

Data Aggregation Service

Brent Miszalski brent.miszalski@mq.edu.au

Simon O'Toole and James Tocknell (AAO Macquarie) Lachlan Marnoch (Macquarie) Stuart Ryder (AAL/Macquarie)





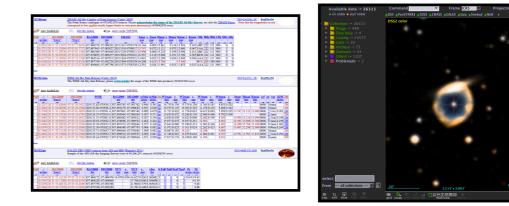
Overview

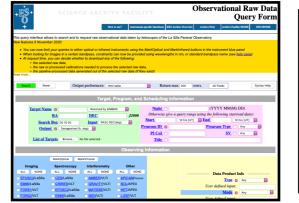
- There are so many astronomical data services online
- Querying each of them: starting to be overwhelming and tedious
- Transient surveys will generate very large lists of targets.
- We can't inspect **ALL** of them, but we may want to inspect a **SUBSET**.
- Focus has been on transient brokers and their alert streams but what happens downstream of this? What software ecosystem do we need?
- Data Central has developed extensible infrastructure to aggregate and visualise information for large numbers of targets. Data Aggregation Service
- **Observatory Archives** (e.g. AAT, ESO) can assume a prominent role:
 - They may contain precious observations for transients lacking data elsewhere.
 - Archives could issue automatic notifications (VOEvents) when transient alerts are cross-matched with archival data (or even future data) — an **active** archive!
 - Data may be reduced on-demand via Pipeline As a Web Service (PAWS) developed by Data Central and built into above infrastructure (e.g. 2dFdr).



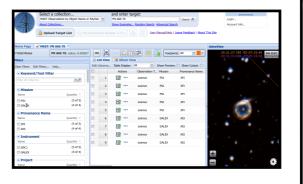
Project motivation

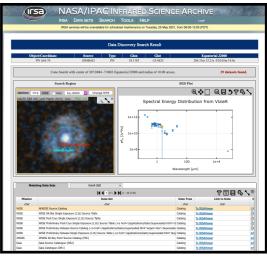
- Data Aggregation: Find all data for a sky position and radius (e.g. images and catalogues).
- **CRAFT survey team**: inspect FRB candidates. Large uncertainty in position. Host galaxy? Redshift?
- Several mature and powerful data platforms: VizieR, CDS Portal, Telescope Archives, MAST Portal, Aladin, IRSA/Firefly, etc.
- Checking each platform can be time consuming and error prone.
 Especially for large numbers of targets.







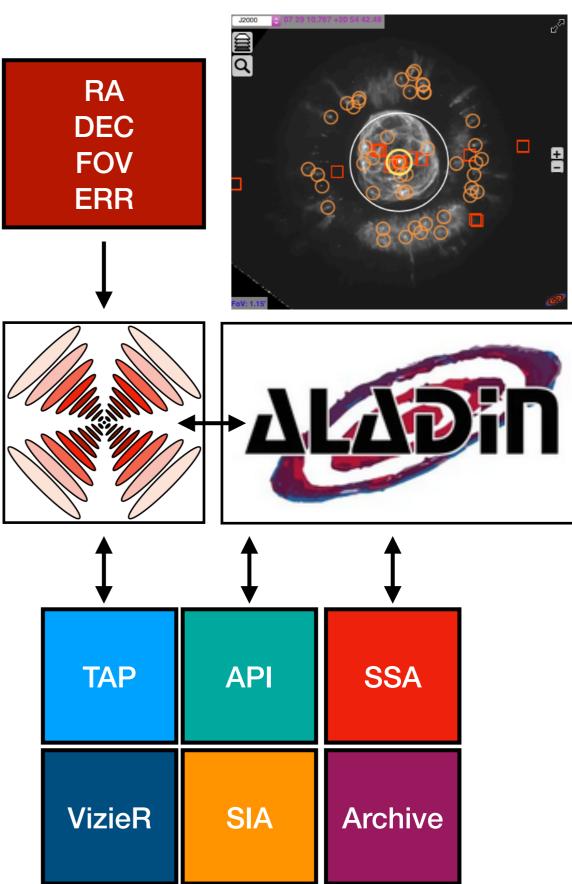






Data Aggregation Service

- A web application that simultaneously queries multiple online data services. Query results are displayed in browser and may be exported (TOPCAT and Aladin via Web SAMP, .csv, .xls, .vot)
- Aladin Lite: interactive visualisation of images and catalogues. Convert FITS images to HiPS on-the-fly.
- Input GET parameters: RA, DEC, FOV and ERR (position uncertainty).
- Fast and asynchronous from the ground up.



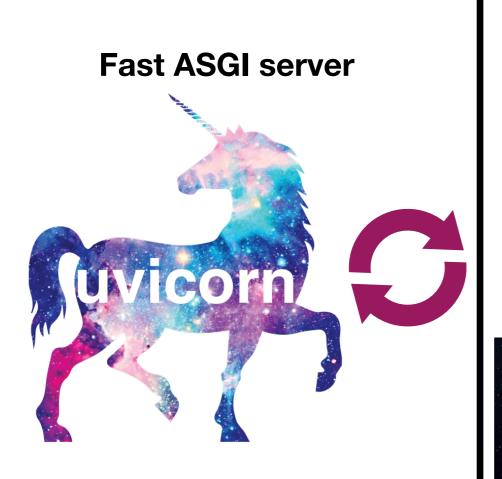
Data Aggregation Service

data Data Aggregation	Catalogues ** DCSSA & VizierTables & WISE & ESO & NVSS & SIMBAD & SkyMapper I TIC & DCSCS & GAIA & PS1 ZTF							
central Service	Catalogue	Status 265 rows	SAMP	Excel	сsv	VOTable	Catalogues	
SAMP Brent Miszalski Overlay Transparency: 100% Enable Logout NONE Image: Comparency of the second s	ZTF	241 rows	Ry R.	샾	쑈	싪	(Vizier, TAP, SSA, API)	
	ESO	139 rows 106 rows	ちょう	문 문	샧 산	쇼 쇼	Toggle on/off	
	TIC GAIA	36 rows	Ry Ry	쇼 쇼	윤	윤 윤	Mouseover: highlights	
	WISE	34 rows	Ś	샾	산	샾	location in image	
	 SkyMapper DCSSA 	32 rows 3 rows	Ś	문 문	쟋 산	문 문	Vizier tables: load on demand	
	Images ¥ LS ℒ Images Statu SMASH 0 row	PS1 SAMP	<u>yMapper</u>	<u>FOR</u>	S2	<u>Gemini</u>		
	LS 6 row DES 0 row HLA 0 row		윤	≞ (SI	д А,	AP	Images I, Custom Pipelines)	
Resolve Name FOV SINBAD Resolve 2.0 Width of the field-of-view (arcmin) Resolve name with Sesame Error radius Name Right Ascension 10.0 Radius of the error circle (arcsec)	HLA 0 rows PS1 5 rows SkyMapper 6 rows FORS2 1 row Trigger downloads: Large images (HST)							
307.888365097 Decimal degrees, "H M S" or "H:M:S" Declination -7.08834222636 Decimal degrees, "D M S" or "D:M:S"	Gemini 4 row DCSIA 0 row DECAPS 0 row	5		C	las	s.da	atacentral.org.au	

DAS structure

ASGI application

(Asynchronous Server Gateway Interface)



uvicorn.org



socket.io



starlette.io

(FastAPI ASGI web framework is built upon Starlette)

socket.io server (python-socketio)

Real-time bidirectional, event-based communication between server and web browser

Starlette ASGI web framework

Jinja2 templates, websockets, async background tasks, session middleware, etc.



Main dependencies

socket.io

websocket message

handling from browser



Aladin Lite FITS image and catalogue display. HiPS management: Hipsgen and MOCPy.



1342

Data Sources

IVOA Data Access Layer (DAL): TAP, SIA, SSA, SCS

VizieR catalogues (ASU interface)

APIs: PanSTARRS, Gemini archive, IRSA, SIMBAD (search and name resolver)



httpx/aiohttp async file downloads



redis VOTable and file caching



Celery async task execution invoked asynchronously from Starlette



MontagePy Mosaic production The Astropy Project

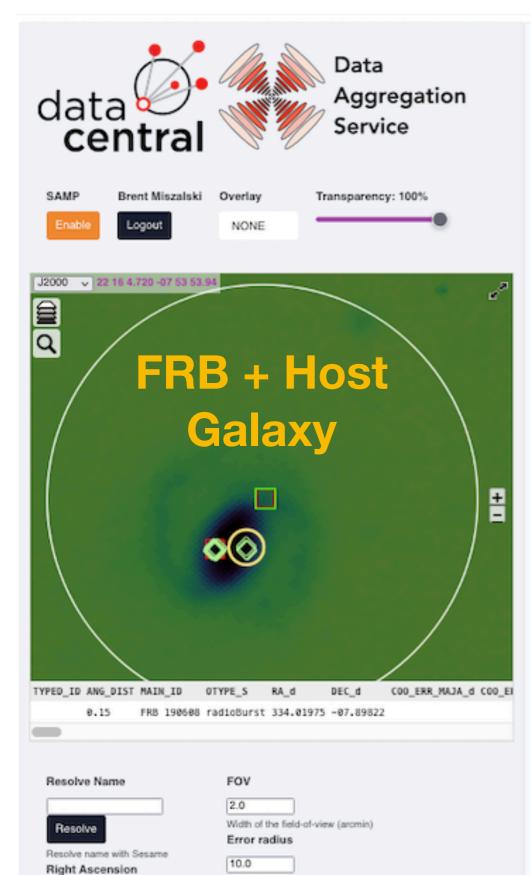


pandas astropy VOTable parsing and filtering

Fast from the ground up: Asynchronous with caching

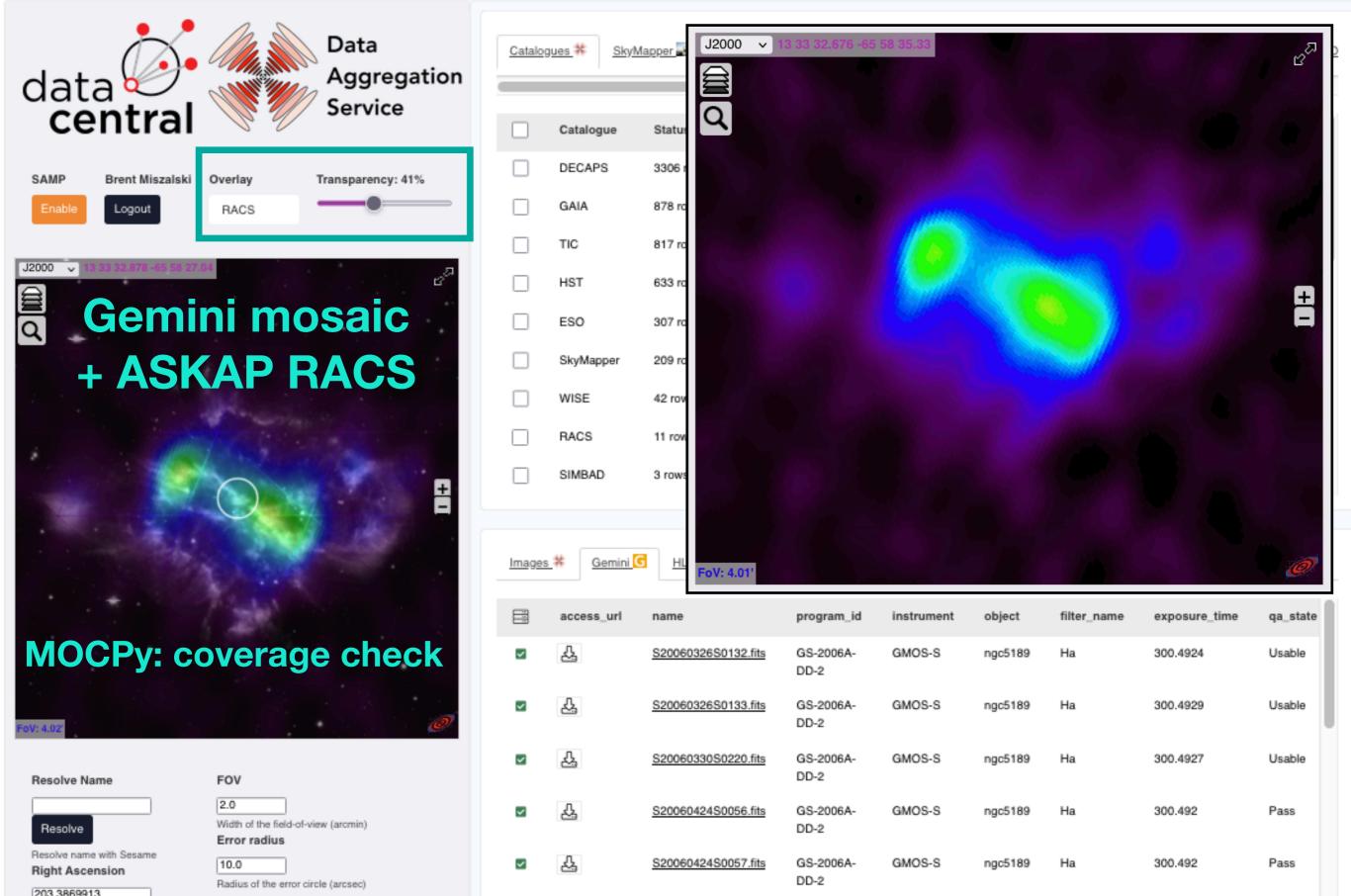
- **Simultaneous queries** of **multiple** services (e.g. TAP, SIA, SSA) from asynchronous functions launched as background tasks (asyncio)
- Tasks may also spawn background tasks (e.g. downloading multiple images) OR more complex pipelines (e.g. creation of mosaics)
- Task results sent to browser via websocket messages. Contain tabular data (json) or image location on disk.
- Browser listens for messages: Results added to page when received.
- VOTable/ASCII query results cached (redis) and converted to pandas dataframes (astropy) for internal handling and filtering
- FITS files are converted to HiPS format using Aladin and cached on disk (location cached with redis)

Custom Redshift Aggregator



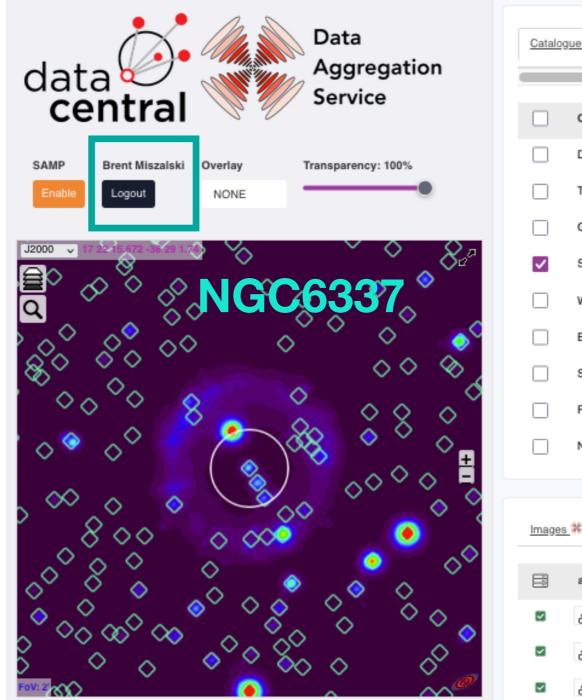
atalogu	105 * <u>Si</u>	Viz kyMapper 🛋			gues:		c.redshi	f
_							_	
_r_arcs	ec RA		DEC	z	ZCOLS	vizier_cat	vizier_cat_description	
2.3436	334	.02	-7.89884	0.1157	z	VII/281/glade2	GLADE v2.3 catalog (Dalya+, 2018)	
2.4804	334	.02	-7.89888	0.1178	z	<u>J/ApJS</u> /223/20/spec	SDSS-DR8 galaxies classified by WND- CHARM (Kuminski+, 2016)	
2.5056	334	.019989	-7.89889	0.1176	z	J/MNRAS /455/2440 /catalog	Gal. 2D phot. decompositions in r, g & i bands (Meert+, 2016)	
2.5056	334	.019989	-7.89889	0.1176	z	J/MNRAS /446/3943 /catalog	galaxies 2D phot. decompositions in SDSS-DR7 (Meert+, 2015)	
Images_	* <u>PS1</u>	<mark>ວ <u>LS</u> L</mark>	<u>SkyMapp</u>	er E FORS2				
	access_ur	1 Object	Filter	Exptime	Date	ProgID	RA	DE
S	4	FRB 19060 HOST	G_HIGH	499.989990234	4 2020-09-21T02:4	1:36.975 <u>105.204W.00</u>	0 <u>1</u> 334.02083305	-7.8
2	쑈	FRB 190600 HOST	G_HIGH	499.977996826	2 2020-09-21T02:5	0:35.321 <u>105.204W.00</u>	<u>)1</u> 334.01803194 ·	-7.8
	쑈	FRB 190600 HOST	G_HIGH	499.981994628	9 2020-09-21T02:5	9:32.606 <u>105.204W.00</u>	0 <u>1</u> 334.01802694 ·	-7.9
5	4	FRB	G_HIGH	499.990997314	5 2020-09-21T03:0	8:31.312 <u>105.204W.00</u>	01 334.023635	-7.9

Multi-wavelength HiPS overlays



Authenticated data access

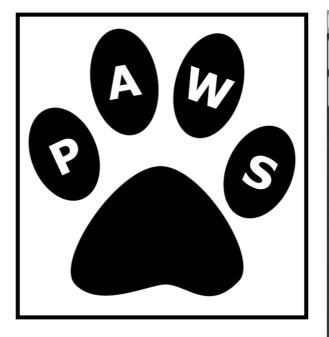
- Data Central's CAS server for authentication (single sign-on)
- Each user may belong to a group (e.g. SkyMapper)
- Seamless data access control based on group membership
- Select SkyMapper TAP and SIA servers: DR3 (Australian) or DR2 (everyone else)
- Lots of other possibilities



Catalo	gues <mark>#</mark> <u>Sky</u>	Mapper 🗾	SIMBAD /	DUST.	
	Catalogue	Status	SAMP	Excel	
	DECAPS	1168 rows	Ś	산	
	TIC	414 rows	C)	샾	
	GAIA	358 rows	C)	샾	
\checkmark	SkyMapper	245 rows		윤	
	WISE	52 rows	C	윤	
	ESO	13 rows	R	윤	
	SIMBAD	2 rows	R	£	
	RACS	2 rows	R	윤	
	NVSS	1 row	Cy)	쑈	
Images	s 🗱 SkyMag	oper 🛃 Ger	nini <mark>C</mark>	DECAPS 💞	
	access_url	images	Filter	dclabel	
	윤	Ľ	g	SkyMapper: g	
2	윤	Ľ	r	SkyMapper: r	
V	윤	Ľ	i	SkyMapper: i	

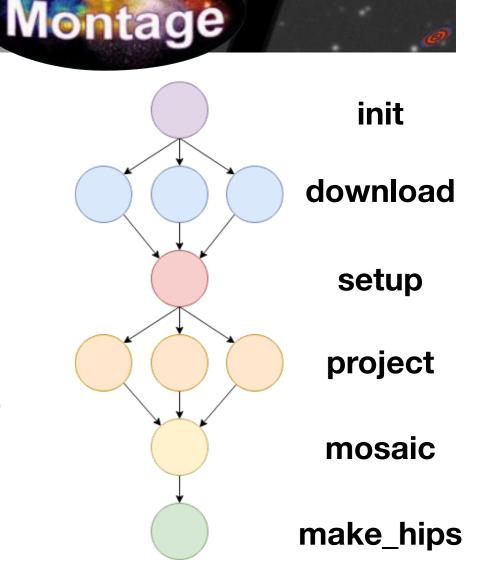
SkyMapper DR3: catalogue (TAP) + mosaic (SIA)

Pipeline As a Web Service (PAWS) Technology Demonstrator



NGC 6326 FORS2 [OII] mosaic

- Async DAS architecture: ideal testbed for PAWS
- Use celery to prepare and schedule complex workflows, e.g. image mosaics (IPAC MontagePy)
- Observatory archive queries input for:
 - Mosaic 10x10 arcmin SkyMapper images (SIA queries in all uvgriz filters)
 - Mosaic GMOS images (Gemini archive json API)
 - Mosaic VLT FORS2 images (ESO archive TAP query)



Browsing NGC6302 VLT FORS2 mosaics (R, I, V and Halpha)



	GAIA	505 rows	Ċ	샾	샩	윤		
	TIC	470 rows	C	£	싪	윤		
	SkyMapper	166 rows		샾	셨	윤		
	WISE	25 rows	Ċ	샾	셨	윤		
	SIMBAD	5 rows	Ś	샾	싰	윤		
	RACS	4 rows	Ľ	샾	싪	샾		
<u>Images</u>	* SkyMapp	er 🗾 🔣	A S FORS2	DECAPS	9			
	access_url	Object	Filter	Exptime	Dat	e	ProgID	F
V	윤	NGC 6302	G_HIGH	0.2579999864	201	8-04-21T04:05:34.553	60.A-9801(C)	2
	샾	NGC 6302	R_SPECIAL	5.0040001869 2018-		8-04-21T04:07:26.115	60.A-9801(C)	14
V	윤	NGC 6302	L_BESS	4.9809999466	201	8-04-21T04:08:39.213	60.A-9801(C)	2
V	윤	NGC 6302	V_HIGH	5.006000042	201	8-04-21T04:09:27.728	60.A-9801(C)	4
S	윤	NGC 6302	H_ALPHA	8.0010004044	201	8-04-21T04:10:26.874	60.A-9801(C)	2
7	섚	NGC 6302	G_HIGH	5.006000042	201	8-04-21T04:12:16.525	60.A-9801(C)	A A

FOV

2.0

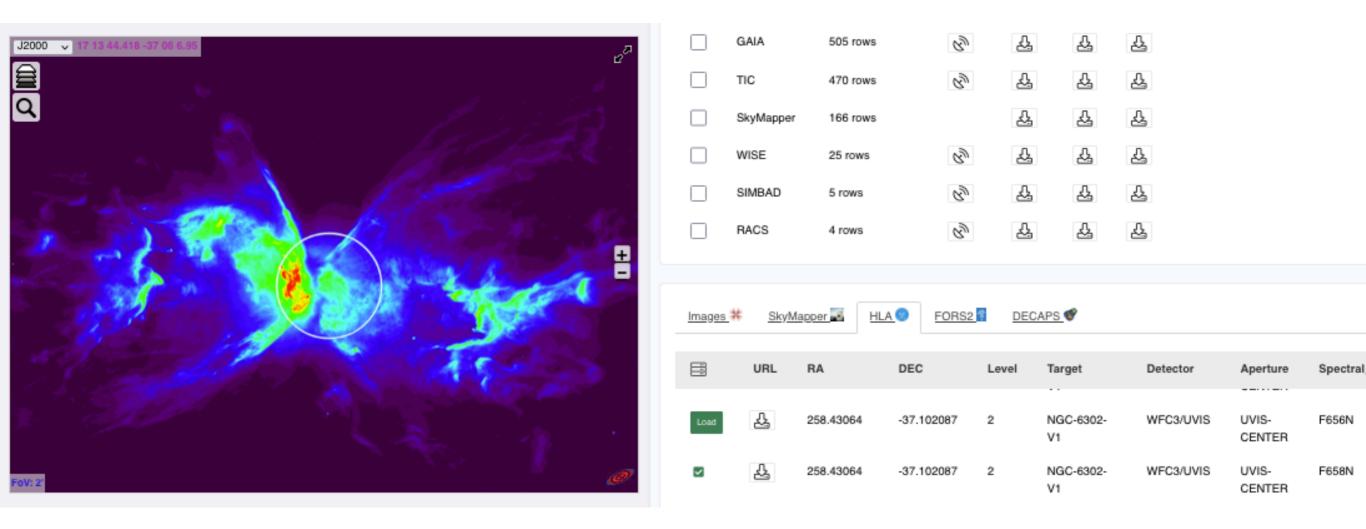
Width of the field-of-view (arcmin) Error radius

10.0

Radius of the error circle (arcsec)

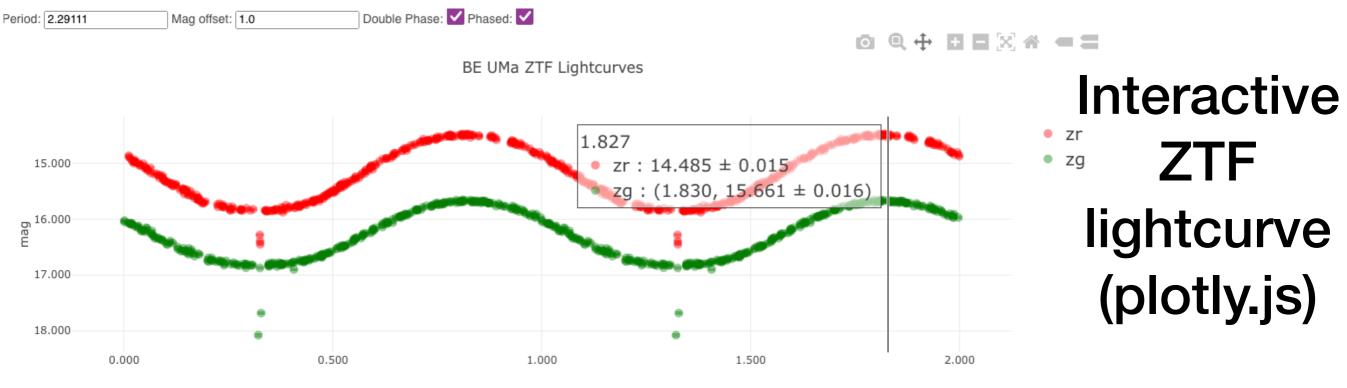
Submit

NGC6302 F658N HST WFPC3

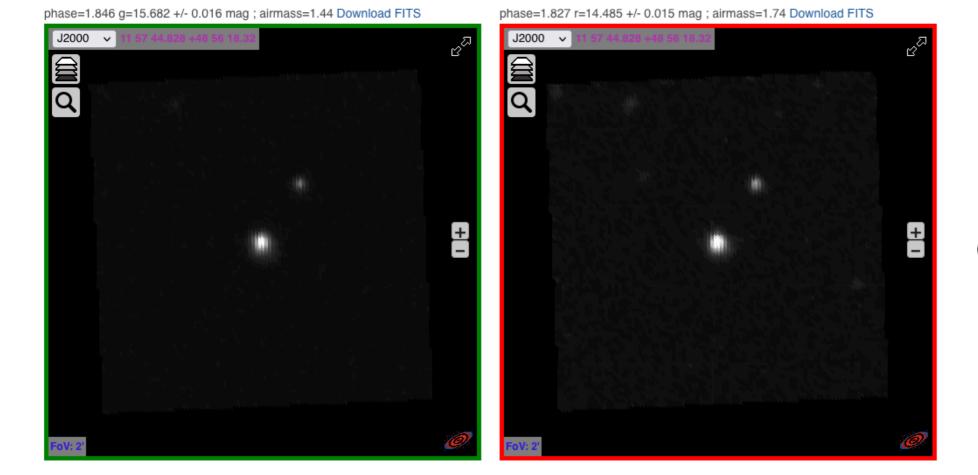


Large image downloads (~500 MB) triggered manually + progress bar (Hubble Legacy Archive SIA)

Todo: add support for time series data



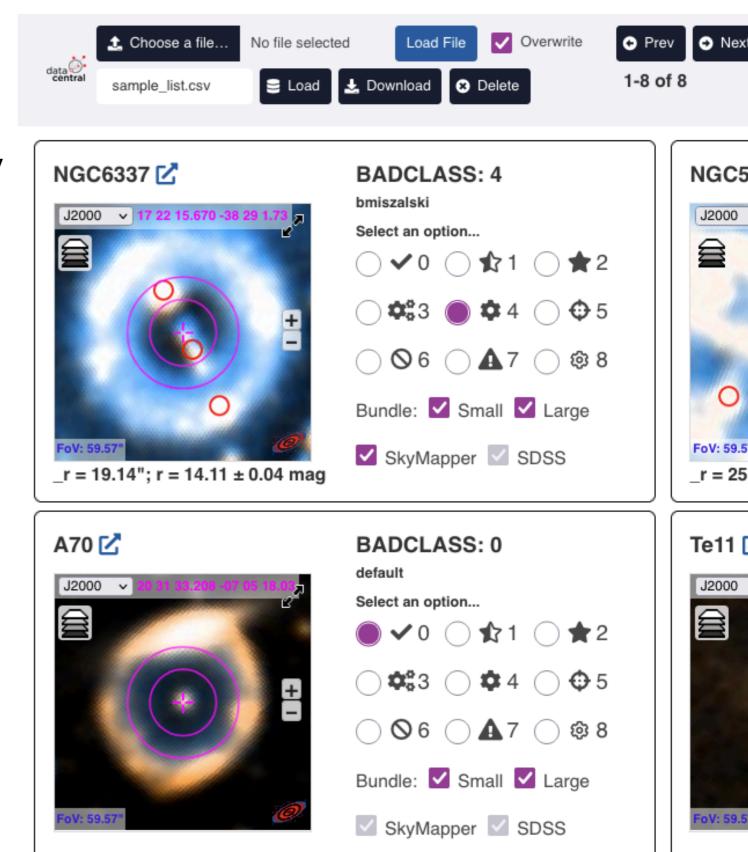
Phase



zg and zr individual epoch images at cursor position

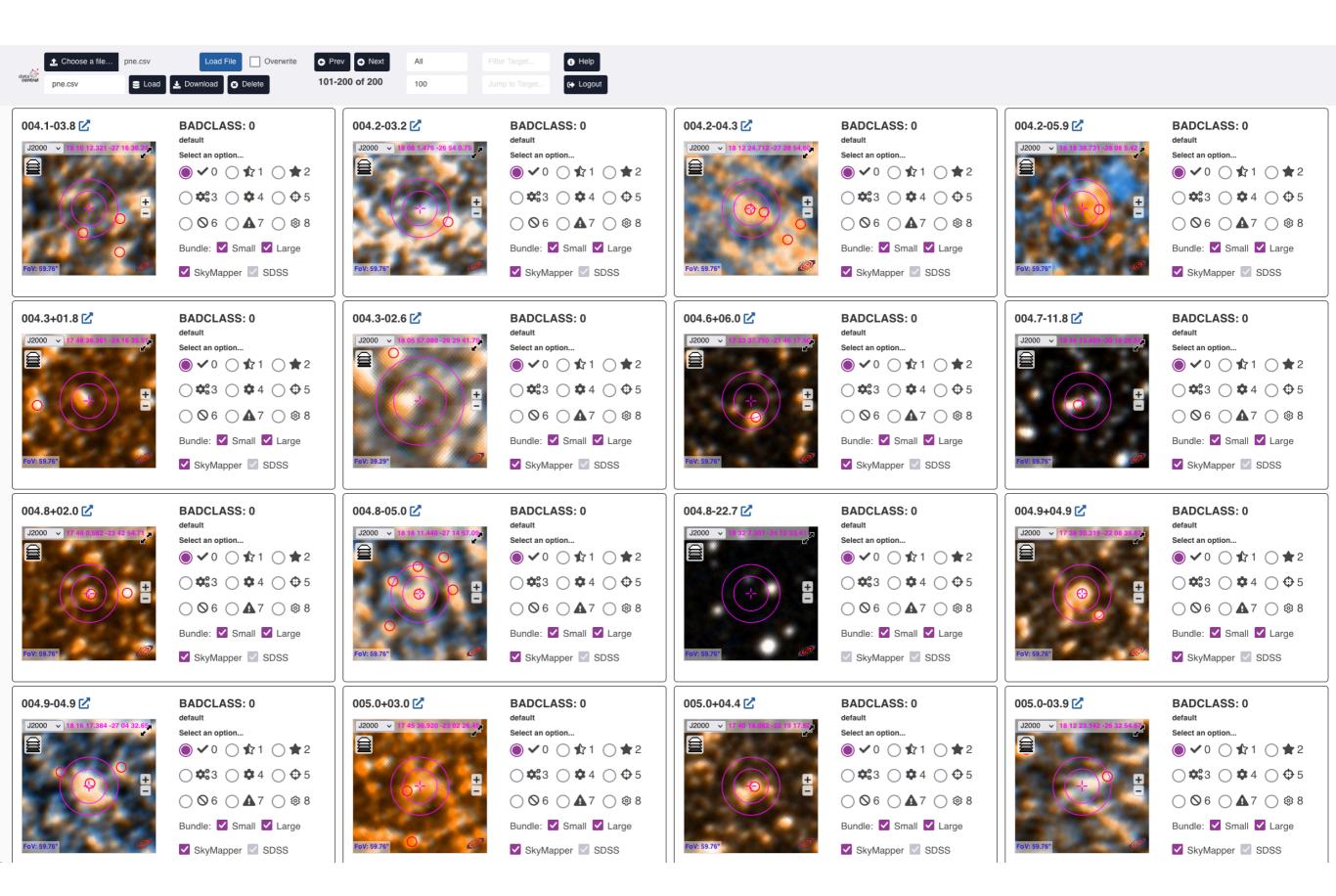
Target Visualisation and Classification

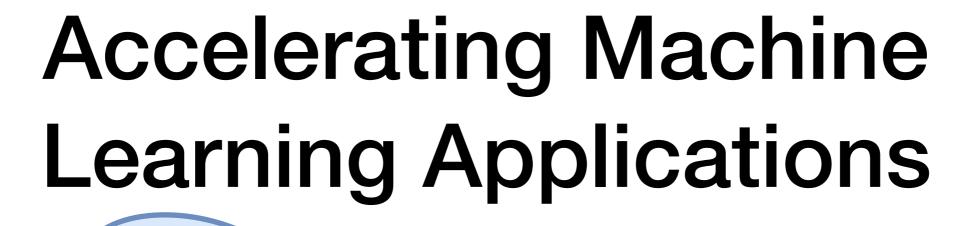
- Target visualisation and classification app: Repurposed DAS for upcoming Hector Survey
- Target lists: Import CSV into MongoDB. Classify targets (keyboard shortcuts) and export results.
- Aladin Lite: Images and catalogues (loaded asynchronously) per target. Display instrument footprints.
- Filter catalogues: Radius and magnitude cuts. Mouseover to view.

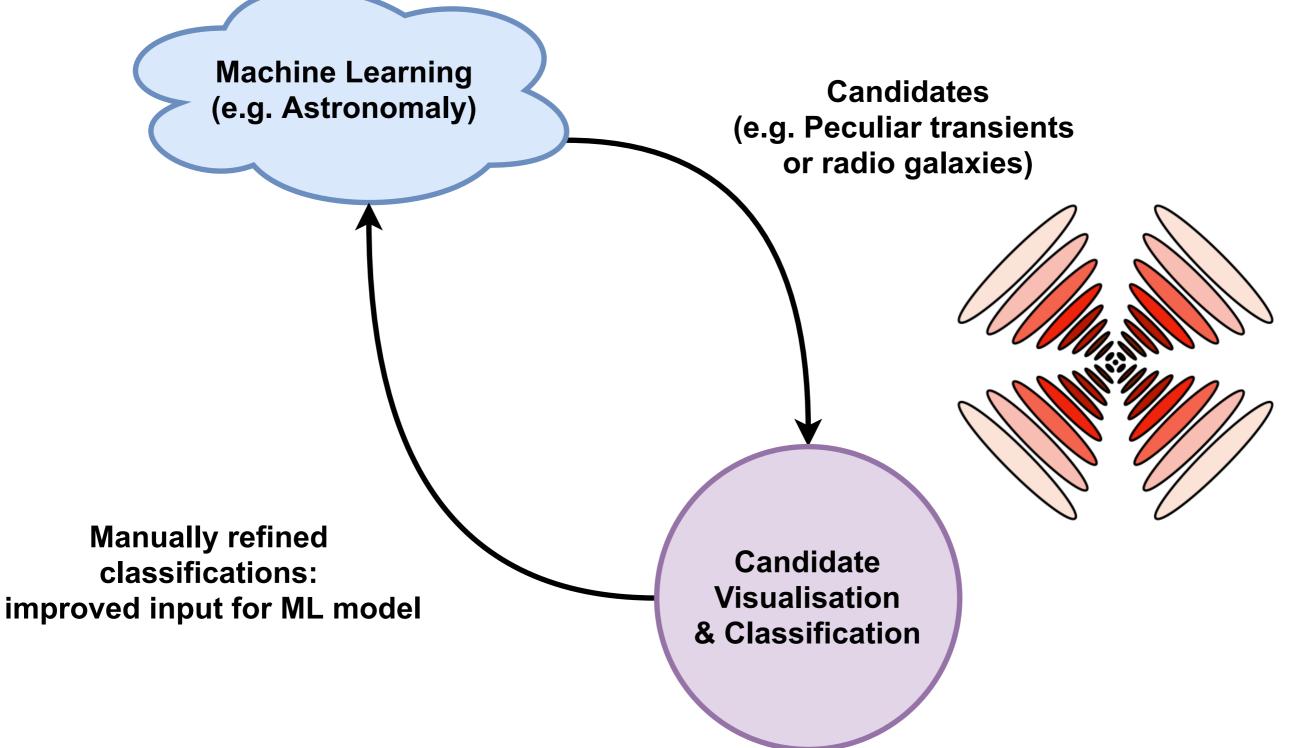


More details: https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpNov2021Apps#Applications_2

Up to a few hundred targets per page

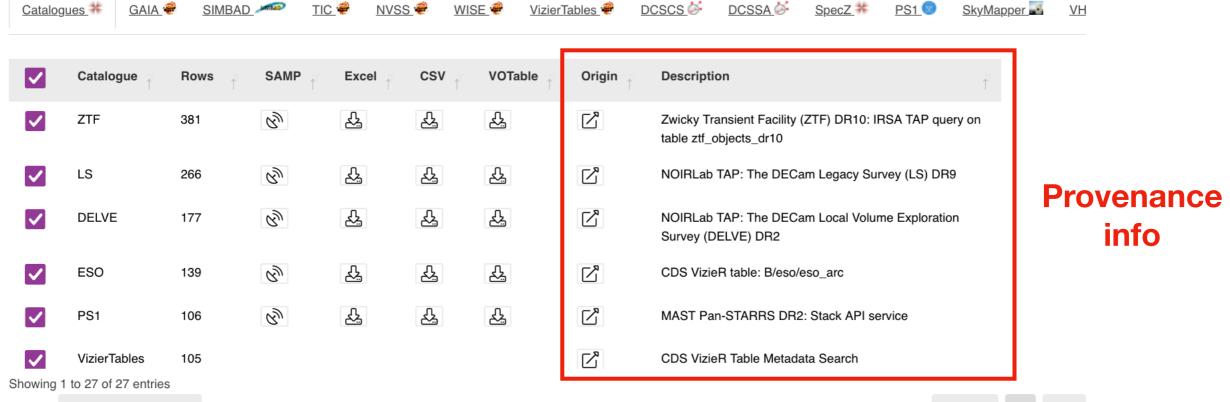






Recent additions to DAS

- GET parameters: allows users to toggle on/off individual data sources
- More detailed provenance information (see below)
- Added VHS catalogue (hosted by Noirlab) and images (VISTA Science Archive)
- Want to see more data added to DAS? Let me know!

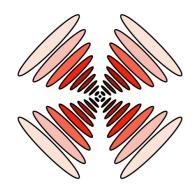


Customising DAS: Examples of GET parameter usage

- Load only image sources: <u>https://das.datacentral.org.au/das?</u> <u>CAT=0</u>
- Load only image sources and the Acker catalogue of Planetary Nebulae: https://das.datacentral.org.au/das?VIZ=acker,V/84
- Load the DEFAULT catalogue sources and the Acker catalogue: https://das.datacentral.org.au/das? CAT=DEFAULT&VIZ=acker,V/84
- Load only the FINK and ZTF catalogues: <u>https://</u> <u>das.datacentral.org.au/das?</u> <u>CAT=ZTF&CAT=FINK&IMG=0</u>

Future DAS plans

- Add API endpoints for different functionality:
 - Fast cross-matching against multiple catalogues
 - e.g. cone search match with WISE, GALEX, SDSS, etc all in one go
 - Redshift aggregator
- Other ideas:
 - Real-time table displaying transient events + triggered cross-matching and data retrieval as soon as events come in
 - Interactive lightcurves and image display for time series data
 - Interactive viewing of spectra
 - Include ESO PAWS for on demand data reduction (e.g. FORS, UVES, X-shooter)



Conclusions data **Central**



- DAS is a powerful and extensible platform for asynchronous data discovery and visualisation
- DAS as a platform for PAWS: Reduction of spectroscopic and imaging data sourced directly from multiple observatory archives. Visualisation and download of reduced data products without having to download raw data OR install pipelines.
- Development version of Target Visualisation and Classification app in place to view many targets using same asynchronous infrastructure as DAS underneath.
- Very interested in supporting transient science: opening up AAT archive: VOEvents and PAWS