The ESO Science Archive: supporting and enhancing science of the La Silla Paranal Observatory

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

[a]ESO, Karl-Schwarzschild-Strasse 2, 84748 Garching bei München, Germany; [b]Pactum Limited, London, UK; [c]TEKOM Industrielle Systemtechnik GmbH, Gautin, Germany; [d]INAF - Osservatorio Astronomico di Capodimonte, Naples, Italy; [e]Terma GmbH, Darmstadt, Germany

ABSTRACT

The archive of the La Silla Paranal Observatory is a powerful science resource for the ESO astronomical community. It stores both the raw data generated by all ESO instruments and selected processed (science-ready) data. We present the 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.

Keywords: ESO Science Archive, Web interface, Programmatic access, Virtual Observatory (VO)

1. INTRODUCTION

The ESO Science Archive began operating in 1998 in its current form, a few months ahead of the start of science operations of the Very Large Telescope, VLT (see Pirenne et al.,1998). Its access point is at http://archive.eso.org. 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 La Silla Paranal Observatory, 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 on 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 Camera (WFI), together with the raw data for the UKIDSS WFCAM survey obtained at the United Kingdom Infrared Telescope (UKIRT) in Hawaii. ESO also hosts and operates the European copy of the ALMA Science Archive (Stoehr et al, 2017). Integration of archive services for LPO and ALMA data is discussed below.

Over the years, it has grown steadily into a powerful science resource for the ESO astronomical community. This is especially the case since the routine inclusion of processed data, which can be used directly for scientific measurements, thus alleviating the need for users to do data processing on their own. An in-depth analysis of the archive usage and user community is presented in Romaniello et al (2016).

1.1 The content of the ESO Science Archive

The archive is populated with processed data through two channels. On the one hand, data-processing pipelines are run at ESO for selected instrument modes to generate products that are free from instrumental and atmospheric signatures and have been calibrated. They cover virtually the entire data history of the corresponding instrument modes and are generated by automatic processing, without knowledge of a specific science case. Checks are in place to identify quality issues with the products. On the other hand, the community contributes data products generated with processing schemes optimized to serve specific science cases and that have, in most cases, been used to derive results in refereed publications (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 mosaicked images, source catalogues and spectra.

Thorough user documentation, detailing the characteristics and limitations of each collection of processed data, is also provided for all data releases. 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 VLT 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 summary overview of the released data is available from the Phase 3 Data Releases and Phase 3 Data Streams pages[1,2].

1.2 The motivation for enhanced archive services

The number of users accessing processed data in the archive has steadily grown in 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 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 (see STC Report 580[3]), the ways in which to access the ESO Science Archive have undergone a major upgrade in order to enhance data discovery and usage, in the face of the increasing volume and complexity of the archive holdings.

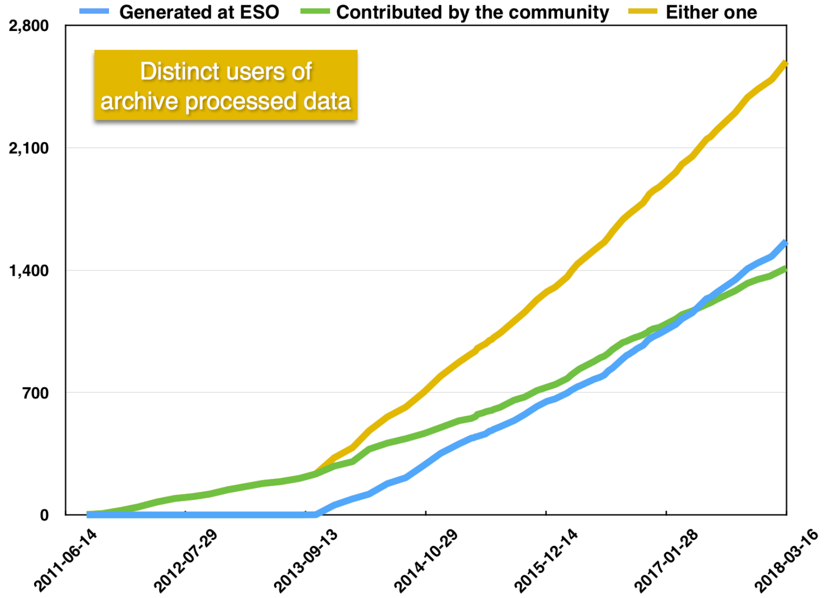

Figure 1. The growth of distinct users of processed data in the archive as a function of time. The green line displays access to products contributed by the community (deployed on 25 July 2011), the blue line shows access to products generated at ESO (deployed on 10 September 2013) and the orange line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

The trajectory of contemporary astronomical research increasingly involves multi-epoch, multi-messenger, multi-wavelength, multi-facility science, in which data is plentiful and varied. At the same time, data acquired from different facilities are becoming ever more complex and 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 as much as possible 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. It is therefore the access layer to the data that has to be as self-explanatory as possible 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).

2. THE NEW ACCESS POINTS TO THE DATA

Different ways of user interaction are supported:

- **Interactive access** via web-based pages through which users can browse and explore the assets with interactive, iterative queries. The results of such queries are presented in real time using various tabular and/or graphic ways allowing an evaluation of the usefulness of the data. Data can, the, be selected for retrieval.
- **Programmatic access,** whereby the 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 handling data capabilities, such as TOPCAT[4] for catalogues, or Aladin[5] 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 centers. 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 all the current ones for La Silla Paranal data, while ALMA will keep maintaining also a dedicated, separate access.

2.1 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[6], using any recent version of the most popular internet browsers. The direct access point is at http://archive.eso.org/scienceportal. 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, like 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 further specialized analysis. To this end, the ESO Archive Science Portal communicates via the SAMP[8] protocol, which is supported by popular astronomical tools like TOPCAP and Aladin, enabling them to receive information easily.

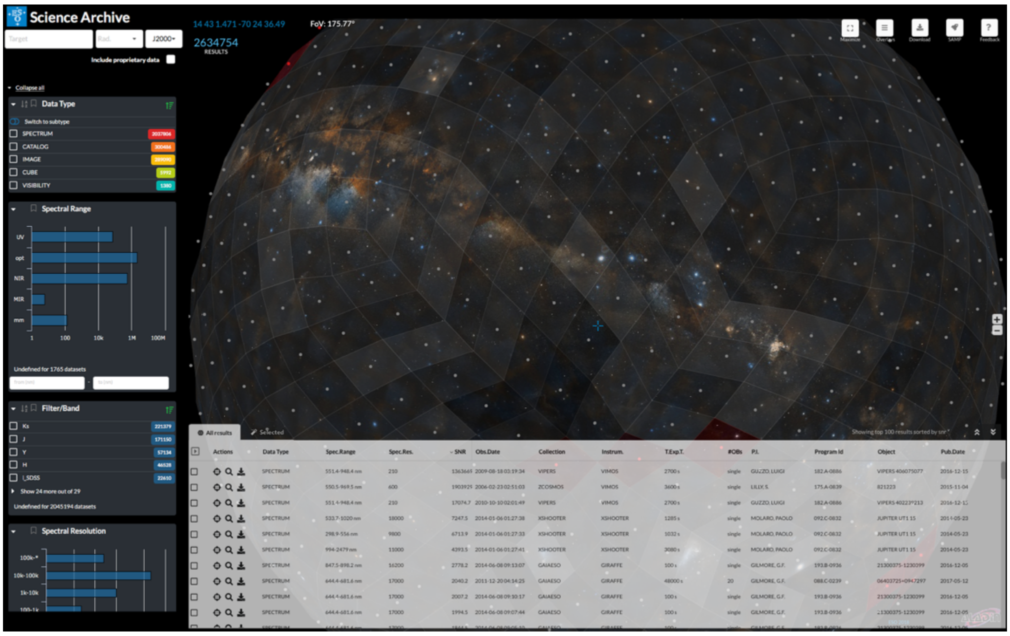

Figure 2. The landing web page of the ESO Archive Science Portal. The celestial sphere is color-coded according to the types of ESO data contained at each position (the sky viewer is CDS' AladinLite[5], the web version of Aladin). The entire content of the archive is presented through aggregations of 17 parameters, which can also be used to enter query constraints in addition to querying by object coordinate or name, as resolved by CDS' Sesame service[7]. 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), of the observational setup (e.g. filter name, exposure time) and of the ESO observing process (e.g. PI name, programme ID).

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), of the observational setup (for example, filter name and exposure time), and of the ESO observing process (for example, Principle Investigator [PI] name and Programme ID). Since many parameters are intrinsically dependent, constraining one parameter typically restricts the meaningful range in 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[9]: "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 of any previous knowledge. For example, as shown in Figure 2, it is immediately apparent in the landing page itself that the archive contains data of several different types, including spectrum, catalogue, 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.

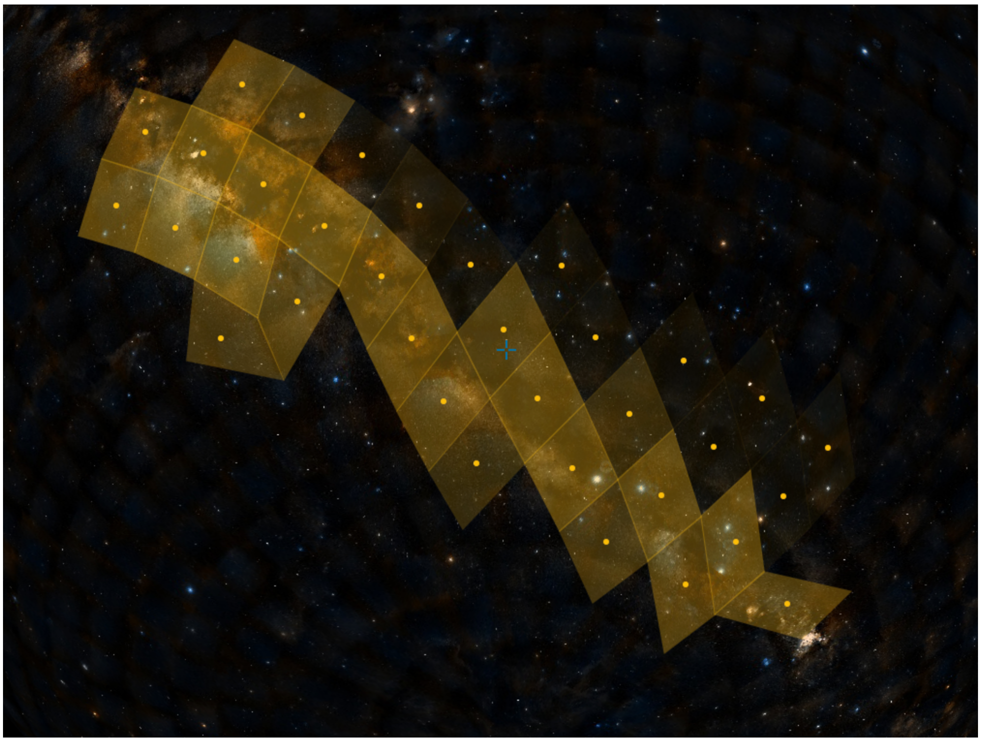

Figure 3. The all-sky search and rendering capabilities of the ESO Archive Science Portal make it easy to find and visualize data collections that span large areas of the sky. In the example above, the footprint of the VVV Public Survey covering 630 square degrees is shown on the all-sky DSS imagery. 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 its usefulness without transferring the full-size file(s). They are needed for a swift, albeit in-depth assessment of the data, beyond the characterization 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 hundreds of thousands of 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 HEALpix pixellation[10] of the celestial sphere (see Figure 3).

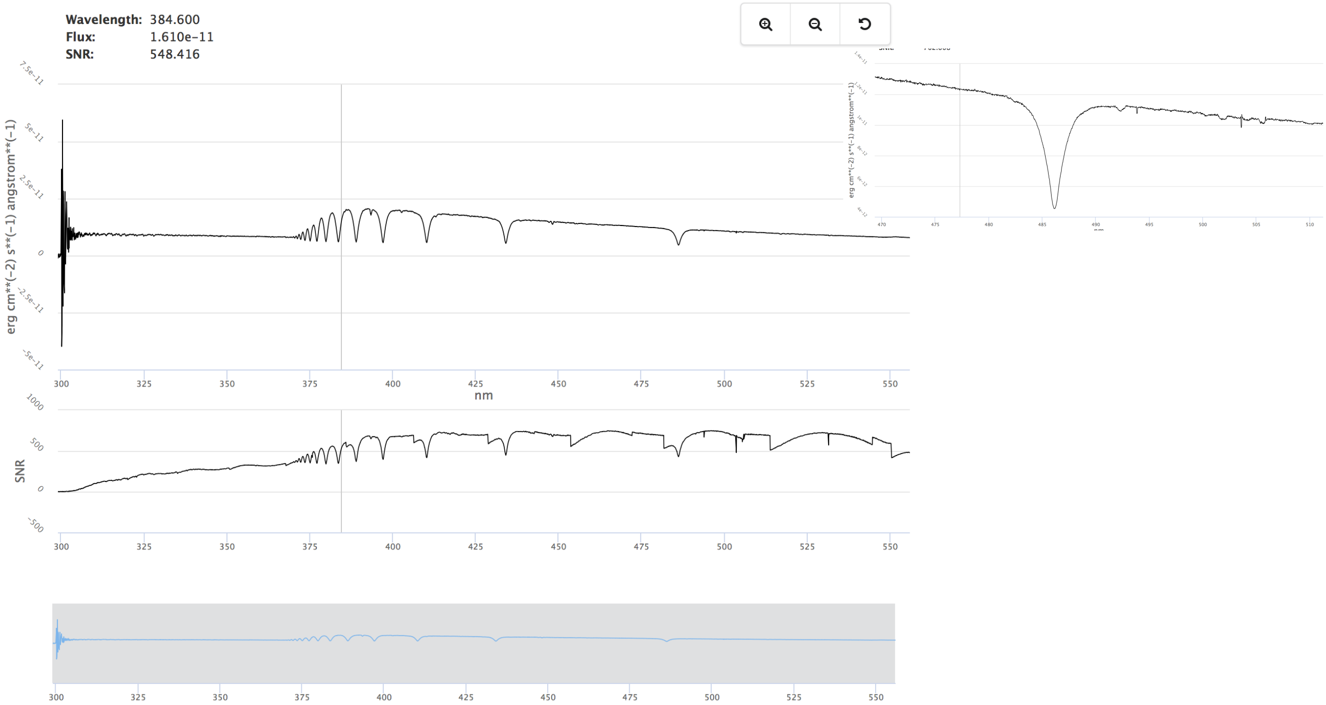

Figure 4. Example of a spectrum preview: the star Hip058859 as observed with the X-Shooter instrument. Dynamic interactions are possible in order to evaluate the quality of the data and it is fit for the intended purpose, e.g. by interactively zooming in on a spectral region of interest (inset).

Customized 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 the preview of a spectrum is shown in Figure 4. Image previews are rendered with a hierarchical tiling mechanism called Hierarchical Progressive Surveys (HiPS)[11], 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 which the preview of a tile from the VISTA Variables in The Via Lactea (VVV) Public Survey is superimposed on an image from the Digitised Sky Survey (DSS).

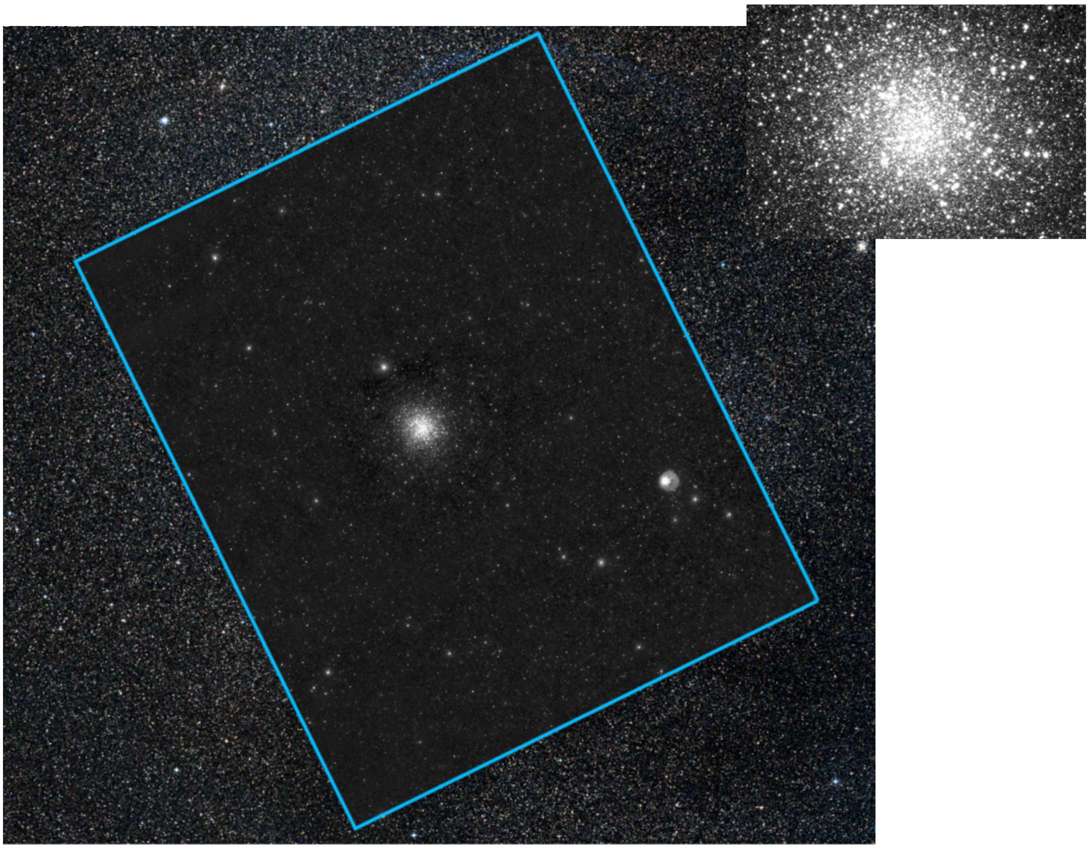

Figure 5. A preview of one tile from the VVV Public survey is shown superimposed on the backdrop of the DSS. The preview itself was generated using the HiPS mechanism and can be interacted with my zooming and panning on it. A full-resolution zoom on the inner-most regions of the stellar cluster in the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

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).

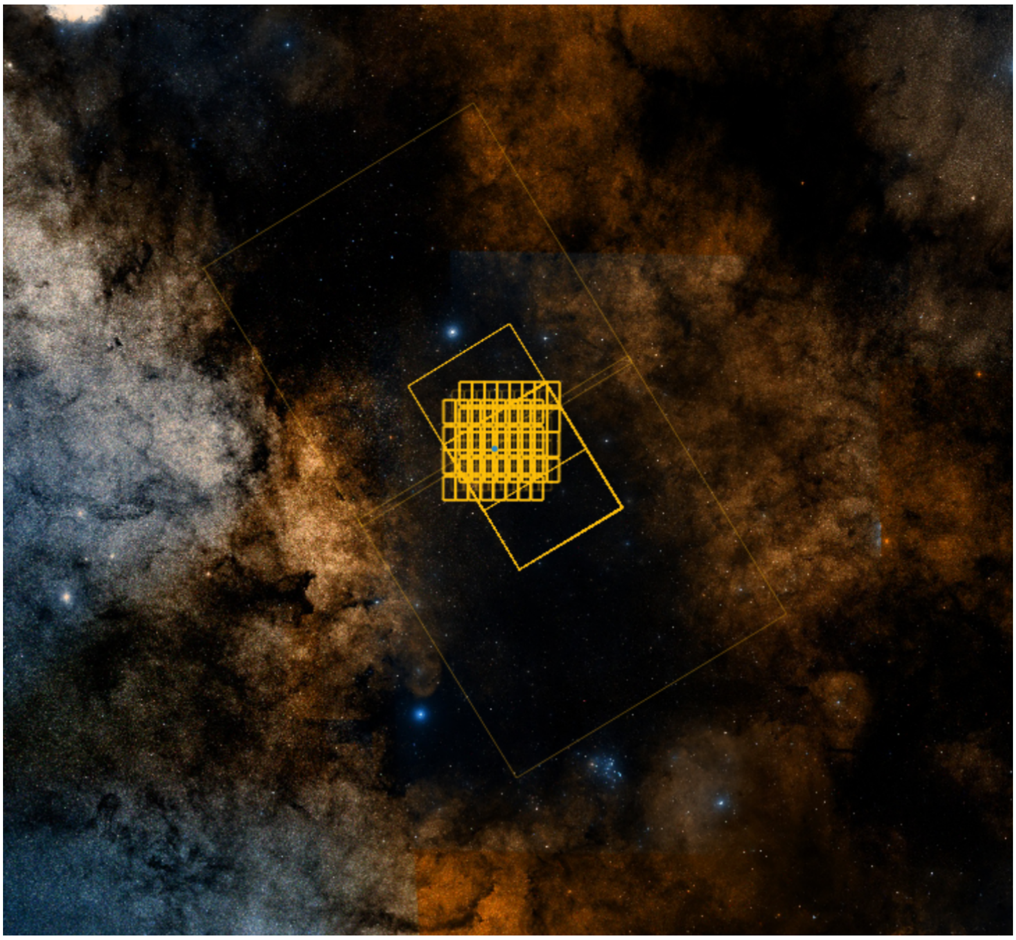

Figure 6. An example of footprints of imaging data: VISTA-VIRCAM (VVV Public Survey), VST-OMEGACam (VPHAS+ Public Survey) and APEX-APEXBOL (ATLASGAL Large Programme) footprints towards the Galactic centre. Background imagery is from the DSS.

2.2 Direct database and Virtual Observatory access

The limitation inherent to the intuitive way that the web interface provides to search and discover archive content 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 mediation of the web interface, thus providing direct access to the ESO database tables. By adhering to widely recognized standards developed by the International Virtual Observatory Alliance (IVOA), the ESO data can be queried alongside data from other observatories and data centers. This brings the ESO data in 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[12]. The service protocol used to accept the queries and return the results is the Tabular Access Protocol (TAP)[13] 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, pyvo, STILTS, 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 example of the types of queries include searching in a cone, to 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), to the ability to find images or source tables in different filter bands whose footprints overlap (enabling the selection of processed data for color-magnitude studies).

The tables exposed in this first release of the ESOtap[14] service are: the IVOA Obscore[15] which fully characterizes the processed products, and ESO tables describing the LPO and Chajnantor raw observations and atmospheric conditions[16,17,18] (for example, seeing, precipitable water vapor, isoplanatic angle). A second TAP server is available at http://archive.eso.org/tap_cat to query the content of more than five billion records of high-level science catalogue data. In order to optimize the response for such a large pool of data, the spatial searches supported in this first release are limited to cones.

A VO data link service[19] has also been implemented. It provides access to scientific data, their ancillary files (for example, weight maps), previews and data documentation. It also lists provenance information, such as the data files that were used to derive these products, and the data they gave origins to, if applicable. The VO Simple Spectral Access (SSA) service provides easy browsing and access capabilities for the 1d spectroscopic data.

Extensive documentation is provided in terms of practical examples, which are intended to provide templates for users to customize and adapt to their specific needs.

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 will be published in one of the IVOA Registries, allowing VO tools to discover them.

2.3 The Archive Community Forum

Finally, open communication with users is crucial to collect precious feedback and provide a forum to exchange individual experiences. To this end, the ESO Archive Community Forum[20] 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.

ACKNOWLEDGMENTS

Providing this comprehensive suite of services would not have been possible without the major efforts to curate the archive assets; 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 and 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, DALI v1.1, SAMP v1.3.
We have made use of *taplib* library by Grégory Mantelet (Astronomisches Rechen Institut, Heidelberg), which implements ADQL, TAP, and UWS. Grégory's support is gratefully acknowledged. The ESO implementation of taplib, providing additional support for the specific MS-SQLServer geographical datatypes and functions, is made available to the community via github.

REFERENCES

[1] Arnaboldi, M. & Retzlaff, J. 2011, The Messenger, 146, 45
[2] Arnaboldi, M. et al. 2014, The Messenger, 156, 24
[3] Boch, T. & Fernique, P. 2014, ADASS XII, Ed. N. Manset, P. Forshay, ASP Conf. Series, 485 277
[4] Bonnarel, F. et al. 2000, A&AS, 143, 33

[5] Dowler, P., Rixon, G. & Tody, D. 2010, IVOA Recommendation
[6] Dowler, P. et al. 2015, IVOA Recommendation
[7] Louys, M. et al. 2017, IVOA Recommendation
[8] Pirenne, B. et al. 1998, The Messenger, 93, 20
[9] Romaniello, M. et al. 2016, The Messenger, 163, 5
[10] Stoehr, F. et al. 2017, The Messenger 167, 2
[11] Tody, D. et al. 2012 IVOA Recommendation

LINKS

[1] ESO contributed processed data: http://eso.org/rm/publicAccess#/dataReleases
[2] ESO pipeline-processed data: https://www.eso.org/sci/observing/phase3/data_streams.html
[3] 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
[4] TOPCAT is accessible at: http://www.star.bris.ac.uk/~mbt/topcat
[5] Aladin and AladinLite: http://aladin.u-strasbg.fr
[6] The ESO Archive Science Portal: archive.eso.org/scienceportal
[7] CDS Sesame service: http://cds.u-strasbg.fr/cgi-bin/Sesame
[8] The IVOA Simple Application Messaging Protocol (SAMP): http://www.ivoa.net/documents/SAMP/20120411/REC-SAMP-1.3-20120411.html.
[9] Wikipedia definition of faceted search: https://en.wikipedia.org/wiki/Faceted_search
[10] HEALpix data analysis and visualisation software: http://healpix.sourceforge.net
[11] The hierarchical tiling mechanism HiPS, developed by the CDS: http://aladin.u-strasbg.fr/hips
[12] The IVOA Astronomical Data Query Language: http://www.ivoa.net/documents/latest/ADQL.html
[13] Table Access Protocol TAP v1.0 (Dowler, Rixon & Tody 2010): http://www.ivoa.net/documents/TAP/20100327/
[14] The ESOtap service can be accessed at http://archive.eso.org/esotap (TBC).
[15] The IVOA ObsCore data model (Louys et al., 2017): http://www.ivoa.net/documents/ObsCore
[16] Ambient conditions for Paranal: http://www.eso.org/asm/ui/publicLog?name=Paranal)
[17] Ambient conditions for La Silla: http://www.eso.org/asm/ui/publicLog?name=LaSilla
[18] Ambient conditions for APEX: http://www.apex-telescope.org/weather
[19] IVOA DataLink v1.0 (Dowler et al., 2015): http://www.ivoa.net/documents/DataLink/index.html
[20] ESO Archive Community Forum: https://esocommunity.userecho.com.