PHOENIX: the production line for science data products at ESO

Reinhard Hanuschik[1], Lodovico Coccato[*]
[*]European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany

ABSTRACT

In the past three years the QC group at ESO has installed a production line for science-grade data products. With the focus on spectroscopic observations, these in-house generated data products are complementary to the externally provided data products from the surveys. The production line combines efficient mass production (more than one million spectra have been generated so far), previews, and quality control. All data products are available to the community on the ESO archive interface.

Keywords: Science-grade data products, pipelines, data standards, data processing, quality control

1. INTRODUCTION

1.1 Setting the stage

The European Southern Observatory (ESO) runs the Very Large Telescope (VLT) at Paranal. With four UTs, the VLT interferometer and two survey telescopes it currently offers 16 operational instruments to its community. ESO has developed a number of tools to support its users with the reduction of the raw data towards data products. Only from the data products it is possible to start the science analysis. Data reduction pipelines are made available, and interactive workflows can be downloaded.

The latest addition to this are science-grade data products, at least for selected instruments and modes. In this article we focus on the production process for these data products. They become increasingly popular and are an important research tool.

1.2 Data calibration and reduction

Data reduction in ground-based astronomy is essential for two reasons:

- Instrumental and atmospheric artefacts need to be disentangled from the scientifically useful information that is collected on the detector.
- The science information needs to be re-gridded from the detector domain to the physical domain, which involves resampling, noise minimization, and calibration to physical units.

In order to separate the instrumental and atmospheric artefacts from the scientific information, dedicated calibrations are taken for all instruments, mostly during daytime. They properly register the instrumental effects, in form of bias, flats, arclamp exposures etc. Some of them are taken during nighttime (e.g. standard stars) or twilight (twilight flats).

At the Very Large Telescope we have a well-defined calibration plan for each instrument. All calibrations that are needed to reduce the science data taken in the night are acquired during the following day. The calibration plan may deviate from the strictly daily pattern if the instrument is known to be stable, or – as applicable to twilight calibrations – if the time slot is too short for daily cadence. In addition to the science-driven calibrations, certain types of maintenance calibrations are also part of the calibration plan. These calibrations (e.g. detector monitoring or efficiency monitoring) are typically acquired at monthly intervals.

[1] rhanusch@eso.org; www.eso.org/sci

1.3 Calibration processing, concepts of Quality Control

At ESO headquarters in Garching, all calibrations are processed by the *Quality Control Group* with pipelines into *master calibrations*. These masters are processed files with *condensed* quality and instrumental information. We extract *quality control (QC) parameters* which measure key properties of the calibration data and of the instruments. These parameters are automatically compared against pre-defined lower and upper thresholds (*scoring* [1]), and also put into the context of earlier similar parameters (*trending*). The trending is particularly important since, in comparison to the environment of industrial production lines, we are often undersampled in terms of data points: typically we have at most one measurement of a given parameter per day (for a bin in the parameter space defined by data type, instrument mode and setup). By comparison to earlier similar data we can not only determine the current instrument status, but also compensate for the undersampling if the variations of the QC parameter are sufficiently slow and small. The up-to-date collection of trending plots is available on the *Health Check monitor*[2].

The QC group reviews all master calibrations for quality and for validity. If accepted, a master calibration is called *certified*. The certified master calibrations are archived. They register the instrument status at the time of the calibration in a valid and complete way. In that sense we *calibrate the instrument*, not the data. The master calibrations are ready to be used for the reduction of the science data.

This approach supports both the needs of archive users, who can download the master calibrations and use them for their data reduction, and of internal processing projects.

1.4 Pipelines and interfaces

Each VLT instrument has its own pipeline. The pipelines can be used for the processing of the calibration data as well as for the science data reduction.

For the science reduction, the `reflex` workflows have been developed by ESO [2]. They allow the user to use the pipeline recipes in an interactive way, review intermediate results, and modify and optimize reduction strategies. They are particularly useful for the PI who wants to process the data in a way that is optimized for a specific science case.

For the standardized and efficient processing of datasets, the `esorex` command-line interface is available. It is particularly efficient for processing batches of any size and is the standard interface to the pipelines. It is used by the QC group for the processing of master calibrations.

2. SCIENCE DATA PROCESSING

2.1 Data processing and Quality Control

At ESO, we have developed two main approaches to support our community for getting to the science data products:

- the `reflex` workflows,
- the in-house generated data products.

The in-house generation of science data products is a major effort undertaken at ESO since 2013.

Key for both are the pipelines. As written by the instrument consortia and integrated into the ESO environment, the pipelines have different levels of maturity. It is key for the successful generation of quality data products that the respective pipeline used for a data processing project is *reviewed* and *certified*. This process includes a code review, the in-depth knowledge of algorithms and a review of the calibration plan. This certification process is done by the Science Data Products group at ESO.

For generating quality science data products, there is a second key component. The certified master calibrations contain the full information about the instrument status and conditions imprinted. These data have been processed, reviewed and certified by the closed-loop QC process, close in time to the original date of the observations. Not only have the data been reviewed for quality and validity, but the on-site staff also had the opportunity to react on QC issues, or data issues, and measure data again, if necessary, within a reasonable, pre-defined and controlled time interval. Hence it is a reasonable assumption that these master calibration data contain the full information about instrumental effects that is

[2] under the URL http://www.eso.org/HC

needed to disentangle the science information from the instrumental signature on the raw data. Furthermore, while the QC review and certification process is always to a certain extent interactive and time-consuming, the processing of the science data with these master calibrations is in principle a mass production, and it is repeatable, compute resources and time permitting.

Finally, for the ingestion of data products into the archive there is the need for defining their content and describing their properties in a standardized way. This is the task of the data product standard, which is developed and maintained by the Archive Science Group. Find the ESO data product standard here[3].

2.2 Processing of science data

A certified pipeline, certified and archived master calibrations, and the data product standard pave the road towards largely automated processing of science data to science-grade data products. This process has been implemented in the past three years at ESO. Crucial for this was the workflow tool PHOENIX.

We have set up a production line for generating science products with the following properties:

- We apply a *standardized* reduction: the reduction strategy is driven by data properties (encoded in the headers).
- Whenever choices about the data reduction strategy could be made (concerning recipes or reduction parameters), we assume a general-purpose approach. This means we do not optimize the data reduction for certain science cases. We aim at the highest possible processing level that can be reached without specific knowledge of the science case.
- We deliver science-grade products that come in physical units (whenever possible or reasonable) and with error estimates. They are essentially free from instrumental artifacts and are ready for scientific analysis.
- The science processing has a quality-control step which is process-oriented, but is not focusing on properties of individual products.

Our goal is to be complete for the entire history of a given instrument or selected mode. This is sometimes limited by the lack of key header features in the early times, or by the need to reconstruct master calibrations which might require resources beyond proportion.

By having processed data for the entire history of a given instrument (for UVES this is 16 years, with the first two years being incomplete), we can offer science data products covering hundreds of different scientific proposals. Also, these data can often be useful for scientific projects other than the original proposal. Chances are high that they can be used for a larger or a different context than for the original (PI-defined) science.

2.3 External and advanced data products

Complementary to our in-house data products, the ESO archive offers also data products processed externally, either by PIs or by data centres. These come either from surveys (VISTA, OMEGACAM) or from Large Programmes [3]. Their reduction has been fine-tuned towards a specific science case. They can be processed to higher levels, e.g. into source catalogues. Most of these *advanced data products* come from imaging modes. They can be considered complementary to the in-house data products.

Find all available data products on the ESO archive interface[4].

3. PHOENIX: WORKFLOW TOOL FOR SCIENCE DATA PRODUCTS

We have designed the workflow tool PHOENIX for the process to create science data products. It controls the whole workflow from the definition of the processing jobs up to the ingestion of products into the archive. It includes components for the bookkeeping and for storing the data model for future reference. It is a wrapper organizing the high-level orchestration of the processing jobs. The processing itself is provided by the pipelines. These are the most important steps:

---

[3] http://www.eso.org/sci/observing/phase3/p3sdpstd.pdf
[4] http://archive.eso.org/cms/eso-data/eso-data-products.html

Table 1. PHOENIX workflow steps

1. The tool is called for a given night. It selects all candidate datasets.
2. The datasets for each processing job, and their associations (the list of master calibrations required for processing) are downloaded from a central repository. In some cases the datasets might need to be re-created or optimized.
3. File headers and OB[5] execution information (like observing comments and quality grades) are downloaded from the nightlog database.
4. The raw science data are downloaded from the archive.
5. The master calibrations are downloaded from the archive.
6. The pipeline processing jobs are launched in an orchestrated way. In most cases they are called in parallel for efficiency on a multi-core machine, using a scheduling engine.
7. In a similar way, the post-pipeline quality-control jobs are scheduled and executed. They extract quality parameters, measure scores, and create preview plots.
8. If configured (per instrument), the data products are offered for review.
9. As necessary, information for the archive upload is added to the headers.
10. The data products are ingested into the archive.

The processing datasets are defined in the same way as the PI originally created them: either by single files (like for UVES or GIRAFFE) or by OB (for XSHOOTER and MUSE) which is a single pointing that might come with a series of N raw files, designed to be stacked. So far we have not combined datasets across OBs or across nights, but this might become a realistic option.

In comparison to the first presentation of PHOENIX [4], we have now added a module that can also process master calibrations for earlier epochs of interest. This is useful if at that time there was already a QC process with scores and certification, but the format of the master calibrations at that time was incompatible with the current pipeline version. Another use case for this PHOENIX mode is for those epochs when master calibrations were not yet archived. Both scenarios assume that the QC process was already implemented earlier and does not need to be repeated.

In order to achieve a high throughput for the whole processing schema, we need to make sure to have:

- stable and performant access to archived raw and master calibration files;
- efficient processing;
- stable and performant ingestion of the products into the archive.

For each PHOENIX project, there are typically two stages:

- The *historical batch*, which includes an exploration phase ("dry runs"); at that stage the processing strategy and the QC schema are developed, tested and fine-tuned. Several iterations might be necessary to develop a high-quality, high-performance and stable scheme. Naturally the historical batch is the most time-consuming and resource-demanding part of each PHOENIX project.
- The *current stream*: once the historical data set is processed, the stream of new data is processed typically once per month and then updates the existing data products dataset in the archive. These data are under the same proprietary policy as the raw data. The maintenance of the current stream is no project anymore but part of the regular operations done by the QC group.

For a given month, the processing of the new slice is started a few days after that month is over, when all applicable master calibrations are certified and archived. Once an OB has produced raw files on the VLT for a PHOENIX-

[5] Observing Block, the fundamental observing unit at the VLT.

supported instrument, the corresponding science-grade data products are available to the PI within a delay of between a few days at best, and about 5 weeks at most (unless exceptional delays occur).

4. CURRENT PHOENIX PROJECTS

4.1 Selection of projects

Not every VLT instrument has a PHOENIX project, and not every VLT instrument will ever have one. One reason for the necessity to select is available manpower. Another selector is the maturity of the pipelines which differs quite a bit between the VLT instruments. A necessary condition for a data processing project is the certification of the corresponding pipeline. Another necessary condition is the historical stability and completeness of the calibration plan. When there are several candidates for a new project, we select by potential impact: an instrument with a rich data heritage and high popularity among users ranks higher than a "niche" instrument with complicated, rarely used and perhaps poorly calibrated modes. Finally, we believe that data products from instruments with complex data reduction chains are more useful for the community than e.g. data from imagers which are fairly straightforward to reduce.

In Table 2 we list the current ESO data product projects. We have so far focused on spectroscopic instruments, either single or MOS modes. The latest addition is IFU spectroscopy (MUSE).

Table 2. Available science data products

Instrument, mode	**technique**	**wavelength range**	**Resolving power**	**Date range**	**Number of spectra/ datacubes**
UVES, Echelle	single slit spectroscopy	300 nm- 1 µ, in two arms (BLUE and RED)	R up to 100,000	2000-now	N ≈ 100,000 single spectra
XSHOOTER, Echelle	single slit spectroscopy	350 nm – 2.5 µ in three arms (UVB, VIS, NIR)	R = 10-20,000	2009-now	N ≈ 48,000 single spectra
FLAMES/Giraffe, Medusa (MOS)	multi-object spectroscopy	VIS	R = 30,000 (HR) or 6000 (LR)	2003-now	N ≈ 1.2 mio. single spectra, extracted from ≈15,000 raw files
MUSE IFU *(coming)*	integral-field spectroscopy	VIS (4750-9300 Å)	3000	2014-10 … 2015-05	*coming:* N ≈ 1,700 datacubes

4.2 UVES Echelle

UVES, the high-resolution single-slit echelle spectrograph, was the first VLT instrument with science data products. In 2013-10 the entire historical batch for the ECHELLE mode became available, with all new data being added regularly as a stream since then. The data products include all ECHELLE modes, including the SLICER data, the ABSORPTION-CELL data and the data using both optical components. To-date more than 100,000 spectra are being offered, coming from two arms that expose simultaneously.

The UVES data products are very popular among archive users. As of May 2016, the entire set of offered UVES spectra has effectively been downloaded more than 3 times. Virtually every single spectrum has already been downloaded. We have seen a total of 3,500 individual archive requests, 180 of them by PIs during the proprietary period. The popularity has not yet flattened out, we still receive more than 3 requests per day on average. Like the other data products, the UVES data have a detailed description of the processing steps and the dataset structure[6]. In Figure 1 we display the entire high-resolution spectrum of an M0.5III star obtained in 2 subsequent exposures and processed from 4 input files.

[6] http://www.eso.org/observing/dfo/quality/PHOENIX/UVES_ECH/processing.html

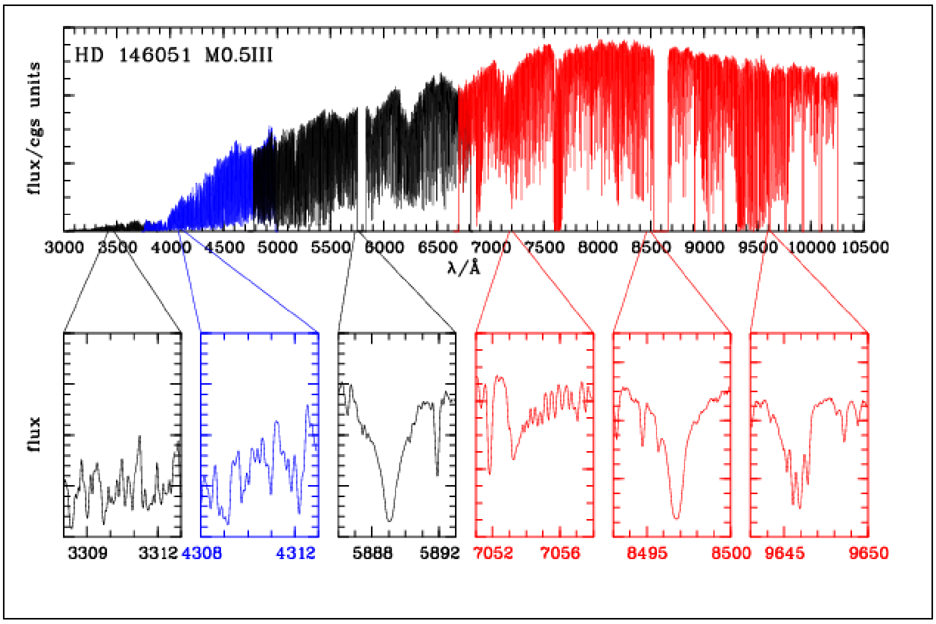

Figure 1. Set of UVES spectra, produced with PHOENIX. This is the flux-calibrated optical spectrum of an M0.5III star, composed of 4 individual spectra ranging from 300nm to 1µ. The spectral components are color-coded black-blue-black-red from left in the upper panel (in different grey tones in the printed version). Some spectral features are displayed at full resolution in the lower panel. The SNR is about 100-200, the “noise” in the upper panel is actually spectral structure. These spectra have been reduced in an entirely automated way.

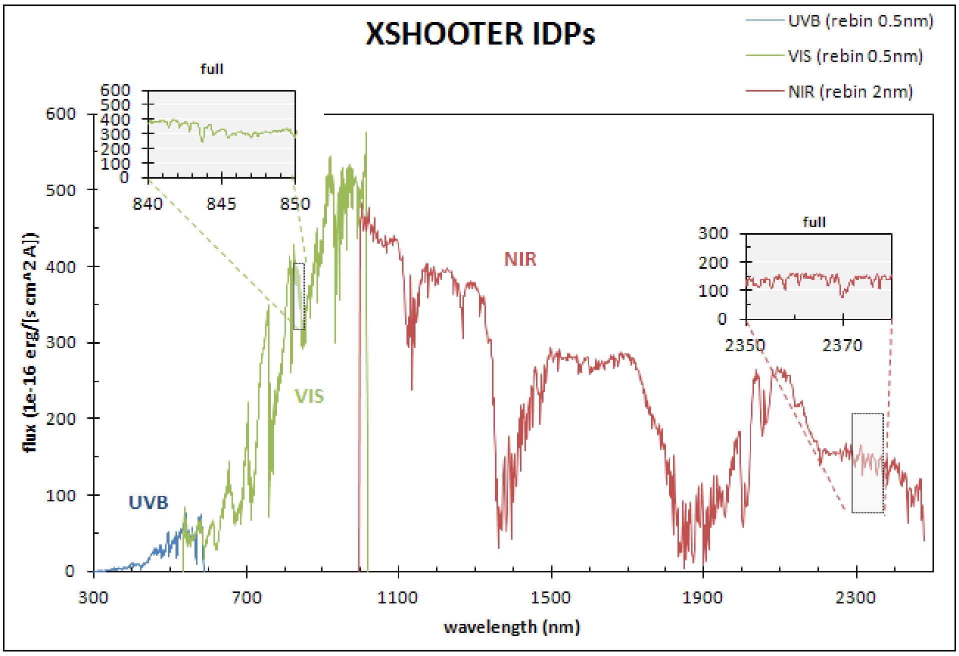

Figure 2. Set of XSHOOTER science-grade spectra. They come from the three arms UVB, VIS and NIR (from left to right). The two small windows display close-ups at full resolution (while the overview plot is rebinned as indicated at upper top). These data products have been processed independently. They nevertheless overlap nicely in the regions around 550 nm, and 1000 nm. The deep absorption troughs e.g. at 1.4 µ are telluric absorption. That kind of atmospheric signature, like also the telluric emission lines on longer exposures, could not be removed with the current automatic processing scheme.

4.3 XSHOOTER

The XSHOOTER spectrograph is a medium-resolution spectrograph (R=10-20,000) with the unique capability to get a full spectrum from 350 nm up to 2.5 µ in one shot. It has three arms, each arm delivers one spectrum. It is operational since 2009-10. The ECHELLE mode is supported with our data products, while the rarely used IFU mode is not supported.

Figure 2 displays a complete set of three XSHOOTER spectra, consisting of data from the UVB, VIS and NIR arm. It also illustrates a limitation of the automated approach for the data reduction. The correction for the telluric absorption and emission features could not be done, it is left to the user. The correction would require interactive selection of correction spectra and optimization of processing parameters.

XSHOOTER spectra are also quite popular. 48,000 spectra are available in total, corresponding to about 16,000 pointings (3 files per pointing). Since they showed up on the archive interface in 2014-06, we have seen about 1,400 download requests until May 2016[7], among them 80 PI requests. There are about 1.5 requests per day, with no signs of flattening. Find the release description here[8].

4.4 GIRAFFE

The FLAMES/GIRAFFE spectrograph is an optical medium-to-high resolution multi-object spectrograph, offering two modes with R ~ 5000 (low-resolution LR) and R ~ 30,000 (high-resolution HR). Up to 130 objects can be observed simultaneously. It is operational since April 2003. Its MOS mode (called Medusa) is fully supported by the PHOENIX process. The spectra are relatively short (one order). Science data products are available for all setups of GIRAFFE-Medusa.

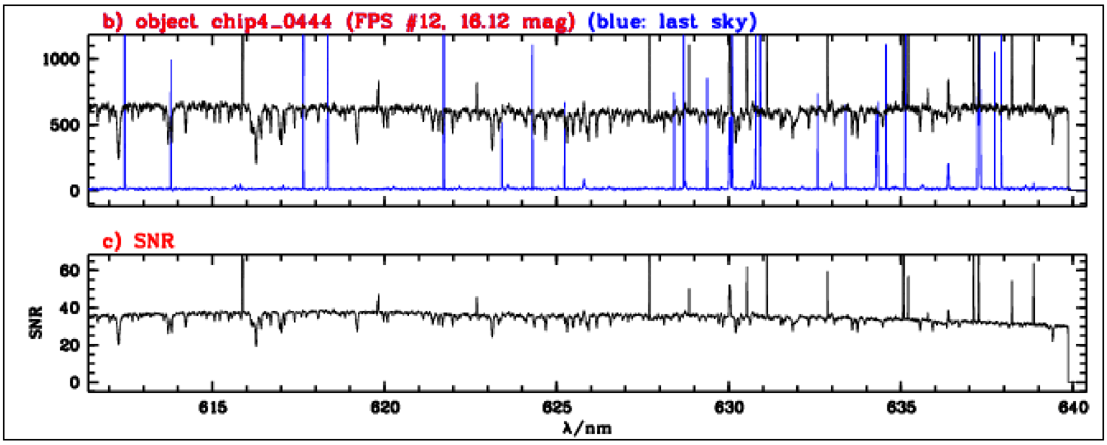

Figure 3. One of 1.2 mio GIRAFFE spectra (top), with its corresponding SNR plot (bottom). Up to N=130 individual fibre pointings can be assigned to science targets in the field-of-view. Each gets its own spectral data product that can be individually selected and downloaded. The blue spectrum in the top panel (the one with values around 0 in the printed version) is one of the sky spectra needed to determine the zeropoint of the flux scale.

A total of 1.2 mio data products is available, processed from about 15,000 input raw files. In Figure 3 we show an example of a GIRAFFE spectrum (upper panel) and its SNR spectrum (lower panel). Since their publication in May 2015, the GIRAFFE spectra received about 400 download requests, slightly less than one per day. Find the release description here[9].

---

[7] Scaled by the factor-2 lower number of spectra offered, and taking into account the shorter time we offer them, the rate of demand is comparably high as for UVES.

[8] http://www.eso.org/observing/dfo/quality/PHOENIX/XSHOOTER/processing.html

[9] http://www.eso.org/observing/dfo/quality/PHOENIX/GIRAFFE/processing.html

4.5 MUSE

The latest, and by far most ambitious, PHOENIX project is for MUSE data products. MUSE is the powerful multi-channel IFU spectrograph installed since 2014 at the VLT. It delivers 3D spectro-images from 24 identical spectrographs with 4Kx4K CCDs each, corresponding to a total of 384 mio pixels. The field of view is 1'x1', mapped with a pixel size of 0.2"x0.2" on a grid of 300x300 spatial pixels and formally delivering 90,000 spectra of 4000 pixels each with every datacube. With its maximum efficiency of 35%, exposure times are often short, so that some nights collect more than 50 of such files, 1 GB each.

Processing these data is a challenge. For the previously mentioned instruments the challenge was to set up the process, but once this was done, it was straightforward to process data created with years-old detector technology by present-day compute hardware. But MUSE is different. It was felt that at least part of our community is not ready to routinely cope with such complex and big data processing. Hence we have started to create MUSE datacubes. At the time of writing they are not yet fully archived, but we expect them to be available very soon[10].

We have created so far more than 1500 MUSE datacubes, most of them combined at the OB level, the others being processed from individual exposures. Many PIs combine 2, 3 or 4 exposures into one OB, with small dithers and rotated field-of-view in order to minimize residual differences between the individual spectrographs. We combine these exposures into a final combined datacube. We have also combined all OBs with more than 4 individual exposures, up to the current hardware-induced limit of N=16 input files.

The MUSE datacubes come fully calibrated for flux, wavelength and astrometry, in physical units and with error estimates. They are science-ready and can be used to fully exploit their scientific information content. This means that there is no source extraction (which would require assumptions on the target nature). One such datacube has about 3-5 GB if all input exposures are fully overlapping, or correspondingly larger size if they only partially overlap. The whole science reduction chain has up to 6 different steps, with 5 pipeline recipes. The resampling of the input pixels to a common grid is done only once, as the latest-possible step, to avoid unnecessary sampling noise.

A standard combined (N=4) datacube (as the one behind the preview displayed in Figure 4) takes about 2 hours to process, on a powerful blade with (at least) 24 cores and (at least) 100 GB memory.

We are currently investigating whether a combination beyond the OB level would be possible, since many PIs design their MUSE OBs such that they correspond to multiple visits of the same targets and the ultimate goal to go deeper than is possible with a single OB.

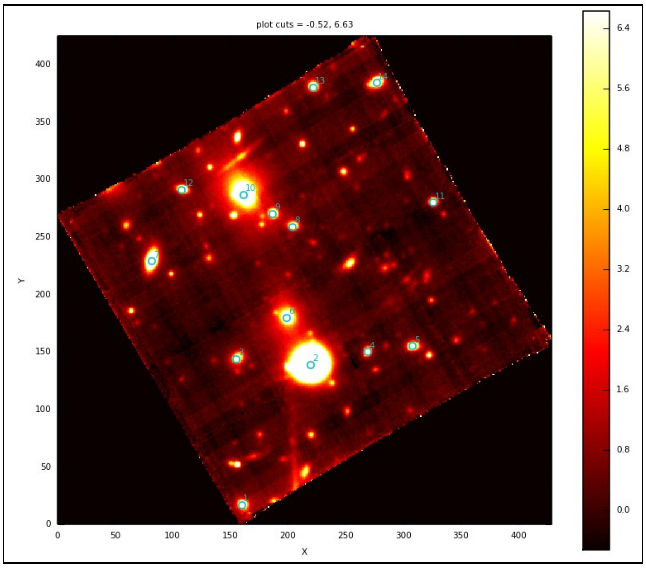

Figure 4. Preview plot of a MUSE datacube. This datacube has been created from N=4 input files. The Figure displays the white-light collapsed (2D) image. The datacube has a full flux-calibrated spectrum for each of the 90,000 pixels visible here.

---

[10] Check out the latest news on http://archive.eso.org/cms/eso-data/eso-data-products.html.

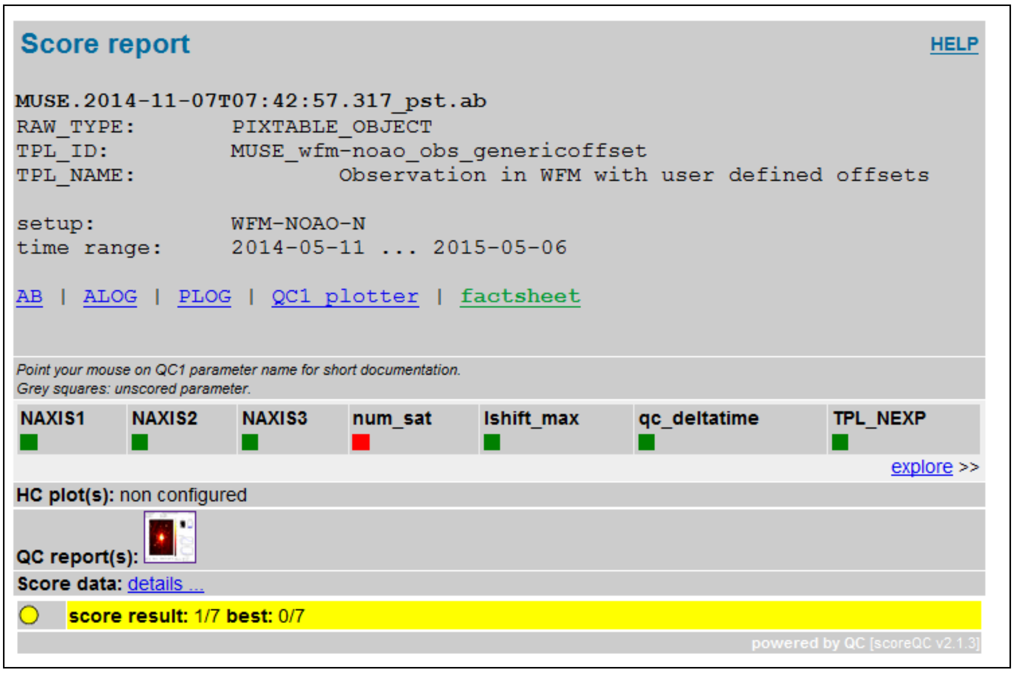

Figure 5. Automatically generated score report for a MUSE data product. The tool has detected a potential issue with saturated pixels, while all other control parameters (7 in total) are within the nominal range and score green.

5. PROCESS QUALITY CONTROL

We have developed the following quality control components for the PHOENIX process:

- QC parameters measuring data properties,
- a database storing these parameters along with other relevant data and processing information,
- scores measuring the compliance of these QC parameters with pre-defined thresholds,
- QC plots and previews for visualization.

As an example for the scores, we show in Figure 5 an automatically created score report. It flags a potential issue with saturation. It is left to the operator of the tool to assess whether there is a quality issue or not. The color boxes can be clicked for dynamic trending plots.

We also use trending plots to check and monitor the performance of key quality parameters. As an example we display in Figure 6 a comparison between user-defined magnitudes and pipeline-extracted fluxes. In theory there should be a narrow correlation, given the short spectral range of the spectra. Despite the uncertainties of the users' input creating some scatter, we can see a correlation confirming that the pipeline indeed extracts the scientific signal in the correct way.

Another automatic check is displayed in Figure 7. This plot shows the SNR vs. the total counts, as one data point per input file (from the brightest fibre). Ideally this curve should be a square-root law shaped by photon shot noise, if no other noise source dominates the noise budget. Additional noise sources would reveal if the curves level off. This plot confirms that the SNR is indeed dominated by shot noise, i.e. by the fundamental limitation of the observation process. This gives confidence that the pipeline processing is correct, even at very high SNR values (up to ~800).

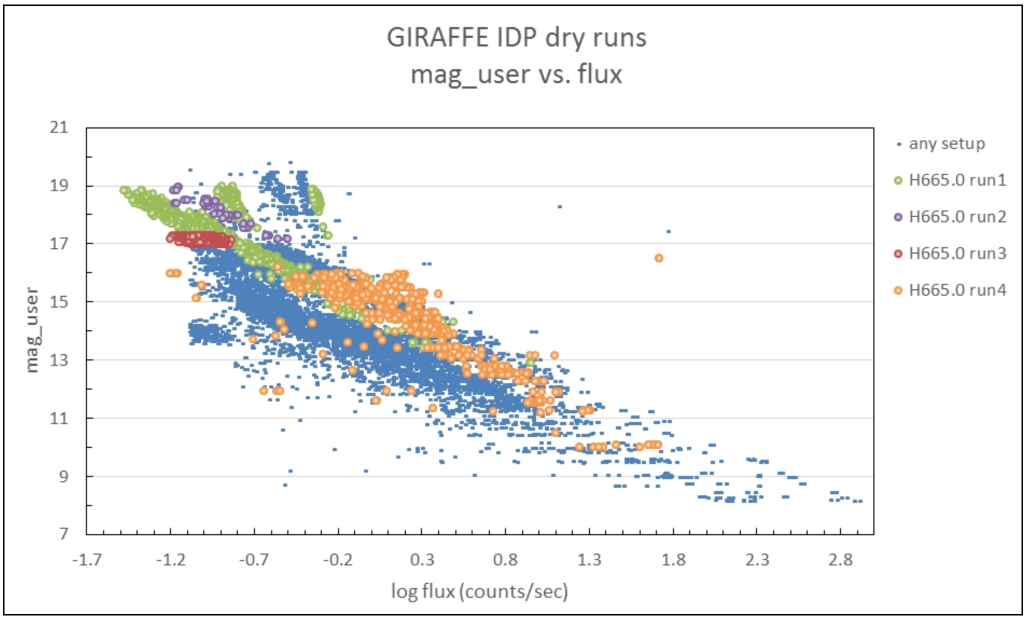

Figure 6. User-defined magnitudes vs. average flux of GIRAFFE spectra, for selected programmes and settings. PIs can enter the magnitudes of their targets in the fibre preparation tool. These numbers are not perfectly defined and constrained (e.g. the users are not forced to use a standard photometric system), but in a statistical sense they can be used as an independent check of the pipeline-delivered extracted fluxes of the data products. We do see a correlation from the upper left part to the lower right, constituted e.g. by the green or purple "clouds" (grey in the print version). Isolated "clouds" likely come from PI input using different photometric systems.

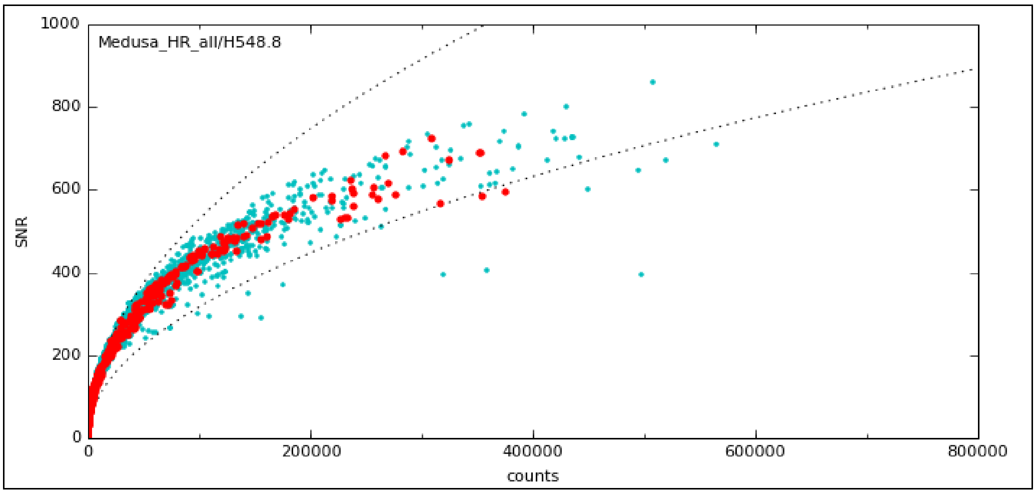

Figure 7. Trending plot to control the quality of the GIRAFFE data processing. We plot the pipeline-extracted signal-to-noise ratio (SNR) versus the total counts in the extracted spectrum. The dark, larger data points (red in the electronic version) refer to a reference setting, the other smaller points show all data points for the high-resolution mode. Only the brightest fibre is displayed, i.e. one point per input file, totaling to more than 14,000 data points.

6. OUTLOOK

We have demonstrated that it is possible to create science-grade data products from selected instruments of a large ground-based observatory, with reasonable and moderate efforts. This has been achieved by the QC group at ESO for the spectroscopic workhorse instruments UVES, X-SHOOTER, and GIRAFFE. We are about to extend this approach to the high-data volume, complex MUSE IFU instrument, with its data cubes coming on-line soon.

Key components for this success have been:

- stable, performant, certified and well-understood pipelines;
- certified and archived master calibrations, created in a quasi-real time quality control loop;
- efficient archive access for downloads and uploads;
- the flexible workflow tool PHOENIX that has been built around existing components from the daily QC workflow for the processing of master calibrations.

The publication of the science-grade data products required also a data validation component, a data products standard and a set of archive interfaces to offer the products to the users.

The PHOENIX production line for data products is quite efficient. Once configured and tested for a new project, it takes only a small fraction of an FTE to maintain the product stream. The investment in hardware is significant only for the MUSE project.

The spectroscopic data products are quite popular among the ESO community. The download rates for UVES and X-SHOOTER are quite high, with an average number of 3 and 1.5 requests per day, respectively. Currently there is no sign of flattening.

As the next project there will likely be IFU datacubes from the near-IR instrument KMOS. We are currently also investigating whether we can create MUSE *deep* datacubes, with a superposition of MUSE datacube across the (rather technical) OB limitation. This would be a useful step since many MUSE projects are designed to go deep, and only the final and correct superposition can fully exploit the scientific content of the MUSE data.

Acknowledgements. The GIRAFFE and MUSE science data projects at ESO are the outcome of the efforts of many people. Specifically we would like to thank Peter Weilbacher from Astronomical Institute Potsdam for many useful discussions, hints and help with the MUSE pipeline, and Ralf Palsa, Alberto Micol, Danuta Dobrzycka, Burkhard Wolff, Nausicaa Delmotte, and Laura Mascetti (all from ESO).

REFERENCES

[1] Hanuschik R., Neeser M., Hummel W. and Wolff B., “Scoring – a novel approach toward automatic and reliable certification of pipeline products”, Proc. SPIE 7016, [70160Q] (2008).

[2] Freudling W. et al., A&A 559, A96 (2013).

[3] Arnaboldi M., Retzlaff J. The Messenger, 146, 45 (2011).

[4] Hanuschik, R., “Phoenix: automated science processing of ESO-VLT data”, Proc. SPIE 9149, [91491V] (2014).