



ASKAP, CASDA, and the Australia Telescope National Facility Archives

Minh Huynh (CSIRO, Group Leader)
CASDA lead, on behalf of the CASDA team

Australia's National Science Agency





CSIRO Space and Astronomy – our sites and telescopes



ASKAP

Murchison
Geraldton



New Norcia

New Norcia
Perth



Parkes

Narrabri
Coonabarabran
Parkes



Australia Telescope Compact Array



Mopra



CDSCC

Sydney
Tidbinbilla

Australia Telescope National Facility (ATNF)

Australian Square Kilometre Array Pathfinder

Collecting area **36 x 12m**

Frequency range **700 MHz - 1.8 GHz**

Bandwidth **300 MHz**

System temp/ η **~ 75 K**

FOV **30 square degrees**

Baselines to **6km**

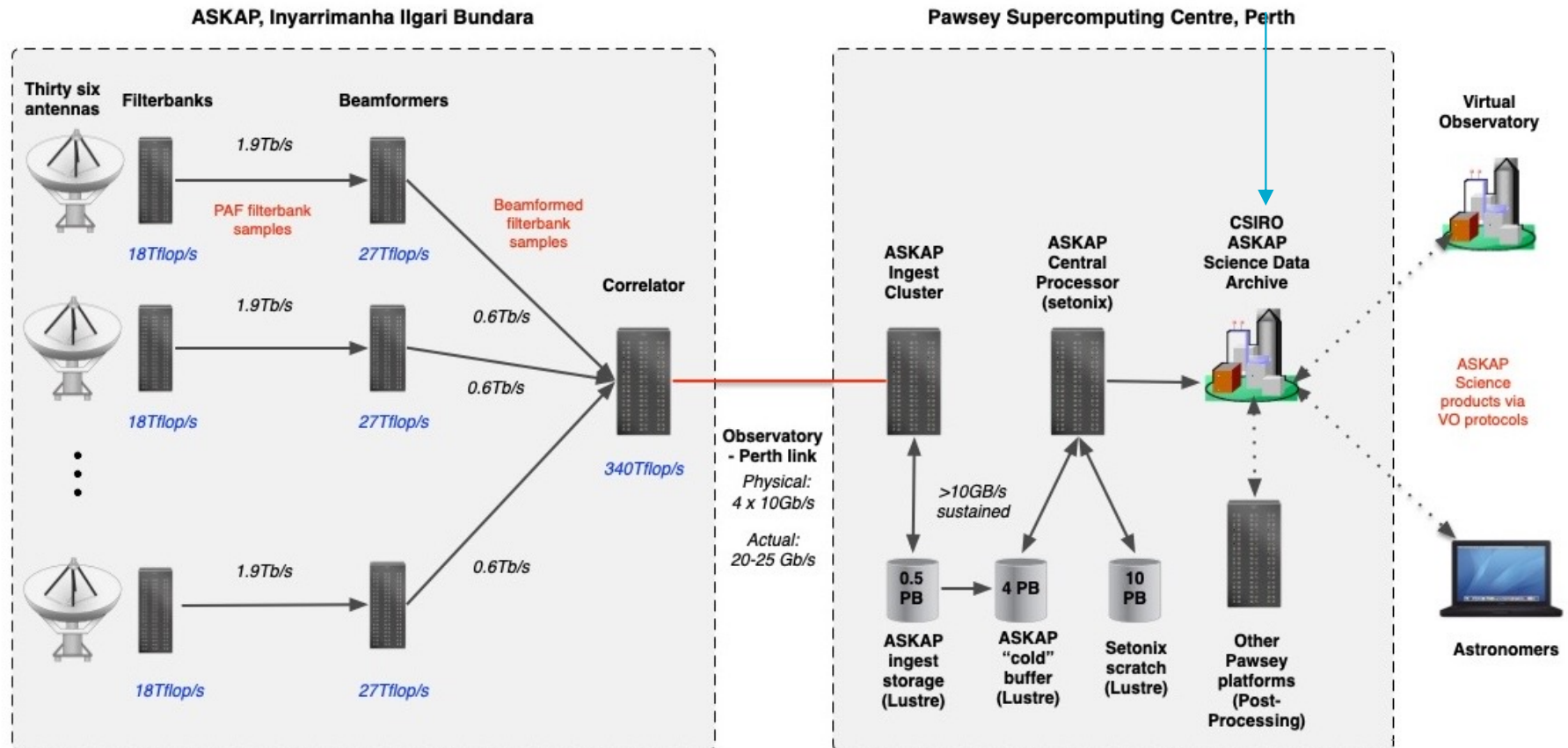
8 Survey Science Projects





ASKAP and all its wonderful data

5 - 6 PB / yr





CSIRO ASKAP Science Data Archive

- CASDA archives science-ready data products from ASKAPsoft /observatory team
- Data formats:
 - Images & Image cubes (FITS, up to ~3 TB cubes)
 - Spectra (FITS)
 - Catalogues (VOTable)
 - Visibilities (CASA Measurement Sets)
- Publish user derived valued-added data from science teams
- Archive ~5 PB per year, current volume ~5.7 PB.
- <https://casda.csiro.au>





CASDA Deployment

	PAWSEY SUPERCOMPUTING CENTRE	CSIRO CANBERRA DATA CENTRE
Location	Perth, WA	Canberra, ACT
Functions	<ul style="list-style-type: none">• Deposit ASKAP data products• Data access• Virtual Observatory Services	<ul style="list-style-type: none">• Interactive search, User Interface• Collections, DOIs• Authentication/Authorisation• Data Validation/Release
Facilities	<ul style="list-style-type: none">• Data access modules and database on NIMBUS VM• Data deposit modules on Setonix• Object storage for long term archiving (Pawsey's Acacia)• Lustre filesystem (shared: /askapbuffer)	<ul style="list-style-type: none">• virtual machines• Integrated with CSIRO Data Access Portal

CASDA on the CSIRO Data Access Portal

CSIRO Data Access Portal

CSIRO.AU Sign In

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


Featured



Published 23 Oct 2023 • 1

Annual woody vegetation and canopy cover grids for Tasmania


This collection provides annual woody vegetation (> 10% canopy cover, > 2 m height) and canopy cover (0 - 100%) grids for Tasmania with a spatial resolution of 10 m. This dataset was developed to improve the availability of information suitable for farm-scale analyses of tree cover using publicly...



Published 30 Mar 2023 • 5

Estimated spatial distribution for Australia's terrestrial threatened species


This data set contains spatial layers of the estimated historical distribution for 1518 Australian terrestrial species listed as Critically Endangered, Endangered or Vulnerable as of July 2022. Each species' historical distribution was derived from broad habitat preferences. Habitat preferences were...



Published 04 Oct 2023 • 5

Modelled deer density and impact on vegetation across the Melbourne drainage and waterway extent, Victoria

This collection contains spatial predictions of deer density and deer impact on vegetation across the Melbourne drainage and waterway extent. Deer density is quantified in units of faecal pellets/m². Deer impacts on vegetation (whole plants) are quantified as a percentage, based on foliage browsed, stem...



Published 26 Sep 2023 • 6

Aggregated Data: Australian Species Occurrences 1900-2022

Aggregated Australian species occurrence data from 1900 to the present using a suite of facets of most importance for environmental assessments. Occurrence records were aggregated and organised by the Atlas of Living Australia (ALA, <https://ala.org.au/>) and include survey and monitoring data...

<https://data.csiro.au>

CASDA Landing Page: <https://casda.csiro.au>



CARTA Visualisation Tool

- *Cube Analysis and Rendering Tool for Astronomy (CARTA)*, a visualisation & analysis tool designed for the SKA pathfinders.
- Provides usability and scalability by utilizing modern web technologies and computing parallelization.
<https://cartavis.org/>
- View and do basic analysis of images/cubes, without downloading to your local computer

The screenshot displays the 'CASDA Data Access' portal. At the top, there is a navigation bar with 'CASDA Data Access', 'CSIRO Data Access Portal', 'Contact', and 'Help' links, along with the CSIRO logo. The main content area is titled 'Visualise Files with CARTA (Beta)'. It contains a message: 'This request submitted 19/02/2024 09:05:25 UTC includes 1 file(s) totalling 815 MB. It may take some time before your data is retrieved. Once your data is retrieved you can request a CARTA session. This page will show the current status of your request.' Below this is a reusable link: https://casda.csiro.au/casda_data_access/requests/76e5ca10-cfde-44de-8217-1ba09487972c/page/1. The section 'CARTA Download Job Status' shows a green bar indicating 'Your data is ready for CARTA visualisation'. Below this, a message states: 'The files in this CARTA Download Job will be available for approximately 167 hours. In this time you will be able to start CARTA Sessions from this page in the section below to visualise these data. This job will expire at 26/02/2024 09:09:12 UTC. After this time you will no longer be able to start CARTA sessions from this page.' A table follows with columns 'Name', 'Size (kB)', and 'Status'. The table contains one row: 'image.i.EMU_1748-64.SB54926.cont.taylor.0.restored.raw.fits' with a size of 835046 kB and a status of 'Available'. The section 'CASDA CARTA Session Status' shows a green bar indicating 'Session Available'. Below this, a message states: 'There is a running CARTA Session for this job. This session will remain active for 176 minutes, expiring at 19/02/2024 12:55:09 UTC, or a few minutes after you disconnect from the session, whichever is first. The session should only be used by one person at a time. After this session has expired you may be able to start new sessions until the CARTA Download Job, detailed in the section above, has expired.' A button labeled 'Connect to CASDA CARTA Session' is located in the bottom right corner.

Name	Size (kB)	Status
image.i.EMU_1748-64.SB54926.cont.taylor.0.restored.raw.fits	835046	Available



Integrated VO services

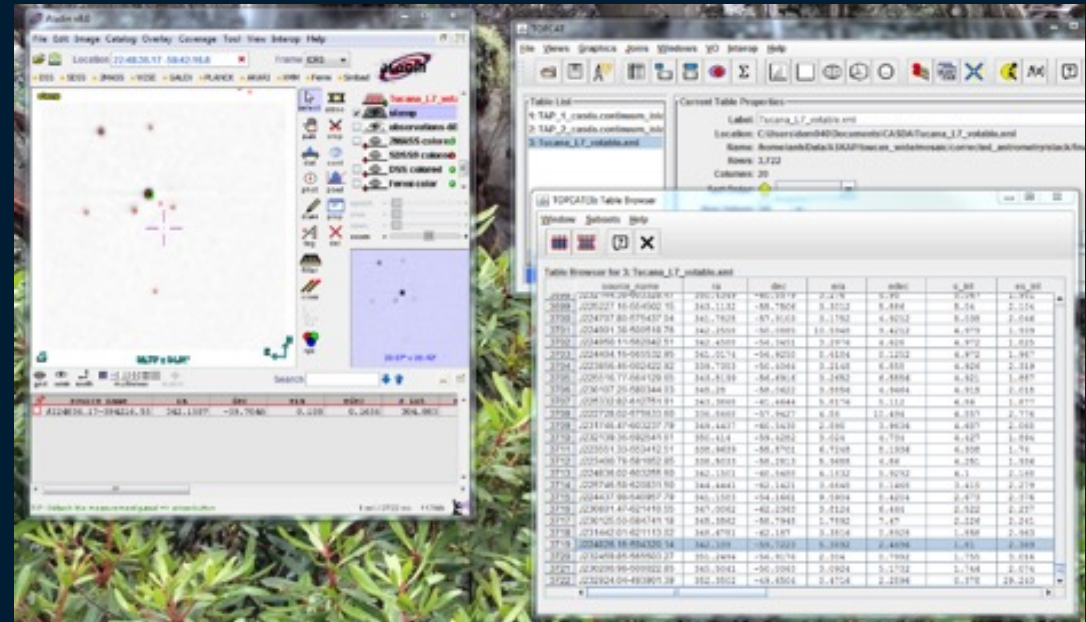
Increase FAIRness of CASDA data.

Access data directly in well known astronomy applications such as TOPCAT and Aladin.

Allows scripted (e.g. python) queries, downloads and cutout generation

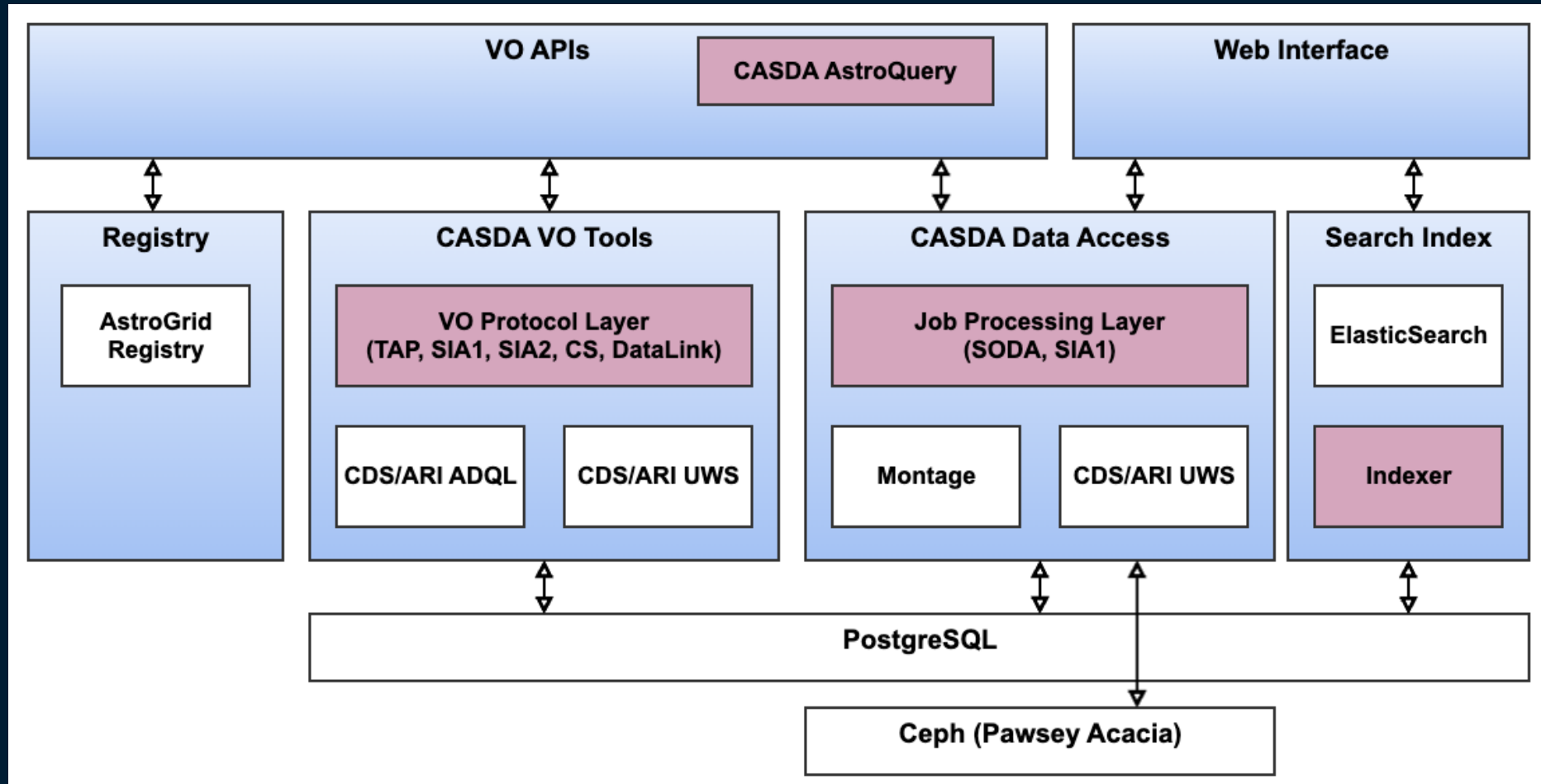
- Python's Astroquery

Backbone of web portal e.g. UI cutout service





CASDA Block Diagram





CSIRO Space and Astronomy – our sites and telescopes



ASKAP

Murchison
Geraldton



New Norcia

New Norcia
Perth



Parkes

Narrabri
Coonabarabran
Parkes



Australia Telescope Compact Array



Mopra



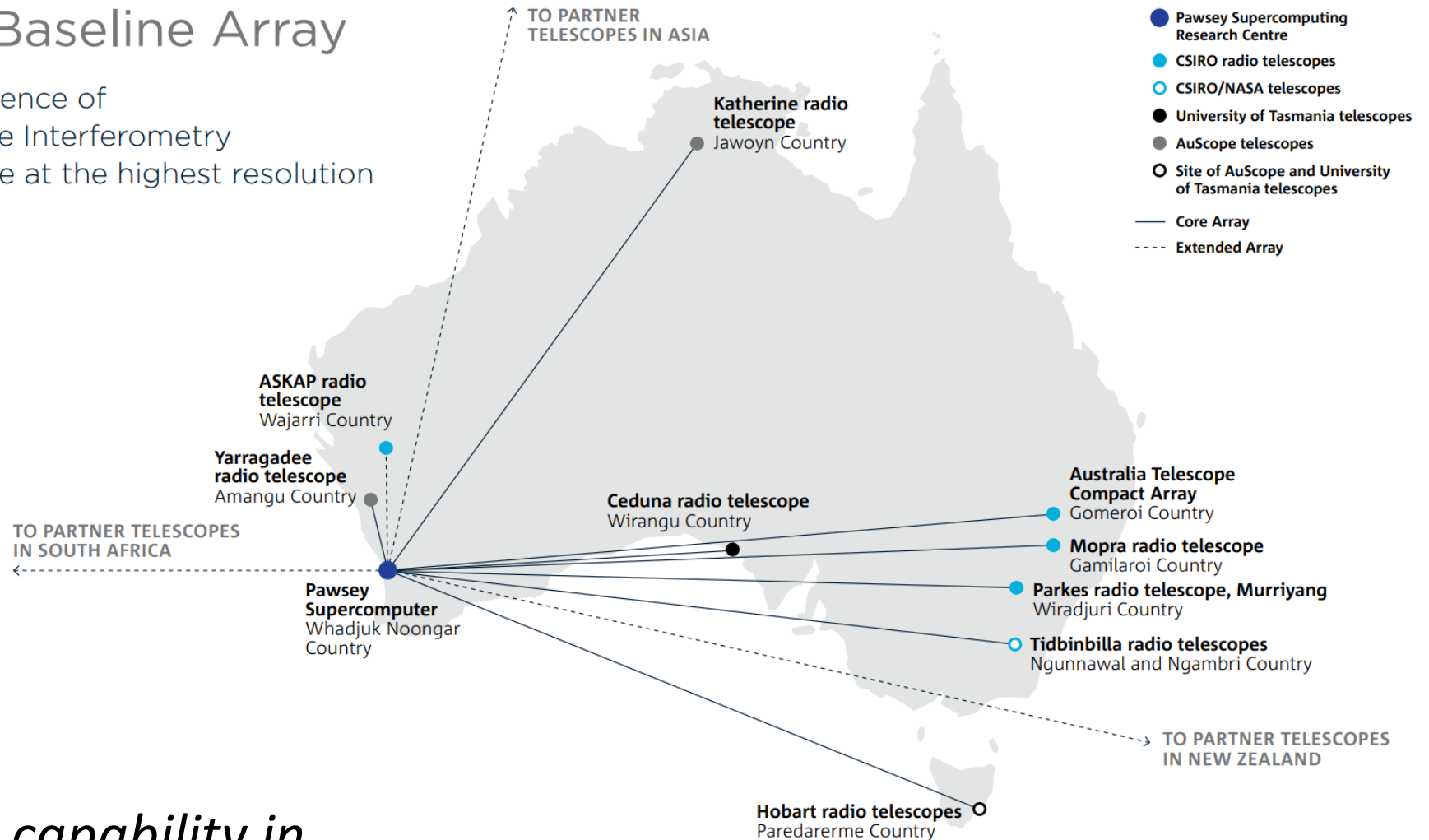
CDSCC

Sydney
Tidbinbilla

Australia Telescope National Facility (ATNF)

The Long Baseline Array

Harnessing the science of Very Long Baseline Interferometry to see our Universe at the highest resolution



Unique VLBI capability in the Southern hemisphere!



Other CSIRO Astronomy Archives

- Australia Telescope Online Archive (Marsfield)
 - Australia Telescope Compact Array, Parkes (spectral line + continuum), LBA
- Parkes Pulsar Archive (Data Access Portal, Canberra)
- New instrumentation -> larger data rates -> need solutions



ATCA



BIGCAT on ATCA

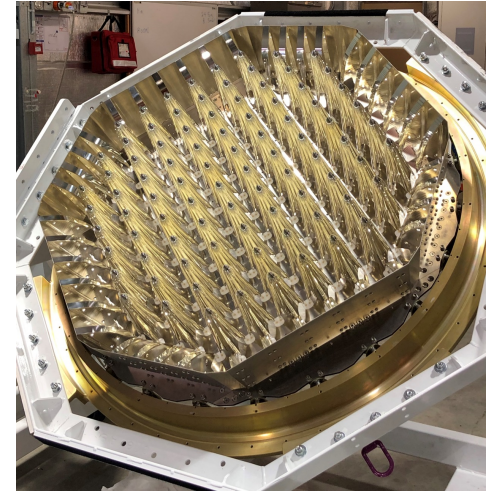
- **BIGCAT: Broadband Integrated GPU Correlator for ATCA**
 - Replacement of CABB digitisers and correlator with a hybrid FPGA+GPU backend
- Double instantaneous bandwidth
- Change of backend, no antenna or receiver changes
- New file format RPFITS -> ASDM
- Shared risk observing in OCT2024 Semester





CryoPAF on Parkes/Murriyang

- Phased Array Feed, cryogenically cooled
 - 98 dual linear polarization elements
- Maximum of 72 beams (approx. 1.5 sq deg FoV)
 - c.f. 36 beams on ASKAP
- Frequency Range: 700 to 1950 MHz
- Processed bandwidth of 600 MHz
- Science includes: pulsars, FRBs, HI surveys, OH, VLBI, SETI and more
- Timeline: On-dish commissioning Q4 2024
- File format: SDHDF for continuum/spectra, (see Toomey et al. 2024 in Astronomy and Computing), PSRFITS for pulsar





ATOA Migration to CASDA

CASDA development to manage multiple telescopes

- Database changes to accept ATOA data
- General changes to be able to accept multiple telescopes
- Data access: UI updates, ability to download data



ATOA

[ATOA Observations](#)

Query the Australia Telescope Online Archive (ATOA) for observations taken by ATCA, Parkes and LBA.



ATOA migration status

Work completed for CryoPAF,
BIGCAT and LBA readiness

Expect further enhancements
necessary to support users

https://data.csiro.au/domain/atoaObservation

https://data.csiro.au/domain/atoaObservation



Parkes Pulsar archive (Data Access Portal)

Domain of CSIRO Data Access Portal

- hosted and maintained by CSIRO IM&T at the CSIRO Canberra Data Centre.

Total ~4.3 PB of data (May 2024)

Access to fold-mode, search-mode and calibration data of single dish pulsar data from Parkes. (PSRFITS)

CSIRO Data Access Portal

SEARCH BROWSE DOMAINS CONTACT US HELP

Search by domain / ATNF Pulsar Observation Search

ATNF Pulsar Observation Search

Clear form Search

Source name / position

Source name:

Right ascension:

Declination:

Search window:

Observation

Project:

Observation date: MJD

Filename:

Frequency / band

Frequency (MHz): to

- OR -

Band name:

How to search ATNF data:

To display all results just click the "Search" button

Enter search parameters

Fill in one or many fields to search for the pulsar observation results you require.

Click the "Configure columns" button to set up your personal view of the results.

Filter search results

The search results will appear on a new page. Filter the results by clicking "Options", then the "Filter results" button to further refine your search results.

Each filter is actioned after clicking "Filter" at the bottom of the "Filter results" dialog. The number of results for each filter is calculated and displayed beside the label.

The "clear all" link at the top of the "Filter results" dialog removes all filters from the search results. The "clear" button for each group of filters removes the filters from within that group. Changes will be applied after clicking "Filter" at the bottom of the dialog.

Download results

You can download files by checking the check box next to the file and either "download selected files as TAR" or "download selected files as ZIP".

New search

Click "Options" and then "Refine search" from the search results page to change the existing search parameters.

Click on "Clear form" to clear all parameters after you have returned to the search page.

Useful links

[Australia Telescope National Facility](#)

<https://data.csiro.au/domain/atnf>



Parkes Pulsar archive (Data Access Portal)

- Recent move to S3/object storage
 - Increase in ingest and publication rates
- For projects > 10 TB:
 - weeks to get data overseas
 - days to get to a local/Aust institution
- Trialing AWS to get data to users
- GLOBUS integration for user downloads





CASDA Challenges I

- Large data files
 - Storage requirements
 - Transfer rates (at center for processing and to users)
 - Looking to implement GLOBUS at Pawsey (GLOBUS already implemented in Canberra for pulsar data)
- Cutouts
 - Using custom IPAC Montage: Multithreaded to get ~8 times speed up for cubes
 - New use cases such as 1 deg cutouts (10s GB) from full continuum spectral cube (100s GB) can take several hours to process
 - Have had users request large cutouts of full ~TB cube



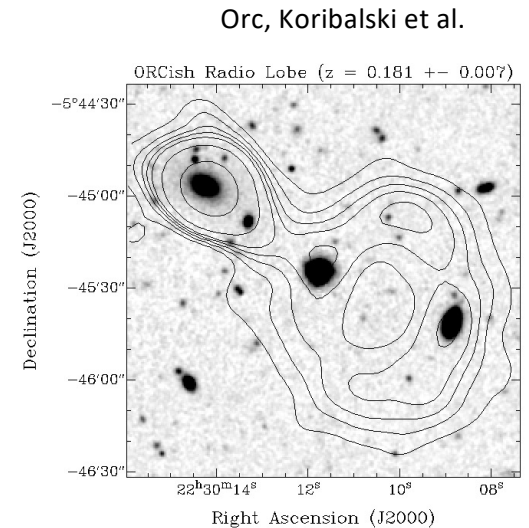
CASDA Challenges II

- Large number of data/files
 - Bulk spectra
 - Obscore now getting to millions of rows due to polarisation spectra
 - How to serve and discover millions of spectra in a useful way?
 - Is FITS the best?
 - Currently serving single 1D spectra with FITS header for metadata (allows for cone search)
 - FITS table containing ~10k spectra per observation also available as separate file
 - Global catalogue tables in TAP now also millions of rows (each RACS all sky catalogue = millions of sources) [**TAP enhancements needed?**]

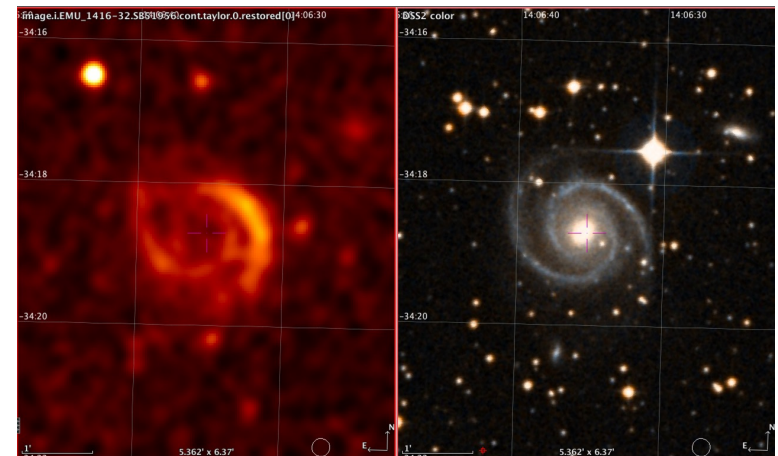


CARTA User Requests/Needs

- CARTA and radio image/cube visualisation
- Users want to crossmatch with multi-wavelength data
 - View optical/IR/radio side by side and overlay contours and catalogue positions
- Integrate HIPS into CARTA?



Example of interesting face-on Spiral in ASKAP-EMU data





CASDA-ATOA User requests/needs

- Cutouts of visibility and single dish data from ATCA and Parkes
 - Ability to download portions of a dataset (e.g. channels, beams, or scans of interest) [**Update SODA? DAL 2 // Splinter on DAP and SODA on Thu**]





Parke Pulsar Data Access Options

CryoPAF pulsar (search) data rates are a challenge (potentially ~100 TB per month)

Options:

- Push to user-defined endpoint (being done now)
 - Amazon AWS cloud S3 (ingress cheap, egress ~\$100/TB, user pays?)
- CSIRO proto-astro-platform (AWS, CSIRO-EASI platform) being explored
- Set of default pipelines at Parkes, pre-process/reduce data for users?



Summary

- CASDA serving ASKAP science data to community with many VO services
- New instrumentation coming online on other ATNF telescopes require new solutions
- Challenges and needs:
 - Around Big Data: e.g. serving bulk spectra
 - Need for multiwavelength data interoperability
 - Need to move towards science platforms not just archives





Thank you

Space and Astronomy

Minh Huynh

Senior Astronomer, Data Scientist and ATNF Group Leader

+61 8 6436 8696

Minh.Huynh@csiro.au

Australia's National Science Agency