



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

ESO VISTA Public Surveys: Data Product Types and Formats

J. Retzlaff (EDP)



Outline

- Phase 3 User support
- Process & Responsibilities
- External Data Products Standard
 - VISTA Data product types
 - EDP keyword definitions
 - The Data release description
- Data format for Catalogues



Phase 3 User support

Contact the EDP team via email:

usd-help@eso.org, subject: EDP-ADP Submission

Phase 3 web page at

<http://www.eso.org/sci/observing/phase3>

- Access to the Release Manager
- Phase 3 Validator

ESO Phase 3 user documentation

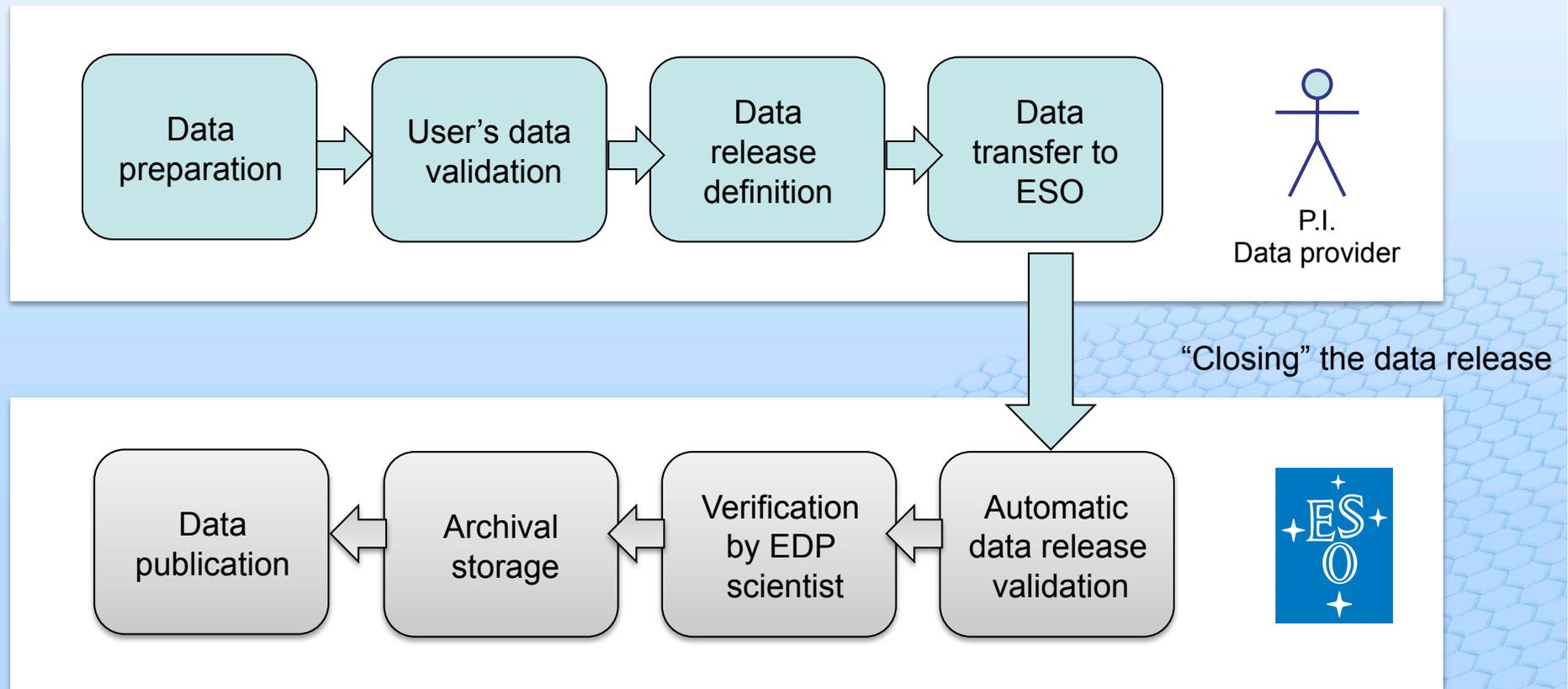
- Policies
- EDP standard (file format, keywords)
- Data submission user guide



EDP scientist's functions:

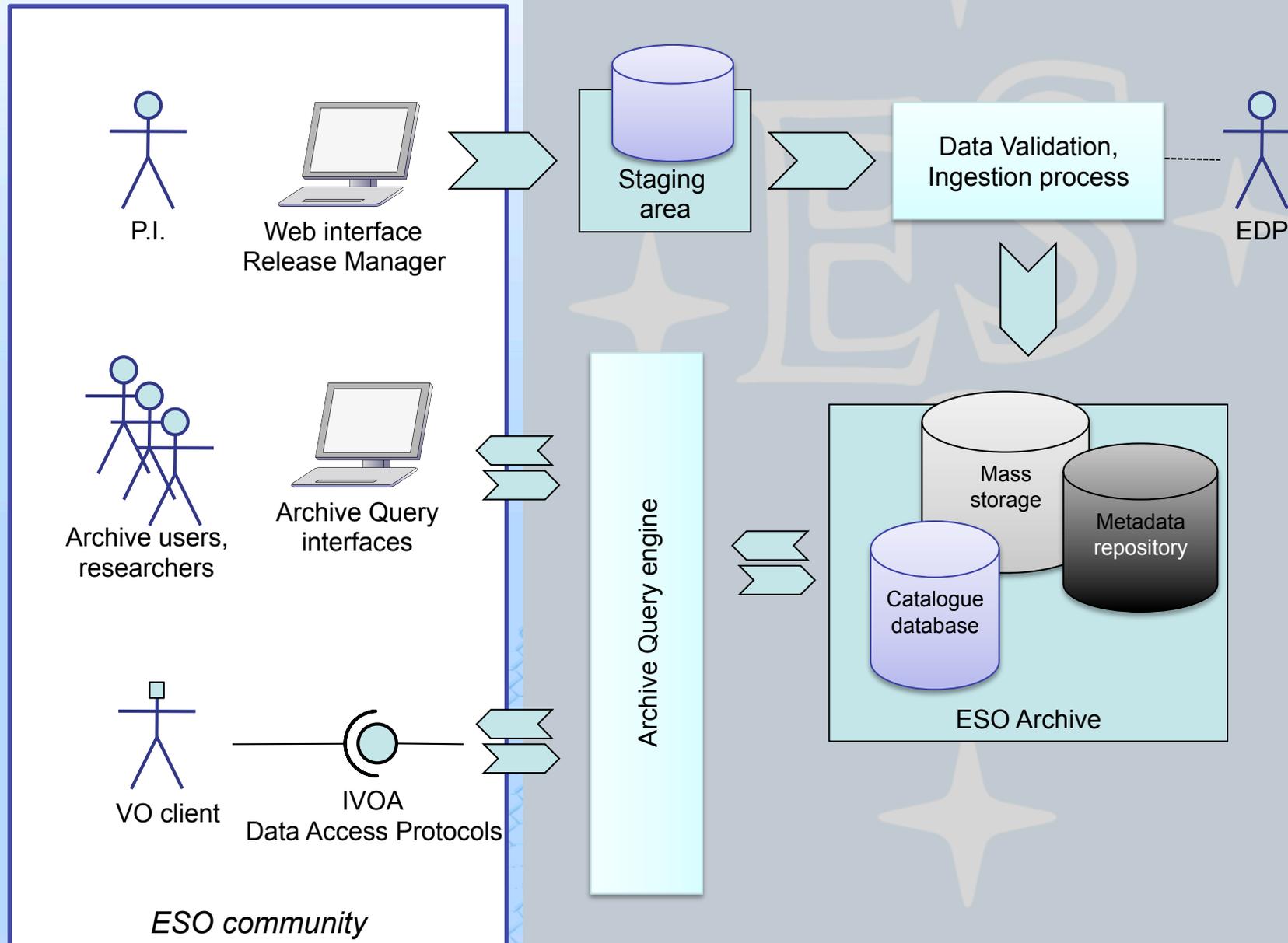
- Phase 3 User support for VISTA Public Surveys
- Checking data submissions with respect to content & format, e.g. against
 - ESO/EDP data standard
 - Survey Management Plans
- Definition of data standards
- User documentation for the phase 3 process

The Phase 3 process & responsibilities



- ESO provides the phase 3 infrastructure (Release validator, Release manager, FTP server).
- The ESO Archive provides data query and access services including data access via VO protocols.

Schematic view



ESO External Data Products Standard



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

Each system component mandates certain data format requirements:

- ESO Archive can ingest data in FITS format
- Standard keywords track provenance allowing ESO to monitor survey progress
- Standard keywords to characterize & query for DPs
- Standard keywords to support VO protocols

Additionally:

- Monitoring the data quality of EDPs



ESO External Data Products Standard

- ❑ Level of data reduction and calibration
- ❑ Data products characterization
- ❑ Data format
 - must generally conform to the FITS standard
 - must comply with the ESO Data Interface Control Document, as far as applicable in the context of EDPs
 - File names
 - Max. 68 characters (incl. suffix, see below).
 - Filenames must be unique for a given data release. Uniqueness within the data collection is required for complementing data release, i.e. if the data is released in a cumulative fashion (see below)
 - Filename suffixes `*.fits` and `*.fits.fz` (using *fpack* compression)
- ❑ Data product types
- ❑ EDP keyword definitions

Version 3.0, July 2008,
<http://fits.gsfc.nasa.gov/iaufwg/>

Doc. No.: GEN-SPE-ESO-19400-0794,
Issue: 4, Date: 8 April 2008
http://archive.eso.org/DICB/dicd_v4/dicd_v4.pdf



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

Data product types



The *VISTA* tile image

- basic building block of VISTA public surveys;
- filled area of sky fully sampled, covering ca. 1 x1.5 square degrees;
- 6 pointed observations (with fixed offsets) required for reasonably uniform coverage;
- observation executed with the VISTA/VIRCAM template “VIRCAM_img_obs_tile6”;
- exposes each piece of sky, except for the edges of the tile, on at least 2 camera pixels;
- stores the data array in the FITS file’s primary HDU
- Requires to associate the respective weight image, also known as confidence map (using the keywords ASSON1 and ASSOC1, see below).
- Two flavors in order to distinguish whether data have been combined from one or from multiple observations (“VISTA *deep* tile image”).

List of header keywords specific to the VISTA tile image including the VISTA deep tile image

```
TELESCOP= 'ESO-VISTA'      / ESO Telescope designation
INSTRUME= 'VIRCAM '        / Instrument name
OBSTECH = 'IMAGE,JITTER'   / Technique of observation
PRODCATG= 'science.image'  / Data product category
IMATYPE = 'TILE'           / Specific image type
```



The *VISTA pawprint image*

- The 16 non-contiguous images of the sky produced by the VISTA IR camera, with its 16 non-contiguous detector chips are termed *VISTA pawprint image*;
- Sampling ca. 0.6 square degrees within a FOV of about 1 x1.3 square degrees;
- According to the usual technique of observation in the NIR regime (jitter mode or offset sky), typically, the pawprint is the result of multiple exposures;
- The “normal” VISTA pawprint is based on a single observation, which has to be identified by the keyword OBID1. If the original data, which has been combined to form the final pawprint image, was obtained in more than one observation block, the data product is termed *VISTA deep pawprint image*, and the complete set of original observations should be listed using the indexed keyword OBIDi;

List of header keywords specific to the VISTA pawprint image

```
TELESCOP= 'ESO-VISTA'           / ESO Telescope designation
INSTRUME= 'VIRCAM '             / Instrument name
OBSTECH = 'IMAGE,JITTER'       / Technique of observation
PRODCATG= 'science.MEFimage'   / Data product category
IMATYPE = 'PAWPRINT'           / Specific image type
ISAMP   =                       T / TRUE if image represents partially sampled sky
```



The *VISTA stripes image*

- consists of 4 vertical stripes sampling a total of about 0.73 square degrees of the 1° by 1.5° patch of one VISTA tile;
- results from the co-addition of 3 pointed observations vertically offset by ca. 1/3 of the detector size;
- The data of the VISTA stripes image are stored in 4 image extensions of a multi-extension FITS file;
- The stripes image is the baseline layout for the ultra deep part of the UltraVISTA public imaging survey.
- According to the observing strategy chosen for this programme, the VISTA stripes image will consist of 3 OBs each at least, corresponding to the 3 pointing positions. Following the terminology introduced earlier for VISTA tiles and pawprints resulting from multiple OBs, the respective UltraVISTA products are called deep stripes image.
- It is not foreseen that VISTA public imaging programmes deliver the stripes image resulting from a single OB.



The *VISTA* source list

- provides the file format for the tabular data of sources extracted from VISTA imaging data products;
- pipeline-produced, using the nightly calibrations and is delivered on an image-by-image basis;
- Typical example: single-band source catalogue extracted from one VISTA tile, or origins from VISTA pawprint image or VISTA stripes image;
- Must be associated to its originating image due to its processing provenance (using the keyword *PROVi*).
- FITS binary table format. Each data array of the originating image gives rise to one binary table extension in the FITS file. Thus, the source list of a VISTA tile contains one single binary table extension while the source list of VISTA pawprint contains 16 binary table extensions.
- VISTA source list adopts a number of keywords that characterize the originating imaging observation to facilitate direct archive queries.

List of header keywords specific to the VISTA source list

```
TELESCOP= 'ESO-VISTA'      / ESO Telescope designation
INSTRUME= 'VIRCAM '        / Instrument name
OBSTECH = 'IMAGE,JITTER'   / Technique of observation
PRODCATG= 'science.srctbl'  / Data product category
IMATYPE =                   %s / Specific image type
ISAMP   =                   %c / TRUE if image represents partially sampled sky
```



Summary of VISTA imaging data products

Data product description	VISTA footprint shape	Single/multiple OBs	Sampled/contiguous sky	Number of disconnected regions sampled	Source list
VISTA tile	Tile	Single	contiguous	1	✓
VISTA deep tile	Tile	Multi	contiguous	1	✓
VISTA pawprint	Pawprint	Single	sampled	16	✓
VISTA deep pawprint	Pawprint	Multi	sampled	16	✓
VISTA stripes	4 vertical stripes	Single	sampled	4	✓
VISTA deep stripes	4 vertical stripes	Multi	sampled	4	✓
VISTA single exposure	Pawprint	Single	sampled	16	n/a

The attribute “deep” indicates if the data product is based on multiple observations. The flag ISAMP indicates if the imaging data represents multiple disconnected regions, i.e. a *sampling* or one *contiguous* fraction of the sky. The column labeled “Source list” indicates the potential availability of source tables extracted from the imaging products.



EDP keyword definitions

External Data Products can be considered at the top of the hierarchy of data products in terms of processing level, and their metadata values must be obtained from the information present in the lower level products. Therefore, it is generally requested for proper archive ingestion that EDPs carry over keyword information from the original raw observational data.



Keyword settings for VISTA imaging data products

Data product description	PRODCATG	IMATYPE	ISAMP	SINGLEXP
VISTA tile VISTA deep tile	science.image	TILE	F	F
VISTA pawprint VISTA deep pawprint	science.MEFimage	PAWPRINT	T	F
VISTA stripes VISTA deep stripes	science.MEFimage	VSTRIPES	T	F
VISTA single exposure	science.MEFimage	PAWPRINT	T	T
VISTA tile's source list VISTA deep tile's source list	science.srctbl	TILE	F	F
VISTA pawprint's source list VISTA deep pawprint's source list	science.srctbl	PAWPRINT	T	F
VISTA stripes's source list VISTA deep stripes's source list	science.srctbl	VSTRIPES	T	F

- PRODCATG is a mandatory keyword for all EDPs.
- The presence of the keyword OBID2 indicates the multi-OB character.



```
SIMPLE = T / Standard FITS format (NOST-100-2.0)
BITPIX = %d / Number of bits per data pixel
NAXIS = 2 / Number of data axes
NAXIS1 = %d / Length of data axis 1
NAXIS2 = %d / Length of data axis 2
EXTEND = T / Extensions may be present
BZERO = %f / real = fits-value*BSCALE+BZERO
BSCALE = %f / real = fits-value*BSCALE+BZERO
BUNIT = %s / Physical unit of array values
BLANK = %d / Value used for NULL pixels
ORIGIN = 'ESO-PARANAL' / European Southern Observatory
DATE = %s / Date the file was written
DATAMAX = %f / Maximum pixel value
DATAMIN = %f / Minimal pixel value
TELESCOP= 'ESO-VISTA' / ESO Telescope designation
INSTRUME= 'VIRCAM ' / Instrument name
FILTER = %s / Filter name
OBJECT = %s / Target designation
RA = %f / Telescope Pointing (J2000.0)
DEC = %f / Telescope Pointing (J2000.0)
EQUINOX = %.0f / Standard FK5 (years)
RADECSYS= %s / Coordinate reference frame
EXPTIME = %f / Total integration time per pixel (s)
TEXPTIME= %f / Total integration time of all exposures (s)
MJD-OBS = %.8f / Start of observations (days)
MJD-END = %.8f / End of observations (days)
DATE-OBS= %s / Date the observation was started (UTC)
TIMESYS = 'UTC ' / Time system used
PROG_ID = %20s / ESO programme identification
OBID1 = %d / Observation block ID
M_EPOCH = %c / TRUE if resulting from multiple epochs
NCOMBINE= %d / # of combined raw science data files
PROV1 = %s / Originating science file
PROV2 = %s / Originating science file
PROV3 = %s / Originating science file
OBSTECH = 'IMAGE,JITTER' / Technique of observation
PRODCATG= 'science.image' / Data product category
IMATYPE = 'TILE' / Specific image type
FLUXCAL = 'ABSOLUTE' / Certifies the validity of PHOTZP
CRVAL1 = %f / Coordinate value at ref pixel
CRVAL2 = %f / Coordinate value at ref pixel
CRPIX1 = %f / Ref pixel in X
CRPIX2 = %f / Ref pixel in Y
CTYPE1 = %s / pixel coordinate system
CTYPE2 = %s / pixel coordinate system
```

List of header keywords for the VISTA tile





Specifying the temporal coverage

Type	Keyword	Description
(R)	MJD-OBS	Specifies the start of the observation in terms of the modified Julian day; the start of the earliest observation if the data product results from the combination of multiple observations. To be adopted from the original data.
(R)	MJD-END	Specifies the end of the observation; the end of the latest observation if the data product results from the combination of multiple observations.

MJD-OBS should be set to MJD-OBS of the first exposure contributing to this data product.

MJD-END—the end of observations can be obtained in an approximate fashion using MJD-OBS of the last exposure contributing to this data product and adding the total exposure time of this exposure, i.e. $MJD-END = MJD-OBS + (DIT * NDIT) / 86400$ in which MJD-OBS, DIT and NDIT refer to the last exposure.

Specifying total exposure time (1)



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

Type	Keyword	Description
(R)	EXPTIME	<p>Total integration time per pixel (in seconds).</p> <p>For an imaging data product resulting from the co-addition of multiple exposures pointing at the same sky position (with a tolerance given by a small fraction of the instrumental field of view), EXPTIME should represent the total integration time per pixel obtained in the centre of the image.</p> <p>If the product has been constructed from exposures whose positions were offset from each other in order to sample a region of the sky being larger than the instrumental FOV then the total integration time may vary across the image array. In this case EXPTIME should be set to the nominal total integration time obtained in at least 50% of the image array taking into account the chosen offset pattern.</p> <p>Note that EXPTIME as given in the original raw data almost never represents the proper number of EXPTIME for the product, specifically if detector sub-integrations are involved.</p>
(R)	TEXPTIME	<p>Arithmetic sum of the integration time of all exposures included in this product (in seconds).</p> <p>Note that an exposures integration time is $DIT \cdot NDIT$ if sub-integrations are involved.</p>



Specifying total exposure time (2)

EXPTIME—for a VISTA tile being filled using the standard pattern of 6 pointed observations with fixed offsets with *NJITTER* exposures per pointing, *EXPTIME* should be set to the product $2 \times \text{NJITTER} \times \text{NDIT} \times \text{DIT}$. The factor 2 reflects that most of the pixels of the final co-added image receive the contributions of at least two observations except for two narrow stripes along the edges, which receive just ‘single’ exposure time.

For a VISTA deep tile image resulting from the co-addition of *N_OBS* observations, each of them using the standard pattern of 6 pointed observations with fixed offsets with *NJITTER* exposures per pointing, *EXPTIME* should be set to the product $2 \times \text{N_OBS} \times \text{NJITTER} \times \text{NDIT} \times \text{DIT}$. The factor 2 reflects that most of the pixels of the final co-added image receive the contributions of at least two observations except for two narrow stripes along the edges, which receive just ‘single’ exposure time.

If the *N_OBS* observations do not share the same individual exposure time, i.e. *NJITTER*, *NDIT*, and *DIT*, then *EXPTIME* should be set to the sum

$$\sum_{i=1}^{N_OBS} 2 \times \text{NJITTER}(i) \times \text{NDIT}(i) \times \text{DIT}(i)$$

If individual exposures were rejected before combination into the tile, *EXPTIME* should be adjusted accordingly.



Specifying total exposure time (3)

TEXPTIME—for a VISTA tile being filled using the standard pattern of 6 pointed observations with fixed offsets with NJITTER exposures per pointing, TEXPTIME should be set to the product $6 \cdot \text{NJITTER} \cdot \text{NDIT} \cdot \text{DIT}$. If individual exposures were rejected before combination into the tile, TEXPTIME should be adjusted accordingly. For a VISTA deep tile image resulting from the co-addition of N_OBS observations, each of them using the standard pattern of 6 pointed observations with fixed offsets with NJITTER exposures per pointing, EXPTIME should be set to the product $6 \cdot \text{N_OBS} \cdot \text{NJITTER} \cdot \text{NDIT} \cdot \text{DIT}$. If the N_OBS observations do not share the same individual exposure time, i.e. NJITTER, NDIT, and DIT, then TEXPTIME should be set to the arithmetic sum of $6 \cdot \text{NJITTER} \cdot \text{NDIT} \cdot \text{DIT}$. If individual exposures were rejected before combination into the tile, TEXPTIME should be adjusted accordingly.

NB: The concept of TEXPTIME is different from the difference $\text{MJD_END} - \text{MJD_OBS}$.

Describing the provenance of data products



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

Type	Keyword	Description
(I)	OBID i	Set of Observation block IDs to identify the original observations this product results from. The Observation block IDs is a unique numeric ID that was assigned to the observation block by the Observation Handling Subsystem. Normally, OBID i should be copied from the keyword HIERARCH ESO OBS ID of the original data. If the product includes data from n observations, OBID i with index i running from 1 to n should be provided. On the contrary, a given Observation block ID must be listed in all the products that are based on this observation.
(S)	PROV i	Processing provenance, i.e. the list of science files originating this data product. i is a sequential number starting from 1. PROV i should appear as many times as needed to identify the complete set of science data files this product has been generated from.

OBID i is mandatory for EDPs because it is used to monitor the delivery of data products in view of the observational progress (SVMT).

PROV1—for a single-band source list the reference to the image data product from which this source list was extracted. If image and source list are being submitted at the same time, PROV1 has to be set to the filename of the image under which it is submitted. If the image has been submitted to and archived by ESO at an earlier time, PROV1 has to be set to the ESO archive identifier, which has been assigned to the respective file at the time of archiving. The ESO archive identifier can be obtained from the special file named `CONTENT.ESO` in the release directory on the phase 3 ftp area.



Ancillary products and data sets

Examples for categories of associated ancillary files given by the indexed keyword ASSOCI.

Category	Description
ancillary.weightmap	Weight map describing the pixel-to-pixel variation of the statistical significance of the image array in terms of the inverse variance. The weight map should be a FITS file having the same structure, i.e. number of FITS extensions, if any, and dimensions as the FITS file that contains the image data array.
ancillary.preview	Preview of the data product normally using one of the common graphics file formats such as JPEG, PNG, GIF, etc. For image data products the preview usually consists of an appropriately downscaled version of the image. For spectra the preview typically consists of a line plot for which the PS or PDF formats may be considered.

Type	Keyword	Description
(S)	ASSON i	The list of files associated to this data product. i is a sequential number starting from 1. If n files are associated to the product, the indexed keywords ASSON i and ASSOCI should appear n times ($i=1, \dots, n$).
(S)	ASSOCI	Specifies the category of the associated file given by ASSON i .

The set of files consisting of the (primary) science file and its associated ancillary files is called a **data set** in the Phase 3 Release Manager.



FITS Keywords for External Data Products

Photometric scale and quality parameters

Type	Keyword	Description
(R)	PHOTZP	Photometric zeropoint that relates the pixel data to total magnitudes (MAG) according to the equation $MAG = -2.5 * \log(data) + PHOTZP$, i.e. any applicable scaling with exposure time should be absorbed into PHOTZP.
(R)	ABMAGLIM	5-sigma limiting AB magnitude for point sources. The quoted magnitude should refer to the total flux of a point source.
(R)	ABMAGSAT	Saturation limit for point sources (AB magnitude).
(R)	PSF_FWHM	Spatial resolution (arcsec). Quality parameter measured from the image. Average size of the point spread function expressed as the full width at half maximum in arcseconds.
(R)	ELLIPTIC	Average ellipticity of point sources defined as $(1-b/a)$ with a and b denoting the major and minor axes of the source profile, resp.



List of header keywords for the VISTA tile

```
SIMPLE = T / Standard FITS format (NOST-100-2.0)
BITPIX = %d / Number of bits per data pixel
NAXIS = 2 / Number of data axes
NAXIS1 = %d / Length of data axis 1
NAXIS2 = %d / Length of data axis 2
EXTEND = T / Extensions may be present
BZERO = %f / real = fits-value*BSCALE+BZERO
BSCALE = %f / real = fits-value*BSCALE+BZERO
BUNIT = %s / Physical unit of array values
BLANK = %d / Value used for NULL pixels
ORIGIN = 'ESO-PARANAL' / European Southern Observatory
DATE = %s / Date the file was written
DATAMAX = %f / Maximum pixel value
DATAMIN = %f / Minimal pixel value
TELESCOP= 'ESO-VISTA' / ESO Telescope designation
INSTRUME= 'VIRCAM ' / Instrument name
FILTER = %s / Filter name
OBJECT = %s / Target designation
RA = %f / Telescope Pointing (J2000.0)
DEC = %f / Telescope Pointing (J2000.0)
EQUINOX = %.0f / Standard FK5 (years)
RADECSYS= %s / Coordinate reference frame
EXPTIME = %f / Total integration time per pixel (s)
TEXPTIME= %f / Total integration time of all exposures (s)
MJD-OBS = %.8f / Start of observations (days)
MJD-END = %.8f / End of observations (days)
DATE-OBS= %s / Date the observation was started (UTC)
TIMESYS = 'UTC ' / Time system used
PROG_ID = %20s / ESO programme identification
OBID1 = %d / Observation block ID
M_EPOCH = %c / TRUE if resulting from multiple epochs
NCOMBINE= %d / # of combined raw science data files
PROV1 = %s / Originating science file
PROV2 = %s / Originating science file
PROV3 = %s / Originating science file
OBSTECH = 'IMAGE,JITTER' / Technique of observation
PRODCAT= 'science.image' / Data product category
IMATYPE = 'TILE' / Specific image type
FLUXCAL = 'ABSOLUTE' / Certifies the validity of PHOTZP
CRVAL1 = %f / Coordinate value at ref pixel
CRVAL2 = %f / Coordinate value at ref pixel
CRPIX1 = %f / Ref pixel in X
CRPIX2 = %f / Ref pixel in Y
CTYPE1 = %s / pixel coordinate system
CTYPE2 = %s / pixel coordinate system
```





The Data Release Description

The release description forms an integral part of any ESO data release. It is essential for ESO/EDP to review and validate any release.

- provides an account of the release content,
- the originating observations,
- the calibration and data reduction procedures,
- the data quality, the data format, and, possibly,
- the scientific context of the programme.

The filename of the release description has to be `release-description.tar` in the main directory of the release on the phase 3 ftp server.

Abstract

A short, broad overview, text-only, referring to ESO programme, instrument, observational setup, filters/bands used, total coverage, resolution etc. when applicable. Possibly touching on the scientific context.

Overview/layout of observations

For imaging data products: A brief listing of the positions of the various fields/objects with an indication of the set of bands used for each, preferably with a finding chart or display of the covered fields/objects.

For spectroscopic data products: If possible, a finding chart, or another illustration that gives an idea of from where on the sky the spectra were taken.

Release content

For imaging data products: An extended listing for each field/object of sky position, filter, exposure time, seeing, observing date etc.

For spectroscopic data products: An extended characterisation of the spectra.

Release notes

Short descriptions of the reduction methods used, the calibration procedures (astrometric, photometric, wavelength etc.), characterization of the data quality, and a comparison with previous releases where applicable. It is recommended that the reference catalogue being used to establish the astrometric calibration is specified, e.g. GSC1, GSC2, USNO, 2MASS.

Data format

A description of the types of files in this release, associated files, and naming conventions used.

Acknowledgements

The acknowledgments to be included when using this data. Usually, a reference to the scientific publication associated with the data is given.



ESO
European Organisation
for Astronomical
Research in the
Southern Hemisphere

Data format for Catalogues resulting from ESO Public Surveys



Source Lists

- ❖ Based on single-night calibrations
- ❖ Pipeline-processed
- ❖ Usually per tile
- ❖ Possibly band-merged
- ❖ Degeneracy due to multiple detections
- ❖ Mostly single-epoch (except for deep stacks)

Catalogues

- ✧ Astrometry/Photometry equalized across the survey
- ✧ Band-merged
- ✧ Multiple detections merged– unique entries
- ✧ Uniform tabular structure including column descriptors
- ✧ Multi-epoch (light curves)



Data Format Specification for Catalogues

- FITS binary table format
- Generic EDP Keywords similar as for other types of data products
- Specific column-related keywords
- Controlled set of (scalar!) data-types
- Uniform Content Descriptors (IVOA standard)
- Facilitating catalogue queries, e.g. query by flux, color etc.
- Unique identifier constraint
- Relational data model based on *foreign keys* allowing to map the links between catalogue records (e.g. VVV)

Model for the VHS source catalogue

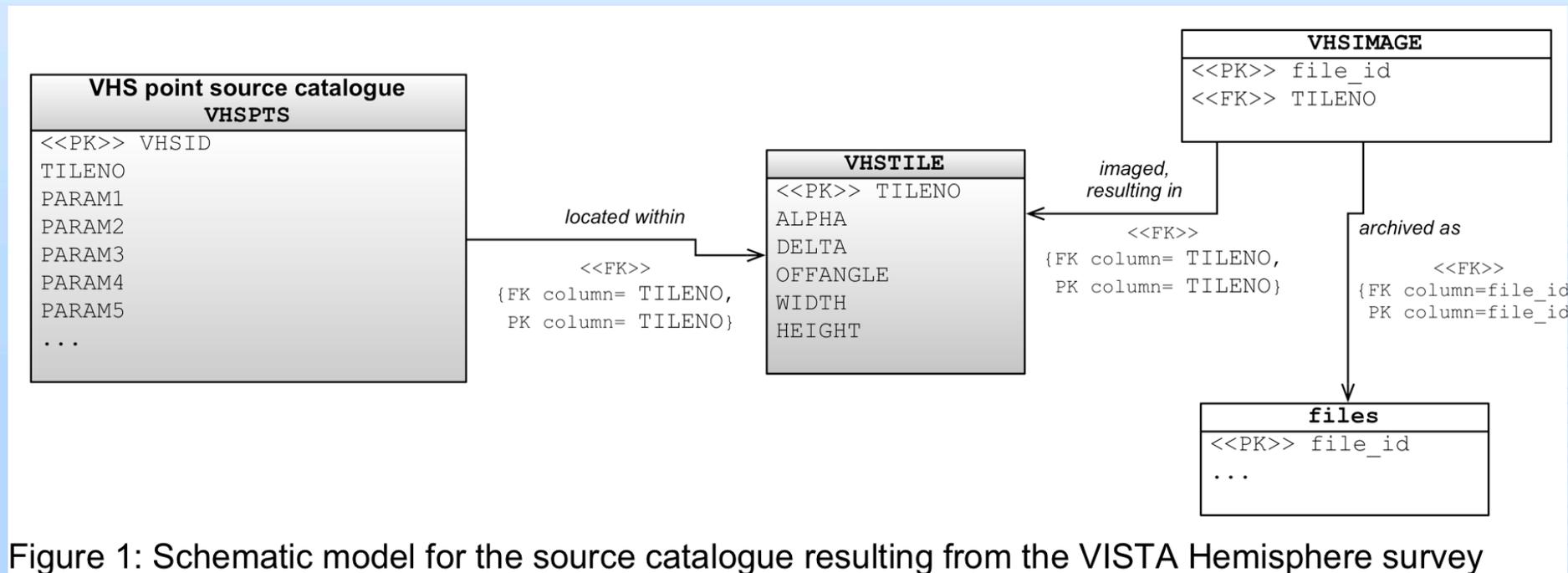
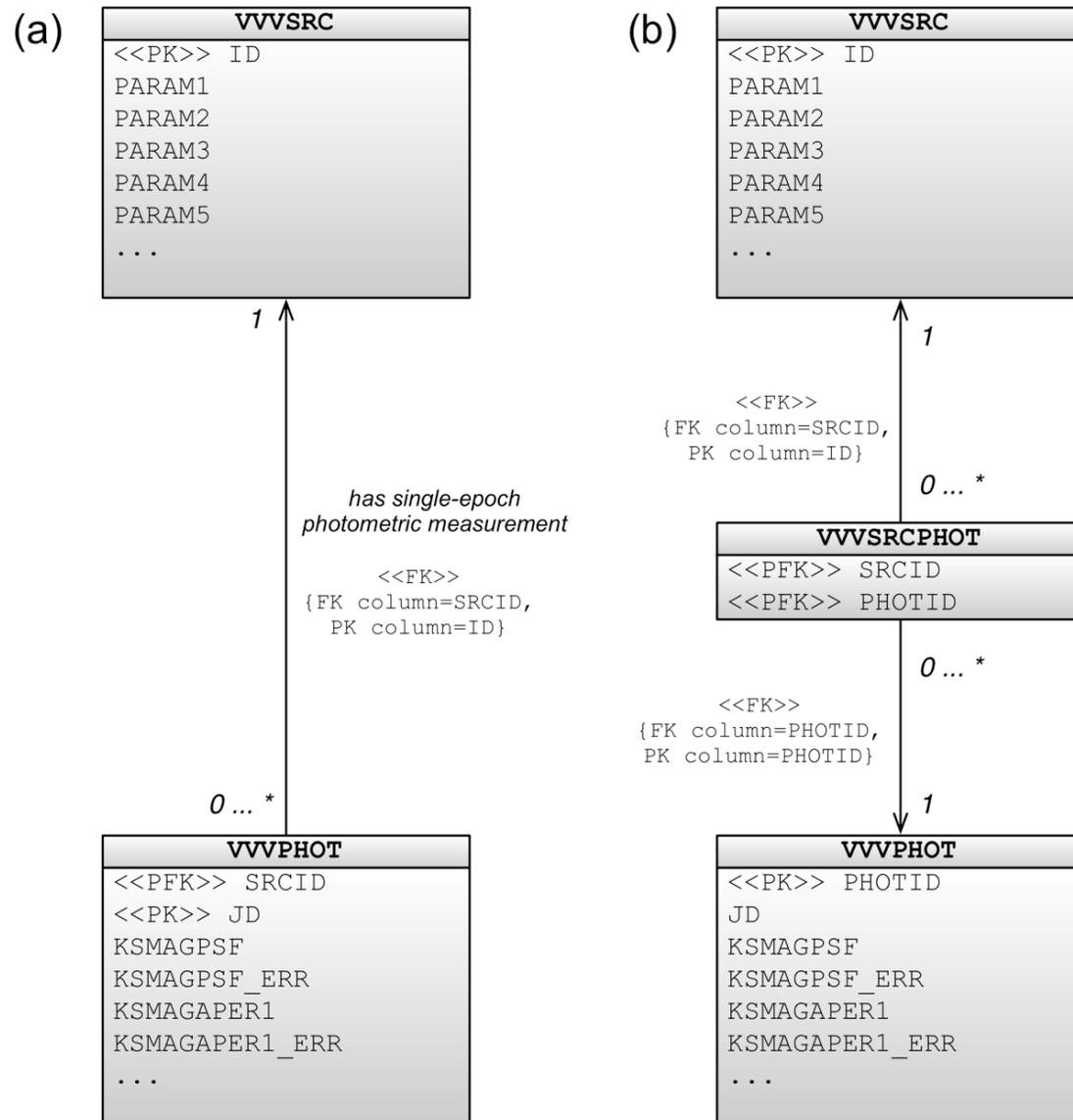


Figure 1: Schematic model for the source catalogue resulting from the VISTA Hemisphere survey

Possible Models for the VVV Catalogues



- describes the relation between the variable source and its photometric data point

- Model (b) is less restrictive than (a) by allowing photometric measurements in **VVVPHOT** which are not associated to any source in **VVVSRC**. For this reason, b is the model of choice if the catalogue of photometric data points is produced and delivered in a first stage, and then the catalogue of variable sources including the associative table at a later date.

The format specification for catalogues defines a *meta-model*.



We are planning to published the data format specification for catalogues on the Phase 3 web pages beginning of 2011.