



ATNF Data Archives

For ATUC

Minh Huynh | Apr 2022

Australia's National Science Agency





CASDA and Storage Migration

- Primary data migration from CXFS/tape to Ceph/S3 disk-based object storage at Pawsey (Acacia) complete
 - More than 0.6 PB available
- Remaining data (~1 PB) to be migrated off tape when Pawsey's new tape system Banksia
 - Transfer started Fri Apr 1!
 - Expect data migration to be complete mid-May.





Recent CASDA Developments

- Work for storage migration
 - Support for data deposit and access with new object storage
- Astroquery enhancements: ability to download non-image data, cutout service awaiting pull request
- TAP and VO service improvements, inc. RACS catalogues in Simple Cone Search
- Various bug fixes



Future CASDA Work: Big Ticket Items

- Migration to new servers at Pawsey (by Sep 2022)
- New cutout UI (by Sep 2022)
- UI-UX refresh (start analysis Q3 2022)
- Installation and integration with CARTA (by Nov 2022)
- ATOA analysis and migration (start analysis Q4 2022, development in 2023)

ATUC: are these the correct priorities?



Future CASDA Work: Other Requests

- Delete function for deposits before data is released
- Improve validation workflow
- Custom DOIs
- Derived data/L7: Ability to deposit FITS tables and other file types
- UI search: show integration time and ability to filter by Stokes/pol

ATUC: are these the correct priorities?



Parke Pulsar DAP

- Major code refactoring underway
- Move to S3 storage planned
 - Accommodate UWL+CryoPAF data rates
- Question of what to archive with CryoPAF? (data rates of up to 80 TB/day)



ATOA

- BIGCAT and CryoPAF will increase data rates in 2023+
 - ATCA BIGCAT up to 0.5 PB per year
 - Parkes UWL and CryoPAF continuum/spectral data 50 to 100 TB(?) per year
 - UWL is currently ~20 TB per semester
- Plan to migrate to CASDA in 2023.



Thank you

CSIRO Astronomy and Space Science

Minh Huynh

Senior Data Scientist and Astronomer, ATNF Science Team
Leader

+61 8 6436 8696

Minh.Huynh@csiro.au

Australia's National Science Agency



ATOA Metrics

Volume of data in ATOA

- 0.48 PB

No. of queries received by ATOA*

- 430 searches per month

Requests successfully delivered*

- 2.8 TB per month

*For Q1 2022



Parke Pulsar DAP Metrics

Volume of data in
DAP

No. of queries
received by DAP*

Requests successfully
delivered*

* ?



CASDA Metrics

Volume of data in CASDA

- 1.6 PB

No. of queries received by CASDA*+

- ~80 web, ~49,000 VO per month

Requests successfully delivered*

- ~900, ~90 TB per month

* For Q4 2021

+ VO includes aggregators and HIPS sessions