*Astronomical Data Analysis Software and Systems XVI*
*ASP Conference Series, Vol. XXX, 2006*
*R. Shaw, F. Hill and D. Bell, eds.*



# Distributed Quality Control of VLT data at ESO

Reinhard Hanuschik

*European Southern Observatory, Data Management and Operations Division, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany; Email: rhanusch@eso.org*

**Abstract.** For the VLT observatory on Cerro Paranal in Chile, ESO has implemented an integrated end-to-end approach which covers all aspects of data flow, from the definition of the observations, through the scheduling, down to the acquisition, archiving and processing of the data. An integral part of this end-to-end data flow model is the Quality Control of the data. This paper deals with the implementation of the QC process, and its challenges met so far.

## 1. Introduction

The Very Large Telescope (VLT) of the European Southern Observatory (ESO) consists of currently 9 instruments on four unit telescopes, two interferometric instruments, and soon two survey cameras on dedicated survey telescopes. The instruments cover all kinds of resolution, in the wavelength range from the optical to the mid-IR. Most of them have several different modes including imaging, single-object and multi-object spectroscopy, and integral-field spectroscopy[1]. The interferometric instruments are fully integrated in the VLT data flow process.

Most of the instrument modes are supported by data processing pipelines enabled for automatic processing. These pipelines have two main functions:

- quality control: measure the instrument status and provide information for controlling their performance;
- data reduction: remove instrument signature from scientific observations.

The pipelines run at the observatory in automatic mode in the background, with their products available for visiting and staff astronomers, and at the headquarter in Garching/Germany in off-line, supervised mode. Here the Quality Control Group is in charge of monitoring the data quality and providing feedback to the observatory staff.

## 2. Data Quality

What are good astronomical data? For science data, good quality means best possible performance: minimal impact by ambient conditions, best S/N and

[1] http://www.eso.org/instruments

resolution, low background etc. Good calibration data measure a valid record of the instrumental and ambient conditions, accurate enough to allow removing the instrumental signature from science observations without degrading them. This means that e.g. a flat should have high-enough S/N to not increase the noise of the flattened science exposure. A valid record is obtained if the instrument status, or that of the atmosphere, at the time of the calibration is close enough to that of the science data.

In many cases this means taking calibrations close in time with the science data, and the accuracy of the calibrations is determined by the difference in conditions between science and calibrations. The more stable an instrument is, the more this requirement can be relaxed. Whenever possible ESO follows the approach of calibrating the instrument, rather than calibrating each single science observation, meaning that all information required to reduce a scientific exposure is known if the instrument status (e.g. temperature) is measured and calibration data exist for that status. ESO follows a calibration plan for each instrument mode which acquires calibrations in a defined frequency and matches predefined accuracy goals.

Data quality control means to ensure that the acquired data follow the above stated requirements, to a known and predictable accuracy. ESO has a team of eight QC astronomers who ensure that these goals are met. They process the VLT data stream, extract QC information, certify the products, feed back QC information to the mountain, and create data packages for the PIs. Currently, this means to process roughly 400 GB of raw data every month, in about 12,000 processing jobs. About 80 data packages are sent out every month[2].

## 3. From QA to QC

### 3.1. Shared Quality Control

The Quality Control process has two main steps: quality assessment (QA) where key parameters are measured and subjected to trending; and quality control (QC) where a feedback loop is established and QC information is fed back to the mountain. For optical instruments, immediate interventions are possible. For the cryogenic IR, they can at least be scheduled.

For QA, the pipelines measure performance numbers (the QC parameters) which are written into the data product headers and into a database. For example, from BIAS frames the mean bias level, the read noise (both in raw and product frames), and some structure parameters are extracted.

The VLT Quality Control is a shared and distributed process. Due to the large data volume, the VLT data are currently transferred on DVDs or hard disks from Paranal to Garching headquarters, a process which takes between 8 and 14 days in total. The QC process therefore must be a shared process: the most fundamental checks must be done on-site and in near-real time, otherwise large data volumes would be affected by a problem discovered too late. These Health Checks can only include a limited set of parameters. All more complete

[2]Find more information under http://www.eso.org/qc

checks, and in particular those requiring advanced studies or larger data sets, are done in Garching.

The main components of the QC process are:

- the pipelines extracting QC parameters;
- the on-site QC process;
- the Health Check process: comparing new to existing QC parameters;
- off-line processing and certification at Garching Headquarters;
- trending.

The on-site QC process checks:

- formal compliance of file format;
- compliance with user constraints (observation grading);
- new raw files against reference files.

The off-line QC process at Garching headquarters involves:

- processing with optimized associations (closest in time, with certified data products and correct dependencies[3]);
- process with best-possible pipeline setup (with optimized, customized parameters, and up-to-date, possibly patched/improved beta versions);
- measuring and assessing the data product quality (using customized QC reports and trending analysis);
- certifying products: release them for archiving and delivery, and for usage in processing other data.

### 3.2. Health Check monitor

The on-site pipelines on Paranal extract the QC parameters from the data and write them in to a database. That information is the nreplicated to Garching headquarters where it is then compared by an automatic process to previous data, a procedure commonly called trending. The trending reports are published on a web site, the Health Check Monitor[4]. The interferometric instruments

This web site is the main QC interface between the mountain and the QC group in Garching. The daytime astronomers, when they take up their duty in the Paranal morning, check these pages first, to see if the most recent calibrations conform to specifications or to the previous performance. Any unusual effect visible in the QC parameters and trending plots should be discovered at that stage. Of course there are sometimes more subtle effects which will become apparent only when the data processing and final certification is done in Garching.

The Health Check (HC) system is planned to be expanded to automatically detect outliers and create alerts or trigger other actions. This will become particularly important for the QC of survey data which will produce up to a 100x more data and QC parameters than the current instruments. Then QC will become a real mass production process, with quality control concepts adopted from industrial procedures.

---

[3] Calibrations often follow a hierarchical scheme which is not necessarily matching their temporal sequence; such dependencies can only be respected when the full data set is available, not in sequential processing as done on the mountain.

[4] http://www.eso.org/qc/ALL/daily_qc1.html

An important part of the HC process is the regular acquisition of dedicated calibrations just for the purpose of health checks, without relation to any science data. With equally spaced data points trend analysis becomes much easier and safer.

### 3.3. Trending and history

After final processing and certification of the calibration data, their products are archived. The QC parameters are also stored. The complete set of QC parameters is a record of the instrument evolution. It helps e.g. to better understand the instrument performance and to analyze whether a current issue is new or had its precursors in history.

The trending history is used for:

- maintenance planning ("by how much do the zeropoints degrade over a year by dust accumulation");
- understanding the complete instrument history ("how often in the past has the high-resolution grating shown that kind of shift");
- instrument scientists, researchers, archive users ("how does my current dispersion rms compare to earlier observations two years ago").

## 4. Critical components, challenges

With two different sites involved, located on different hemispheres and separated by time zones and duty shifts, one of the main challenges in keeping the shared QC concept alive is information exchange. This requires formalized procedures like regular video and tele conferences, structures like the instrument operation teams (with key members from all ESO parties involved), and tools like the Health Check Monitor.

In terms of data processing, the main issue is not data volume but data complexity. Over the past two years, QC Garching has been fighting hard to align all procedures and processes across the instruments. The QC process must be further pushed to be based on automatic evaluation whenever possible. We plan a scoring system which will analyze new products in terms of compliance to specifications and earlier trends: ideally a full day of calibrations is scored with a single green light, and certified by the QC scientist without further detailed analysis. A red light will trigger an analysis one level deeper, to identify those data with unusual performance which can then be studied at even deeper level. In that way the QC scientist is led to the real problem within just three mouse clicks and can then focus on the detailed analysis of a problem, rather than certify thousands of compliant products.