The CSIRO ASKAP Science Data Archive Progress and Plans

Jessica Chapman, CSIRO on behalf of CASDA team

CSIRO ASTRONOMY AND SPACE SCIENCE
www.csiro.au

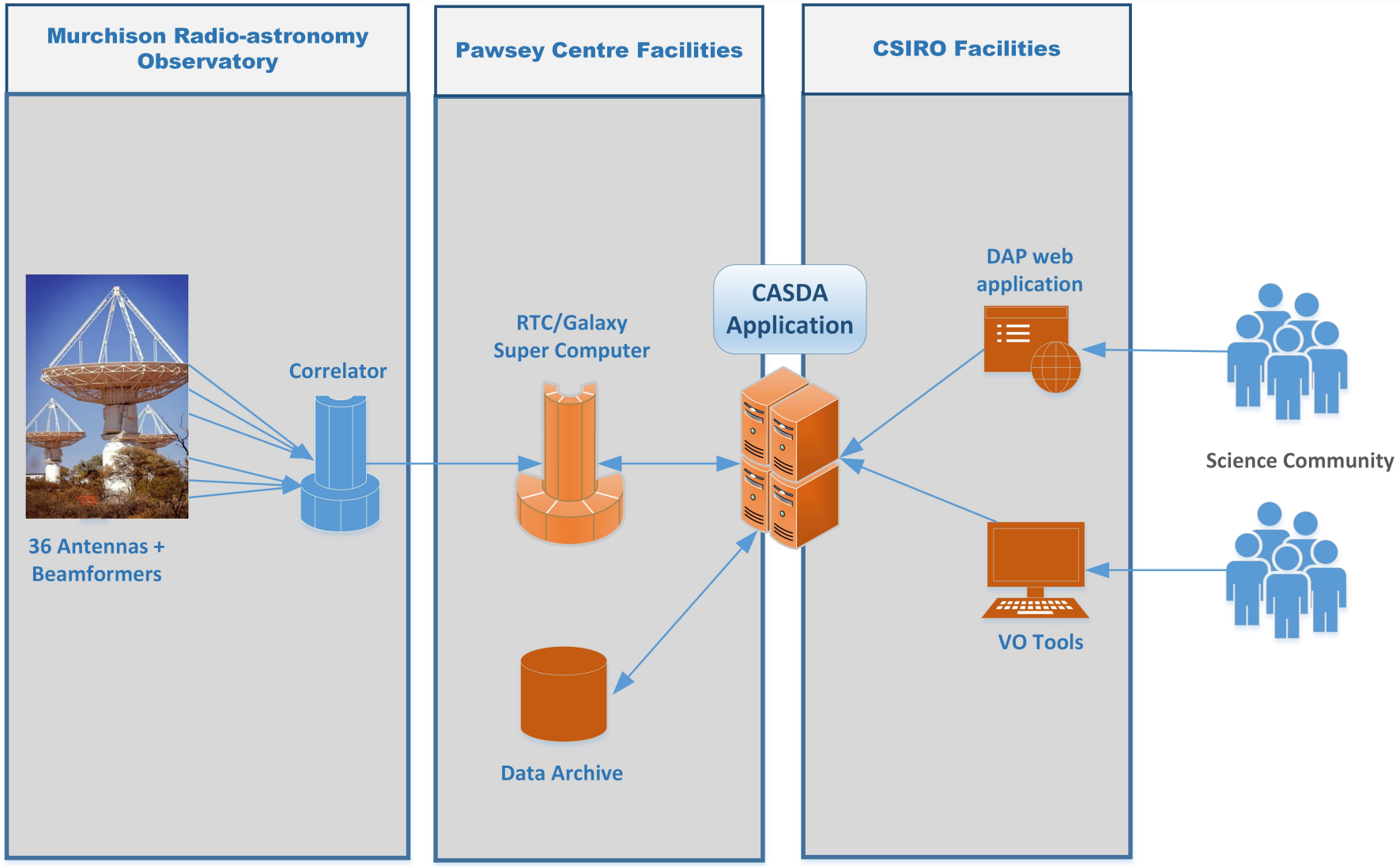
CASDA Timeline

Start	End	Project Stage	Notes
Aug 2013	Jun 2014	0	Analysis stage
21 May 2014	-	0	User Requirements doc to community
1 July 2014	End-Feb 2016	1	20 Sprints
8 Mar 2016	End-Jun2016	1.5	3 Sprints
Jul 2016	~ Dec 2017	2	tbd

Production releases

1.0	5 Nov 2015
1.1	15 Feb 2016
1.2	21 Apr 2016 (with software)
1.3	*23 Jun 2016 (tbc)*

ASKAP Data Management



ASKAP Data Products

- **Calibrated visibility data files**

  archived for all continuum data (as 300 channels x 1 MHz)

- **Images (2-d) and image cubes (3-d and 4-d)**

  continuum, spectral line, polarisation, moment maps ..

- **Catalogues**

  source detection tables (with data quality information)

  project-related information

ASKAP data products from the Survey Science projects are made publically available for general use after data quality validation.

CASDA data rate for full operations: 15 TB per day, 5 PB per year.

The CSIRO ASKAP Science Data Archive (CASDA)

CASDA provides a 'big data' science archive for ASKAP Data Products. The application supports:

- ❑ Long term data storage at Pawsey Centre
- ❑ Searches and data access using CSIRO Data Access Portal and VO services
- ❑ Data uploads of science catalogues provided by Science Teams
- ❑ Tools for setting data validation flags and quality information
- ❑ User authentication and tools to manage team members
- ❑ Digital Object Identifiers (DOIs)
- ❑ Archive administration tasks: user mgt, queue control, monitoring and reports etc.

CASDA web access is through the CSIRO Data Access Portal

SEARCH DATA | BROWSE DATA | DOMAIN SEARCH

Welcome Jessica Chapman Logout

How to use the Domain Search Tools to search CSIRO Data Access Portal

The Data Access Portal Domain Search facility allows Data Access Portal users to search for data collections that relate to specific scientific research areas using search criteria that are particular to that area of scientific research.

The Data Access Portal Domain Search facility also allows for searching for individual data files as well as for data collections of multiple data files.

The list of available domain search facilities will be expanded over time.

DOMAIN SEARCH TOOLS

AAHL MICROSCOPY SEARCH

Query microscopy images from Australian Animal Health Laboratory.

About the Australian Animal Health Laboratory...

ATNF PULSAR OBSERVATION SEARCH

Query pulsar observations taken at the Parkes radio telescope.

About the Australia Telescope National Facility...

CASDA OBSERVATION SEARCH

Query the CSIRO ASKAP Science Data Archive (CASDA) for observations taken by the ASKAP radio telescope.

About the Australian Square Kilometre Array Pathfinder (ASKAP)...

SENSOR NETWORK SEARCH

Query CSIRO's Sensors and Sensor Networks empirical data about the natural world. Sensors measure rainfall, temperature and many other phenomena across Australia.

About the Sensor and Sensor Networks Capability...

CASDA Virtual Observatory (VO) Services

CASDA includes VO data services that make use of international protocols:

VO Service	Notes
Table Access Protocol	Access to catalogues. Allows complex ADQL queries and easy cross-matching between catalogs from different surveys.
Cone Search Protocol	Simple cone-searches around given positions.
Simple Image Access Protocol (v2) SODA	Provide access to images and image cubes. 2-d, 3-d, 4-d image cut-out service where user selects regions of interest within one or more cubes

All VO services can be accessed programmatically – i.e. with python scripts

CASDA in Production

First release (v1.0: Nov 2015) included:

- Pawsey infrastructure established (4 environments)
- Open access to science data products from BETA observations
- Ready for use for start of ASKAP Early Science
- Established service agreement between CSIRO and Pawsey
- User support through helpdesk system
- Provide user documentation, example python scripts and training sessions
- VO TAP and cone searches and some early image-access development

CASDA v1.1: Feb 2016

- VO: Using scripts for data access
- VO : authenticated access to unreleased to unreleased products
- VO: 3-d image cut-outs
- VO: Example python scripts for bulk cut-outs
- DAP: Support for direct transfers within Pawsey Supercomputing Centre for users with Pawsey accounts
- DAP: Administration of project roles (e.g. allocation of validation rights for a project)

CASDA v1.2: April 2016

- DAP: Many small changes to improve usability, based on user feedback (~ 30% development time)
- DAP: Multiple cone searches
- DAP + VO: Handling of multiple scheduling blocks
- DAP + VO: Support image ‘types’
- VO: Improved cut-out services (2-d, 3-d and 4-d)

CASDA Software Open Source release: May 2016

Open release of CASDA VO software available:

- CSIRO Data Access Portal (binary code)
- GitHub (source code)

The VO application software can be configured using configuration files and/or a user interface.

Other major CASDA applications are also available on GitHub.

Enquiries welcome: atnf-datasup@csiro.au

External interest

CASDA is now generating interest from external groups:

Group	Interest
MeerKAT	VO services – downloaded and trying out software
Swinburne	VO software for TAP
FAST	VO services – downloaded and trying out software Potential small contract for FAST data feasibility study
Murchison Widefield Array	Some initial discussions (mainly with AAL) on potential for supporting MWA science data products
Breakthrough Project	Request for data support (JER handling this)

Careful management will be needed for any contracts to maintain our reputation and ability to deliver.

CASDA Stage II Planning (in progress)

In 2016/2017 – plan to extend Data Access Portal and Virtual Observatory data services:

Key goals

- ASKAP 1-d spectra and time domain data products
- ASKAP images, cubes, spectra, time series, cross-matched catalogues etc provided by Science Survey Teams as ASKAP 'post-archive' data products
- Extend CASDA to include legacy data products (images / cubes/ spectra/ catalogues) from surveys taken with the Australia Telescope Compact Array and Parkes radio telescope.
- Continue to respond to User Feedback

CASDA support for ATNF legacy surveys

- About a dozen large-scale surveys already carried out with Parkes and/or Australia Telescope Compact Array AND future surveys

- Science teams have produced high quality data products (images, spectra, catalogues)

- Their results are published and data products are stored but with limited access

- Archive services for these ATNF data products would significantly extend their ‘re-use’ for science research.

CASDA Stage II Resources

Some funding is available but probably insufficient to cover the full project plan

	Full plan	Possible descope
	IM&T (6.0 FTE)	IM&T (3.8 FTE)
Project Manager	0.5	0.3
Business Analyst	1.0	0.5
Software Developers	3.0	2.0
Testers (costed to project)	1.5	1.0
	CASS (in kind, 1.2 FTE)	CASS (1.2 FTE)
Operations/admin (TA)	0.6	0.6
CASS Data Scientist	0.4	0.4
Consultants (IH/MW/JCG)	0.2	0.2

ATUC advice

Given restricted funding – What priority order would ATUC give to the CASDA Stage II key goals?

Key goals

- Support ASKAP 1-d spectra and time domain data products

- Support ASKAP images, cubes, spectra, time series, cross-matched catalogues etc provided by Science Survey Teams as ASKAP ‘post-archive’ data products

- Extend CASDA to include legacy data products (images/ cubes/ spectra/ catalogues) from surveys taken with the Australia Telescope Compact Array and Parkes radio telescope.

To get started with CASDA see the CASDA User Guide:

http://www.atnf.csiro.au/observers/data/casdaguide.html

Position available: Senior Data Scientist in Astrophysics

For enquiries – see Simon Johnston or Jessica Chapman