Introduction to the CSIRO ASKAP Science Data Archive

Minh Huynh, for the CASDA team | Sep 2020

Australia's National Science Agency

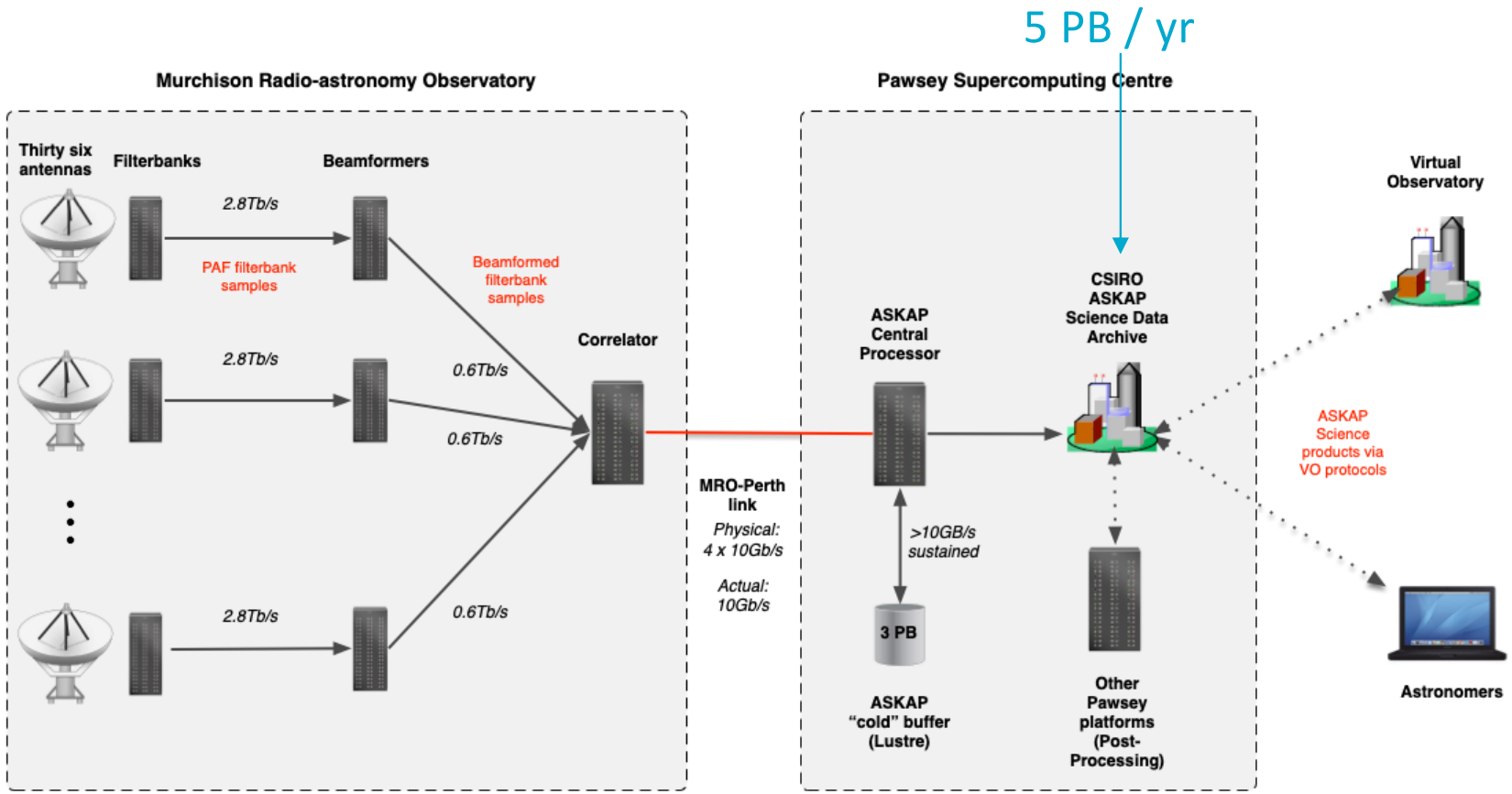
Credit: ANU & CSIRO

CSIRO ASKAP Science Data Archive

- Archive for Australian Square Kilometre Array Pathfinder (ASKAP)
- Science-ready data products
- Data formats:
  - Images & Image cubes (FITS)
  - Spectra (FITS)
  - Catalogues (VOTable)
  - Visibilities (CASA Measurement Sets)

Credit: CSIRO

ASKAP and all its wonderful data



ASKAP and CASDA Glossary

SST = Survey Science Team. There are 8 survey science teams. See resources on next slide.

SBID = Scheduling Block ID. Observations are performed and processed in scheduling blocks.

Project ID = OPAL and CASDA project codes. ASKAP Project codes are in the form “AS---”.

DAP = Data Access Portal. This is CSIRO’s enterprise wide portal for data.

VO = Virtual Observatory.

Observational Data = Science data processed by the observatory with ASKAPsoft. Can be released or unreleased.

Unreleased Data = Deposited and available to project team members.

Released Data = Deposited and validated data. Publicly available to all.

Derived Data = Value-added data uploaded by users.

Resources

- CASDA webpage https://casda.csiro.au
- ATNF ASKAP webpage https://www.atnf.csiro.au/projects/askap/index.html
- ASKAP newsletter https://www.atnf.csiro.au/projects/askap/commissioning_update.html
- ASKAP Survey Science confluence space https://confluence.csiro.au/display/askapsst/

CASDA Aims

Data Discovery

Find data useful to you

Minimal radio astronomy knowledge assumed

Query data products

Query source catalogues

Data Access

Download entire data products

Extract small parts of very large cubes (~800GB)

Data Publication

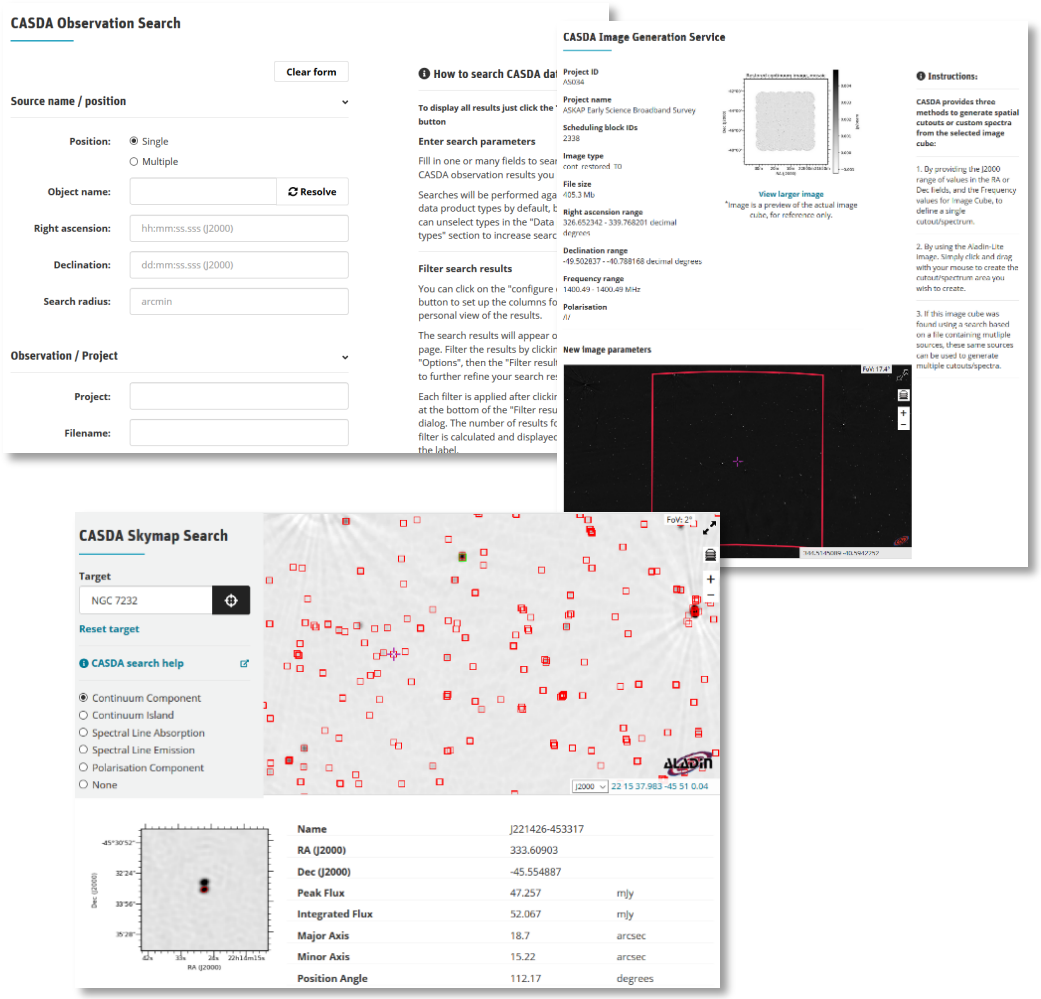
Science team ‘value-added’ data

Legacy surveys

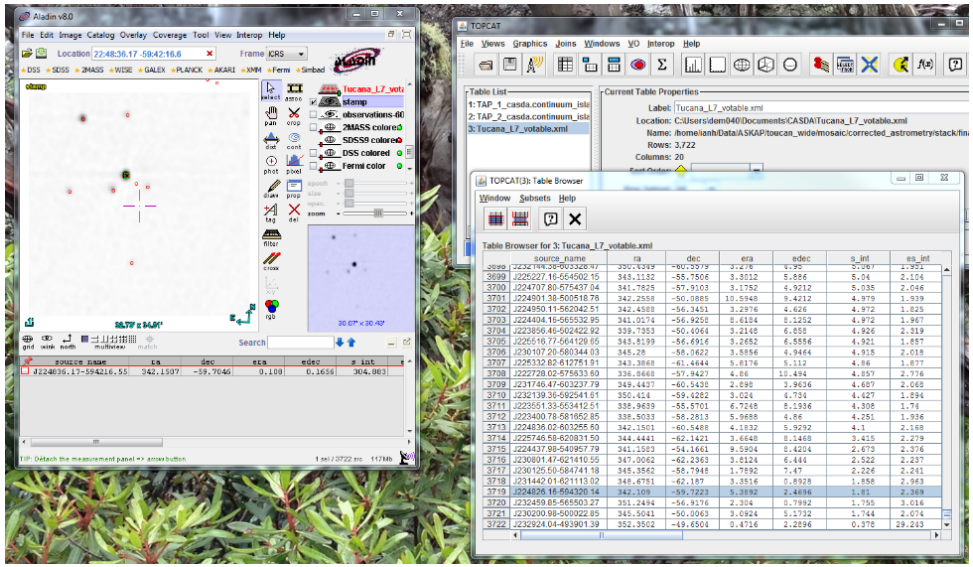
DOIs assigned

ASKAP Data on DAP

- ASKAP data is served through CSIRO's Data Access Portal
- These web tools also interact with the VO services
  - Downloads
  - Cutouts
  - Spectra generation
  - Catalogue queries



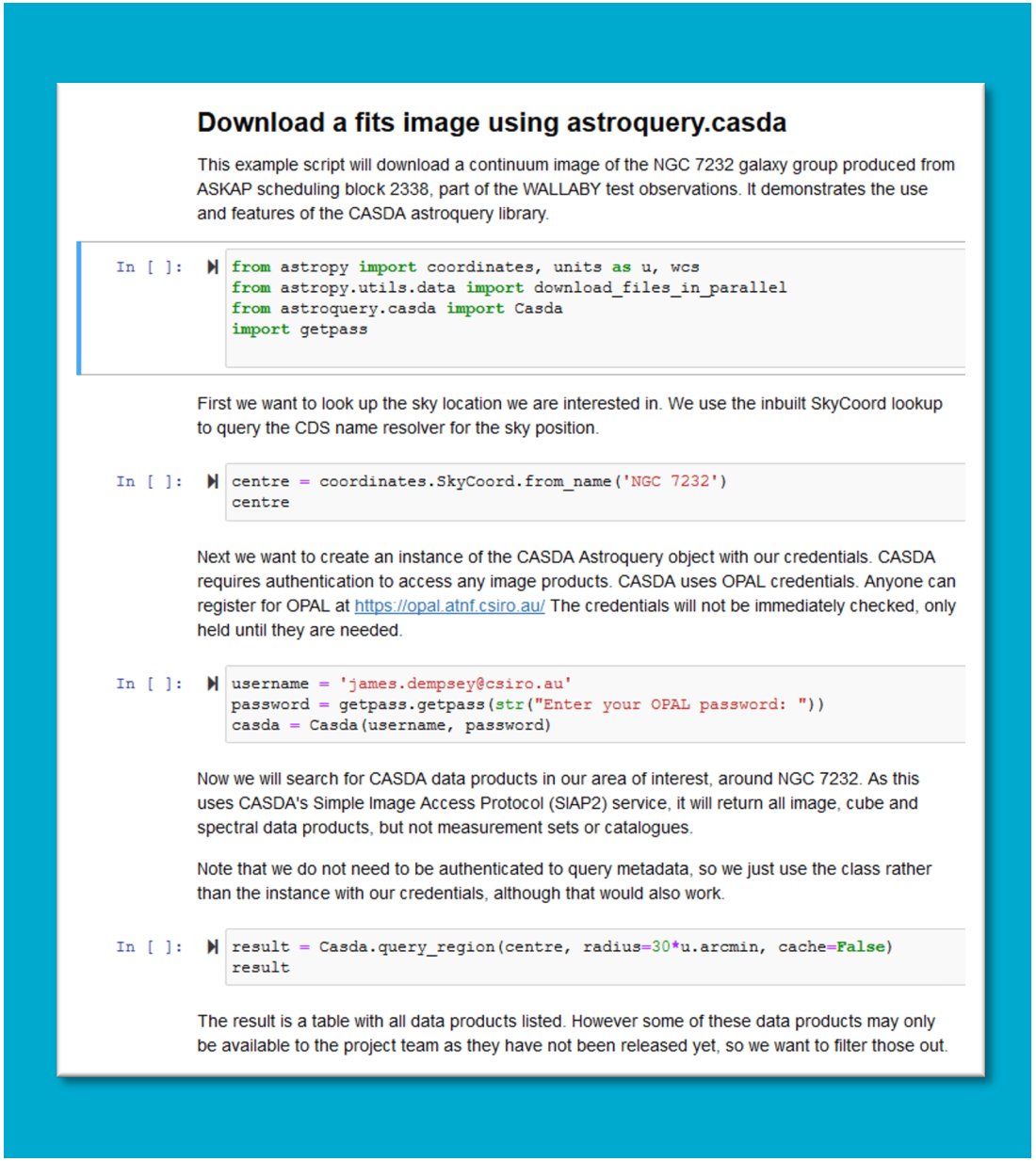
VO Applications



- Access data directly in
  - TOPCAT
  - Aladin
  - SPLAT-VO
- Example python scripts at:
  - https://github.com/csiro-rds/casda-samples/

CASDA Astroquery Module

- Python module
- Discover and download data
- Uses following VO protocols:
  - SIA2, DataLink, SODA
- https://astroquery.readthedocs.io/en/latest/casda/casda.html



Download a fits image using astroquery.casda

This example script will download a continuum image of the NGC 7232 galaxy group produced from ASKAP scheduling block 2338, part of the WALLABY test observations. It demonstrates the use and features of the CASDA astroquery library.

In [ ]:
```
from astropy import coordinates, units as u, wcs
from astropy.utils.data import download_files_in_parallel
from astroquery.casda import Casda
import getpass
```

First we want to look up the sky location we are interested in. We use the inbuilt SkyCoord lookup to query the CDS name resolver for the sky position.

In [ ]:
```
centre = coordinates.SkyCoord.from_name('NGC 7232')
centre
```

Next we want to create an instance of the CASDA Astroquery object with our credentials. CASDA requires authentication to access any image products. CASDA uses OPAL credentials. Anyone can register for OPAL at https://opal.atnf.csiro.au/ The credentials will not be immediately checked, only held until they are needed.

In [ ]:
```
username = 'james.dempsey@csiro.au'
password = getpass.getpass(str("Enter your OPAL password: "))
casda = Casda(username, password)
```

Now we will search for CASDA data products in our area of interest, around NGC 7232. As this uses CASDA's Simple Image Access Protocol (SIAP2) service, it will return all image, cube and spectral data products, but not measurement sets or catalogues.

Note that we do not need to be authenticated to query metadata, so we just use the class rather than the instance with our credentials, although that would also work.

In [ ]:
```
result = Casda.query_region(centre, radius=30*u.arcmin, cache=False)
result
```

The result is a table with all data products listed. However some of these data products may only be available to the project team as they have not been released yet, so we want to filter those out.

Data in CASDA

ASKAP BETA datasets	LEGACY HI datasets	AS034 ASKAP Early Science Continuum (EMU)	AS035 ASKAP Early Science Spectral Line (WALLABY)	Pilot Surveys …
• PKS2252-089 HI absorption (Allison et al.) • Tucana continuum image (Heywood et al.)	• HIPASS (Parkes, Koribalski et al.) • SGPS (ATCA + Parkes, McClure-Griffiths et al.)	• NGC 7232 and surrounding area • 10 EMU cosmology fields • GAMA G23 (36 dishes) • Derived data: combined mosaic GAMA G23	• NGC 7232 HI Spectral Line Cubes (2 SBIDs) • Derived data: - T. Reynolds et al., K. Lee-Waddell et al., For et al., Kleiner et al.	• See next slide

Pilot Survey Status (Phase 1, early Sep 2020)

AS101 EMU: all 10 SBIDs deposited and released

AS102 WALLABY: Hydra Cluster data released, other fields pending processing

AS103 POSSUM: 10 SBIDs deposited, pending validation and release

AS104 DINGO: pending processing

AS107 VAST: all SBIDs deposited, pending validation and release

AS108 GASKAP: 8 SBIDs deposited, pending validation and release

AS109 FLASH: pending processing

AS111 Gravitational Wave Followup: 7 SBIDs deposited, 5 released

Publishing and Acknowledgements

- CASDA (released) data is available to all astronomers.
- Please follow guidelines at: https://www.atnf.csiro.au/research/publications/Acknowledgements.html
- Publications using data accessed through CASDA should include: "*This paper includes archived data obtained through the CSIRO ASKAP Science Data Archive, CASDA (http://data.csiro.au).*"
- Individual SSTs have their own data and publication policies.
- Also please include DOIs to the dataset you use in your papers.

Thank you

CSIRO Astronomy and Space Science
Minh Huynh
Senior Data Scientist, ATNF Science Team Leader

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