

Data Central's Adventures in the VO

datacentral.org.au

Simon O'Toole
AAO Macquarie

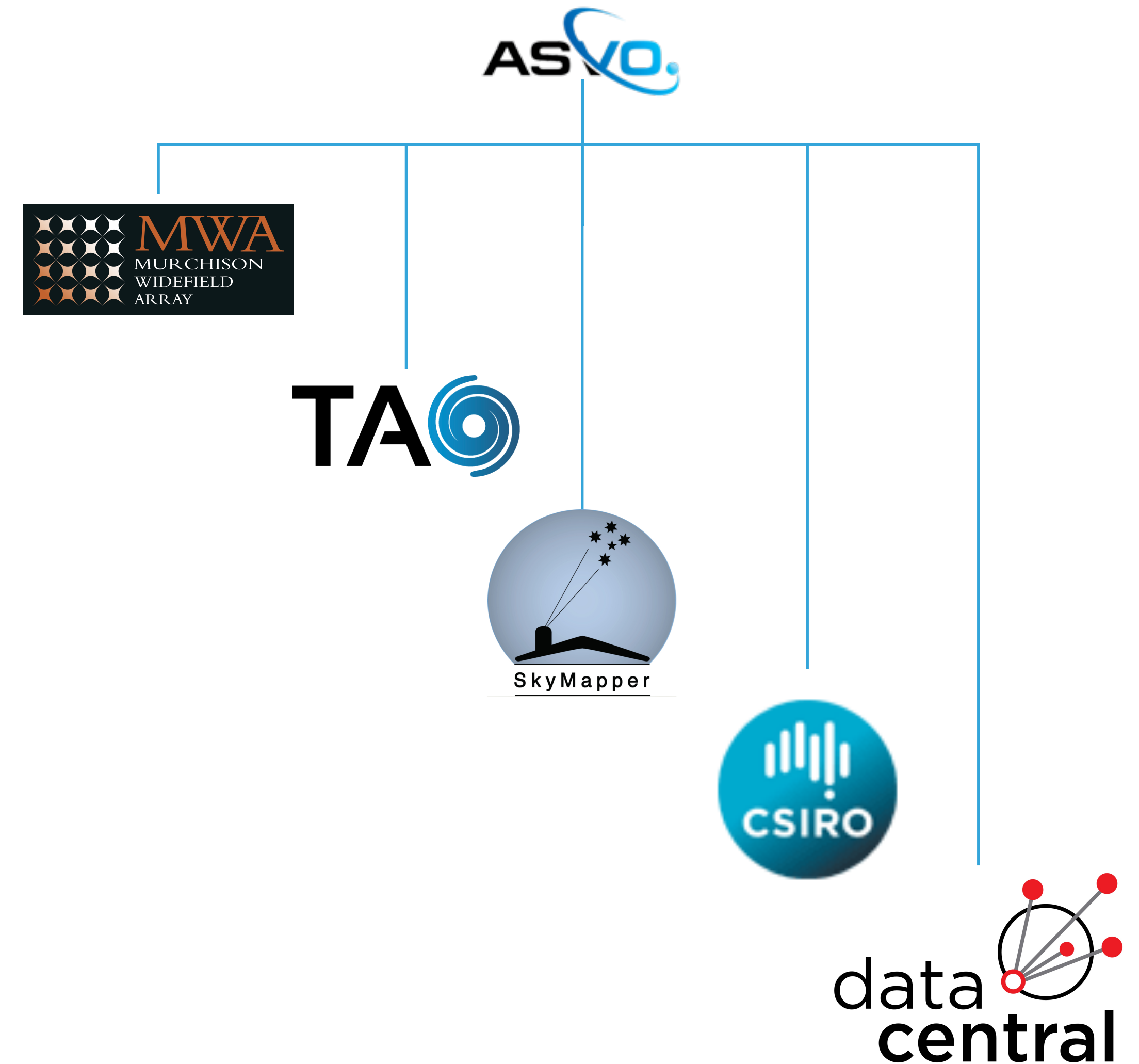


MACQUARIE
University
SYDNEY · AUSTRALIA



The All-Sky Virtual Observatory

- Murchison Widefield Array
- Theoretical Astrophysical Observatory
- SkyMapper
- CASDA (incl ASKAP archive)
- Data Central (incl AAT archive)





Explore. Collaborate. Science.

Incorporating the Optical Data Centre, an AAL-funded initiative

DC v2.0

We aim to give astronomers, students etc the data and tools they need to do research without needing to download all the data

The Data Central Science Platform

A data ecosystem that provides a large variety of different services

- Data services, API, and VO services
- Cloud storage (ownCloud)
- Remote desktops for ESO/MUSE data reduction (Apache Guacamole)
- JupyterHub - Python, R and Julia notebooks
- Data Aggregation Service

All the data!

- GAMA DR 2 and 4
- SAMI DR 1, 2, 3
- GALAH DR 1, 2, 3
- OzDES DR 1, 2
- DEVILS DR 0, 1 (soon)
- RAVE DR 5
- WiggleZ Final DR
- 2dFGRS Final DR
- 6dFGS Final DR
- S7 DR 2
- GLEAM-X DR 1
- LoBES DR 1
- S5 DR 1
- Deeper Wider Faster DR 1 (soon)
- MAGPI DR 1 (soon)
- **Perth Obs Digitised Plates**

Authentication and Authorisation

- Primarily use Apereo CAS service at auth.datacentral.org.au for internal services
- Migrating to more recent version at auth-mfa.datacentral.org.au
 - Includes MFA if required – this service is used by DC admins and non-astro projects
- Also use OAuth2.0 for external services such as CSIRO DAP
- Both CAS and OAuth2.0 allow authorisation via attributes (usually group membership)
- Looking at API tokens to allow programmatic access

FAIR requirements

- Funding agencies in Australia (and everywhere) are requiring data to be FAIR
- IVOA standards get us 95% of the way (see O'Toole & Tocknell, ADASS XXXI)
- But there are still things to focus on:
 - Data licensing – default to CC BY 4.0
 - DOI minting – should return IDs with query results

IVOA Standards implemented

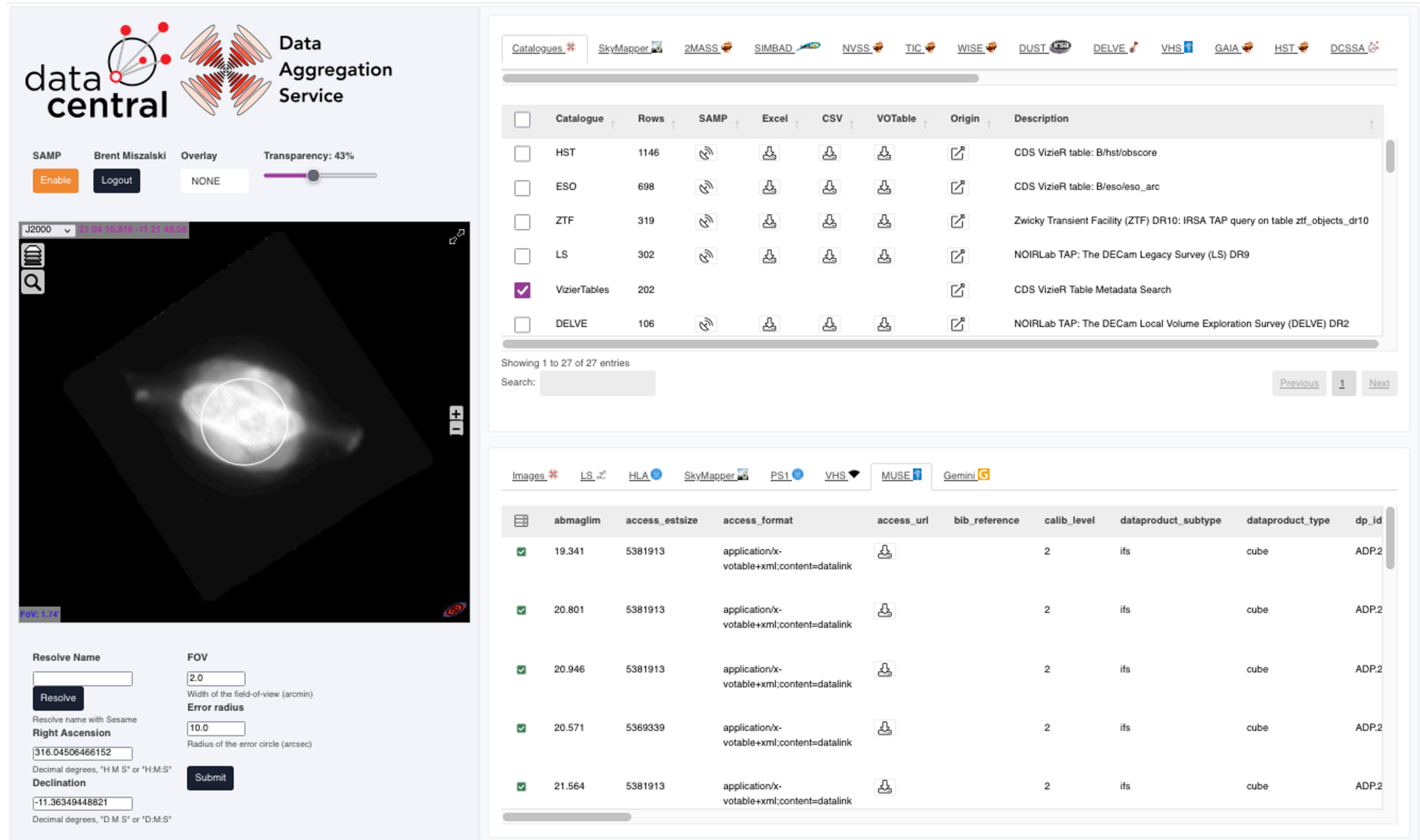
- *Table Access Protocol + ADQL*: query catalogue data in Data Central, now including Table Uploads
- *Simple Cone Search*: Web portal version includes aggregated cone searches
- *Simple Image Access*: query multi-wavelength images, including cutout service
- *Simple Spectrum Access*
- *MOCs*: used in Data Aggregation Service
- *DataLink* connects data products
- *UCDs* tie everything together

Data Aggregation Service

- Integrates Aladin Lite, SCS, SIA with async python
- Uses some non-VO services

Only possible with VO!

das.datacentral.org.au



The screenshot displays the Data Aggregation Service interface. The top left shows the 'data central' logo and 'Data Aggregation Service' text. Below this, there are user controls for 'SAMP' (Brent Miszalski) and 'Overlay' (NONE), along with a 'Transparency: 43%' slider. The main content area is divided into two panels. The left panel shows a visualization of a galaxy with a white circle overlaid, and a search form with fields for 'Right Ascension' (316.04506466152) and 'Declination' (-11.36349448821). The right panel shows a table of search results for catalogues.

Catalogue	Rows	SAMP	Excel	CSV	VOTable	Origin	Description
<input type="checkbox"/> HST	1146						CDS VizieR table: B/hst/obscure
<input type="checkbox"/> ESO	698						CDS VizieR table: B/eso/eso_arc
<input type="checkbox"/> ZTF	319						Zwicky Transient Facility (ZTF) DR10: IRSA TAP query on table ztf_objects_dr10
<input type="checkbox"/> LS	302						NOIRLab TAP: The DECam Legacy Survey (LS) DR9
<input checked="" type="checkbox"/> VizieRTables	202						CDS VizieR Table Metadata Search
<input type="checkbox"/> DELVE	106						NOIRLab TAP: The DECam Local Volume Exploration Survey (DELVE) DR2

abmaglim	access_estsize	access_format	access_url	bib_reference	calib_level	dataproduct_subtype	dataproduct_type	dp_id
<input checked="" type="checkbox"/> 19.341	5381913	application/x-votable+xml;content=datalink			2	ifs	cube	ADP.2
<input checked="" type="checkbox"/> 20.801	5381913	application/x-votable+xml;content=datalink			2	ifs	cube	ADP.2
<input checked="" type="checkbox"/> 20.946	5381913	application/x-votable+xml;content=datalink			2	ifs	cube	ADP.2
<input checked="" type="checkbox"/> 20.571	5369339	application/x-votable+xml;content=datalink			2	ifs	cube	ADP.2
<input checked="" type="checkbox"/> 21.564	5381913	application/x-votable+xml;content=datalink			2	ifs	cube	ADP.2

Connecting with containers

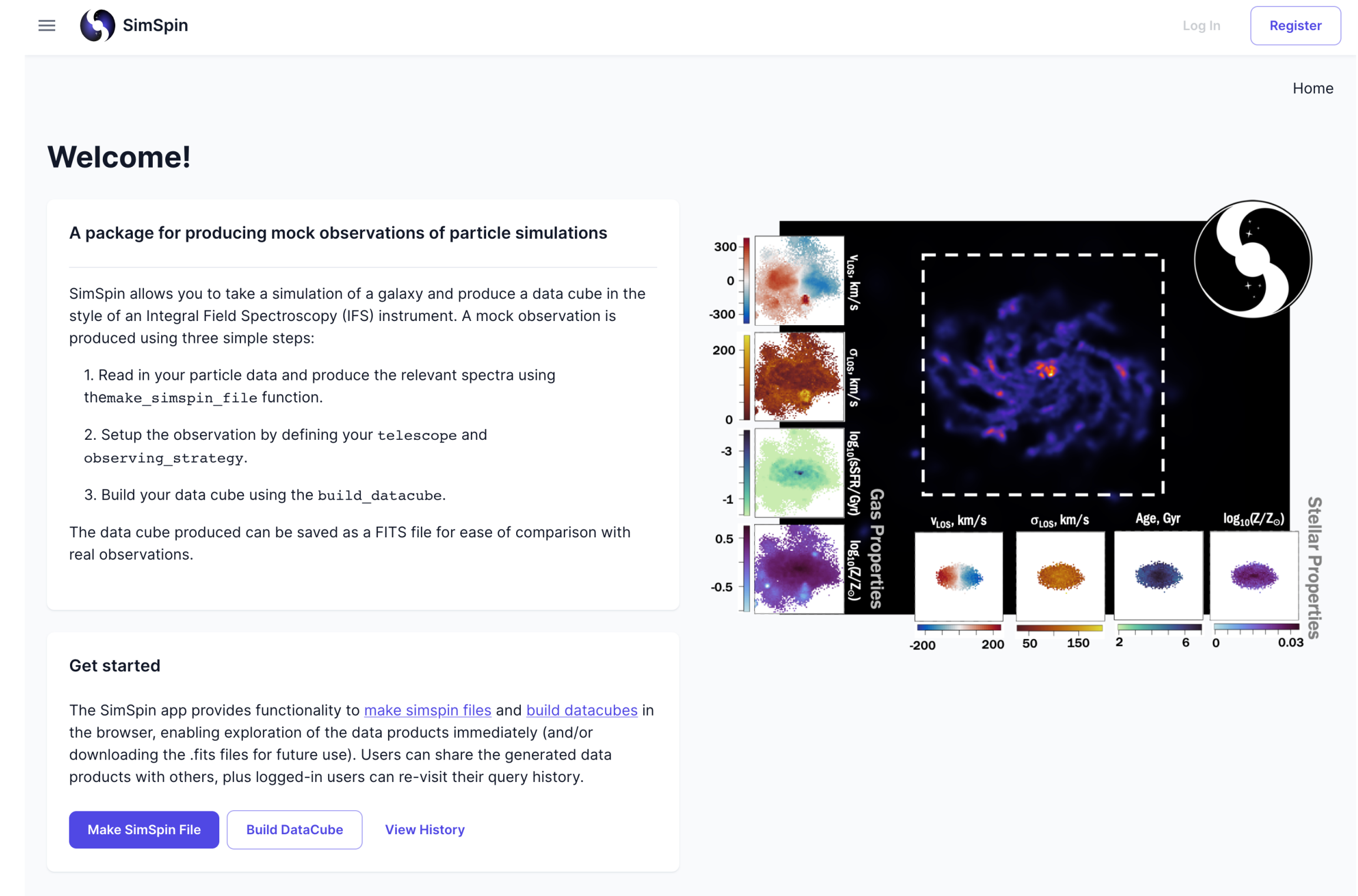
- Much of Data Central runs in containers, making it somewhat portable
- Self-hosted Kubernetes cluster running JupyterHub
- Also make extensive use of docker-py for orchestration, e.g.
 - Pipelines As a Web Service – ideal for telescope archives
 - New data validation service – team-managed data ingestion

Challenges and roadblocks

- Biggest challenge is connecting observations with theory/simulations
- We are looking for ways to link these data together - DataLink?
- How do we even describe simulation data?
 - UCDs are typically observation focused
- We have several ESO Large Programs asking us for this
 - See Theory IG session for more discussion

Linking Observations and simulations

- We are developing a web framework for generating accessing simulated data
- Prototype is SimSpin, which includes an API
- Want to implement SimDM underneath
- Still need vocabularies and discoverability



SimSpin

Log In Register

Home

Welcome!

A package for producing mock observations of particle simulations

SimSpin allows you to take a simulation of a galaxy and produce a data cube in the style of an Integral Field Spectroscopy (IFS) instrument. A mock observation is produced using three simple steps:

1. Read in your particle data and produce the relevant spectra using the `make_simspin_file` function.
2. Setup the observation by defining your telescope and `observing_strategy`.
3. Build your data cube using the `build_datacube`.

The data cube produced can be saved as a FITS file for ease of comparison with real observations.

Get started

The SimSpin app provides functionality to [make simspin files](#) and [build datacubes](#) in the browser, enabling exploration of the data products immediately (and/or downloading the .fits files for future use). Users can share the generated data products with others, plus logged-in users can re-visit their query history.

[Make SimSpin File](#) [Build DataCube](#) [View History](#)

Gas Properties: v_{los} , km/s; σ_{los} , km/s; $\log_{10}(\text{SSFR}/\text{Gyr})$; $\log_{10}(Z/Z_{\odot})$

Stellar Properties: v_{los} , km/s; σ_{los} , km/s; Age, Gyr; $\log_{10}(Z/Z_{\odot})$

We're hiring!

- Fancy working in Australia?
- We have a range of roles available on astronomy and non-astronomy projects
- Come and speak to me!



Summary

The Data Central Science Platform:

- provides access to ever increasing volumes of multi-wavelength data
- makes extensive use of VO standards
- allows users to bring their code near the data in multiple ways
- increasingly allows portability of components with containers