Data publishing at ESO: How to publish protected astronomical data via VO protocols?

Alberto Micol, Vincenzo Forchì

30 January 2019

Asterics

DADI/OBELICS A&A face-to-face Meeting

Osservatorio Astronomico di Trieste

ESO Science Archive made better

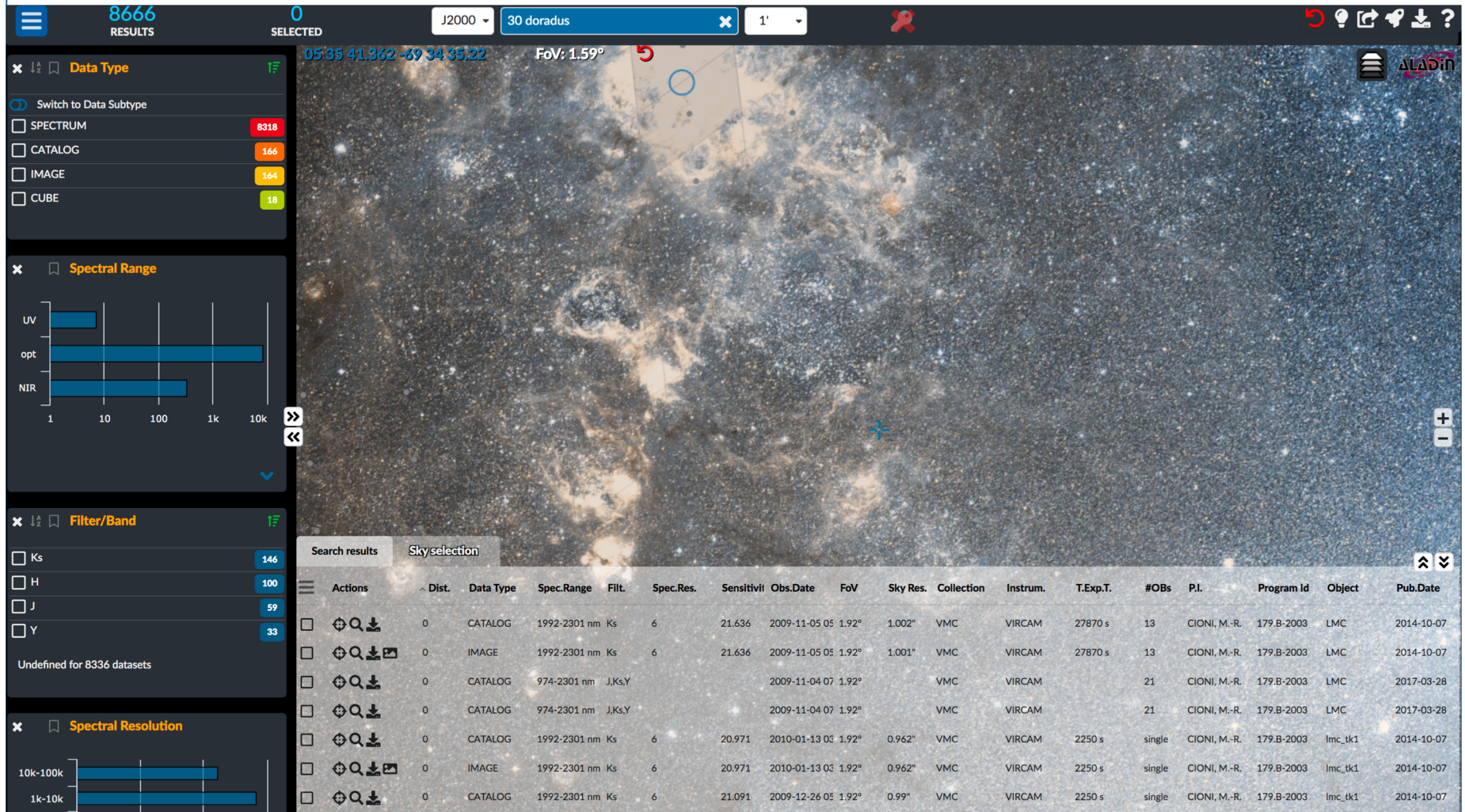In 2018 public with:

Web application (we call it: *science portal*)
   http://archive.eso.org/scienceportal

Programmatic and Tool Access:
   http://archive.eso.org/programmatic

http://archive.eso.org/scienceportal

Alberto Micol, Archive Science Group, Data Management and Operations Division, ESO

VO in the Science Portal

NoSQL database (read: Non-ADQL queries)

VO components used:

- **SAMP** (to export **ObsCore** votable)

- **STC-S** footprints

- **HiPS**

- **Datalink** used internally to expand a dataset
  into its components (science file + ancillary files)

- VO-aware tool used as a component:

- **Aladin Lite / CDS**

Programmatic and Tool access

Release 1: direct database & VO access

Two TAP servers:

- ***tap_obs***: TAP server for observational data

  Powerful spatial queries (ADQL on SQLServer) + ObsCore

- ***tap_cat***: TAP for catalogs provided by ESO observers

  Limited spatial queries (cone searches),

  Tables with more than 4E9 records (ADQL on SYBASE IQ)

- **github** *taplib* (ADQL, TAP, and UWS) by Grégory Mantelet (ARI)

  ADQL for SQLServer and SYBASE IQ by V.Forchì (ESO)

SSAP server to easily access the 1d spectra

- **github** https://github.com/vforchi/SSAPServer (ESO)

DataLink to easily find and access public files: ancillary files, provenance and derived files, previews, etc.

Programmatic and Tool access

Next releases: 2019, 2020

Proposal (subject to approval in February):

- SIAP v2
- Cut-outs + SODA on public files (2019)
- DataLink for calibration files (ESO's calSelector)
- Support downloading of proprietary data:
  - *Direct* access to proprietary data (*direct:* no archive requests)
  - DataLink for proprietary data
- Support for special data:
  - "*special data*" := both metadata and data are protected
- TAP UPLOAD
  - DB space for users
  - authentication

A&A Use Case 1 (1)

Data are of 3 types:

1. public := Public metadata, public data
2. proprietary := Public metadata, proprietary data
3. special := Proprietary metadata, proprietary data

Via a Python script, a user submits a TAP query to find data of certain characteristics (e.g.: ObsCore query)

A&A Use Case 1 (2)

TAP must return:

- ■ If unauthenticated: all public metadata records
- ■ If authenticated: all metadata records the user has access to.

How to achieve this?

- ■ Authentication: yesterday's discussion: timeline?
  - ➢ for now: TAP 1.1, correct?
  - ➢ Read /capabilities to find supported TAP securityMethods
- ■ Authorisation: Injecting into the ADQL the necessary authorisation logic?
  - ➢ Ok for simple queries
  - ➢ but for complex ones (subqueries, joins, etc)?

A&A Use case 1 (3)

■ Each of the TAP returned records provides a datalink

  ➢ Datalink for *public* or *proprietary* data:

    • no authentication is requested to execute it, nor to get the results

  ➢ Datalink for *special* data:

    • Authentication is necessary to execute it;

    • results are sent back only to authorised users

How to achieve it?

■ Datalink for public and proprietary data on a public endpoint (no auth)

■ Datalink for *special* data on a different endpoint which requires authentication?

■ Do I need to describe this in the DataLink /capabilities

■ Can Pat's "one URL to urle them all" still be achieved?

A&A Use case 1 (4)

■ Datalink resulting VOTable contains accessURLs to either public, proprietary, or special data files

- ➤ **Permanent links**: **The URL of a file under proprietary period should not need to change once the file becomes public**: that is, if the user repeats the query after a proprietary file has emerged from its proprietary period the URL is still the same (permanent URL). ESO supports this via redirection: datalink URL for downloading is the same URL of the datalink, with the addition of the token &eso_download=file, e.g.:

- ➤ http://archive.eso.org/datalink/links?ID=ivo://eso.org/ID?ADP.2018-06-06T09:07:35.844&eso_download=file

- ➤ http://archive.eso.org/datalink/links?ID=ivo://eso.org/ID?ADP.2018-06-06T09:07:35.844&eso_download=preview

Datalink, when eso_download provided, redirects to the system that provides files, or to the system that provides previews

Therefore, not only each file has a permanent link, but also its datalink url

A&A Use Case 2: TAP UPLOAD

User, after authentication, TAP-uploads her own table(s)

■ How to support private tables in TAP?

➤ Is there a standard way to do that?

➤ E.g.:

- Adding owner column in TAP_SCHEMA.tables, and
- Matching owner with credential of authenticated user

?

➤ TAPLIB does not seem to support private tables, any other library available?