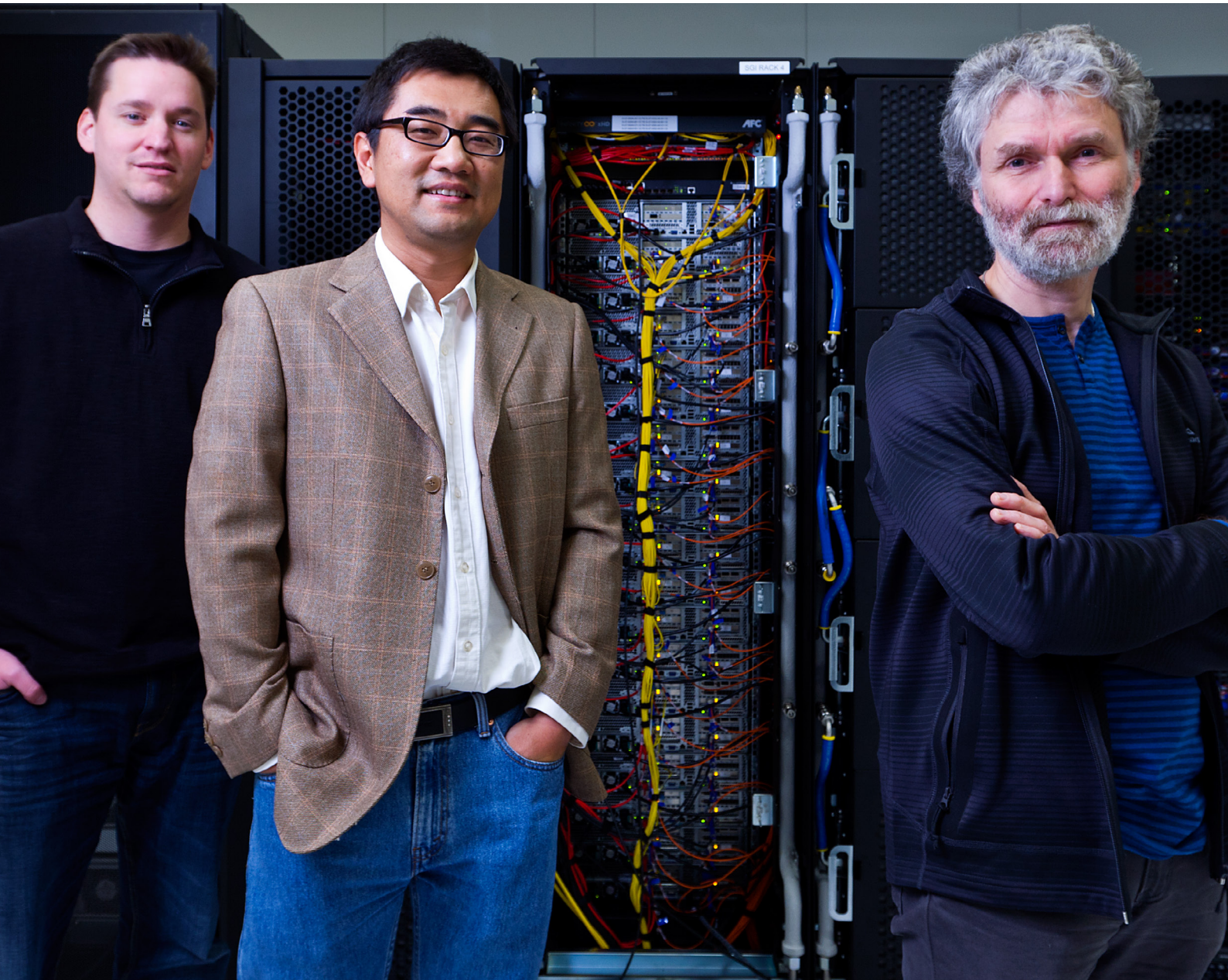
ICRAR

International Centre for Radio Astronomy Research

DATA INTENSIVE ASTRONOMY

Imagining leading edge solutions for unimaginable volumes of data.



Data Intensive Astronomy –a new industry partner

In this brochure we introduce the capabilities, partners and current projects of our Data Intensive Astronomy team.

Exploring the entire Universe through space and time, from now to the very first stars and galaxies that existed more than 10 billion years ago, is an unparalleled feat of human scientific endeavour.

The volume of data generated by new and planned observatories is currently doubling every six to 12 months—faster than the rate of increase in performance of computer chips (Moore's Law).

Such a challenge is costly and creating the biggest astronomical research facilities in the world is beyond the funding capabilities of individual universities, research organisations and even nations.

For this reason, collaborative alliances of organisations and nations are being formed to fund, build and manage the data of the next generation of telescopes.

This expansion and globalisation of research raises a number of major technical and organisational challenges that need to be tackled and solved by researchers, funding bodies, industries and governments.

In all cases, the challenges of managing, exploring and sharing the huge volumes of digital information flowing from these new global facilities is focusing and leading the international discussion.

We are seeking opportunities to work collaboratively with industry partners who are facing similar challenges as they explore the natural resources of our planet.

Cover L→R Dave Pallot, Dr Chen Wu and Professor Andreas Wicenec from ICRAR's Data Intensive Astronomy Team

Introducing our Data Intensive Astronomy & Science teams

DATA TEAM
PROFESSOR ANDREAS WICENEC

Professor Andreas Wicenec completed a PhD in Physics and Astronomy at the University of Tübingen in Germany.

Having worked for some of the world's premier astronomical research facilities, Andreas has a wealth of experience in the design, implementation and operation of large volume, global data flow and management systems.

For the European Space Agency's Hipparcos satellite Andreas developed data management systems as well as photometric and astrometric data reduction techniques. He then went on to implement the back-end archive system for the European Southern Observatory's Very Large Telescope.

In Chile, for the Atacama Large Millimeter Array, Andreas designed the radio telescope's archive sub system, and for the Murchison Widefield Array located in outback Western Australia, he designed and implemented the archive and data distribution system.

Now, Andreas is leading ICRAR's Data Intensive Astronomy program to research, design and implement data flows and high performance computing systems for current and future observatories and large-scale astronomical surveys.

DATA TEAM
MARKUS DOLENSKY

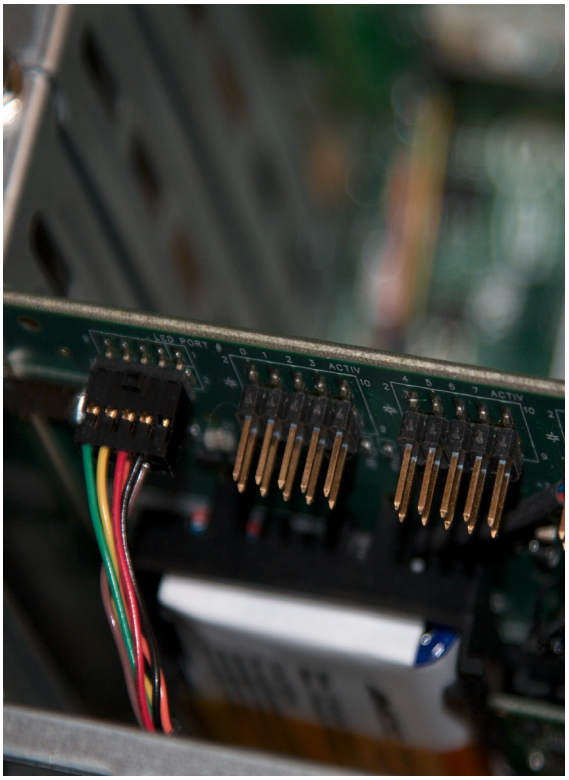
Markus Dolensky is a computer scientist and the Technical Leader for ICRAR's Data Intensive Astronomy Group.

Originally from Austria, Markus has held positions in industry and in several of the world's premier astronomy research institutions, working on projects involving ground and space based missions.

Top Prof. Andreas Wicenec
Below Markus Dolensky

The SKA & ICRAR

The Square Kilometre Array (SKA) will be the biggest and most capable radio telescope ever built. It will expand our understanding of the Universe and drive technological development worldwide.



The International Centre for Radio Astronomy Research (ICRAR) is a joint venture between Curtin University and The University of Western Australia with additional funding from the State Government of WA.

ICRAR was designed to be a multi-skilled institute of astronomers, engineers and data specialists that could support the build up of the SKA in Australia through design, construction and ultimately operations.

In 2014, Deloitte Access Economics identified ICRAR as being one of the top five centres of its kind in the world.

ICRAR's Data Intensive Astronomy (DIA) team, located at the Centre's UWA node, is comprised of researchers from astronomy and industry who have led the development of data and operations systems for billion-Euro astronomical infrastructures in Europe and South America.

In Australia, our ability to lead science projects on pathfinder facilities like the Australian SKA Pathfinder (ASKAP) and the Murchison Widefield Array (MWA), and to design and deploy end-to-end prototype SKA systems into the field, has placed the Centre at the forefront of the international SKA project on several key developments.

The SKA will come online in the next decade and when it does, it will produce a relentless stream of science data products that are exascale in terms of their storage and processing requirements.

ICRAR's DIA team are now leading the international effort to address the challenges surrounding the flow of data within the SKA observatory.

After graduating from the Vienna University of Technology, Markus took a position with GePARD as a software engineer, playing an important role in the development and integration of the Meteosat-6 Anomaly Correction Software deployed at ESOC and the EUMETSAT ground segment that saved the mission.

Then, as a scientific systems analyst for the European Space Agency, Markus went on to design and develop a novel on-demand data processing framework for the Hubble Space Telescope, before pioneering the Virtual Observatory as technical lead and project manager at the ESO.

Markus recently moved to Australia to take on one of the biggest challenges in data intensive computing - designing the data layer for the exascale Science Data Processor of the Square Kilometre Array.

DATA TEAM
MARK BOULTON

Over the past 25 years Mark has held positions in several large multinational companies, working on systems for national and international defence and other large government projects.

Prior to joining ICRAR in 2011, Mark helped design and integrate the front-end systems for Auckland Transport's smart card public transport ticketing system, as well as the ground support systems for the Australian Multirole Helicopter, and was the lead designer for an Intelligent Transportation System network for Queensland Motorways.

Mark is now the Senior Systems Engineer for ICRAR's Data Intensive Astronomy team. He is responsible for the management and support of ICRAR's Petabyte storage servers and is helping to design the data layer for the Square Kilometre Array.

Right Mark Boulton

The Data Deluge

The widening gap between the end user's capabilities to download and process data and the size of the source data sets is driving a major paradigm shift in the way large research data should be processed and accessed.

EACH YEAR, THE SKA WILL PRODUCE 100 TIMES MORE DATA THAN THE WORLD'S INTERNET TRAFFIC IN 2010.

Large data sets and computational resources of the future are likely to be concentrated at data centres located around the world with end users transparently interacting with them.

Data will no longer be migrated from one place to another but rather accessed, processed and explored remotely across a distributed network.

To manage this environment and coordinate these interactions, new software tools and techniques are needed and the Next Generation Archive System, or NGAS as it's commonly called, is one such solution being developed, evolved and used for this purpose.

Originally created under the leadership of Professor Andreas Wicenec, while he was working with the European Southern Observatory, NGAS is being used by many of the world's major astronomical observatories.

Today, ICRAR's version of NGAS has been optimised and finetuned to deal with far higher data rates and volumes than initially possible. It has also been adapted to run in both supercomputing and cloud based environments, and to support the integrated usage of cloud resources for scientific analysis.

The NGAS infrastructure is currently used to control many Petabytes of data stored in hundreds of millions of individual files, distributed and mirrored across four continents, while providing transparent access for end users.

These innovations are not limited to astronomy and have the potential to overcome similar challenges faced by industry now, and in the future.

DATA TEAM
IAN COOPER

Before joining ICRAR, Ian Cooper worked in the satellite communications industry for more than 20 years.

As a Product Manager he led a team of engineers responsible for the certification of Land, Maritime & Aeronautical mobile satellite communications products, which provided both commercial communications and distress and safety systems.

In 2010 Ian started the SwiftBroadBand Satellite (SB-SAT) program, jointly funded by DARPA and the European Space Agency. This system used geostationary satellites to provide an 'always-on' data relay capability for Low Earth Orbit satellite payloads, with applications including weather, imaging and high speed connectivity.

Ian is now ICRAR's Deputy Project Manager for the SKA Science Data Processor.

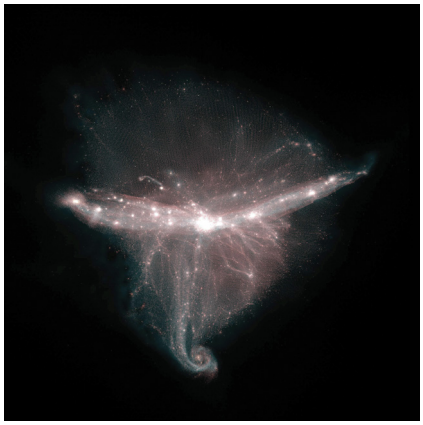
DATA TEAM
DR RICHARD DODSON

Dr Richard Dodson has held research positions at several Australian universities and worked for a number of overseas institutes including the Japan Aerospace Exploration Agency, the National Observatory in Spain and the Korea Astronomy and Space Science Institute.

He is currently investigating ways to export the computing effort required to complete the science analysis of SKA-scale datasets. This is important, as it is probably the only effective way to handle the computing requirements.

Richard is the Project Scientist for ICRAR's Data Intensive Astronomy team and with expertise encompassing both astrophysics and computer science, he is well suited to acting as an interface between big science and big data.

Top Ian Cooper
Below Dr Richard Dodson



Current Projects

SKA SCIENCE DATA PROCESSOR

The Square Kilometre Array Science Data Processor (SDP) work package is responsible for the data reduction, long term archiving and dissemination of the vast data streams delivered by the SKA front-end signal processing systems.

Essentially we are connecting an enormous sensor network directly to a configurable and flexible HPC system and streaming data through a Wide Area Network (WAN) from hundreds of kilometres away into memory and high performance non-volatile storage.

Researchers and industry partners from 14 countries are involved in this significant project for the SKA and our team at ICRAR is responsible for the 'Data Layer'. This is the part of the SDP system which will handle all aspects of data management, from receiving the data streams through to distributing them to thousands of individual compute nodes, triggering the processing steps, collecting the intermediate and final results and then providing access to those results for the global astronomical community.

At its core, the architecture for the Data Layer is a fairly generic design and is applicable to a whole range of data intensive and data driven applications.

SURVEY SCIENCE SUPPORT

ICRAR'S DIA team directly supports the various survey science activities being conducted by our scientists. This research uses data gathered from telescopes imaging the distant Universe across a broad spectrum of wavelengths. This in turn implies a large variety of different algorithms and data formats and results in challenges for the optimal deployment and usage of the available IT infrastructure. The team provides expertise to solve and optimise algorithms and the tasking of available computing resources.

SYSTEM SUPPORT

To support the work of our science teams we operate a data and computing lab. The equipment in this lab includes standard compute servers, a small GPU cluster with Infiniband interconnect and an ample amount of storage.

Through our collaborations with high performance computing providers and vendors we often have access to new and exciting technologies that we trial and evaluate.

DATA TEAM
DAVE PALLOT

Over the past 16 years Dave Pallot has held various engineering positions in the telecommunications and resource sectors with experience in IT administration, software and systems engineering.

Prior to joining ICRAR in 2010, Dave designed process control solutions for BHP Billiton, as well as designing and commissioning protocol stacks for voice and data solutions, and was a telecommunications officer for the Department of Defence.

Dave is now an ICRAR Engineer working on various projects including the data archive for the Murchison Widefield Array and the Square Kilometre Array's Science Data Processor.

DATA TEAM
KEVIN VINSEN

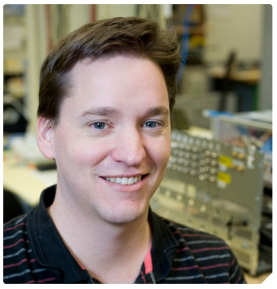
Before coming to ICRAR, Kevin Vinsen was a chief software engineer for Thales, one of Australia's leading defence contractors.

Kevin's research interests now relate to the big data issues facing the SKA. These include high speed data ingest, software engineering for large distributed software teams, data life cycle and optimising high performance computing code.

Currently, he is developing new techniques for the automated classification of galaxies using multi-wavelength data and advancing Spectral Energy Distribution calculations.

Kevin is also providing data support for large galaxy surveys such as GAMA and CHILES and is the project scientist for theSkyNet, a citizen science initiative that links thousands of computers around the world to simulate a powerful supercomputer.

Top Dave Pallot
Below Kevin Vinsen



Top An artists impression of what the SKA's low frequency antennas will look like.

Middle ICRAR headquarters at The University of Western Australia.

Bottom A simulated image of a dwarf galaxy forming several billion years ago. **Credit** Dr Alan Duffy (ICRAR), Paul Bourke (iVEC@UWA), Dr Robert Crain (Leiden Observatory).

DataDirect™ NETWORKS

"Our collaborative work with ICRAR has included the development of a software platform in virtual space within a storage controller to enable a reduction of latency for filtering and processing large data sets.

We feel strongly that the development of any product must directly address specific needs and cannot be done strictly in a laboratory environment. Our collaboration with ICRAR ensures that our product will address the needs of an organisation that is tasked with efficiently reducing, managing and distributing large data sets to a worldwide user community."

Dave Fellinger, Chief Scientist - DDN

ThoughtWorks®

"ThoughtWorks has been proud to work with ICRAR bringing agile software development methods to their data intensive applications. The breadth of work undertaken by ICRAR has given us the ability to adapt our tools and processes to address a wide range of data intensive applications.

We are using these adaptations already, as data intensive applications are becoming more mainstream in the business world, and not just in the scientific computing realm."

Rebecca Parsons, Chief Technology Officer - ThoughtWorks

SYSTEMIC

"Our work with ICRAR has involved many aspects of Big Data and Data Sciences systems engineering, especially in the area of managing the preparation and distribution of the massive quantities of data generated by the SKA.

The enormous data sets involved have given our organisation insights and experience beyond what is typical in our industry. This gives us a head start and a significant competitive advantage as data volumes increase across other industry sectors and Big Data and Data Science in general become more pervasive."

Greg Salotti, Chief Executive Officer - Systemic

DATA TEAM
DR CHEN WU

Dr Chen Wu grew up in China and worked as a software engineer and a researcher at the Chinese Academy of Sciences before moving to Perth to undertake a PhD in data retrieval and storage and then postdoctoral research in pattern recognition.

Currently, Chen is developing a coherent approach towards cost/performance optimal data management in accordance with the data life cycle requirements of various radio astronomy science cases.

Chen is leading the Data Lifecycle Management and Persistent Storage work package for the SKA's Science Data Processor international consortium.

SCIENCE TEAM
PROFESSOR PETER QUINN

Professor Quinn received his BSc (Hons) in Mathematics and Physics from the University of Wollongong in 1978 where he received the University Medal in Physics. He conducted graduate studies in astronomy and astrophysics at the Australian National University and received his PhD in 1982.

During postdoctoral appointments in Theoretical Astrophysics at the California Institute of Technology (1982-1985) and the NASA Space Telescope Science Institute (1985-1989) Peter pursued his research interests in galaxy formation and dynamics, cosmology and dark matter. His work on computational astrophysics was awarded a NASA High Performance Computing and Communications Grand Challenge Award.

In 1995, Peter accepted a position as Division Head of the newly formed Data Management and Operations Division at the

Top Dr Chen Wu
Below Professor Peter Quinn

CASE STUDY: THE NATIONAL RADIO ASTRONOMY ORGANISATION

The Next Generation Archive System (NGAS) was adopted by the National Radio Astronomy Observatory (NRAO) as a strategic partnership with the international Atacama Large Millimeter/submillimeter Array radio telescope while Professor Andreas Wicenec was the lead architect for the telescope's data archive sub-system at the European Southern Observatory (ESO).

NRAO also deployed NGAS with great success to support the observational data archive for the Karl G. Jansky Very Large Array and the Very Long Baseline Array, and is now in the process of implementing at the Green Bank Telescope.

NGAS has provided over three petabytes of aggregate archive storage at NRAO, implemented with diverse site storage for increased availability, enabling NRAO to fulfil its ongoing commitment to preserve all standard observations in perpetuity — even as the data rates from its instruments have increased exponentially.

This unique architecture has the advantage of embedding the processing for archive data management within the storage subsystems, allowing for near-infinite scalability with commodity component storage hardware. This in turn has allowed for an exceptional price/performance value proposition to be realised at a time of ever increasing budget pressure. The insight and support of Professor Wicenec and his team at ICRAR continue to be of substantial benefit to the NRAO and the broader astronomical community.

Left Top The Atacama Large Millimeter/submillimeter Array (ALMA) in northern Chile. **Credit** ESO/NAOJ/NRAO.

Right Bottom The ALMA telescope. **Credit** P. Carillo - ALMA (ESO/NAOJ/NRAO).

The Green Bank Telescope operated by the NRAO in West Virginia, USA. **Credit** ESO/NAOJ/NRAO.

European Southern Observatory headquarters in Munich. While at ESO, Peter led the efforts to design, implement and operate the science data flow system for the 1 billion Euro Very Large Telescope–the world's largest optical and infrared observatory. This work was awarded a Computerworld 21st Century Achievement Award for Science.

During his time at ESO, Peter co-founded the International Virtual Observatory Alliance, directed the FP-5 Astrophysical Virtual Observatory project and coordinated the formation of the EURO-VO as a program to realise VO-enabled science for Europe.

In 2005, Peter was awarded a Western Australian Premier's Fellowship and took up the position of Professor of Astronomy and Astrophysics at The University of Western Australia. In 2008, Peter was appointed inaugural director of the new International Centre for Radio Astronomy Research.

Since then Peter has been named the WA Scientist of the Year and made a Fellow of the Australian Academy of Technological Sciences and Engineering.

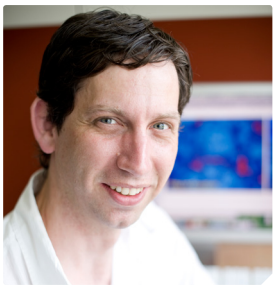
SCIENCE TEAM

DR MARTIN MEYER

Dr Martin Meyer is Senior Research Fellow at The University of Western Australia node of ICRAR. He completed his PhD thesis "Neutral Hydrogen in the Local Universe" at the University of Melbourne using data from the Parkes radio telescope, before moving to Space Telescope Science Institute in the United States, where he worked on data from the Spitzer and Hubble space observatories.

Martin is an expert in large surveys of neutral hydrogen and since returning to Australia has focused his research efforts on the development of new deep surveys to understand the evolution of neutral gas across cosmic time. He is the Principal

Right Dr Martin Meyer



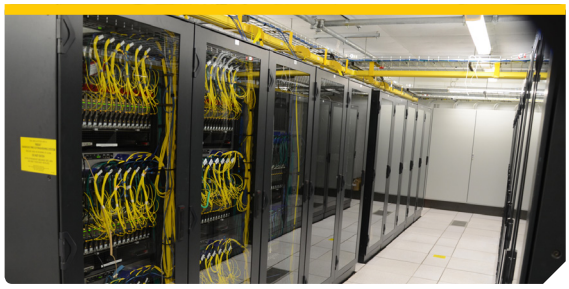
CASE STUDY: CSIRO

The CSIRO ASKAP Science Data Archive (CASDA) is one of three systems that make up the core computing component for CSIRO's newest radio telescope. This telescope, the Australian Square Kilometre Array Pathfinder (ASKAP), is located at the Murchison Radio-astronomy Observatory (MRO) in Western Australia.

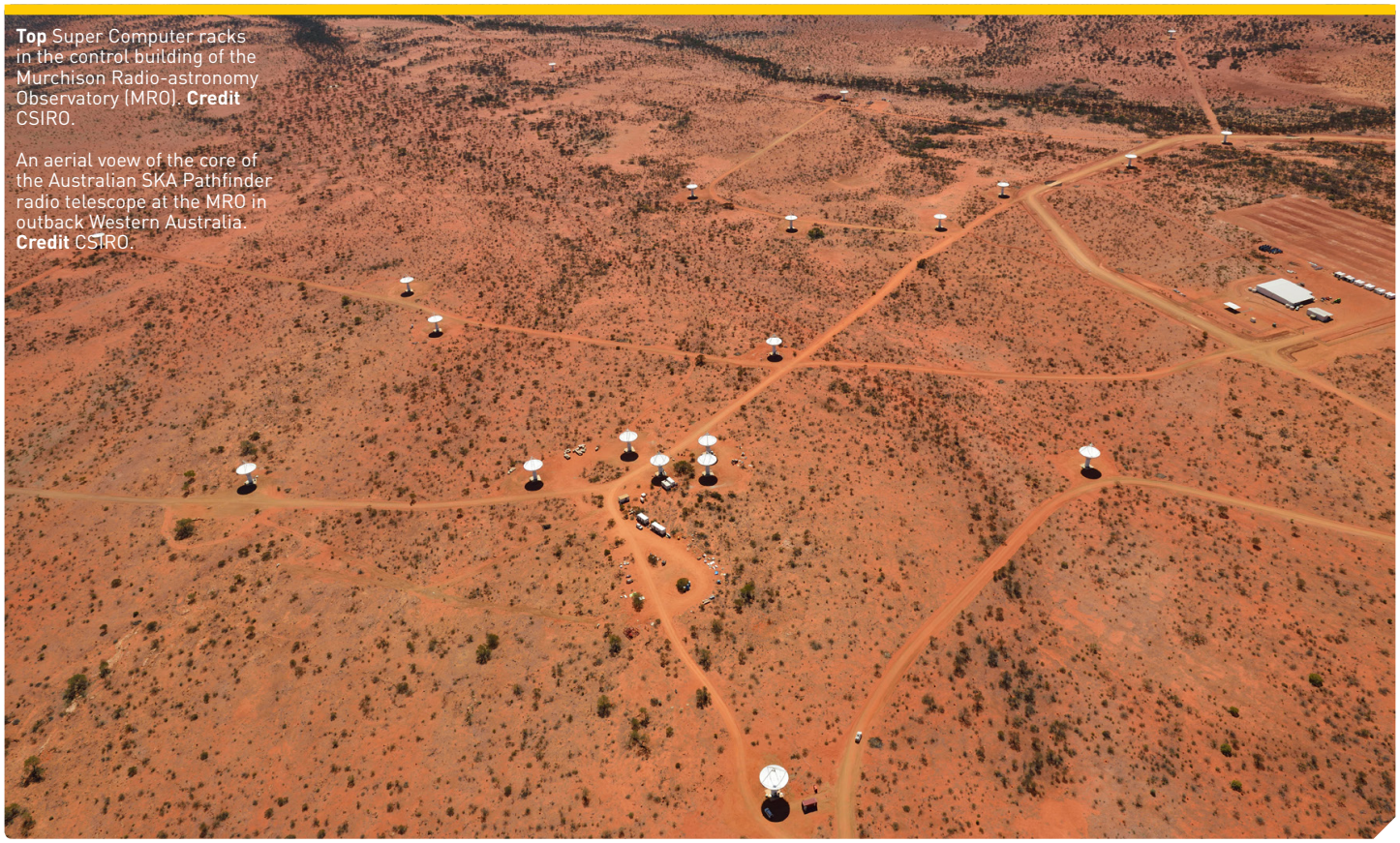
CASDA is the primary point for storing, managing and sharing fully calibrated and science-ready data products. It will also provide the ASKAP Survey Science teams with access to processed data products for analysis.

ASKAP data will arrive at the Pawsey Supercomputing Centre in Perth at a rate of approximately 2.5 GB/s, equivalent to 75 Petabytes per year. ASKAP data processing must therefore be carried out in quasi-real time using automated pipelines to produce data products and associated metadata that are stored and made available through the science archive. The total volume of archive data is expected to reach 5 PB per year.

As part of its multi-tiered system, CASDA makes use of ICRAR's Next Generation Archive System (NGAS) to meet data storage and retrieval needs. The CASDA team has been working closely with Professor Andreas Wicenec and his team to extend NGAS's integration with the hierarchical storage layer and to support the large files produced by the ASKAP pipelines.



Top Super Computer racks in the control building of the Murchison Radio-astronomy Observatory (MRO). **Credit** CSIRO.



An aerial voew of the core of the Australian SKA Pathfinder radio telescope at the MRO in outback Western Australia. **Credit** CSIRO.

Investigator of the DINGO HI survey on the ASKAP radio telescope, a project that surveys the gas content of galaxies over the past 4 billion years, and is a team member of corresponding deep HI surveys on MeerKAT and the Very Large Array.

SCIENCE TEAM
DR ATTILA POPPING

Dr Attila Popping has been with ICRAR at The University of Western Australia since the end of 2010. He completed his PhD thesis in Astronomy at the University of Groningen in the Netherlands, spending a large fraction of his time at CSIRO Astronomy and Space Sciences (CASS) in Sydney.

His research is focused on the distribution of neutral hydrogen (HI) emission in the Universe and he has experience with many major radio observing facilities.

Currently Attila is actively involved in several ongoing and future HI surveys that will all process large data volumes. The most significant is the CHILES survey on the Jansky Very Large Array (JVLA), for which he has developed the imaging pipeline.

Attila is a member of the ASKAP early science and commissioning team and a member of the SKA HI Science Working Group.

SCIENCE TEAM
DR NATASHA HURLEY-WALKER

Dr Natasha Hurley-Walker joined the Curtin University node of ICRAR in 2011, having completed a PhD in Radio Astronomy at the Cavendish Laboratory at the University of Cambridge.

During her PhD studies, Natasha helped to commission the Arcminute Microkelvin Imager, a 15-GHZ radio interferometer, and to carry out surveys for extragalactic radio sources, including clusters of galaxies.

Natasha's expertise in commissioning led to

Top Dr Attila Popping
Below Dr Natasha Hurley-Walker

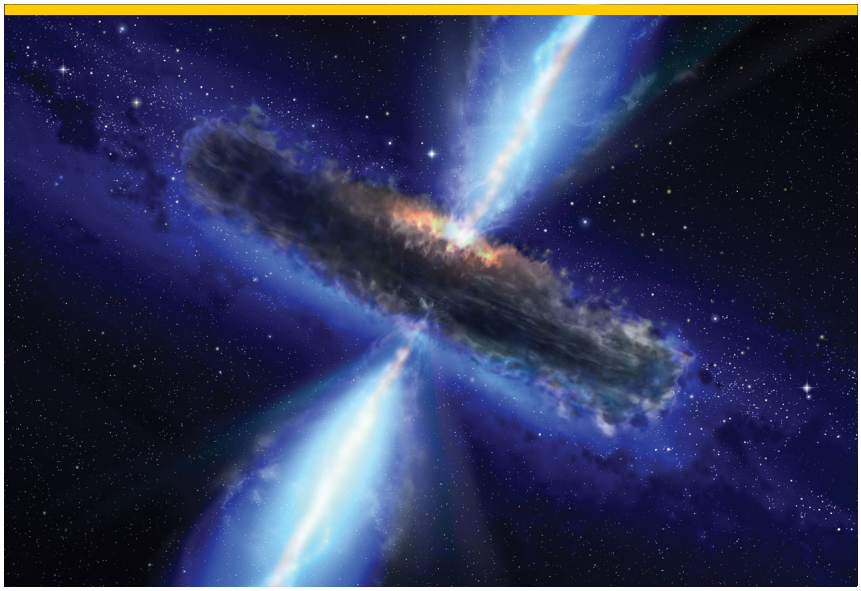
CASE STUDY:
THE INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

The Virtual Observatory framework developed by the International Virtual Observatory Alliance (IVOA) is providing efficient access to astronomical data and services for the world's astronomers. After more than a decade of development this international community-based initiative is transforming the way modern astronomical research is being done.

The Director of ICRAR, Professor Peter Quinn, played a major role in the inception of the Virtual Observatory by leading the first European based project, one of the three inaugural initiatives for the alliance. Peter has been instrumental in defining the IVOA, its role, membership and organisation.

The head of ICRAR's Data Intensive Astronomy team, Professor Andreas Wicenec, was a member of the early Virtual Observatory team at the European Southern Observatory and later re-joined the alliance as the Australian representative having moved to Western Australia and ICRAR.

Collaboration with large astronomical projects is essential for the Virtual Observatory – new projects bring new kinds of data that in turn require new tools and techniques, and ensure the Virtual Observatory initiative remains relevant and sustainable. SKA is of course a prominent element in the landscape and ICRAR obviously has a major role to play in the VO-SKA collaboration.



her performing a critical role in getting the Murchison Widefield Array (MWA) on-line, carrying out the first wide-area radio survey with the telescope and establishing the supercomputing pipelines to process its large volumes of data.

Natasha is the MWA Galactic and Extragalactic Group project scientist and a member of the SKA continuum science working group.

SCIENCE TEAM
DR TOBIAS WESTMEIER

Dr Tobias Westmeier is a Research Fellow at The University of Western Australia node of ICRAR.

Tobias studied physics and astronomy at the University of Bonn in Germany where he completed his doctoral thesis on a study of the high-velocity clouds around the Andromeda Galaxy in 2007. He then moved to Sydney to join CSIRO Astronomy and Space Science as a Bolton Fellow before assuming his current position at ICRAR.

Tobias is the initiator and leader of the SOFIA project, a new source finding pipeline designed to automatically detect and parameterise galaxies in spectroscopic radio surveys. SOFIA is a collaborative effort by radio astronomers from around the world using several novel algorithms to reliably detect signals from very faint galaxies hidden in large data volumes.

Tobias' research interests include the study of neutral hydrogen gas in the Milky Way and beyond, with the aim of understanding its role in the structure and evolution of galaxies. He is involved in several large survey science projects to be carried out with the Australian SKA Pathfinder (ASKAP) and has worked as the ASKAP project scientist in the past.

Right Dr Tobias Westmeier

Far Left An artists impression of a dust torus around a super-massive black hole. **Credit** ESA/NASA, the AVO project and Paolo Padovani.

Right Page The Pawsey Supercomputing Centre in Perth, Western Australia.

CASE STUDY: CHILES

The COSMOS HI Large Extragalactic Survey, or CHILES, is an international collaboration involving astronomers in six continents and ICRAR's Data Intensive Astronomy team.

Using the Very Large Array radio astronomy observatory in New Mexico, this survey will study the same region of sky for 1000 hours spread over several years. By imaging galaxies up to a third the age of the Universe and 4.6 billion light years from Earth, CHILES hopes to provide the most sensitive observations ever made of neutral hydrogen and help researchers better understand how galaxies evolve.

Radio astronomy is becoming extremely data intensive and CHILES is one of the first examples of this new kind of project. The data rate, a Terabyte for every 10 hours of observing, is larger than anything before it.

Reducing and processing such data volumes requires a new approach with large and dedicated computing facilities. This makes the CHILES project not only exciting for astronomers but also an interesting challenge for computer scientists.

Although the project is in its early stages with 200 hours of data in hand so far, ICRAR'S DIA team are playing a crucial role in developing and executing an imaging 'pipeline' for this important work.

CASE STUDY: VICTORIA UNIVERSITY

Victoria University of Wellington (VUW) in New Zealand is one of the partner institutions in the Murchison Widefield Array (MWA) radio telescope and is home to one of the three international MWA NGAS nodes – end points to the data chain that see images and raw data collected from the telescope pushed seamlessly to remote locations from the Pawsey Supercomputer Centre in Perth.

The NGAS node was established in 2012 during MWA commissioning and has been functioning smoothly pushing large volumes of data across Australia and the Tasman to New Zealand.

The provision of the NGAS node in collaboration with ICRAR staff has played a tremendous role in improving the research productivity of the astrophysics researchers in NZ.

Professor Melanie Johnston-Hollitt, Director of Astrophysics at VUW and Chair of the MWA Executive Board says "Working with Andreas and his team to set up this data system has been incredibly easy and efficient and has allowed us to work directly on the 'big data' produced by the MWA. It has been a real pleasure to develop the partnership with ICRAR on the technical side of the big data equations and then to take those data and discover new and exciting things about the Universe."

International
Centre for
Radio
Astronomy
Research

ICRAR HQ
7 Fairway, 2nd Floor
The University of WA
Crawley, Perth

Postal Address
The University of WA
ICRAR M468
35 Stirling Hwy, Crawley
Perth, WA 6009

Contact
Dr Renu Sharma,
Associate Director &
Chief Operating Officer

E Renu.Sharma@icrar.org

T +61 (0)8 6488 7871