DOI: 10.18727/0722-6691/5073

Enhanced Data Discovery Services for the ESO Science Archive

Martino Romaniello[1]
Stefano Zampieri[1]
Nausicaa Delmotte[1]
Vincenzo Forchì[1]
Olivier Hainaut[1]
Alberto Micol[1]
Jörg Retzlaff[1]
Ignacio Vera[1]
Nathalie Fourniol[1]
Mubashir Ahmed Khan[1]
Uwe Lange[1]
Devendra Sisodia[2]
Malgorzata Stellert[3]
Felix Stoehr[1]
Magda Arnaboldi[1]
Chiara Spiniello[1,4]
Laura Mascetti[5]
Michael Fritz Sterzik[1]

[1] ESO
[2] Pactum Limited, London, UK
[3] TEKOM Industrielle Systemtechnik GmbH, Gautin, Germany
[4] INAF–Osservatorio Astronomico di Capodimonte, Naples, Italy
[5] Terma GmbH, Darmstadt, Germany

The archive of the La Silla Paranal Observatory is a powerful scientific resource for the ESO astronomical community. It stores both the raw data generated by all ESO instruments and selected processed data. We present new capabilities and user services that have recently been developed in order to enhance data discovery and usage in the face of the increasing volume and complexity of the archive holdings. Future plans to extend the new services to processed data from the Atacama Large Millimeter/submillimeter Array (ALMA) are also discussed.

Background and motivation

The ESO Science Archive[1] began operating in its current form in 1998, a few months ahead of the start of science operations of the Very Large Telescope, the VLT (see Pirenne et al., 1998). It is the operational, technical and science data archive of the La Silla Paranal Observatory (LPO). As such, it stores all of the raw data, including the ambient weather conditions, from the LPO, i.e., the telescopes at Paranal and La Silla, and the Atacama Pathfinder Experiment (APEX) antenna at Chajnantor. Also available through the archive are data from selected non-ESO instruments at La Silla, for example, the Gamma-Ray burst Optical/Near-infrared Detector (GROND), the Fibre-fed Extended Range Echelle Spectrograph (FEROS) and the Wide Field Imager (WFI). It also includes raw data for UKIDSS, the infrared deep sky survey using the wide-field camera WFCAM at the United Kingdom Infrared Telescope (UKIRT) in Hawaii. In addition, ESO hosts and operates the European copy of the ALMA Science Archive (Stoehr et al., 2017). The integration of archive services for LPO and ALMA data is discussed here.

Over the years, the archive has steadily grown into a powerful scientific resource for the ESO astronomical community. As processed data are routinely included, they can be used directly for scientific measurements, thus alleviating the need for users to carry out data processing on their own. An in-depth analysis of the archive usage and user community is presented in Romaniello et al. (2016).

The archive is populated with processed data through two channels. For the first channel, data-processing pipelines are run at ESO for selected instrument modes to generate products that are free from instrumental and atmospheric signatures and that have been calibrated. They cover virtually the entire data history of the corresponding instrument modes and are generated by automatic processing, with no knowledge of the associated science case. Checks are in place to identify quality issues with the products. The second channel involves data products that have been contributed by the community, which have been generated with processing schemes optimised to serve specific science cases. In most cases, they have already been used to derive published results (see Arnaboldi et al., 2014). These contributed datasets, which are validated via a joint effort between the providers and ESO before ingestion into the archive, often include advanced products like mosaiced images, source catalogues and spectra.

Thorough user documentation is also provided for all data releases, detailing the characteristics and limitations of each collection of processed data. This is particularly important, as it enables users to decide whether the data are suitable for their specific science goals. The systematic archive publication of such processed data dates back to 25 July 2011, with the first products produced by the Public Surveys conducted with the Visible and Infrared Survey Telescope for Astronomy (VISTA) infrared camera VIRCAM (Arnaboldi & Retzlaff, 2011). Processed data that were generated at ESO have been available since September 2013. An up-to-date overview of the released data is available online for contributed and pipeline processed data[2,3].

The number of users accessing processed data in the archive has grown steadily over time (Figure 1). At the current rate, an average of 2.2 new users are added every working day, with each user placing 11 requests on average. Given the growing popularity within the community, and the increasing volume and complexity of the archive holdings — and taking into account the recommendations of advisory bodies, such as the Users Committee, the Public Survey Panel, and the results of the community working group report on science data management (STC Report 580[4]) — it has become necessary to upgrade the ways in which users can access the ESO Science Archive in order to enhance data discovery and usage.

The trajectory of contemporary astronomical research increasingly involves multi-epoch, multi-messenger, multi-wavelength, multi-facility science, in which data are plentiful and varied. At the same time, data acquired from different facilities are becoming ever more complex, yet have to be combined in order to tackle increasingly challenging scientific questions. In this context, the role of science data archives is to lower the access threshold that separates researchers from acquiring and being able to work with the data that they are interested in. The average astronomer cannot be expected to be intimately familiar with the details of each archive and, even less, with the details of the instruments that produced the datasets concerned. The access layer to the data therefore has to be as self-explanatory as

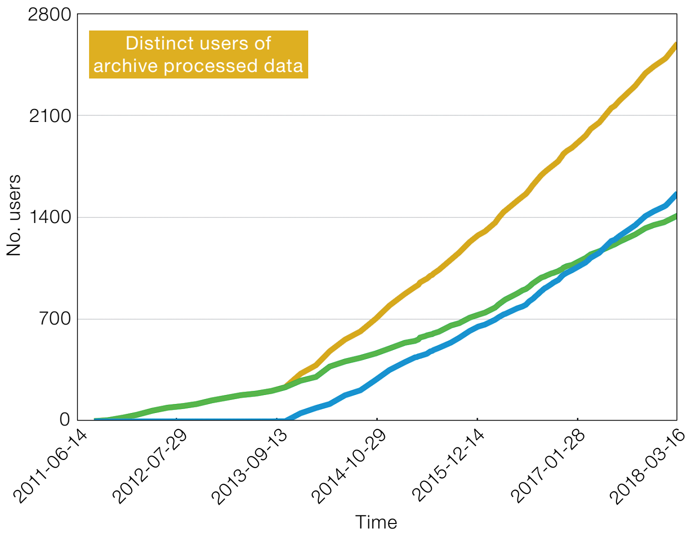


Figure 1. The growth of distinct users of processed data in the archive as a function of time. The green line displays access to products contributed by the community (deployed on 25 July 2011), the blue line shows access to products generated at ESO (deployed on 10 September 2013), and the orange line is for access to either type of products.

- Generated at ESO
- Contributed by the community
- Either one

possible so as to present the data in a user-friendly way, rather than couched in the technical terms used within the archive itself (for example, as calibrated fluxes and wavelengths, rather than detector counts, or an engineering description of the instrument setup).

Access points to the data

Different types of user interaction are supported:

- Interactive access: web pages through which users can browse and explore the assets with interactive, iterative queries. The results of such queries are presented in real time in various tabular and/or graphic forms, allowing an evaluation of the usefulness of the data which can then be selected for retrieval.
- Programmatic access: whereby users are able to formulate complex queries through their own programmes and scripts, obtain the list of matching assets, and retrieve them.
- Access by tools: whereby data are discovered, selected and accessed through tools normally developed by third parties, which are external to the web access channel. These tools often implement sophisticated data handling capabilities, such as TOPCAT[5] for catalogues, or Aladin[6] for images.

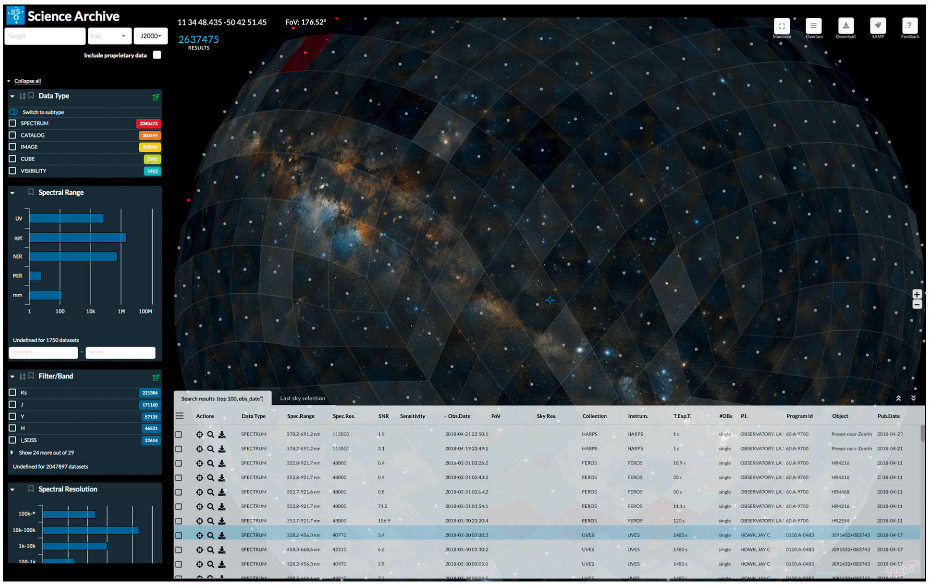
Furthermore, in order to fulfil the potential of multi-wavelength, multi-messenger science, ESO data need to be easily discoverable and handled so that they can be used together with datasets from other observatories and data centres. The natural framework for this is within the Virtual Observatory (VO); compatibility and interoperability with the VO is therefore a high-level goal for this project.

In this first release, processed data from the LPO are supported. Future plans include expanding the support to ALMA processed data and raw data from the LPO. It is planned that these new access points will gradually replace the current ones for La Silla Paranal data, while ALMA will keep maintaining a dedicated, separate access.

The ESO Archive Science Portal

The most immediate way to access the new archive services is through a web application, the ESO Archive Science Portal[7], using any recent version of the most popular internet browsers. A screenshot of its landing page is shown in Figure 2. The window is divided into three main sections: a sky view in which the content of the ESO archive is displayed together with background imagery such as the DSS; a table in which details of individual datasets are shown and from which further actions can be triggered, such as accessing previews; and a section in which query constraints can be specified, by explicitly entering them and/or by selecting values or ranges of values arranged in facets. Query results can be sent to suitable external applications for

Figure 2. The landing web page of the ESO Archive Science Portal. The celestial sphere is colour-coded according to the types of ESO data contained at each position (the sky viewer is CDS' AladinLite[6], the web version of Aladin). The entire content of the archive is presented through aggregations of 17 parameters, which can also be used to enter query constraints, in addition to querying by object coordinate or name, as resolved by the CDS's Sesame service[8]. In order to serve different use cases, they are a combination of physical characteristics of the data (e.g., signal-to-noise ratio, sensitivity, spectral range covered, spectral resolution), the observational setup (e.g. filter name, exposure time) and the ESO observing process (e.g. PI name, Programme ID).



Data Type

Switch to subtype

□ SPECTRUM	2040473
□ CATALOG	300499
□ IMAGE	289095
□ CUBE	5995
□ VISIBILITY	1413

Undefined for 1750 datasets

Filter/Band

□ Ks	221384
□ J	171160
□ Y	57135
□ H	46531
□ I_SDSS	22614

Show 24 more out of 29

Undefined for 2047897 datasets

	Actions	Data Type	Spec.Range	Spec.Res.	SNR	Sensitivity	Obs.Date	FoV	Sky Res.	Collection	Instrum.	T.Exp.T.	#OBs	P.I.	Program Id	Object	Pub.Date
□		SPECTRUM	378.2-691.2 nm	115000	1.9		2018-04-21 22:50:1			HARPS	HARPS	1 s	single	OBSERVATORY, LA	60.A-9700	Preset-near-Zenith	2018-04-23
□		SPECTRUM	378.2-691.2 nm	115000	3.1		2018-04-19 22:49:2			HARPS	HARPS	1 s	single	OBSERVATORY, LA	60.A-9700	Preset-near-Zenith	2018-04-21
□		SPECTRUM	352.8-921.7 nm	48000	0.4		2018-03-31 03:26:2			FEROS	FEROS	18.9 s	single	OBSERVATORY, LA	60.A-9700	HR4216	2018-04-11
□		SPECTRUM	352.8-921.7 nm	48000	0.4		2018-03-31 02:43:2			FEROS	FEROS	30 s	single	OBSERVATORY, LA	60.A-9700	HR4216	2018-04-11
□		SPECTRUM	352.7-921.6 nm	48000	0.8		2018-03-31 02:04:2			FEROS	FEROS	30 s	single	OBSERVATORY, LA	60.A-9700	HR4468	2018-04-11
□		SPECTRUM	352.8-921.7 nm	48000	71.2		2018-03-31 01:54:1			FEROS	FEROS	11.1 s	single	OBSERVATORY, LA	60.A-9700	HR4216	2018-04-11
□		SPECTRUM	352.7-921.7 nm	48000	156.9		2018-03-30 23:20:4			FEROS	FEROS	120 s	single	OBSERVATORY, LA	60.A-9700	HR2554	2018-04-11
□		SPECTRUM	328.2-456.3 nm	40970	3.4		2018-03-30 03:30:3			UVES	UVES	1480 s	single	HOWK, JAY C	0100.A-0483	J091432+083743	2018-04-17
□		SPECTRUM	458.3-668.6 nm	42310	6.6		2018-03-30 03:30:2			UVES	UVES	1480 s	single	HOWK, JAY C	0100.A-0483	J091432+083743	2018-04-17
□		SPECTRUM	328.2-456.3 nm	40970	3.9		2018-03-30 03:05:0			UVES	UVES	1480 s	single	HOWK, JAY C	0100.A-0483	J091432+083743	2018-04-17

further specialised analysis. To this end, the ESO Archive Science Portal commu–nicates via the SAMP[9] protocol, which is supported by popular astronomical tools like TOPCAT and Aladin, enabling them to receive information easily.

Multi-dimensional faceted search

In order to serve a broad range of use cases, 18 query parameters are openly available. They are a combination of positional parameters (cone search around a given position on the sky), physical characteristics of the data (for example, signal-to-noise ratio, sensitivity, spectral range, spectral and spatial resolution), the observational setup (for example, filter name and exposure time), and the ESO observing process (for example, Principal Investigator [PI] name and Programme ID). Since many parameters are intrinsically interdependent, constraining one parameter typically restricts the meaningful range of one or more of the others. As a simple example, specifying a PI restricts the choices of Programme IDs to their programmes.

In order to cope with this, the query parameters and search results are grouped according to facets, so the user can easily be exposed to and navigate the multidimensional space of the archive. Wikipedia defines facets as follows[10]: “A faceted classification system classifies each information element along multiple explicit dimensions, called facets, enabling the classifications to be accessed and ordered in multiple ways rather than in a single, pre-determined, taxonomic order”. This concept may be familiar from most e-commerce sites. In practice, at any given time the user is presented with the available parameter space accounting for the previously specified constraints. In our simple example of specifying a PI name, the choices in the facet of the programme ID will be limited to the programmes by that PI.

Two additional features are offered to ease navigation. Where appropriate, entering the constraints is supported by auto-completion. Also, the possible values that a query parameter can take are grouped and presented as histograms or lists, as appropriate. In this way, the system communicates its content to users at all times, without the need for any previous knowledge. For example, as shown in Figure 3, it is immediately apparent that the archive contains data of several different types, including spectrum, catalogue, image, image cube and visibility (the counts for each of these categories are also provided). The equivalent information and grouping are available for all other search parameters.

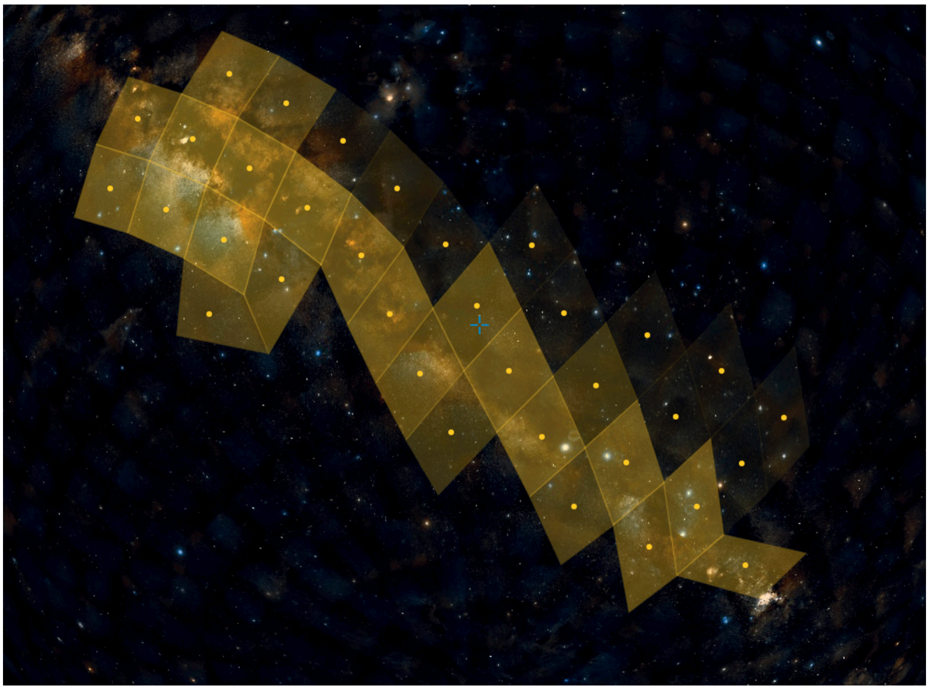


Figure 3. The all-sky search and rendering capabilities of the ESO Archive Science Portal make it easy to find and visualise data collections that span large areas of the sky. In the example above, the footprint of the VVV Public Survey covering 630 square degrees is shown on the all-sky DSS imagery. The level of transparency reflects the relative number of VVV images in the different locations on the sky.

With this approach, searches flexibly adapt to input from the users, guiding them through the content of the archive, rather than limiting them to a pre-defined set of possible paths.

Previews, hierarchical views and footprints

A preview is a lightweight, faithful representation of the actual data, which allows the user to evaluate their usefulness without transferring the full-size file(s). They are needed for a swift but in-depth assessment of the data, beyond the characterisation provided by the faceted query parameters described above.

Data exposed through the ESO Archive Science Portal display a great variety. For example, the range in images spans a few million to several hundred million pixels; in spectra it covers a few hundred to several hundred thousand spectral channels; and data cubes provide simultaneous 3D information. In terms of spatial extent, the ESO archive contains datasets that range from individual pointings to covering significant fractions of the celestial sphere — the whole hemisphere in the case of the VISTA Hemisphere Survey (VHS) public survey. This large spatial dynamic range is handled by adopting the Hierarchical Equal Area isoLatitude Pixelation (HEALpix) pixellation[11] of the celestial sphere (see Figure 3).

Customised previews are offered for different data types, which include the possibility of user interaction (for example, zooming and panning) to navigate through the different spatial and spectral scales within the data. An example of a preview of a spectrum is shown in Figure 4. Image previews are rendered with a hierarchical tiling mechanism called Hierarchical Progressive Surveys (HiPS)[12], which adaptively provides the appropriate spatial scale at any given zoom level, resulting in a responsive and satisfactory user experience. An example is shown in Figure 5, in

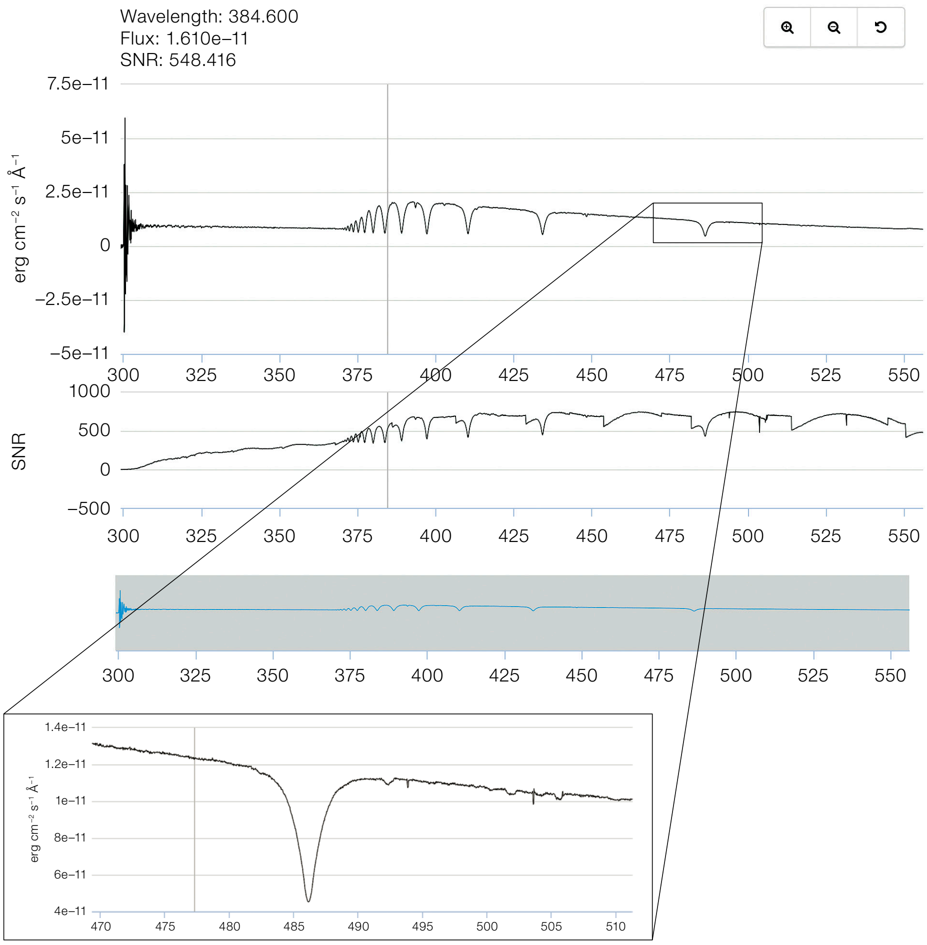


Figure 4. Example of a spectrum preview: the star Hip058859 as observed with the X-shooter instrument. Dynamic interactions are possible in order to evaluate the quality of the data and determine that they are fit for the intended purpose, e.g., by interactively zooming in on a spectral region of interest (inset).

which the preview of a tile from the VISTA Variables in The Via Lactea (VVV) Public Survey is superimposed on an image from the Digitized Sky Survey (DSS). On-sky footprints can be superimposed on an image of the celestial sphere to place the data in context and assist in browsing and selection (see Figure 6 for an example).

Direct database and Virtual Observatory access

The inherent limitation in the intuitive way that the web interface enables archive content to be discovered is that it is unsuited to more complex queries, such as those that include sequences with logical statements like “and”, “or” and “not”, or queries that join different sources of information. This restriction can be overcome by bypassing the web interface, thus providing direct access to the ESO database tables[13]. By adhering to widely recognised standards developed by the International Virtual Observatory Alliance (IVOA)[14], the ESO data can be queried alongside data from other observatories and data centres. This brings the ESO data into the appropriate general context of multi-wavelength, multi-messenger science.

Programmatic access

Users can specify their own custom queries via a standard service protocol using the IVOA’s Astronomical Data Query Language, ADQL[15]. The service

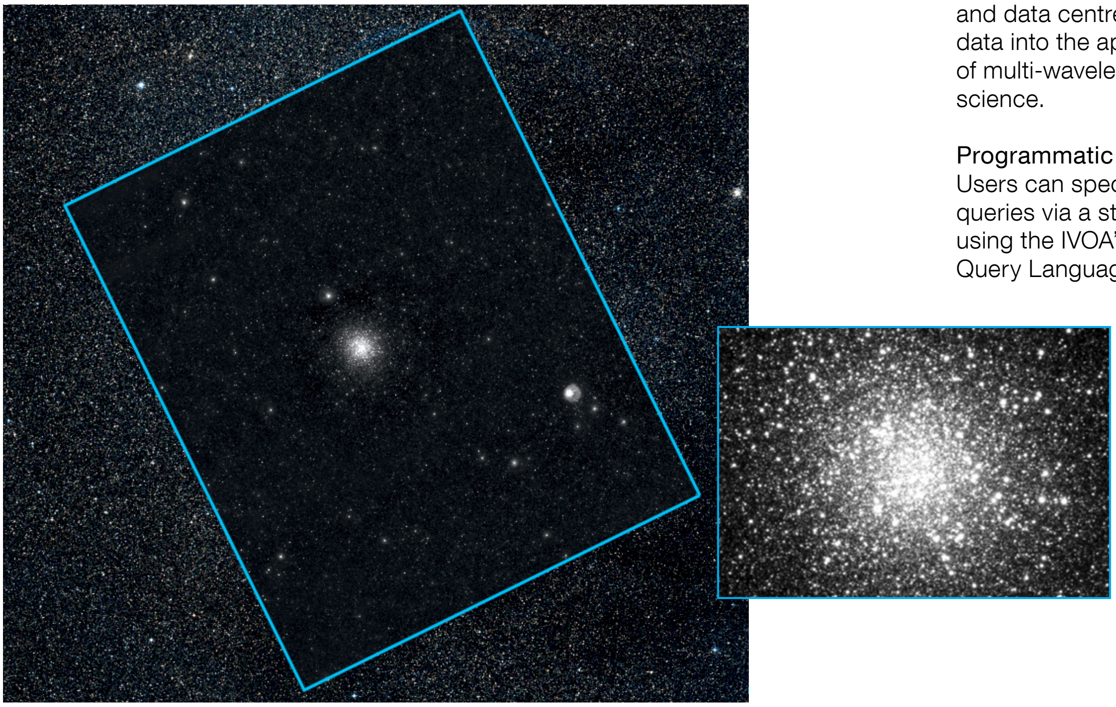


Figure 5. A preview of one tile from the VVV Public survey is shown superimposed on the backdrop of the DSS. The preview itself was generated using the HiPS mechanism and can be interacted with by zooming and panning on it. A full-resolution zoom on the inner-most regions of the stellar cluster in the tile is shown in the inset. Zooming in dynamically loads the appropriate spatial hierarchy, which provides for a responsive and satisfactory user experience.

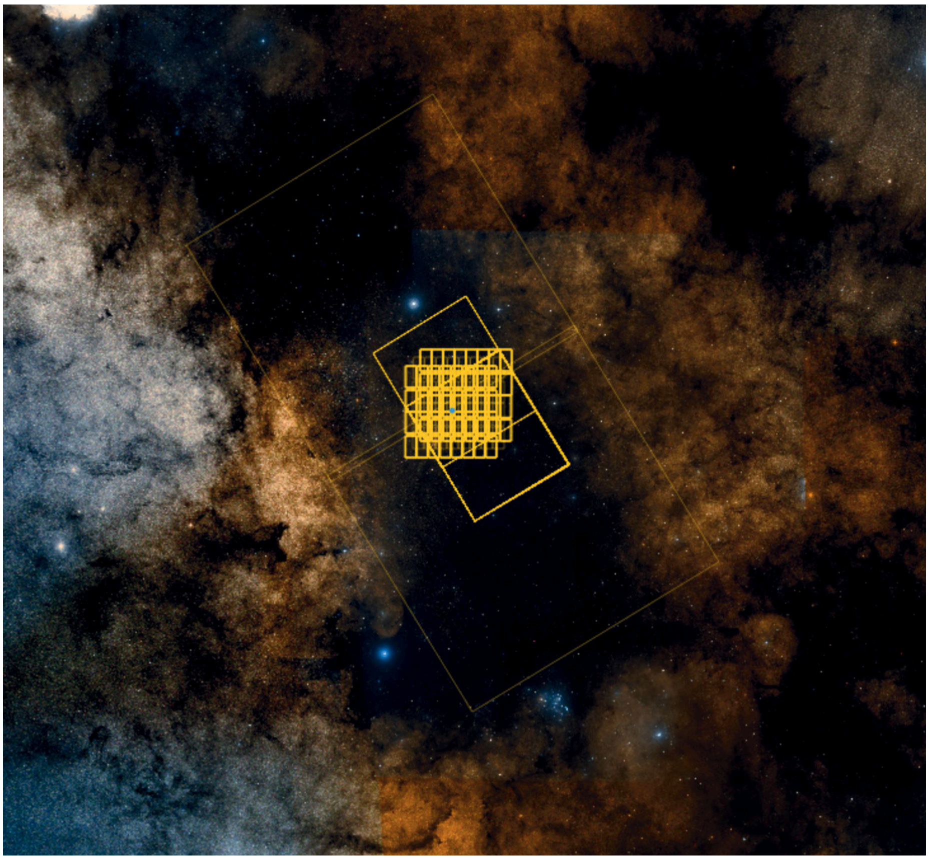


Figure 6. Examples of footprints of imaging data towards the Galactic Centre: VIRCAM data from the VVV Public Survey, OmegaCAM data from the VST Photometric Hα Survey of the Southern Galactic Plane and Bulge (VPHAS+ Public Survey), and APEXBOL data from the APEX Telescope Large Area Survey of the Galaxy (ATLASGAL Large Programme). Background imagery is from the DSS.

protocol used to accept the queries and return the results is the Tabular Access Protocol (TAP)[16] of the VO. Existing public domain software libraries providing TAP-client capabilities can be used to implement full programmatic access to the ESO science archive (for example, astroquery[17], pyvo[18], STILTS[19], to name but a few).

The ability to access the archive with scripts allows users to efficiently and reliably run long and/or repetitive sequences of queries, such as those needed to quickly access data from monitoring or other time-critical programmes. The capabilities of ADQL allow queries on the spatial footprints of the processed data. Some examples of the types of queries include searching in a cone, a more sophisticated "point in footprint" query (for example, if a user wishes to find the progenitor of a supernova that had previously been imaged in one of the 16 non-contiguous VIRCAM detectors), and the ability to find images or source tables in different filter bands whose footprints overlap (enabling the selection of processed data for colour-magnitude studies).

The tables exposed in this first release of the ESOtap[20] service are the IVOA Obscore[21] which fully characterises the processed products, and ESO tables describing the LPO and Chajnantor raw observations and atmospheric conditions[22,23,24] (for example, seeing, precipitable water vapour, isoplanatic angle). A second TAP server[25] is available to query the content of more than five billion records of high-level science catalogue data. In order to optimise the response for such a large pool of data, the spatial searches supported in this first release are limited to cones. Extensive documentation is provided in terms of practical examples, which are intended to provide templates for users to customise and adapt to their specific needs[26].

A VO data link service[27] has also been implemented. It provides access to scientific data, their ancillary files (for example, weight maps), previews and data documentation. It also lists information related to provenance (such as the data files that were used to derive these products, and any data that were produced using these files). The VO Simple Spectral Access (SSA) service[28] provides easy browsing and access capabilities for the 1D spectroscopic data.

Tool access

The same basic infrastructure behind programmatic access allows users to browse the ESO archive from VO-aware applications. This enables users to discover and access ESO data through stand-alone tools, which have powerful generic and/or specific capabilities that cannot be implemented in a general interface. Examples of such external tools include TOPCAT and Aladin, as well as tools like SPLAT-VO and other clients that implement the Simple Spectral Access Protocol (SSAP) of the VO. To achieve this, all ESO VO-compliant data services are published in the IVOA Registries, allowing VO tools to discover them.

The Archive Community Forum

Finally, open communication with users is crucial in order to collect precious feedback and exchange individual experiences. To this end, the ESO Archive Community Forum[29] is available for users to post comments, questions and suggestions addressed to ESO, or intended for the community at large. Posts are moderated by ESO and, provided they meet basic standards of relevance and etiquette, are made openly visible.

Acknowledgements

The first step towards providing this comprehensive suite of services involves the curation of archive assets, which is an integral part of the Phase 3 process (Arnaboldi et al., 2014). We would like to warmly thank the PIs and their teams who have engaged with us in this process, which has greatly enhanced the value of the ESO Science Archive.

We would like to gratefully acknowledge the very fruitful collaboration with Centre de Données astronomiques de Strasbourg (CDS). This research has made use of the Aladin sky atlas developed at CDS, Strasbourg Observatory, France (Bonnarel et al., 2000, Boch & Fernique, 2014).

Many crucial aspects of the work presented here would have not been possible without the results of the sustained, distributed efforts of the VO community. The following IVOA standards were used: ADQL v2.0, DataLink v1.0, ObsCore v1.1, SSAP v1.1, TAP v1.0, UWS v1.1[30], DALI v1.1[31], SAMP v1.3.

We have made use of the taplib library[32] by Grégory Mantelet (Astronomisches Rechen Institut, Heidelberg), which is a collection of Java libraries implementing ADQL, TAP, and UWS. Grégory's support is gratefully acknowledged. The ESO implementation of taplib[33], providing additional support for the specific Microsoft-SQL Server geographical datatypes and functions, and the implementation of the SSA protocol[34], are made available to the community via github.

References

Arnaboldi, M. & Retzlaff, J. 2011, The Messenger, 146, 45
Arnaboldi, M. et al. 2014, The Messenger, 156, 24
Boch, T. & Fernique, P. 2014, ADASS XII, ed. Manset, N. & Forshay, P., ASP Conf. Series, 485, 277
Bonnarel, F. et al. 2000, A&AS, 143, 33
Dowler, P., Rixon, G. & Tody, D. 2010, IVOA Recommendation
Dowler, P. et al. 2015, IVOA Recommendation
Louys, M. et al. 2017, IVOA Recommendation
Pirenne, B. et al. 1998, The Messenger, 93, 20
Romaniello, M. et al. 2016, The Messenger, 163, 5
Stoehr, F. et al. 2017, The Messenger, 167, 2
Tody, D. et al. 2012, IVOA Recommendation

Links

[1] ESO Science Archive: http://archive.eso.org
[2] ESO contributed processed data: http://eso.org/rm/publicAccess#/dataReleases
[3] ESO pipeline-processed data: https://www.eso.org/sci/observing/phase3/data_streams.html
[4] Report of the ESO Working Group on Science Data Management (STC-580): https://www.eso.org/public/about-eso/committees/stc/stc-88th/public/STC_580_Data_management_working_group_report_88th_STC_Meeting.pdf
[5] TOPCAT is accessible at: http://www.star.bris.ac.uk/~mbt/topcat
[6] Aladin and AladinLite: http://aladin.u-strasbg.fr
[7] The ESO Archive Science Portal: http://archive.eso.org/scienceportal
[8] CDS Sesame service: http://cds.u-strasbg.fr/cgi-bin/Sesame
[9] The IVOA Simple Application Messaging Protocol (SAMP): http://www.ivoa.net/documents/SAMP/20120411/REC-SAMP-1.3-20120411.html
[10] Wikipedia definition of faceted search: https://en.wikipedia.org/wiki/Faceted_search
[11] HEALpix data analysis and visualisation software: http://healpix.sourceforge.net
[12] The hierarchical tiling mechanism HiPS, developed by the CDS: http://aladin.u-strasbg.fr/hips
[13] Programmatic and tool access overview: http://archive.eso.org/cms/eso-data/programmatic-access.html
[14] International Virtual Observatory Alliance (IVOA): http://www.ivoa.net
[15] The IVOA Astronomical Data Query Language: http://www.ivoa.net/documents/latest/ADQL.html
[16] Table Access Protocol TAP v1.0 (Dowler, Rixon & Tody 2010): http://www.ivoa.net/documents/TAP/20100327/
[17] Astroquery: http://www.astropy.org/astroquery
[18] pyvo: https://pyvo.readthedocs.io/en/latest
[19] STILTS: http://www.star.bris.ac.uk/~mbt/stilts
[20] The ESOtap service: http://archive.eso.org/tap_obs
[21] The IVOA ObsCore data model (Louys et al., 2017): http://www.ivoa.net/documents/ObsCore
[22] Ambient conditions for Paranal: http://www.eso.org/asm/ui/publicLog?name=Paranal)
[23] Ambient conditions for La Silla: http://www.eso.org/asm/ui/publicLog?name=LaSilla
[24] Ambient conditions for APEX: http://www.apex-telescope.org/weather
[25] ESOtap server to catalogue data: http://archive.eso.org/tap_cat
[26] Programmatic and tool access demonstration page: http://archive.eso.org/programmatic
[27] IVOA DataLink v1.0 (Dowler et al., 2015): http://www.ivoa.net/documents/DataLink/index.html
[28] VO SSAP (Tody et al., 2012): http://www.ivoa.net/documents/SSA/20120210/REC-SSA-1.1-20120210.htm
[29] ESO Archive Community Forum: https://esocommunity.userecho.com
[30] IVOA standards: http://www.ivoa.net/documents/UWS/20140527/WD-UWS-1.1-20140527.html
[31] DALI: http://www.ivoa.net/documents/DALI/
[32] taplib: https://github.com/gmantele/taplib
[33] ESO implementation of tablib (G.Mantelet): https://github.com/vforchi/taplib
[34] ESO implementation of SSA on github: https://github.com/vforchi/SSAPServer

ESO/Juan Carlos Muñoz

The final image taken by VIMOS before it was decommissioned on 24 March 2018 was of the interacting galaxies NGC 5426 and NGC 5427, which form Arp 271.