

VLT Instruments Pipeline System Overview

Yves Jung, Pascal Ballester, Klaus Banse, Wolfgang Hummel, Carlo Izzo, Lars K. Lundin, Andrea Modigliani, Ralf M. Palsa

European Southern Observatory

Derek J. McKay

Rutherford Appleton Laboratory

Michael Kiesgen

Michael Bailey Assoc.

Cyrus Sabet

Tekom GmbH

Abstract. Since the beginning of the VLT operations in 1998, substantial effort has been put in the development of automatic data reduction tools for the VLT instruments. A VLT instrument pipeline is a complex system that has to be able to identify and classify each produced FITS file, optionally retrieve calibration files from a database, use an image processing software to reduce the data, compute and log quality control parameters, produce FITS images or tables with the correct headers, optionally display them in the control room and send them to the archive. Each instrument has its own dedicated pipeline, based on a common infrastructure and installed with the VLT Data Flow System (DFS). With the increase in the number and the complexity of supported instruments and in the rate of produced data, these pipelines are becoming vital for both the VLT operations and the users, and request more and more resources for development and maintenance. This paper describes the different pipeline tasks with some real examples. It also explains how the development process has been improved to both decrease its cost and increase the pipelines quality using the lessons learned from the first instruments pipelines development.

1. Introduction

The VLT instruments pipelines development, integration and maintenance are tasks under the responsibility of the Data Flow System (DFS) group at ESO.

Seven operational instrument pipelines (ISAAC, FORS1, FORS2, NACO, UVES, FLAMES, VINCI) are currently under maintenance, and seven others are being developed and are due over the next two years (CRIRES, VISIR,

SINFONI, VIMOS, GIRAFFE, AMBER, MIDI). On top of that, the second generation instruments pipelines will have to be developed from 2005 on at a rate of one pipeline every two years.

2. Pipelines Infrastructure

The different pipeline tasks are achieved by a number of different software components that can be classified in two distinct categories: those that are common to all instruments and those which are specific to each instrument pipeline package.

2.1. The Instrument Independent Components

The instrument independent components are mainly responsible for the data flow handling.

1. The Data Organizer (DO) is responsible for the data flow part. It recognizes the incoming FITS files by cross-checking some header keywords with `rules` defined in the configuration files delivered with the pipeline package (see 2.2.). It finds out which calibration files are needed and retrieves them from the calibration database (also delivered with the pipeline package). The DO then decides, according to the frames type, which data reduction recipe should be launched, writes all these information in a Reduction Block and sends it to the Reduction Block Scheduler (RBS).
2. The RBS retrieves the created Reduction Blocks, parses them and gives all the necessary informations to the Data Reduction System (DRS).
3. Some DFS tools are used to send the products to the archive, to write the computed Quality Control (QC) parameters into the database, etc..

2.2. The Instrument Pipeline Package

Each pipeline package contains three software components.

1. The rules needed by the DO to classify the frames, to know which calibration frames must be retrieved and which reduction recipe must be used are defined in the DO configuration files (or `rules`).
2. The calibration frames are contained in the calibration database, which is the second component of the pipeline package. Both the calibration database and the rules actually contain instrument specific information, that cannot be reduced any further.
3. The main software component contained in the Instrument Pipeline Package (by far) is of course the DRS. Although many different kinds of data have to be reduced across the different instruments (see 4.), the same low level data structures and data reduction tools are used in many different pipelines. Moving these common tasks to a common library would greatly reduce the size of the different DRS, which is exactly what we are aiming to.

Of course, the instrument specific components have to be developed for each different pipeline, and it is clear that our efforts are put on the reduction of the instrument dependent software size.

3. The Common Pipeline Library

The Common Pipeline Library (CPL) (Banse et al. 2004) is a C library that contains all these common functionalities. It has been developed during the last 2 years, and the first public release will take place in December 2003. It is one additional Instrument Independent Component that will reduce the size of the different DRS delivered with the various pipelines packages.

CPL was developed using the C library used for the ISAAC and NACO pipelines: `eclipse` (Devillard 2001) and the C code developed by ESO for the VIMOS pipeline integration. Particular effort has been put on the documentation and on API clarity, as external consortia will have to base their pipeline development on this library. From now on, every new instrument pipeline will have to be based on CPL. Besides that, the already existing pipelines written in C (ISAAC, NACO, VIMOS, etc..) will have to be converted to use CPL.

The current content of CPL approximately reflects the needs of the pipelines developed so far.

4. Data Reduction

The VLT instruments are each very specific, and their data reduction requirements vary greatly. The following is a short description of the different data reduction modes that can be identified, with their associated instruments.

4.1. Imaging Mode in Infrared (ISAAC, NACO, VISIR)

The high background in infrared data must be carefully estimated to retrieve the science information. In imaging mode, the observations are done in `jitter` mode, with small offsets around a central position for each exposure, to allow to estimate the sky background variations directly by filtering the images, and separate astronomical from sky signal. Apart from this difficult sky estimation, the frames are recombined with some cross correlation techniques to precisely determine the offsets between the images.

4.2. Long Slit Spectroscopy in Infrared (ISAAC)

Like for the imaging mode, the high background is removed using special observation techniques. In long slit spectroscopy, shifts are applied along the slit (`nodding`) or a tip/tilt is applied to the secondary mirror (`chopping`). The pipeline must classify and recombine the frames, and then apply various calibration corrections like, e.g. wavelength calibration, distortion estimation and correction, flat fielding, etc.. The brightest spectrum is then automatically detected and extracted.

4.3. Echelle Spectrograph (UVES, CRIRES)

The UVES science data contain tilted spectra with the different orders in the same image. To extract all these tilted spectra, a precise spectral format definition (spectra position and wavelength calibration) is needed for all of them. This format is determined by different calibration recipes using lamp images and physical model solutions.

4.4. Fiber Mode (FLAMES/UVES)

In FLAMES, a series of fibers illuminate the slit. They all generate a spectrum similar to the common UVES observation. These are all stored together in one common image. This requires an even more precise spectral format definition to extract each fiber than for UVES. The reduction for each fiber is then the same as the standard one applied for UVES.

4.5. Multi Object Spectroscopy (VIMOS, FORS1, FORS2)

The data produced by Multi Object Spectrograph (MOS) contain a huge number of spectra (up to 800 in VIMOS science images). They all must be identified and wavelength calibrated. The source spectra are then individually sky subtracted, flat fielded and integrated.

4.6. VLT (MIDI, VINCI, AMBER)

In the case of VLT instruments, the data compression rate is very high. In the case of MIDI, a 2 gigabytes data set is used to obtain a single measurement of the fringes visibility. This means that around 10 measurements can be obtained from a night of observations. The production of a reliable error estimation on the fringes visibility measurements is a very important task of the pipeline.

4.7. Integral Field Unit Mode (VIMOS, GIRAFFE, SINFONI)

The Integral Field Unit (IFU) mode is using fiber bundles or an image slicer to observe different parts of extended objects. The different fibers' observations are contained in the same science data, and a very precise calibration (like for the Fiber mode) is needed to extract the correct signals.

5. Quality Control

The different recipes of the different instrument pipelines produce some quality control (QC) parameters (Hanuschik et al. 2003) that are automatically written in a central log file and a common database. The health of the instruments is then automatically monitored with e.g. the zero point values, the dark current, the strehl ratio, etc..

References

- Banse K. et al. 2004, this volume, 392
- Devillard N. 2001, in ASP Conf. Ser., Vol. 238, ADASS X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 525
- Hanuschik R., Hummel W., Sartoretti P., Silva D. 2003, Quality Control of the ESO-VLT instruments, Observatory Operations to Optimize Scientific Return III. Edited by Quinn, Peter J. Proceedings of the SPIE, Volume 4844, pp. 139-148 (2003)