# 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

New Norcia

Australia Telescope Compact Array

Mopra

Parkes

CDSCC

Murchison
Geraldton
New Norcia
Perth
Narrabri
Coonabarabran
Parkes
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
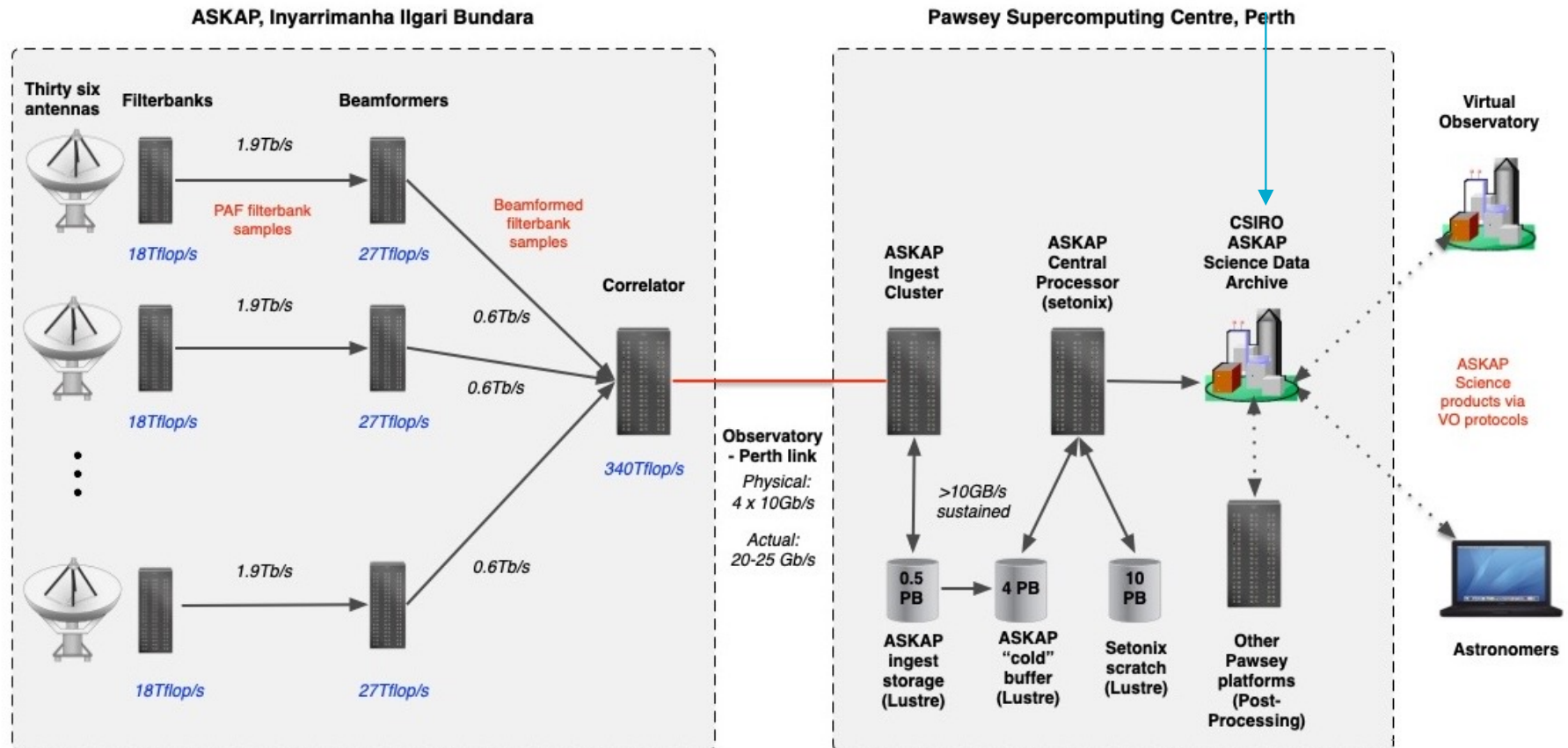
**8 Survey Science Projects**

# 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 | • Deposit ASKAP data products<br>• Data access<br>• Virtual Observatory Services | • Interactive search, User Interface<br>• Collections, DOIs<br>• Authentication/Authorisation<br>• Data Validation/Release |
| Facilities | • Data access modules and database on NIMBUS VM<br>• Data deposit modules on Setonix<br>• Object storage for long term archiving (Pawsey's Acacia)<br>• Lustre filesystem (shared: /askapbuffer) | • virtual machines<br>• Integrated with CSIRO Data Access Portal |

# CASDA on the CSIRO Data Access Portal

# 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

# 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
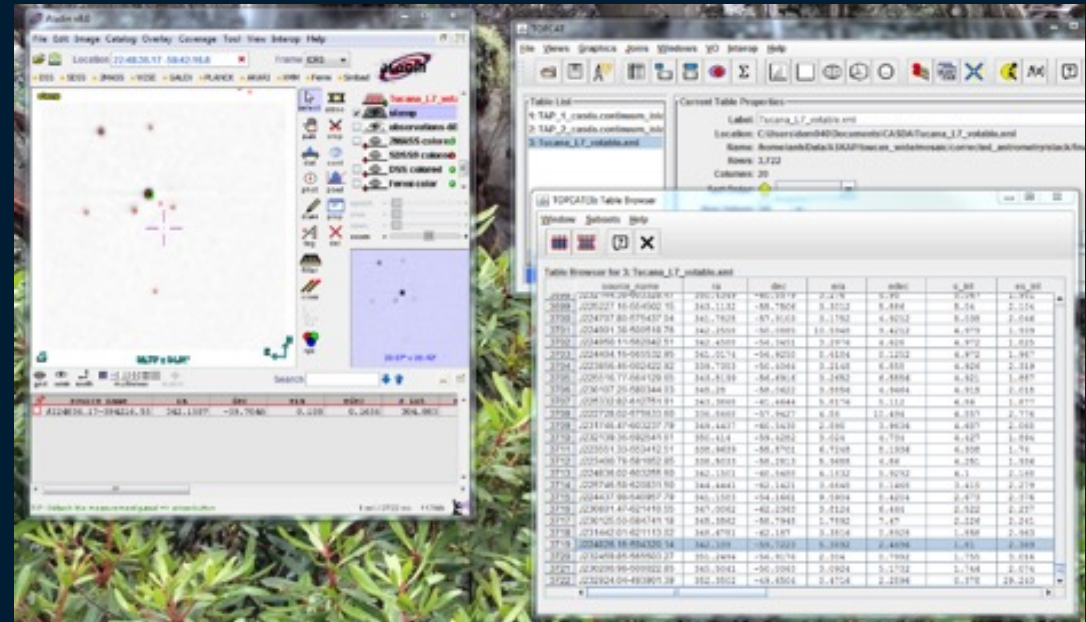
# CASDA Block Diagram
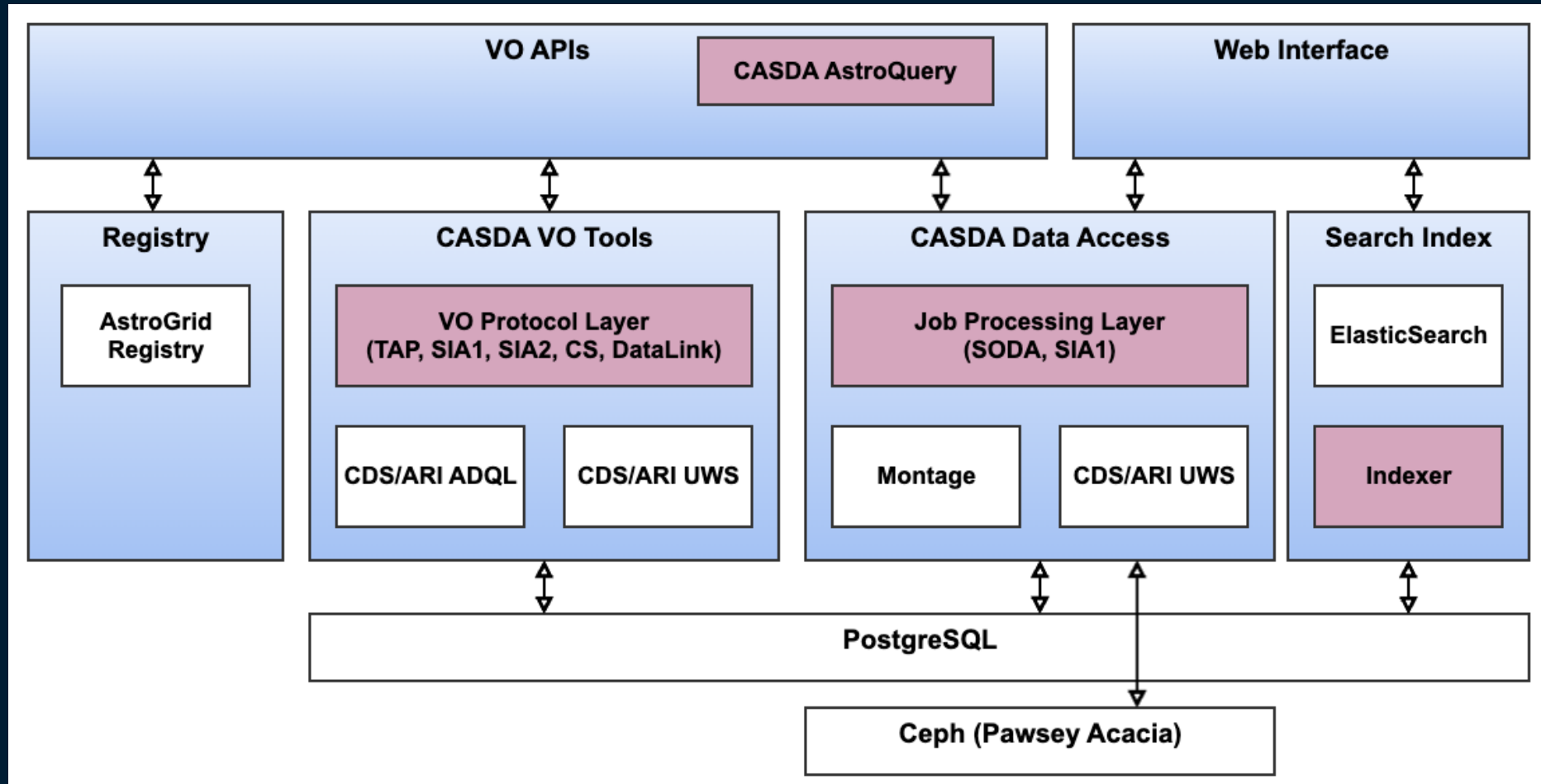
CSIRO Space and Astronomy – our sites and telescopes

# The Long Baseline Array

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

TO PARTNER
TELESCOPES IN ASIA

Katherine radio
telescope
Jawoyn Country

ASKAP radio
telescope
Wajarri Country

Yarragadee
radio telescope
Amangu Country

Ceduna radio telescope
Wirangu Country

TO PARTNER TELESCOPES
IN SOUTH AFRICA

Pawsey
Supercomputer
Whadjuk Noongar
Country

Australia Telescope
Compact Array
Gomeroi Country

Mopra radio telescope
Gamilaroi Country

Parkes radio telescope, Murriyang
Wiradjuri Country

Tidbinbilla radio telescopes
Ngunnawal and Ngambri Country

TO PARTNER TELESCOPES
IN NEW ZEALAND

Hobart radio telescopes
Paredarerme Country

**Legend:**
- Pawsey Supercomputing Research Centre
- CSIRO radio telescopes
- CSIRO/NASA telescopes
- University of Tasmania telescopes
- AuScope telescopes
- Site of AuScope and University of Tasmania telescopes
- —— Core Array
- ---- Extended Array

*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: **B**roadband **I**ntegrated **G**PU **C**orrelator for **AT**CA
  - 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



AI interpretation of BIGCAT (thanks to Bing/Copilot + Stefan) Giant cat + extra antennas not part of upgrade

# 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
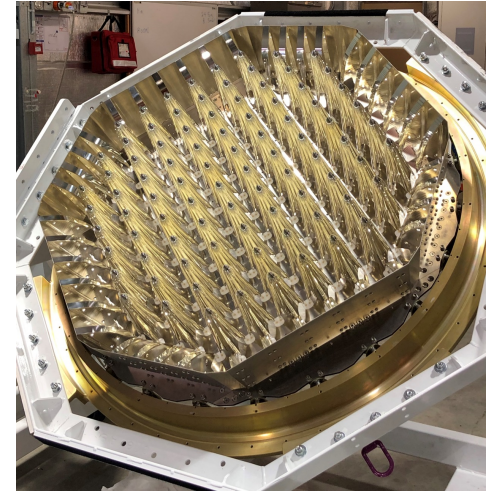
# ATOA Migration to CASDA

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

# 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)



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?

Orc, Koribalski et al.
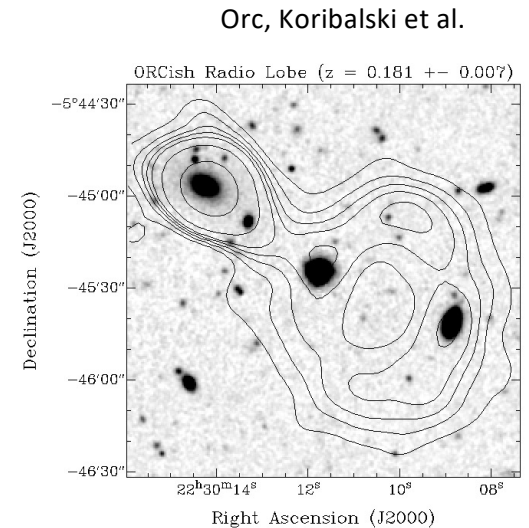
Example of interesting face-on Spiral in ASKAP-EMU data
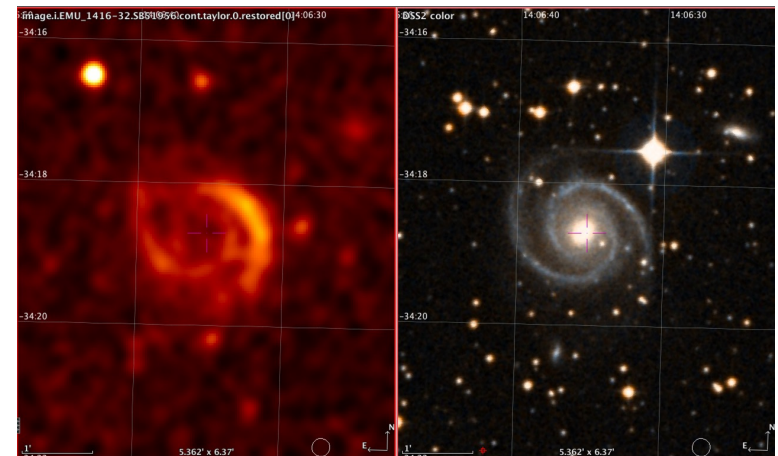
# CASDA-ATOA User requests/needs

- Cutouts of visibility and single dish 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]**

# Parkes 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