



European Organisation
for Astronomical
Research in the
Southern Hemisphere

Organisation Européenne
pour des Recherches
Astronomiques dans
l'Hémisphère Austral

Europäische Organisation
für astronomische
Forschung in der
südlichen Hemisphäre

Data Management and Operations Division


PHOENIX – the process to create science-grade data products

Doc. No.: VLT-MAN-ESO-19600-6104

Issue: 1

Date: 2014-03-03

Prepared by:	Reinhard Hanuschik	2014-03-03	
	Name	Date	Signature
Validated by:	Martino Romaniello		
	Name	Date	Signature
Approved by:	Michael Sterzik		
	Name	Date	Signature

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 2 of 19
--	---	----------------------------------	--

CHANGE RECORD

ISSUE	DATE	SECTION/PARAGRAPH AFFECTED	REMARKS
1.0	2014-03-03	new document	Originated from Internal Memo "Largely automated science processing (working title: <i>phoenix</i>)" R. Hanuschik, 2012-05-10



TABLE OF CONTENTS

1	Introduction	4
1.1	Purpose	4
1.2	List of Abbreviations & Acronyms	4
1.3	Dictionary	4
1.4	List of Applicable and Referenced Documents	5
2	Phoenix: the process	6
3	Preconditions	7
3.1	Process criteria	7
3.2	Instrument mode selection	8
4	Phoenix: Workflow and tools	9
4.1	Steps	9
4.2	Breakpoints	11
4.3	The workflow tool <code>phoenix</code>	11
4.4	Other tools used	12
4.5	Processing	12
4.6	Data management	13
4.7	Quality control and scoring	14
4.8	Previews	16
4.9	Process monitor	17
4.10	Product ingestion	17
5	Day-to-day processing (stream) vs. history processing (batch)	18
6	Supported instruments	18
7	Limitations of PHOENIX	19



PHOENIX – the process to create science-grade data products

1 Introduction

1.1 Purpose

This document describes the PHOENIX process. It has been created and maintained by the QC group. PHOENIX produces quality-controlled science-grade products and offers them to the community via the phase 3 archive interface.

1.2 List of Abbreviations & Acronyms

AB	Association Block
ASG	Archive Science Group
DFS	Data Flow System
EDP	External data product
IDP	Internal data product
NGAS	Next-generation archive system: the ESO archive for science and calibration data (raw and products)
PI	Principal Investigator (the owner of a programme or a run)
PSD	Pipeline Systems Department
QCG	Quality Control Group
SDP	Science Data Products Group


1.3 Dictionary

Association Block (AB) - text file that contains all the information required to process a self-consistent set of input raw files. In the PHOENIX context, historical science ABs are downloaded from the central QC repository (on the QC web server *qcweb*) and processed by the pipeline.

dfos (data flow operations system) – set of workflow tools used by QCG to control data flow operations

PHOENIX – the process to create quality-controlled science products. PHOENIX is not an acronym but a name, reminding of a process which had been terminated but has been revived, with better performance and improved usefulness.

phoenix – the name of the workflow tool supporting the PHOENIX process

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 5 of 19
--	---	----------------------------------	--

PI packages – data packages created by QCG until September 2011 in which all data related to a specific run (raw and product, calibration and science) were delivered to the PI.

1.4 List of Applicable and Referenced Documents


[1] Generation and SAF Publication of Internal Data Products: Deployment Plan. VLT-PLA-ESO-19300-6075, M. Romaniello et al., 2014-02-14

[2] PHOENIX-XSHOOTER automatic science reduction. Report from the exploration phase (dry runs). R. Hanuschik, 2014-02-11

[3] UVES-Echelle IDP release description, version 1.1. R. Hanuschik, 2013-12-06,
http://www.eso.org/sci/observing/phase3/data_releases/uves_echelle_v1.1.pdf

[4] “Science data products standard”, GEN-SPE-ESO-33000-5335. Version 5, 2013-01-11,
<http://www.eso.org/sci/observing/phase3/p3sdpstd.pdf>

[5] “DFO handbook”, VLT-SPE-ESO-19600-3986. Version 4.0, R. Hanuschik, 2012-07-02.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 6 of 19
--	---	----------------------------------	--

2 Phoenix: the process

PHOENIX is the process created and maintained by the QC group to produce quality-controlled science-grade pipeline products and to offer them to the community via the phase-3 archive interface. These internally produced data products are called IDPs (*Internal Data Products*). The phase-3 archive interface is also fed with a second type of products, the *External Data Products* (EDPs) mandatorily provided by PIs of Public Surveys or Large Programmes and open to other contributors as well.

The PHOENIX process replaces an older process in which the QC group processed science data and delivered them in *PI packages* to the users (PIs), along with raw and processed calibrations. The main reason for terminating the old process was that it was felt that the product quality not always matched the user's expectations. Also, it was not possible to deliver these science products to the whole community.

It was decided to create a new process for science products, focussing on known quality and therefore on science-grade pipelines. Rather than attempting to achieve completeness across all instruments, the new process is focussing on selected instrument modes¹. The new science products are available to all users while respecting the proprietary restrictions inherited from the corresponding raw files, as per ESO policy.

This new process is called PHOENIX and is in place since November 2013, with the creation of the UVES data stream by the Quality Control Group.


The differences between the old and the new process are:

- only selected instruments (or instrument modes) are supported by PHOENIX;
- the PHOENIX process feeds an archive IDP stream (internal data products), through which the data products become available in principle to any archive user²;
- the PHOENIX process is much more efficient and to a large extent automatic;
- it can support multiple (re-) processing.

PHOENIX is not an acronym but a name for a process which had been terminated but has been revived, with better performance and improved usefulness.

¹ A more complete coverage across instruments is achieved with the interactive reflex workflows.

² The IDPs follow the ESO access policy: after the end of the proprietary period they are public to anyone, while before their access is restricted to the PI.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 7 of 19
--	--	----------------------------------	--

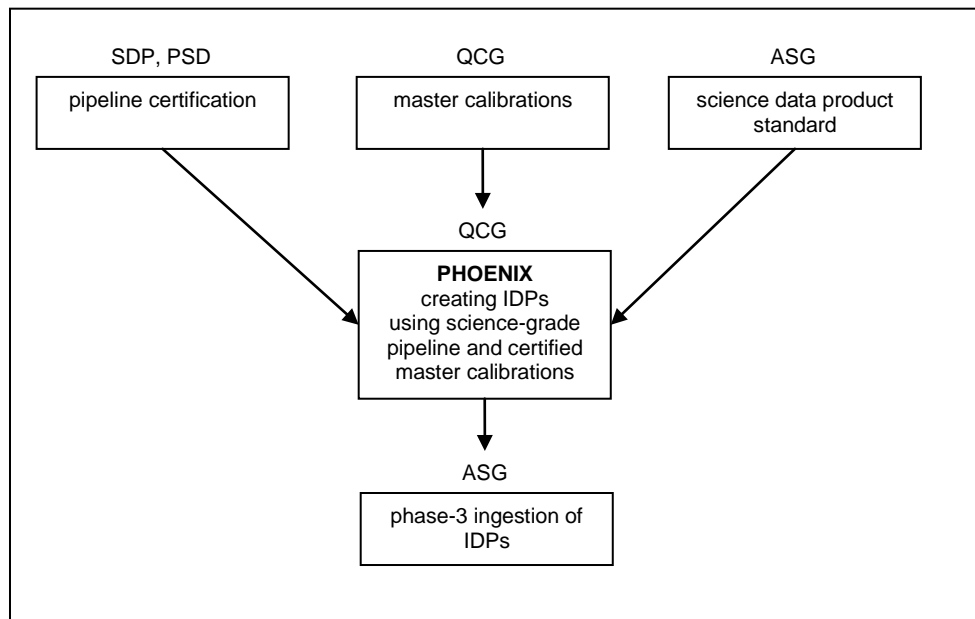



Figure 1. Components and groups contributing to the IDP workflow produced with PHOENIX.

3 Preconditions

3.1 Process criteria

For an instrument mode to be selected for PHOENIX processing, a number of criteria must apply (see also Figure 1).

- A certified pipeline must exist. The certification is done by the Science Data Products Group (SDP) in close collaboration with the Pipeline Systems Department (PSD). They normally review the algorithms used by the pipeline recipes and the calibration plan. The error propagation needs to be supported from the calibration data up to the final products. Usually the science products have to come in physical units. If a pipeline matches these criteria, it is called science-grade and qualifies for the PHOENIX process.
- Master calibrations must be available for the history of the considered instrument mode. The creation of master calibration by QCG involves a certification process for the calibrations. Effectively this process takes into account all kinds of interaction like analyzing trends, investigating issues, communication with Paranal staff. Once certified, the master calibrations contain all the knowledge about the calibrations and then can be used for an *automatic* science reduction process.
- The ingestion of the products into the phase-3 archive is an integral part of the PHOENIX process. The products need to be phase-3 compliant, and a standard for these data products must be defined by the Archive Science Group (ASG). The currently existing standards are defined in [4]. The compliance with the required data standard is either provided by a conversion tool or by the pipeline.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 8 of 19
--	---	----------------------------------	--

- To be efficient and largely unsupervised, all associations (ABs) between science and calibration products need to already exist. They are harvested from the standard QC workflow. The corresponding master calibrations are downloaded from a central repository. Since the day-to-day quality control of the calibrations is an interactive process and requires frequent investigations and decisions, it is of paramount importance for the PHOENIX process to build upon this work and not repeat the calibration certification process. The easiest way of doing this is to preserve the associations and to re-use the existing master calibrations.

The idea behind PHOENIX is that, if all the beforementioned conditions apply, the processing of science data does not require further decisions: all QC checks have been done, instrument events and data quality issues have been taken into account with the certification/rejection mechanism of the QC review, and the pipeline has been reviewed for quality and stability. Then, the raw science data can be processed to the science-grade level without the need to know the observational constraints, prevailing observing conditions or scientific goals.

This is the processing level of the Internal Data Products (IDPs), in contrast to advanced data products (External Data Products) which are often optimized to their specific scientific goal.


3.2 Instrument mode selection

For the PHOENIX process, we consider only selected instrument modes:

- instrument modes with homogeneous and well-known data properties (science data types, calibration plans, headers stable over years);
- instruments with sufficiently demanded modes (we do not aim at completeness but at optimal “return on investment”).

So far the PHOENIX process has been implemented for spectroscopic modes only, but there is no principal reason for this constraint³. The future strategy for IDPs is described in [1].

³ The selection is based on the certification status of the pipelines and the availability of the phase-3 data standards.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 9 of 19
--	---	----------------------------------	--

4 Phoenix: Workflow and tools

The fundamental unit for the PHOENIX workflow is a full observing date with science data being taken. The workflow for a complete date worth of science data is described in the following Table 1.

The workflow is supported by the workflow tool `phoenix`. This tool is a shell script wrapper around existing tools from the operational `dfos` tool suite⁴.

4.1 Steps

The first step is the download of ABs from the central repository. For the science ABs this is the QC web server `qcweb`. The exact directory is configured in the configuration file of `phoenix`. It is the operational directory `public_html/<instr>/logs/<date>` where `<instr>` is the operational instrument name. The ABs have been fed during the historical processing from the day-to-day QC operations.

A logical filter as defined in the configuration file makes sure to select the science data type and possibly setup selected for `phoenix`, since in general not all science data from a given instrument will be `phoenix`-supported.

The module for AB editing is important. It brings the historical ABs to a common standard. The historical ABs reflect the history of increasing knowledge, pipeline improvements, calibration plan changes, and association rule modifications. Without such a module, the `phoenix` processing would e.g. suffer from imperfect early calibration strategies while now better ones are known. Or the ABs would even fail because historical pipeline versions used parameters that are not used by now.

It is in the core of the PHOENIX concept that existing associations, though modified, are accepted as *is*. They are not re-done. This is on purpose and reflects the fact that the historical ABs always contain the correct association to *certified master calibrations*. All known instrument events were taken into account when the historical associations were done. It is the intention of `phoenix` to build on that knowledge.

The download modules are purely technical, they help with the efficient, multi-threaded download of raw files and master calibrations before and during the processing.

The processing itself is done on `muc08` with the `condor` system, with up to 30 processing jobs running in parallel. This is true for the pipeline processing itself, as well as for the preparation of the QC reports.

In the QC report step, customized QC parameters and other information (including OB grades and comments) are extracted from the data and inserted into a QC1 database. These parameters are used for the quality control of the PHOENIX process (not of individual products).

The final step is the automatic distribution of the products to their final directories:

- the science product fits files go to the final data directory `$(DFO_SCI_DIR)/<date>`;


⁴ `dfos` is documented in the ESO-internal URL <http://www.eso.org/~qc/dfos/>.



Part 1: Preparation			
<ul style="list-style-type: none"> download the science ABs from the central repository apply the configured filters (on data type, setups) edit the ABs with a configured plug-in, customized to the instrument mode prepare the raw file download, mcalib download, condor⁵ execution jobs 	phoenix -d <date> -C ↓	phoenix -d <date> -P ↓	phoenix -d <date>
Part 2: Execution			
<ul style="list-style-type: none"> call the raw download job and the mcalib download job call the AB execution file with condor, in parallel mode call the QC job execution file with condor, in parallel mode refresh the processing monitor 		↓	↓
Part 3: Finishing			
<ul style="list-style-type: none"> move all products (fits, logs, plots) to final directories 		phoenix -d <date> -M ↓	↓
<ul style="list-style-type: none"> rename fits files 			
<ul style="list-style-type: none"> update statistics, processing monitor 			
Part 4: Conversion and ingestion			
<ul style="list-style-type: none"> call conversion tool (if configured) call ingestion tool 		ingestProducts -m SCIENCE -d <date>	

Table 1. Sketch of the workflow for phoenix. The normal flow is the one in the rightmost column. To maximize flexibility, this workflow can be interrupted at defined breakpoints.

⁵ Condor is the batch-queue system used by QCG to schedule and manage parallel execution of processing jobs. See <http://www.eso.org/~qc/dfos/condor.html> for more.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 11 of 19
--	---	----------------------------------	---

- the log files and modified ABs go to $\${DFO_LOG_DIR}/<date>$ and to the *qcweb* server;
- the QC reports go to $\${DFO_PLT_DIR}/<date>$ and to the *qcweb* server;
- the fits files are finally ingested as IDPs with the archive Phase 3 infrastructure.

4.2 Breakpoints

The workflow has three possible breakpoints (Table 1). The first one is after AB download and modification. At that point the QC scientist may want to inspect individual ABs for proper edited content, or execute individual ABs (rather than doing the bulk-processing with *phoenix*). The workflow tool can stop at that breakpoint⁶ with the option `-C`. This breakpoint is relevant only in the commissioning phase of a new *phoenix* instance.

The second breakpoint is after processing. With option `-P`, the workflow tool will stop here⁷. The QC scientist can inspect pipeline products one by one, or fine-tune the QC report or QC1 parameters. If the second breakpoint has been set, the tool can be called later with option `-M` to resume the final data distribution⁸. Again, this breakpoint is relevant only in the commissioning phase.

The third breakpoint is after the final data distribution. The *phoenix* tool will always stop here. This breakpoint is needed before the final ingestion of the IDPs, because experience shows that it is a good idea to de-couple processing and ingestion. The ingestion tool might have an issue, or the conversion is not yet properly done, or checks need to be implemented. The conversion into the science data products format (if needed), and the final ingestion are provided by a separate ingestion tool, the standard *dfos* tool *ingestProducts*⁹ wrapped around the DFS provided phase-3 ingestion tool.

The *phoenix* tool called without option can execute the whole workflow in one go, without the first and second breakpoint. This is the standard way of calling it.

The fundamental unit for the workflow execution is a day. The tool can also be called for a full month (parameter `-m <month>`).

4.3 The workflow tool *phoenix*

The workflow tool *phoenix* is built around off-the-shelf components from the *dfos* tool suite. The main components are:


- `createJob` (to create the download and execution jobs);
- `processAB` (the interface to the *esorex* calls of the pipeline recipe);

⁶ C stands for (AB) *creation*.

⁷ P stands for *processing*.

⁸ M stands for *moving* (final distribution).

⁹ In its version 2.x which is enabled for IDP ingestion. See more under <http://www.eso.org/~qc/dfos/ingestProducts.html>.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 12 of 19
--	---	----------------------------------	---

- `processQC`, `scoreQC` and the QC scripts for science data (providing the QC reports, QC1 parameters, scoring)
- `moveProducts` (to distribute the pipeline products, logs and plots)
- `renameProducts` (to rename them to their final product name)

Furthermore there are helper tools like

- `ngasClient` (download of raw and master calibration files),
- `updateAB`,
- `monitor tools` (`dfoMonitor`, `mucMonitor`).

For a complete list check the tool documentation¹⁰ and document [3].

On the short side, all versions of used *dfos* need to be frozen, because most *dfos* updates relate to operational changes, while the phoenix environment is decoupled from operations and requires stability. Some unavoidable updates will be needed for the future nevertheless. This will create some additional maintenance workload but it seems minor at the moment.

4.4 Other tools used

The bookkeeping of a mass production process is important. It is provided by a customized version of the *dfos* tool `histoMonitor`, called `phoenixMonitor`. It essentially creates a hierarchical set of overview pages¹¹ that contain daily information about processed ABs, failure or success, setups and modes used and overall statistics. They are essential for controlling the process health.

Metrics of the `phoenix` process like number and size of created IDPs and execution times are collected and stored in a database. They are monitored on the WISQ monitor¹².

4.5 Processing

The multi-core machine `muc08` is currently used for the PHOENIX processing. It can process as many as 30 jobs in one go. So far, the available processing power for IDP processing is more than sufficient. To give an impression, the historical batch of the entire UVES-Echelle dataset (2000-2013) took about 5 days of net processing time on `muc08`, including overheads for data download (but not for ingestion). In Figure 2 the cascade performance of a set of XSHOOTER jobs is displayed which took in this example a total of 18.8 min for processing the 85 XSHOOTER science spectra for date 2013-11-08.

¹⁰ <http://www.eso.org/~qc/dfos/phoenix.html>

¹¹ UVES: http://qcweb/UVESR_2/monitor/FINISHED/histoMonitor.html;
XSHOOTER: http://qcweb/XSHOOT_R/monitor/FINISHED/histoMonitor.html

¹² http://www.eso.org/observing/dfo/quality/WISQ/FULL/trend_report_products_ALL_HC.html



For each instance of PHOENIX, there is one user on muc08. For the current two supported instances, there are users *uves_ph* and *xshooter_ph* (where *ph* stands for *phoenix*). There is also the generic user *scipro* (science processing) managing, among other things, the central tool distribution.

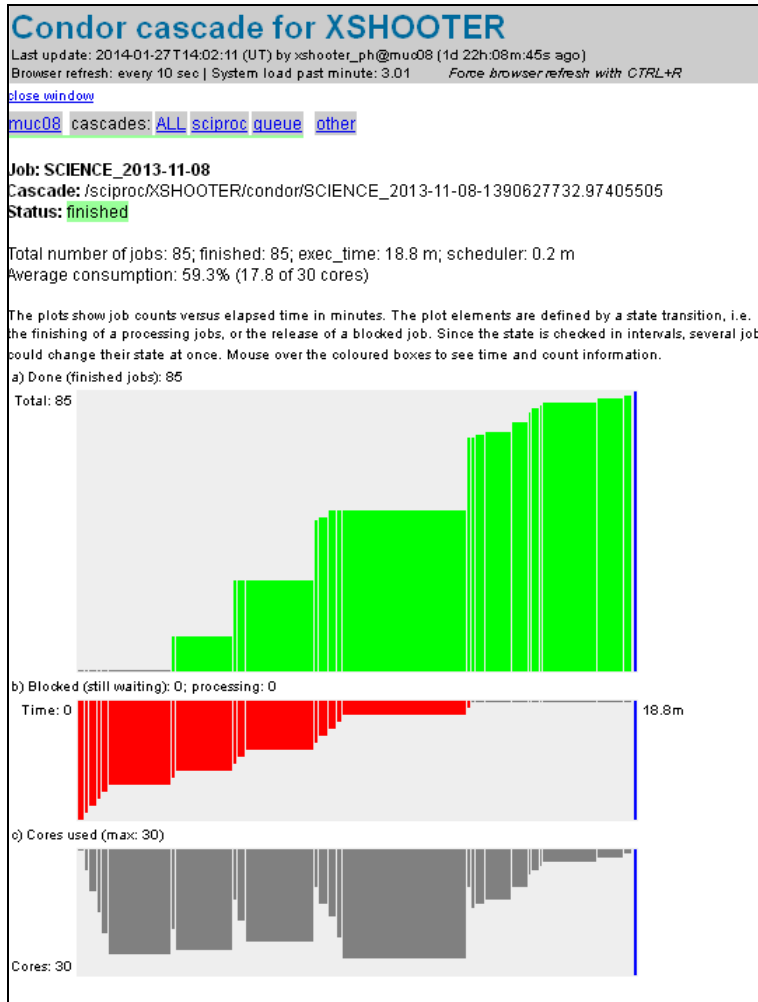



Figure 2. Processing of a condor batch of 85 XSHOOTER science ABs. Horizontal axis is time in minutes, vertical axis is number of jobs. With each job requiring about 2-4 minutes, the total execution time was 18.8min (including overheads). Green marks the finished jobs, red the scheduled ones, grey the currently used cores (maximum N=30).

4.6 Data management

Raw data are downloaded from the archive in up to 10 parallel streams. They are immediately deleted at the end of a *phoenix* day batch. See Figure 3 for a complete picture of the *phoenix* data flow.

The master calibrations are also downloaded from the archive in parallel batches. The strategy for their management is different since they can be used for a certain range of days. They are deleted in a moving window of +/- 28 days, meaning if there is a new *phoenix* job for a nearby date, it can make

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 14 of 19
--	--	----------------------------------	---

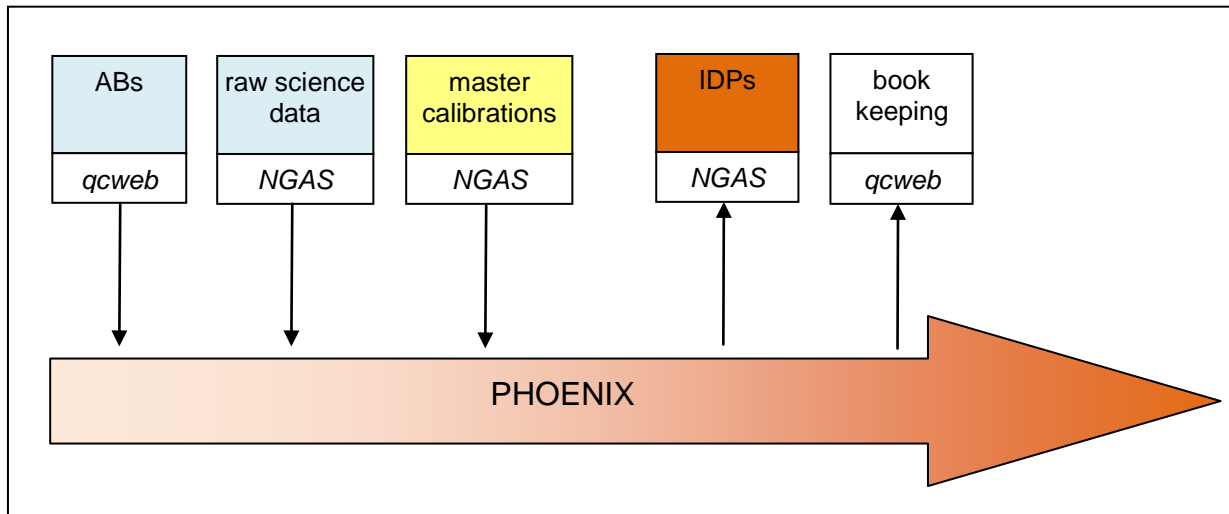


Figure 3. PHOENIX data flow.

use of the already downloaded masters¹³. If a new date is called which is off the current calibration dates by more than the specified range, then these master calibrations are deleted and the new ones are downloaded. This strategy provides a reasonable and self-organized compromise between quick availability of master calibrations and disk space management.

4.7 Quality control and scoring

The main components of the quality control are already implemented by the process itself: usage of certified master calibrations, and usage of a certified pipeline. Still, it is very useful to have an automatic quality control of the science data products. There are a few useful QC parameters to be measured and monitored by the `phoenix` process:

- saturation (normally saturation comes only for parts of the raw file, and the science products can still be potentially useful, provided the saturation is marked somehow);
- for certain observing strategies, products can show negative signal; as for the example of XSHOOTER, negative flux, and a negative SNR, can occur for NODDING or OFFSET observation and indicate an issue with the SKY subtraction, or with the nature of the signal; this cannot be solved by the pipeline and is flagged;
- incomplete templates might indicate an issue related to the observing conditions and therefore are flagged.

On the other hand, a net flux close to zero is not considered a quality issue, since it might correspond e.g. to a valid non-observation, or an emission-line object.

For the flagging `phoenix` uses the standard `scoreQC` tool used by the `dfos` suite.

¹³ They never come strictly day by day but a frequent pattern is that some come daily (like BIAS frames), others every third day (like time-consuming flats), others when conditions are perfect (like standard stars).



AB NAME	INDEX	COMPL	AB LOG	RECIPE	RAW_TYPE	SETUP	AB STATUS	P LOG	T_EXEC	QC REPORT	SCORE	RE	COMMENT	
SHOOT 2010-03-28T23:57:16.831	lpl.ab	sci001	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_UVB	100k/1pt/hg_UVB_1.0x11_SLOT	OK	FLOG	3.4+0.1	QC	✓	(0/5)	
SHOOT 2010-03-28T23:57:21.832	lpl.ab	sci002	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_VIS	100k/1pt/hg_VIS_0.9x11_SLOT	OK	FLOG	8.6+0.2	QC COVER	✓	(0/5)	
SHOOT 2010-03-28T23:57:24.912	lpl.ab	sci003	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_NIR	50.0000000_NIR_0.9x11_SLOT	OK	FLOG	2.1+0.1	QC	✓	(0/5)	
SHOOT 2010-03-29T01:23:13.678	lpl.ab	sci004	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_UVB	100k/1pt/hg_UVB_1.3x11_SLOT	OK	FLOG	2.1+0.1	QC	✓	(0/5)	
SHOOT 2010-03-29T01:23:18.908	lpl.ab	sci005	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_VIS	100k/1pt/hg_VIS_1.2x11_SLOT	OK	FLOG	5.2+0.2	QC COVER	✓	(0/5)	
SHOOT 2010-03-29T01:23:22.080	lpl.ab	sci006	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_NIR	1200.0000000_NIR_1.2x11_SLOT	OK	FLOG	1.5+0.1	QC	○	(1/5)	more than 20000 pixels saturated
SHOOT 2010-03-29T02:41:14.894	lpl.ab	sci007	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_UVB	100k/1pt/hg/1x2_UVB_0.8x11_SLOT	OK	FLOG	1.7+0.1	QC	✓	(0/5)	
SHOOT 2010-03-29T02:41:19.894	lpl.ab	sci008	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_VIS	100k/1pt/hg/1x2_VIS_0.7x11_SLOT	OK	FLOG	4.2+0.2	QC COVER	✓	(0/5)	
SHOOT 2010-03-29T02:41:22.926	lpl.ab	sci009	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_NIR	1800.0000000_NIR_0.6x11_SLOT	OK	FLOG	1.4+0.1	QC	○	(1/5)	more than 20000 pixels saturated
SHOOT 2010-03-29T03:58:44.839	ab	sci010	compl.	OK	xsh_scired_slit_stare	SCI_SLIT_STARE_UVB	400k/1pt/hg_UVB_0.5x11_SLOT	OK	FLOG	1.9+0.2	QC	✓	(0/3)	
SHOOT 2010-03-29T03:58:50.100	ab	sci011	compl.	OK	xsh_scired_slit_stare	SCI_SLIT_STARE_VIS	400k/1pt/hg_VIS_0.4x11_SLOT	OK	FLOG	4.1+0.2	QC COVER	✓	(0/3)	
SHOOT 2010-03-29T03:58:52.908	ab	sci012	compl.	OK	xsh_scired_slit_stare	SCI_SLIT_STARE_NIR	1.0000000_NIR_0.4x11_SLOT	OK	FLOG	1.4+0.2	QC	✓	(0/3)	
SHOOT 2010-03-29T03:58:42.795	lpl.ab	sci013	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_UVB	400k/1pt/hg_UVB_0.5x11_SLOT	OK	FLOG	8.4+0.2	QC	✓	(0/5)	
SHOOT 2010-03-29T03:58:48.015	lpl.ab	sci014	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_VIS	400k/1pt/hg_VIS_0.4x11_SLOT	OK	FLOG	11.7+0.3	QC COVER	✓	(0/5)	
SHOOT 2010-03-29T03:58:51.137	lpl.ab	sci015	compl.	OK	xsh_scired_slit_nod	SCI_SLIT_NOD_NIR	2.0000000_NIR_0.4x11_SLOT	OK	FLOG	2.2+0.2	QC	✓	(0/5)	

Figure 4. Processing monitor for XSHOOTER, scores and comments provided by *phoenix*.

For the UVES *phoenix* process only saturation is scored. For XSHOOTER the scoring scheme is more advanced: it scores up to seven parameters, includes them in a score flag and writes it as a 7-bit character into the product headers¹⁴. There is also some automatic, score-based commenting which helps to mark potential issues on the processing monitor (Figure 4). For future instruments, the scoring scheme will be customized as well.

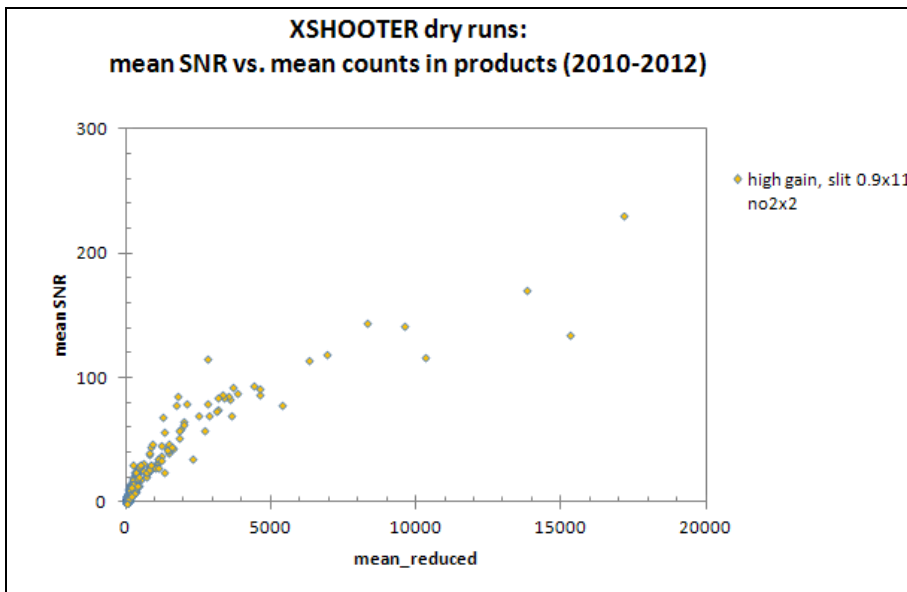


Figure 5. PHOENIX QC parameters and process health. Mean SNR vs. mean counts, for all XSHOOTER VIS arm spectra between 2010 and 2012-12-31, high-gain, slit 0.9x11 dataset (with 1x1 or 1x2 binning). This plot helps monitoring the health of the extraction procedure. From [2].

¹⁴ The score flags cannot be displayed on the archive interface but they are accessible in the header of the downloaded files.



With the QC parameters it is possible to monitor the overall quality of the `phoenix` process for a given instrument. One of the most important quality parameters for spectroscopic products is the signal-to-noise ratio (SNR) that can be plotted versus the mean signal of the input data in order to monitor the extraction quality (see Figure 5 and document [2]). This parameter constituting a square-root law, demonstrates the health of the extraction, and more generally, reduction algorithms.

The QC1 parameters are inserted into a QC1 database table (specific for each instance of the PHOENIX process).

4.8 Previews

Related to quality control and scoring is the production of QC reports which provide mainly a preview of the final product (Figure 6, Figure 7). They are useful to assess e.g. the quality of the flux calibration, and for the investigation of outliers.

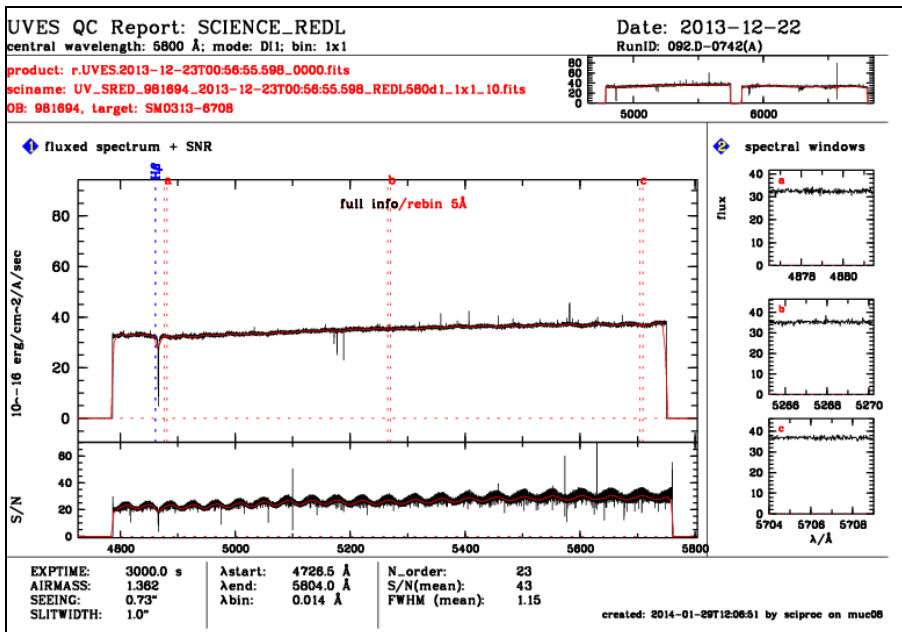


Figure 6. QC report for the UVES IDPs, with the flux-calibrated spectrum, the error file, close-up windows and QC parameters.

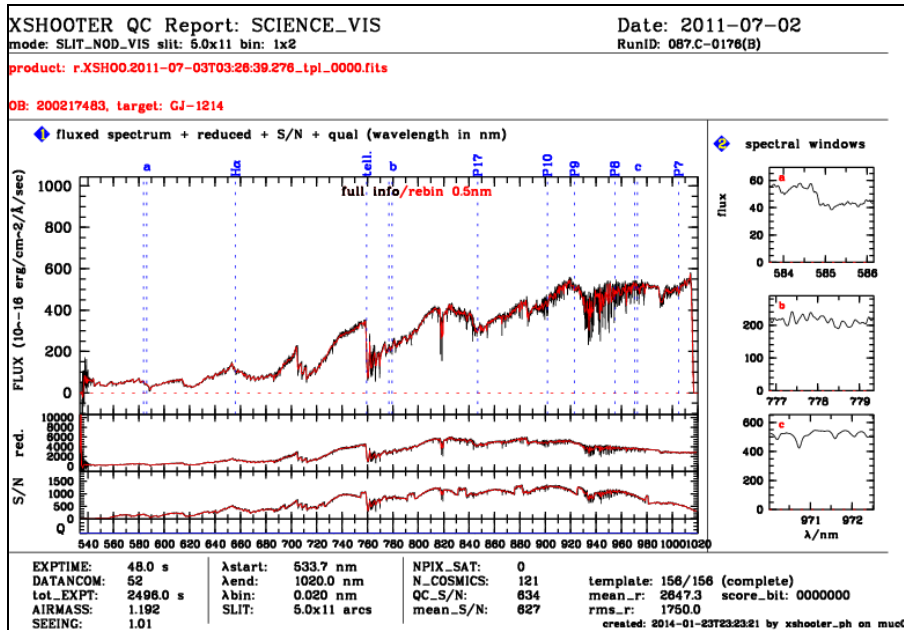



Figure 7. QC report for the XSHOOTER IDPs, with an overview of a VIS-arm spectrum, the SNR, close-up spectral window, and QC parameters. More in [2].

4.9 Process monitor

The process monitor for the PHOENIX process is provided by the tool `phoenixMonitor`. It displays an overview of all so far processed dates, with a hierarchical scheme in which one can start from an annual overview and cascade down to individual products, their QC flags, pipeline logs or previews.

4.10 Product ingestion

The IDPs as generated by `phoenix` are finally ingested by the `ingestion-tool` which is part of the phase-3 infrastructure. The final selection which product(s) to be ingested depends a bit on the instrument mode. A common pattern is to have the flux-calibrated product (if available), the corresponding error, sky spectrum (if available), quality flag (same), in one single multi-column table which constitutes the IDP. Inclusion of ancillary files into the data package is supported by defining appropriate header keywords. The ingestion is fully automatic and is started as a batch, typically for a full month. This is the last step in the workflow, see Table 1.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 18 of 19
--	---	----------------------------------	---

5 Day-to-day processing (stream) vs. history processing (batch)

The typical PHOENIX project has two parts: a batch related to the processing of the historical (so far accumulated) data, and a stream related to the processing of the current data.

The batch processing is done entirely detached from the current data stream. It is executed as a project, with an exploration phase (exploration of the data history, investigation of the backwards compatibility of the pipeline and of the ABs), and a mass-production phase. It is important to have a monitoring of the product data quality, in the above sense of a process quality.


The stream processing can be seen as part of the operational workflow of a QC scientist. As soon as a month worth of calibration data has been completely finished, with all master calibrations certified and archived, a new monthly `phoenix` batch can be called. This seems a good compromise between the requirement “calibrations first” for the QC scientist, and the commitment to deliver IDPs close in time to the community. The latter commitment is a relaxed one anyway since the data will be accessible to the PIs only for the first year.

6 Supported instruments

Currently there is the UVES Echelle (slicer and non-slicer) process operational¹⁵. Next the XSHOOTER IDPs will follow, with their dry runs already finished¹⁶. The currently planned next steps are listed in the IDP deployment plan [1].

¹⁵ Monitored under http://qcweb/UVESR_2/monitor/FINISHED/histoMonitor.html.

¹⁶ Progress under http://qcweb/XSHOOT_R/monitor/FINISHED/histoMonitor.html.

	PHOENIX – the process to create science-grade data products	Doc: Issue: Date: Page:	VLT-MAN-ESO-19600-6104 1 2014-03-03 19 of 19
--	---	----------------------------------	---

7 Limitations of PHOENIX

The PHOENIX process, as it is currently implemented, is limited by the following conditions:

- The *existence and validity* of historical ABs: the current structure of ABs has been fairly stable over the years. The content (the associations) can be edited up to a certain extent but the creation of new ABs is currently not supported by PHOENIX. This means in particular a natural earliest date for *phoenix* processing corresponding to the oldest date with ABs stored on *qcweb*. To go beyond that historical mark requires new concepts (e.g. using *calSelector* rules for new associations) which currently are not implemented.
- The *existence* of historical master calibrations in NGAS: The master calibrations have not been stored for all instruments from the begin of operations on. This is also true for static master calibrations which have been provided by the pipeline developers. The creation of historical master calibrations with *phoenix* is not supported currently.
- The *applicability* of historical master calibrations: due to pipeline improvements the quality of early masters might be substantially lower than of current masters. It is then up to a detailed judgement if the processing of science data with the current pipeline quality but old master calibrations, with sub-standard quality, is an acceptable compromise. If not, the re-creation of historical master calibrations is currently not supported by *phoenix*.
- The *backwards compatibility* of the pipeline: while in theory all pipelines are required to be fully backwards compatible (based on the argument that for a given science case VLT data from 2000 are most likely as valuable as those from 2014), there are limitations to this principle, sometimes because of unavoidable changes, or because of desired improvements. A good example is the FORS2 spectroscopic pipeline that saw a substantial improvement of the wavelength calibration together with the extraction of the calibration information in the flat field data. With the current pipeline, the science data from 2014 can be processed using the current master calibrations without problems. But the same does not apply for science data from 2006: the master calibrations from that date had a completely different format and are not accepted anymore by the current pipeline. The only way of processing FORS2 science data from 2006 is with interactive tools like *reflex*, processing the entire cascade. In that environment the pipeline is backwards compatible, but not for *phoenix*.