High Performance Radio Astronomy @Pawsey: past, present and future

Swiss SKA Days 2023

Ugo Varetto, *Chief Technology Officer - ugo.varetto@csiro.au*

Pawsey Supercomputing Research Centre

Agenda

Pawsey

Radio astronomy

R&D

Collaboration

The Pawsey Supercomputing Centre

pawsey

Pawsey Supercomputing Centre

200+ Research Projects

42 Research Institutions

15+ Scientific Fields

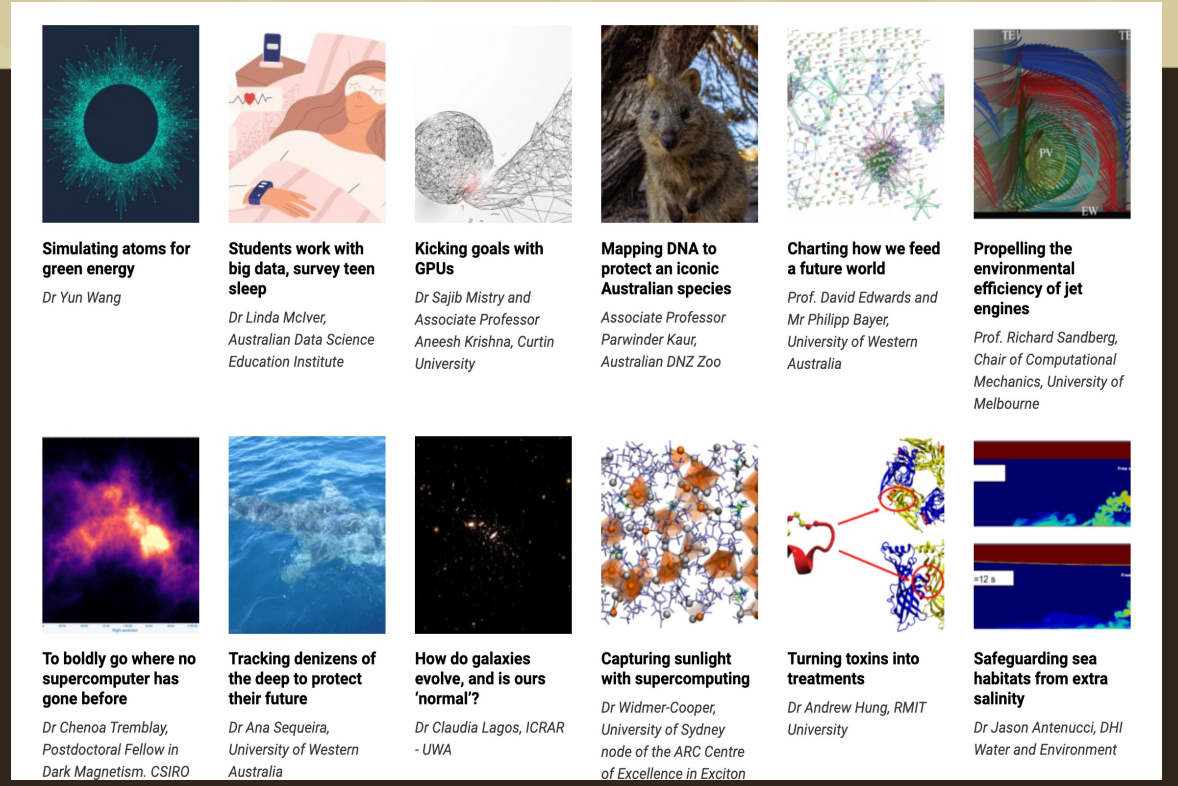
194 Training Programs & Events

International collaborations

Quantum technology hub

Radio astronomy operations

24/7 Data Ingest



Simulating atoms for green energy
Dr Yun Wang

Students work with big data, survey teen sleep
Dr Linda McIver, Australian Data Science Education Institute

Kicking goals with GPUs
Dr Sajib Mistry and Associate Professor Aneesh Krishna, Curtin University

Mapping DNA to protect an iconic Australian species
Associate Professor Parwinder Kaur, Australian DNZ Zoo

Charting how we feed a future world
Prof. David Edwards and Mr Philipp Bayer, University of Western Australia

Propelling the environmental efficiency of jet engines
Prof. Richard Sandberg, Chair of Computational Mechanics, University of Melbourne

To boldly go where no supercomputer has gone before
Dr Chenoa Tremblay, Postdoctoral Fellow in Dark Magnetism. CSIRO

Tracking denizens of the deep to protect their future
Dr Ana Sequeira, University of Western Australia

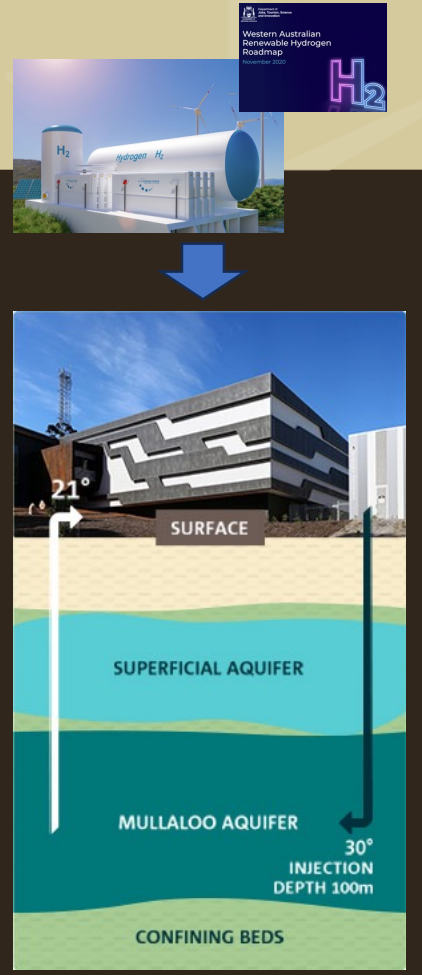
How do galaxies evolve, and is ours 'normal'?
Dr Claudia Lagos, ICRAR - UWA

Capturing sunlight with supercomputing
Dr Widmer-Cooper, University of Sydney node of the ARC Centre of Excellence in Exciton

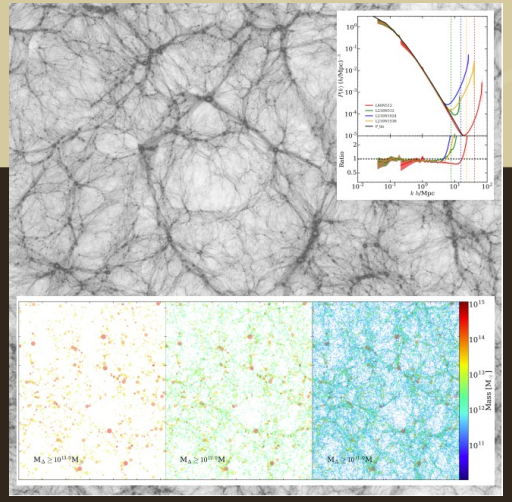
Turning toxins into treatments
Dr Andrew Hung, RMIT University

Safeguarding sea habitats from extra salinity
Dr Jason Antenucci, DHI Water and Environment

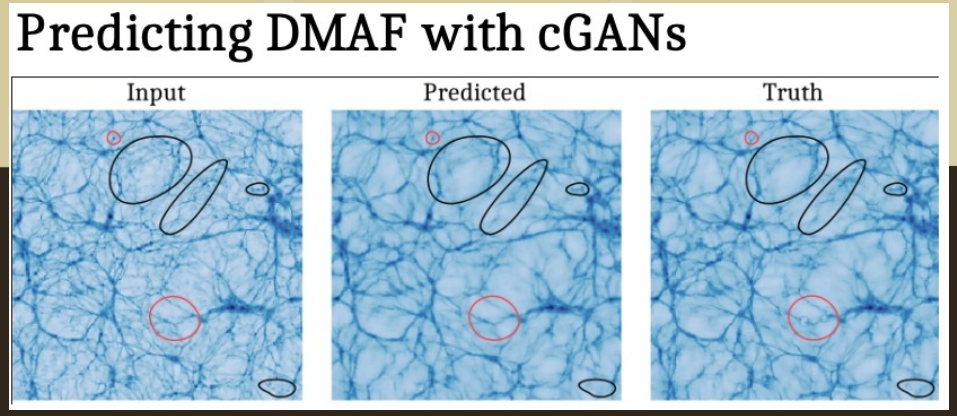
https://pawsey.org.au/case_studies

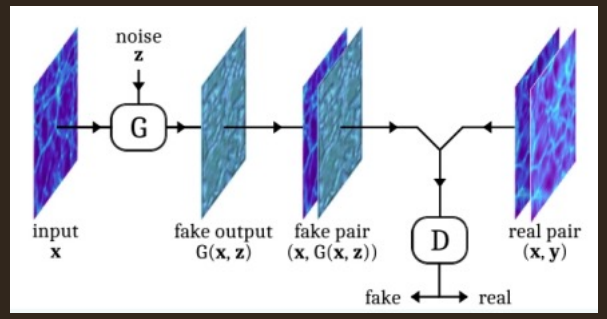


Astronomy @Pawsey

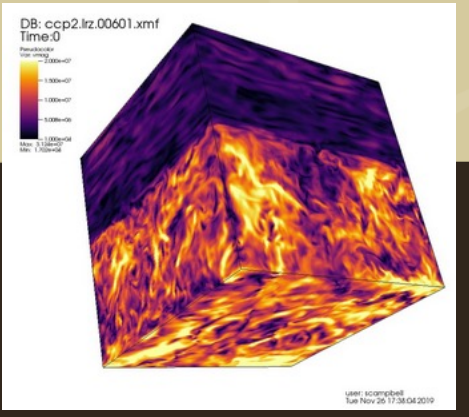


SURFS: Riding the waves with Synthetic UniveRses For Surveys
Pascal J. Elahi et al.

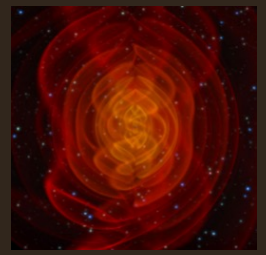




A Novel Scheme for Dark Matter Annihilation Feedback in Cosmological Simulations
Floarian List, Geraint F.Lewis, Nikolas Iwanus, **Pascal Elahi**, Ishaan Bhat

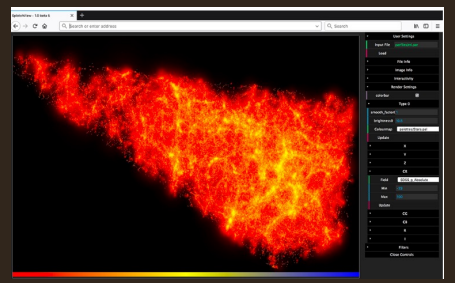


Convective-Reactive Nuclear Burning and Turbulence Boundaries in Stars
Simon Campbell - Monash University



Real-Time Gravitational Wave Search
Prof. David Blair & Prof. Linqing Wen, University of Western Australia

Real-time Web-based Remote Interaction with Active HPC Applications
Tim Dykes, Ugo Varetto et al.



Advanced architecture

Nimbus (Cloud)

100 Gbps

Setonix Supercomputer

100 Gbps

ASKAP Ingest

10 Gbps

Garrawarla (MWA)

10 Gbps

400 Gbps

100 Gbps

100 Gbps
External Firewall

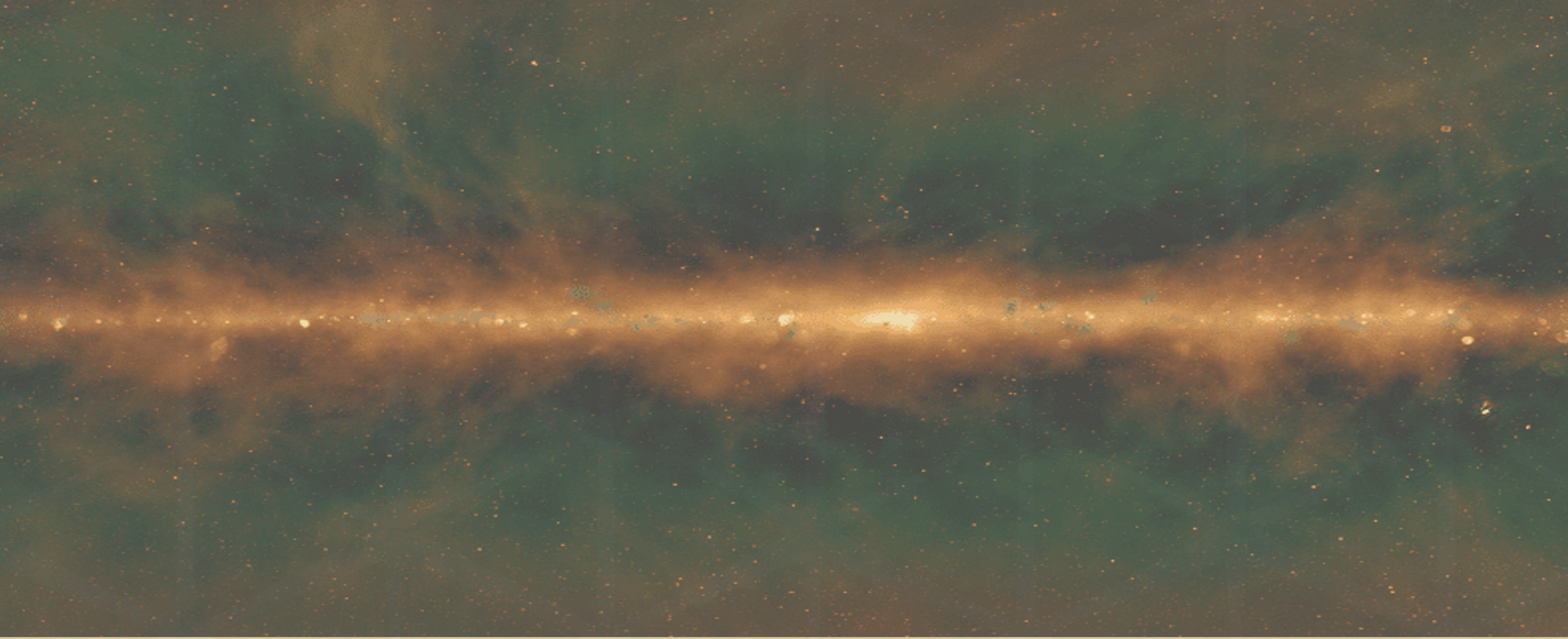
S3 Interface

Fibre Channel

200 PB tape system

2x100Gbps

130 PB Ceph Object Store



https://public.nrao.edu/radio-astronomy/the-science-of-radio-astronomy/

Radio astronomy

pawsey

Radio astronomy services

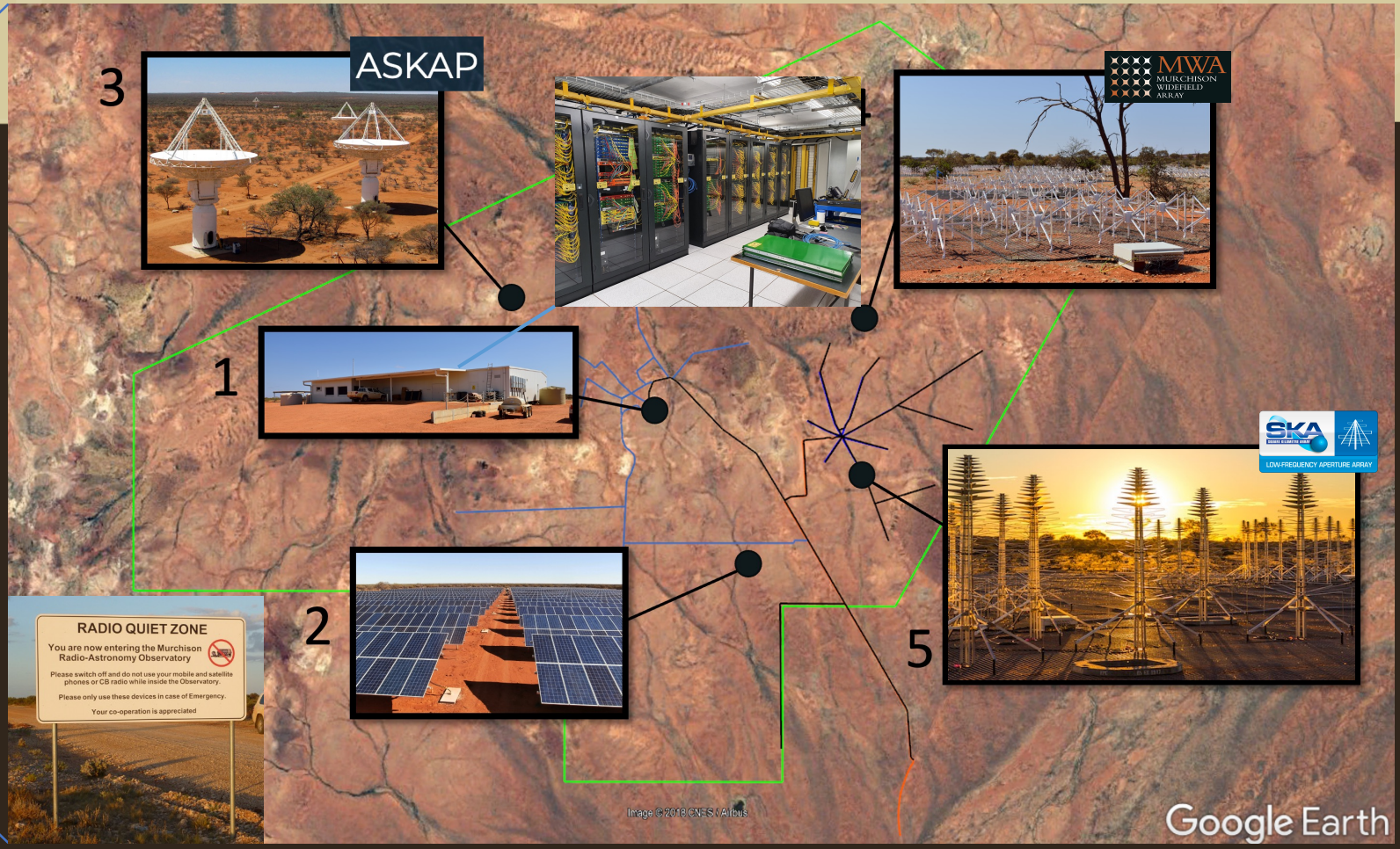
- Data ingest
- Pre/post-processing
- Storage
- Data lifecycle management
- Data sharing (FAIR)

Operations

- Visualisation
- Data analysis

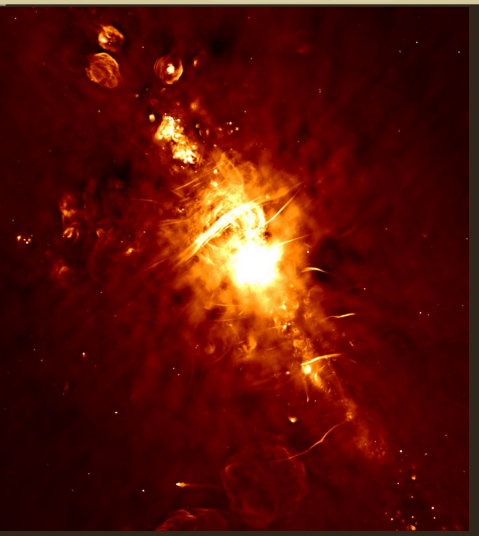
Science

Murchinson Widefield Array



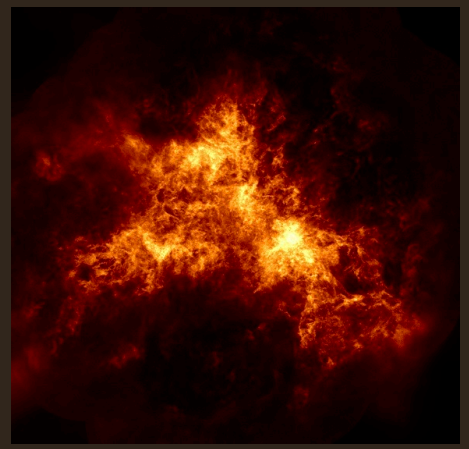
Australian Square Kilometre Array Pathfinder

700 MHz – 1.8 GHz, ~cm wavelength

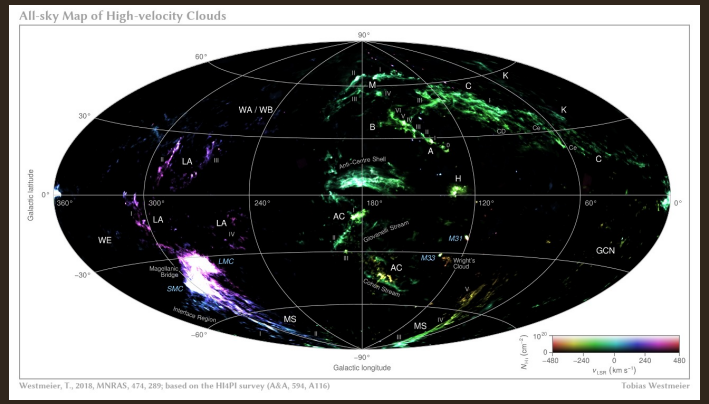


Centre of the Milky Way Galaxy, 28 antennas: *ASKAPSoft*/Wasim Raja.

Credit: *Wasim Raja and Pascal Jahan Elahi, CSIRO, Pawsey*
https://www.theguardian.com/australia-news/2022/aug/11/australian-supercomputer-produces-fantastic-picture-of-supernova-remnant

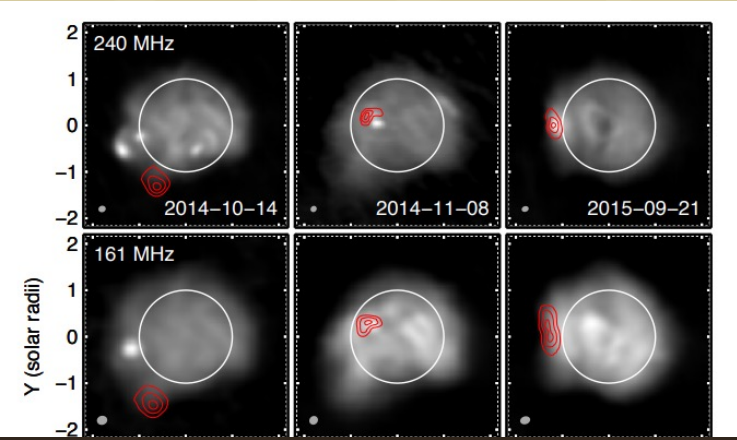


Atomic hydrogen gas in the Small Magellanic Cloud as imaged with CSIRO's Australian Square Kilometre Array Pathfinder (ASKAP). Credit: ANU and CSIRO

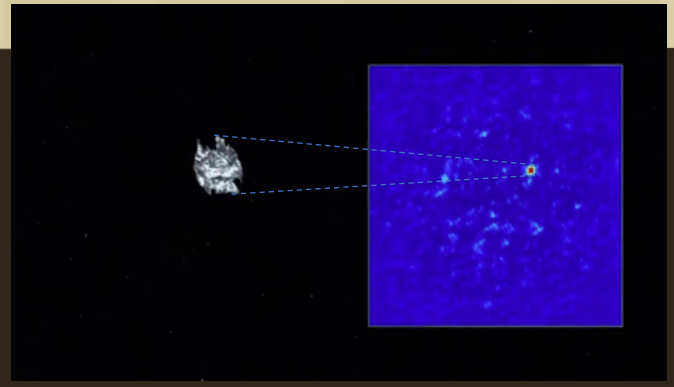


Murchinson Widefield Array

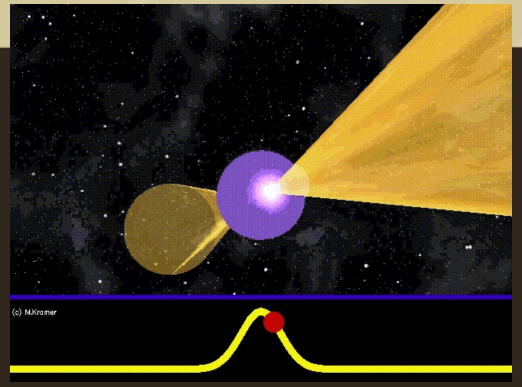
70-300 MHz, ~m wavelength



Solar Imaging

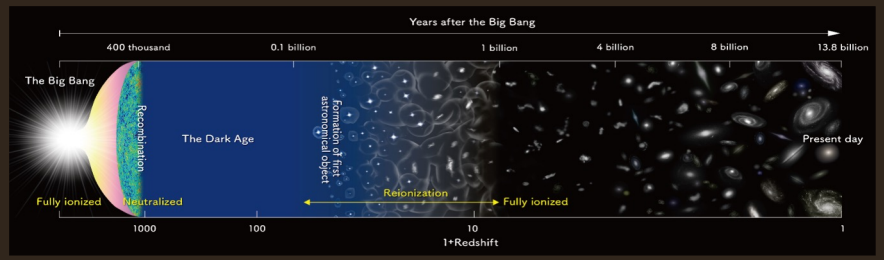


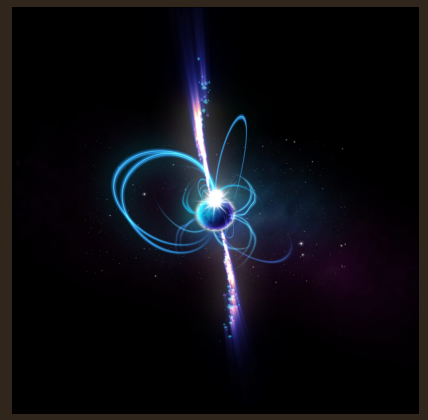
Space Situational Awareness



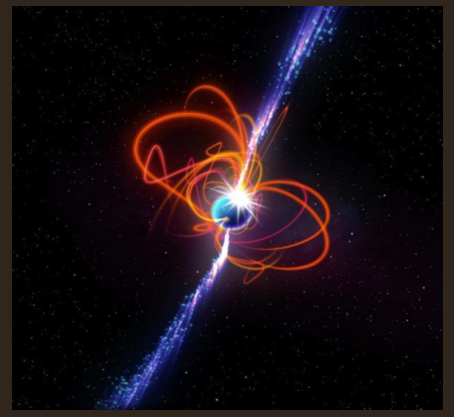
Fast radio bursts

Epoch of Re-ionisation

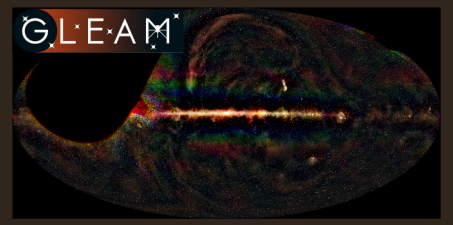




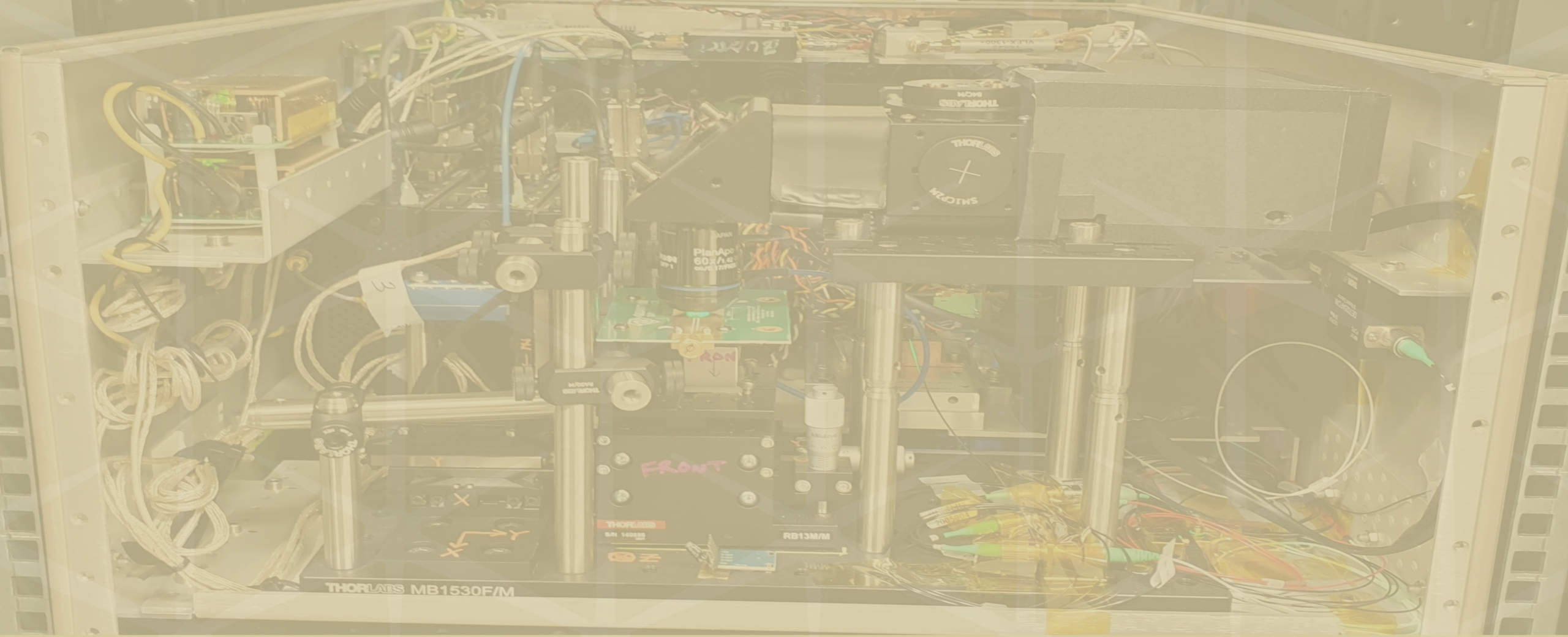
Credit: Natasha Hurley-Walker (ICRAR/Curtin University)
https://www.icrar.org/repeating-transient/



Credit: Natasha Hurley-Walker (ICRAR/Curtin University)
https://thewest.com.au/news/wa/wa-team-lead-by-dr-natasha-hurley-walker-discover-what-could-be-rare-star-magnetar-hidden-in-plain-sight--c-11325087



Galactic and Extragalactic All-Sky MWA Survey



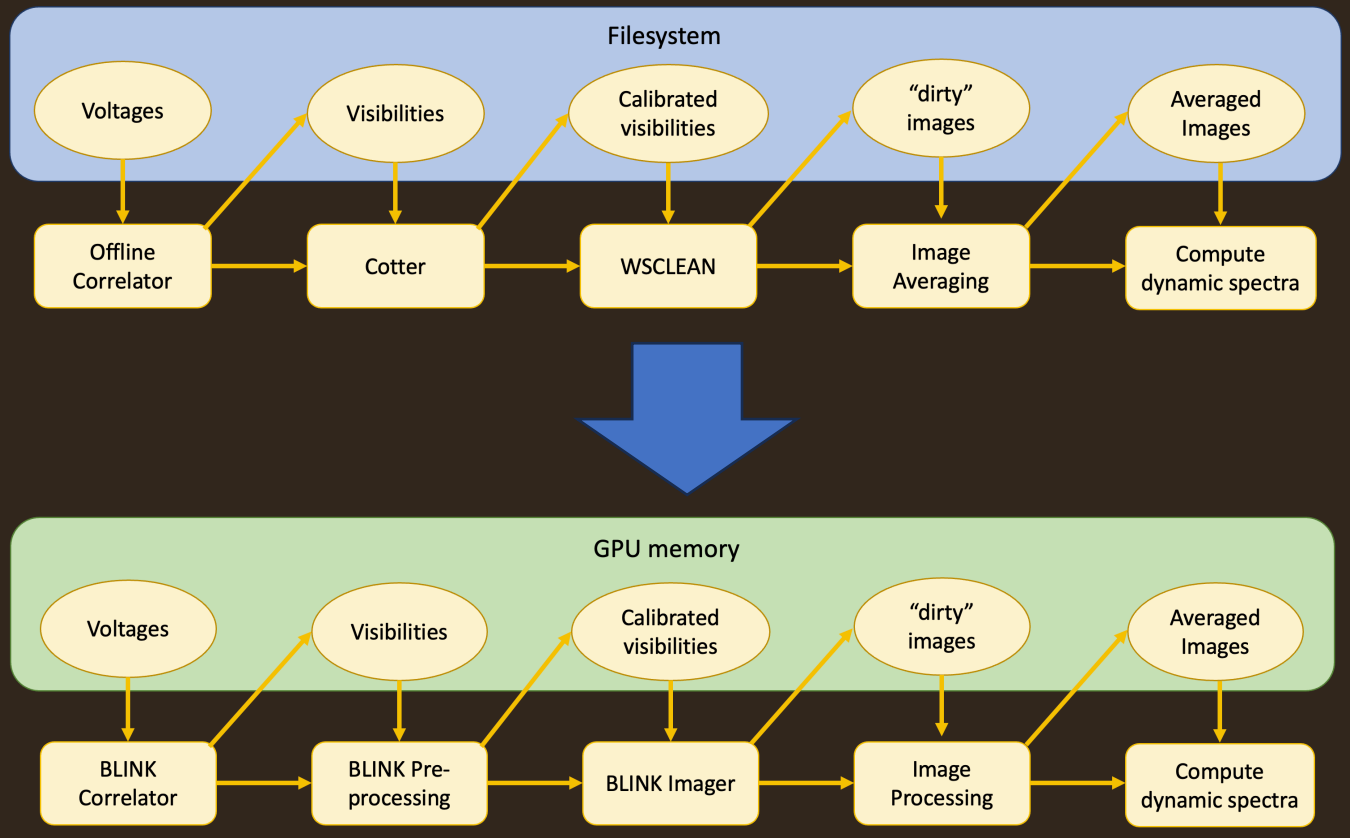
Research and Development

pawsey

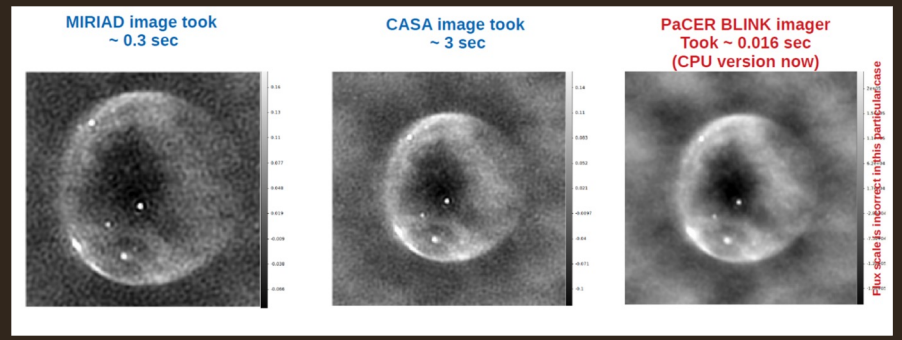
BLINK

BLINK and you'll miss it -
Breakthrough Low-latency Imaging with Next-generation Kernels

Science goal: to find FRBs (ms) in the MWA Phase 1 archival data stored at Pawsey.



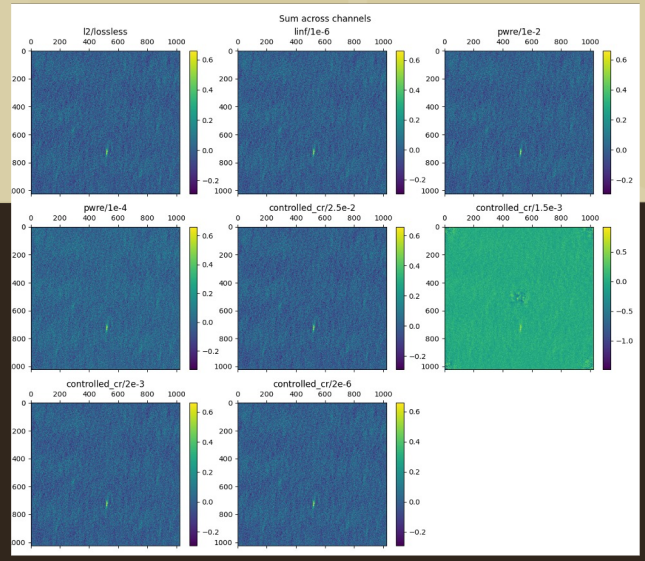
- High time resolution (10ms) and high frequency channel resolution (10KHz) required.
- High resolution requirements translate into ~30 million images generated per 100s of MWA observation. Image data has to be rearranged in the dynamic spectra (image to the left).
- Current data processing pipeline is too slow to keep up with the data rate (40TB per 1hr observation).
- Producing 1024x1024 images over all the 3072 frequency channels of MWA at 10ms will result in 1.29 TB/s



I/O

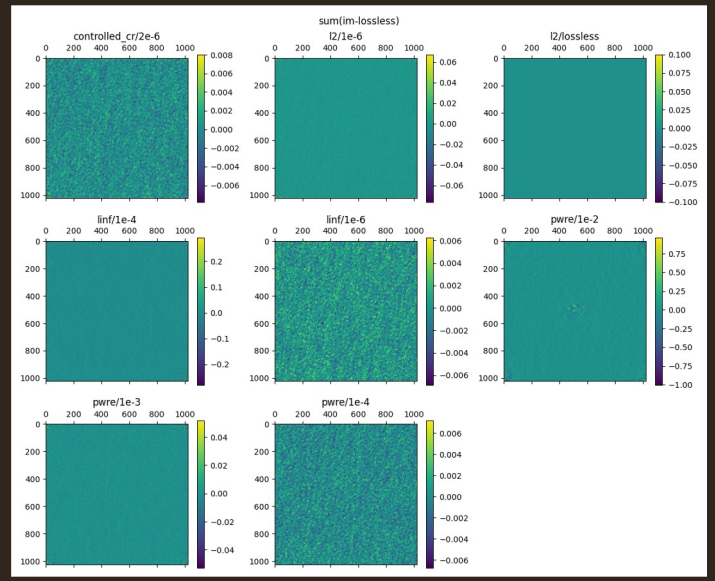
Compression rate:
Lossless: 6x → 8x
Lossy: 15x → 20x

- Optimise and minimise
- Integration with ADIOS2
- In memory processing
- Lossless and lossy (de-)compression as close as possible to the data
- Scheduling and resource allocation (PhD project, DALiuGE built-in)
- Parallel data access
- Feedback of workload characterization at run-time

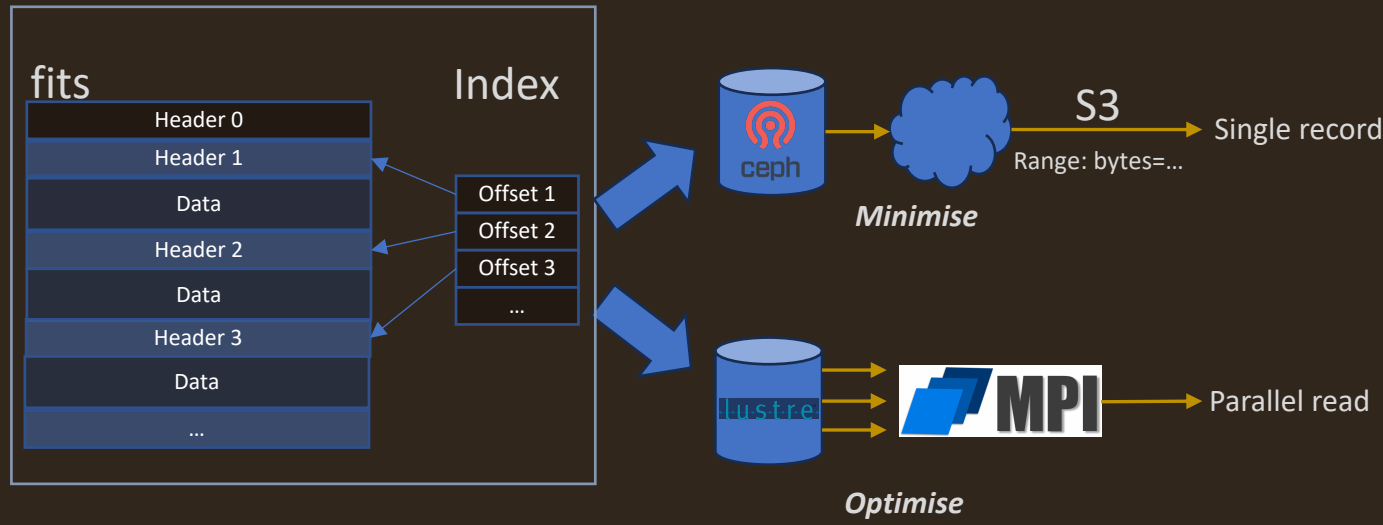


Actual data

Credit: Andreas Wicenec, ICRAR/UWA
https://www.icrar.org/study-with-icrar/postgraduate-opportunities/postgraduate-research-projects/data-intensive-astronomy/delivering-the-goods-data-intensive-ska-scale-astronomy-copy/

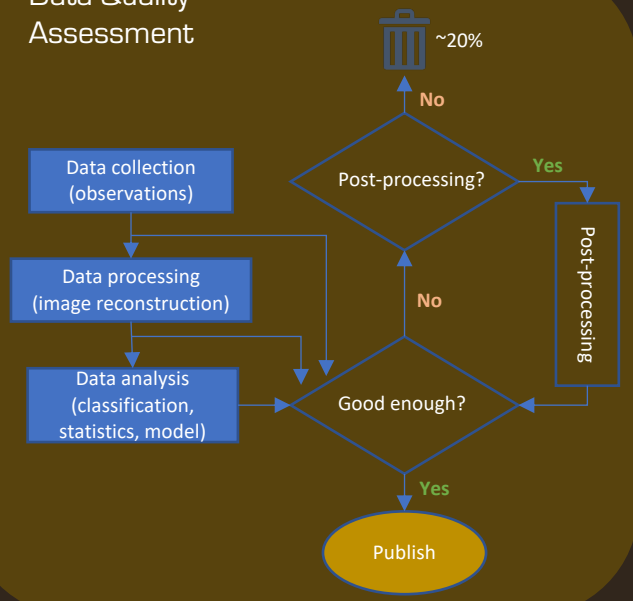


Residuals

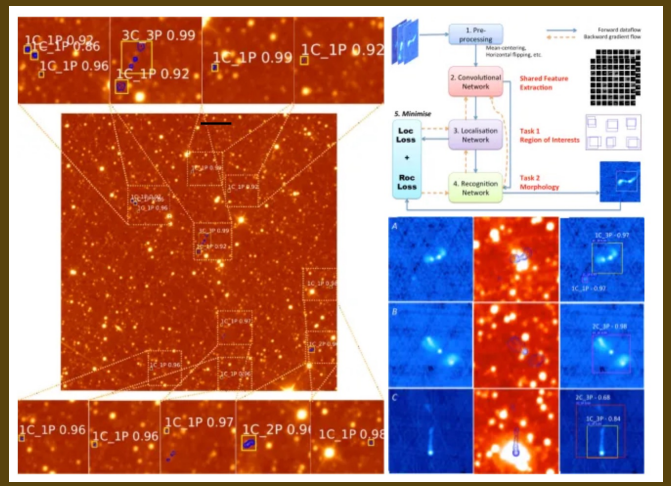


Machine Learning

Data Quality Assessment

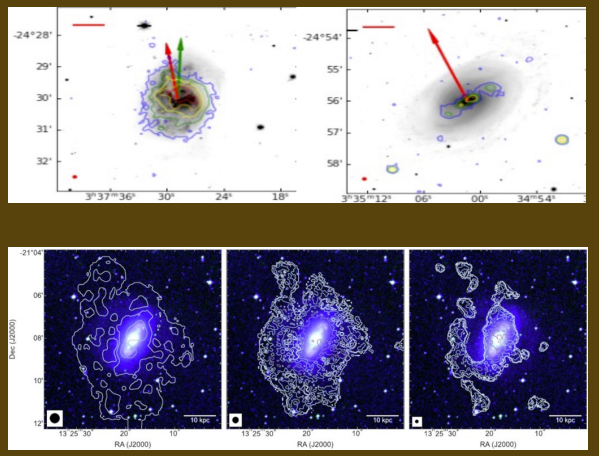


Automated classification & identification



http://arxiv.org/abs/1805.12008

Post-processing



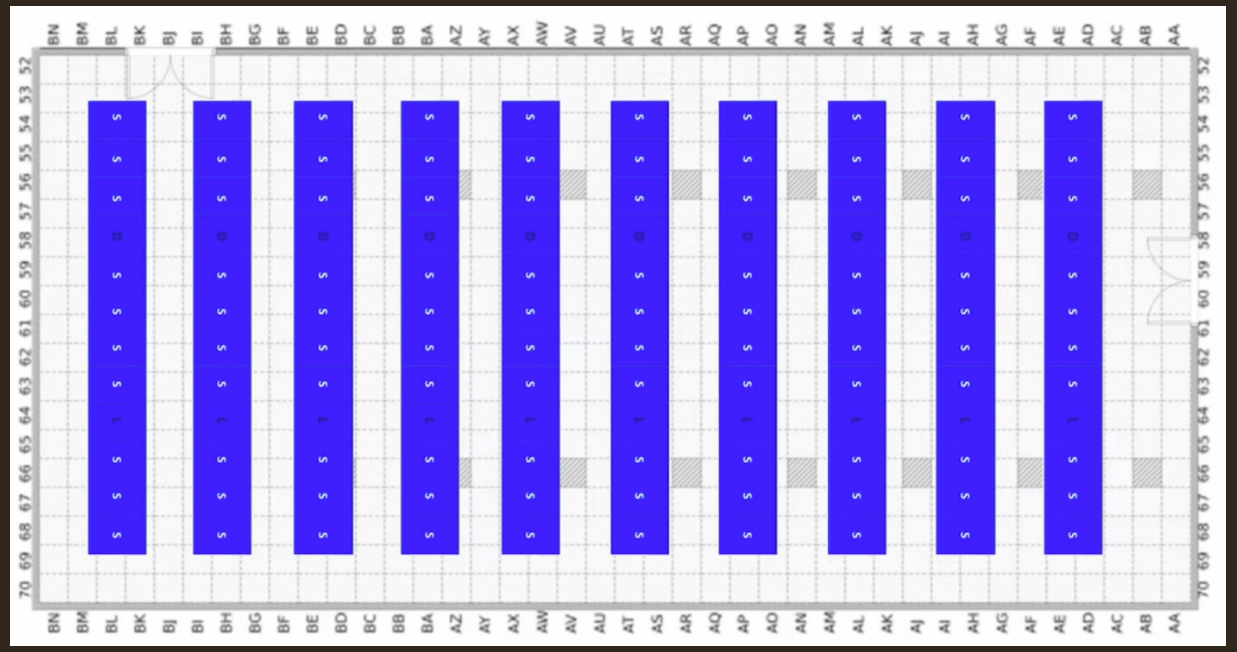
Credit: Joe Grundy, Chandra Murugeshan (CSIRO)

Storage

200 PB mirrored, uncompressed
1 PB/day ingestion rate
5 PB cache
S3 Interface

30+ EB
10+ PB/day ingestion rate
Per-library cache
~Computational storage/DPU
5-10 PB cache
S3 Interface

11.4m x 24m (273.6m²)

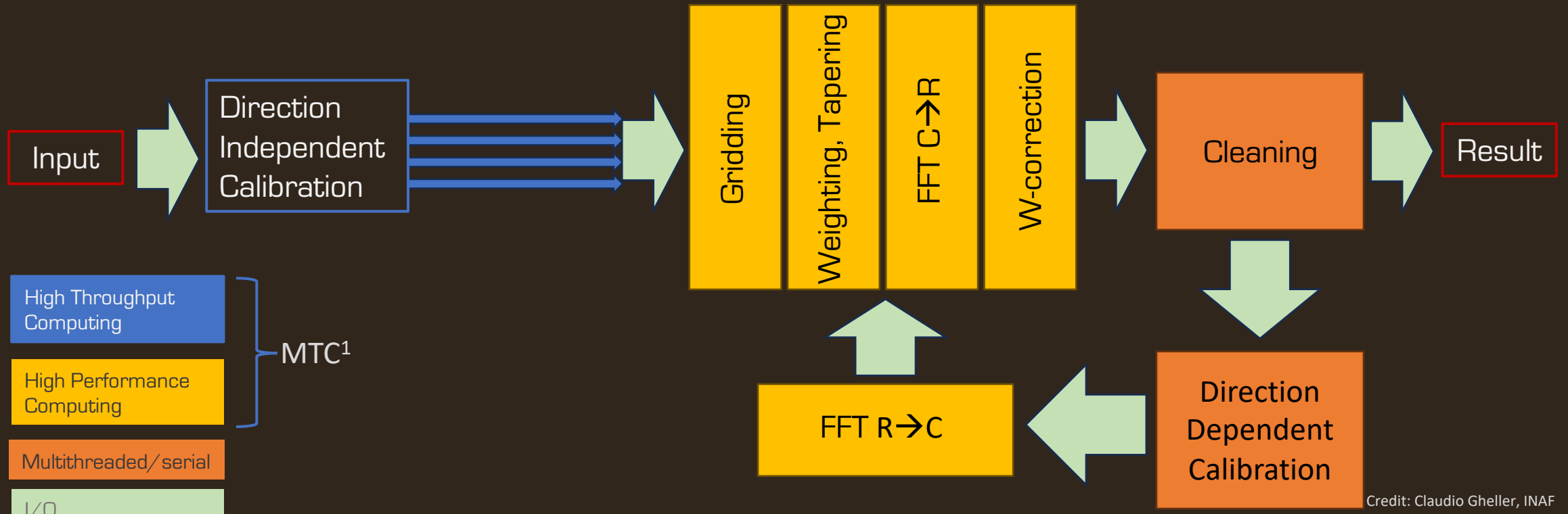


Total JF tape capacity of 26.3 Exabytes (EB) compressed or 8.76 Exabytes (EB) uncompressed

High Performance Workflows

StreamFlow: https://streamflow.di.unito.it/

Common Workflow Language: https://www.commonwl.org/



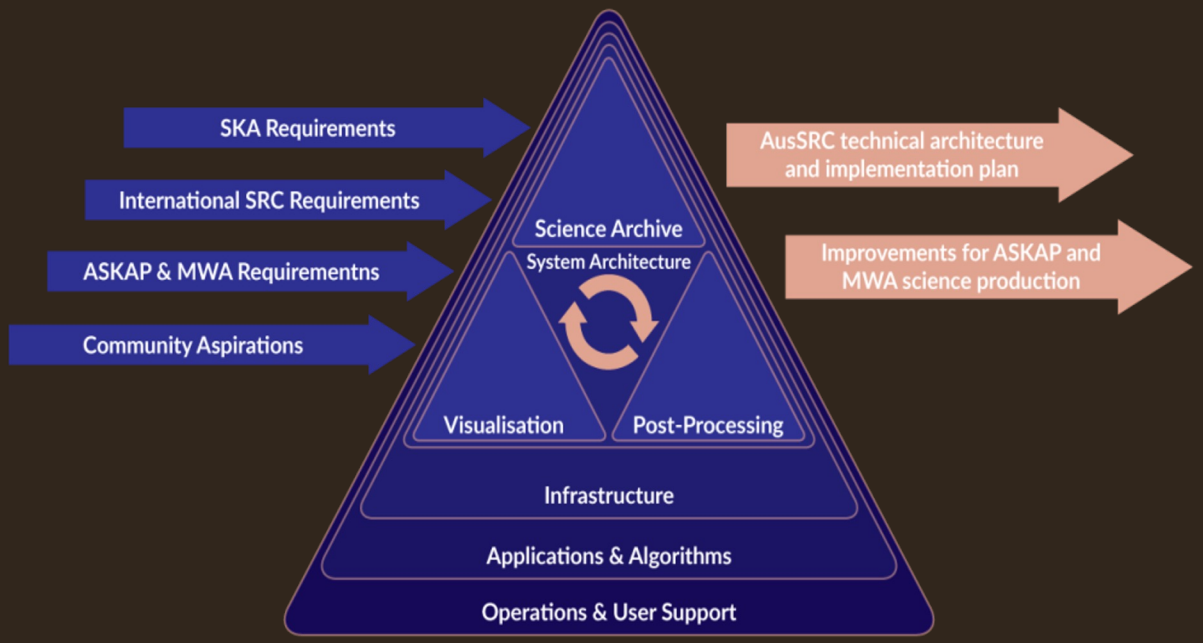
Credit: Claudio Gheller, INAF

[1] MTC = (HTC U HPC); Many Task Computing = High Throughput Computing U High Performance Computing; I. Raicu, University of Chicago, Doctorate Dissertation, March 2009

I. Colonnelli, B. Cantalupo, I. Merelli, and M. Aldinucci, "StreamFlow: cross-breeding cloud with HPC," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, iss. 4, p. 1723–1737, 2021. doi:10.1109/TETC.2020.3019202

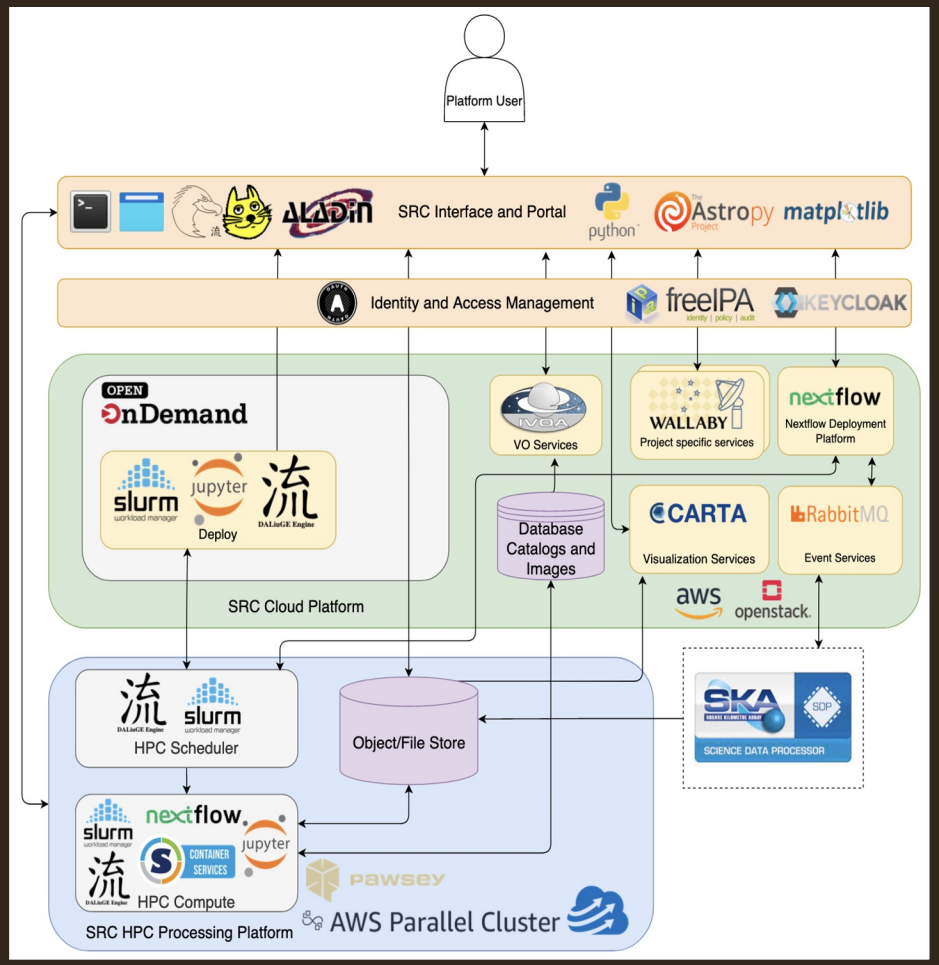
AusSRC

Design study



vCPUs	2040
Memory	8 TB
Storage	0.5 PB

Prototype



SKAO

~Cluster received (AAVS3, AA0.5)

Storage space allocated

Network configured

Collaboration agreement in progress

Stage	#	SKA-Low (Date)	SKA-Mid (Date)
Aperture Array Verification System 2 (AAVS2)	1 STATION (256 ANTENNAS)	NOW	N/A
Aperture Array Verification System 3 (AAVS3)	1 STATION (256 ANTENNAS)	~Sept 2023	N/A
Array Assembly 0.5 (AA0.5)	4 DISHES 6 STATIONS	2024 Aug	2024 Dec
Array Assembly 1 (AA1)	8 DISHES 18 STATIONS	2025 Oct	2025 Nov
Array Assembly 2 (AA2)	64 DISHES 64 STATIONS	2026 Sep	2026 Oct
Array Assembly * (AA*)	144 DISHES 307 STATIONS	2028 Jan	2027 Aug
Operation Readiness Review		2028 Apr	2027 Nov
End of Staged Delivery Programme		2028 Jul	2028 Jul
Array Assembly 4 (AA4)	197 DISHES 512 STATIONS	TBD	TBD

Collaboration

- FRB imaging pipelines
- Observation of diffuse cluster emission
- Platform design and deployment schemes (SRC)
- Data storage
- AI
- Data reduction and analysis
- Distributed processing (Serverless, Microservices):
  1. Send processing pipeline to Pawsey
  2. Compute and trigger transfer to remote site (CSCS?)
  3. Continue computation locally
  4. (Sync results with other sites)
- Quantum computing
- Staff exchange

Credit: Dr. Emma Tolley, EPFL

THANK YOU