Data Central’s Data Aggregation Service

Casting a wide net in the ocean of astronomical data

Dr Brent Miszalski
brent.miszalski@mq.edu.au
• 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
- VizieR
- IRSA/Firefly
- MAST Portal
- Aladin
- 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

- **SIA**
- **API**
- **TAP**
- **Archive**
- **SSA**

**DAS**

Images and mosaics

Catalogues: tables, overlays & download

Other: datacubes, spectra, time series, etc.

Aladin Lite
DAS interface

Catalogues (Vizier, TAP, SSA, API)

Aladin Lite

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
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.
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.