

Data Products in the ESO Science Archive

Jörg Retzlaff

Archive Science Group (ASG)

Data Management and Operations Division

European Southern Observatory

M. Arnaboldi, N. Delmotte, V. Forchi, M. Klein Gebbinck, S. Geier, J. Lockhart, A. Micol, C. Moins, I. Vera Sequeiros, , S. Zampieri





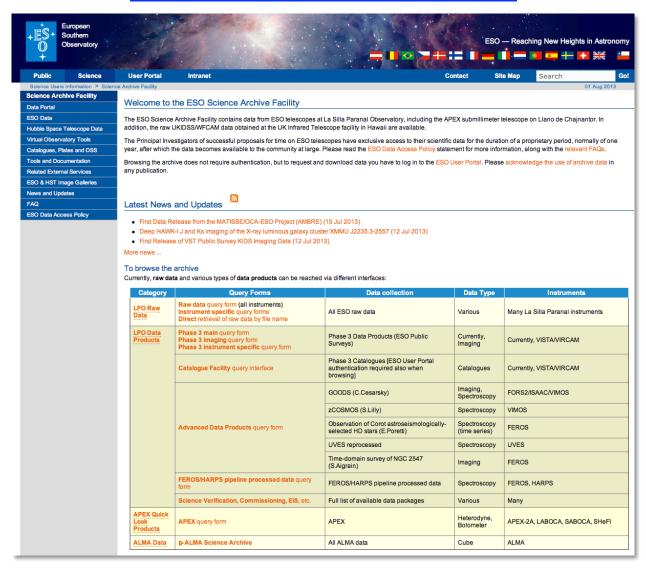
Data Products in the ESO Science Archive

- 1. Archive Contents
- 2. Phase 3 Operations
- 3. Data Releases
- 4. Internal Data Products



ESO Science Archive Facility (SAF)

http://archive.eso.org



ESO Archive Contents

Category	Query Forms	Data collection	Data Type	Instruments		
LPO Raw Data	Raw data query form (all instruments) Instrument specific query forms Direct retrieval of raw data by file name	All ESO raw data	Various	Many La Silla Paranal instruments		
LPO Data Products	Phase 3 main query form Phase 3 imaging query form Phase 3 instrument specific query form	Phase 3 Data Products (ESO Public Surveys) Currently, Imaging		Currently, VISTA/VIRCAM		
	Catalogue Facility query interface	Phase 3 Catalogues [ESO User Portal authentication required also when browsing]	Catalogues	Currently, VISTA/VIRCAM		
		GOODS (C.Cesarsky)	Imaging, Spectroscopy	FORS2/ISAAC/VIMOS		
		zCOSMOS (S.Lilly)	Spectroscopy	VIMOS		
	Advanced Data Products query form	Observation of Corot astroseismologically- selected HD stars (E.Poretti)	Spectroscopy (time series)	FEROS		
		UVES reprocessed	Spectroscopy	UVES		
		Time-domain survey of NGC 2547 (S.Aigrain)	Imaging	FEROS		
	FEROS/HARPS pipeline processed data query form	FEROS/HARPS pipeline processed data	Spectroscopy	FEROS, HARPS		
	Science Verification, Commissioning, EIS, etc.	Full list of available data packages	Various	Many		
APEX Quick Look Products	APEX query form	APEX	Heterodyne, Bolometer	APEX-2A, LABOCA, SABOCA, SHeFI		
ALMA Data	BALMA Science Archive	All ALMA data	Cube	ALMA		

Total volume: >300 TB predominantly raw data (rate in: 3.6TB/month)

ESO Archive Contents

		5 () ()				
Category	Query Forms	Data collection	Data Type	Instruments		
LPO Raw Data	Raw data query form (all instruments) Instrument specific query forms Direct retrieval of raw data by file name	All ESO raw data	Various	Many La Silla Paranal instruments		
LPO Data Products	Phase 3 main query form Phase 3 imaging query form Phase 3 instrument specific query form	Phase 3 Data Products (ESO Public Surveys)	Currently, Imaging	Currently, VISTA/VIRCAM		
	Catalogue Facility query interface	Phase 3 Catalogues [ESO User Portal authentication required also when browsing]	Catalogues	Currently, VISTA/VIRCAM		
		GOODS (C.Cesarsky) Imaging, Spectroscopy		FORS2/ISAAC/VIMOS		
	Advanced Data Products query form	zCOSMOS (S.Lilly)	Spectroscopy	VIMOS		
		Observation of Corot astroseismologically- selected HD stars (E.Poretti)	Spectroscopy (time series)	FEROS		
		UVES reprocessed	Spectroscopy	UVES		
		Time-domain survey of NGC 2547 (S.Aigrain)	Imaging	FEROS		
	EROS/HARPS pipeline processed data query form FEROS/HARPS pipeline processed data		Spectroscopy	FEROS, HARPS		
	Science Verification, Commissioning, EIS, etc.	Full list of available data packages	Various	Many		
APEX Quick Look Products	APEX query form	APEX	Heterodyne, Bolometer	APEX-2A, LABOCA, SABOCA, SHeFI		
ALMA Data	BALMA Science Archive	All ALMA data	Cube	ALMA		

Data Products

ESO Archive Contents

Category	Query Forms	Data collection	Data Type	Instruments
LPO Raw Data	Raw data query form (all instruments) Instrument specific query forms Direct retrieval of raw data by file name	All ESO raw data	Various	Many La Silla Paranal instruments
LPO Data Products	Phase 3 main query form Phase 3 imaging query form Phase 3 instrument specific query form	Phase 3 Data Products (ESO Public Surveys)	Currently, Imaging	Currently, VISTA/VIRCAM
	Catalogue Facility query interface	Phase 3 Catalogues [ESO User Portal authentication required also when browsing]	Catalogues	Currently, VISTA/VIRCAM
		GOODS (C.Cesarsky)		
		zCOSMOS (S.Lilly)	Spectroscopy	VIMOS
	Advanced Data Products query form	Observation of Corot astroseismologically- selected HD stars (E.Poretti)	Spectroscopy (time series)	FEROS
		UVES reprocessed	Spectroscopy	UVES
		Time-domain survey of NGC 2547 (S.Aigrain)	Imaging	FEROS
	FEROS/HARPS pipeline processed data query form	FEROS/HARPS pipeline processed data	Spectroscopy	FEROS, HARPS
	Science Verification, Commissioning, EIS, etc.	erification, Commissioning, EIS, etc. Full list of available data packages		Many
APEX Quick Look Products	APEX query form	APEX	Heterodyne, Bolometer	APEX-2A, LABOCA, SABOCA, SHeFI
ALMA Data	BALMA Science Archive	All ALMA data	Cube	ALMA

Phase 3 data represent the largest collection of data products in the ESO archive:

In total: 13 TB, >230k files, and growing...

Science Data Products – Characteristics and Requirements

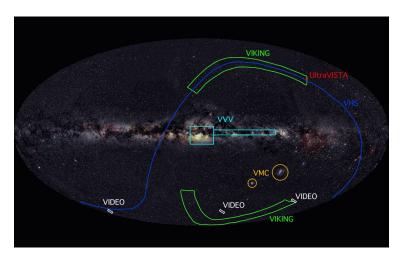
- Science-grade data, i.e. "ready" for scientific exploitation— to enable the users to do what they want.
- Instrumental signatures to be removed/corrected for;
- Data calibrated to physical units;
- Data uncertainties/errors quantified.
- Must adhere to the ESO Science Data Product Standard regarding data and metadata format (ie. required keywords)
 - Uniformity of products across the ESO/SAF in terms of data format, level of characterization, search capabilities, and data documentation.

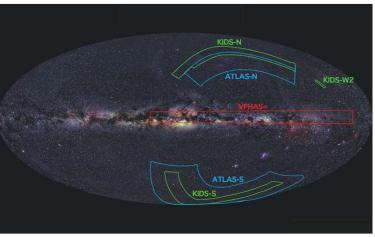


ESO Archive Data Products

PHASE 3 OPERATIONS

ESO Public Surveys





VISTA: 6 surveys started Apr 2010 (P85)

- VHS 20000 deq² YJHKs (Ks<20 AB)
- VIKING 1500 deg² ZYJHKs (Ks<21.2 AB)
- VIDEO 3 Deep Extragalactic Fields
- Ultra-VISTA Ultra-deep ZYJHKs + NB118 in the COSMOS field
- VVV Variability study of 520 deg² in bulge +plane plus multi-color map
- VMC Magellanic Survey

VLT Survey Telescope: 3 surveys, started 15 Oct 2011 (P88)

- VST-Atlas 4500 deg² UVRIz, like SDSS
- KIDS 1500 deg² UVRI, 2.5 mag deeper than SDSS
- VPHAS+ 1800 deg² UVHaRI in the Southern Galactic Plane

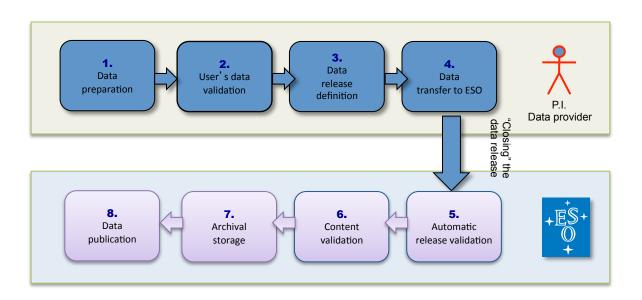
Spectroscopic Surveys: started 1 Jan 2012

- PESSTO: 30+60n on NTT (SOFI+EFOSC)
- Gaia-ESO: 30+30n on VLT-UT2 (FLAMES)

The Phase 3 process has been setup in 2011 to implement ESO's policies for public surveys.

Phase 3 Process & Responsibilities

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



ESO defines the required data format, provides dedicated tools, user documentation and direct support for Phase 3 data providers.

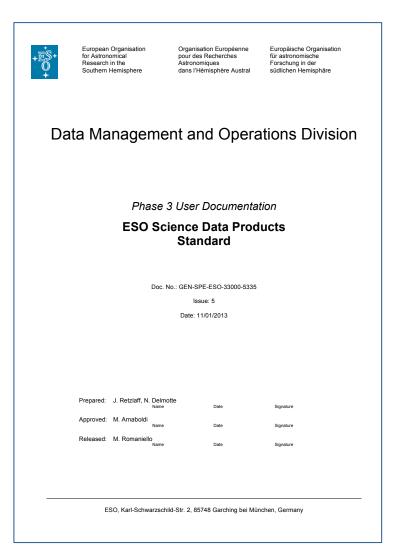
The data provider ie. survey P.I. is responsible for the quality of the reduced data products and the associated data release documentation.

Content validation: audit of submitted data, including checks of

- consistency of data content against documentation,
- format of data and metadata.
- completeness and quality of the metadata (keywords).

http://www.eso.org/sci/observing/phase3.html

ESO Science Data Products Standard



Identifies product types, their data and metadata formats depending on instrument/mode.

Defines relevant keywords for data characterization, quality, processing provenance (history) etc.

- Supports archive operations, setup of query forms, data request etc.
- To minimize (meta)data curation during archive ingestion.
- Current Issue: #5, Date 11/01/2013 covers all PS data products: imaging & spectroscopy

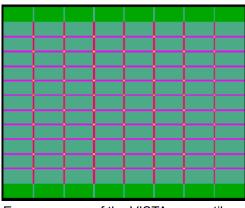
Types of Data Products

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**

Phase 3 System Components



The **Phase 3 Validator** is a command-line application that verifies the data's compliance with the format standard and the validity of the FITS header keywords against predefined rules.

Start of operations: March 2011



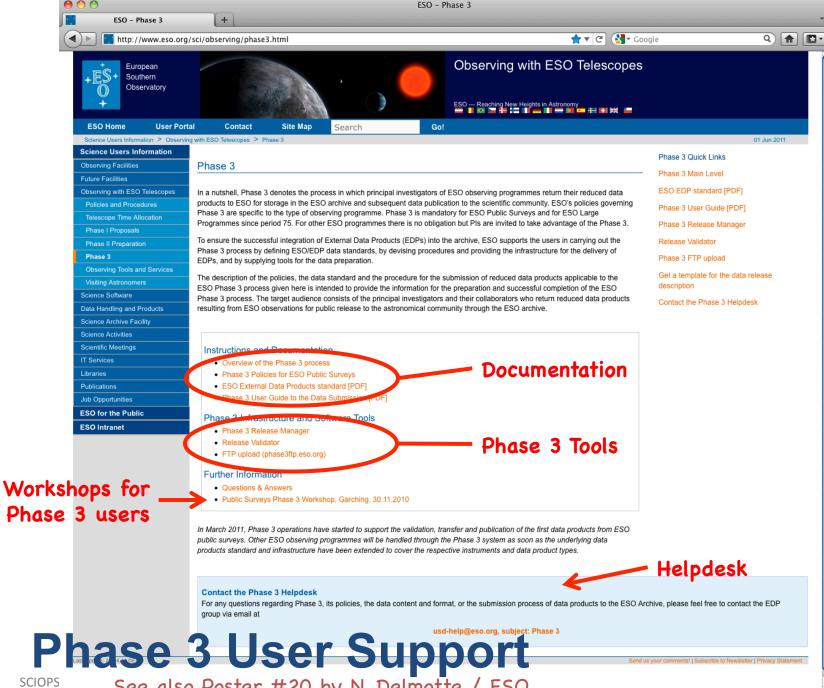


Any FTP client like lftp, filezilla etc. Note: the requirement for SSL/TLS was dropped (Jun 2012).

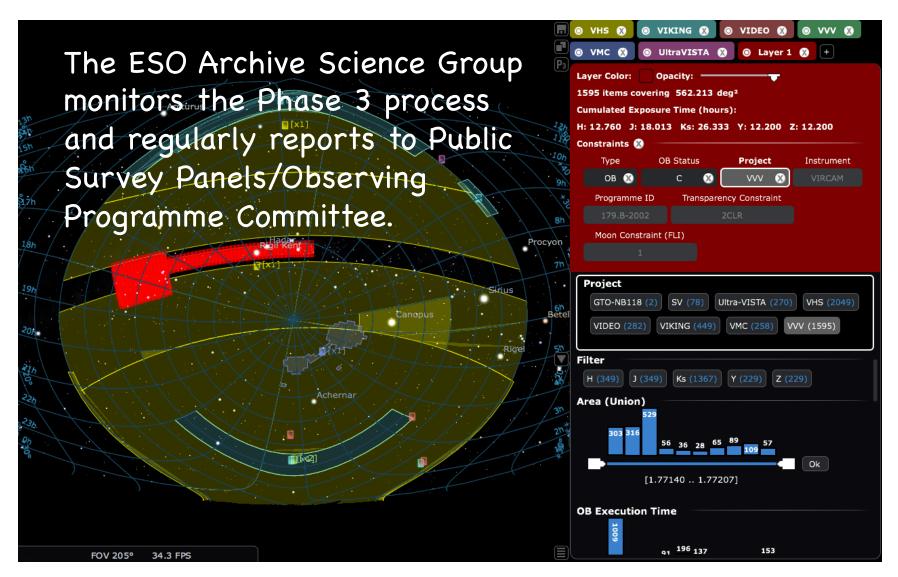
The **Phase 3 Release Manager** is a web application that allows the P.I. to define data collections and releases and to manage the Phase 3 <u>delegation</u> to co-investigators.

http://www.eso.org/rm

The P.I. or a delegate transfers the data via standard FTP to the dedicated staging area at ESO headquarters.



Monitoring the Phase 3 Progress





ESO Public Survey Data Products

PHASE 3 DATA RELEASES

Status of Phase 3 Data Releases

http://www.eso.org/sci/observing/data_releases.html

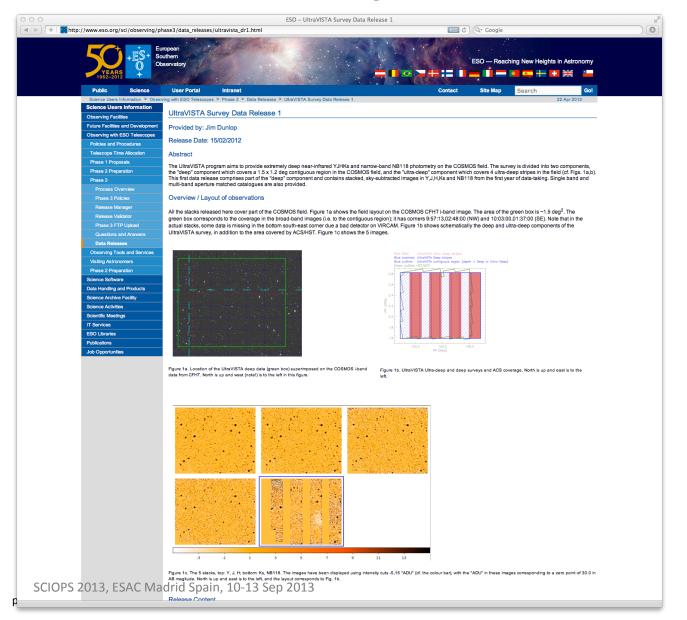
Data Release	Data Type	Release Type	Release Date
VVV: VISTA Variables in the Via Lactea Survey - Data Release 1	Image	obsolete	25.07.2011
VIDEO: VISTA Deep Extragalactic Observations Survey - Data Release 1	Image	latest	25.07.2011
VMC: VISTA Magellanic Survey - Data Release 1	Image	obsolete	25.09.2011
VHS: VISTA Hemisphere Survey - Data Release 1	Image	obsolete	17.10.2011
Ultra-VISTA: an Ultra Deep Survey with VISTA - Data Release 1	Image	latest	15.02.2012
Ultra-VISTA Catalogue - Data Release 1	Catalogue	latest	24.09.2012
VVV: VISTA Variables in the Via Lactea Survey - Data Release 2	Image	latest	28.06.2013
VHS: VISTA Hemisphere Survey - Data Release 2 (*)	Image	latest	28.06.2013
VIKING: VISTA Kilo-degree Infrared Galaxy Survey - Data Release 1	Image	latest	28.06.2013
VMC: VISTA Magellanic Survey - Data Release 2	Image	latest	28.06.2013
KiDS: the Kilo Degree Survey - Data Release 1	Image	latest	11.07.2013
60.A-9284H: XMMU_J2235_JK_imaging - Data Release 1	Image	latest	11.07.2013
MATISSE/OCA-ESO Project (AMBRE) - Data Release 1	Catalogue	latest	15.07.2013
VPHAS+: VST Photometric Ha Survey of the Southern Galactic Plane and Bulge - Data Release 1	Image	latest	20.08.2013

(*) Note: Phase 3 supports *incremental* and *superseding* data releases.

Link to the associated data documentation

Link to the data

ESO Phase 3 – Data Release Documentation Example: Ultra-VISTA



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 -

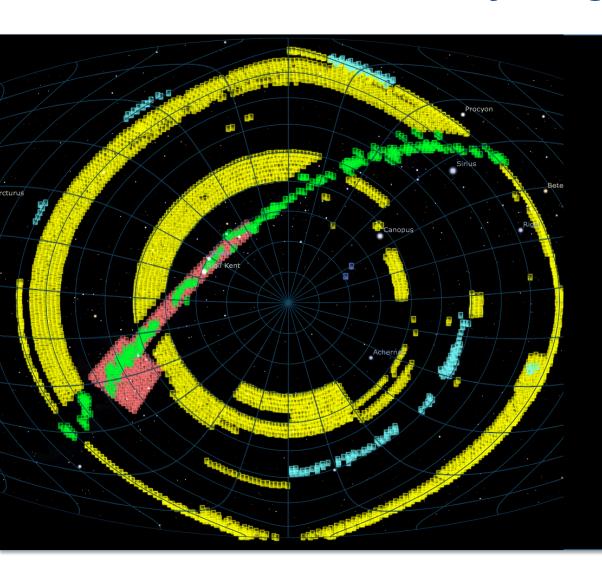
 Bibliographic reference to be included when using these data.

Data Releases from Public Imaging Surveys (Aug. 2013)

Survey	Submission Date	Date of Observatio ns	Release Content	Pass- bands	Sky coverage (sq.deg)	Type of Data Products	Total volume	Total number of files	VISTA tile images	Pub. date
VVV	06.05.2013	Feb 2010 – Oct 2012	Contiguous patch of bulge and disk region including multi-epoch data in Ks	ZYJHKs	~520 (348 tiles)	Tiles, single-band source lists	2.8 TB	~16000	5327	28.06.2013
VIDEO	03.05.2011	Nov 2009 – Feb 2010	XMM-LSS field	YJHKs	1.5	Tiles, single-band source lists	24 GB	291	97	25.07.2011
VMC	03.06.2013	Nov 2009 – Nov 2010	2 pointings in the LMC: one overlapping the 30 Doradus and the other the South Ecliptic Pole region	YJKs	3.5	Stacked tiles, single- OB tiles and pawprints, single- band source lists	27 GB	252	50	28.06.2013
VHS	25.02.2013	Nov 2009 – Sep 2011	VHS DES: 120 secs in JHK VHS ATLAS: 60 secs in YJHK VHS GPS: 60 secs in JK	YJHKs	~3800	Tiles, pawprints, single-band source lists	8.3 TB	200000	10000	28.06.2013
UltraVISTA	06.10.2011 - 30.01.2012	Dec 2009 – Apr 2010	Deep imaging of the COSMOS field	YJHKs, NB118	1.5	Stacked images, SExtractor catalogues including Ks-selected multi-band catalogue	87 GB	19	5	15.02.2012
VIKING	20.05.2013	Nov 2009- Feb 2011	Sub-areas overlapping with the GAMA09/12/14, SGP stripe and CFHTLS-W1 regions	ZYJHKs	226	Tiles, single-band source lists	288 GB	7000	151	28.06.2013

Survey	Submission Date	Date of Observatio ns	Release Content	Pass- bands	Sky coverage (sq.deg)	Type of Data Products	Total volume	Total number of files	VST tile images	Pub. date
KiDS	28.06.2013	Oct 2011– Sep 2012	Non-contiguous area in KiDS South and North	u, g,r,i	50	Tiles, weight maps and mask images, single- band source lists	700 GB	800	200	11.07.2013
VPHAS+	16.05.2013	Dec 2011 – Sep 2012	Bulge and disk of the MW	u, g,r,i Hα	334	Pawprints, single-band source lists	750 GB	291	3489	20.08.2013

Status of ESO Survey Imaging Products



ESO Public Survey data products released up to Aug'13 cover about 4900 square degrees in total.

- VHS (yellow) 4210 sq.deg.
- VIKING (cyan) 235 sq.deg.
- VVV (red) 563 sq.deg.
- VMC (blue) 3.6 sq.deg
- VIDEO (light green) 1.8 sq.deg.
- VPHAS+ (green) 240 sq.deg. UltraVISTA, KIDS not shown.

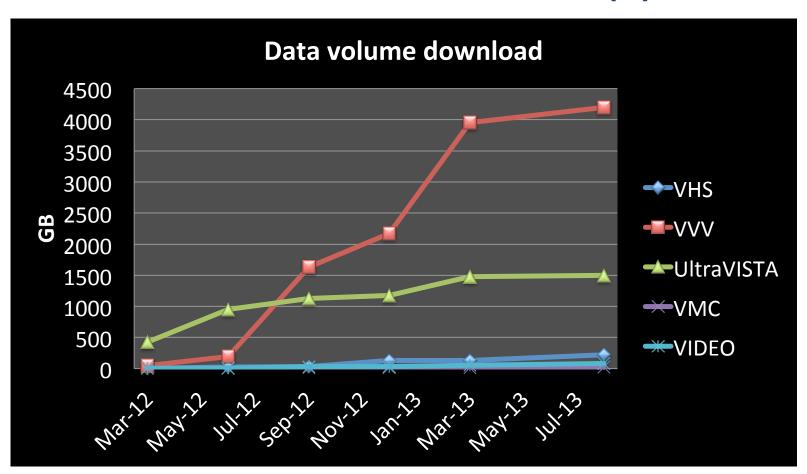
Total data volume: ~13 TB, 230k+ files

Accessing Phase 3 Data: Query Forms

Imaging observing programme								
Any VVV VIDEO VMC VHS Any VVV VVDEO_XMM3 VMC VHS Any VVV VIDEO_XMM3 VMC VHS Release version:	default: latest							
▼ Run/Program ID: PPP.C-NNNN(R) (eg 179.B-2003) □ Phase3 user:								
Target Information SIMBAD name SIMBAD name								
	urs, decimal/degrees							
Search Box: 02 09 00 Equatorial Output Format Sexagesimal \$								
Input Target List: Choose File no file selected								
Observation Parameters								
Any ESO-VISTA ESO-VIT-U4 ESO-VST Telescope: Any VIRCAM HAWKI OMEGACAM OMEGACAM OMEGACAM OMEGACAM OBSTECH: Any IMAGE,JITTER IMAGE IMAGE IMAGE,DITHER Filter:								
DATE OBS. : UT in YYYY-MM-DD HH:MM:SS format								
✓ MJD OBS: Modified Julian Date ✓ EXPTIME: Total integration time per pixel [sec]								
▼ TEXPTIME: Total Exposure Time [sec]								
MULTI EPOCH : Any ÷								
<u>MULTI OB</u> : Any ÷								
Any SCIENCE.IMAGE SCIENCE.SRCTBL SCIENCE.MAGE SCIENCE.SRCTBL SCIENCE.MEFIMAGE Data Product Format Sky coverage sky solid angle [deg^2] e.g. 0.61.8 ISAMP Any Sky Sampling Spatial Sampling [arcsec] ABMAGGIM Limiting Depth [mag] e.g. 2022.5 ABMAGSAT Saturation Limit [mag] e.g. 13.515	 Supports queries by high-level quality parameters: MAGLIM PSF_FWHM For spectra (2013Q4): 							
✓ PSF_FWEM. : Spatial Resolution [arcsec] (e.g. 0.91.1) ORIGFILE. : Original File Name ✓ ARCFILE. : Archive File ID	By wavelength (range),RP, SNR etc.							

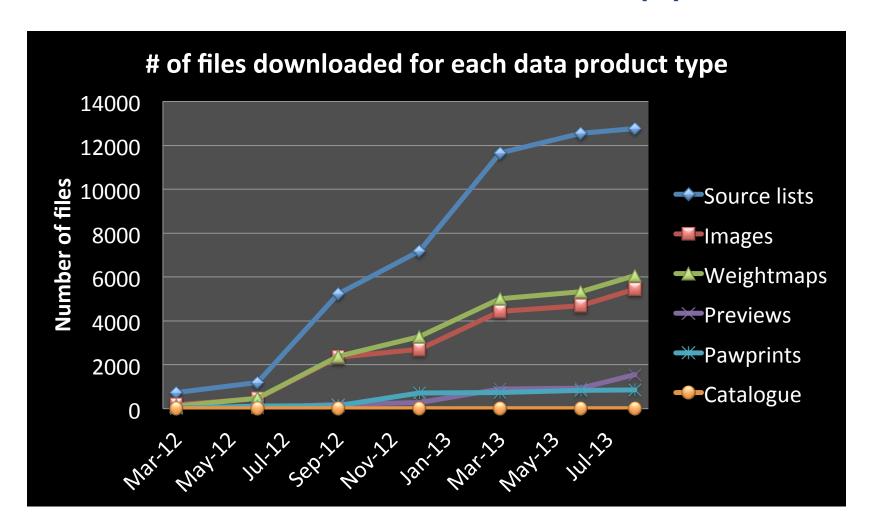
http://archive.eso.org/wdb/wdb/adp/phase3_main/form

ESO SAF Data Access Statistics (1)



6.8 TB of data products and more than 27000 files downloaded from the ESO SAF since December 2011

Data Access Statistics (2)



➤ High popularity of high-level products (source lists = single-band tile-based catalogues)



New data resulting from the MATISSE/OCA-ESO Project (AMBRE) released on 15.07.2013

- In collaboration between the Observatoire de la Côte d'Azur (OCA, Nice) and ESO, stellar radial velocities, atmospheric parameters including effective temperature, surface gravity, mean metallicity and enrichment in alpha-elements have been obtained by analyzing 21551 FEROS spectra from the ESO data archive with the MATISSE parametrization algorithm
- Available through the ESO/SAF Catalogue Facility query interface

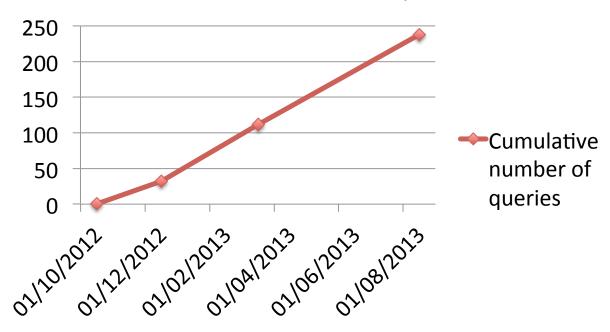
Catalogue Facility Query Interface



See also: poster #26

- Serves science catalogue data
- Flexible: source catalogues, photometric light curves, redshift catalogues, stellar parameters,...
- Integrated with the ESO
 Phase 3 system (process, responsibilities, data standard)
- Scalability: supports large data volumes as resulting from large-area surveys.
- Deployed in 2012. Example for the further development of ESO/SAF services.

Evolution of CF User Queries



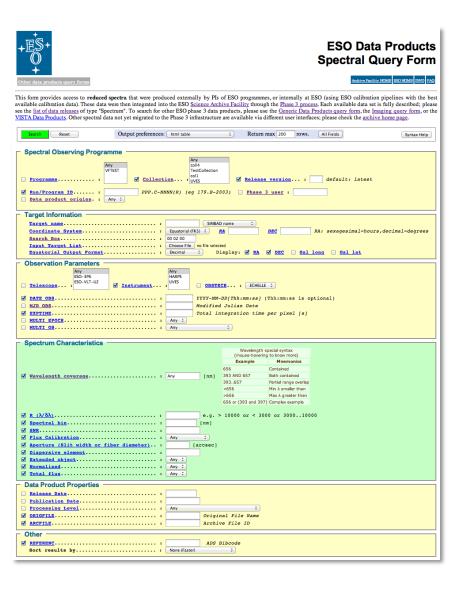
Current Status of Phase 3 Data Submissions

- VST-ATLAS: 3 TB submitted- Phase 3 content validation on-going
- VMC: updated band-merged catalogue data submitted
- VIKING: 14.5+ million sources band-merged catalogue data revised
- VIDEO: feedback on content validation provided
- VVV: band-merged catalogue data submission revised (11/09/2013), in content validation phase

Upcoming: Internal Data Products (IDPs)

- Why? The Users Committee recommended [UC36.R.3]: The ESO data archive should contain calibrated data where at least instrumental signatures are removed to increase the value of the archive for the ESO users.
- Plan: Science-grade DPs being produced by ESO in-house and integrated in Phase 3-compliant fashion for seamless access of internal and external data products through the ESO archive user interfaces.
- To enable IDP publication the existing Phase 3 system has been extended:
 - Concept of the Phase 3 Data Stream added: content grows over time; publication on fileby-file basis; supports proprietary data (according to ESO's data access policy)
- <u>Development</u> started in April '13. Phase 3 System to be upgraded on 16 Sep.
- The ESO/SDP Standard plays a pivotal role! Output format of ESO data reduction pipelines must comply (or conversion must be done).
- Operations: Data are being processed by the ESO/Quality Control group as part of the data flow operations.
- First IDPs to be integrated: UVES/Echelle data archive (ca. 91,000 spectra) + the current data stream; see also Poster #11 by R.
 Hanuschik / ESO

Internal Data Products (cont'd)



The Phase 3 upgrade includes query forms, one dedicated to spectra, to support both, the upcoming spectroscopic PS data and IDPs.

Further planning:

- To extend the production of IDPs to more (existing) instruments and modes;
- To ensure that upcoming instrument pipelines are "IDP ready"

Next steps (preliminary):

- UVES slicer: processing to start once batch processing of UVES/ Echelle point-like data is completed.
- GIRAFFE-MEDUSA: data products of science quality appropriate for archive publishing exist for 2004-04 until 2011-10. Pipeline upgrade planned.
- X-Shooter/Echelle: pipeline upgrade to be compliant to SDP standard planned for next year.
- HARPS: specification for the data conversion into SDP standards-compliant form currently on-going.
- HAWK-I and VIMOS imaging: delivery of compliant data products in October 2014 by the UK in-kind team.
- FORS2 and VIMOS spectroscopy: SDP-compliant pipeline planned. Processing of backlog and stream thereafter.
- In parallel: extension of the ESO/SDP standard to data cubes to support IFU data (KMOS, MUSE, as well as GIRAFFE-Argus); and the extension to interferometric data

Conclusions

- ☐ Transition of the ESO archive is currently taking place: from a repository of predominantly raw data into a scientific resource of immediately utilizable data products.
- ☐ The ESO Science Archive Facility is being populated with science data products including calibrated images, spectra, and catalogues
- ☐ Currently, predominantly resulting from ESO Public Surveys; complemented by other programmes
- Regular production and publication of science data products for selected instrument modes is about to start (with UVES/Echelle data first)
- Next: VST-Atlas DR1, Spectroscopic Public Surveys, VISTA band-merged catalogues (VMC, VIKING, VVV, VHS), VISTA DR3, LPs, VIMOS SPSs, more IDPs.