

Ensuring the legacy value of ESO Public Surveys

Community data access through the Science Archive Facility

J. Retzlaff

On behalf of the ESO **A**rchive **S**cience **G**roup M. Arnaboldi, N. Delmotte, S. Geier, L. Mascetti, A. Micol





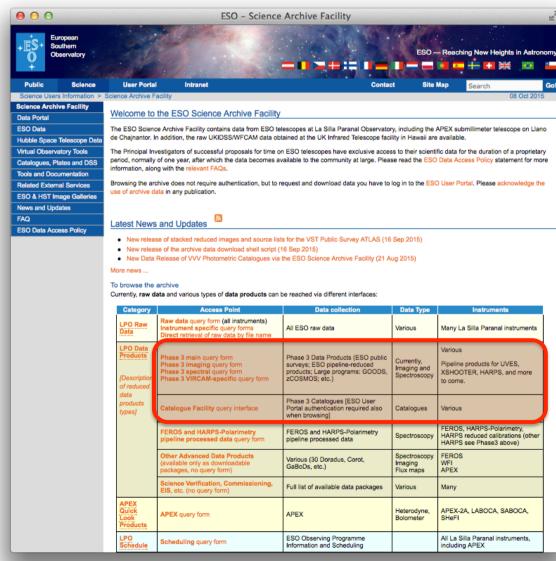
Outline

- 1. Overview of Public Survey data releases
- 2. Large programmes & pipeline data reduction
- 3. Data access statistics
- 4. Phase 3 process



Science Archive Facility

http://archive.eso.org



- The ESO SAF is the collection point for the survey products and the primary point of publication/availability of these products to the ESO community.
- Delivery of the data products to the ESO archive from an ESO Public Survey is the responsibility of the PI, who certifies the scientific quality and accuracy of the data products.

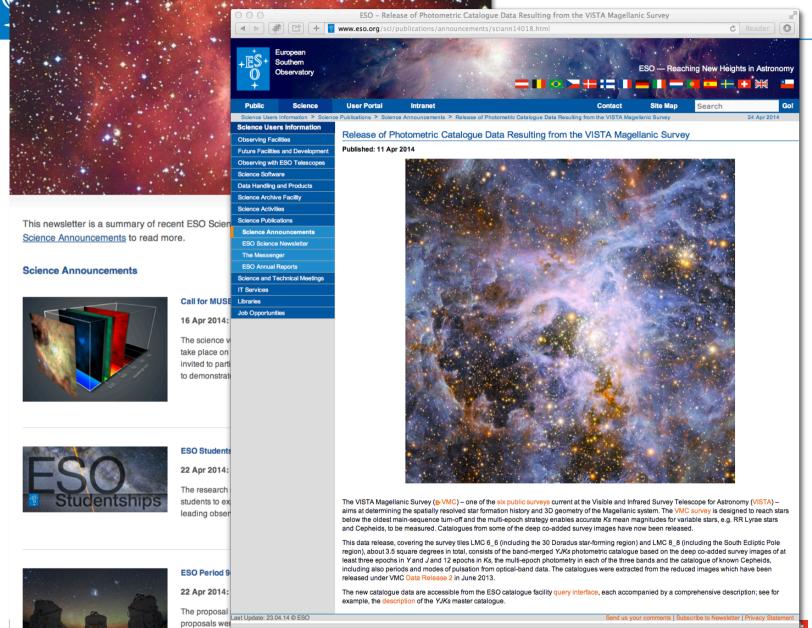


Public survey data releases 2015

Date	Survey	Data
Sep 2015	VST ATLAS	Stacked ugriz images and source lists. 2000-3000 sq.deg. observations until Sep 2013. 5 TB.
Aug 2015	VVV	ZYJHKs photometric catalogues. 418 Mio. sources. 2.5 Billion Ks photom. points. 13 Mio. variables. Including observations until Sep 2011.
Aug 2015	PESSTO	1728 EFOSC2 and SOFI spectra and 392 reduced SOFI images. Associated catalogs.
Jul 2015	VPHAS+	Band-merged catalog based on DR2 imaging includes observations until Sep 2013.
Jul 2015	Gaia-ESO	27359 spectra of point-like sources obtained with FLAMES until Dec 2013.
May 2015	VPHAS+	Images and single-band source lists for data acquired until Sep 2013.
Mar 2015	VIDEO	ELAIS-S1 field
Mar 2015	VVV	Images and source lists including observations until Sep 2013. Ks variability survey.
Feb 2015	VIDEO	CDFS & XMM-LSS field
Feb 2015	VMC	PSF photometry
Feb 2015	VHS	14 TB imaging data including observations until Sep 2013 + band-merged catalog of 590 Mio sources
Feb 2015	KiDS	Images and catalogues including data acquired until Sep 2013. 148 sq.deg. source catalog 17 Mio. sources.
Jan 2015	VMC	YJKs photometric catalogue from 7 completed tiles (10.5 sq.deg) + multi-epoch photometric data + catalogs of Cepheids and Eclipsing binaries.
Oct 2014	VIKING	Images and catalogues including data until Sep 2013. 690 sq.deg + 46 Mio. sources catalog with band-merged photometry.

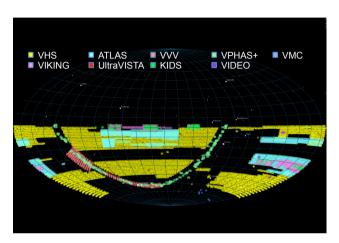
ESO Science Newsletter April 2014

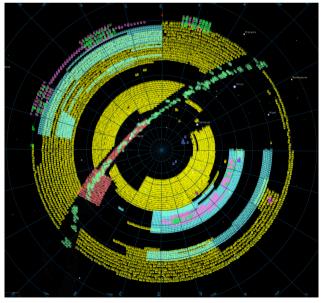
Announcement of Phase 3 Data Releases to the Community





Science Archive Facility (1) Public Survey data products





- > 11 public surveys
 - VISTA DR3, VST DR2
 - SPSs: GaiaESO, PESSTO DR2
- Sky coverage >11500 sq.deg
 - Opt./NIR: 4336 / 9445 sq.deg
- Tot. data volume: >35 TB
 - 270k+ files
 - >29k spectra
- High-level science catalogs for all surveys



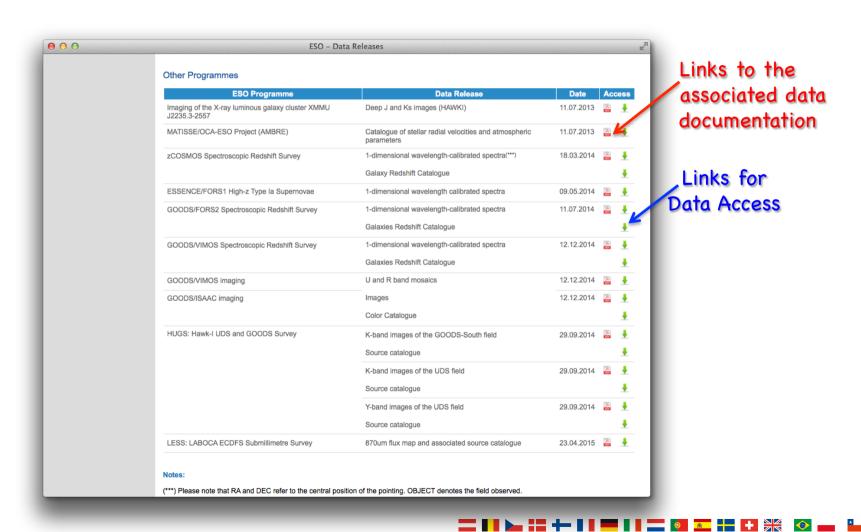
data,





Phase 3 Release Overview (cont'd)

ESO large programs, normal programs etc.

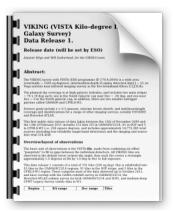




atlas_dr1.pdf



video_cdfs_dr1.pdf





Data Release Description

Provide short broad overview of the program, with an overview/layout of the observations

Essential input for data content validation.

Release content

Extended listing for each sky position, filters, exposure times, seeing

Release notes

Reduction method used, calibration procedures, data quality

Data format

Description of files in this data release, associated files, and naming conventions

Acknowledgements

 Bibliographic reference to be included when using these data.







vhs_dr2.pdf



viking_cat_dr1.pdf



viking_dr1.pdf vmc_dr2.pdf vphasplus_dr1.pdf vvv_dr2.pdf vvv_dr2.pdf



Data Product Types

Survey Tile Image

- Astrometrically & photometrically calibrated FITS image with associated confidence/weight map;
- Quality params. (limiting magnitude, PSF size, etc.)
- Generally: processing provenance (keyword PROVi) to trace back to the original (raw) data.

1-dim. extracted wavelength-calibrated spectrum

- FITS binary table format (compliant to IVOA Spectral DM, v1)
- Support for 2d spectral frames as ancillary files

Exposure map of the VISTA survey tile.

Survey Source List

- Single-band source catalogues directly extracted from the (tile) image -associated to its originating image (provenance keyword PROVi)
- Based on nightly calibrations, degenerate w.r.t. physical sources

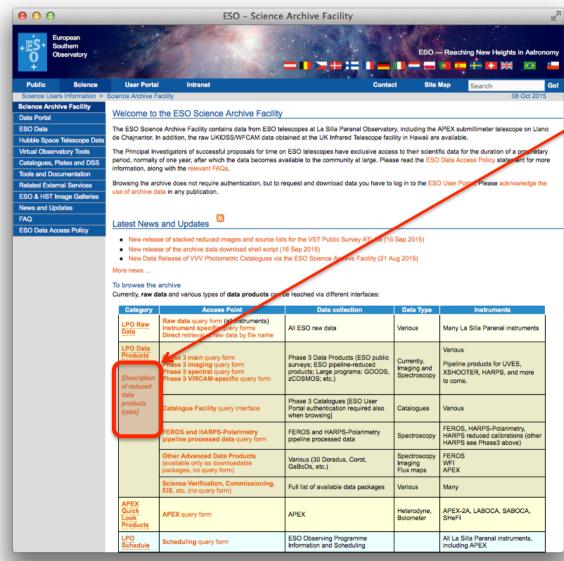
Science Catalogues

- Homogeneous merged multi-band catalogue for each survey (possibly per region)
- Global astrometry/photometry; cross-calibrated using overlapping tiles and across bands
- Multiple detections merged, i.e. unique entries (ultimately)
- Uniform tabular structure including content descriptors (employing UCDs)
- Supports variability surveys: multi-epoch photometric catalogues (i.e. light curves)
- Tile-by-tile scheme supported for data delivery
- Dedicated query interface: ESO Catalogue Facility

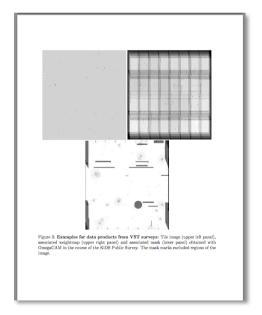


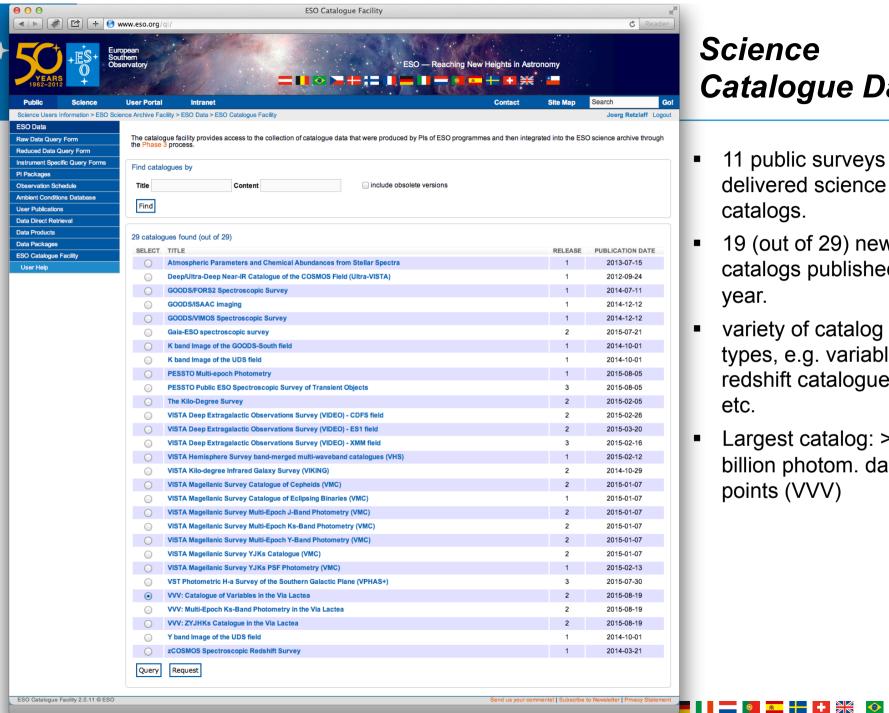
Science Archive Facility

http://archive.eso.org



New: Documentation of Reduced data products in the ESO Phase 3 archive





Science Catalogue Data

- 11 public surveys delivered science catalogs.
- 19 (out of 29) new catalogs published this year.
- variety of catalog types, e.g. variables, redshift catalogues etc.
- Largest catalog: >2.5 billion photom. data points (VVV)



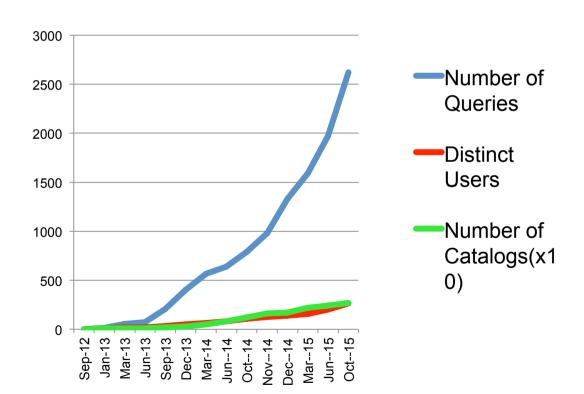
Catalogue Facility Query Interface





Access to catalog data

Access to the ESO archive catalogue query interface - All catalogues have been queried for!



Collaboration with CDS

- Based on the independent evaluation of catalogue data (format and metadata) and quality criteria by ESO/ ASG.
- CDS adds user value to the science catalogue content provided by ESO, i.e. by cross matching with other wavelengths/data, and further dissemination in the VO.



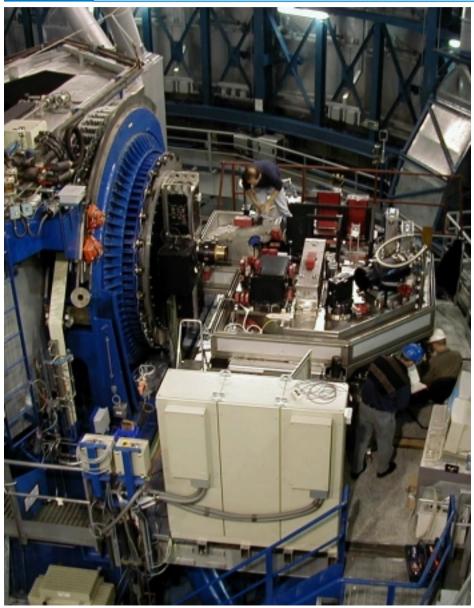


Origin of Data Products in the ESO/SAF

- 1 Data Processing by P.I./Co-I for a particular programme and delivery to the ESO archive (Public Surveys, Large programs)
- 2 Uniform processing of data from selected instruments/modes integrated with ESO data flow operations independent of the original science goals



Release of UVES-Echelle SDPs



- As of Sep 2013: Reduced scientific data obtained with the UVES@VLT available.
- Extracted wavelength- and, whenever possible, flux-calibrated 1-dim. spectra in tabular format following the established ESO SDP standard.
- Observations from the beginning of 2000 until now including the current stream of UVES data; reduced products becoming available in monthly release cycles
- Currently ~94000 spectra of pointlike sources.



Publication of ESO pipelineprocessed science data

- Production of science-grade data products for selected instruments & modes based on an evaluation of the quality of data reduction and legacy value of the respective data products.
- Phase 3-compliant fashion for seamless access of all DPs whether provided by PI's or processed by ESO.
- ESO Data Access Policy: Data products of normal and large programmes become publicly accessible once the proprietary period (usually one year) has expired; before that date, access to the data products is restricted to the P.I. of the observing programme and her/his collaborators.

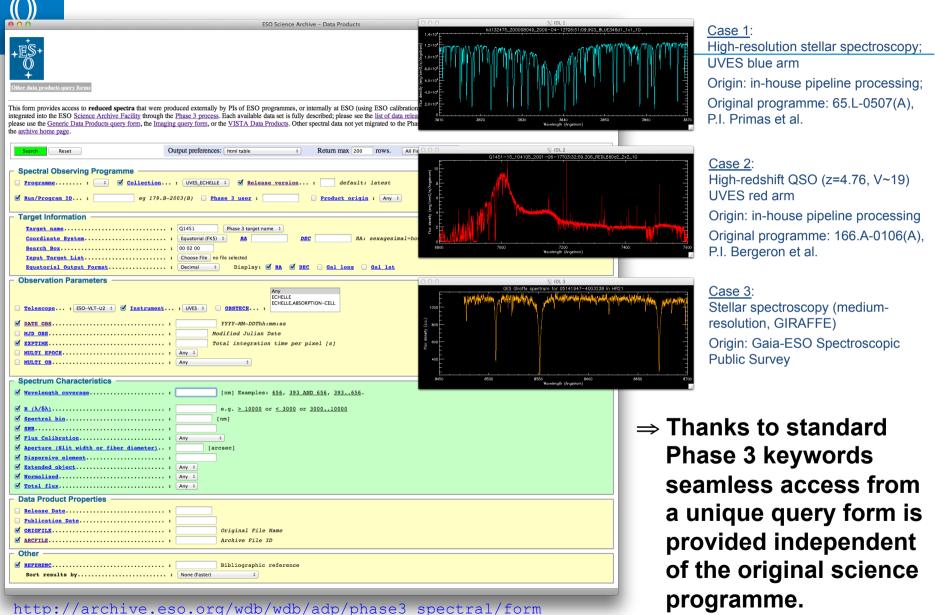
17

	No. of spectra	GB
UVES-Echelle	105,666	274
X-Shooter	46,487	855
HARPS	235,295	2,343
Giraffe-Medusa	1,130,268	5,260

=> Synergy effects due to one common archive for data produced by public surveys, large programs and ESO pipeline data reduction.



Spectral Science Data Products in the ESO Archive



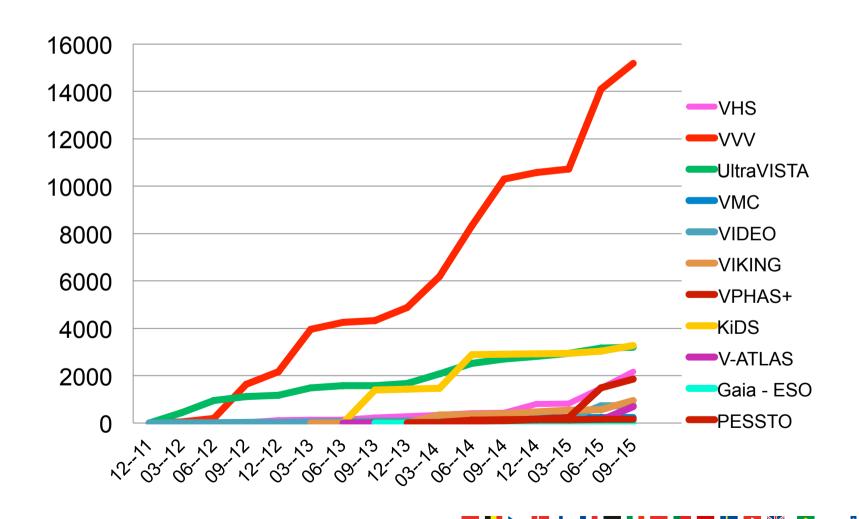


DATA ACCESS STATISTICS



Access per survey

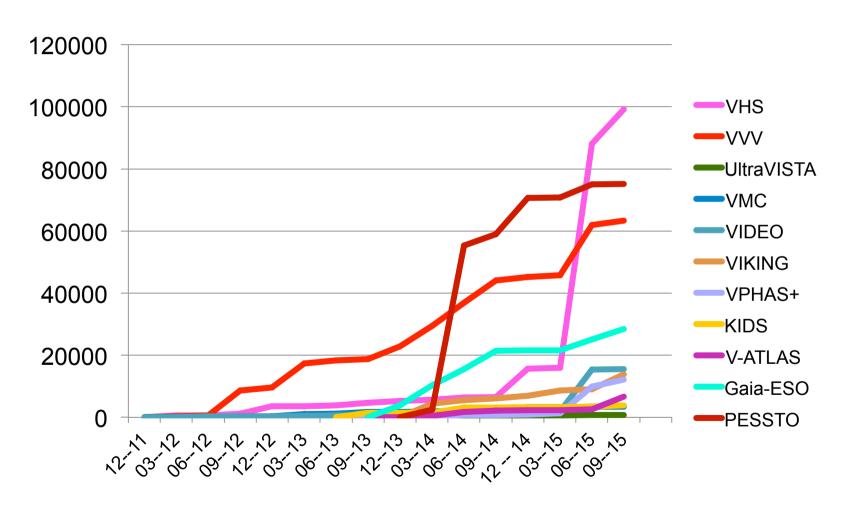
Data volume download from ESO SAF of public survey products





Access per survey (2)

Number of files download from ESO SAF of public survey products



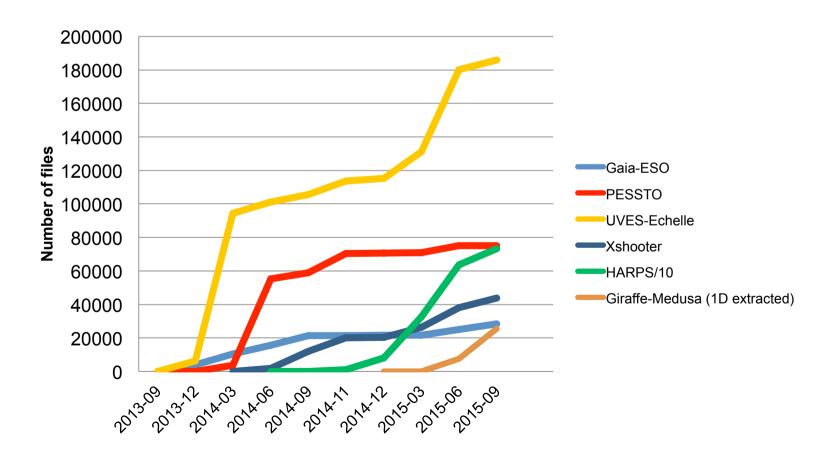




Access to Spectral Data

Spectroscopic surveys (Gaia-ESO, PESSTO)

Pipeline-processed (UVES-Echelle, Xshooter, HARPS, Giraffe-Medusa)

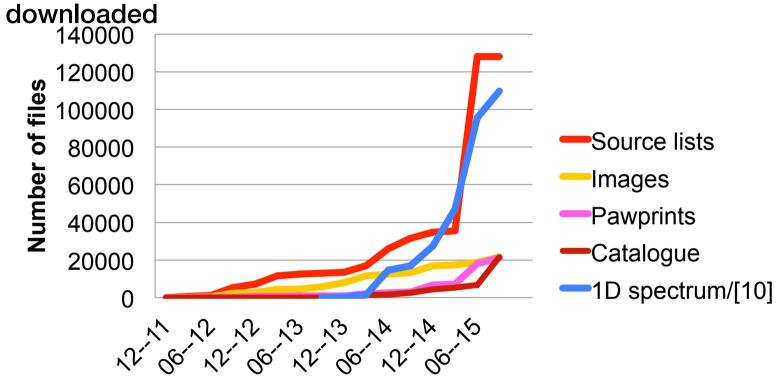






Access per data type

Number of files and products download from ESO SAF for public surveys & IDPs: all types of SDPs have been accessed/



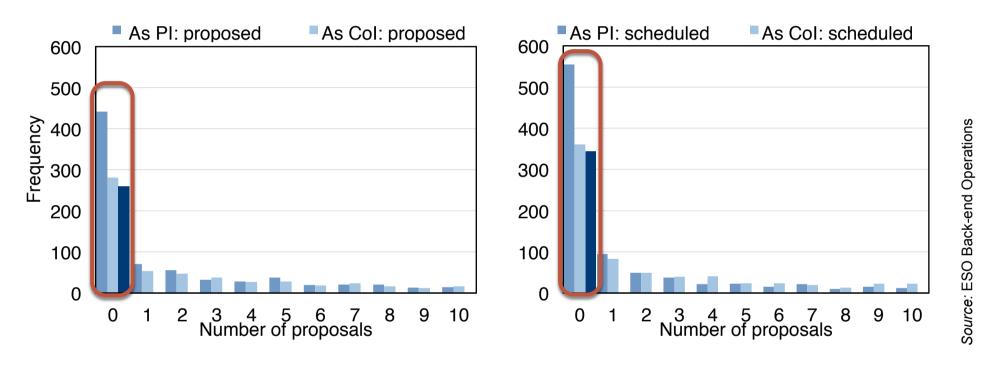
>25 TB and one million files downloaded since December 2011!





Phase 1 habits of archive users

Data products



- SAF data products users: 28% (260) have never applied for time, neither as PIs nor co-Is; 37% (344) have never gotten time, neither as PIs nor co-Is
 - For comparison, 1/3 of the Phase 1 PIs have never got time



How do we operate?

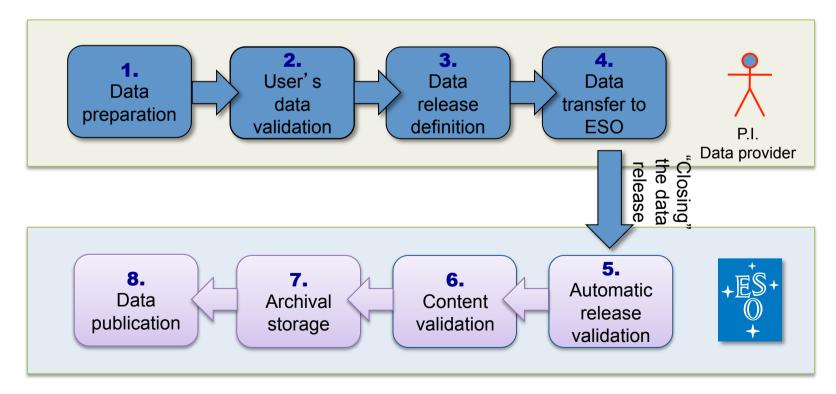
PHASE 3 IMPLEMENTATION



Phase 3 Implementation

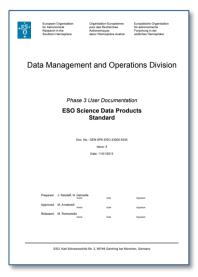
Phase 3 process: preparation, submission, validation and ingestion of science data products for storage in the ESO Science Archive Facility, and subsequent publication to the scientific community.

- The data provider i.e. survey P.I. is responsible for data reduction and final QC, delivering reduced data to ESO.
- ESO is responsible for the definition of the data products standard, archiving the data and publication to the community.





Phase 3 Data preparation



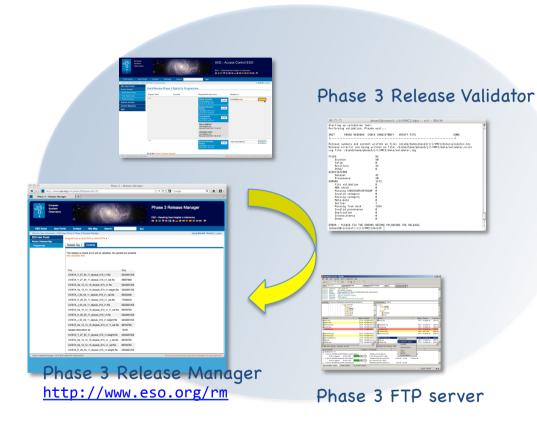
Phase 3 data must adhere to the **ESO**Science Data Product Standard regarding format and required metadata, i.e. relevant keywords for data characterization, quality, processing provenance etc.

http://www.eso.org/sci/observing/phase3/p3sdpstd.pdf

- Uniformity of products across the ESO/SAF in terms of data format, level of characterization, search capabilities, and data documentation.
- Current Issue covers all PS data products, imaging, spectroscopy; Sub-mm flux maps (e.g. APEX/LABOCA)
- Format for IFU data cubes (MUSE etc.) available



Phase 3 Key Features



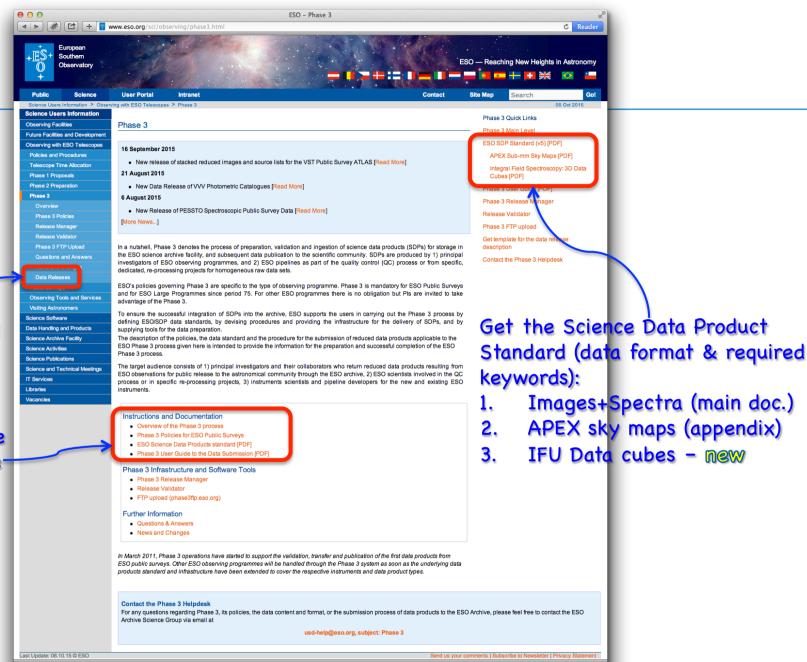
- Phase 3 User Delegation
- Support for incremental and superseding data releases.
- Association of ancillary data files with science data products
- Processing provenance to trace original science data
- Data documentation following standard structure ("releasedescription.pdf")

Phase 3 content validation includes consistency checks of submitted data content w.r.t. documentation, check of completeness and quality of the metadata (keywords) prior to archive ingestion.



Overview of data published via Phase 3

How to prepare and submit the data..







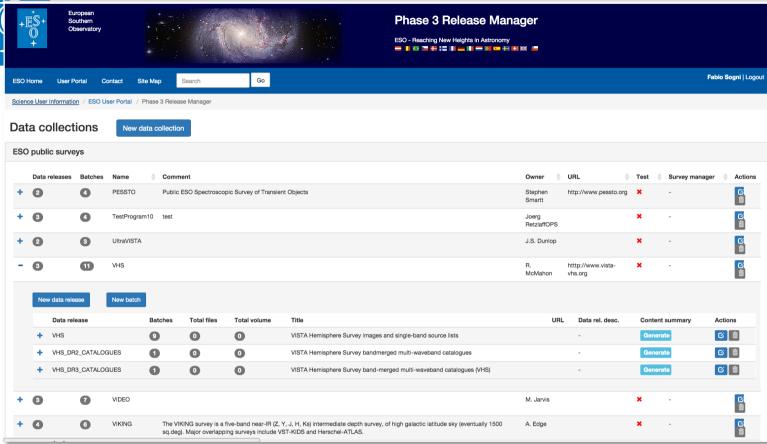
Phase 3 Upgrade

Main Goals

- Minimizing the overall time to data publication
 - Optimize the feedback loop with the data provider regarding the data validation status
 - and uploaded data content
 - Reducing workload for data providers as much as possible
- Adding missing operational requirements
 - Better support for very large data releases using submission via multiple 'batches'
 - New Survey manager role
 - Improved communication and tracking of data-related issues between data providers and ESO ("Phase 3 RIX'es mgt.")

+ES+

New Phase 3 Release Manager



- Phase 3 Programme merged with "Data Collection"
- Local validation fully replaced by server side validation; validation on demand also possible
- Flexible management of Data Releases by ESO
- Up-to-date statistics of uploaded data (content summary)
- New Survey Manager role for ESO Public Surveys
- Flexible management of Release Documentation (Description) via the new Release Manager interface



New Content Summary – 1

Content summary of Data Release VHS

All files summary

Data collection: VHS
Data Release: VHS
Total files: 568088
Total size: 25,559.40 GB

All files by category

Category	#files
ANCILLARY.WEIGHTMAP	189367
SCIENCE.IMAGE	26861
SCIENCE.MEFIMAGE	162506
SCIENCE.SRCTBL	189354

All files by filename extension

Filename extension	#files
fits	189354
fits.fz	378734

All files by filename extension and category

Filename extension	Category	#files	Size (GB)
fits.fz	ANCILLARY.WEIGHTMAP	189367	6,837.23
fits	SCIENCE.SRCTBL	189354	1,872.82
fits.fz	SCIENCE.MEFIMAGE	162506	11,873.81
fits.fz	SCIENCE.IMAGE	26861	4,975.54

+ES+ Science

New Content Summary – 2

Science files summary

Total files: 378721

Start period:

End period:

Start date: Wed Nov 04 00:29:13 UTC 2009 **End date:** Sat Oct 01 04:35:41 UTC 2011

Sky area:

Science files by PRODCATG

PRODCATG	#files
SCIENCE.IMAGE	9943
SCIENCE.MEFIMAGE	60231
SCIENCE.SRCTBL	70139

Science files by OBSTECH

OBSTECH	#files
IMAGE,JITTER	140313

Science files by INSTRUMENT

Instrument	#files
VIRCAM	140313

Science files by MULTOB

Multob	#files
false	
true	

Science files by FILTER

Filter	#files
Н	27658
J	49424
Ks	48632
Υ	14599

Science files by ESO period

Period	#files
TODO	0

Science files by PROCSOFT

PROCSOFT	#files
TODO	0







Summary

- The Science Archive Facility is the central repository for ESO Public Survey data products.
- >35 TB of data from 11 public surveys are currently available for download.
- The SAF contains science data products including calibrated images, spectra, source lists and high-level catalogues.
- These data are complemented by pipeline-reduced science-grade data products for selected instruments/modes processed by ESO and by other data deliveries, e.g. by large programmes.
- Download statistics demonstrate a large ongoing interest by the user community.
- Offering PS data and archival data in general extends ESO's user community.





Many Thanks to PS and LP teams for having returned their data products!

ESO/ASG is ready to support your data submission. Please contact the Phase 3 helpdesk at usd-help@eso.org, subject: Phase 3