partner Institution:

APPF Australian Plant Phenomics Facility
The Plant Accelerator® | University of Adelaide

Improving crop performance by seeing plants in a new light

Project Leaders: *Dr Bettina Berger and Mr George Sainsbury, University of Adelaide*
Areas of science: *Plant Biology*
System: *Nimbus*

The Australian Plant Phenomics Facility (APPF) allows academic and commercial plant scientists from Australia and around the world to measure phenomic information (physical and biochemical traits) about thousands of plants a day over the course of their life cycle. In combination with controlling environmental conditions for growth, this information is being paired with plant genetic information to identify the most promising traits for growing crops tolerant to heat, drought, salinity or limited nutrients for our changing climate. But as plant trials get larger and phenotype measurement systems get more powerful, data management and storage requirements also grow.

THE CHALLENGE

The APPF is a national research infrastructure platform funded by the National Collaborative Research Infrastructure Strategy (NCRIS) that provides large-scale plant growth research facilities to address complex problems in plant and agricultural sciences. A major focus is developing improved crops and more sustainable agricultural practices in the face of declining arable land and the challenges of climate change.

Dr Bettina Berger, Scientific Director of the APPF node at the University of Adelaide, explains: "We explore how the genetic makeup of a plant determines its appearance, function and performance. At the APPF, our high-throughput, automated phenotyping systems use imaging technologies to phenotype over 2,000 plants a day moving on conveyors in our 'smart' greenhouses with watering, nutrient and temperature control."

Visible light and fluorescence imaging have previously provided measures of plant height and width, canopy density, leaf colour, and evidence of ageing, nutrient limitations or disease. Complementing these, the advent of hyperspectral imaging (collecting images across a much wider range of discrete electromagnetic frequency bands) and automation of camera operation on tractors or conveyors to increase measurement speed has greatly expanded the volume of phenotypic data available, and enabled repeat measurement of plant properties over several stages of plant growth.

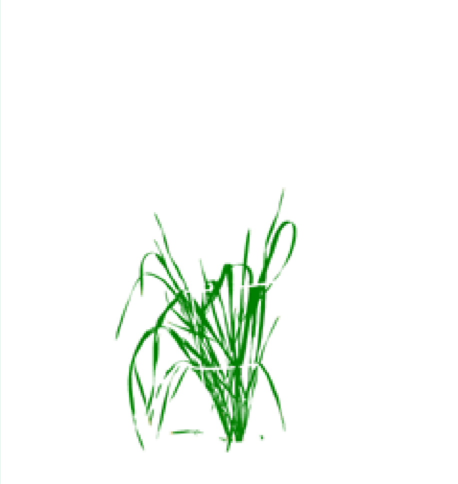
"Our experiments can now record colour images of thousands of plants and hyperspectral images and X-ray Computerised Tomography (CT) scans of hundreds of plants over the course of an experiment," says Dr Berger. "As a national research facility allowing researchers to run their plant trials on this scale, we're now having to partner with another national research facility like Pawsey to support our researchers in managing, accessing and analysing these data sets."

THE SOLUTION

Mr George Sainsbury, a data architect and software engineer at the APPF, notes that because plants grow at a steady rate, and because images are being collected at a steady rate, the ongoing real-time analytical and computational requirements of an experiment are still manageable. "We don't need supercomputing to analyse our image sets – yet. But as a time-series of images accumulates, data storage and management become challenging."

The APPF now uses Pawsey's Nimbus cloud service for server infrastructure and the data portal for data storage and sharing. Mr Sainsbury explains: "As soon as we acquire an image it is now sent straight to a data store at Pawsey. We access it from there for analysis and the results are then added to another database at Pawsey. The whole process is automated."

The Field Explorer is the APPF's newest phenotyping system. It combines RGB and hyper spectral imaging with LIDAR and INS location data to measure traits such as digital biomass and spectral reflection over field trials. (Picture: Naomi Jellicoe). One of the thousands of images taken each day at The Plant Accelerator. This is an 8-week old wheat plant, part of an experiment investigating the effect of different soil types on plant growth. Images provided by George Sainsbury, Australian Plant Phenomics Facility.



A visualisation of the segmentation of the plant above. The green represents the parts of the image automatically determined to be plant, and the white is the background. This "mask" is the first step to inferring the plant's growth over time and other traits. Images provided by George Sainsbury, Australian Plant Phenomics Facility

30	60200	15
Nimbus cores across 5 virtual machines	*GB of storage*	*TB of archival storage per year*

Nimbus is used as the active data store for experiments, but as experiments finish the images and databases are also archived at Pawsey.

"We're using around 30 cores across five virtual machines on Nimbus to operate our various server software and analytical platforms, and our archival storage is growing at around 15–20 TB each year."

OUTCOME

"We've got to the point that our plant researchers don't have to worry about how all their experimental data is stored and analysed," says Dr Berger. "We support plant research across the country and internationally, so researchers outside of Adelaide can now follow their experiments remotely and use data for decision making while the experiment is still running."

"We're using machine learning and other advanced computational techniques to interpret the hyperspectral images as they're generated, and then applying a range of smoothing methods to extract growth traits over time that the researchers can then use in their genetic analyses."

> *"As a national research facility allowing researchers to run their plant trials on this scale, we're now having to partner with another national research facility like Pawsey to support our researchers in managing, accessing and analysing these data sets."*

Mr Sainsbury points out that the APPF's use of Pawsey facilities is only likely to increase. "We're still learning how to extract the most plant growth- and plant health-relevant hyperspectral, x-ray, and other remote-sensing data at scale. As the datasets become more layered and complex, the number of plants in trials increase and the resolution of the images increases, we'll reach a point where we may need to also incorporate Pawsey supercomputing into our automated data pipeline to analyse hundreds of gigabytes of imagery at a time."

Dr Berger sums it up: "Between Pawsey and the APPF, we can now bring massive amounts of genomic and phenomic information together, to better understand what specific genes of a plant contribute to crop success in the face of various environmental stresses. This supports plant breeding and crop improvement efforts worldwide."

CASE STUDY 7

Partner Institution:

Department of Primary Industries and Regional Development
GOVERNMENT OF WESTERN AUSTRALIA

DNA detectives to combat animal disease

Project Leader: *Mr Sam Hair, WA Government – Department of Primary Industries and Regional Development*
Area of science: *Biology*
System: *Zeus*
Applications used: *Nextflow*

When birds fall out of the sky, fish float lifeless in the river, or bats drop dead from trees, we scramble to find a cause. When mystery illnesses spread through our aquaculture fisheries or sheep flocks sicken, the race to identify a cause and a cure is even faster, given how quickly disease can spread in our intensive farming industries. Mr Sam Hair at the WA Department of Primary Industries and Regional Development (DPIRD) is using next-generation sequencing methods and Pawsey supercomputing capabilities for data analysis to quickly identify the pathogens causing mystery disease outbreaks, reducing the time to respond.

THE CHALLENGE

The Animal Pathology Laboratory at DPIRD works to protect the WA community and agriculture and food sector from the impacts of diseases and pests that affect animals and aquatics. When a mystery illness affects a herd or a sudden fish kill chokes a river, Mr Hair uses all of the tests at his disposal to identify the cause. But standard diagnostic tests are time-consuming, and sometimes fail.

"If the illness is caused by a brand-new strain of a virus, our standard diagnostic tests based on older strains may not identify it," explains Mr Hair. "And when time is of the essence, testing for one potential cause after another and progressively ruling them out is time-consuming. Sometimes we can exhaust our battery of tests and still not find a cause."

Image courtesy of Mr Sam Hair, WA Department of Primary Industries and Regional Development (DPIRD)

"The problem is that assembling all of the genetic material and comparing it to a database of millions of genetic sequences also takes time. Data analysis is the real bottleneck."

100000
core hours

24
hrs max metagenomic sequence

4
team members

46
cases of unknown diseases tested

Next-generation sequencing raises the possibility of amplifying and expressing all of the genetic material found in an animal sample, including lurking bacteria and viruses. If the fragments can be stitched into complete genomes, these can be compared against international genome databases to identify the mystery pathogens that shouldn't normally be there.

"The problem is that assembling all of the genetic material and comparing it to a database of millions of genetic sequences also takes time. Data analysis is the real bottleneck."

THE SOLUTION

Mr Hair undertook a Pawsey Uptake project, working with supercomputing applications specialist Dr Marco De La Pierre to speed up the analytical work required to turn gene fragments into a disease diagnosis.

"We still do the genome sequencing in the DPIRD laboratory, but instead of processing those fragments on our local desktop computers, we port the data over to the Pawsey systems. With Pawsey expertise, we created an automated workflow that assembles the genomes and then compares them against existing databases."

The workflow reduces the effort Mr Hair spends on repetitive analysis, genome assembly, quality control and database searching to just a few computer commands to analyse a collection of samples. The result is either a match for a known disease pathogen, or the complete genome of something new, which can become the basis for a more targeted and routine diagnostic test.

THE OUTCOME

"If an animal is sick or dead from something completely unknown, we need a diagnostic protocol that covers all of the possibilities and identifies everything there, not just a test that may confirm what we might suspect. Next-generation sequencing gives us that capability," points out Mr Hair. "But it is only useful if we can analyse enough samples and process the data fast enough to make a difference to a real-world scenario."

Investigating a recent fish-kill incident highlights the power of the DNA detective approach. Whereas Mr Hair would have previously tested for likely disease culprits for weeks until he got a positive result, next-generation sequencing was used to very quickly rule out the involvement of any pathogen at all, allowing the Department to turn its focus immediately to environmental factors instead.

Dr De La Pierre, Pawsey supercomputing applications specialist, is continuing to work with Mr Hair to streamline the workflow, improve its speed and ease of use, and enable data from multiple genetic sequencing technology platforms to be used. The workflow is available to any research organisation accessing Pawsey facilities, so it can be used by other researchers to quickly assemble and identify any genomes of interest, regardless of where the sample came from.

But for Mr Hair and DPIRD, it's Western Australia's newest tool in bio-surveillance and disease investigation to support animal welfare, food safety and agricultural production.

Sheep at Murdoch University. Image courtesy of Mr Sam Hair, WA Department of Primary Industries and Regional Development (DPIRD)

CASE STUDY **8**

Partner Institution:

Department of Primary Industries and Regional Development
GOVERNMENT OF WESTERN AUSTRALIA

Local tests for local viruses

Project Leader: *Dr Monica Kehoe, WA Government – Department of Primary Industries and Regional Development*
Areas of science: *Plant Pathology*
System: *Magnus, Zeus*
Applications used: *Nextflow, Singularity and various Bioinformatics software containers.*

Plant viruses cause diseases in food crops that result in losses in quality and production valued at over $30 billion a year worldwide. Just like the common cold virus, plant viruses evolve over time into different strains or new species, making their early detection and identification an ongoing challenge. Dr Monica Kehoe at DPIRD is using next-generation sequencing methods and Pawsey supercomputing for data analysis to develop diagnostic tests that can keep up with local plant virus variations.

THE CHALLENGE

Grapevine leaf-roll viruses (GLRV) infect grape vines and are a major problem for grape growers across Australia, causing poor fruit quality and reducing yield between 10 and 70 per cent. There are at least four different GLRVs, and more are being identified each year in grape growing countries. These viruses are typically transmitted between plants by mealy bugs. However, each different GLRV can produce a range of different symptoms and severities. Some plants may be asymptomatic but contribute to the spread of the virus in a vineyard, while other plants may display similar symptoms but be suffering from poor environmental conditions, like limited nutrition. In addition, symptom expression can vary between different varieties of grape.

Unlike the common cold in people, once a plant becomes infected by a virus, it has it for life. Therefore, growers can only contain the virus by managing its spread or removing and destroying all infected plants. Early detection of virus infection then becomes critical when considering the health of a block of 40-year old shiraz vines, or when setting up a brand-new block of chardonnay vines.

Unfortunately, the available diagnostic tests for GLRVs are mostly produced overseas, and are a bit 'hit and miss', according to Australian growers.

"The genetic diversity of GLRV 1 and 3, the two major viruses of concern in Australia, is largely unknown, especially in Australia," explains Dr Kehoe, a plant virologist and molecular plant pathologist for WA DPIRD. "There are 56 complete genomes available in the public database for GLRV 3, but only one is Australian. And there are only seven genomes of GLRV 1 available, none from Australia. This is potentially a problem because we don't know if the primers used for our current tests, which are based on those known genetic sequences, will pick up all of the GLRV 1 and GLRV 3 strains circulating in our vineyards. Also, if strains emerge here that are different, those tests could fail."

250000
core hours

30
mins potential for virus detection

1
week for genome results instead 18 mths

The size of that task shouldn't be underestimated, warns Dr Kehoe: "My earlier PhD with other virus genomes relied on 24 samples, and that analysis took me about 18 months." With supercomputing, Dr Kehoe estimates she could now do that in a week.

Dr Kehoe is using next-generation genetic sequencing technologies to develop new diagnosis methods and faster tests for GLRV 1 and 3. She collected 264 samples from vineyards across south-west Western Australia for analysis, 103 testing positive for GLRV 1, and 123 positive for GLRV 3. The next step was to sequence the genomes of these Australian virus isolates.

The size of that task shouldn't be underestimated, warns Dr Kehoe: "My earlier PhD with other virus genomes relied on 24 samples, and that analysis took me about 18 months." With supercomputing, Dr Kehoe estimates she could now do that in a week.

THE SOLUTION

Pawsey supercomputing was needed to speed up the assembling, mapping and phylogenetic analysis of Dr Kehoe's genetic samples, which resulted in 13 complete genomes for GLRV 3 and 12 complete genomes for GLRV 1. Although analysis suggests that all of the known GLRV 1 genomes so far are similar, the number available for comparison worldwide is small. However, the 13 new GLRV 3 genomes, added to the existing worldwide database of 56, suggest that the diversity of this virus is quite high.

These Australian genome sequences were then used to adjust the testing procedures used for GLRV 3 in WA, changing the primers to ensure that screening for GLRV 3 in WA is up-to-date and less likely to miss local strains.

OUTCOME

The DPIRD Diagnostics and Laboratory Service now has improved diagnostic capability for GLRV 1 and 3, and can test large numbers of samples with relatively short turnaround times. Vineyard owners can now identify and remove infected vines before the virus spreads, and can also be confident in planting clean vines when developing new vineyard blocks. Dr Kehoe has also developed a portable molecular diagnostic test for those specific virus strains, with the potential to be used in-field by growers and agronomists to identify GLRV 1 and 3 within 30 minutes. "This would have taken years without the help of Pawsey supercomputing, and I can be confident that my new tests cover all the known diversity of GLRV 1 and GLRV 3, which now includes local WA virus genomes."

"The new virus genomes will soon be added to the GenBank database," adds Dr Kehoe, "so they're available to other researchers across Australia and internationally to include when they design new molecular tests for their local conditions and viral strains, as well as for broader research and monitoring of viruses in our crops that impact on our food security."

The molecular techniques developed and the workflows created at Pawsey to speed up the analysis required is giving DPIRD the ability to routinely test for more viruses with locally-specific diagnostic tools, not just in vineyards, but also in our grains industry, and in our cucurbit (zucchini, cucumber and melon) growing regions. By making the vast amounts of information unlocked by next-generation genetic sequencing technologies accessible in realistic timeframes, supercomputing is allowing DPIRD to protect agricultural industries across Western Australia from the impacts of viral disease.

CASE STUDY 9

Partner Institution:

THE UNIVERSITY OF MELBOURNE

Mapping arteries to treat heart attacks before they happen

Project Leaders: *Prof. Andrew Ooi, Dr Eric Poon, Dr Shuang Zhu, Prof. Peter Barlis and Dr Vikas Thondapu, University of Melbourne*
Areas of science: *Biomedical Engineering, Fluid Physics*
System: *Magnus, Zeus*
Applications: *OpenFOAM*

According to the Australian Bureau of Statistics 2017–18 National Health Survey, an estimated 1.2 million (six per cent) of Australian adults aged 18 years and over had one or more conditions related to heart or vascular disease, including stroke.

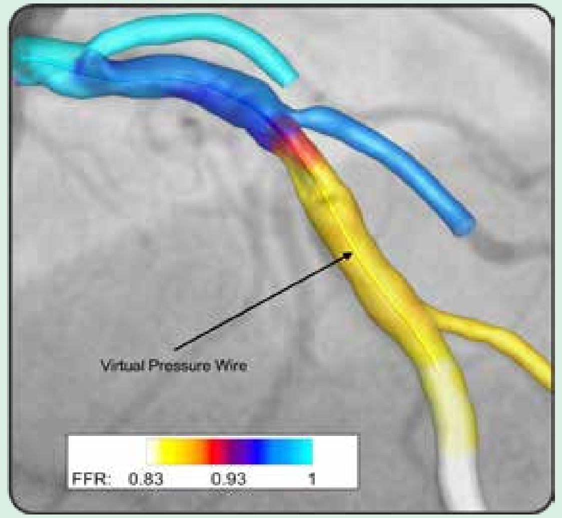
University of Melbourne's researchers Professor Andrew Ooi, Dr Eric Poon and Dr Shuang Zhu have partnered with clinicians Professor Peter Barlis and Dr Vikas Thondapu in pioneering virtual tools using mathematics and Pawsey supercomputing to develop a method to rapidly and accurately detect coronary artery disease and inform treatment well before fatal heart attacks occur.

Their approach is determining the extent of arterial blockages and inform the best course of treatment recommendations.

THE CHALLENGE

Coronary artery disease, or narrowing of the coronary arteries through plaque buildup, is the most common cause of death globally through causing heart attacks. Many medical approaches to assess the degree of artery narrowing are invasive and provide limited information to guide treatment.

The current gold standard to assess arterial blockages involves catheter-based coronary angiography, which takes low-resolution 2D pictures of the artery. This doesn't give a functional assessment of blood flow though, which requires an even more invasive procedure to insert a pressure wire into the artery.



Computer reconstruction of an artery. Image processing, AI and computational fluid dynamics to work towards a new gold standard providing cardiologists with detailed information to guide them in the diagnosis and treatment of coronary artery diseases

> *"We're combining skills in image processing, Artificial Intelligence (AI) and computational fluid dynamics to work towards a new gold standard that will provide trained cardiologists with more detailed information to guide them in the diagnosis and treatment of coronary artery diseases."*

1250000
core hours allocated

17
journal articles

22
peer-reviewed conference papers

The wire measures the pressure at just two points, but the pressure drop between the points can be used to infer the size of the blockage and hence the volume of the blood flow getting through. Neither technique gives the exact location or geometry of the blockage.

To improve on this, Professor Ooi is using information collected by interventional cardiologist Professor Barlis during angiography and combining it with Optical Coherence Tomography (OCT) medical imaging data, the highest resolution 2D arterial cross-section imaging available, to build accurate 3D computer models of diseased arteries. Computational fluid dynamics simulations can then be run to predict blood flow in the artery, giving a complete picture of flow restriction and pressure at every point, accurately mapping blockages in much greater detail to guide treatment.

The technology has already been demonstrated and compares extremely well against existing diagnostic methods. The challenge, Professor Ooi explains: "is to provide the answers fast enough to be clinically useful for cardiologists, who often need to make decisions to operate on a critical patient within minutes, not days".

THE SOLUTION

Reconstructing the artery inside a computer is the first part of the process. "It takes a few days on a local computer, but it's not convenient to build the simulations much faster on the Pawsey supercomputers yet," admits Professor Ooi. "Reconstruction currently requires a lot of manual work so it's easier to do that locally."

Pawsey supercomputing facilities are being used to run the computational fluid dynamic simulations on the complete artery reconstructions. "We can get the full flow field, pressure at every point, turbulence, wall shear stress – all of the blood flow behaviour that is physiologically significant, created from the detailed shape of the artery. But even using hundreds of processors at Pawsey, it's taking too long to do the computations for the cardiologists to use it. They want answers in minutes," says Professor Ooi.

"Right now we're working out the minimum information that they need – probably just a pressure map rather than a full flow field – and developing algorithms to predict only that, much faster."

OUTCOME

The project aims to develop an efficient computational tool to give cardiologists the information they need to diagnose and recommend treatment in a clinically useful timeframe. With information available about artery shape and obstruction, and assessment of blood flow around the blockage, they can decide whether a patient can be treated with medication, or if expansion of the artery is required via the insertion of a stent.

"We're combining skills in image processing, Artificial Intelligence (AI) and computational fluid dynamics to work towards a new gold standard that will provide trained cardiologists with more detailed information to guide them in the diagnosis and treatment of coronary artery diseases", Professor Ooi forecasts. "If we can automate and simplify the process enough to enable fast, high-volume use in a clinical setting, we can avoid invasive pressure wire tests and reduce the risk of complications associated with diagnostic tests, while providing rapid and accurate detection of coronary artery disease and improved treatment well before fatal heart attacks occur."

CASE STUDY 10

Partner Institution:

RMIT UNIVERSITY

Improved chronic pain relief treatments

Project Leader: *Dr Andrew Hung, RMIT*
Partner Institution: *RMIT University, University of Wollongong, University of Queensland, Monash University*
Areas of science: *Biology*
System: *Topaz, Magnus*
Applications used: *GROMACS*

250000 core hours allocated on Magnus

16 publications since 2018

20000 core hours allocated on Topaz

260 toxins studied

One in five Australians lives with chronic pain, a debilitating disorder which also afflicts adolescents and children, and this rises to one in three people over the age of 65. When physical pain is routinely felt in the absence of an external stimulus, it becomes a problem of our central nervous system.

Researchers have turned to toxins produced by venomous animals to find an alternative treatment for chronic pain and other neurological disorders. Dr Andrew Hung from the College of Science, Engineering and Health at RMIT University is using Pawsey Supercomputing Centre to discover exactly how the chemicals known as conotoxins interact with our nervous system, guiding drug design to turn these toxins into a stable new pain relief drug.

THE CHALLENGE

Toxic substances found in marine cone nails, known as conotoxins, contain small proteins (peptides) that interact very strongly with our nervous systems. This makes them promising starting materials for drug development to treat neurological disorders, including chronic pain. As Dr Hung explains: "Conotoxins attach strongly to a range of different receptors in the nervous system, but we need to modify them so that they attach only to very specific targets. If we can 'tune' their binding to be selective to only a particular receptor, we can control their action to achieve the effect we want from a drug – pain relief – without other side effects".

Thousands of natural conotoxins are known, and more can be created through chemical modification, so the potential to tune their selectivity and activity is high. This fine-tuning requires an understanding of exactly how the toxins interact with our nervous system, the 'molecular handshake' that occurs between the peptide and various receptors in the nervous system. Subtly altering the shape and chemical properties of the peptide can then affect how strongly and selectively they attach to a specific receptor.

Other modifications also need to be made to improve peptide stability in the human body, as this defines how they can be used. Peptide-derived drugs, such as insulin, are usually broken down and digested if swallowed, so they are often only injectable directly into the bloodstream. Making changes to improve peptide stability with respect to digestion raises the possibility of creating an ingestible painkiller, but any changes made can potentially also affect the strength and specificity of the drug binding.

THE SOLUTION

Dr Hung is using molecular modelling and simulation methods to identify the specific binding interactions that define the molecular handshake between conotoxins and their known receptor targets. "We build an atomic model of the toxin and of the receptor in the cell membrane and add all of the water molecules and ions that would be there as well. Fortunately, the structures of the toxins and receptors are already quite well known, so that gives us a good starting point to explore how they behave. We then run molecular dynamics simulations to see how they all move and interact, which parts of the structures contact most closely, and how strong that binding is."

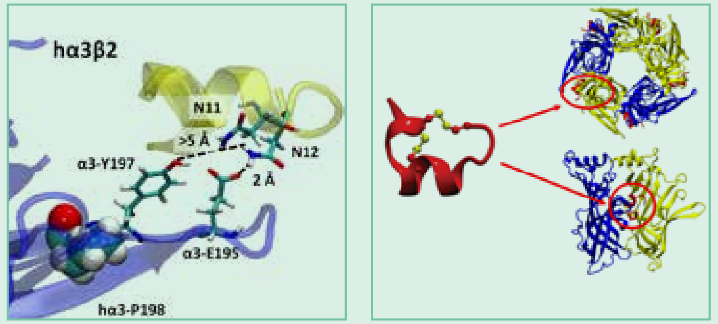
A typical model contains about 300,000 atoms, and atomic movements are tracked for hundreds of nanoseconds, so supercomputing facilities are essential to managing the number of computations involved. Dr Hung explains: "There's a trade-off between how big a system you can model and how long you can run it for, to reasonably see all of the interactions that are likely to happen in nature. And we must run each peptide–receptor system from several slightly different starting positions to make sure we capture all the likely interactions. This then gives us the range of movements that occur in the molecular handshake, the specific points of contact that are important, and how the toxin affects the receptor".

OUTCOME

By understanding the binding mechanism through simulation, Dr Hung can then modify the structure of the peptide models to investigate which modifications lead to more selective or stronger binding.

These results are shared with researchers at the Illawarra Health and Medical Research Institute, led by Professor David Adams. His team creates and tests promising toxin variants in real biological systems to measure the strength and selectivity of their action. Their real-life results then confirm the most promising variations to explore further in the next round of molecular modelling.

"It's an iterative approach to develop a new peptide-based drug," says Dr Hung. "Understanding these structure–activity relationships is letting us create novel peptides with characteristics more and more suitable for therapeutic testing. A stable oral analgesic for chronic pain is the goal."



Improved chronic pain relief treatments Dr Hung, RMIT

PROJECTS AND PUBLICATIONS

1,691 researchers

175 projects

332 publications

36% chemistry and material science

21% radio astronomy

18% engineering and fluid mechanics

During the reporting period, the work of 1,691 researchers and 15 interns across 175 projects translated into 332 research publications showcasing their supercomputing-driven findings and breakthroughs across the scientific and broader community. This number of publications represents an increase of eight percent over the previous reporting period, even considering the limitations that COVID-19 imposed on some researcher projects.

International journals Pawsey researchers have published in include *Nature, Nature Communications, Nature Genetics,* the *Journal of the American Chemical Society, Advanced Materials, Astrophysical Journal,* and *Physical Review.*

Chemistry and material science projects represent the second biggest allocation on the Magnus supercomputer, and translated into the largest group of scientific publications supported by Pawsey allocations, accounting for just over a third of the publication list this financial year. Some project findings, although only recently published, have already been cited more than 20 times in the broader research literature in the areas of solar cells, and catalysts for clean energy generation, nitrogen fixation and other feedstock chemical transformations. As an example of the activity in these fields, 'Transition-metal single atoms anchored on graphdiyne as high-efficiency electrocatalysts for water splitting and oxygen reduction' in *Small Methods* has already been cited 33 times.

Astronomy is another one of Pawsey's focus areas, and radio astronomy projects resulted in 69 papers during the reporting period, representing 21 per cent of the total publication list. The majority of publications were in *Astrophysical Journal,* the *Monthly Notices of the Royal Astronomical Society*, and the *Publications of the Astronomical Society of Australia.*

Investigation of fast radio bursts represented a significant research effort by the radio astronomy community, exemplified by 'Measurement of the rate distribution of the population of repeating fast radio bursts: implications for progenitor models' published in *Astrophysical Journal Letters.* The same team using Australian Square Kilometre Array Pathfinder (ASKAP) subsequently solved a decades-long mystery finding 'missing matter' in a project published as 'A census of baryons in the Universe from localized fast radio bursts' in *Nature.*

Researchers using the GaLactic and Extragalactic All-sky MWA survey (GLEAM) from the Murchison Widefield Array (MWA) have been investigating dark matter ('Searching for dark matter signals from local dwarf spheroidal galaxies at low radio frequencies in the GLEAM survey' in the *Monthly Notices of the Royal Astronomical Society*) and have also discovered 27 new surpernova remnants in the Milky Way using new data from the GLEAM survey.

Engineering and fluid mechanics remain recognised strengths and were well-represented with 59 publications, accounting for 18 per cent of total publications. Energy and resources are also a recognised strength with 30 publications listed, although this is not a true reflection of the size of this research effort, as many of the chemistry and material science publications grouped above have direct application in clean energy developments.

Life sciences is an area supported at Pawsey that is continuing to grow, this year accounting for 16 publications across bioinformatics and the biosciences, and another nine publications directly related to human health outcomes. These have ranged from 'A reference genome for pea provides insight into legume genome evolution' in *Nature Genetics* to 'Maternal smoking during pregnancy induces persistent epigenetic changes into adolescence, independent of postnatal smoke exposure and is associated with cardiometabolic risk' in *Frontiers in Genetics* and 'A computational framework to investigate retinal haemodynamics and tissue stress' in *Biomechanics and Modeling in Mechanobiology.*

The full list of publications arising from research challenges that have been tackled with Pawsey support can be found online at **www.pawsey.org.au.**

Meet the people accelerating scientific breakthroughs at Pawsey

Voices of Science

OUR PEOPLE

Paul Nicholls Board Member

Building research infrastructure and research communities

Although Paul has been Curtin University's representative on the Pawsey Board since 2016, his association with Pawsey extends back to his commencement at Curtin in 2010. As the then Director of Strategic Projects, he was involved in the development of Information Technology (IT) capability at Curtin, particularly its partnership in the supercomputing joint venture iVEC. In 2012/2013 Paul became the interim CEO overseeing the transition of iVEC into a national supercomputing facility, with the construction and commissioning of the $80 million Pawsey Supercomputing Centre.

Since then, as the Director of Research Partnerships at Curtin, he has continued to build research capability in computation and data intensive science at Curtin, alongside its partnership with Pawsey.

"I've historically had a very strong focus on building research communities," says Paul. "It's critical that strong relationships are developed between universities, the State and Federal governments and industry to support investment and the creation of critical research infrastructure such as the Pawsey Supercomputing Centre. And then we need to make sure our research communities can maximise the benefit that can be gained from it."

Having 'sat in the seat' at Pawsey and now chairing the procurement of the current $70 million Capital Refresh, Paul has more detailed insights about the operations at Pawsey than is usual for a Board member. "It's not just buying new equipment and creating a step-change in terms of our supercomputing capability, it's also making sure we support our research communities to transition into making the best use of these new technologies," Paul explains.

Paul sees Pawsey's support for research communities as central to the services Pawsey offers. "We aren't just providing infrastructure and hoping that researchers can use it. At every stage, and particularly during the Capital Refresh, we're being proactive in ensuring the infrastructure we're developing here meets our user community needs, and we're helping our user community to build their skill sets to make the best use of it. Equally, we're always extending the range of research communities we're able to support, most recently in the areas of bioinformatics and space sciences in support of the recently established Australian Space Agency."

From its beginnings with iVEC, Paul is very proud of the work that has been done to position Pawsey as a critical part of our national HPC infrastructure. "It's incumbent upon the Board to ensure we maintain these strong working relationships with other national infrastructure like the National Computational Infrastructure (NCI)."

"As part of the national infrastructure, we now have a federal government commitment to a regular cycle of upgrades like the one we're working through now. Ensuring that Australian researchers have ongoing access to the newest and most powerful supercomputing facilities, whether at Pawsey or the NCI, is really important when the technological capabilities are expanding so rapidly."

Pawsey's strengths also provide the biggest challenges, as Paul notes:

> *"We aren't just providing infrastructure and hoping that researchers can use it. At every stage, and particularly during the Capital Refresh, we're being proactive in ensuring the infrastructure we're developing here meets our user community needs.*
>
> **Paul Nicholls**
> *Board Member*

"There is such a diverse range of research community needs, and HPC technology is moving very fast – matching the technology to the needs of the community in the right timeframe is an ongoing challenge!"

Through his long association with Pawsey, one idea has remained central to Paul's commitment. "It's about providing Australian researchers with a really critical piece of infrastructure to support them in undertaking world leading, transformative research. As an example, Pawsey was intrinsic in turning radio astronomy into a competitive advantage for the State and Australia. Transitioning Pawsey-scale supercomputing to full Square Kilometre Array (SKA) operations with the radio astronomy community will be a huge achievement for us in the next decade. But that's only one example – helping really smart people solve problems faster will ultimately drive benefit for the broader Australian community."

OUR PEOPLE

Professor Alan Mark
Professor, School of Chemistry and Molecular Biosciences, University of Queensland
Area of Science: Simulation of biomolecular systems

Seeing in atomic detail: from molecules to functional cells

Professor Alan Mark studies how biological systems function in atomic detail, from how cell membranes form to how viruses penetrate them and drugs affect them. Since it is not possible to directly observe the properties of individual molecules as they interact with cellular structures, he uses theoretical approaches and computer simulation techniques to build detailed computational models that can predict the properties of real systems.

ABOUT PROFESSOR ALAN MARK

Alan has always wanted to know how living systems work, in particular how different components within cells can spontaneously self-assemble into functional structures. Understanding that complexity means delving into more and more detail, right down to individual atomic interactions.

He studied chemistry and biochemistry before looking at how insulin binds to receptors for his PhD research. He then began working to determine the 3D shapes of proteins using nuclear magnetic resonance imaging. This led him to move to the Netherlands and then Switzerland to join one of the world-leading groups developing molecular simulation techniques to model the key elements of cellular function, such as how proteins fold. Since then he's used computer modelling to explore other critical cellular processes such as how receptors on the surface of a cell recognise specific drugs and how protein misfolding can lead to disease.

WHAT DREW HIM TO SCIENCE?

"I was always interested in science," says Alan. "How do things work? How do leaf cells respond to some combination of temperature and day length to recognise when it's time to fall off a tree in autumn? How do molecules get in and out of cells? It's all just interactions between atoms and molecules. If we can model these interactions correctly, we can understand all aspects of a system. We've progressed a lot but are still incredibly naïve about how biological systems really work at an atomic level."

RESEARCH WITH SUPERCOMPUTERS

Molecular dynamics simulation techniques involve calculating the interactions between every pair of atoms in a particular system. Once all the forces acting on each atom are known, the system can be propagated step-wise through time using Newton's laws of motion. Millions of these small steps in time are combined to make a molecular 'movie' of the system. These movies allow us to see how molecules interact and move, how they find each other and assemble into a working structure.

"The challenge is to describe the interactions between atoms with enough accuracy to be able to examine the properties of interest, and to reproduce measurable properties of real systems," explains Alan. "Biomolecular systems are large and complex, involving up to millions of atoms, and must be modelled over significant timescales to explore all of the possible interactions. It's so computationally intensive it's only possible using a supercomputer."

Alan has spent his career improving the reliability of these models, and validating them wherever possible against experimental data so that their predictions can be used with confidence.

REAL WORLD SOLUTIONS

Alan has used molecular dynamics simulation techniques to demonstrate how membranes self-assemble, how different lipids (the oils that make up cell membranes) contribute to antimicrobial resistance, how receptors get triggered by drugs and how peptides bind to membranes. He and his team have also created a globally-recognised molecular modelling tool and database, the Automated Topology Builder. This is used by researchers around the world to generate atomic interaction parameters for novel molecules as required for computational drug design and the design of new materials.

Alan has previously used Pawsey supercomputing to simulate the fusion-active forms of the Ebola virus and the Hendra virus to study exactly how they infect cells, as well as inform drug design efforts to block their formation. He is currently determining the structure of the fusion-active intermediate of the SARS-CoV-2 protein spike as a potential target for drug design. His team is also working to ensure all researchers wanting to test how existing drugs may be repurposed to treat COVID-19 have ready access to the highest quality interaction parameters possible.

OUR PEOPLE

Dr Chenoa Tremblay
Postdoctoral Fellow in Dark Magnetism at CSIRO
Area of Science: Astrochemistry

Studying the spaces between to understand star formation

Dr Chenoa Tremblay uses modern telescopes and supercomputing power to find and study atoms and molecules in interstellar space. Her astrochemistry observations are improving our understanding of star and galaxy formation, and underpin a search for signs of life beyond our own here on Earth.

ABOUT DR CHENOA TREMBLAY

After university, Chenoa came across the idea of studying chemistry in space, and it has since led her around the world. After completing a degree majoring in chemistry with a minor in mathematics in New Hampshire, she moved to New Mexico and completed more study in mathematics, physics and radio astronomy, while helping to build and test a new radio telescope, the Long Wavelength Array. She joined Curtin University as a research associate in 2011 to study high-mass star formation and then completed a PhD, using the Murchison Widefield Array (MWA) to search for molecules around stars. Now at CSIRO, she is a postdoctoral fellow commissioning and using telescopes around Australia to study the early stages of star formation.

WHAT DREW HER TO SCIENCE?

"I was always interested in maths," recalls Chenoa. "I'd be working on the family farm with my grandfather and he'd always challenge me with maths problems to do in my head." Her broader interest in science was sparked in high school once she could try dedicated subjects like chemistry and physics.

Surprisingly, a career in science was never an ambition. "Growing up in a very small rural community I just didn't have those sorts of role models – science was what you saw smart people doing on TV," Chenoa explains. "I was familiar with farming and small business, so I thought I could funnel my maths interest into accounting and business. But I didn't enjoy business school at all. My university roommate was studying biology and chemistry, and I realised I was spending more time helping her with her homework than doing my own! My friends eventually convinced me I could be a scientist too, so I started studying chemistry at university and absolutely loved it."

RESEARCH WITH SUPERCOMPUTERS

Chenoa uses spectroscopy, the interactions of light and matter, to look for molecules in the gas layers around stars. As molecules in space are bombarded with energy from nearby stars, they can absorb and re-emit that energy at specific frequencies based on their structure – each molecule's energy signature is unique. Chenoa uses radio telescopes to collect signals across a range of narrow-frequency bands to build up and identify those full-spectrum energy signatures.

"That's why we need supercomputers to do this sort of research," explains Chenoa. "We image the sky over a lot of very narrow frequency bands, and process all of that information independently to pinpoint the energy absorptions and emissions to specific frequencies with the resolution we need to identify the molecules out there."

"It's only with this next generation of telescopes like MWA and the Australian Square Kilometre Array Pathfinder (ASKAP) that we can view entire constellations at very high resolution, and the data volumes are massive, so supercomputing is essential to process the information. If I'd had to do my PhD research on a laptop computer, it would have taken me roughly 25 years. Now I'm using Pawsey to create over half a million images per year per project."

REAL WORLD SOLUTIONS

Chenoa is pushing the study of molecules in space to lower radio frequencies than ever before. "In the colder regions of space, much larger molecules should be more stable, and emit at these lower frequencies. It's a very sensitive frequency range to study the chemistry of cold gases, but the emissions are not very intense or energetic, so we need very powerful telescopes like the MWA, ASKAP, and eventually the SKA to find them."

Chenoa's work is defining the relative importance of different mechanisms of star formation, and how they influence how galaxies evolve. She's also laying the groundwork to identify 'biotracers' – more complex molecules like amino acids and proteins in space with the SKA.

"We already know so much about star formation and the molecules in space, but about 85 per cent of the molecules we've found in space are based on carbon," notes Chenoa. "And we don't know why. It may be because stars produce a lot of carbon, and it's easy to make lots of molecules out of it. But we've only been looking in the higher frequencies – are we missing information that could be important in understanding how stars evolve, die, and spread atoms around the galaxy? Carbon may really be centrally important to life in general, but our understanding may be biased because we've only looked in certain ranges of the electromagnetic spectrum. Looking for atoms and molecules at lower radio frequencies could give us a more complete picture of our place in the Universe."

OUR PEOPLE

Professor Phil Bland
2019 WA Premier's Science Awards Scientist of the Year
Director, Space Science and Technology Centre, Curtin University
Area of Science: Planetary science

John Curtin Distinguished Professor Phil Bland tracks meteorites to explore the origin and early evolution of the solar system, and develops systems that can track not only meteorites, but also satellites and space debris.

ABOUT PROFESSOR PHIL BLAND

Previously from Imperial College in London, Phil came to Australia in 2012 and created the Desert Fireball Network. This system of observatories in the Western Australian outback takes photographs of the night sky to find and track fireballs, allowing researchers to identify where meteorites land, and trace their orbits back to work out where they've come from.

As Director of Curtin's Space Science and Technology Centre, Phil now works across a range of projects in space and planetary science, contributing to research teams that are finding and analysing meteorites, studying impact craters on Earth, the Moon and Mars, tracking objects in near-Earth orbit and contributing to space missions.

WHAT DREW HIM TO SCIENCE?

"I've always been into hiking and mountaineering," says Phil. "So the landscape, the geology really fascinated me. I wanted to go to university and learn more about geology. And then in my final year at university I saw the research side of it – I really got 'the bug' for science when I experienced that feeling when you realise you're discovering things that no-one knew before."

Phil had also always had an interest in space and spaceflight, so he made the most of the opportunity to analyse meteorites in his first geological job. That subsequently led to him pursuing a research career in planetary science, combining his original interests in both space and rocks.

RESEARCH WITH SUPERCOMPUTERS

As the Desert Fireball Network developed, the need for large amounts of computer storage became apparent. Phil explains: "We've got 50 cameras in Australia, and now even more world-wide. Each night each camera takes around a thousand 45 megapixel images of the sky – so the images accumulate very fast. Pawsey stores all of that for us, which now accounts for almost two petabytes of data."

The processing power of supercomputers is also being used, with machine learning harnessed to automatically identify meteorite streaks and moving objects between the thousands of images in near real time. Datasets of orbits for objects in our solar system can then be created, and run backwards through time to work out where meteorites originally came from.

REAL WORLD SOLUTIONS

The Desert Fireball Network has been responsible for locating nine meteorites so far, and identifying their point of origin. This is contributing to a geological 'map' of our solar system and increasing understanding of how our planetary system formed.

Based on the experience gained from the Desert Fireball Network, Phil's team is using the same technologies, image analysis and data processing algorithms to track satellites in near-Earth orbit. "There's no equivalent to air traffic control for satellites," explains Phil. "Space is shared, with many countries sending up

Finding falling stars and safeguarding satellites

multiple satellites. The guidelines for tracking and collision warning are not formalised between countries, and different countries use different datasets to map what is up there."

If satellites collide, the cloud of debris generated is likely to cause even further impacts, creating a runaway collision event with the potential to destroy everything in low-Earth orbit. This would profoundly disrupt global communications, navigation and Global Positioning Systems (GPS), weather forecasting and storm tracking, and take trillions of dollars and several years to restore.

Phil's team, in partnership with Lockheed Martin, has shown that they can track orbiting objects as small as a shoebox, analysing the data to give positional details in real time. The satellite tracking system they have built in Australia sees 6,000 square degrees of sky, tracks 1,000 objects and updates their positions every 10 seconds. It's a demonstration that could be extended to a global network capable of tracking the entire catalogue of satellites worldwide and updating their positions every few hours, making space traffic control a possibility.

And it developed from finding falling stars.

OUR PEOPLE

Professor Debra Bernhardt
ARC Laureate Fellow, Australian Institute for Bioengineering and Nanotechnology and School of Chemistry and Molecular Biosciences,
The University of Queensland
Area of Science: Nonequilibrium statistical mechanics, quantum chemistry and molecular simulation

Computational chemistry for clean energy applications

Professor Debra Bernhardt uses a range of theoretical and computational chemistry techniques to study and develop new materials for clean energy applications. From better catalysts for hydrogen fuel cells to more effective electrode and electrolyte materials for rechargeable batteries and supercapacitors, she is helping reduce the limitations of existing clean energy technologies.

ABOUT PROFESSOR DEBRA BERNHARDT

Debra began her career in wave function quantum computation methods, and quickly developed an interest in using computational chemistry approaches to develop a fundamental understanding of the behaviour of matter. Through further study at the University of Basel in Switzerland and then at the Australian National University (ANU) she extended her computational expertise to molecular dynamics simulations and non-equilibrium statistical mechanics to study systems involving fluid flow.

Now with a joint appointment with the Australian Institute for Bioengineering and Nanotechnology and the School of Chemistry and Molecular Biosciences at the University of Queensland, she is combining her expertise in both quantum chemistry and molecular dynamics simulation in clean energy materials research, investigating the energy storage and transfer behaviour of materials at the atomic level.

WHAT DREW HER TO SCIENCE?

"I always really enjoyed chemistry in high school," admits Debra. "Particularly the laboratory experiments – I thought it was amazing to make different chemicals with different colours or smells, or that even gave off light. I was always curious to know why different chemicals behave the way they do."

While she studied a broader range of physical sciences at university, it was the hands-on chemistry laboratories that 'kept her coming back'. Debra remembers: "It's ironic that I ended up being a theoretical chemist – I got busy with the theory and computation while I was waiting a long time for some experimental equipment to arrive for my PhD project, and never looked back."

RESEARCH WITH SUPERCOMPUTERS

"One of the advantages of doing chemistry on a computer is the control you can have on the molecular structure," she explains. "In real physical experiments, your graphite electrode may not have a completely ordered structure, or the distribution and concentration of dopant metals may not be uniform. Using computation, we can make our samples very pure and structurally uniform, we can control the doping and crystalline defects. That allows us to get a much better understanding of exactly which factors affect the behaviour we observe in these materials experimentally."

Computation and simulation are also one of the only ways to study systems while they're operating without disturbing them. "You can't watch a capacitor discharge directly, or see the ions diffuse between electrodes under an applied voltage, for example. So working on energy storage systems, you need input from both experiment and computation. You just get different information, and some of that operational information, from the calculations."

"Supercomputing lets us do our calculations at even higher levels of theory, and on bigger systems – for example, you can only look at electrode doping concentrations as low as one in a million if you're simulating over a million atoms to start with. Faster computations also let us follow systems for longer periods of time, so we can see double layers form and discharge on capacitors, or the slow diffusion of charge carriers through very viscous ionic liquid electrolytes."

REAL WORLD SOLUTIONS

The challenge is to create better energy storage and transfer options for renewable energy sources.

"Looking at the binding energy of atoms to electrodes can tell us about a battery's capacity, whereas measuring the diffusion coefficient or conductivity can tell us how fast we can charge and discharge it. If we see the material structure start to distort over several charge/discharge cycles, it corresponds to batteries that lose capacity and stop working over time, as their components irreversibly break down at the atomic level."

Debra continues to be driven to solve problems and do things better, from developing new theories to predict outcomes of processes, to optimising computer code, so her simulations at Pawsey run faster and using them to develop cheaper, more stable and more efficient materials for batteries and supercapacitors.

OUR PEOPLE

Associate Professor Nicola Armstrong
Associate Professor, Mathematics and Statistics, Murdoch University
Area of Science: Statistical bioinformatics

Sifting through the pieces to pinpoint the origins of disease

Associate Professor Nicola Armstrong analyses genetic data to help medical researchers understand the genetic basis of diseases like cancer and degenerative disorders like dementia.

ABOUT ASSOCIATE PROFESSOR NICOLA ARMSTRONG

Nicola brings meaning to data, mostly coming from medical research. With a background in theoretical and applied statistics, she undertook her PhD at the University of California in Berkeley during the late 1990s when new genetics technologies were revolutionising our understanding of biology and the human genome. Being immersed in this rapidly developing medical research environment, she could see real-world applications of statistics, integrating computational and statistical methods into experimental and clinical research that was generating increasing volumes of novel data using new genetics technologies.

Now at Murdoch University, Nicola develops and applies statistical methods for the analysis and integration of high-throughput genomic and epigenomic data to understand how the development of complex diseases and ageing processes vary between individuals.

WHAT DREW HER TO SCIENCE?

"I've always been good with numbers, and enjoyed playing with them," remembers Nicola. "And I had a really good maths teacher in high school, who made the subject really enjoyable and interesting. When I went to university, I actually wanted to be an actuary – looking at the statistics of risk and uncertainty that underpins insurance – but I couldn't study it locally at the time. So I studied maths and statistics, which was the next best thing."

"Now, my research is enabling other people to work on really amazing medical problems. I love that statistical bioinformatics lets me pull together information and experts from lot of other scientific disciplines to try and make a difference to people's health outcomes."

RESEARCH WITH SUPERCOMPUTERS

Nicola explains how genetics research has changed since the 1990s: "We might have done controlled experiments with 200 mice, and amplified particular sections of their DNA to study very specific regions of their genome. It was slow, and the data produced was limited and very specific. But now with next-generation sequencing technology, we can sequence the entire genome of a mouse overnight – we can correlate the onset of a disease with any genetic differences we see genome-wide."

"Now the issue is these next-generation, or massively parallel, sequencing technologies give you millions of very small fragments of DNA overnight that still need to be pieced back together and mapped to the complete reference genome, whether mouse or human. That 'assembly' or processing of the genetic data requires a supercomputer, especially if you're looking at 500 whole genome sequences from a large study."

Nicola is also now looking at epigenetic data on a large scale – evidence of which particular genes are being turned on and off under different conditions to cause changes in certain cells. It's yet another aspect to cell regulation and function that can now be measured at a genome-wide scale, with the resulting data then integrated with all of the genetic, diagnostic and clinical information that makes up a modern medical study.

REAL WORLD SOLUTIONS

Nicola's current research at Pawsey is working to identify individuals at high risk of developing dementia, particularly late-onset Alzheimer's disease. The cascade of events that lead to dementia currently begin many years before the disease is clinically diagnosed, and may only be tentatively identified through magnetic resonance imaging of the brain. Even if structural changes in the brain are seen, the underlying causes of dementia are still not comprehensively understood.

As part of an international genetics of brain imaging project involving over 25,000 individuals across the US, Europe and Australia, Nicola is exploring the genetic and epigenetic variations that correlate with structural brain alterations identified through non-invasive brain scanning in people prior to the onset of clinical symptoms.

"The genetic data should let us identify people that might be at risk later," notes Nicola. "But the epigenetic data, looking at how the DNA is modified and which genes are expressed as a result of both an individual's genetic makeup at birth and the effects of environmental factors, that information is potentially what will underpin future treatment options. Exploring the genes expressed in the development of dementia is an essential step for identifying therapeutic drug targets or lifestyle interventions."

OUR PEOPLE

Turning information into knowledge

Dr Yathu Sivarajah
Pawsey Team Lead, Senior Visualisation Specialist

Dr Yathu Sivarajah began his research career in geophysics, focusing on how people interact with complex datasets, and particularly in quantifying human–data interactions to understand the techniques used by experts in analysing complex datasets. He subsequently joined the Pawsey visualisation team, and worked on a range of projects supporting remote visualisation, large-scale display systems and virtual reality. Yathu now leads the visualisation team, and helps other researchers with their large-scale scientific data visualisations.

"It's one thing to generate a large dataset with lots of information about a particular research problem," says Yathu, "but as the datasets get bigger, understanding what they really mean – seeing the patterns, the trends, the connections – becomes more difficult. The visualisation team helps our users gain insights from their data, which helps them to understand what is really going on in their research. Good visualisations also help the researchers present their findings in a clear and understandable form to others."

Working with experts from a range of research domains and often crossing disciplinary boundaries, it is a regular challenge for Yathu to learn about users' specific research and data problems, and understand their visualisation needs. He then draws on techniques developed across traditionally data-intensive domains to create useful representations of their data for analysis.

"Almost all of our projects require cross-team collaboration across Pawsey. It enables us to achieve the best outcomes for our users, but at the same time helps us to share knowledge and learn from the many other experts within Pawsey. My team may know a lot about data visualisation, but we can also tap into expertise here about the characteristics, structure and subtleties of bioinformatics data, or radio astronomical data, or dynamic modelling data, for example."

Each day brings different visualisation challenges, admits Yathu, but that's what he likes the most about working at Pawsey: "It's really rewarding to be a collaborative part of so many different research outcomes. Especially when I can work with a researcher who hasn't traditionally used HPC, and help them gain so much more insight from their data."

Yathu is particularly proud of the development of Pawsey's web-based remote visualisation service. "Our users can now remotely visualise their large-scale data via the web, without having to copy the data to their local computer. Using our dedicated nodes and high-end graphics cards, they can use their laptops, and even their mobile devices, to explore their data."

Yathu is seeing firsthand how HPC is diversifying into new areas of research across Australia, and can only see this progressing into a much broader day-to-day uptake of HPC across the wider research community. He looks forward to helping even more researchers understand and interpret the wealth of information they now have at their fingertips. "Between developing remote visualisation, visualisation on large-scale display systems, and continuing to explore how people can best interact with these massive datasets, I'll be staying just as busy as our researchers are, extracting meaning from 'the matrix'."

Ann Backhaus
Pawsey Education and Training Manager

Developing knowledge and skills to harness HPC for research

Ann Backhaus has a passion for lifelong learning. Ann lives this passion, as exampled by her doing a 'refresher' Teaching and Learning Graduate Certificate "for fun" before joining Pawsey in 2019, even though she already had previous degrees as well as 25 years of experience in training and talent development in private industry and academia. "I enjoy teaching and learning at all levels and in all forms, from designing programs to developing content across a range of platforms," says Ann.

Since Pawsey exists to enable science and accelerate discovery, skill development and knowledge building are central to its vision and purpose. It develops the expertise Pawsey staff need to run its ever-evolving infrastructure, and builds the skills and confidence of its researchers to use Pawsey resources effectively to do more ambitious research, more quickly.

As a result, Ann works with everyone. She helps individual staff map out and plan their professional development, works with entire teams to upskill and strengthen capabilities, and implements Pawsey-wide educational programs supporting change management for new operations. Many learning opportunities have arisen this year due to the current Capital Refresh of Pawsey's supercomputing systems.

Ann also collaborates with researchers and users from other organisations, creating learning journeys for them that are REAL (relevant, engaging, meaningful and learner-centred), to help them do the best research they can with Pawsey facilities. I may not know the specifics of the many research domains and technologies that intersect at Pawsey," explains Ann, "but I do know about teaching and learning, and I enjoy working closely with these individuals to help them to build the skills and knowledge to use Pawsey in the most direct and meaningful way possible".

"There's never enough time in the day to explore, develop, deliver or collaborate," admits Ann. "My days are always varied and interesting. I work with colleagues to design learning events, and work with Subject Matter Experts to develop training content. I work with our new equipment vendors to design training programs for staff and Pawsey users, and help researchers develop learning outcomes for our Pawsey Summer Interns. Further afield, there's collaborations with Pawsey partners to ignite interest in STEAM subjects in the next generation of students, and partnerships with national and international colleagues to develop content, programs, conference presentations and workshops."

"I'll never, ever, ever tire of the 'aha!' moments that happen with learning," enthuses Ann. "I love seeing 'learning' progress, from the introduction of a topic all the way through to applying that learning in real life, whether it's a researcher seeing first-hand a tenfold increase in productivity using a workflow manager or Graphics Processing Units (GPUs), or an intern confidently presenting research outcomes about a topic they knew nothing about 10 weeks prior."

Ann's own learning journey is also continuing at Pawsey: "There's never a dull moment. With all of the amazing people working and using the facilities here, every day I learn something new about a research domain or a technology, I try something new, or I think about things in a new way. Pawsey staff and users are figuratively (and sometimes literally) reaching for the stars, and that is inspiring each and every day."

OUR PEOPLE

Translating a user focus into seamless connections for researchers

Dr Lachlan Campbell
Pawsey Research Data Specialist

Dr Lachlan Campbell likes problem solving and helping other people solve problems. At Pawsey, he does that every day. With a research background in astronomy, Lachlan originally joined Pawsey as part of the Astronomy Data and Computing Services program to provide astronomy-specific training to help astronomers access Australia's data and computing infrastructure for their research. Now as a full-time member of Pawsey's services team, he is a vital link connecting astronomy researchers to Pawsey.

"I'm not doing astronomy research, although I've done that in the past and I know how to talk to researchers about what they need and what they're trying to do," says Lachlan. "I'm also not a supercomputing or data specialist actively assisting with coding and setting up workflows, but I know how they work and I know what is possible at Pawsey. I basically coordinate and bridge the gaps between the two groups."

Lachlan is essentially a facilitator, managing regular meetings with the astronomers using Pawsey facilities in their work with the MWA and ASKAP. "Communication is the key," says Lachlan. "For example, our astronomers have been using our distributed filesystem Lustre at Pawsey for at least the last six years. We might assume they know exactly how to use it and interpret the filesystem reports, but in reality, they often have a generalised understanding that allows them to get on with the research they want to do, but don't necessarily know how to use it in the best way from a computational point of view. It can cause frustration and inefficiency on both sides. So, I worked with the astronomers to gauge the level of their knowledge and identify the gaps, and how best to fill those gaps. It's about getting the right information to people in the right form for them so we can optimise how we work together."

Facilitation means communication, but it often also means training. Lachlan also develops training resources for people using Pawsey facilities, most recently for the Nimbus cloud service. While the regular face-to-face training 'road trips' around Australia are paused, even more effort is being put into online training.

"Giving training to Nimbus users, I interact with all sorts of researchers, not just astronomers. I still love science and research, so I relish the challenge of very quickly getting to a reasonable understanding of a particular subject, being able to ask sensible questions about it, and then helping bring that user focus into what we do at Pawsey, tailoring our training and our systems to their needs."

"I've got the best of both worlds at Pawsey because I see what we do from both the internal and external perspectives. It combines my loves of teaching and assisting people, while still being able to immerse myself in the research and stay involved in astronomy."

Audrey Stott
Pawsey Bioinformatics Platforms Specialist

Building supercomputing into bioinformatics

As Pawsey's first bioinformatics platforms specialist, Audrey Stott is attuned to what biologists need to support their research, and how to help them get it. With an established background in human biology, podiatry and clinical research, Audrey moved into biomedical research by completing a Masters degree in biomedical science. This exposed her to supercomputing for biological analyses, particularly bioinformatics and data analysis for genomics research, and subsequently became a turning point for her career.

Audrey now builds, optimises and deploys containers for bioinformatics, deploys specialist bioinformatics software on HPC and the cloud, develops instructions for using HPC, and collaborates with researchers needing larger and faster data processing. "I assist researchers in installing the software they need, getting workflows written and running, and help them make full use of the resources they require," Audrey says.

Helping biologists process their research data in the best and most efficient way possible means that Audrey works with experts across Pawsey. "There's fairly constant two-way communication with the help desk staff and the infrastructure and system administrators to build on their expertise in hardware information and software deployment," she elaborates. "I also work with the education and training team to develop training material, documentation, webinars and workshops for researchers wanting to use Pawsey resources. Our bioinformatics team works across every aspect of raising awareness with biologists and making it easier for them to use HPC in their research."

"The easier and faster it is for them to process their data, the more time they can spend on experimentation and writing, and the sooner they can get their research out to where it can make a difference."

Audrey considers the relationships with her colleagues and maintaining strong research communities of Pawsey users as one of the best parts of her job. "There's a great work culture here, everyone is very supportive and cohesive, and we are all about looking outwards and helping the researchers. It's really exciting to be building supportive and dedicated expertise for the bioinformatics and biology domains, and collaborating with both the research and HPC communities to optimise HPC resources for their needs."

Job satisfaction comes from helping more bioinformatics users feel more equipped and confident in approaching and using HPC. And although Audrey has been with Pawsey for less than a year, her ambitions for the future continue to grow. "Just like we're doing for bioinformatics, I'd love to get to the point where every scientific domain is supported at Pawsey by a dedicated team with a structured way to use the infrastructure and resources that suit the specific research needs. Every researcher in Australia should be able to confidently sign up to use HPC to upscale their research."

Growing to upscale our scientific ambitions, outcomes and impact

A New Era of Research

PAWSEY CAPITAL REFRESH

To ensure the Pawsey Centre can continue to provide internationally-competitive supercomputing and data services, a $70 million capital refresh project funded by the Federal Government began in 2018 to secure the next generation of supercomputers, data and supporting infrastructure for Australia's research communities.

This upgrade is allowing Pawsey to keep pace with global advances in supercomputing technology, and provide significantly increased compute power and speed, expanded data storage capacity, and more efficient high-performance computing services for Pawsey users.

This complex upgrade is being undertaken as a staged process. In the previous financial year, additional tape storage was installed to expand the existing tape libraries for long-term storage, and four procurements were begun to source high-speed storage, long-term storage and ancillary systems.

DELIVERED
VENDOR SELECTED
EVALUATION

01 GOVERNANCE

Project Manager appointed
Reference groups established
Engagement strategy planned
Procurement plan updated.

02 TAPE STORAGE

Additional tape storage has been procured to expand the existing tape libraries from 50 to 63 Petabytes in each library.

03 ASTRONOMY HIGH SPEED STORAGE

3x more storage and performance. The existing Astro filesystem was expanded to service the MWA community. Powered by HPE, it has been upgraded to 2.7 PB of usable space and capable of reading/writing at 30 GB/s. The New buffer filesystem, a dedicated resource for ASKAP researchers, provides 3.7 PB of usable space and is capable of reading/writing at 40 GB/s. It is manufactured by Dell.

04 CLOUD/HIGH THROUGHPUT COMPUTING

Pawsey partnered with Dell EMC to expand its cloud system with 5x more memory and 25x more storage to form a cutting-edge flexible compute system. This expansion provides better service to emerging research areas and communities who benefit more from a high throughput compute.

05 MWA CLUSTER

Garrawarla, the 546 TeraFlops MWA cluster, is a resource tuned to MWA's needs, powered by HPE. Procured ahead of the Main Supercomputer, this cluster allows ASKAP to use the full CPU partition of Galaxy.

06 HIGH-SPEED STORAGE FILESYSTEMS

Designed to deal with thousands of users accessing them at the same time. The Pawsey high speed filesystems will be procured as part of the main supercomputer system to increase speed and storage capability to general purpose science.

07 NETWORK INTERCONNECT

Pawsey is moving to a CISCO spine-leaf architecture with a 400Gbps backbone and 100 Gbps links to host endpoints. The network has been designed to be easily expandable to support the object storage platform being purchased as part of the Long-Term Storage procurement as well as integration with the new Pawsey supercomputer.

08 LONG TERM STORAGE

Server and hard disk storage and networking will be refreshed, along with the tape library total maximum storage capacity. Both ingest and egress data transfer services will be upgraded, improving upload and download times.

09 PAWSEY SUPERCOMPUTER SYSTEM

PSS will be built using HPE Cray EX supercomputer architecture, will deliver 30x more compute power than its predecessors and will be at least 10x more power efficient.

NEW HIGH THROUGHPUT COMPUTE AVAILABLE

The upgrade to Pawsey's Nimbus cloud infrastructure was awarded to Dell Technologies in Q2 2019. The expansion of the existing cloud system represents three times more memory and ten times more storage and provides improved computational flexibility, accessibility and speed. The new system features 58 compute nodes supporting up to 3,700 cores, 9 PB of storage, 58 TB of RAM and 100 Gb ethernet networking.

The upgrade enables researchers to process and analyse large amounts of data through additional object storage and the Kubernetes container orchestrator, building on Pawsey's existing container technology for its HPC systems. The system was tested and evaluated by researchers during Q1 2020 and was opened generally to Australian researchers in April 2020.

ACCELERATING UNDERSTANDING OF THE UNIVERSE

The new Astronomy high-speed storage was delivered in two parts. In August 2019 the Astronomy Filesystem, now dedicated to the Murchison Widefield Array (MWA), was upgraded from 1.8 PB to 2.7 PB of usable space and is capable of reading/writing at 30 GB/s. This was followed by the procurement of a new Lustre filesystem, Buffer, to become a dedicated resource for Australian Square Kilometre Array Pathfinder (ASKAP) operations. This provides 3.7 PB of usable space and is capable of reading/writing at 40 GB/s. Together these procurements have tripled both the capacity and performance for Australia's two SKA pathfinder telescopes.

Pawsey's Galaxy supercomputer previously serviced the computational demands of both the MWA and ASKAP, but with the data processing needs of both instruments growing; a new MWA compute cluster was procured ahead of the main supercomputing system at Pawsey, providing the latest generation technologies tailored for their workflows. After a thorough consultation with key stakeholders to determine the best system to respond to the specific needs of MWA telescope users, the tender went to market in Q3 2019, and HPE was selected to deliver the $2 million compute cluster in February 2020.

The new 546 TeraFlops 78-node cluster will be dedicated to processing in excess of 30 PB of MWA telescope data. It features 156 of the latest generation Intel Central Processing Units (CPUs) and 78 cutting-edge Graphics Processing Units (GPUs) with more high-bandwidth memory, internal high-speed storage and more memory per node. 100 Gb/s of bandwidth is available per node for accessing data on the global Lustre filesystems.

Stress testing of the installed system is scheduled for August 2020, and general MWA users will commence migration to the new cluster at the end of Q3 2020. The name chosen for the new cluster is 'Garrawarla', meaning spider in the Wajarri language; whose land the MWA sits on. This reflects the appearance of the MWA antennas, which the Wajarri people have likened to spiders on the landscape.

A NEW SCALE FOR GENERAL-PURPOSE RESEARCH

The tender for the new Pawsey Supercomputing System was released in November 2019. It will replace the functionality of Magnus, Galaxy and Zeus supercomputers in a single system, providing one large-scale, balanced High Performance Computing (HPC) system to support the broad scope of research users at Pawsey.

The Statement of Requirements for the system went to market in Q4 2019 and evaluation of tenders is progressing. Delivery of the new system is expected to occur in two phases.

Phase one, due to be commissioned in mid-2021, will provide researchers with a system that is at least equivalent in capacity to what they are currently using, with the latest generation of processors and increased memory per node. During this phase, researchers with an active allocation Magnus will transition to the new system.

Phase two is expected to be in production by mid-2022. It will provide an exponential expansion in capacity and state-of-the-art technology.

With the upgrade, Pawsey will allow Australia's researchers to remain globally competitive, upscale their research ambitions and accelerate scientific discovery.

The final pieces of the puzzle, the replacement of the Long Term Storage system, firewall and network backbone will be procured and installed during the next financial year.

SYSTEMS AND SERVICES

Allocation in numbers

Almost 500 million
hours were requested across all allocation schemes

275 million
core hours were allocated

69 projects
were granted a National Computational Merit Allocation Scheme (NCMAS) allocation

76 projects
accessed Magnus via Partner Allocation

33 projects
received an allocation via Energy and Resources scheme

ACCESS TO PAWSEY SUPERCOMPUTERS

Pawsey provides supercomputing access to researchers predominantly via competitive merit allocation schemes. Through these merit allocation schemes, the Centre strives to:

- maximise the overall research impact of Pawsey supercomputing resources
- promote scientific advantage in priority domains such as radio astronomy, energy and resources
- provide leading-edge supercomputing resources for researchers in Pawsey partner institutions
- enable wider adoption of, and benefit from supercomputing across Australia.

The Galaxy supercomputer is only available for radio astronomy-focused operations. Specifically, it is used to support ASKAP and MWA operations, which are two of the Square Kilometre Array (SKA) precursor projects currently underway in the north-west of Western Australia. For ASKAP, Galaxy acts as a real-time computer and is a part of the telescope, allowing direct processing of data delivered to the Pawsey Centre from the Murchison Radio Observatory.

RESOURCE ALLOCATIONS

Over 487 million core hours were requested on the supercomputer Magnus during the period, almost twice as many hours as were available. 275 million core hours were allocated.

As part of the 2020 allocation, released in January 2020, 178 projects received an allocation on Pawsey supercomputers.

Supercomputer Allocation Scheme

THE NATIONAL COMPUTATIONAL MERIT ALLOCATION SCHEME (NCMAS)

Twenty-five per cent of the supercomputer time available at Pawsey is allocated through this scheme, along with compute-time allocations from other Australian HPC centres including the National Computing Infrastructure (NCI), the University of Queensland, and Monash University. The scheme operates annual allocation calls, which are open to the Australian research community and which provide substantial amounts of compute time for meritorious, computational-research projects.

THE RADIO ASTRONOMY SCHEME

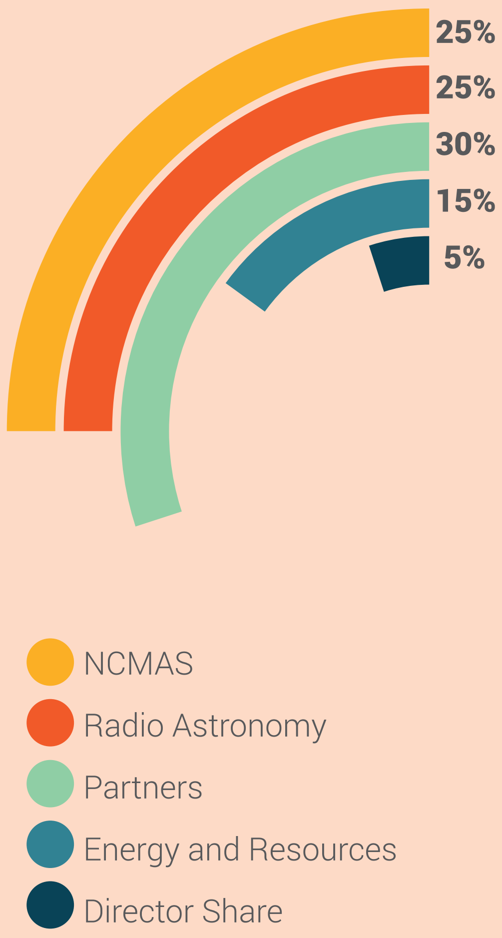
Galaxy supercomputer, representing 25 per cent of Pawsey's total HPC resources, are fully allocated to support ASKAP and MWA operations. During the reporting period a new cluster named Garrawarla was procured as part of the $70m Capital Refresh project, and it is still to be commissioned. When available it will provide a dedicated system for astronomers to process in excess of 30 PB – equal to 399 years of high definition video – of MWA telescope data using Pawsey infrastructure.

THE PAWSEY PARTNER MERIT ALLOCATION SCHEME

Thirty percent of the whole supercomputer time available at the Centre is allocated through this scheme. Calls are open to researchers in Pawsey Partner institutions – CSIRO, Curtin University, Edith Cowan University, Murdoch University and The University of Western Australia – and provide significant amounts of compute time for meritorious computational research projects. This scheme operates annual calls, with an out-of-session application process for newly eligible project leaders.

THE PAWSEY ENERGY AND RESOURCES MERIT ALLOCATION SCHEME

Fifteen percent of Pawsey's total supercomputing resource is allocated through this scheme. Open to the Australian energy and resources research community, it provides significant amounts of compute time for meritorious research projects in these domains. Energy examples include generation, storage and distribution, while resources examples include exploration, minerals extraction and processing, groundwater, and waste management. This scheme operates annual calls, with an out-of-session application process for newly eligible project leaders.



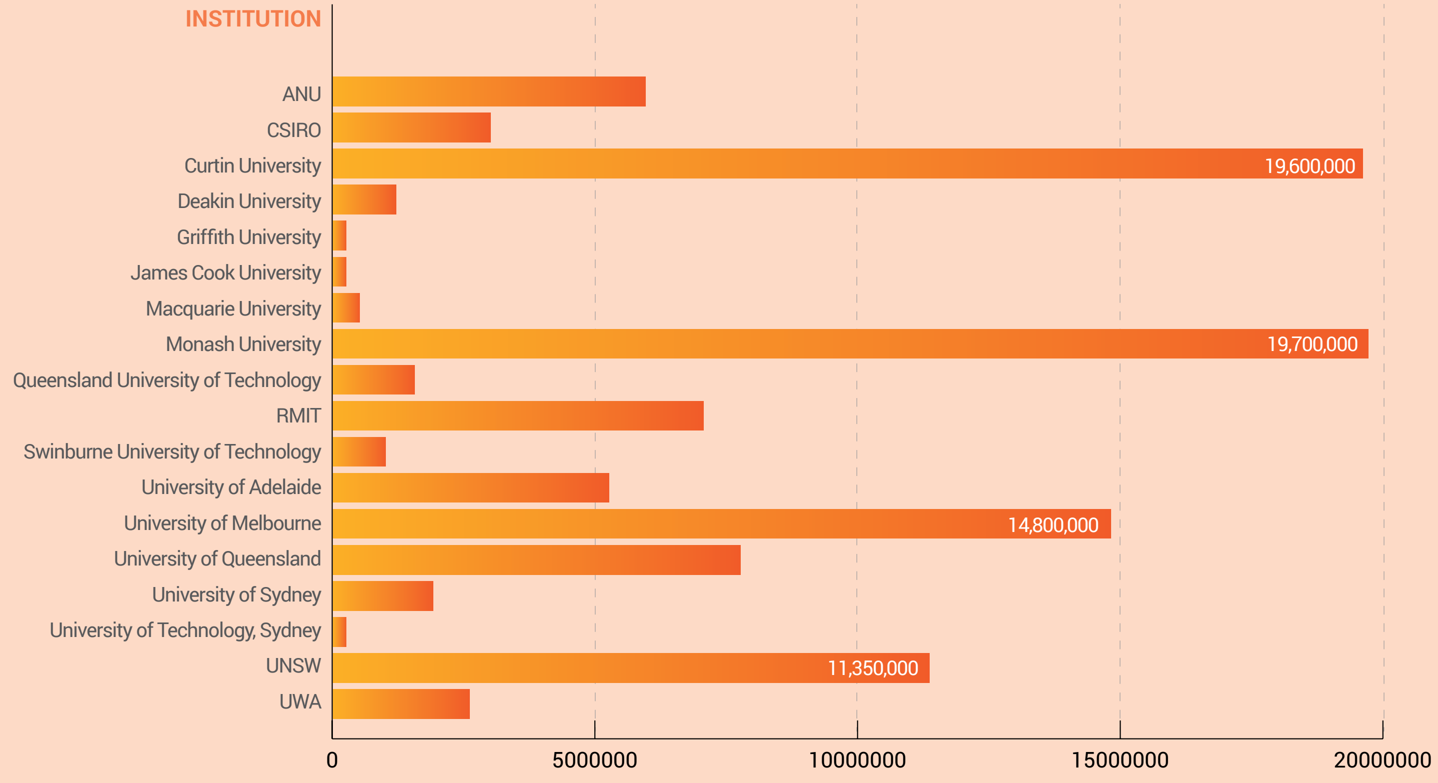
National research facility

The Pawsey Supercomputing Centre is a Tier 1 supercomputer facility, and supports researchers at universities across the nation throughout the NCMAS meritorious allocation. The Centre provides 100 million core hours to NCMAS for allocation to national research projects every year.

During the reporting period, the top four institutions benefiting from Pawsey's contribution to the national allocation resources were Monash University, followed by Curtin University, the University of Melbourne, and the University of New South Wales.

National access to Pawsey resources is also provided via the Energy and Resources Merit Allocation Scheme and the Pawsey Director's share.

NCMAS ALLOCATION BY INSTITUTION



Engineering
38.6%
core hours used in Magnus

Physics
16.9%
core hours used in Magnus

Chemistry
16.9%
core hours used in Magnus

Earth science
16.6%
core hours used in Magnus

Supporting Australian Science Priorities

The Centre supports the National and State Science Priorities. Among the 178 projects that Pawsey supercomputers supported for the 2020 calendar year, Magnus uptake for engineering projects remained the same as previous years, accounting for 38.6 per cent of its use. This year the gap between physical, earth and chemical sciences reduced, with physics and earth science projects increasing to represent 16.9 and 16.6 per cent of the Magnus total allocation respectively, while chemistry projects reduced six per cent compared to last year's allocation. It is still the second-most allocated area of science, accounting for 16.9 per cent of usage.

Galaxy is Pawsey's real-time radio astronomy supercomputer, connected to the telescopes in the Murchison, and its usage is 100 per cent consumed by radio astronomy projects. As reported as part of the Capital Refresh project, Garrawarla, a new GPU-based supercomputer, was delivered to expand radio astronomy support for the MWA which when available will provide a better tuned system for that community. This will also allow researchers using ASKAP telescopes to take full advantage of the Galaxy supercomputer, in preparation for ASKAP approaching its full surveys of the sky milestone.

During the period there were several significant system upgrades and deployments for radio astronomy. Most of those deployments were part of the Capital Refresh project that the Centre is currently undertaking and have been previously reported.

Topaz, a new GPU expansion for the Zeus system, was deployed for users in early 2020 and has seen enthusiastic uptake.

Supercomputer Specifications

GALAXY

- Peak performance in excess of 200 TeraFLOPS
- Real-time system for the SKA pathfinders and radio astronomy projects
- Cray XC30 system with over 9,000 Intel processor cores
- Over 30 TB of memory
- 64 NVIDIA Kepler K20X GPUs.

MAGNUS

- World-class supercomputer in excess of 1 PetaFLOP
- Cray XC40 featuring the 72 Gb/s Cray Aries interconnect
- Over 35,000 Intel 'Haswell' processor cores
- 3 PB of scratch file storage
- Over 95 TB of memory.

GARRAWARLA

- Peak performance in excess of 750 TeraFLOPS
- 78 compute nodes with a total of 3120 Intel 'Cascade Lake' cores, 78 NVIDIA Tesla V100 Tensor Core 32 GB GPUs and almost 30 TB of memory
- Fat Tree network topology with InfiniBand HDR100 100 Gb ConnectX-6
- Local storage: 240 GB SSD boot drive and 960 GB NVMe drive per node.

ZEUS

- Manufactured by HPE/SGI (USA)
- 92 nodes with 28 'Broadwell' Intel Xeon cores with a 100 Gb/s Omni-Path interconnect and 128 GB of memory per compute node
- 80 nodes with 64 Intel Xeon Phi 7210 cores with a 100 Gb/s Omni-Path interconnect and 96GB of memory per compute node
- 11 nodes with four NVIDIA Tesla P100 SXM2 GPUs with a 100 Gb/s InfiniBand interconnect and 256 GB of memory per compute node
- Six nodes with 16 'Broadwell' Intel Xeon cores and 1 TB of memory per compute node
- SUSE Linux (SLES) 12 operating system.

TOPAZ (ZEUS EXPANSION)

- Peak performance in excess of 330 TeraFLOPS
- 22 compute nodes with a total of 320 Intel 'Cascade Lake' processor cores, 40 NVIDIA Tesla V100 16 GB GPUs and 4.6 TB of memory
- 20 visualisation nodes each with NVIDIA QUADRO RTX5000 16 GB GPUs
- Interconnect based on Mellanox ConnectX-6 HDR100 (100Gb/s) InfiniBand.

OTHER SYSTEM SPECIFICATIONS

TAPE LIBRARIES

- Up to 100 PB of storage
- Two duplicate libraries for added resilience
- Connected at up to 40 GB/s.

NIMBUS CLOUD

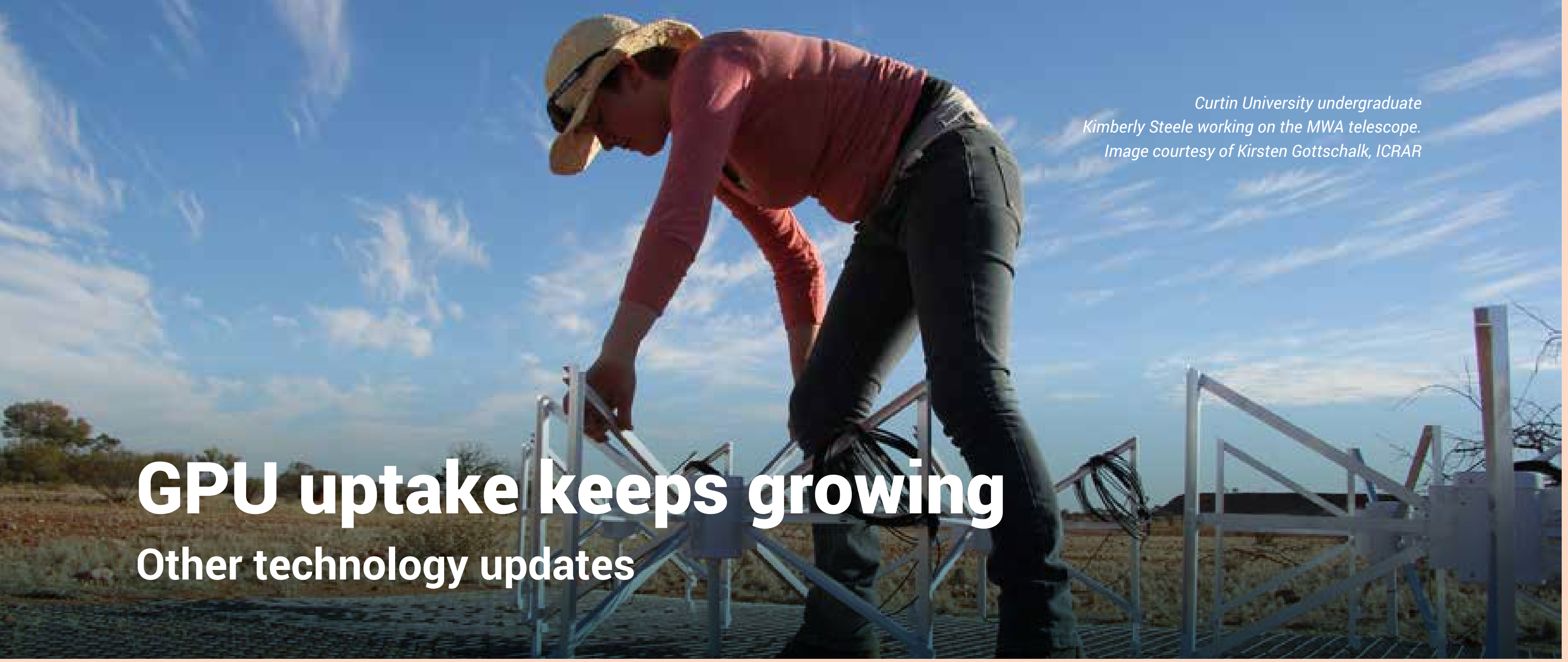
- 3,700 cores and 58 TB of RAM across 58 compute nodes
- 12 Nvidia V100 GPU nodes
- 9 PB of volume storage
- 100 Gb Ethernet networking.

VISUALISATION WORKSTATIONS

- Two workstations, Intel(R) Xeon(R) CPU E5-2697 V2 @ 2.70 GHz
- Two processors
- 256 GB RAM
- Two GPU components; Quadro K5000 and Tesla K20c
- 512 GB SSD
- 20 remote visualisation Topaz nodes, dual Intel Xeon 2.5GHz 8 core processors.

NETWORK

- 50 devices supporting in excess of 6,000 ports
- ~1,400 non-blocking 10 Gbps ports
- ~70 non-blocking 100 Gbps ports
- The remainder are 1 Gbps ports, mainly used for management and monitoring purposes, and staff access.



Curtin University undergraduate Kimberly Steele working on the MWA telescope. Image courtesy of Kirsten Gottschalk, ICRAR

GPU uptake keeps growing

Other technology updates

Pawsey staff and researchers have been working together during the past few years to understand the opportunities GPU accelerators represent for Australian researchers.

Historically, Pawsey offers to researchers have been CPU heavy, as has the code researchers have been using. It is now a global trend that GPU uptake is increasing among research facilities, as the technology has demonstrated to be both cost-effective and energy efficient.

As a result of the opportunities GPU accelerators offer, Pawsey continues to provide test beds for Australian researchers to develop their skills and understand its benefits. During the reporting period, Topaz became available, expanding Pawsey's GPU capabilities. Topaz is complementary to the existing Zeus cluster, with Topaz providing researchers with enhanced visualisation and GPU compute capabilities and increased batch processing speeds.

Topaz consists of 42 nodes, and its peak computational speed exceeds 330 teraflops.

Making new technology available is an important part of what Pawsey does, but it is equally important that users can take advantage of it. Pawsey staff work closely with their research communities to help them make the most out of what is provided.

This year, Pawsey expertise supported researchers to port and optimise their codes for GPU architecture. Among them, researchers from MWA received Pawsey expert support with porting codes to GPU architecture and Curtin University researchers developed a new GPU-enabled version of their Convergent Close Coupling (CCC) code which has seen an impressive speedup.

Several research projects were migrated to Topaz with Pawsey support, and planning for MWA researchers to migrate to Garrawarla has begun.

Among the Topaz allocation for the year were researchers fighting COVID-19 using molecular modelling of virus targets to help repurpose existing research on potential antiviral effects, using the molecular dynamics package Gromacs.

Containerisation technologies played an important role during this period. Containerisation of software and research workflows have been identified as a technique that can enable research communities to readily leverage HPC and the power provided through its effective use. Pawsey is currently supporting OpenFOAM containers.

Opportunities to also work with Kubernetes container orchestration, building on Pawsey's existing container technology for its HPC systems, were identified with the deployment of the refreshed Nimbus cloud. Kubernetes is a widely-used general-purpose container orchestration solution for automating application deployment, scaling, and management. It has evolved to result in a powerful solution for HPC and Artificial Intelligence (AI) workflows, suitable to complement and interact with the current supercomputing and HPC capabilities of Pawsey.

To extend the existing services to more straightforward and more contemporary frameworks and workflow managers, the operations and end-user support teams at Pawsey were trained in the Kubernetes open-source container orchestration system.

NIMBUS CLOUD

During the period, as part of the Pawsey Capital Refresh, the Nimbus Cloud was refreshed and now contains approximately 3,700 cores with 9 PB of storage, 58 TB of RAM, 12 Nvidia V100 GPU nodes, and 100 Gb Ethernet networking. The new high-throughput computing infrastructure represents a significant improvement from the original Nimbus Cloud in CPU, memory and storage, providing enhanced computational flexibility, accessibility and speed.

Nimbus Cloud allows researchers to process and analyse larger volumes of data through additional storage and opportunities to work with Kubernetes container orchestration, building on Pawsey's existing container technology for its HPC systems. Researchers with memory-heavy workloads will find the new infrastructure better suited to their research needs, with the capacity to support more computationally challenging projects.

Nimbus has seen significantly increased uptake during the reporting period with 113 different projects benefiting from its upgraded version.

The cloud has been used as a stepping stone for researchers who find their computing capacity is not enough but have not identified a need for supercomputers or don't know how to use them. In some cases, Nimbus has been used by researchers as a resource additional to Magnus. Nimbus uptake from bioinformaticians has increased as the GPU technology included as part of the system has been demonstrated as extremely valuable to running their workflows as well as the container orchestration available.

DATA STORAGE INFRASTRUCTURE

As part of the data services provided at the Centre, Pawsey hosts a series of significant data research collections from national universities and government departments across Australia.

Pawsey procured 26 Petabytes of tape storage, increasing each of the two libraries by 13 Petabytes.

In addition to housing in excess of 60 Petabytes of data storage resources, Pawsey's research collection service provides data sharing facilities to support scientific collaboration among researchers nationally and internationally.

During the financial year, 51 projects have access to Pawsey long-term storage. Although many projects are radio astronomy related, Pawsey also supports eight universities and CSIRO; Landgate and the Government Departments of Mines and Petroleum, Parks and Wildlife, Health, and Transport; and the Telethon Kids Institute.

VISUALISATION

Pawsey's data visualisation services help researchers gather real-time insights from their data and allow the data to be more easily interpreted by translating it into more meaningful or recognisable forms. About 200 projects benefited from Pawsey's visualisation services during the reporting period.

During this period Zeus based remote visualisation service was migrated to newly procured Topaz nodes, which provided better resources for remote visualisation. In addition to this, several visualisation tools were compiled and provided to our users. There were a total of 2925 remote visualisation sessions performed by 162 users.

A new Windows cluster was procured to replace the current high-end Windows nodes. Staff are working on providing web-based remote access to the new cluster. During the reporting period, there were 1095 sessions with 2966 hours of remote visualisations performed on high-end Windows nodes.

Leading visualisation expert Dr Jean Favre from CSCS visited Pawsey and shared his experiences in visualising large datasets. This has helped in implementing some of the new services and training on advanced visualisation topics for our users. The visualisation team also worked on designing a start-of-the-art Visualisation Lab, which will enable researchers to perform large-scale visualisations. This is expected to complete within the next financial year.

Extending the reach of Australia’s HPC expertise and its scientific endeavours

A World of Difference

Increasing Our Reach

As the Capital Refresh allows Pawsey to continue delivering the services Australian researchers need, new opportunities for researchers across Australia are being identified and new collaborations established that will also benefit from Australia's expanding High Performance Computing (HPC) capability.

REACHING FOR OUTER SPACE

Pawsey's long-term engagement with the astronomy community, the Square Kilometre Array (SKA) project and its precursors, and more recently the Australian Space Agency, continues to provide a basis to further expand activity in space research and industry.

In partnership with the WA Data Science Innovation Hub, and with support from both the Commonwealth and State governments, Pawsey is establishing a new national space data analysis facility in Perth. The Australian Space Data Analysis Facility (ASDAF) will support the growth of the Australian space industry, and unlock the value of space observational data for industries such as agriculture, mining, emergency services and maritime surveillance. By supporting researchers and small to medium enterprises with space data analytics and expertise, improving access to space data products, and lowering the cost and risk of exploratory use of space data, it aims to accelerate the commercialisation of new products and services.

THE BIGGEST SCIENCE PROJECT IN HUMAN HISTORY

While Pawsey continues to support the operation of two SKA precursors, the Murchison Widefield Array (MWA) and the Australian Square Kilometre Array Pathfinder (ASKAP), Pawsey is equally contributing to the development of the SKA. A group of Pawsey delegates this year visited the Murchison Radio Observatory (MRO) seeing firsthand the infrastructure that establishes the optical fibre connections between the Pawsey Supercomputing Centre, ASKAP, the MWA, and the SKA-low site, an uncleared area of land that will eventually host anywhere between 100,000 to a million antennae.

Mr Mark Stickells attended a SKA Governance Board Meeting, and the 2019 SKA Meeting in Shanghai, where the full designs of the telescopes were presented. He had meetings with the SKA Director and Head of Operations, and many of the lead scientists and managers from across the SKA project. Executives from the SKA Organisation subsequently visited Pawsey as part of the orientation meetings for SKA's new Head of Operations.

At OzGrav, the ARC Centre of Excellence for Gravitational Wave Discovery in Canberra, Mr Mark Stickells presented an update on Pawsey operations and capabilities to SKA stakeholders and to local Consuls-General at an event organised by the International Centre for Radio Astronomy Research (ICRAR) and the Department of Foreign Affairs and Trade's WA office.

Pawsey remains engaged in developing a regional processing model for SKA data. Pawsey's role in the planning and development for the future SKA is discussed at bi-monthly meetings with the management committee for the Australian Science Regional Centre (AusSRC). As part of the Science Data Processor (SDP) Consortium, Pawsey continues to explore SDP specifications for future installation at Pawsey.

INTERNATIONAL COLLABORATIONS

Pawsey continues to collaborate at scale across geographic boundaries to tackle global problems. Strengthening ties across the UK, Pawsey staff visited the HPC Centre at the University of Cambridge, the EPCC at the University of Edinburgh, SKA Global Headquarters at Jodrell Bank, and participated in the HPC User Forum in the UK.

Staff also participated in the HPC User forum at Argonne National Laboratory in the US, and then hosted the Chief Technical Officer of the Lawrence Livermore National Laboratory on his visit to Pawsey.

Locally, Pawsey is also working towards an Australian Research Environment collaboration together with AARNet and NCI, to provide one central research environment to access data cloud compute and HPC.

DIVERSITY AND INCLUSION

Diversity and inclusion have continued to be a key focus and vital part of Pawsey's strategy, its plan for growth and its license to operate.

More than half of Pawsey's staff have a diverse background, over 20 percent are female, and the management team and Board have over 30 percent and 25 percent female representation respectively.

Pawsey is working with the National Computational Infrastructure (NCI) and New Zealand eScience Infrastructure (NeSI) to establish an Australasian Women in HPC chapter, and hosting diversity and inclusion sessions. The Centre also supported the Women in HPC Summit in Canada in April, although this was postponed until 2021 as a result of COVID-19 restrictions. As in-person Harmony Day activities were also cancelled, a virtual space was made available to share stories and pictures of where Pawsey's community is from and the strength of its cultural diversity.

Training Staff and Researchers

The Pawsey Supercomputing Centre continues to develop a critical mass of advanced computing knowledge in the research community, making supercomputing more accessible to researchers across Australia. The expertise of Pawsey's research community is being broadened to encompass the new systems and services afforded by the Capital Refresh, and new user communities are being supported to access Pawsey facilities. Through training and development, the true impact of Pawsey supercomputing is being realised across Australia.

RESEARCHER DEVELOPMENT

TRAINING

To ensure researchers can effectively use Pawsey facilities as part of their work, Pawsey training programs support users across all levels of experience. As the Capital Refresh progresses, Pawsey is also developing and providing training specific to the new infrastructure to assist users transition to these more powerful systems.

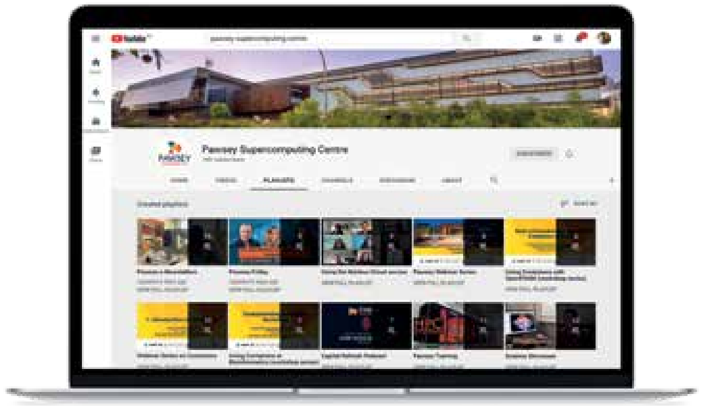
While in-person training was conducted nationally in capital cities around Australia, there has been a significant move to redevelop training material for online delivery in light of the ongoing restrictions related to COVID-19. In Q4, Pawsey created an eight-part, modular training series, providing foundational knowledge about using Pawsey's cloud, compute and visualisation resources. An online containers and webinar-and-workshop series was also launched, which trialled various 'hands-on' techniques for virtual training. A significantly large number of learners have been reached through these 'live' events and have viewed and downloaded the available resources. Pawsey is continuously developing these modes of training. For example, in the Containers series alone, 463 individuals attended 'live' virtually, and another 1,088 viewed recordings. This training series alone almost doubles the training outreach of in-person/ webinars in FY 2019/2020 (789 persons).

Foundational and core skill training provided over the year has included:

- introductory and intermediate courses on supercomputing
- advanced sessions on MPI
- OpenMP and GPU programming
- introductory courses on Unix and pshell
- Nimbus research cloud training
- remote visualisation courses.

Targeted training to meet specific user group needs included:

- NextFlow (training and hackathon)
- CUDA Programming Workshop
- Introduction to Singularity training
- Python for HPC
- iRODS Advanced (open source data management software)
- Cosmic Machines (machine learning in astronomy)
- Containers in HPC (with OpenFOAM workshops)
- Using Containers in HPC for Bioinformatics
- Scalable Bioinformatic Workflows in Snakemake and Nextflow.



To supplement these training sessions, additional technical reference documentation and information videos have been developed, which are accessible via the Pawsey Confluence page (https://support.pawsey.org.au/documentation/display/US/Training+Material) and YouTube pages (https://www.youtube.com/pawseysupercomputingcentre). Pawsey also releases a monthly technical newsletter to connect Pawsey's research community and provide useful tips and information on how to best use Pawsey resources (https://support.pawsey.org.au/documentation/display/US/Pawsey+Technical+Newsletter).

In the move online, Pawsey introduced an 'Ask. Me. Anything.' virtual hour, a weekly opportunity for researchers and users to join in a discussion with Pawsey expert staff and an online community of peers. These sessions have been useful for new researchers wanting to understand if Pawsey services are for them, as well as for current users who have specific questions about Pawsey's infrastructure and expertise or have particular challenges in using Pawsey resources for their research. These sessions also provide an opportunity for targeted, Pawsey-led discussions.

RESEARCH COMMUNITIES

Pawsey has continued to grow supercomputing expertise in the astronomy community through its involvement in the Astronomy Data and Computing Services (ADACS) project. In this collaboration Pawsey has embedded supercomputing specialists in astronomy research teams to work with researchers to better use supercomputing technologies in support of their research.

A newer user community that Pawsey has worked closely with is the emerging domain of bioinformatics. In consultation with Australian BioCommons, Pawsey has led a multi-part webinar and workshop training series, provided NextFlow training specific to bioinformatics projects, and supported uptake projects in the bioinformatics domain to help researchers better use Pawsey's resources.

Other new user communities supported to improve their supercomputing access include:

- materials sciences, with Pawsey representatives presenting a training session for computational material scientists needing HPC to support their simulations work, during the International Conference on Materials Science and Engineering 2019
- the phenomics user community, working with the Australian Plant Phenomics Society on uptake projects
- the geosciences domain, through attendance at the National Geoscience Week in conjunction with NCI and NeSI.

ENRICH CoP, an international community of trainers and educators in data and compute topics, spearheaded by Pawsey during this period, is now an active, self-led community of over 100 people. The new community creates space for national and international training collaboration opportunities, and provides an incubator platform for discussions on virtual training techniques and strategies. This is an important platform for Pawsey and others, especially as in-person industry gatherings, such as C3DIS, eResearch Australasia, ISC2020 and SC2020 are recreated as online events.

The Pawsey Summer Internship program received 250 applications from across Australia. The 12 successful Pawsey interns were this year also joined by ICRAR interns and Australian Mathematical Sciences Institute (AMSI) recipients in the Internship's first week of immersive training. Pawsey interns spent 10 weeks delving deeper into their scientific areas with support from Project Supervisors, Principal Investigators, and Pawsey HPC, data analytics and visualisation teams. Internship projects ranged across genomic mapping, sampling and genetic analyses, automatic seismic interpretation, critical infrastructure monitoring using deep learning, molecular dynamics simulations, quantum statistical algorithms, and impact crater detection.

PAWSEY TEAM DEVELOPMENT

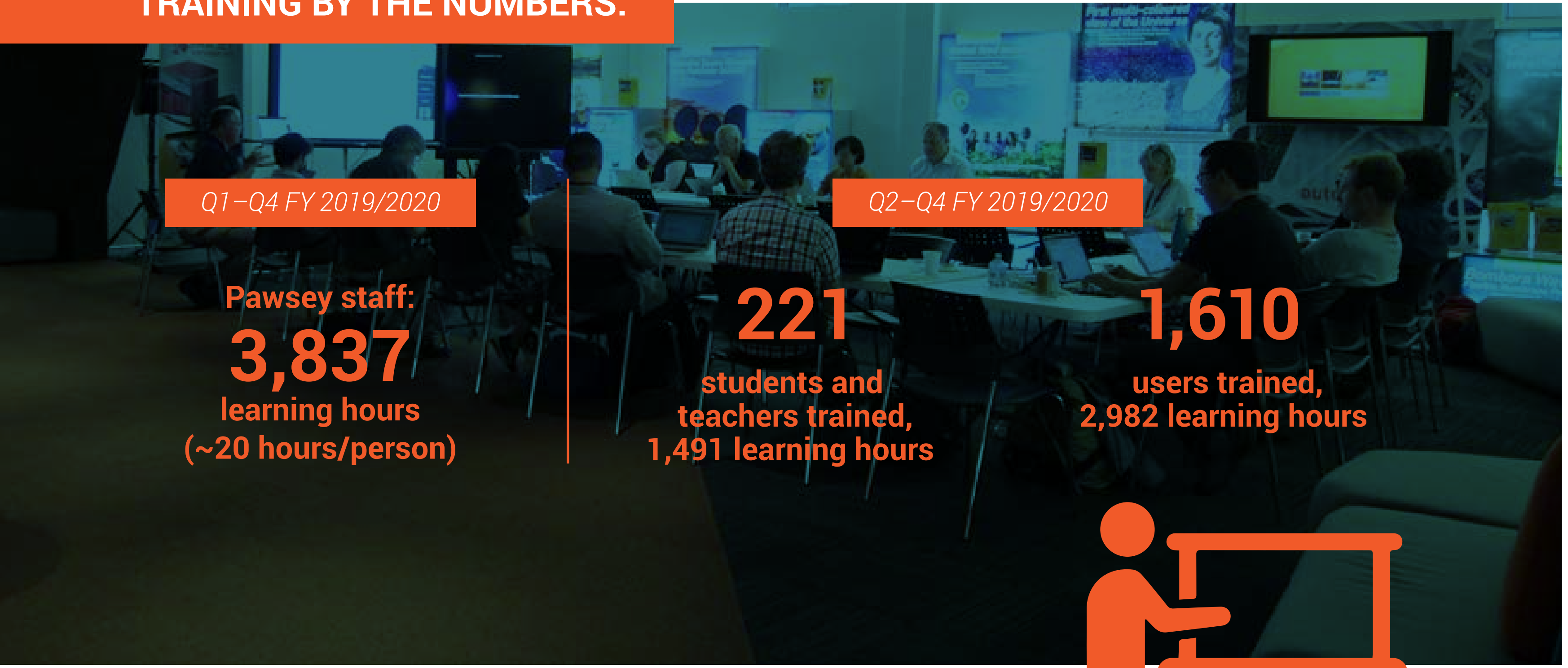
To enable Pawsey staff to provide the best supercomputing, cloud, visualisation and data management services and expertise to the research community, all staff access regular training and education opportunities. As part of the Pawsey Training and Education Strategy for 2020–2025, the internal training program ensures that existing required skills are maintained, and new skills that support and expand individual and team capabilities are developed. A significant focus for staff development has been to ensure readiness for the new systems and services coming into operation as a result of the Capital Refresh.

In preparation for the arrival of Pawsey's new supercomputing system, staff have attended emerging Graphics Processing Unit (GPU) technologies and advanced features developer presentations. Specific training to further develop skills around high performance computing and large-scale data analysis during the period included:

- Puppet
- Starfish Storage
- Snakemake
- Paraview (Visualisation)
- CUDA/GPU
- Singularity
- ISO
- Lustre
- Xilinx
- NVIDIA advanced training
- Kubernetes
- Directive-based GPU Programming (OpenACC and OpenMP).

Staff have also undertaken the CSIRO Prosci Practitioner course to assist with user-facing aspects of the Capital Refresh, and participated in courses covering essential skills, including emotional intelligence, diversity and inclusion, workplace leadership, customer service and change management.

TRAINING BY THE NUMBERS:



Q1–Q4 FY 2019/2020

Pawsey staff:
3,837
learning hours
(~20 hours/person)

Q2–Q4 FY 2019/2020

221
students and teachers trained, 1,491 learning hours

1,610
users trained, 2,982 learning hours

Engagement and Outreach

Hon Karen Andrews, Minister for Industry, Science and Technology with Pawsey CTO Ugo Varetto during her visit to the Centre

Pawsey is dedicated to generating awareness of the supercomputing, large-scale data and scientific visualisation capabilities available within Western Australia and highlighting the possibilities available for researchers and industry to harness Pawsey resources and expertise. The Centre actively engages with potential users, collaborators and partners. STEM outreach has also become an important part of Pawsey activities, with emphasis on its scalability and reproducibility.

Various events and activities were organised and attended to connect Pawsey more closely with researchers, the science community, government and industry.

- The Hon Karen Andrews, Federal Minister for Industry, Science and Technology, visited the Centre to receive an update on Pawsey's Capital Refresh Project and find out more about the impact facilities such as Pawsey have on science, industry and the nation.
- The Hon Dave Kelly MLA, Minister for Water; Forestry; Innovation and ICT; Science; Youth also visited the Centre to launch the Index of Marine Surveys for Assessments (IMSA) portal, a platform developed by the Department of Water and Environmental Regulation (DWER) and the Western Australian Marine Science Institution (WAMSI). This new online platform provides, for the first time, free access to vast amounts of environmental impact assessment data.
- Pawsey hosted more than 300 visitors touring the Centre facilities from supercomputing centres, government representatives and scientific facilities. These included an ASEAN delegation, Chinese delegates from a LNG training fund program, and visitors from the Centre for Data Health Linkage.
- As a result of COVID-19, physical visits halted in March 2020, with Pawsey expanding online virtual experiences. These visits and online experiences provided an opportunity to highlight Pawsey services to a local, national and international audience and explore how Pawsey may be able to support their work.
- Over the year Pawsey hosted four Pawsey Fridays events, two of them virtually. Topics included big data analytics, crater counting and Artificial Intelligence (AI), quantum computing, diagnostic workflows, analysis of environmental DNA (eDNA) data and more.
- Pawsey continues to lead the International HPC Best Practices newsletter released twice a year, before the two major global supercomputing conferences, SC and ISC. Collaborations with NCI and NeSI resulted in contributions on storage systems and cybersecurity.
- Regular monthly newsletters sharing scientific breakthroughs enabled by high performance computing and data science at scale were sent out and reached over 4,000 people throughout the year. A new technical newsletter was also launched in March 2020, distributed specifically to all users of Pawsey systems to inform and engage the research community through the Capital Refresh period of system change and expansion and develop and share best practices.

Supporting and promoting data intensive sciences, and STEM more generally, to the next generation of researchers and the general public remains a focus for Pawsey.

- With She Codes, Pawsey supported an all-women coding week in Geraldton in June, where 15 women learned how to script, code and program to achieve their own business and personal goals.
- Australia's future innovators were supported via the Game Changers Award in November, with Pawsey acting as both an event sponsor and a mentor for the 78 teams and 60 schools participating to showcase their innovative solutions to the world's challenges.
- Pawsey also took part in the 2019 Carnival of Computing, a multi-location, multiple day professional development event for teachers across Western Australia to attend masterclasses in digital systems, representations of data, collecting, managing and analysing data, and ICT capabilities.
- Pawsey and ICRAR partnered to host a professional development event for secondary school teachers, Making Data Accessible for Students, in February. This event built data teaching strategies for teachers and data analysis skills for year 10–12 students.
- PULSE@Parkes in February was an opportunity to introduce students to radio astronomy. Students from three high schools spent a day with CSIRO radio astronomers to operate the Parkes radio telescope in NSW and make observations in real time from the Pawsey Centre.
- Pawsey supported the 24th International Conference on Computing in High Energy and Nuclear Physics – CHEP2019, in Adelaide in November, which provided an engagement opportunity with CERN, the European Organisation for Nuclear Research.
- Pawsey supported the Human Genome Meeting 2020, which moved online in April with 505 participants from 37 countries.

Throughout the year Pawsey attended and presented at several international events to increase the exposure of the work Pawsey is involved in, as well as promoting Australia's supercomputing, large-scale data and scientific visualisation capability more generally.

Dr Natasha Hurley- Walker, ICRAR Radio Astronomer, during her presentation as part of Game Changers Award at ECU

Prof Federico Silla, Universitat Politecnica de Valencia, during his presentation at 2019 Perth HPC.AI Advisory Council conference

INTERNATIONAL CONFERENCES

- In August, Pawsey again partnered with the HPC AI Advisory Council Conference to host their third annual international conference. Over 120 delegates explored emerging domains in the industry, ranging from the impact of AI on HPC to architectural challenges for the convergence of big data analytics. Pawsey staff presented throughout the conference on a variety of topics, contributed to GPU and Singularity training and hosted the Systems Monitoring and Application Benchmarking Symposium.
- At the largest international supercomputing conference, SC19, in Denver, Pawsey's joint booth with NCI attracted at least 400 visitors in November. As part of the program, Pawsey staff delivered a Containers in HPC tutorial to more than 180 participants and contributed to two Birds of a Feather (BoFs) focusing on system administrators and outreach teams. During the conference, the 2019 HPCwire Readers' and Editors' Choice Awards were announced, and Pawsey won the HPCwire Editor's Choice Award for the Best Use of HPC in Physical Sciences for a project undertaken with Professor Brett Harris at Curtin University to map and model Perth's aquifers.
- Several other conferences where Pawsey planned to participate were cancelled, and ISC20 scheduled for June 2020 in Frankfurt took place online, with Pawsey engaging with more than 4,000 participants through BoF, presentations and a workshop on HPC education and training for emerging technologies.
- Pawsey staff also presented at or hosted many national events across a broad range of research disciplines and stakeholder groups.

NATIONAL CONFERENCES

- At eResearch Australasia in Brisbane in October, Pawsey and NCI showcased Australia's HPC capability to the world, welcoming visitors to a joint booth. Pawsey staff presented at five sessions, including perspectives on scientific software at scale, simplifying researcher experiences at HPC centres via federation and tools for reproducible and portable research workflows, and contributed to a BoF session about diversity and training.
- AstroFest '20, an astronomy festival of epic proportions celebrating Australian science, was held in February, with hundreds of visitors to the Pawsey booth participating in the 'ball challenge' to learn about supercomputing and the concept of parallelism.
- Data Science Week, a week-long series of events that celebrate everything data science organised by Pawsey, was scheduled to take place nationally in May. The event was postponed until 2021. In the leadup to the event, Pawsey had 27 organisations and events confirmed to take part in the week.
- Pawsey presented at Perth's Research Bazaar (RezBaz) festival at Curtin University in July, upskilling attendees in next generation digital research tools.
- Pawsey hosted the launch of the WA State Government hackathon #GovHack in September and submitted a challenge for participants to undertake as part of the event.
- Pawsey co-convened the Australia Leadership Computational Symposium, ALCS 2019, together with the NCI in November. The event targeted over 150 researchers benefiting from supercomputing to share their experiences.

Users

11.44%

23,366 vs 20,967

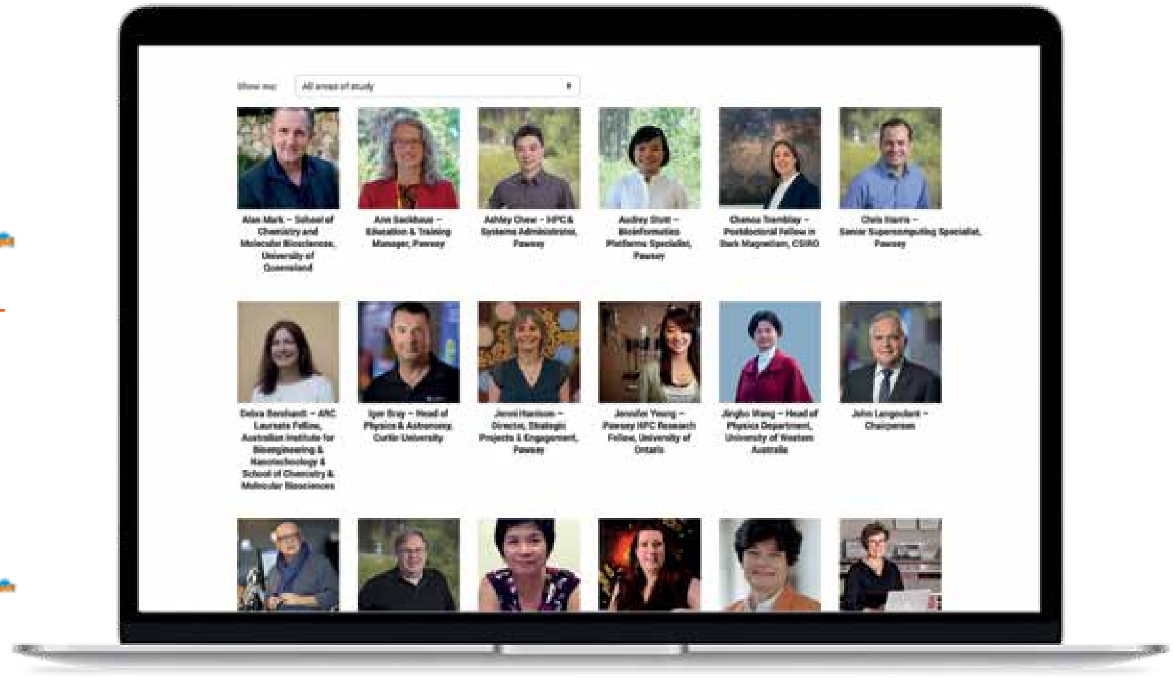
FY19 *FY18*

New users

11.71%

23,939 vs 20,535

FY19 *FY18*



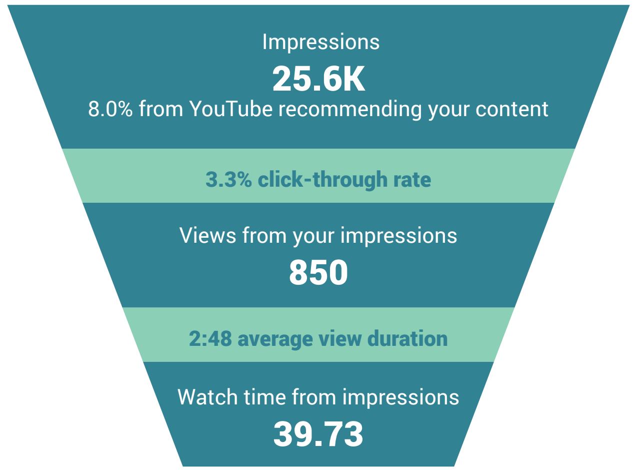
DIGITAL ENGAGEMENT

- With the drastic shift to a digital landscape, Pawsey's website has seen more than a ten per cent increase in traffic, particularly with new users. Case studies, research profiles and news articles, regularly added to the website, highlighted the work of researchers who use the Centre's infrastructure and services and benefit from Pawsey supportive staff. A Capital Refresh Update page was also created on the website explaining the main components of the Capital Refresh and its current status.
- Pawsey's YouTube channel continues to grow as the Centre focuses on its video production. Popular videos focus on Pawsey-enabled research on subjects such as climate change, genomic research, earth science and quantum computing. Due to COVID-19, training video production has also increased to ensure Pawsey's research community continues to access important resources and information to support their work. A Capital Refresh podcast series was added to the channel in November 2019 to communicate more targeted updates and progress on the Capital Refresh program. Viewership for both recently uploaded and historical videos is increasing as the channel becomes a recognised source of information on Pawsey activities.

YOUTUBE IMPRESSIONS

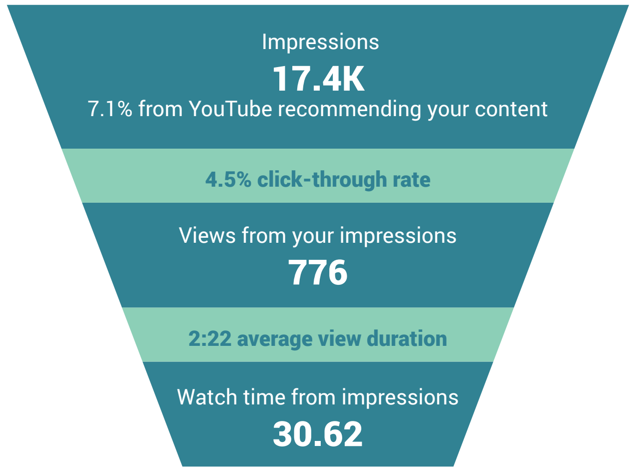
IMPRESSIONS AND HOW THEY LED TO WATCH TIME

Data available 1 July 2019 – 30 June 2020



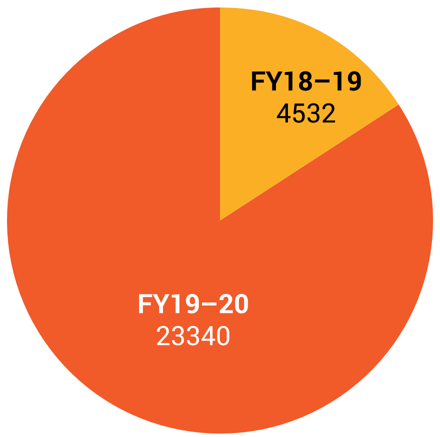
IMPRESSIONS AND HOW THEY LED TO WATCH TIME

Data available 1 July 2018 – 30 June 2019

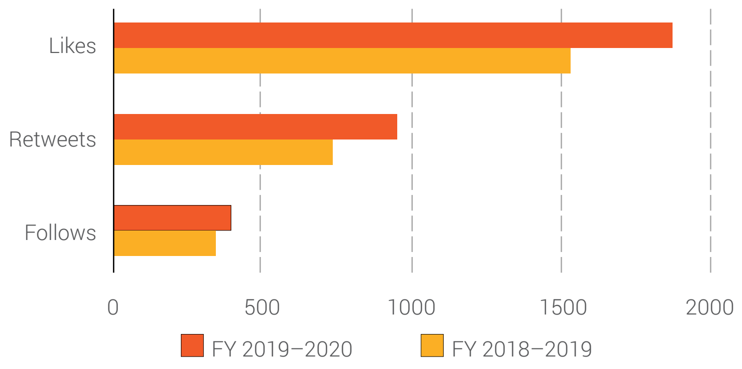


- Facebook engagement with the general public grew, with posts showcasing the impact of the services and expertise Pawsey provides, and our work towards training the next generation of researchers in Australia. A mix of science showcases, researcher profiles and event announcements has seen in a 515 per cent increase in the Centre's Facebook reach in this period.
- Twitter continues to be Pawsey's platform of choice when engaging with Australia's scientific community. Throughout COVID-19, Pawsey's community paid increasing attention to Twitter, with a steady growth of likes, retweets and follows. The COVID-19 accelerated access initiatives advertised on Twitter engaged 2,492 individuals.

FACEBOOK REACH



TWITTER GROWTH



Users and collaborations across the globe

AMERICA

EUROPE

ASIA

Financials

FINANCIAL REPORT 2019-20

The Pawsey Supercomputing Centre is an unincorporated joint venture (UJV) between five institutions: The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Curtin University of Technology (CUT), Murdoch University (MU), Edith Cowan University (ECU) and The University of Western Australia (UWA). CSIRO, the appointed Centre Agent of Pawsey, holds and manages its assets and finances. With this arrangement, Pawsey Centre is required to adhere to CSIRO's governing legislation, constitution or adopted policies including its reporting, budgeting and auditing framework.

The Commonwealth Government provided the initial capital investment of $80 million in 2009 for the construction of the Pawsey Centre's building and High-Performance Computing infrastructure. In April 2018, the Pawsey Supercomputing Centre received a $70 million Capital Refresh Grant from the Australian Federal Government to secure the next generation of supercomputers, data infrastructure and service upgrades. This is financially managed and reported by Curtin University, one of the UJV members of Pawsey Supercomputing Centre. A strong governance process over the procurement has been implemented to involve all members of the UJV. The progress during this period included the delivery of the Cloud/High Throughput Computing and Astronomy high speed storage procurements with other elements of the refresh including the MWA Cluster, Remote Visualisation, and the Pawsey Supercomputer System. Approximately a total $4.8 million of completed asset procurements through this Capital Refresh program have been transferred to Pawley's Fixed Assets Register as at the end of June 2020.

The Commonwealth Government continues to support the infrastructure maintenance and operation of the Pawsey Centre through the National Collaborative Research Infrastructure Strategy (NCRIS) program run by the Department of Education, Skills and Employment (formerly known as the Department of Education and Training).

The Western Australian Government along with the UJV partners, provided the other portion of the operational funding for the running and maintenance of the Pawsey Centre. Recently, Pawsey was able to attract additional funding from the Federal Government for the Australian Space Data Analysis Facility in Western Australia, co-funded by the Western Australian Government.

Pawsey, through CSIRO, employs professionals and experts on High Performance Computing to run operations of the facility and provide exceptional services on supercomputing access, cloud, data storage and visualisation, training and consulting. As a Tier-1 High Performance Computing facility in Australia, Pawsey Centre's function is to accelerate scientific research for the benefit of the nation enabling research in areas of astronomy, bioinformatics, space and other national and state science priorities. The continuance of this funding support is essential for its existence.

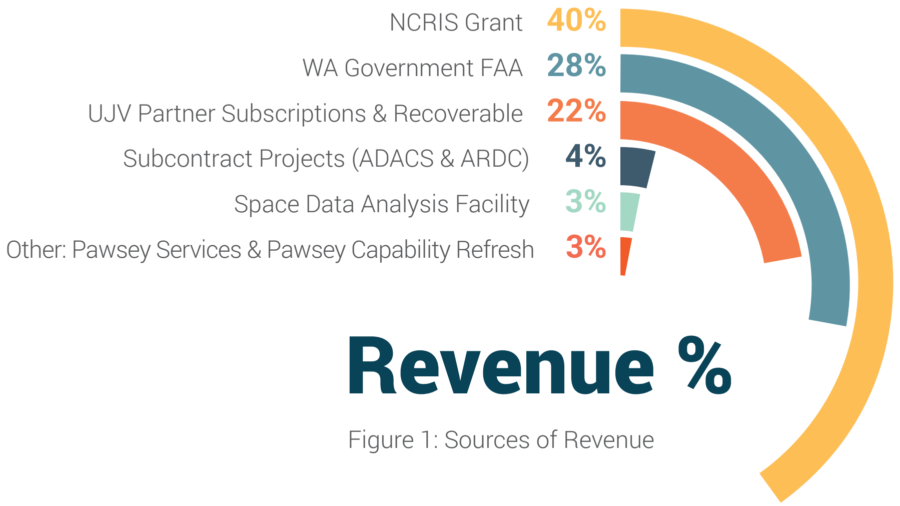


Figure 1: Sources of Revenue

Pawsey Supercomputing Centre
Statement of Income and Expenditure
01 July 2019 to 30 June 2020

	$'000
Revenue	
External Project Funding	12,282
Internal Joint Venture Partner Subscriptions and Recoverable	3,522
Total Revenue	**15,804**
Expenditure	
Labour staffing	7,704
Asset Procurement (Capital Expenditure)	3,013
Machine Maintenance Contracts	1,398
Operating costs	
Utility charges	1,709
Other Operating expenses	2,167
Total Expenditure	**15,991**
Surplus / (Deficit)	**(187)**

*Note: Deficit refers to carried forward activities funded in 18/19 and completed this period.

The funding model for Pawsey aims to reflect the proportionate usage of the machine on research projects. The Commonwealth Government through the NCRIS program provided around 40% of the total funding source. The Financial Assistance Agreement from the Western Australian State Government contributed around 28%, whilst 22% of the total source of funds came from the UJV partners membership subscription and share of site operations costs. These are the three main pillars of income streams for Pawsey Supercomputing Centre. Minor subcontract projects both from ADACS and ARDC also contributed in total around 4% of income and subsequently 3% equally from Other income sources: Pawsey Services/Pawsey Capital Refresh and the newly established Space Data Analysis Facility in Western Australia.

CSIRO, as the Centre agent also contributed to the running of the facility by providing administrative support services and facility maintenance. This contribution is considered in-kind as it is not paid through Pawsey's external funding sources and is part of CSIRO's appropriation budget. For the year, CSIRO's Enterprise Support Distribution $5.16 million to the Pawsey Supercomputing Centre. This is not reflected in the above Statement of Income and Expenditure.

For this reporting period, the major cost driver for the Centre is labour which is 48% of the total expenditure. This represents the staffing costs for technical experts and allied support services which are essential services to run the facility. Asset procurement for minor hardware and network upgrades comes next at 19% which is mainly funded from the NCRIS grant. These are essential minor upgrades to the existing infrastructure to continue usage while the Capital Refresh is still ongoing. Other Operating expenses covers 14% of the total costs whilst utility cost is around 11% and includes electricity, gas and water usage. The Centre receives electricity credits from the solar panels installed at the building.

48% Labour staffing
19% Asset Procurement (Capital Expenditure)
14% Other Operating expenses
11% Utility charges
9% Machine Maintenance Contracts

Expenditure %

Figure 2: Expenditure Breakdown

Cover image: Jesse Pickett | Yamaji Art | *Rainbow Serpent and the Moon*

The Pawsey Supercomputing Centre is supported by the Australian Government through a $70 million grant made under the Industry Research and Development Act and administered by the Department of Industry, Innovation and Science. Pawsey is also supported by the Australian Government under the National Collaborative Research Infrastructure Strategy (NCRIS) and related programs through the Department of Education. The Centre would also like to acknowledge the support provided by the Western Australian Government and its Partner organisations.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

GOVERNMENT OF WESTERN AUSTRALIA

CSIRO Curtin University ECU Edith Cowan University Murdoch University The University of Western Australia

pawsey.org.au