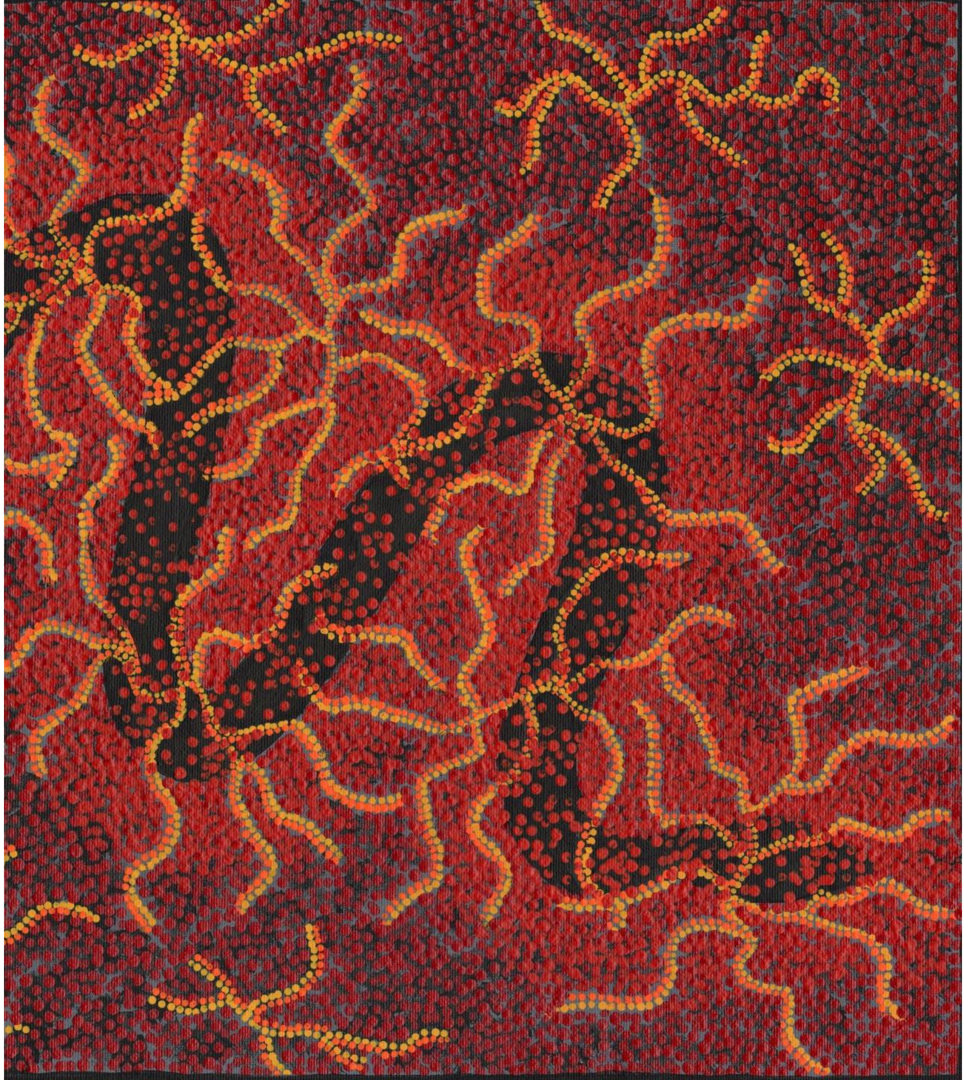
ATUC: Session 4 Software, computing, data and pipelines

ATUC meeting
9 April 2025

Susan Merry, *Data,* 2019



ASKAP Key Capabilities Project
Aidan Hotan

(7 + 3 minutes)

The ASKAP Key Capabilities Project

- ASKAP made its first attempt at full survey operations for the combined Survey Science Projects in November 2022
  - Issues transitioning from the Galaxy supercomputer to its successor Setonix forced many halts in observing, some extending to several months
    - *ASKAP's observing is tightly coupled to processing due to limited storage*
  - Ongoing consolidation and development required alongside operations
- The Key Capabilities project was launched on 1 September 2024 to prioritise and resource essential development tasks
  - We can’t make up all the lost time, but we can focus effort where needed

Project roadmap key areas

Meet our 70% uptime KPI	• Firmware-level data integrity, domino watchdog & masking, raw data monitoring, error avoidance, beamforming improvements, drives efficiency, antenna exclusion logic
Reduce imaging artifacts	• Bright source subtraction, self-cal using the RACS sky model, improved CLEAN masking, use of wide-band information when CLEANing narrow channels, mitigation of solar interference
Improve calibration	• Implement field-based bandpass calibration using RACS sky model, fix astrometry with routine field-based phase calibration, increase overall quality and efficiency of calibration process
Increase data throughput	• Optimise ASKAPsoft, collaborate with Pawsey to improve workflow and Setonix platform performance so processing keeps up with observing in all modes
Commensal fast-imaging	• New feature required for transient science goals. Includes creating Stokes V source catalogues

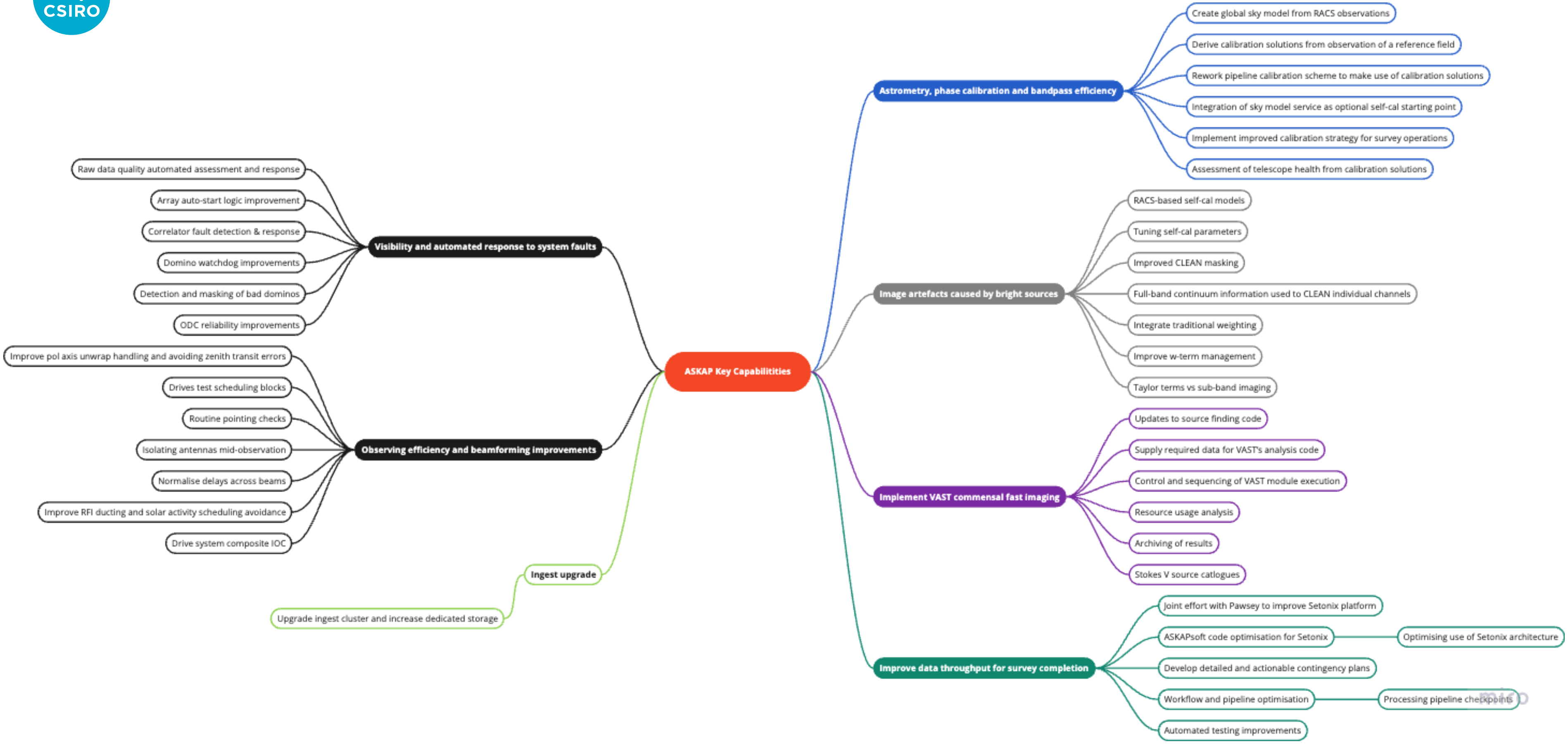


Image provided by Eric Bastholm

Project roadmap key areas

Meet our 70% uptime KPI

- Firmware-leve[illegible]y, domino watchdog & masking, raw data monitoring, error avoidance, bea[illegible]provements, drives efficiency, antenna exclusion logic

Reduce imaging artifacts

- Bright source subtraction, self-cal using the RACS sky model, improved CLEAN masking, use of wide-band information when CLEANing narrow channels, mitigation of solar interference

Improve calibration

- Implement field-based bandpass calibration using RACS sky model, fix astrometry with routine field-based phase calibration, increase overall quality and efficiency of calibration process

Increase data throughput

- Optimise ASK[illegible]rate with Pawsey to improve workflow and Setonix platform performance [illegible]keeps up with observing in all modes

Commensal fast-imaging

- New feature [illegible]ansient science goals. Includes creating Stokes V source catalogues

Meet our 70% uptime KPI	•Firmware-level data integrity, domino watchdog & masking, raw data monitoring, error avoidance, beamforming improvements, drives efficiency, antenna exclusion logic
Increase data throughput	•Optimise ASKAPsoft, collaborate with Pawsey to improve workflow and Setonix platform performance so processing keeps up with observing in all modes
Commensal fast-imaging	•New feature required for transient science goals. Includes creating Stokes V source catalogues

Goals achieved

- Observation KPI exceeded (Vanessa’s talk)
  - Major improvements to PAF watchdog software system
  - More improvements planned, but we're **already exceeding our uptime KPI**
- Fast imaging mode integrated into the standard pipeline
  - Currently active on all EMU observations
- Data throughput greatly improved on Setonix
  - Pawsey worked with platform vendor to improve performance in Q4 2024
  - Optimisation of ASKAPsoft tools and pipelines has further improved throughput
  - No data backlog since November 2024 with spectral line modes active

Improving ASKAP science ready data quality

- Extensive feedback from SSPs received and triaged
  - We continue to learn more about ASKAP as surveys proceed
- ASKAPsoft updates available now or in testing (Matthew’s talk)
  - RACS sky model used to improve the first round of self-cal
  - Traditional weighting available as a selectable option
  - GASKAP-HI deep joint imaging improvements
  - Offset source subtraction to reduce artifacts in the observed field
  - Multiple improvements to ASKAPsoft CLEAN performance and quality

Improve calibration

- Implement field-based bandpass calibration using RACS sky model, fix astrometry with routine field-based phase calibration, increase overall quality and efficiency of calibration process

Improving ASKAP's calibration system

- Calibration is currently time consuming and constraining
  - Requires B1934-638 to be visible
  - Needs a few minutes per beam for a total of about 2.25 hrs
  - Must be done whenever beam weights are updated (e.g. switching frequency bands)
  - Must be done for every beam configuration (footprint) used
- With a high-quality sky model, we can improve significantly
  - Calibrate all beams simultaneously, reducing the total time to ~15 mins
  - Calibrate at any time on a variety of fields with no LST constraint
  - Calibrate more often to better track changing parameters and learn more about the system

Improve calibration

- Implement field-based bandpass calibration using RACS sky model, fix astrometry with routine field-based phase calibration, increase overall quality and efficiency of calibration process

Workflow steps

New weights trigger holography & 1934 bandpass	SAURON, a few times a year
Regular short observations of reference fields	SAURON, as needed, e.g. after a band switch, replacing bandpass observations. ~15 mins each.
Measure complex gains from a reference field	Processing Manager launches calculation when observation ends
Create updated bandpass solution	Processing Manager merges new gain solution with original 1934 bandpass and publishes an updated solution
Science pipelines use most recent update	Processing Manager tracks two classes of bandpass, the original 1934 and updated solutions

- Development work well underway, initial end-to-end testing should begin in May, with operational integration and commissioning planned for June / July
- *However, this is a new technique and involves a research component with some risk*

Sky model and astrometry referencing

- RACS_low2 is over 90% complete
  - The last epoch required for full frequency coverage in EMU closepack footprint
- The RACS team is investigating ways to tie RACS astrometry to other wavelengths and reference catalogues
  - RACS epochs are already being used to ensure *consistent* astrometry in EMU, the next step will also ensure *correct* astrometry
- This work leads towards a global sky model using all RACS epochs
  - Understand frequency evolution of sources across the entire ASKAP band

Future capabilities: the Flint pipeline

- We are developing an alternative to the ASKAPsoft pipeline
  - Motivation: test different design philosophies, new workflow management libraries and alternative imaging tools with a focus on ease of understanding, portability, user engagement and community contributions (Tim's talk)
    - Tested at RACS scale with promising results, improvement in RMS map
    - Compatible with VAST, EMU and POSSUM, other modes under investigation
    - Support for CASDA upload needs to be developed
  - Operational integration planned (selectable for GSPs and SSPs) this year
    - Reference field calibration will take priority

ASKAP Key Capabilities Summary

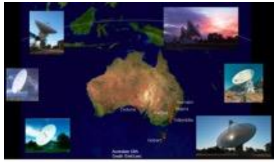
- Several factors blocking sustained survey progress are now resolved
  - Data throughput now keeping pace with observing
  - Time on-sky is exceeding our 70% KPI
- Two-part approach to ensuring data quality
  - Triaged improvements to ASKAPsoft with many improvements under test
  - Active development of Flint pipeline provides flexibility and future proofing
- New calibration system should improve survey efficiency
  - Reduce calibration overheads by 10x, abolish calibrator LST constraints

Data and data archiving

Minh Huynh

(7 + 3 minutes)

ATNF Archives Current Status



Continuum and Spectral Line

Pulsar

ATOA

- At Marsfield
- Volume: 0.6 PB
- Growth: 50 TB in last year

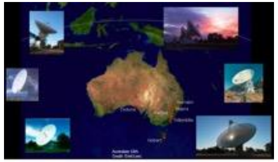
Parkes Pulsar DAP

- Part of CSIRO Data Access Portal
- Canberra, CSIRO Data centre
- Volume: 4.7 PB
- Growth: 0.6 PB in last year

CASDA

- Data and Services at Pawsey
- Uses DAP front end (Canberra CDC)
- Volume: 8.8 PB
- Growth: 3.3 PB in last year

ATNF Archives Future Status



Continuum and Spectral Line

Pulsar

CASDA - ATOA

- Data and Services at Pawsey
- Uses DAP front end (Canberra CDC)
- 5 - 6 PB/yr ASKAP surveys
- up to 0.8 PB/yr Murriyang/Parkes, ATCA, LBA
- Current ATOA data migrated by end of 2025 (expected)

Parkes Pulsar DAP

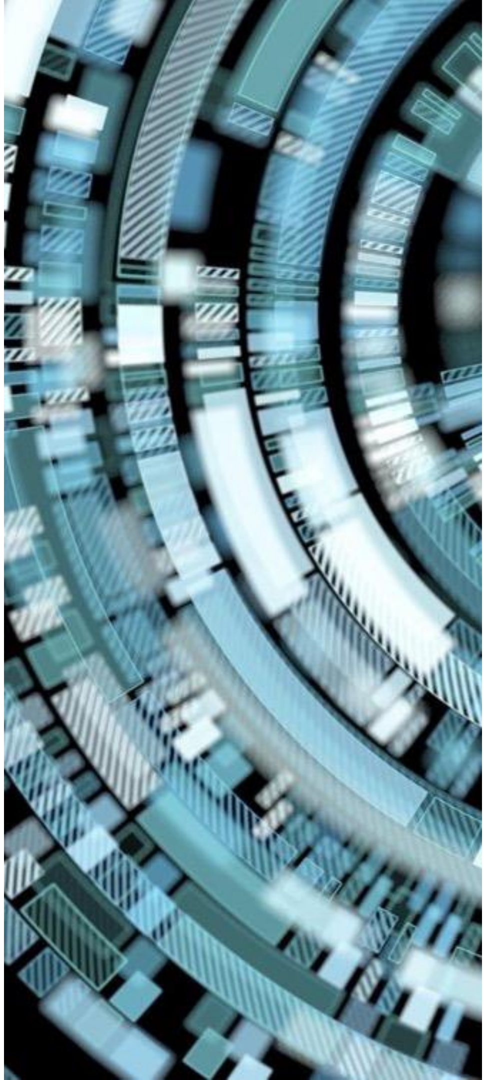
- CSIRO Data Access Portal
- Canberra, CSIRO Data Centre
- Up to 2 PB/yr with CryoPAF pulsar search mode

Users: What To Expect

- ATCA data:
  - Transition is now (so BIGCAT data from CASDA-ATOA)
  - CABB data from current ATOA will be migrated

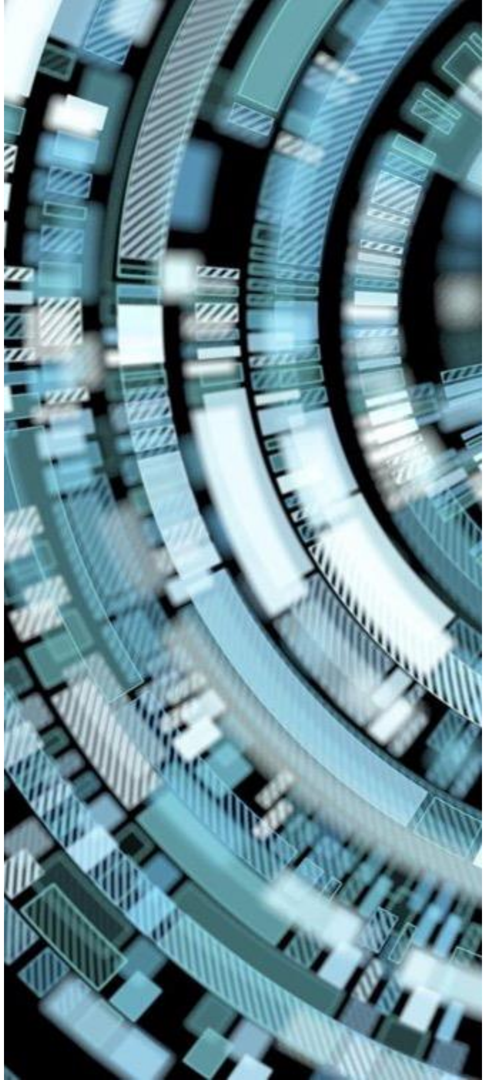
- Parkes Spectral Line / Continuum Data
  - Switch to CASDA-ATOA will happen in May (?)

- LBA (correlated) data:
  - Transition to CASDA-ATOA has happened
  - Observations from end of 2018 now in CASDA

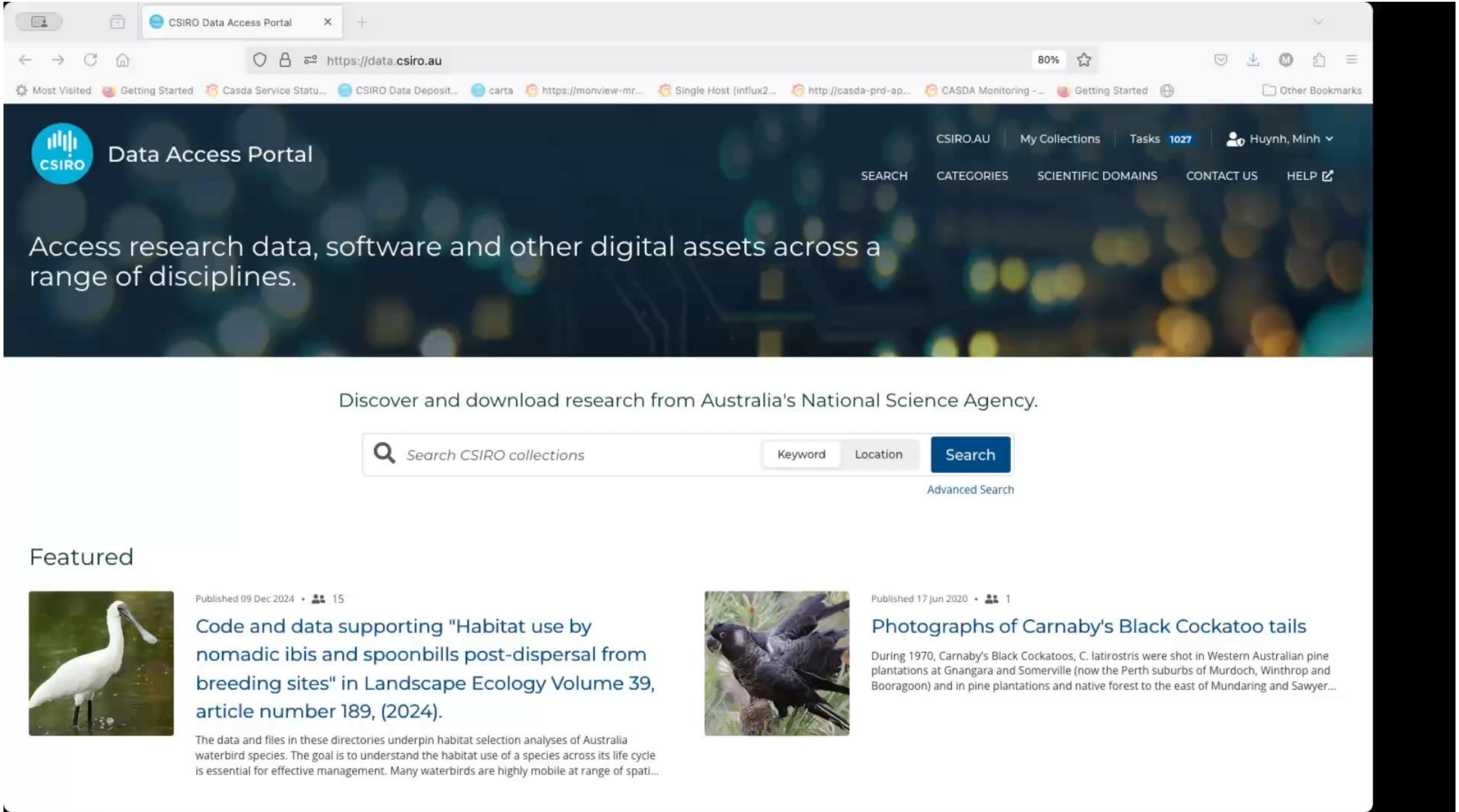


Users: What To Expect

- Implement the same data access policies e.g. proprietary period
- Team Access is modifiable by team PI
  - OPAL accounts for data access and Team admin
- Current/old ATOA to be migrated into CASDA-ATOA, and then eventually de-commissioned.
- Parkes Pulsar Data in DAP
  - no change



ATOA on DAP



CSIRO Data Access Portal
https://data.csiro.au
80%
Most Visited
Getting Started
Casda Service Statu...
CSIRO Data Deposit...
carta
https://monview-mr...
Single Host (influx2...
http://casda-prd-ap...
CASDA Monitoring -...
Getting Started
Other Bookmarks
Data Access Portal
CSIRO.AU
My Collections
Tasks 1027
Huynh, Minh
SEARCH
CATEGORIES
SCIENTIFIC DOMAINS
CONTACT US
HELP
Access research data, software and other digital assets across a range of disciplines.
Discover and download research from Australia's National Science Agency.
Search CSIRO collections
Keyword
Location
Search
Advanced Search
Featured
Published 09 Dec 2024 • 15
Code and data supporting "Habitat use by nomadic ibis and spoonbills post-dispersal from breeding sites" in Landscape Ecology Volume 39, article number 189, (2024).
The data and files in these directories underpin habitat selection analyses of Australia waterbird species. The goal is to understand the habitat use of a species across its life cycle is essential for effective management. Many waterbirds are highly mobile at range of spati...
Published 17 Jun 2020 • 1
Photographs of Carnaby's Black Cockatoo tails
During 1970, Carnaby's Black Cockatoos, C. latirostris were shot in Western Australian pine plantations at Gnangara and Somerville (now the Perth suburbs of Murdoch, Winthrop and Booragoon) and in pine plantations and native forest to the east of Mundaring and Sawyer...

CASDA Updates

- Hitting our storage allocation but expect to have an increase of 5 to 10 PBs in ~month(s).
- Will be deleting Rejected data to make room on the Acacia Storage
  - More than 2 PB of Rejected ASKAP data
- Longer term storage: working with Pawsey on NDRI bid

Acknowledging Our Data Archives

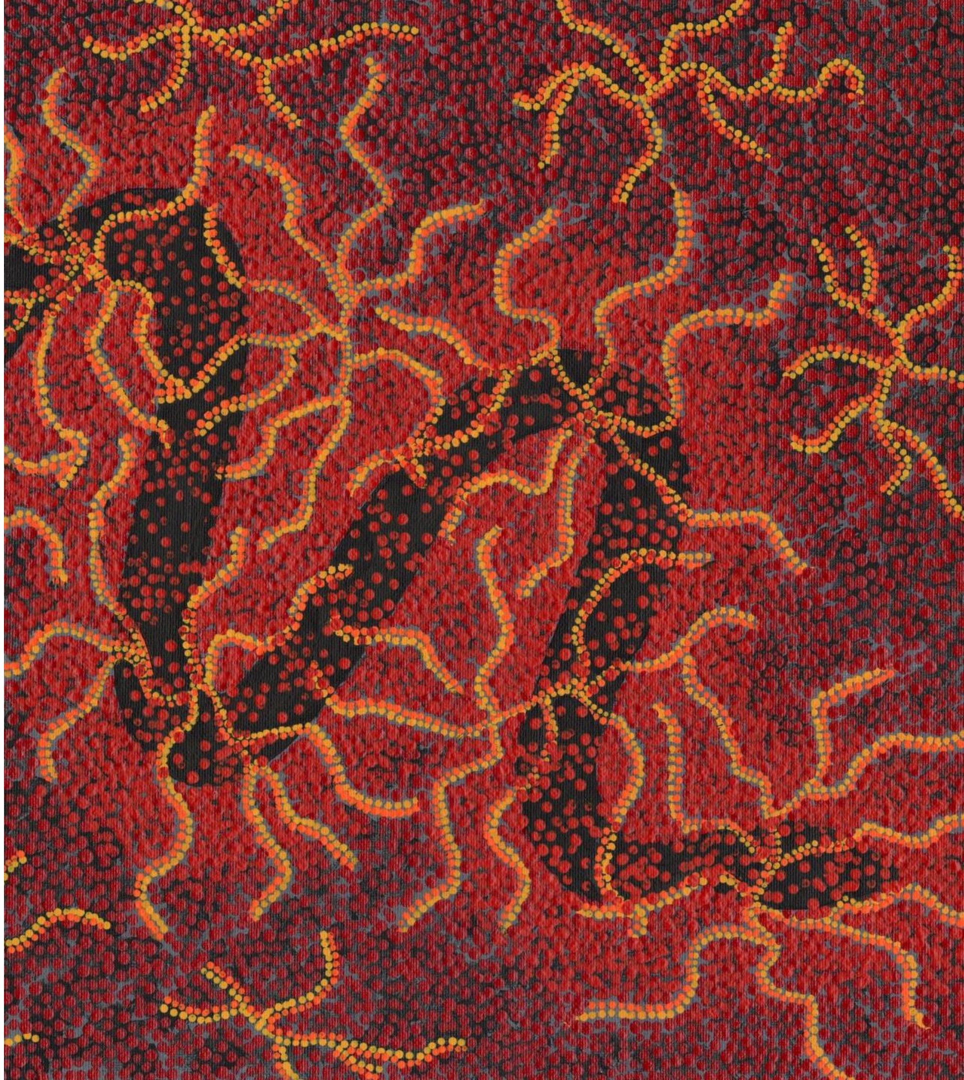
- Measuring archive impact can be challenging:
  - 32 refereed papers with "CASDA" in text in 2024
  - 42 refereed papers with "data.csiro.au" in text in 2024
  - 11 with "ATOA" in 2024
  - 118 refereed papers with "CASDA" since 2014
  - 143 refereed papers with "data.csiro.au" since 2011
  - 132 with "ATOA" since 2007

- Please follow the Acknowledgements request at https://www.atnf.csiro.au/resources/publications/atnf-publication-acknowledgement-statements/

Thank you

Space and Astronomy
Minh Huynh
Deputy Head of ATNF Science / CASDA Project Lead

+61 8 6436 8696
Minh.Huynh@csiro.au



Susan Merry, *Data,* 2019

New features in ASKAPsoft

Matthew Whiting

(3 minutes)

ASKAP processing – new & recent features

ASKAPsoft and pipelines under active development with new features guided by feedback from science teams

Testing is on-going alongside operational work – resources can be limited

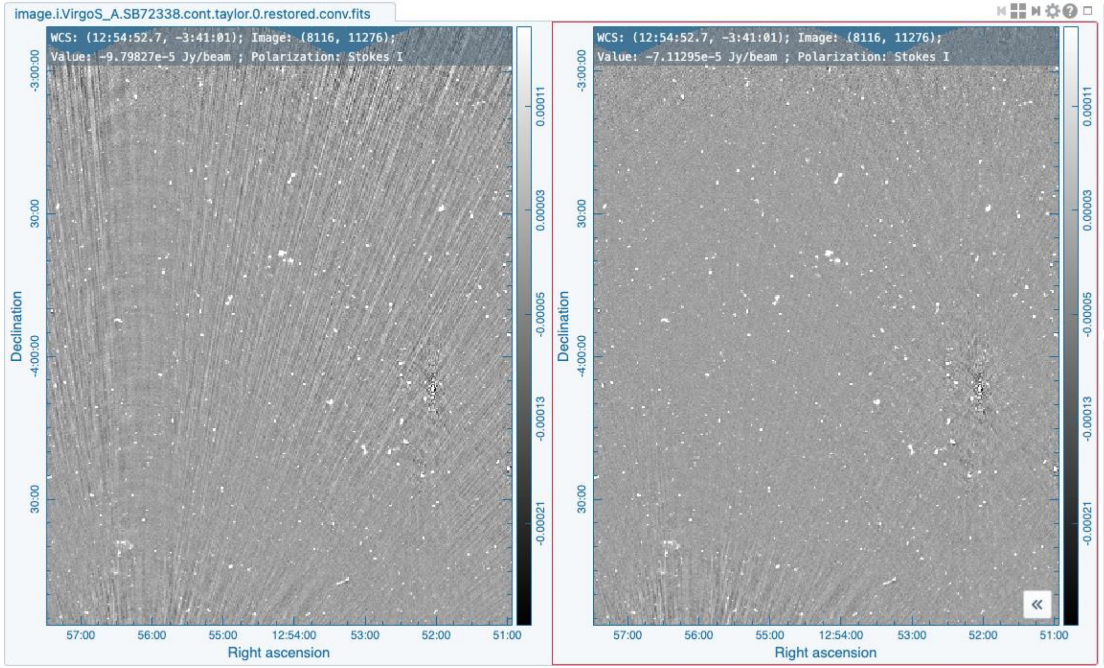
Emphasis on deliberate and considered changes to operational pipelines

Providing some datasets for science-team validation & acceptance

- Phase calibration against RACS sky model
  - Available, used by some teams regularly and looking to expand its use
  - Allows for consistent astrometry from beam to beam
- Offset-field imaging of nearby bright sources
  - Remove their artefacts from a given beam's image at imaging/selfcal stage
  - Testing nearing completion and expect to have available soon
- Various imaging improvements:
  - Traditional weighting as an option
  - Use of MFS model for starting spectral cleaning
  - More control over deconvolution (S/N threshold, divergence, pixel-lists)
  - Masking of model in selfcal and cleaning
  - Positivity of multi-scale clean components (new)

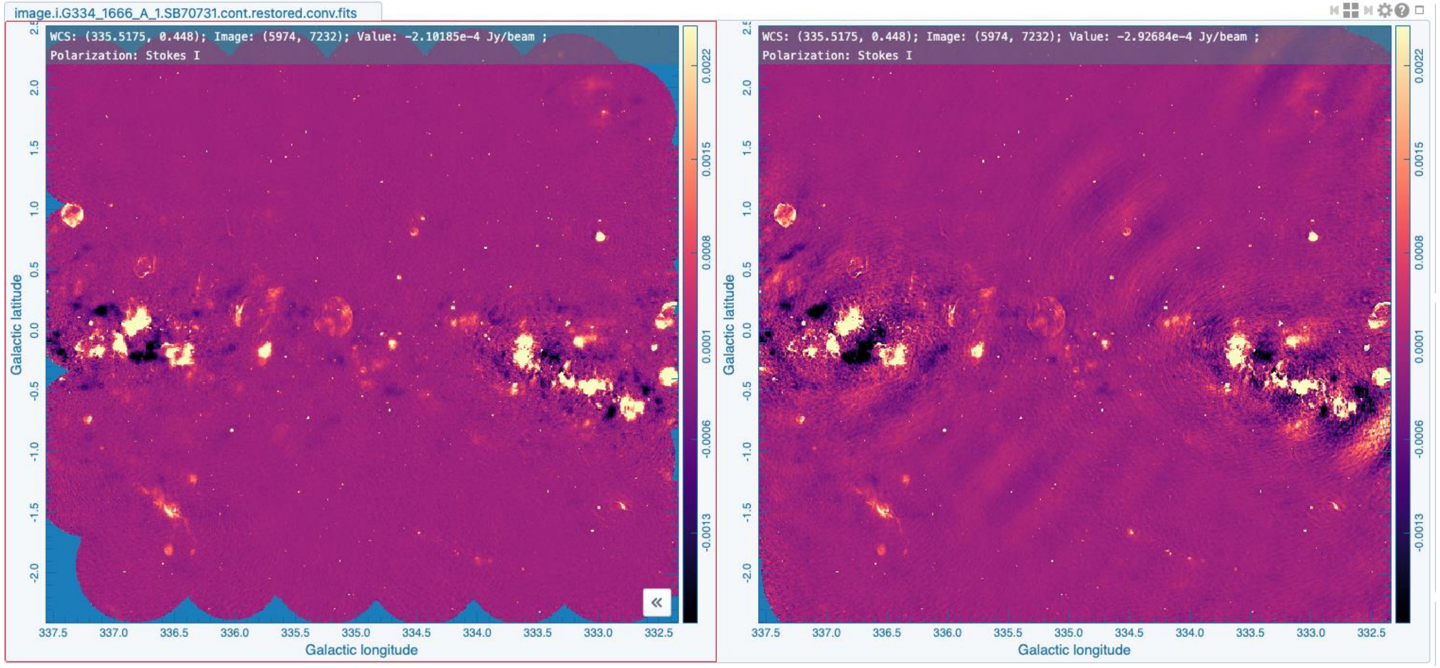
Offset-field imaging - example

8Jy source in centre of field – causes streaks in outer beams where it isn't in the clean model



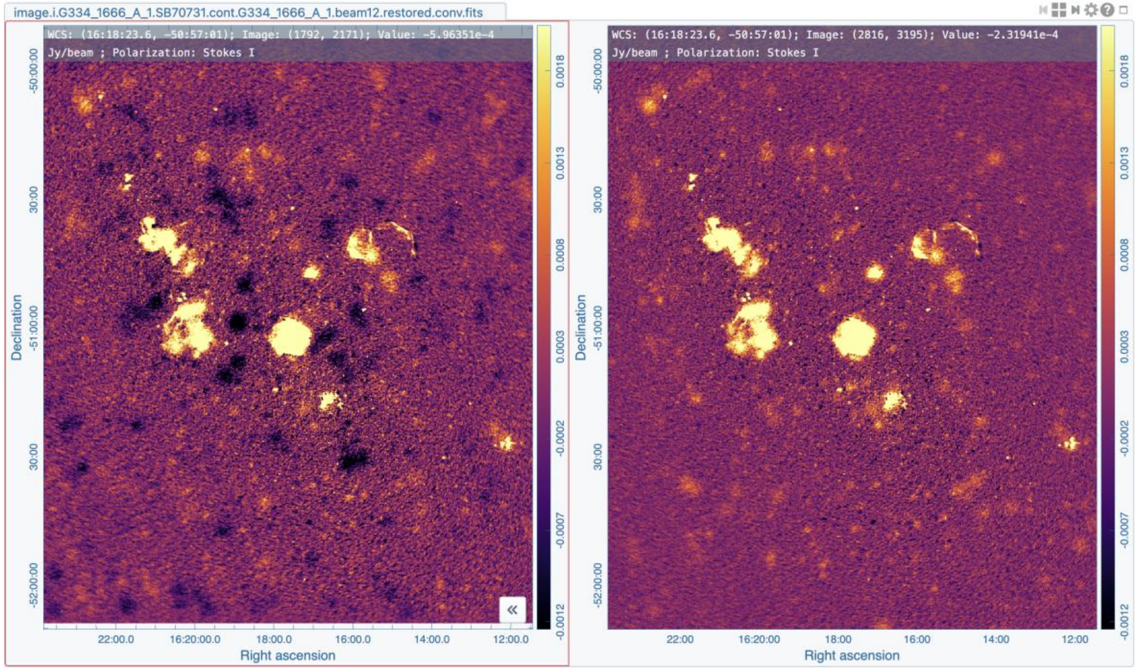
Cleaning an offset field placed on this source at the same time cleans up the artefacts

Complex field imaging improvements – GASKAP-OH



Range of improvements (left) compared to original template (right): better deconvolution strategies, avoiding divergence and cleaning to noise + using appropriate linmos cutoffs for new set of beams

Complex field imaging improvements – GASKAP-OH



Recent testing of new positivity feature in clean, to prevent prominent negative features

Flint and ASKAPsoft Comparison

Tim Galvin

(3 minutes)

Flint: A python ASKAP pipeline

Tim Galvin | ATUC 9 April 2025
Slide credits: A. Thomson, S. Duchesne

What is Flint?

Flint *is*:

- Python module(s) to process ASKAP data
- Built on Prefect & Dask frameworks
- Containerised native code
- Pirate themed

Capabilities

Join the flint crew!
Code is available on GitHub
Lots of dev tools to assist with contributions

Flint *can*:

- Process:
  - Raw visibilities, or
  - CASDA deposited MSs
- Run flows to do:
  - Bandpass calibration
  - Self-calibration & imaging
  - Continuum-subtraction
  - Polarisation imaging *
  - Timestep imaging *

(* contributed flows)

Flint *does*:

- Rely on community software
- Use Prefect and Dask
- Use WSClean
- Provide CLI entry points

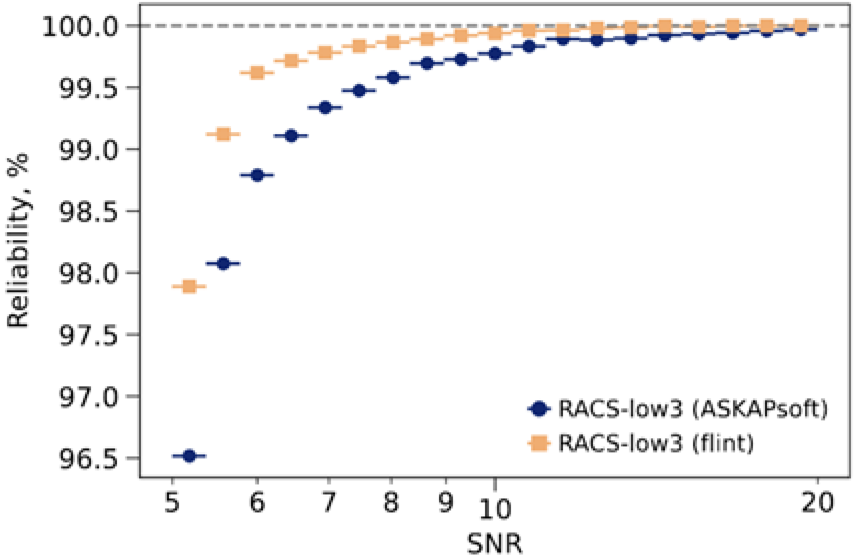
Flint *does not*:

- Map 1:1 with ASKAPsoft pipeline
- Rely on MPI for inter-node acceleration

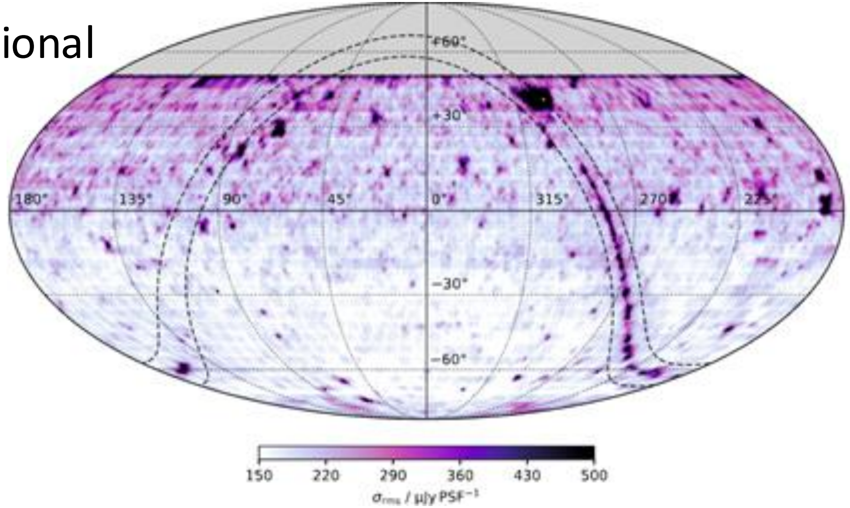
RACS-Low3 Results

Analysis by Stefan Duchesne:

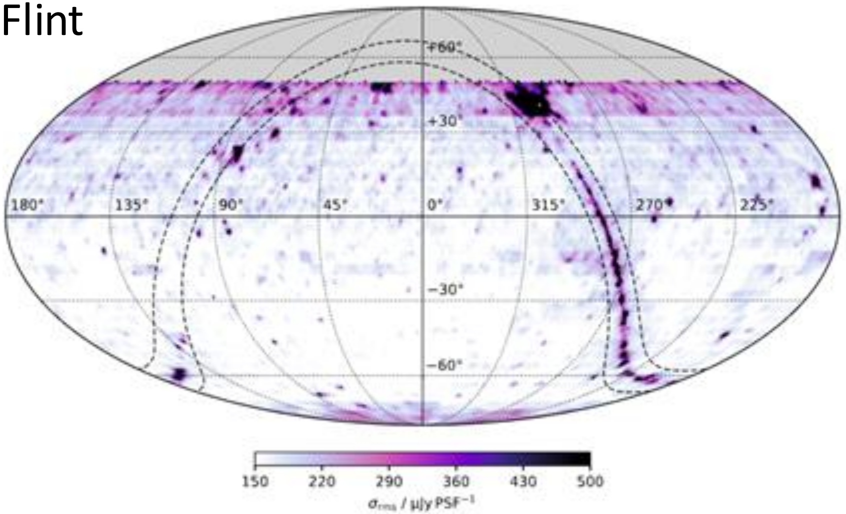
- Flux scale 1 to 1
- 3.9M to 4.4M detected islands
- 6% improvement in reliability at 5σ



Operational

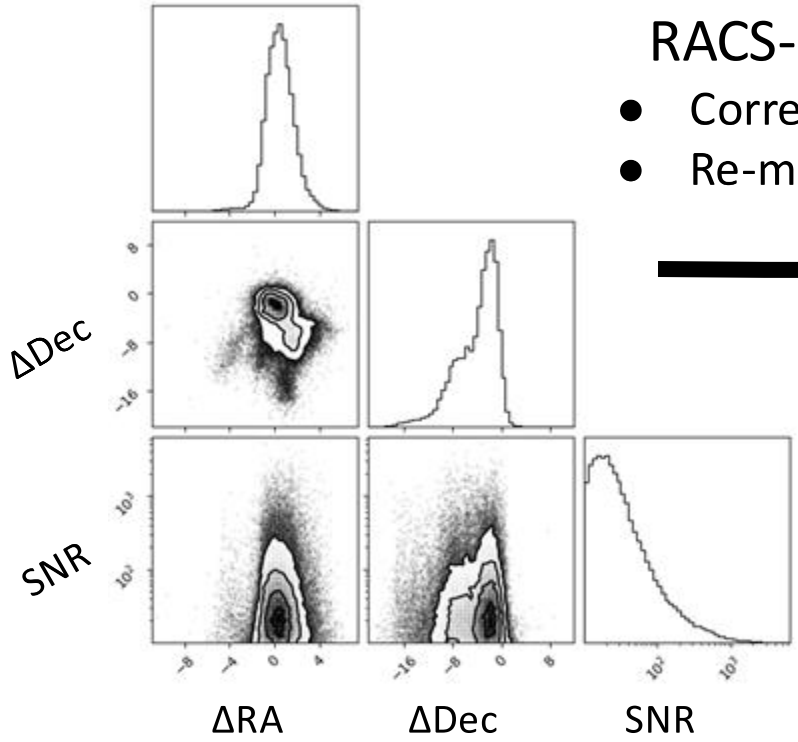


Flint



Astrometry Correction for Global Sky Model

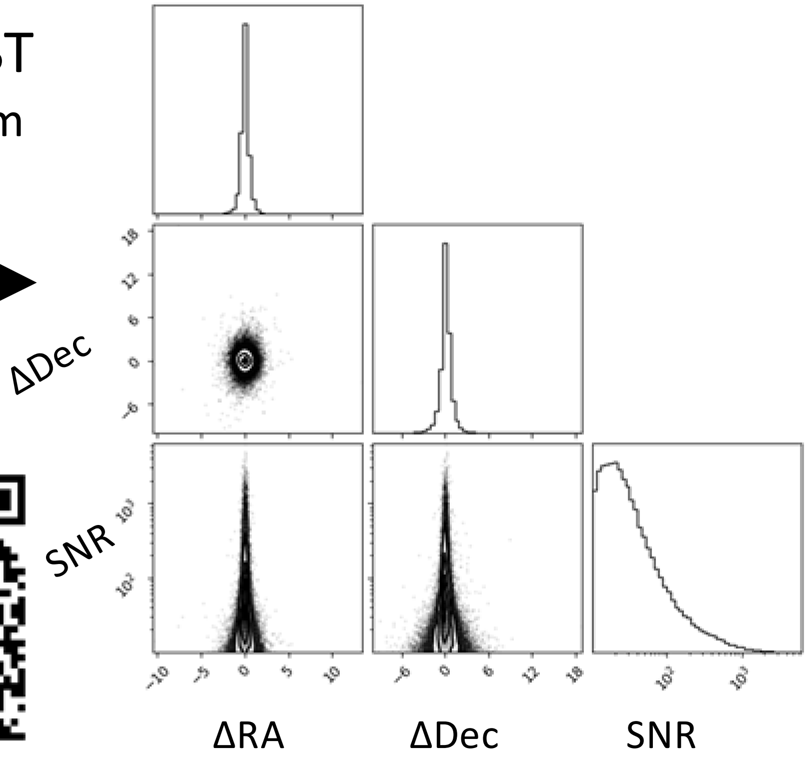
Before



RACS-low3 ╳ FIRST

- Corrected per beam
- Re-mosaicked

After



(Credit S. Duchesne)

https://github.com/flint-crew/cross-bones/

ATCA Data Reduction Pipeline Ideas

- Goal is to generate quick-look data products
  - Continuum image with reliable flux / bandpass / gain calibration applied
  - Sufficient for basic science output (e.g. measure flux of a source)
  - High fidelity calibration / imaging: probably want human eyes on the data
- ATCA is flexible: Hard to automate everything reliably!
  - A classic completeness / reliability problem
  - Incorporating observation metadata with BIGCAT / ASDM may help
  - Still leaves difficulty providing automation of archival CABB processing
- Future opportunities with BIGCAT?
  - Need to weigh up opportunities with development resources
  - We should make use of community knowledge / tools

cal-cabb: a miriad wrapper and continuum pipeline

- Implements a basic continuum workflow
- Automatic flagging and application of bandpass / flux / gain solutions
- Optional inclusion of a polarisation calibrator
- Interactive flagging if heavily affected by RFI
- Generates basic diagnostics and warnings
- **Likely many equally / more capable pipelines are hiding in this crowd!**

http://github.com/joshoewahp/cal-cabb

```
→ atca_calibration atca-cal --help
Usage: atca-cal [OPTIONS] DATA_DIR PROJECT_CODE

Options:
  -B, --band TEXT                 [default: L]
  -p, --primary-cal TEXT          [default: 1934-638]
  -s, --gain-cal TEXT
  -t, --target TEXT
  -l, --leakage-cal TEXT
  -m, --mfinterval TEXT           Time interval to solve for antenna gains in
                                  bandpass calibration.  [default: 1.0]
  -b, --bpinterval TEXT           Time interval to solve for bandpass in
                                  bandpass calibration.  [default: 1.0]
  -g, --gpinterval TEXT           Time interval to solve for antenna gains in
                                  gain calibration.  [default: 0.1]
  -f, --nfbin TEXT                Number of frequency subbands in which to
                                  solve for gains/leakage.  [default: 4]
  -n, --num-flag-rounds INTEGER   Number of rounds in each autoflagging /
                                  calibration loop.  [default: 1]
  -r, --refant [1|2|3|4|5|6]      Reference antenna.  [default: 3]
  --int-freq TEXT                 Intermediate Frequency (IF) to select (only
                                  valid for L-band)
  --shiftra TEXT                  Offset between pointing and phase centre
                                  right ascension in arcsec.  [default: 0]
  --shiftdec TEXT                 Offset between pointing and phase centre
                                  declination in arcsec.  [default: 0]
  -P, --strong-pol                Solve for absolute XY phase and leakage
                                  (requires good leakage calibrator)
  -F, --noflag                    Disable birdie and rfiflag options in atlod
                                  and avoid target flagging.
  -I, --interactive               Run calibration pipeline interactively with
                                  manual flagging.
  -o, --out-dir PATH              Path to store calibrated MeasurementSet.
                                  [default: .]
  -S, --skip-pipeline             Skip execution of flagging/calibration
                                  pipeline.
  -L, --savelogs                  Store processing logs.
  -d, --diagnostics               Generate diagnostic plots.
  -k, --keep-intermediate         Store intermediate files produced by miriad.
  -v, --verbose
  --help                          Show this message and exit.
```

Parkes Pipeline Ideas

Lawrence Toomey

(3 minutes)

Providing pulsar searching as a service for the Murriyang community

- A real-time FRB detection mode with transient buffer is currently under-development for the CryoPAF receiver
- We are also exploring the feasibility of providing an end-to-end pulsar search pipeline:
  - CSIRO users – will need to arrange a resource allocation and use the container
  - Non-CSIRO users will receive a candidate list only - raw data will be deleted after processing, but will be available in DAP (on a best effort basis)

Observing best practices: give everyone a fair go

- Please adhere to your proposal statements with regards to data rate - the telescope and data archive resources are planned based on all users’ requirements

Lawrence Toomey, Shi Dai, ATUC 2025