

Phase 3 Status and Access to Science Data Products from ESO Public Surveys

Magda Arnaboldi¹
 Nausicaa Delmotte¹
 Stephan Geier¹
 Laura Mascetti¹
 Alberto Micol¹
 Jörg Retzlaff¹
 Martino Romaniello¹

¹ ESO

By January 2014, the eleven currently ongoing ESO Public Surveys had successfully completed their initial submissions and publication of their science data products through the ESO Science Archive Facility. The submission and validation process for the science data products through the Phase 3 process is summarised. The science data products available in the archive and their usage by the astronomical community are presented in view of the legacy value of these projects.

Introduction

The policies to manage the ESO Public Survey projects cover the delivery of data products for ingestion and publication via the Science Archive Facility (SAF). An overview of the wavelength and area coverage of the nine ESO imaging surveys on the VISTA telescope and the VLT Survey Telescope (VST) was presented in Arnaboldi et al. (2007). Summaries of the progress of the nine imaging and two spectroscopic Public Surveys are presented in the special section of *The Messenger* 154, except for the Variables in the *Via Láctea* (VVV) survey presented in Hempel et al. (2014). Additional allocation of telescope time to Public Surveys is conditional on the submission of science data products via Phase 3 and their successful publication through the SAF.

Phase 3 concludes the process started with the submission of the letters of intent, followed by Phase 1 (proposal preparation and submission) and the preparation and submission of observing blocks (OBs) for observations in service mode, i.e., Phase 2. Phase 3 is a partnership between the principal investigators (PIs) of the Public Survey and ESO and is based on the definition of a science data

product standard, thereby guaranteeing uniformity in terms of data format and characterisation across the entire archive. As a result of Phase 3 the community can access and download the data products from the SAF and is able to carry out independent science projects in addition to those targeted by the survey teams (c.f., Arnaboldi et al., 2011).

From data upload to archive publication

When the PI, or their co-investigator(s) (Co-Is), uploads the data products to the dedicated Phase 3 ftp staging area and closes a release, checks are made and a valid status signals that the data are formally compliant with the ESO science data products standards. This step certifies that the files are standard FITS, i.e., were not corrupted during transfer, that the mandatory keywords are present, the provenance keywords are correctly mapped to products already archived in the SAF, or within the current upload, and the associations with the ancillary files are correct and complete. Being formally compliant with the standard is a necessary condition for publication via the SAF, but it does not guarantee that the data content is consistently documented to the required level of detail that would allow further scientific analysis independent of the survey team. The process that brings the formal compliance to published science data products via the SAF is the content validation of a Phase 3 submission.

The content validation of a Phase 3 submission by an ESO Public Survey team is carried out by the Archive Science Group (ASG). The release description is the starting point to assess whether the submitted data are consistent with the associated data documentation in terms of quantity (i.e., number of files), science quality and types of products, i.e., OB level products (those on the level of a single OB, such as pawprints, source lists or 1D spectra) or high-level products (such as deep stacks, aperture-matched, band-merged source lists and catalogues). ASG members make use of *ad hoc* scripts to check the values of the mandatory keywords, and the consistency of interconnected keywords. For example total exposure time values are

checked against the Julian dates and the consistency of the NCOMBINE keyword for number of combined images and spectra is checked against the number of provenance (PROV) keywords given in the header. The content validation also includes spot checks for corrupted data.

For high-level data products, e.g., deep stacked images and catalogues, extensive tests are carried out to ensure the overall consistency of the science content in the archive. For the high level products, the PROV keywords are essential to ensure consistency within the hierarchy of products, from OB level products (pawprints and source lists) already published through the SAF, or delivered in the same data release, to the deep stacked images and catalogues. An additional test for catalogues covers statistics on the distribution of the values of the records in the different catalogue columns to check for outliers that may prevent the ingestion into the database, which supports the query interface of the ESO Catalogue Facility. ASG members also check for consistency of the Unified Content Descriptor (UCD) identifiers, the uniqueness of the name identifier for source records and presence of the primary source tag. For 1D spectra, specific tests are carried out on the FITS tabular format to ensure that the wavelength solution is monotonic. Further details can be found in the Phase 3 user guide¹.

In Figure 1 we plot the time to publication for the second VISTA submission (#2), VST submission #1 and the spectroscopic surveys submission #1. In this figure, $T = 0$ refers to the start of content validation by ASG and it accounts for both ASG and the PI/Co-I interventions. In Figure 2 we report the number of ASG content validation reports prepared for the survey submissions prior to publication of the data products in the archive.

The figures of merit shown in Figures 1 and 2 illustrate that the partnership of the survey teams and ESO is essential in order to achieve publication within two/three months from closing a Phase 3 submission. The leadership role of the PIs of the surveys is a cornerstone of the policies for Public Surveys, and follows from the agreement signed by the ESO Director General prior to the allocation of

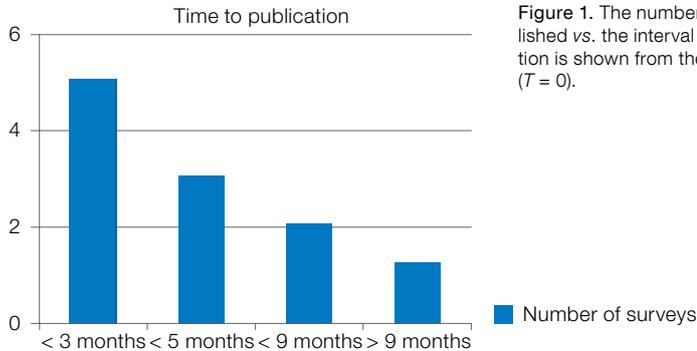


Figure 1. The number of survey releases published vs. the interval of time (months) to publication is shown from the start of content validation ($T = 0$).

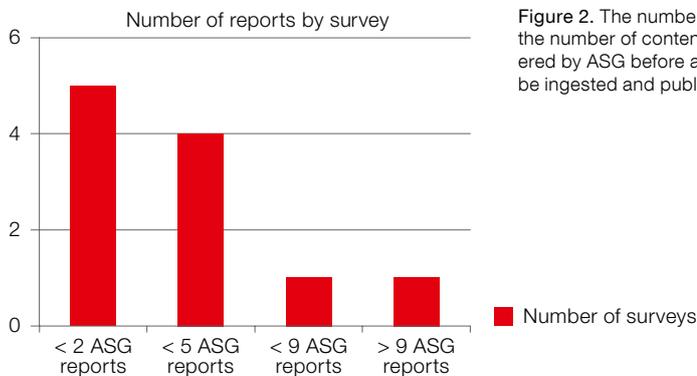


Figure 2. The number of surveys plotted against the number of content validation reports delivered by ASG before a Phase 3 submission can be ingested and published in the archive.

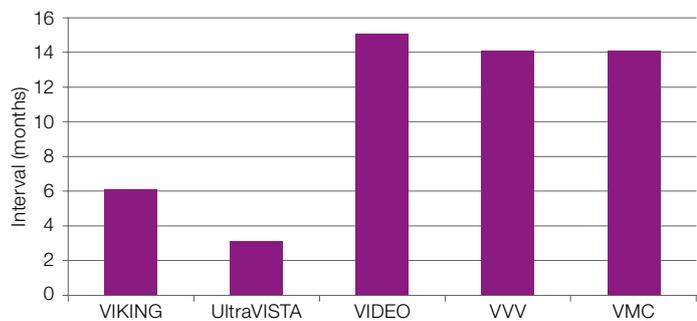


Figure 3. Interval between ingestion of catalogue products and archive publication is shown for the first catalogue submission.

telescope time to a survey project. Hence ESO assists the survey teams to define and package their data products in a manner consistent with the SAF and Virtual Observatory standards, enabling them to be integrated into the SAF.

Once the teams have ascended the learning curve on the science quality of the products, their data format and metadata, provenance information and content, another major delay can occur whenever there is a lack of consistency among the quantity and quality of delivered products, or the content of the uploaded Phase 3 submission is unclear,

i.e., there are inconsistencies between the release description and uploaded files.

Content validation of catalogue products

Most catalogue products that were uploaded during the first catalogue submission did not fulfil the scientific quality required by the Phase 3 standard, primarily with respect to the qualifying attribute of the uniqueness of the source identifier. This issue is relevant for all surveys, except the UltraVISTA survey, for those sources in the region of overlap

among tiles. Following discussion with the survey teams and the Public Survey Panel in October 2012, a strategy was identified so that catalogues now include an additional column which lists the primary source tag. Since catalogues are clearly the highest level science products from imaging surveys, it was necessary that ESO and the survey teams work together towards the common goal of achieving the required uniqueness of sources in the catalogue records. In Figure 3 we show the time from submission to ingestion for the catalogue products.

Dissemination of the catalogues from the ESO Public Surveys is further supported by an agreement between ESO and the Centre de Données astronomiques de Strasbourg (CDS). Only those catalogues from the Public Surveys that have been validated and published via the SAF are curated and published at CDS (in addition to those published in the refereed journal *Astronomy & Astrophysics*).

Current published science data products assets

The core deliverables of data products from the Public Surveys to the SAF are:

- Astrometrically and photometrically calibrated, co-added, re-gridded tiles, along with their respective confidence maps, in all of the project-relevant filters.
- Source lists for each tile based on individual, co-added bands. The single band catalogues will be ingested and accessible from the SAF as FITS files.
- 1D wavelength-calibrated, and flux-calibrated or continuum-normalised, spectra, including the 1D sky spectrum used for subtraction of the object spectrum, and an error map.
- Catalogues containing a list of parameters of individual objects across all of the observed filter bands. Their precise content will depend on the scientific goals and exploitation possibilities of each Public Survey.

The content of the Phase 3 submissions, completed during 2013, for the VST and VISTA Public Surveys (Tables 1 and 2 respectively) and for the spectroscopic Public Surveys are described.

ESO Public Survey data submissions

In 2011 Phase 3 operations began with the first submission of ESO Public Survey data products from five of the six VISTA Public Surveys (VISTA submission #1). A total of 5.5 TB of compressed imaging data including single-band source lists were submitted and published in the SAF. VISTA submission #2 (see Table 2) is a subsequent data release following VISTA submission #1. It included logical dependencies among archived and newly submitted data, which are mapped in the database that supports the Phase 3 workflow and the Phase 3 query interfaces of the SAF. This functionality of the Phase 3 infrastructure was fully tested and deployed to support archive operations during this process. It ensures that science data products can be searched and downloaded conjointly across all active data releases.

VISTA survey PIs were invited to deliver high-level catalogue data products before the reference date of 1 October 2012. The primary deliverables are aperture-matched multi-band catalogues, which were provided by five of the six VISTA surveys. The Ultra-VISTA band-merged catalogue of the COSMOS field is based on the previously released deep stacked images. The VIDEO team delivered the multi-band catalogue based on the deep stacked data of the VIDEO-XMM3 field. In order to support variability studies, the VVV and VISTA Magellanic Clouds (VMC) survey teams delivered the requested multi-epoch photometric data (light curves) and catalogues of variable objects.

The first submission of the VST Public Surveys took place in 2012 leading to the publication of 4.4 TB of imaging data covering over 2700 square degrees in the *u,g,r* and *i* bands, including *z* and *H α* bands according to each survey (Table 1).

The first submissions of data from the spectroscopic Public Surveys, Gaia-ESO and Public ESO Spectroscopic Survey of Transient Objects (PESSTO), took place in 2013. They include as primary deliverables the 1D wavelength-calibrated spectra from those observations, which were completed before the reference

Table 1. Content of the VST submission #1. All data products based on observations completed before the reference date of 15 September 2012 on a tile-by-tile basis.

Survey	Deliverable 1 — tiles & confidence maps	Deliverable 2 — single band source list for each tile
V-ATLAS	<i>u,g,r,i,z</i> (2341 deg ²)	<i>u,g,r,i,z</i>
KIDS	<i>u,g,r,i</i> (56 deg ²)	<i>u,g,r,i</i>
VPHAS+	<i>u,g,r,i,Hα</i> (375 deg ²)	<i>u,g,r,i,Hα</i>

Table 2. Content of the VISTA submission #2. All data products based on observations completed before the reference date of 1 July 2012 on a tile-by-tile basis.

Survey	Deliverable 1 — tiles & confidence maps	Deliverable 2 — single-band source list for each tile	Deliverable 3 — co-added deep stacked tile
UltraVISTA	<i>Y,J,H,Ks,NB118</i>	Single bands for each stack (1.5 & 0.74 deg ²)	<i>Y,J,H,Ks,NB118</i>
VHS	<i>Y,J,H,Ks</i> (3800 deg ²)	—	—
VIDEO	<i>Z,Y,J,H,Ks</i> (9 deg ²)	—	Deep stacks for XMM field
VVV	<i>Z,Y,J,H,Ks</i> (564 deg ²)	—	—
VIKING	<i>Z,Y,J,H,Ks</i> (235 deg ²)	—	—
VMC	<i>Y,J,Ks</i> (3.5 deg ²)	—	—

Table 3. Summary of VISTA and VST Public Survey products in the ESO SAF. The quoted sky coverage in column 2 is the total geometric area covered by the image products, which normally differs from the nominal survey area.

Survey	Bands	Sky coverage ¹ (sq. deg.)	Data volume (GB)	Sky coverage of raw data at Oct. 2013 (sq. deg.)
VHS	<i>YJHKs</i>	4210	8511	8260
VIKING	<i>ZYJHKs</i>	235	288	700
VVV	<i>ZYJHKs</i>	564	2877	562
VMC	<i>YJKs</i>	3.6	26	105
Ultra-VISTA	<i>YJHKs</i>	1.8	86	1.8
VIDEO	<i>YJHKs</i>	9.0	70	10.2
ATLAS	<i>ugriz</i>	2341	3015	2410
VPHAS+	<i>ugri,Hα</i>	375	747	600
KIDS	<i>ugri</i>	56	701	360

Table 4. Summary of the Gaia-ESO and PESSTO Public Survey products in the ESO SAF (status at 15 April 2014).

Survey	1D spectra	Total number of files	Data volume (GB)
Gaia-ESO	Spectra and weight maps	5654	1.3
PESSTO	Spectra, weight maps and images	4991	18

date 15 December 2012. As ancillary products, the PESSTO team further delivered 2D spectral frames and associated images.

The complete sky coverage of the science data products from the Phase 3 submissions of the ESO Public Surveys is shown in Figure 7 of Arnaboldi et al. (2013).

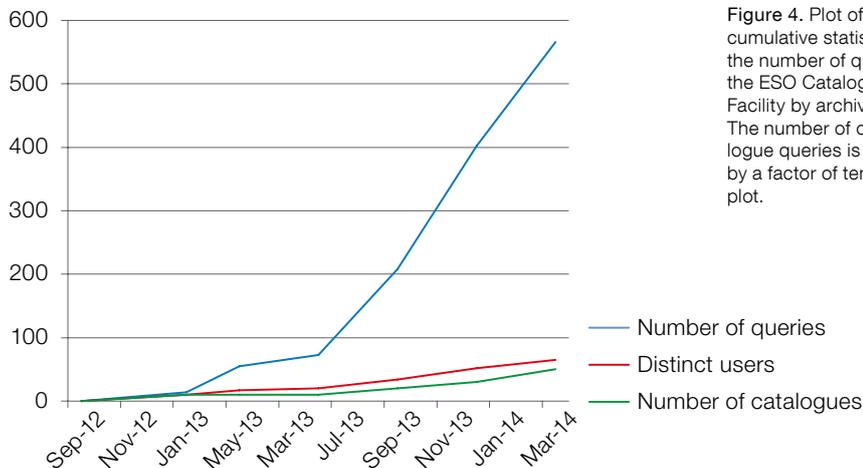


Figure 4. Plot of the cumulative statistics of the number of queries of the ESO Catalogue Facility by archive users. The number of catalogue queries is scaled by a factor of ten in this plot.

Publication of science data products from the ESO Public Surveys

The year 2013 was a very important one for Phase 3 activities, as all eleven ESO Public Surveys submitted and published their data products via the SAF. The milestones for Phase 3 were the second VISTA submission for images and source lists, and the first submission for catalogues. The first data release of the VST surveys was announced in September 2013. The spectroscopic Public Surveys completed the publication of their data products in January 2014. Thus far, a total volume of 16 TB of data products – images, weight maps, source lists, 1D spectra and catalogues – is now available and fully searchable via dedicated query interfaces². In Table 3 we provide an overview of the data volume, wavelength and sky coverage of the data releases from the imaging surveys. In Table 4 we provide a summary of the submission by the Public Spectroscopic Surveys. Further information and detailed descriptions of the data releases from the ESO Public Surveys are available³.

Public Survey data are published through the ESO Archive interfaces conjointly with other products, such as the stream of UVES echelle data that result from in-house generation of science data products. All Phase 3 data products comply with the established standard for ESO science data products⁴, thereby guaranteeing uniformity in terms of data format and characterisation across the entire ESO archive.

In Figure 4 we show the cumulative statistics of the number of queries carried out by archive users at the Catalogue Facility since September 2012. The increase in the cumulative curves signals a “critical mass” effect: as more catalogues become available, more users are interested in the catalogue products available from all the Public Surveys. The current results from the publication of science data products from ESO Public Surveys show that we are moving in the right direction towards making these products available to the community.

In general, the delivered OB-level products cover smaller areas than those for which the raw data are available, as shown by comparing the areas covered by the science products (Column 3 of Table 3) and the areas covered by the raw products from successfully executed OBs (Column 5 of Table 3). The comparison between science data products and delivered raw data also shows that we are only at the beginning of this endeavour: more science data products will become available in the future Phase 3 calls for submission and in the course of the Public Survey observations.

A new cycle of Phase 3 calls is taking place in 2014. It will comprise the VISTA submission #3, VST submission #2 and submission #2 for the spectroscopic surveys. The VST submission #2 and spectroscopy survey #2 will also entail the delivery of high-level products, e.g., catalogues.

Refereed publications from Public Survey data products

Merit parameters for the Public Surveys are the number of refereed publications by the survey teams and archive users, and the cumulative download of data products from the SAF. There are now 135 refereed publications from the survey teams; a significant increase on the number of refereed publications (up 100%) since October 2013. There are also 15 refereed papers from archive users who are not members of the survey teams, such as the analysis of data products from the VVV survey by Wegg & Gerhard (2013).

The statistics on archive downloads shown here, either by volume or number of files, does not include download by operational accounts of the ESO User Portal. We define as operational accounts those associated with PIs and Co-Is of the survey teams that are delegated for the Phase 3 data upload, and those of ESO users who have functional duties related with Phase 3 and the SAF.

The parameters on the data download by the community also demonstrate a strongly increasing interest. The cumulative downloads from the SAF since December 2011 amount to more than 10.5 TB of data products and ~ 50 000 files. In Figures 5a and 5b these numbers are differentiated by survey project and in Figure 6 by data product type. From the cumulative trend plots shown in Figures 5a and 5b, it is obvious that the community is most eager to access data from the VVV, UltraVISTA, KiDS (with the largest volume downloaded) and Gaia-ESO surveys. The largest data volume downloaded for products is for the source lists, followed by tile images and 1D spectra; see Figure 6. Since we believe catalogues to represent very valuable archival assets, since they are the highest level products for surveys, we are working hard to ingest the remaining VVV and VISTA Hemisphere Survey (VHS) catalogues so that the community can benefit even more from the joint efforts of ESO and the survey teams. Figure 7 shows the download of 1D spectral products.

We can also map the number of distinct archive users and the number of requests

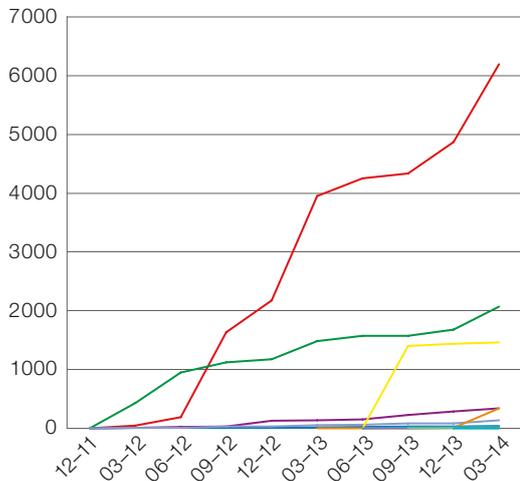


Figure 5a. Data volume downloaded (GB) for the ESO Public Surveys.

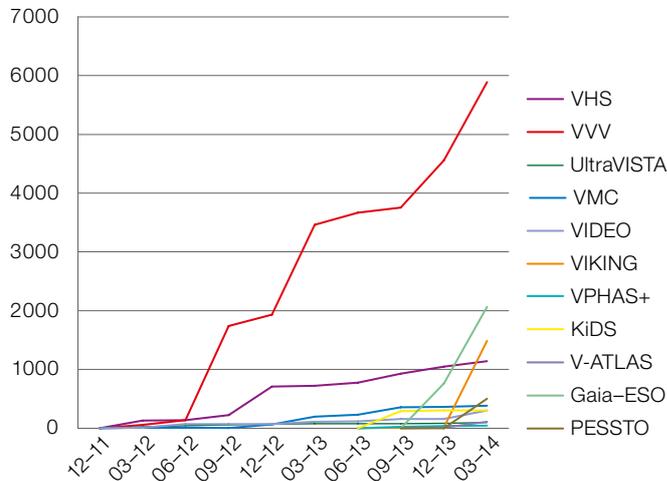


Figure 5b. Number of files downloaded for the ESO Public Surveys.

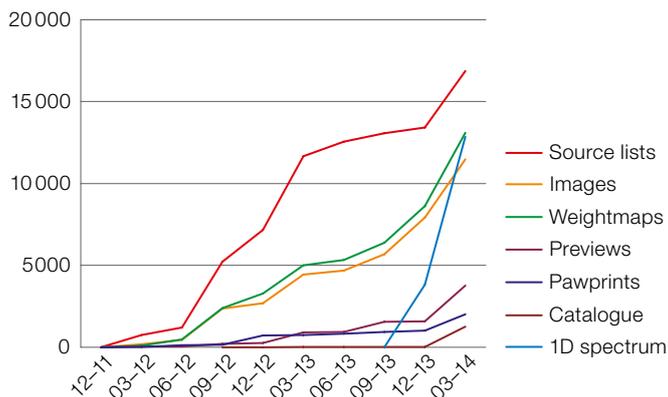


Figure 6. Number of files downloaded from the SAF for the different data product types. For catalogues, the number of files refers to a whole catalogue download. This figure of merit is complementary to the number of queries shown in Figure 4, in the sense that queries for catalogue records may not require the full download of a catalogue FITS file.

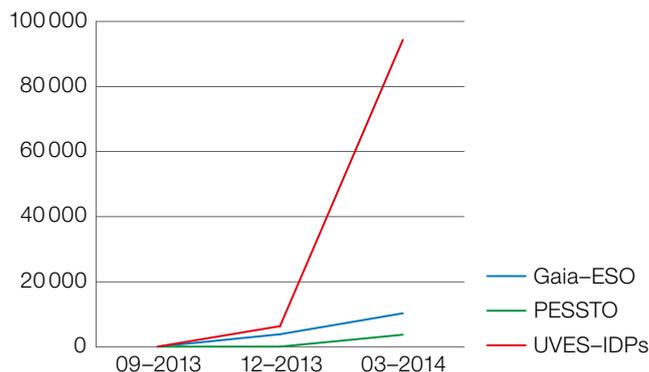


Figure 7. Download of 1D spectral products, including the UVES echelle science data products and the Gaia-ESO and PESSTO Public Survey products. The strong increment in the first quarter of 2014 is also associated with the download of all the available UVES data products by a single user.

per user. Again these statistics are carried out for users who are not associated with operational accounts. The access to data products in the SAF was performed by 517 distinct users, with about 50% having executed at least two requests resulting in a successful download of data products.

Concluding remarks

By monitoring the access and download of data products from the SAF we are witnessing a “critical mass” effect caused by the successful publication of homoge-

neous science data products across the Public Surveys. We note that since October 2013 and the completion of the first Phase 3 submission cycle from all the eleven Public Survey projects, the download of science data products (number of files and volume) from the SAF has doubled. There is also a “threshold” effect in the interest/awareness of the astronomical community: the dedicated section in the December 2013 issue of the *The Messenger* on Public Surveys triggered interest and an increased download of science data products from the SAF. There is evidently a strong scientific community interest in the study

of the Milky Way (beyond the science goals of the Public Survey teams) as shown/triggered by the increased download of VVV products after the ESO Release on the intrinsic peanut shape of the Milky Way Bulge (Release 1339) and by the synergies created between the ESA Gaia mission and the Gaia-ESO spectroscopic survey.

Synergies are also present between products from different surveys. Since Public Survey data are published through the ESO archive interfaces conjointly with other science data products, such as the stream of UVES echelle spectra

(uniform in terms of data format and characterisation across the archive), archive users can exploit the links between products. The statistics indicate that the VISTA Kilo-degree Infrared Galaxy (VIKING) survey download increased after publication of catalogues and KiDS (Kilo-Degree Survey) images; UltraVISTA download increased after catalogue and second release of high level data products; and the publication of UVES echelle data products, which showed in six months a cumulative download of one fifth of all science data products downloaded in terms of number of files (Figure 7), also increased the visibility of 1D spectral products from the Public Surveys.

It is clear that the community involved in the science carried out with the ESO Public Survey projects is expanding: further to the nearly three hundred astronomers involved in the eleven Public Survey teams as PIs and Co-Is, there is now a community of archive users which is similar in size, or even larger, and is accessing these data for their own independent science.

Acknowledgements

The Archive Science Group would like to thank all the Public Survey teams for their hard work and their support in making the ESO Public Survey projects a success.

References

- Arnaboldi, M. et al. 2007, *The Messenger*, 127, 28
- Arnaboldi, M. et al. 2011, *The Messenger*, 144, 17
- Arnaboldi, M. et al. 2013, *The Messenger*, 154, 18
- Wegg, C. & Gerhard, O. 2013, *MNRAS*, 435, 187

Links

- ¹ Phase 3 User Guide: <http://www.eso.org/sci/observing/phase3/p3userguide.pdf>
- ² Phase 3 generic query interface: http://archive.eso.org/wdb/wdb/adp/phase3_main/form
- ³ Descriptions of the data releases from the ESO Public Surveys: http://www.eso.org/sci/observing/phase3/data_releases.html
- ⁴ Science Data Products standard: <http://www.eso.org/sci/observing/phase3/p3sdpstd.pdf>

ESO/G. Brammer



A fish-eye lens view of the Paranal Observatory, showing the four Unit Telescopes (UTs) and four Auxiliary Telescopes. The laser guide star emerges from the dome of UT4 (Yepun). The Large and Small Magellanic Clouds can be seen at the top of the image and the Galactic Plane near the horizon.