DATA INTENSIVE ASTRONOMY

Imagining leading edge solutions for unimaginable volumes of data.
Data Intensive Astronomy
Observing 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.

But 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.

As such, collaborative alliances of organisations and nations are being formed to fund and build 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. Some of these challenges will be in common amongst organisations and industries exploring our world and the Universe in which we live.

In all cases, the challenges of managing, exploring and sharing the huge volumes of digital information flowing from these new global facilities and programs, is focusing and leading the international discussion.
THE SKILLS AND EXPERIENCE DEVELOPED WITHIN OUR DATA INTENSIVE ASTRONOMY TEAM ARE FOCUSED ON EXPLORING THE UNIVERSE USING BIG DATA.

BUT MANY OF THE CHALLENGES WE FACE WILL ALSO BE SHARED BY THOSE EXPLORING OUR PLANET TO FIND NEW RESOURCES.

ICRAR IS SEEKING THE OPPORTUNITY TO WORK COLLABORATIVELY WITH EXPLORERS FROM INDUSTRY.

DATA TEAM

PROFESSOR ANDREAS WICENEC

Professor Wicenec and his team are working with astronomers and astrophysicists from around the world to design and develop innovative solutions for database management, data storage and high performance computing.

After completing a PhD in Physics and Astronomy at the University of Tübingen in Germany, Andreas went on to develop photometric and astrometric data reduction systems for the European Space Agency’s Hipparcos satellite. Then, as archive scientist for the European Southern Observatory (ESO), Andreas was involved in the implementation of the archive for ESO’s Very Large Telescope. Since then Andreas has led the archive subsystem development group for the ALMA radio telescope in Chile and been involved in the International Virtual Observatory Alliance.

Currently 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.

MARKUS DOLENSKY

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

Originally from Austria, Markus started as an IT consultant and 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.

After graduating from the Vienna University of Technology, Markus took a position with GePARD as a software engineer. A highlight was the development and integration of the
The Astronomical Data Deluge

The volume of data generated by new and planned observatories is currently doubling every six to 12 months. This is faster than the measured rate of increase in performance of computer chips (Moore’s Law), which doubles every 18 months. More critically, storage hardware access rates (megabytes/second) are at a relative standstill, presenting a major bottleneck to the transfer of large data sets.

The widening gap between the end user and the source of the data (in terms of processing capabilities and download time) is driving a major paradigm shift in the way large research data sets should be processed and accessed. This new paradigm will likely mean that large data sets and computational resources are concentrated at a number of data centres.

Through a new software infrastructure, end users will transparently interact with these distributed resources in a similar manner to the transparency of data access for the existing World Wide Web.

However, unlike the Web, data will not be migrated to end users but rather accessed, processed and explored remotely across a network of distributed data and computational service providers.

The explosion in data volume is driving the development of new software tools and mathematical algorithms that will operate in this distributed resource and service environment.

Astronomy has been one of the leading communities in the international effort to develop these distributed data access and processing infrastructures through initiatives like the International Virtual Observatory Alliance (IVOA) and, more recently, through the prototyping of cloud-based large-scale data management and data processing.

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 ICRAE 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 Multirrole Helicopter, and was the lead designer for an Intelligent Transportation System network for Queensland Motorways.

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

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.

Mark Boulton has a wealth of expertise encompassing IT security, systems engineering, systems integration and engineering management.
ICRAR
The International Centre for Radio Astronomy Research (ICRAR) was founded in 2009 as a joint venture of Curtin University and The University of Western Australia with additional funding from the State Government of Western Australian.

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 in the top five centres of its kind in the world.

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

In particular, ICRAR has developed a world-leading team of data intensive innovators and scientists who are now leading the effort to design the SKA data system. The SKA will be capable of producing a stream of science data products that are exascale in terms of their storage and processing requirements. This Google-scale enterprise is attracting considerable international interest and excitement from within the industrial and academic communities.

Within ICRAR, the Data Intensive Astronomy (DIA) team is leading the international effort to address the technical challenges surrounding the flow of data within the SKA observatory. The DIA team is composed of researchers from astronomy and industry who have led innovative international efforts like the IVOA, as well as the development of data and operations systems for billion-Euro astronomical infrastructures in Europe and South America.

THE SKA WILL PRODUCE 100 TIMES MORE DATA THAN THE CURRENT GLOBAL INTERNET TRAFFIC.

DATA TEAM
IAN COOPER
Before joining ICRAR, Ian Cooper was based in the UK and worked in the satellite communications industry for more than 20 years.

As a Product Manager Ian 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 DARA 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 ICRAR’s Deputy Project Manager for the SKA Science Data Processor.

DATA TEAM
DR RICHARD DODSON
Dr Dodson is the Project Scientist for ICRAR’s Data Intensive Astronomy Group.

His expertise encompasses astrophysics and computer science, making him well suited to act as the interface between the ‘science’ and the ‘big data’ focus of the DIA.

With such a versatile skill set, Richard 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.

Currently, Richard is investigating possibilities of exporting 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.
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 Engineer working at ICRAR on various projects including the Murchison Widefield Array Data Archive and the Square Kilometre Array’s Science Data Processor.

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

Kevin is also developing new techniques for the automated classification of galaxies using multi-wavelength data and working on new advances in Spectral Energy Distribution calculations.

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

As well as contributing to the design process for the SKA’s science data processor, Kevin is providing data support for large galaxy surveys such as GAMA and CHILES and is the project scientist for theSkyNet, a citizen science project.
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 conduct a PhD in data retrieval and storage and postdoctoral research in pattern recognition.

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

Chen is leading the Data Lifecycle Management and Persistent Storage WP-tasks for the SKA-Science Data Processor (SDP) 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.


In 1995, Peter accepted a position as Division Head of the newly formed Data Science team that links thousands of computers around the world to simulate a powerful supercomputer.

AMAZON
CISCO
DDN
INTEL
SGI
SYSTEMIC
THOUGHTWORKS

Current Projects

SKA SCIENCE DATA PROCESSOR
The SKA 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.

Our team are responsible for something called 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.

The SDP Data Layer architecture at its core is a fairly generic design and is applicable to a whole range of data intensive and data driven applications.

SURVEY SCIENCE SUPPORT
The 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. Our 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 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.

Our current list of collaborating partners includes Amazon, CISCO, DDN, Intel, SGI, Systemic and Thoughtworks.
CASE STUDY:
THE NATIONAL RADIO ASTRONOMY ORGANISATION

ICRAR’s 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.

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.

Management and Operations Division at the European Southern Observatory headquarters in Munich. While at ESO, Peter lead 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 in June 2005.

During his time at ESO, Peter co-founded the International Virtual Observatory Alliance (ivoA) in 2002, directed the FP-5 Astrophysical Virtual Observatory project and coordinated the formation of the EURO-VO as a program to realise VO-enable 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 2006. In 2008, Peter was appointed inaugural director of the new International Centre for Radio Astronomy Research (ICRAR). Peter became WA Scientist of the Year in 2012 and was made a Fellow of the Australian Academy of Technological Sciences and Engineering in 2013.

During his time at ICRAR, Peter co-founded the International Virtual Observatory Alliance (ivoA) in 2002, directed the FP-5 Astrophysical Virtual Observatory project and coordinated the formation of the EURO-VO as a program to realise VO-enable 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 2006. In 2008, Peter was appointed inaugural director of the new International Centre for Radio Astronomy Research (ICRAR). Peter became WA Scientist of the Year in 2012 and was made a Fellow of the Australian Academy of Technological Sciences and Engineering in 2013.

CASE STUDY: THE NATIONAL RADIO ASTRONOMY ORGANISATION

SCIENCE TEAM
DR MARTIN MEYER

Dr 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 focussed his research efforts on the development of new deep surveys to
understand the evolution of neutral gas across cosmic time. He is the Principal Investigator of the DINGO HI survey on the Australian SKA Pathfinder 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 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, for which he spent a large fraction of his time at CSIRO Astronomy and Space Sciences (CASS) in Sydney, Australia. 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 he is actively involved in several ongoing and future HI surveys that will all process large data volumes. The most significant ongoing survey 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 TOBIAS WESTMEIER**

Dr Westmeier is an astronomer at the University of Western Australia, working as a Research Fellow at ICRAR. He 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 in 2010.

Tobias is the initiator and leader of the 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.
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 VOSKA collaboration.

**CASE STUDY:**

**THE INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE**

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.

**SCIENCE TEAM**

**DR NASTASHA HURLEY-WALKER**

Dr Natasha Hurley-Walker joined the Curtin University node of ICRAR in 2011.

During her PhD studies at the University of Cambridge, 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 performed a critical role in bringing the Murchison Widefield Array (MWA) online, before conducting 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.
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 – endpoints 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.”

CASE STUDY: VICTORIA UNIVERSITY

Lorem ipsum dolor sit amet, sed ei praesent theophrastus. Vel torquatos interpretaris no, pri liber tritani ad, vis nibh brute dolorem id. Eam cu novum salutandi, lorem tibique deterruisset nam eu. Pri ut dolor congue, nam illud facete officis ex. Alii equidem debitis ius et, id mel brute luptatum.

Eam eu euismod delicata, ad justo aperian sed. Eos stet essent vituperatoribus ut, autem harum delenit vel at. Vis dolores argumentum at. Eum ei alia sonet corpora, eam quot utroque molestiae no. Eu sea dicta simul, summo iusto dolorum et duo.

Solum pericula reprimique pro ea, sit an stet petentium, eam mandamus deterruisset necessitatibus id. Mel nominavi prodesset percipitur cu, qui melius menandri ea, ei nec doctus commodo. Vel in facer copiosae petentium. Est diam dicunt ex, equidem consequat nam in. Debet deserunt urbanitas cum ad, qui magna blandit convenire ea.

Prima mnesarchum ei mea, an possess diligendi cotidieque nec. Nibh graece virtute ad sed. Ut illum tamquam vim. Nisl movet scribentur vel eu. Sea again quas torquatos ei.