

# TAPPING IN HADOOP WITH AAO DATA CENTRAL

Simon O'Toole – Australian Astronomical Observatory



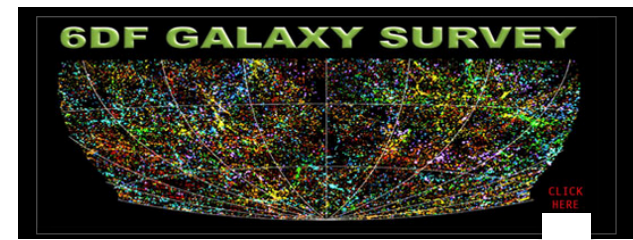
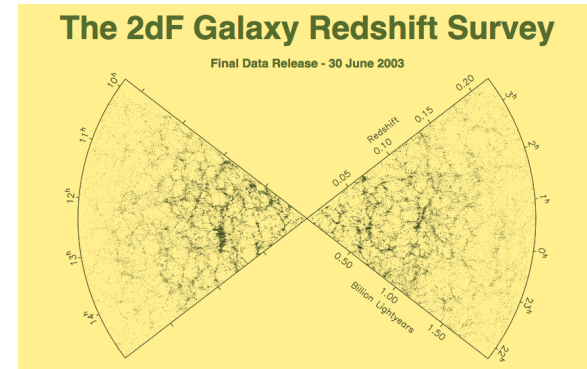
# Data Central Team

- Prof Andrew Hopkins (Sponsor)
- Dr Katrina Sealey (Project Manager)
- Dr Simon O'Toole (Project Scientist)
- Mr Lloyd Harischandra (Database Specialist)
- Dr Liz Mannering (UX/UI Specialist)
- Mr Minh Vuong (AAO Software Engineer)



# AAT Survey History

- AAT survey teams have been at the forefront of data access, starting with the 2dF Galaxy Redshift Survey
- Many surveys have continued the trend: 6dFGS, WiggleZ, RAVE & GAMA all have portals
- Historically no VO-compliance
- Long-term maintenance is difficult
- The data stand in isolation and not easily cross-matched



# Motivation

- A central repository for AAT & UK Schmidt data is needed
  - Long-term availability
  - Multiple heterogeneous data sets
  - Cross-matching capability
  - IVOA compliant services
- All Sky Virtual Observatory is the logical fit (formerly Aus-VO)

**⇒ the AAO Data Central project**

# The Problem

- Data Central is required to:
  - serve heterogeneous data (and datastores)
  - query semi-structured and structured data
  - access remote data
  - be IVOA compliant wherever possible
  - be interoperable with other ASVO and IVOA nodes

# Pilot launched August 2017



HOME

SURVEYS

DATA

DOCUMENTATION

ABOUT

HOME

**i Important Notice** AAO Data Central v2 launches in August 2017. We emphasize that all survey data served through ADC are the *final versions*, but the web interface, access methods and API are subject to change.



Australian  
Astronomical  
Observatory

Query Australian Astronomical Observatory Data Central (ADC) and retrieve data from current and legacy surveys. Bring your code to the data.

GitHub-like tools for Astronomers

SURVEYS

View the Surveys hosted at ADC.

RAW OBSERVATIONS

View observations collected from the AAT or Schmidt telescope (coming soon).

## WHAT'S NEW

### SAMI DATA RELEASE 1.0

View the data release page [here](#).

### Coming soon: Siding Spring Southern Seyfert Spectroscopic Snapshot Survey (S7)

View the survey description [here](#).

## GET STARTED

### Surveys Overview

View all surveys hosted at ADC.

### Data Access Overview

Information about accessing data at ADC.

### Single Object Viewer

Browse data products per Survey object.

### Examples

Help using ADC.

# The Solution

- Use Hadoop/HDFS distributed data storage
- “Federated” data access without the cost: PrestoDB (<https://prestodb.io>)
- Allows distributed queries across remote heterogeneous datastores
  - Apache Hive
  - MySQL
  - PostgreSQL
  - MongoDB
  - Redis
- **BUT** we need a new spherical geometry layer to allow ADQL queries through TAP



# TAP Service

- CDS implementation with modifications for PrestoDB
  - New spherical geometry functions
  - Built using PrestoDB user defined functions
  - UDFs implemented per ADQL function permutation (e.g. intersects\_circle\_circle, contains\_point\_circle)
  - Map to INTERSECTS, CONTAINS, etc
- With thanks to Gregory Mantelet (Heidelberg) for help
- With TAP\_SCHEMA for remote datasets, we can query all of them at once
  - e.g. GAMA DR2 stored in Sydney (Hive), AAT raw observation log at Siding Spring Observatory (MySQL)
- <https://datacentral.aao.gov.au/adc/tap>



# Next steps

- Finalise ADQL functions
- Implement TAP upload
- Simple Cone Search endpoints (alpha)
- Simple Image Access (in development)
- Register Data Central Services
- VOSpace 2.1 (under investigation)
- Simple Spectrum Access (under investigation)

# New data coming soon

- OzDES DR1 (2017)
- 2dFLens (2017)
- S7 Galaxy Survey DR2 (2017)
- Anglo-Australian Planet Search (2017)
- Taipan Galaxy Survey (2017+)
- FunnelWeb Stellar Survey (2017+)
- DEVILS Galaxy Survey (2017+)
- Hunstman Telephoto Array (2018+)
- GALAH DR2 (2018)
- SAMI DR2 (2018)
- GAMA DR3 (2018)
- HECTOR (2019+)

# Data Central V1.5 (November)

HOME SURVEYS SERVICES API DOCS VERSION 1 SUPPORT LOG IN

data central

Query, explore and retrieve data from current and legacy surveys.  
Bring your code to the data.

SURVEYS

Survey	Logo	DR Status
GAMA	GAMA logo	DR2 PDR
SAMI galaxy survey	SAMI logo	DR1
GALAH	GALAH logo	DR1
S7	S7 logo	DR1
2dFGRS	2dFGRS logo	DR1
6dFGS	6dFGS logo	DR3