ESO Science Archive Facility

VO-protocol based access to the ESO Science Archive Facility (release 1)

Micol, Forchi', Fourniol, Hainaut, Lange, Mubashir, Retzlaff, Romaniello, Sisodia, Stellert, Stoehr, Vera, Zampieri
Archive Services Project: Top Level

- Interactive Access
  - Web-based user interface
  - Explore by scientific characteristics of the data
  - In a very interactive graphical way
  - Evaluation of usefulness of the data before downloading (footprints, previews)

- Programmatic and Tool Access
  - Allow complex queries
  - Allow scripted access, to automate repetitive tasks
  - Allow discovery and access via astronomical tools
  - Discoverability via Virtual Observatory tools and services
A 3 year project

- Staged approach

Balance between: **what** (new capabilities) *vs when* (to deploy them)

First release (Q1 2018): significant step forward
Next releases (2019, 2020): improving, extending
Programmatic and Tool Access: Release 1

Based on VO Protocols:

- Tabular Access Protocol (TAP) [no UPLOAD]
- Astronomical Data Query Language (ADQL)
  - Complex Queries
- Simple Spectral Access Protocol (SSA)
- ObsCore (based on ESO Science Data Product Standard)
- DataLink: to access the data and their related files (ancillary files, progenitors, derivations, previews, etc.)
- Other tables (non standard) available in TAP:
  - Raw data
  - Provenance (to link data products and raw data)
  - Ambient measurements
Technology stack

Two TAP Servers for two different DBMSes:

- **MS SQLServer**:
  - Small tables (millions of records), full spatial capabilities

- **Sybase IQ**:
  - Big catalogs (billions of records), limited spatial capabilities;
  - custom solution: HTM index, direction cosines;
  - column-based, petabyte scale, relational database.

**TAPLIB (G. Mantelet)**

- ESO: ADQL translator to support SQLServer

**TAPLINT (M. Taylor)**

**ESO Simple Spectral Access**:

- Spring Boot application developed in house, open source: [https://github.com/vforchi/SSAPServer](https://github.com/vforchi/SSAPServer)
  - Implements SSA in 2 different ways:
  - Translating user’s input into ADQL, using then TAP
  - Translating user’s input to query directly the database
Tool Access: ObsCore query (Aladin)
Aladin displays ObsCore and DataLink
DataLink

That was a spectrum… but for images?

ESO previews of individual images: HiPSes

■ How to serve an HiPS from DataLink?
■ HiPS mime-type needed!
Complex Queries

Colour-magnitude diagram

- Required: images covering the same sky region, but taken with different filters

- ObsCore allows such query, though with some limitations

- Solved adding an ESO User Defined Function

- Tools get confused by this query
Colour-magnitude query

```sql
SELECT *
FROM
(select * FROM ivoa.Obscore WHERE dataproduct_type='image'
AND INTERSECTS(CIRCLE('ICRS',266.42,-29.0,5),s_region) = 1
AND em_min < 1.265E-6 AND em_max > 1.265E-6
) J,
(select * FROM ivoa.Obscore WHERE dataproduct_type='image'
AND INTERSECTS(CIRCLE('ICRS',266.42,-29.0,5),s_region) = 1
AND em_min < 1.66E-6 AND em_max > 1.66E-6
) H
WHERE INTERSECTS(J.s_region, H.s_region)=1 and
ESO_INTERSECTION(J.s_region,H.s_region) > 1.79

Tools will not recognise the output as ObsCore!
```
Without constraint on extent of intersection
New standard functions?

- **Colour-magnitude diagram**
  
  ESO_INTERSECTION(J.s_region, H.s_region) > 1.79
  
  Should actually be: AREA(INTERSECTION(a,b)) > 1.79
  
  whereby **INTERSECTION** would return the polygon of the intersection.
  
  It seems generic enough… *Standardizing it?*

- **Find the ambient conditions of the raw data that participated in the generation of a data product**
  
  - ADQL joining ObsCore and Ambient tables, via Provenance and timestamp (start time of observation)
  
  **ESO_DATEADD_SEC(-360, start_date)**
6. **100 best spatially-resolved J, H, or Ks images intersecting a 5 deg cone around the Galactic Centre**

Find the imaging products (dataproduct_type='image') intersecting a cone of 5 degrees around the Galactic Centre (266.42,-29.0) extending the spectral axis over the central wavelengths of either J, H, or Ks photometric bands. Sort results by the spatial resolution (from best to worse), and return only the first such 100 results.

```sql
SELECT TOP 100 instrument_name, em_min, em_max, s_resolution, distance(point(",s_ra,s_dec),point(",266.42,-29.0)) as dist_from_GC
FROM ivoa.ObsCore
WHERE dataproduct_type='image'
AND INTERSECTS(CIRCLE("ICRS",266.42,-29.0,5),s_region) = 1 -- intersecting (aka overlapping) a 5 deg cone around the Galactic Centre
AND (em_min < 1.25E-6 AND em_max > 1.25E-6) -- Contains lambda(J) or
  OR (em_min < 1.65E-6 AND em_max > 1.65E-6) -- Contains lambda(H) or
  OR (em_min < 2.2E-6 AND em_max > 2.2E-6)) -- Contains lambda(Ks)
ORDER BY 4 -- 4 is the position of the s_resolution within the SELECT list
```

---

**List of tables involved in the query:** ivoa.ObsCore

**URL query string:**

```url
REQUEST=doQuery&LANG=ADQL&MAXREC=200&FORMAT=text&QUERY=SELECT%20TOP%201%2C%20em_min%2C%20em_max%2C%20s_resolution%2C%20distance(point("%2Cs_ra%2Cs_dec"),point("%2C266.42%2C-29.0"))%20as%20dist_from_GC%20FROM%20ivoa.ObsCore%20WHERE%20dataproduct_type%3D'image'CIRCLE("ICRS%2C266.42%2C-29.0%2C5"),s_region)%20AND%20((em_min%20%6%20AND%20em_max%20%3E%201.25E-6)%20OR%20((em_min%20%3C%201.65E-6%20AND%20em_max%20%3E%202E-6%20OR%20em_min%20%3C%202E-6%20AND%20em_max%20%3E%202E-6))%20AND%20(s_region)%20AND%20(distance(point("%2C266.42%2C-29.0"),point(",s_ra,s_dec"))%20%3D%207.5E-6))
```

---

**Execute this example**  **Edit this example**  **Reset this example**
Challenges & Improvements

Challenges encountered during development

Suggesting possible improvements to VO standards
Spatial Support

- ADQL/DALI seems crafted around pgsphere
  - ADQL and DALI geometrical data types:
    - Point, Circle, Polygon, Interval
  - MS SQLServer extended list of geography data types:
    - Point, LineString, CircularString, CompoundCurve, Polygon, CurvePolygon, MultiPoint, MultiLineString, MultiPolygon, GeometryCollection

- ADQL:
  - DISTANCE(point, point)
  - AREA(circle | polygon)

- MS SQLServer:
  - DISTANCE(geography, geography)
  - AREA(UNION(POLYGON … POLYGON …))
Dropping REGION? Please don’t!
No REGION, No Multi-chip!

```xml
<FIELD
    ararraysize="*"
datatype="char"
    name="s_region"
    ucd="pos.outline;obs.field"
    xtype="adql:REGION"/>

<TD>
UNION J2000 (POLYGON J2000 266.3870080584 
-30.2063917509 266.3868600144 -30.1303732103 266.3006456891 
-30.1304481436 266.3007272851 -30.2064668154 266.3870080584 
-30.2063917509 POLYGON J2000 266.31158900549997 
-30.2063857788 266.3115717899 -30.1303750718 266.2254010285 
-30.1303396166 266.2253518365 -30.2063502569 
266.31158900549997 -30.2063857788 POLYGON J2000 266.3866300575 
-30.1415730026 266.3864527576 -30.0655893992 
266.3003345775 -30.0656892393 266.3004455775 -30.1416729925 
266.3866300575 -30.1415730026 POLYGON J2000 266.31135866479997 
-30.1415995932 266.311218201 
-30.0655922248 266.2250734718 -30.0656609103 266.2251475943 
-30.1416682921 266.31135866479997 -30.1415995932…)
</TD>
```
ESO implements non (yet?) standard attributes:

- `n_obs` Number of observations contributing to product
- `filter_name` To help readability
- `gal_lat` To allow searches on galactic coords
- `gal_lon`
- `ecl_lat` To allow searches on ecliptic coords
- `ecl_lon`
- `local_id` To allow joins with local tables
- `snr` Signal to Noise Ratio (spectra, cubes)
- `abmaglim` Limiting magnitude (images)

A part from last two, the rest seems generic enough to permit standardisation…?
ObsCore defines:
UCD(obs_publisher_did):
  meta.ref.uri;meta.curation

UCD1+ defines:
  P | meta.curation | Identity of man/organization responsible for the data

(votlint complains at build time)

- **ObsCore (or UCD1+) needs to be amended.**
Next Releases

- Lessons learned: improvements
- Raw+products under same data model (ObsCore)
- Support for Cutouts (images, cubes, etc.), SODA (VO Protocol)
- Expose more tables in TAP (scheduling information, calibration master files)
- Upgrading tap and adql to latest version of the standard (if beneficial)
- TAP Upload functionality
Conclusions (1/2)

- First year of Archive Services Project
- WebUI: Novel user’s experience, Aggregations, Previews (jpeg, pdf, HiPS), HiPS for collections
  - Angular, elasticsearch, plugin (ignacio), aladin lite (CDS), aladin hips generator (CDS, CADC)
- Programmatic: TAP, ObsCore, SSA, DataLink
  - VO-ware, taplib (Mantelet), adaptation for SQLServer, ssa (homegrown, github)
- Tools: topcat, aladin, splat-vo, etc.
- Some challenges, some proposals
Items for discussion

- REGION needed to support complex footprints
- Utility functions? Users’ creativity bounces against the limited ADQL support for utility functions:
  - string functions (*substring*, *char_length*, etc)
  - date functions (*dateadd*, etc.)
  - spatial functions (compute *intersection* of geographies)
- Additional standard ObsCore attributes?
- Complex queries not well supported by tools
- Mime type for HiPSes
- Wrong UCD for obs_publisher_did