ATNF Data Archives Update

For ATUC

Minh Huynh | Nov 2020

ATNF Data Archives

Parkes Pulsar DAP

Parkes pulsar data:

- fold mode
- search mode

ATOA

ATCA raw data

Mopra raw data

Parkes spectral line data

VLBI correlated data

Some processed survey data (Mopra)

Parkes UWL spectra

CASDA

Archive for ASKAP, science-ready processed data

Science team value-added products

Legacy data products

ATOA Updates

- Scope expanded last year to include UWL spectral line data
- Now supports Single Dish Hierarchical Data Format (SDHDF) from Parkes UWL

- Large data rates expected in future (UWL, BIGCAT, + CryoPAF):
  - Future options for ATOA have been explored
  - Preferred option is into CASDA/DAP
  - Need IM&T long term support, in discussions

ATOA Metrics

Volume of data in ATOA	No. of queries received by ATOA*	Requests successfully delivered*
• ~350 TB	• ~600 per month	• ~2.4 TB per month

* For 6 months, Apr to Aug 2020, when most recent statistics are available

Parkes Pulsar DAP Updates

- Improvements to data verification and ingest pipelines
- Improved access for HPC users
  - Scripted WebDAV access

- Backlog due to huge UWL data rate
  - Improved ingest and publication rates, but need to increase even more
  - Data are available 2 hours after observation from CASS servers
  - However there is currently ~3 month wait to access data through DAP
  - Optimisation of chosen observing mode could help make data rates more manageable

Parkes Pulsar DAP Metrics

Volume of data in DAP	No. of queries received by DAP*	Requests successfully delivered*
• ~2.8 PB total • ~1.5 PB of public data	• ~800 per month	• ~200 per month

* For Q3 2020, Jul to Sep 2020. Lower limits as REST services not included.

CASDA Update (9th July Prod Release)

- Extended astroquery python module
- Primary images/cubes highlighted in UI search (ancillary data in separate tab)
- Validation reports by SSTs hosted by CASDA, clickable html link
- Speed up deposit for many small files
  - RACS and VAST deposits ~5 times faster
- Integrated catalogue and image/cube upload for derived data (Level 7)
- Channel selection added for SODA (allows scripted cutouts by channel)

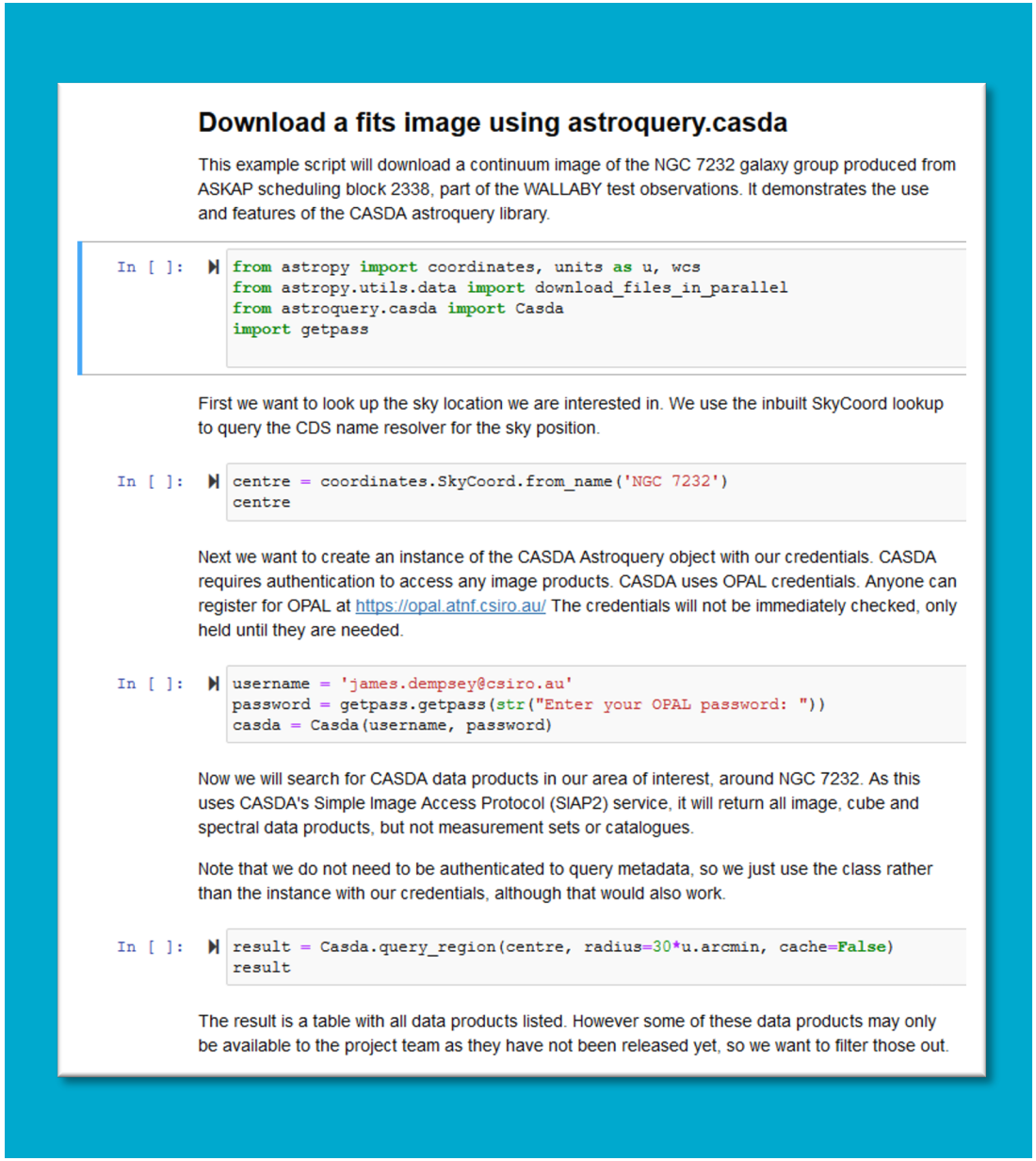
CASDA Update (12th Nov Prod Release)

- Validation and release moved to new DAP UI
- Script to do bulk validation and release
- Lightweight URL endpoint for CASDA events (using VOEvent)
- Support for new data-types, e.g. background/median image/cube

- Next development period Q2 2021
  - User feedback/requests welcome!

CASDA Astroquery Module

- Python module
- Discover and download data
- https://astroquery.readthedocs.io/en/latest/casda/casda.html
- Example notebooks available from https://research.csiro.au/casda/services/



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.

CASDA VOEvent URL Endpoint

- Another way to discover data in CASDA
- VOEvent .xml produced for deposits, validations, releases and re-deposits (updates)
- https://casda.csiro.au/casda_data_access/observations/events
- Can filter by time, project, SBID, and event type
  - Documentation: https://research.csiro.au/casda/services/

Summary of Data in CASDA

- Legacy Datasets (HIPASS, SGPS)
- ASKAP Beta and Early Science
- ASKAP Pilot Surveys (Phase 1)
  - Varying processing/validation/release states, see next slide
- Rapid ASKAP Continuum Survey (RACS)
  - First all-sky continuum survey with ASKAP at 888 MHz
  - 356 fields deposited, remaining 547 to come
- Survey With ASKAP of GAMA-09 + X-Ray (SWAG-X)
  - Observations of GAMA-09, in collaboration with eROSITA
  - Initial 888 MHz continuum observations released
    - Spectral cubes and 1.3 GHz data to come

ASKAP Pilot Survey Phase 1 Status

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: 2 SBIDs deposited, pending validation and release

AS107 VAST: all SBIDs deposited, pending validation and release

AS108 GASKAP: 8 SBIDs deposited, 2 released

AS109 FLASH: pending processing

AS111 Gravitational Wave Followup: 7 SBIDs deposited, 5 released

CASDA Metrics

Volume of data in CASDA	No. of queries received by CASDA*	Requests successfully delivered*
• 340 TB	• ~200 web, ~3,800 VO per month	• ~470, ~19 TB per month

* For Q3 2020, Jul to Sep 2020

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