

Data Central's Data Aggregation Service



*Casting a
wide net in
the ocean of
astronomical
data*

Dr Brent Miszalski
brent.miszalski@mq.edu.au



MACQUARIE
University

data  central



data
central

Explore. Collaborate. Science.

DC v1.12

- Data Central (datacentral.org.au) is an e-research platform and data archive developed at Australian Astronomical Optics (AAO), Macquarie University, that facilitates cutting-edge science.
- It provides web-based tools and archive functionality for scientists from a range of disciplines to explore, collaborate and make new discoveries.
- New SIA (July 2020) and **SSA** (February 2021) services: **Thursday May 27, 06:30 UTC**

Project motivation



- **Dr Stuart Ryder** (Astronomy Australia Limited/Macquarie University) and **Lachlan Marnoch** (PhD student, Macquarie University)
- The Commensal Real-time ASKAP Fast Transients (**CRAFT**) Survey:
 - High-time resolution survey for fast radio transients (700-1500 MHz) with the Australian SKA Pathfinder (ASKAP) e.g. Fast Radio Bursts (FRBs)
- Given an FRB candidate position and its uncertainty, CRAFT want to:
 - Find out what data are available near position: Host galaxy redshift? Deep images? (e.g. DES DR2) Catalogues? Archival spectra?
 - Check quickly: decide whether to coordinate follow-up observations
 - Check many: repeat for large numbers of candidates

Can we use existing applications?

CDS Portal

CDS Portal

VizieR

VizieR

IRSA/Firefly

IRSA/Firefly

MAST Portal

MAST Portal

Aladin

Aladin

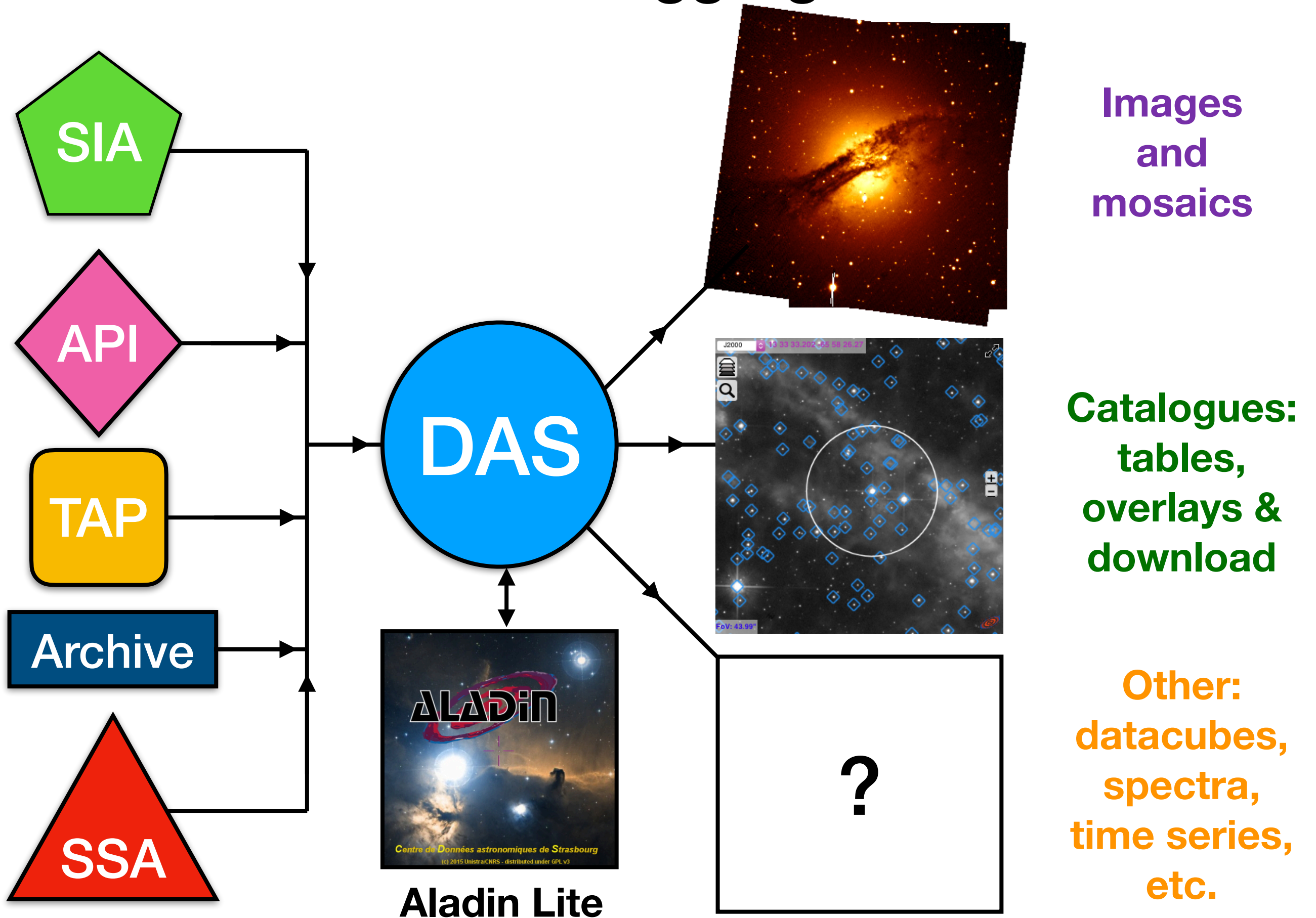
Telescope archives (e.g. ESO, Gemini, etc.)

Telescope archives (e.g. ESO, Gemini, etc.)

Not exactly! Introducing the Data Aggregation Service...


- Available portals for data discovery are **very powerful** and **mature** platforms.
- **Aladin/CDS Portal** have a broad focus, but others typically have a narrow, specialised focus (e.g. **VizieR**: catalogues; **IRSA**: IR; **MAST**: UV and optical).
- Check each portal sequentially? Time consuming and may be hard to manage with many manual steps (especially for many candidates).
- **Our approach**: We leverage **IVOA DAL** + **web services** to perform queries and collate their results with help of **Aladin Lite** from within a single web application.
- **Data Aggregation Service**: Collates data from multiple sources inspired by CRAFT team requirements. (Dockerised Django Python3/Javascript application)
- **Input GET parameters**: RA, DEC, field-of-view and radius (position uncertainty).
- **Service abstraction**: Users do not have to know how to query each service: we handle everything under the hood so it **Just Works™**

Schematic of Data Aggregation Service



DAS interface

Catalogues (Vizier, TAP, SSA, API)



**CRAFT
Data
Aggregation
Service**

SAMP

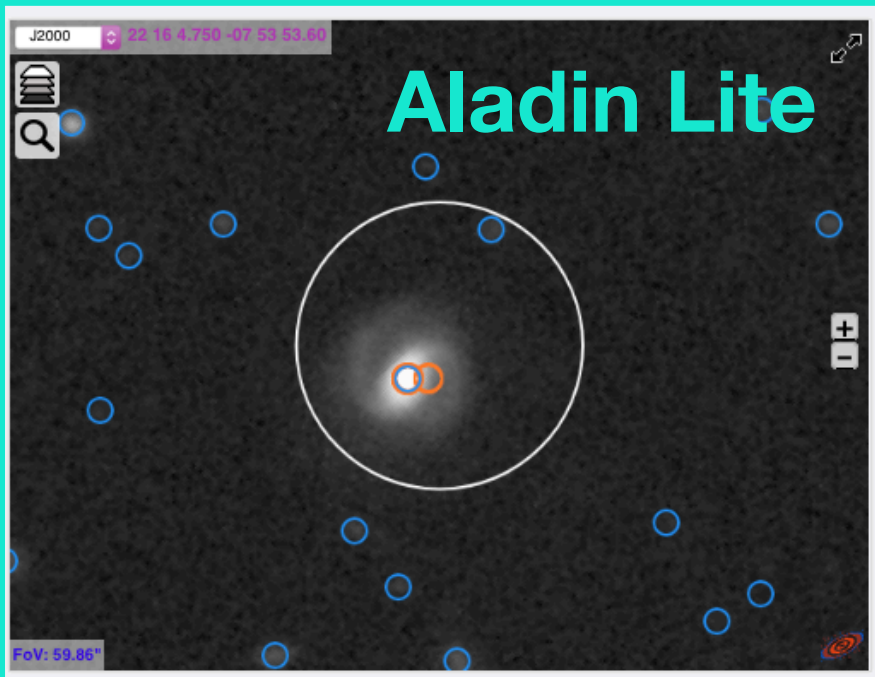
Enable

Click to enable SAMP

Overlay survey: RACS

Overlay transparency: 0%

J2000 22 16 4.750 -07 53 53.60
Aladin Lite



FoV: 59.86"

Resolve Name

 Resolve

Resolve target name with CDS Sesame service

Right Ascension

Decimal degrees, "H M S" or "H:M:S"

Declination

Decimal degrees, "D M S" or "D:M:S"

FOV




Width of the field-of-view (arcmin)

Error radius

Radius of the error circle (arcsec)

Submit

_r	RA	DEC	Z	ZCOLS	vizier_cat	vizier_cat_description	vizier_tab_description
2.3436	334.020000	-7.898840	0.1157	z	VII/281/glade2	GLADE v2.3 catalog (Dalya+, 2018)	GLADE v2.3 catalog (Galaxy List for the Advanced Detector Era)
2.4804	334.020000	-7.898880	0.1178	z	J/ApJS/223/20/spec	SDSS-DR8 galaxies classified by WND-CHARM (Kuminski+, 2016)	Morphological catalog of SDSS objects with spectra
2.5056	334.019989	-7.898890	0.1176	z	J/MNRAS/455/2440/catalog	Gal. 2D phot. decompositions in r, g & i bands (Meert+, 2016)	General catalog data (band-independent data) (UPenn_PhotDec_CAST)
2.5056	334.019989	-7.898890	0.1176	z	J/MNRAS/446/3943/catalog	galaxies 2D phot. decompositions in SDSS-DR7 (Meert+, 2015)	General Catalog Data (band-independent data used in fitting and analysis)
3.0996	334.020375	-7.898861	0.1180	z	J/ApJ/667	Mass function of active black	The sample

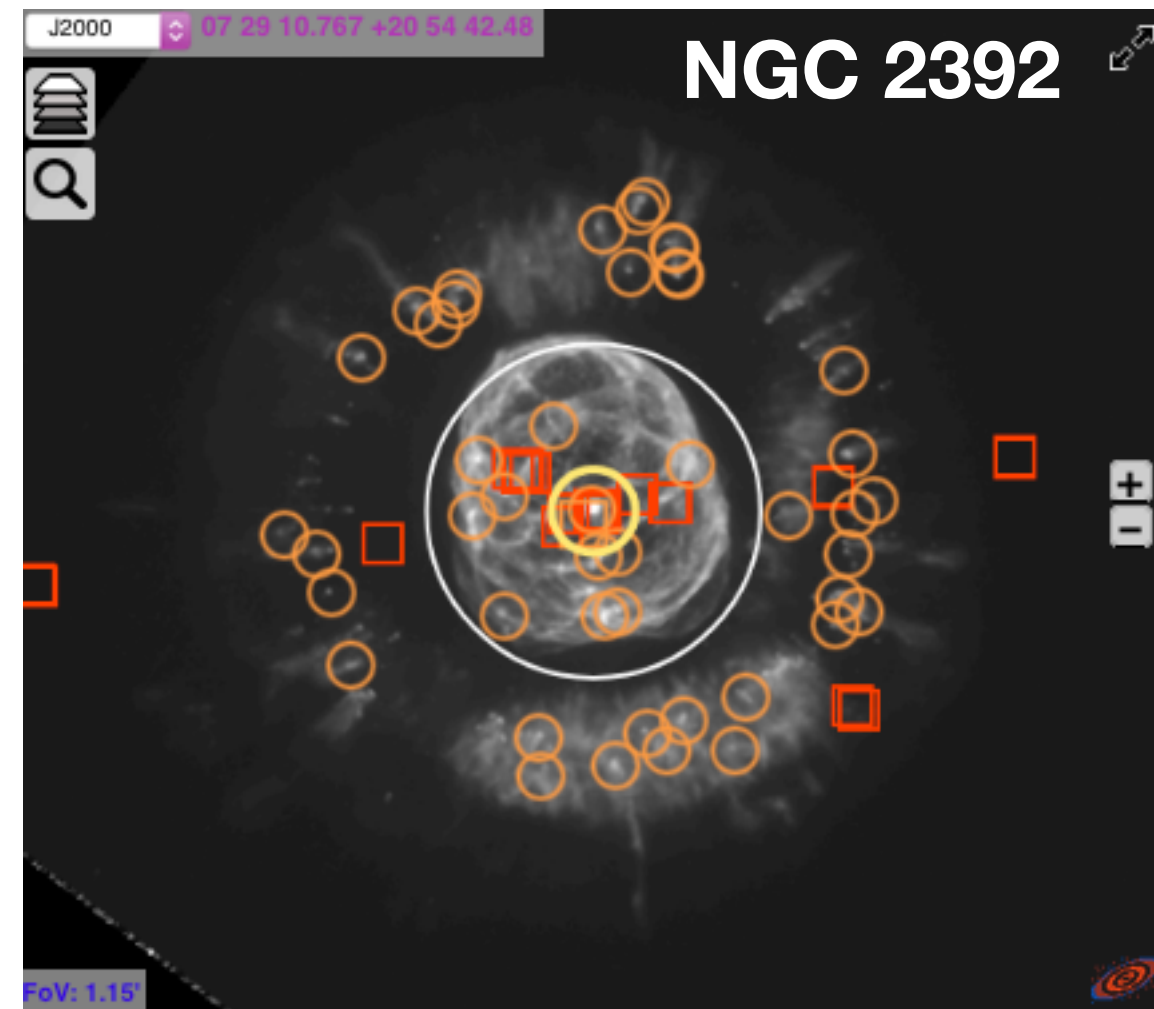
obs_id	instrument_name	obs_bandpass	preview	s_ra	s_dec	em_min	em_max	em_res_power	object	airmass	project_code
ls_dr8	DECam	g DECam SDSS c0001 4720.0 1520.0		333.908	-7.99998	0.0	0.0	nan		0	DECaLS DR8
ls_dr8	DECam	r DECam SDSS c0002 6415.0 1480.0		333.908	-7.99998	0.0	0.0	nan		0	DECaLS DR8
ls_dr8	DECam	z DECam SDSS c0004 9260.0 1520.0		333.908	-7.99998	0.0	0.0	nan		0	DECaLS DR8

Images (SIA, API)

SIMBAD Name resolver

Catalogue handling

- **Query TAP / SSA / API (e.g. VizieR ASU) services:** return VOTable format results. TAP ADQL query encoded into query urls.
- Results added as Aladin Lite catalogue layers. Empty catalogues are removed.
- Javascript callback creates HTML table. Mouseover on table rows: Yellow highlight appears in Aladin Lite.
- **Redshift aggregator:** custom endpoint that queries and collates all VizieR catalogues with non-empty columns where UCD = src.redshift or src.redshift.phot.
- Toggle display of catalogues via checkboxes (individually or as group).
- Web SAMP export (e.g. to TOPCAT, javascript adapted from VizieR), download formats: Excel, CSV, VOTable



















<input checked="" type="checkbox"/>	Catalogue	Status	Link	SAMP	Excel	CSV	VOTable
	SAMI	No sources	SAMI				
<input checked="" type="checkbox"/>	HST	1116 sources	More info				
<input checked="" type="checkbox"/>	GAIA	46 sources	More info				
	SDSS	No sources	More info				
<input checked="" type="checkbox"/>	WISE	4 sources	More info				
	SUMSS	No sources	More info				
<input checked="" type="checkbox"/>	NVSS	1 source	More info				

Image handling

- Adapted Aladin Lite function to (directly) load FITS images via **fits2hips** web service (CDS)!
- **SIA services are ideal:** Data Central (DEVILS and GAMA swarps), NOIRLab Astro datalab (e.g. DES DR2), Hubble Legacy Archive
- **APIs:** PANSTARRS DR2, DECAPS, Gemini archive (json).
- **Mosaics-on-demand:** Archival imaging data often needs mosaicking before Aladin Lite (~15 sec). **Custom mosaicking endpoints created:**
 - Multi-extension FITS (Gemini GMOS; json archive query),
 - Two files for one image (ESO FORS2; ESO TAP_OBS query),
 - Three images over ~10 arcmin (SkyMapper DR3; SIA query).
- **Pipeline As a Web Service (PAWS)** could fully reduce images to ensure proper WCS (FORS2 and GMOS raw data only have approximate WCS).
- **Image selection:** Mouse over in Images section or Aladin Lite directly.



NGC5189: Archival Gemini GMOS image navigation



CRAFT Data Aggregation Service

SAMP

Enable

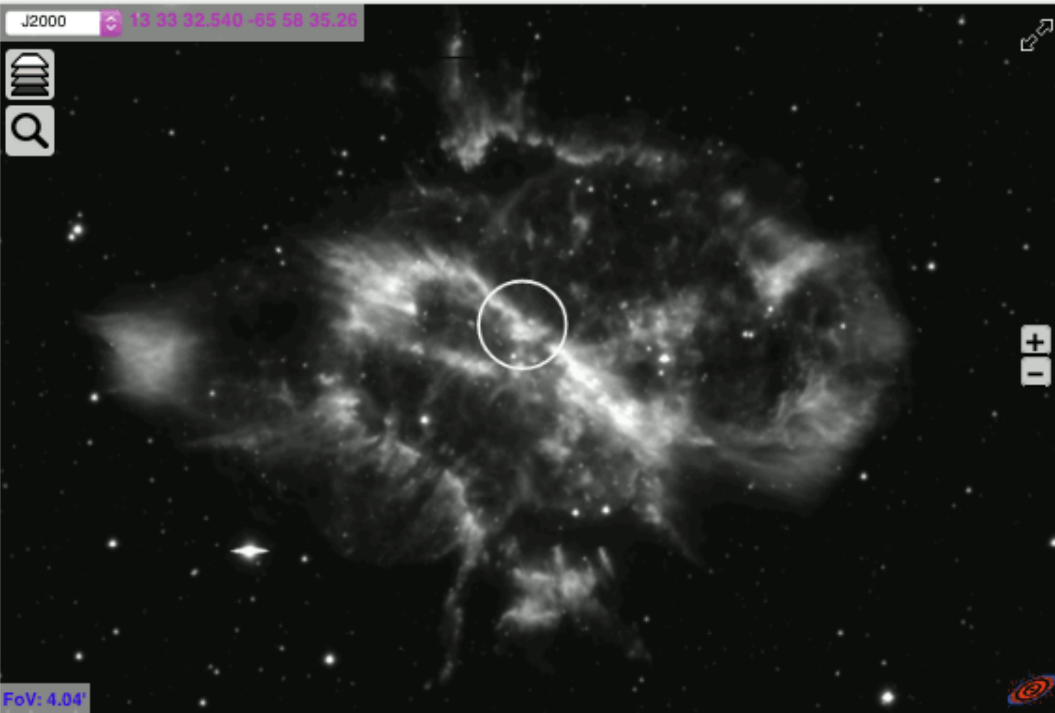
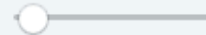
Click to enable SAMP

Overlay survey

RACS

Overlay

transparency: 0%



Resolve Name

Resolve target name with CDS Sesame service

Right Ascension

Decimal degrees, "H M S" or "H:M:S"

Declination

Decimal degrees, "D M S" or "D:M:S"

FOV

Width of the field-of-view (arcmin)

Error radius

Radius of the error circle (arcsec)

Submit

Catalogues

ESO

HST

GAIA

WISE

RACS

DECAPS

DUST

IRSA

<input type="checkbox"/>	Catalogue	Status	Link	SAMP	Excel	CSV	VOTable
<input type="checkbox"/>	SAMI	No sources	SAMI				
<input type="checkbox"/>	ESO	286 sources	More info				
<input type="checkbox"/>	HST	483 sources	More info				
<input type="checkbox"/>	SDSS	No sources	More info				
<input type="checkbox"/>	PhotRedshifts	No sources	More info				
<input type="checkbox"/>	GAIA	236 sources	More info				
<input type="checkbox"/>	DELVE	No sources	More info				
<input type="checkbox"/>	NVSS	No sources	More info				

Images

SIA

SkyMapper

DECAPS

PS1

Gemini

DES

LS

SMASH

SDSS

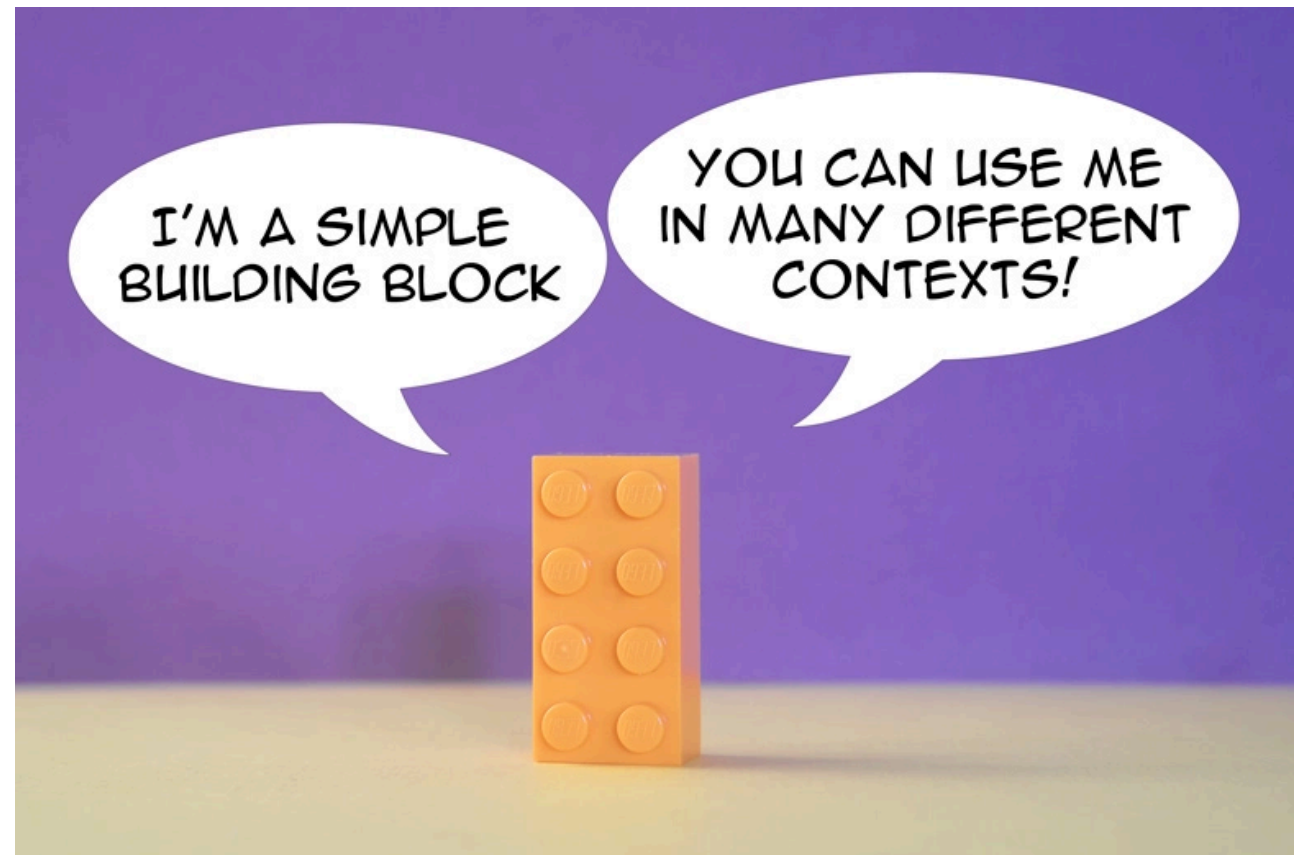
object	ra	dec	azimuth	elevation	cass_rotator_pa	airmass	filter_name
ngc5189	203.390667080331	-65.973968976745	164.163956944444	51.259098611111	-54.583982231924	1.278	Ha
ngc5189	203.390641571578	-65.973977122724	167.528206944444	52.5220875	-62.699552783934	1.257	SII
ngc5189	203.390536099272	-65.973893129679	171.262058333333	53.476652777778	-71.176596747396	1.242	OIII
ngc5189	203.390462917282	-65.973972693683	189.672016666667	53.262933333333	-110.912469487987	1.249	Ha

Future plans

- Currently released to CRAFT for testing...plan to release **DAS** as general service
- Add authenticated user profiles:
 - Customise: What services to query. What columns to display.
 - Make available restricted data sets (e.g. SkyMapper DR3; Private team data)
- Discoverability of services via registries (IVOA DAL; HiPS, etc)
- Broader usage of MOCs: Check survey coverage. Only query services with target.
- Improve service query performance
 - Launch multiple queries asynchronously and cache results (Celery/redis)
- Expand PAWS (Pipeline As a Web Service) usage beyond simple mosaics
 - Reduction of archival images and spectra from raw files, etc.

Final thoughts

- Consider IVOA services as building blocks for web applications or meta-services.
- Greater integration of IVOA services by web portals: enhance interoperability.
- Abstracting IVOA services: helpful to improve accessibility for average users.
- Intermediary web services could play a large role in opening up IVOA to people with a disability (André Schaaff talk on Wednesday) and young people (Education interest group on Wednesday).
- What other meta-services should we be building? Time domain explorer? (find all the time series data for an object). Themed explorers? High energy. Radio.



thevaluable.dev