# CSIRO ASKAP Science Data Archive

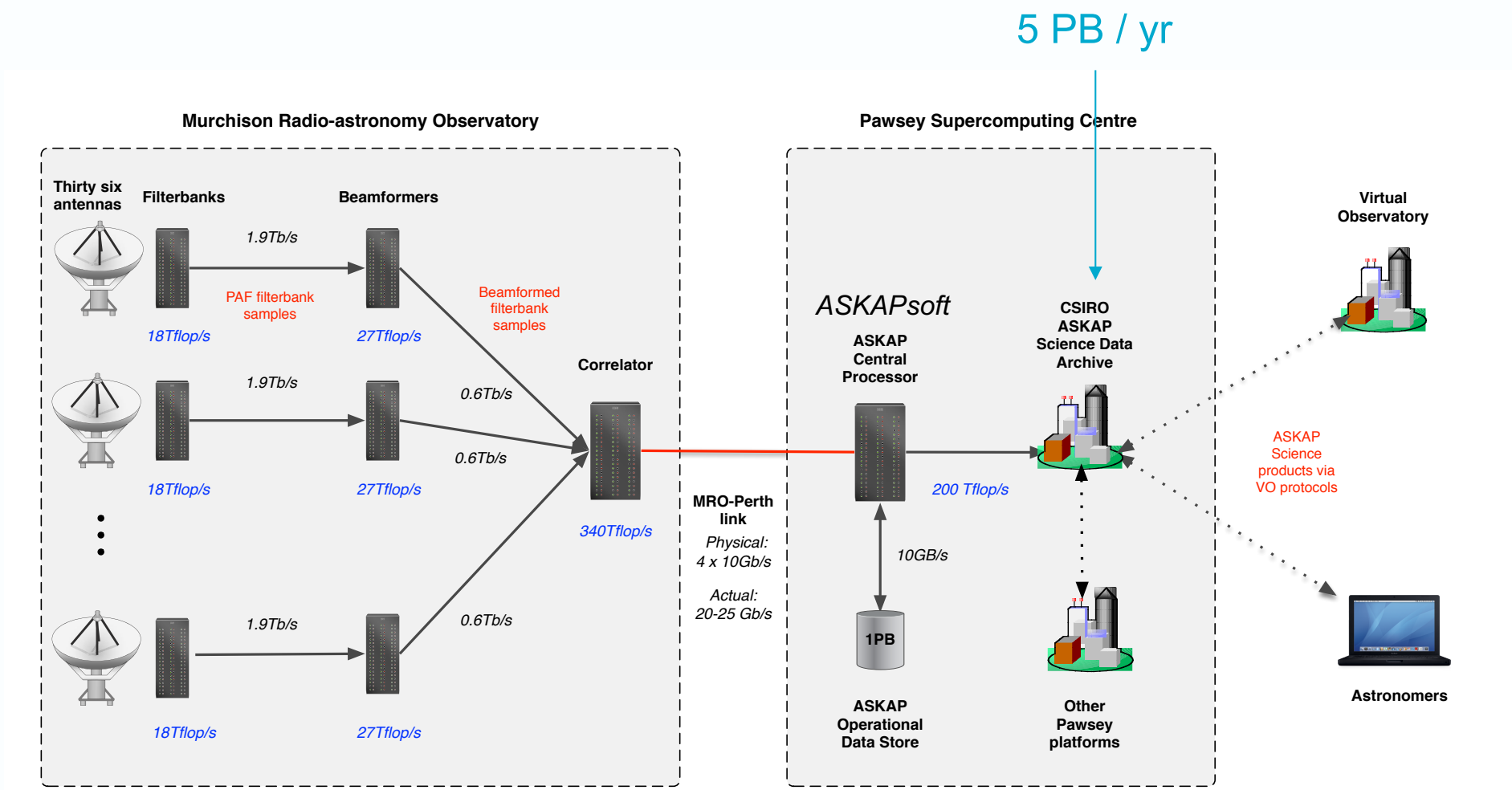
**Minh Huynh (CSIRO, ICRAR/UWA)**
**James Dempsey, Matt Whiting and the CASDA team**

# ASKAP – A next generation radio telescope

Australian SKA Pathfinder (ASKAP):

- 36 element interferometer
- Steerable parabolic dishes, 12m in diameter.
- Operating frequency: 700-1800 MHz.
- Wide instantaneous BW: 300 MHz.
- **High Spectral resolution**: 16200 channels (~18.5kHz)
- **Phased Array Feeds: Very Wide Field of view** (~30 sq deg)
- Roll-axis: High dynamic range imaging
- **High survey speeds**
- **Also very High Data Volume/rates**

# ASKAP and all its wonderful data



# CASDA Deployment

| | PAWSEY SUPERCOMPUTING CENTRE | CANBERRA |
|---|---|---|
| Location | Perth, WA | Canberra, ACT |
| Functions | •Deposit ASKAP data products<br>•Data access<br>•Virtual Observatory | •Interactive search<br>•Collections<br>•Authentication/Authorisation<br>•Data Validation/Release<br>•User interface |
| Facilities (Prod) | •2 dedicated servers<br>•2 x 10 PB tape library (shared)<br>•Lustre filesystem (shared)<br>•256 TB disk (initial allocation) | •5 virtual machines<br>•Integrated with CSIRO Data Access Portal |
| Facilities (Dev, Test, AT) | •5 dedicated servers<br>•Lustre filesystem (shared)<br>•384 TB disk (initial allocation) | •15 virtual machines<br>•Integrated with CSIRO Data Access Portal |

# CASDA Software Environment

Java

Tomcat

PostgreSQL

NGAS

SLURM

Elastic Stack (formally ELK)






Open source libraries

# CASDA Functionality

CASDA provides long-term archiving of and access to the large scientific datasets taken by ASKAP.

Functionality includes:

- Long term storage of ASKAP science data products
- Searches and data access via web (CSIRO Data Access Portal) and Virtual Observatory services
- Validation of ASKAP observations
- Upload of value-added science catalogues and image cubes by users
- Digital Object Identification (DOI) for all datasets
- Archive administration, inc. team member access to unreleased data

Approach:

- Agile Scrum software development

# CASDA Data Products

CASDA stores the following data products:

## Calibrated visibilities

- Archived long-term for continuum data only (frequency averaged)

## Images and Image cubes, Spectra and Moment Maps

- Basic image products produced by the pipelines
- Derived image products, such as extracted spectra for detected sources, moment maps, etc

## Catalogues

- Detected sources and their parameters

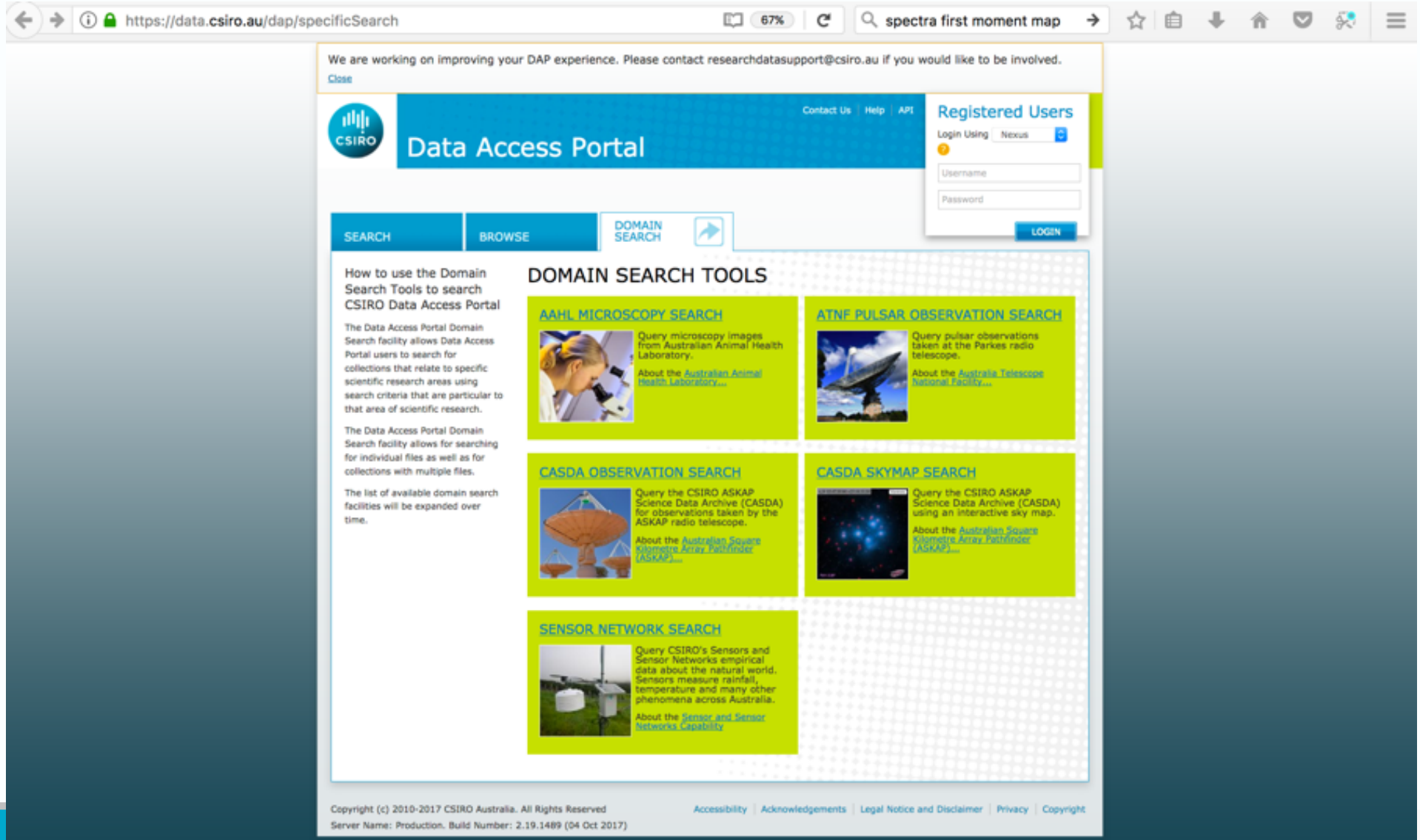
## Project and Observation information

## Data quality information

- Used to evaluate the quality of an observation
- Evaluation metric files

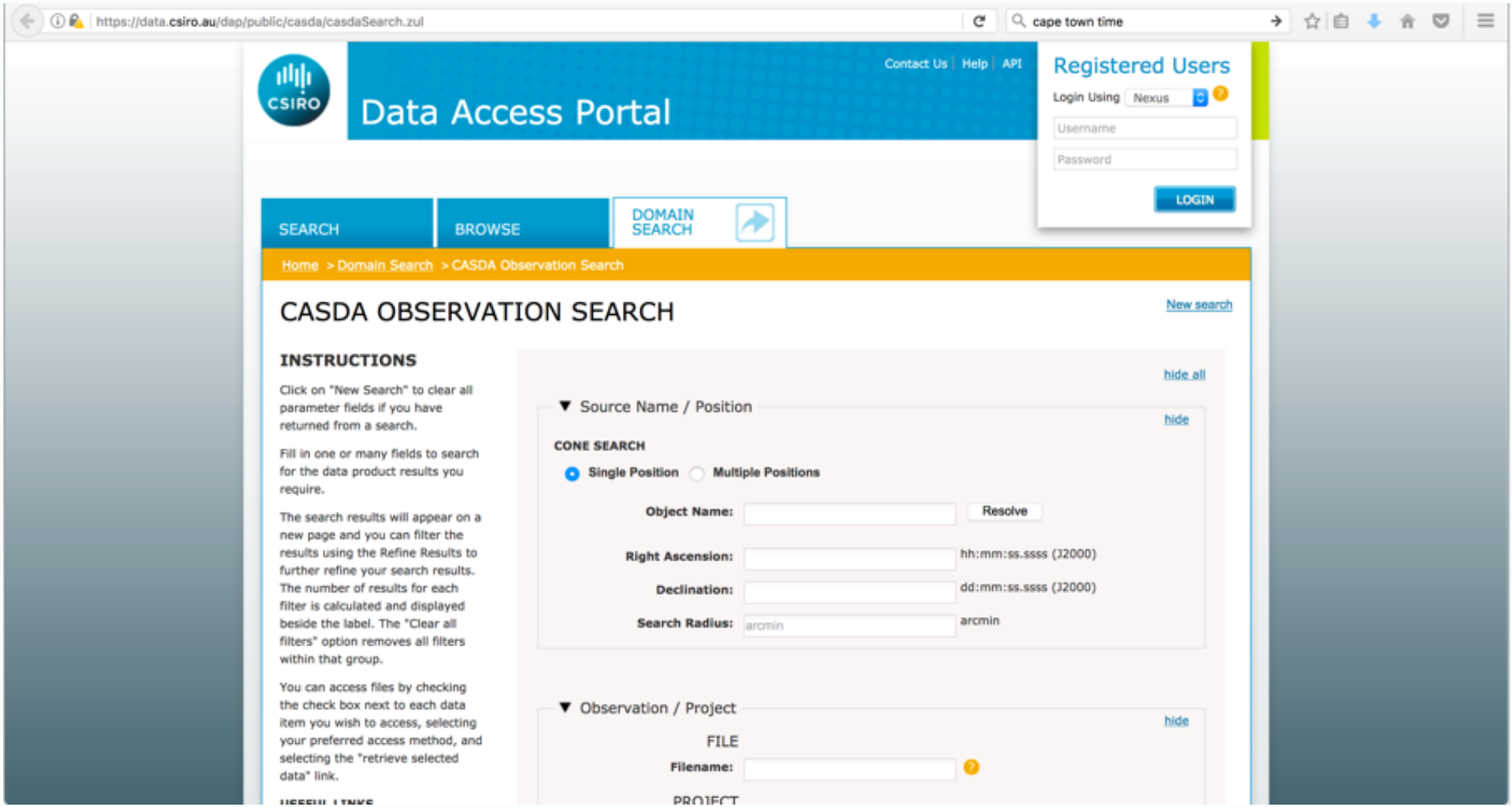
# CSIRO Data Access Portal

https://data.csiro.au/dap/specificSearch



# Data Access Portal query interface

https://data.csiro.au/dap/public/casda/casdaSearch.zul



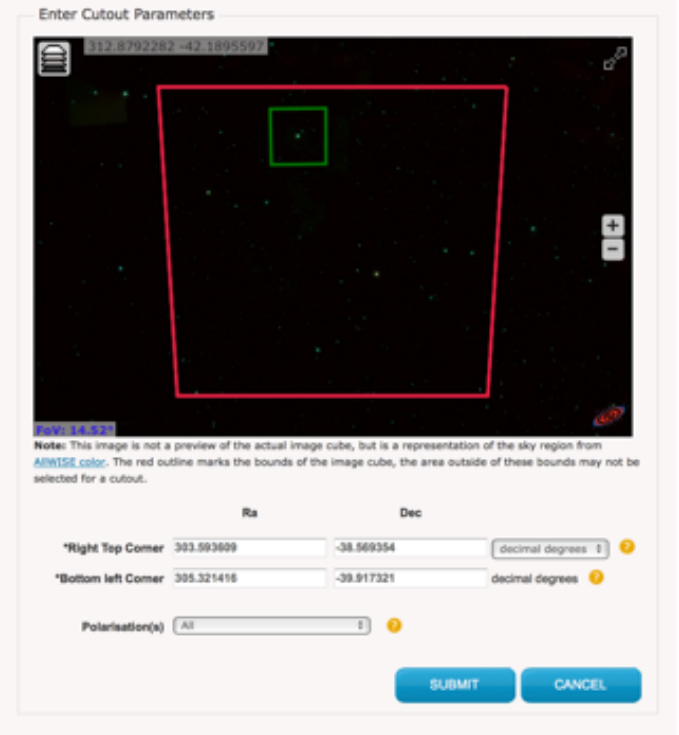
# Data Access Portal Capabilities

- Search for images/cubes and catalogues via
  - Coordinates, cone search
  - Object name
  - Project or Scheduling Block ID
  - filename

- Generate image/cube cutouts via Aladin Lite tool

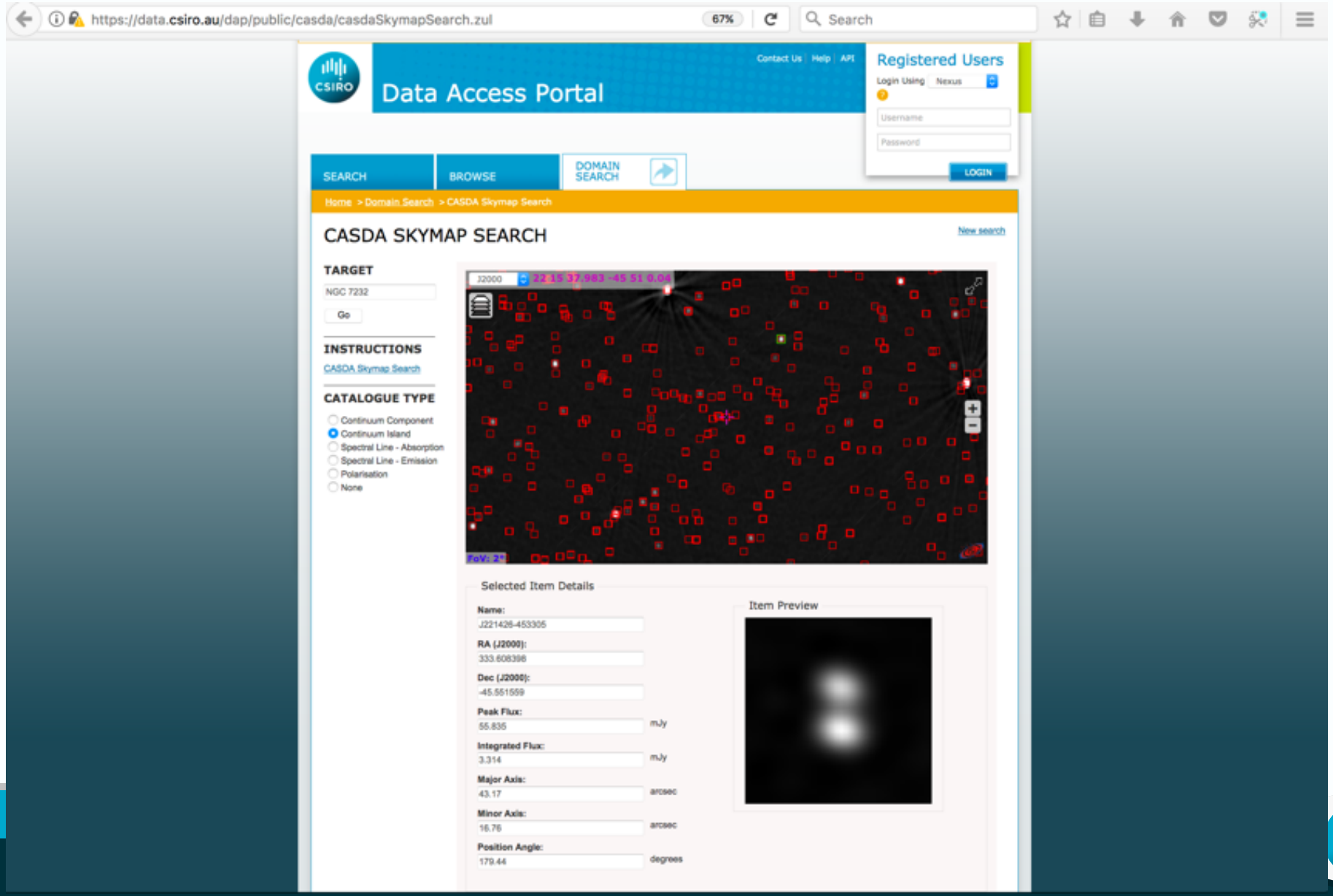
- Data validation and release

- Upload of User Generated Catalogues, Images and Cubes



# Data Access Portal skymap search with Aladin Lite

https://data.csiro.au/dap/public/casda/casdaSkymapSearch.zul



# CASDA and Virtual Observatory

CASDA work is continuing to build CSIRO's skills, knowledge and international reputation in VO protocols and services.

Main Protocols

- Table Access Protocol (TAP)
- Simple Cone Search
- Simple Image Access Protocol (SIA v2)
- Server Operations for Data Access (SODA)
- Simple Spectral Access Protocol (SSAP)

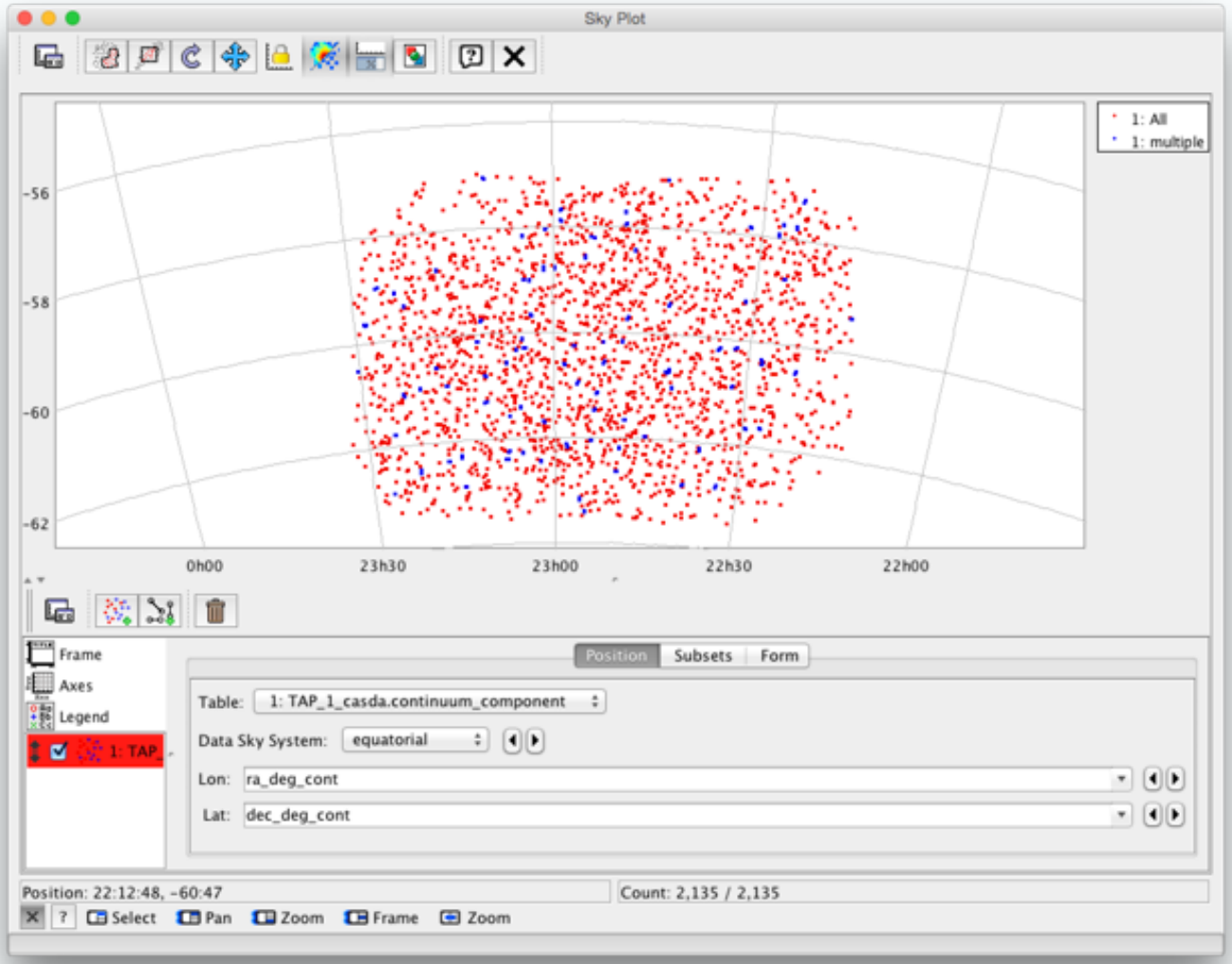
Automated data discovery and retrieval via scripts

Also used as a service layer:

- Cutouts
- File access
- Catalogue Retrieval

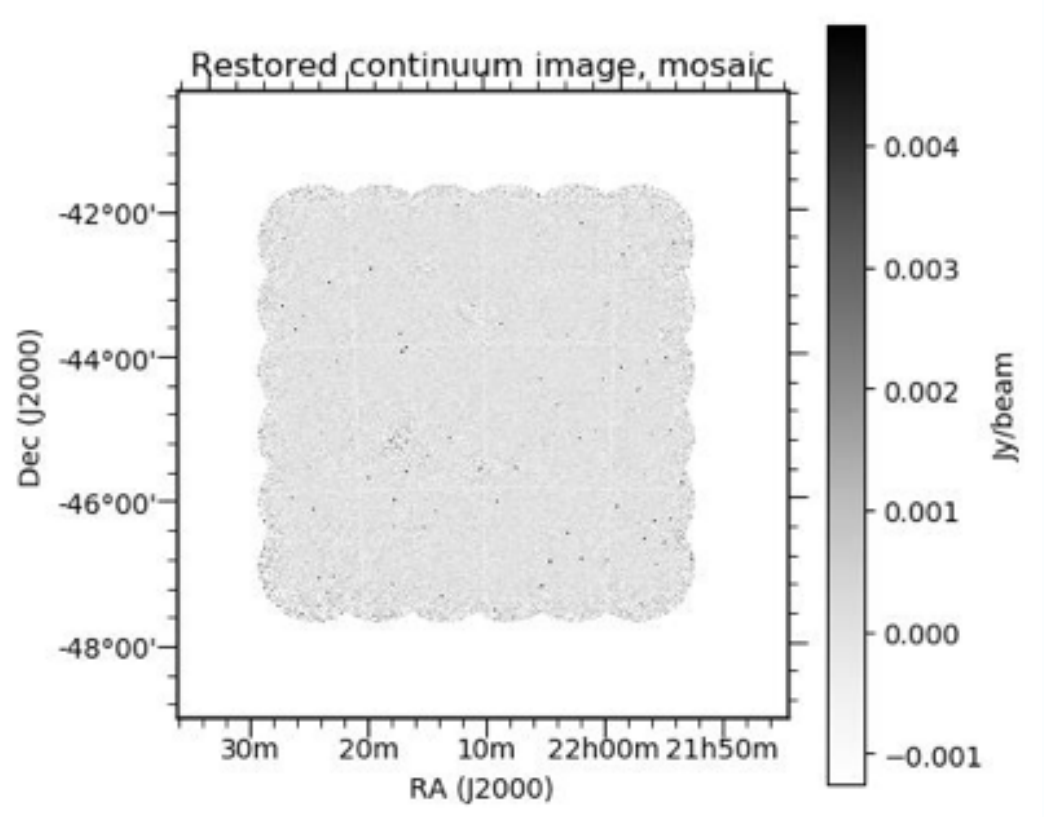
CASDA Virtual Observatory Tools available: http://doi.org/10.4225/08/59dc56fda2997

# CASDA VO via TOPCAT



# First ASKAP Early Science CASDA Data Release

- Observations of NGC 7232 and surrounding area
  - 12 MkII PAFs, 48 MHz bandwidth, 36 beams
- Eleven scheduling blocks (or days)
  - not combined, this is left to the user
- Data released:
  - ASKAPsoft continuum images
  - Selavy continuum source catalogues
  - Calibrated visibilities (freq binned) as CASA measurement sets



# Summary

ASKAP will produce ~5 PB a year of processed data products

CSIRO ASKAP Science Data Archive (CASDA) provides:

- Long term storage for science-ready data products - images, cubes, catalogues, calibrated visibilities (continuum) , spectra and moment maps
- Search and download via web (Data Access Portal) and Virtual Observatory services

ASKAP Early Science has started, first dataset now available.

- Expect more datasets soon!