



EUROPEAN SOUTHERN OBSERVATORY

Organisation Européenne pour des Recherches Astronomiques dans l'Hémisphère Austral
Europäische Organisation für astronomische Forschung in der südlichen Hemisphäre

ESO - EUROPEAN
SOUTHERN
OBSERVATORY

QC Handbook

DMO Data Processing and Quality Control Group

VLT-SPE-ESO-19600-3986

Issue: 5.2

Date: 2020-01-24

52 pages

Prepared (version 1): D. Silva, R. Hanuschik 2005-10-12
(this version): R. Hanuschik 2020-01-24

	Name	Date	Signature
Approved:	M. Romaniello	2006-06-09	

	Name	Date	Signature
Released:	P. Quinn	2006-06-09	

	Name	Date	Signature
--	------	------	-----------

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	2 of 52

CHANGE RECORD

Issue	Date	modified	By whom
0.7A	2003-06-01	ALL	Revised based on DFO comments
0.7C	2003-06-12	ALL	Input by R. Hanuschik
1.1	2005-10-11	updated to include dfos tools and update process description	Revised and updated by R. Hanuschik
2.0	2008-05-05	updated and revised (includes SCIENCE ingestion; OCA; scoring; header download and modification); most open points from previous edition resolved, some new added; reprocessing added	Revised and updated by R. Hanuschik
3.0	2009-11-30	updated and revised (includes fast data transfer, ancillary files, package ingestion, incremental processing, science-grade processing); La Silla packages removed	Revised and updated by R. Hanuschik
3.1	2010-12-20	update information about data packages and User Portal	Revised and updated by S. Moehler and R. Hanuschik
4.0	2012-07-01	removed: science processing, pipacks; new: XXLight	Revised and updated by R. Hanuschik
5.0	2015-03-20	Updated: IDPs, calSelector; muc servers. Removed: ancillary files; opslogs; AB harvesting	Revised and updated by R. Hanuschik
5.1	2017-01-24	Removed: delivery of master calibrations (PAR mode) Added: DEEP process, rawdisp2reference, WISQ IDP	Revised and updated by R. Hanuschik
5.2	2020-01-24	Updated to serve as reference for QC1 (CALIB) and renamed to "QC Handbook"	Revised and updated by R. Hanuschik

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	3 of 52

TABLE OF CONTENTS

1.	INTRODUCTION	8
1.1	Applicable Documents	8
1.2	Reference Documents	8
1.3	Glossary	9
1.4	Abbreviations & Acronyms	11
1.5	Acknowledgements	11
2.	HIGH-LEVEL OPERATIONS CONCEPTS	12
2.1	QC Goals & Objectives	12
2.2	Calibration Plan	12
2.3	QC Scientists.....	13
2.4	Data Concepts	13
2.4.1	Data Frames: the fundamental QC data unit	13
2.4.2	Data complexity and data volume.....	14
2.4.3	QC frame life-cycle.....	15
2.4.4	QC comments.....	16
2.5	QC operational models	16
2.5.1	Incremental processing	16
2.5.2	Daily workflow	16
3.	INTERFACES	18
3.1	Interfaces: Operational.....	18
3.1.1	Paranal Observatory.....	18
3.1.2	ESO Science Archive.....	19
3.1.3	User Support Department (USD)	19
3.1.4	Instrument Operations Team (IOT)	20
3.1.5	Science Data Products Group (SDP).....	20
3.1.6	Archive Science Group (ASG).....	20
3.1.7	End-users	20
3.1.8	ESO Management	20
3.1.9	QC Internal	21
3.2	Interfaces: Technical.....	21
3.2.1	Data Interface Control.....	21

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	4 of 52

3.2.2	Science Operations Software SCS	21
3.2.3	Data Products Board	21
3.2.4	Technical support groups	21
4.	OPERATIONAL WORKFLOWS	22
4.1	Daily Workflow	22
4.1.1	Header Processing and Report Creation.....	22
4.1.2	Data File Retrieval and Calibration Processing	22
4.1.3	Calibration Data: Quality Control, scoring and certification	23
4.1.4	Calibration Data: Final organization	24
4.1.5	Science ABs	24
4.1.6	Final Clean-up.....	24
4.1.7	Possible ERROR conditions in the daily workflow	25
4.1.8	The Association Block	25
4.1.9	Raw data lifetime	25
4.1.10	Product data lifetime	25
4.1.11	History record	26
4.2	PHOENIX workflow	26
4.3	Special workflows.....	26
4.3.1	Pre-imaging workflow	26
4.3.2	Health Check Monitor Workflow	26
5.	PROCESS INFRASTRUCTURE	27
5.1	Compute Infrastructure	27
5.1.1	QC processing infrastructure	27
5.1.2	Personal workstations	27
5.2	DFS Installation	27
5.3	User Account and Directory Layout	28
5.3.1	Process Management Directory Tree	28
5.3.2	Data Management Directory Tree.....	28
5.4	Backup scheme	29
5.5	Daily Workflow Process	29
5.5.1	General pre-processing	29
5.5.2	CALIB: create ABs.....	30
5.5.3	Processing CALIB ABs	30
5.5.4	CALIB: QC reports and updating HC plots.....	31

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	5 of 52

5.5.5	CALIB: Post-processing	32
5.5.6	SCIENCE: AB creation	33
5.5.7	Final clean-up.....	33
5.5.8	Summary of Commands.....	33
5.5.9	Status flags.....	34
5.6	PHOENIX workflow	34
5.6.1	Description of the workflow	34
5.6.2	The phoenix workflow tool.....	34
5.6.3	PHOENIX DEEP	35
5.6.4	IDP ingestion	36
5.6.5	PHOENIX projects	36
5.6.6	PHOENIX master calibrations	36
5.7	Special workflows.....	37
5.7.1	Pre-imaging workflow	37
5.7.2	Health Check Monitor.....	37
5.7.3	calChecker workflow	37
5.7.4	Workflow for management of <i>calSelector</i> rules	38
5.8	DFS pipelines and infrastructure.....	38
5.9	Infrastructure: monitors and GUIs	40
5.10	Infrastructure: workflow tools	40
5.11	Supporting information: nightlogs, problem reports, history	40
5.12	QC web server qcweb.....	41
5.13	Web server stargate1	41
6.	QUALITY CONTROL.....	41
6.1	Basic Principles.....	41
6.1.1	QC0.....	41
6.1.2	QC1.....	42
6.1.3	Scoring.....	42
6.2	QC Reports	43
6.2.1	General Overview	43
6.2.2	QC1 Database and Tools.....	43
6.2.3	QC1 plots	43
6.2.4	Health Check Monitor and tools	44
6.3	Certification	44

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	6 of 52

6.4	QC parameters not computed by pipelines	45
7.	QC WEB PAGES AND DOCUMENTATION	45
7.1	Basic principles	45
7.2	Structure	46
7.3	Maintenance and issues	46
8.	PROCESS MANAGEMENT & CONTROL	47
8.1	Goals	47
8.2	Current status	47
8.2.1	Process metrics	47
8.2.2	Workflow information system WISQ	47
8.2.3	Process management tools	48
8.2.4	Process development	48
8.2.5	Process Implementation	49
9.	FUTURE ISSUES, FUTURE DIRECTIONS	50
9.1	Data processing	50
9.1.1	Data rates	50
9.1.2	Processing strategies	51
9.1.3	Compute server strategies	51
9.2	Re-processing	52

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	7 of 52

LIST OF TABLES

Table 1: Information hierarchy for QC 14

Table 2: Structure of process management directory tree 28

Table 3: Structure of data management directory tree 29

Table 4: Steps of the daily workflow 34

Table 5: Current IDP projects 36

Table 6: DFS pipelines and infrastructure 39

Table 7: Statistics parameters for daily workflow 47

Table 8: QC metrics in the workflow information system 48

Table 9: Current QC compute server model 52

LIST OF FIGURES

Figure 1. Frame flow diagram 15

Figure 2. QC daily workflow 17

Figure 3: Information flow 18

Figure 4: Daily workflow 23

Figure 5. Score report 45

Figure 6: AB monitor with score flagging. 45

Figure 7: Evolution of VLT(I) data rates (compressed, relevant for data transfer). Red: science data; green: calibration data; black: sum of both..... 50

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	8 of 52

1. INTRODUCTION

This document describes the operational concepts and processes developed by the DMO Data Processing and Quality Control Group (QCG). The goal is to provide a global overview for anyone trying to understand the QC environment (e.g. newly appointed QC scientists).

This document mainly describes instrument-independent workflows which are served by a common toolset. Instrument-specific issues are encapsulated in configuration files, pipeline recipes, and database structures.

The term 'DFO' has changed meaning several times during the history of this document. When it was originally written in 2003, DFO was the name of the group which is now called the Data Processing and Quality Control Group, or QC Group for short (as already between 1998 and 2002). Until 2008 DFO was the name for the department which hosted, among other groups, the QC group. Between 2008 and 2011, this department was called DPD (Data Products Department). Currently it is called BOD (Back-end operations department). With ever-changing names of organizational structure, we decided that this document should keep its name as it is well established. It describes the processes of the QC group.

1.1 Applicable Documents

The following documents form part of this document:

Reference	Document Number	Issue	Date	Title
[AD01]	DFO internal memo	1.5	2005-10-07	Guidelines for QC Scientists
[AD03]	DFO internal memo	1.2	2003-02-03	DFO Data Association: I. General rules & concepts
[AD04]	DFO internal memo	1.1	2003-02-12	DFO Data Association: II. The Association Block
[AD05]	DFO internal memo	1.1	2003-02-12	DFO Data Association: III. General association tools
[AD06]	DFO internal memo	1.1	2007-05-20	QC guide for Paranal missions
[AD07]	DFO internal memo	1.0	2006-10-11	Codes and Practices

1.2 Reference Documents

The following documents are referenced in this document:

Reference	Document Number	Issue	Date	Title
[RD03]			2003-06	VLT Quality Control & Trending Services (Hanuschik & Silva, 2002, Messenger 108, p. 4)
[RD04]	GEN-SPE-ESO-19940-794	4.0	2008-04	Data Interface Control Document
[RD05]			2008-05	Scoring: a novel approach towards automated and reliable certification of pipeline products (Hanuschik et al. 2008, SPIE)
[RD06]	VLT-PRO-ESO-10300-4953	1.0	2011-12-15	Charter for IOTs (Smette, Siebenmorgen, Kaeufl)
[RD07]	VLT-MAN-ESO-19600-6104	1.0	2014-03-03	PHOENIX – the process to create science-grade data products (Hanuschik)
[RD08]	VLT-235058	2.0	2017-01-14	Generation and SAF Publication of Internal Data Products: Deployment Plan (Romaniello, Arnaboldi, Freudling, Hanuschik)

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	9 of 52

1.3 Glossary

Archive - see Science Archive Facility.

ASG – Archive science group: the group within BOD in charge of handling external and internal data products

Association Block – text file that contains all the information needed by QC to process one or more raw files. Contents include: names of raw files, names of master calibration products needed to process those files, pipeline recipe used to process the raw files, etc. For further details, see [AD04].

Autrep (automatic reporting system) - Paranal database hosting the complete VLT technical data stream, and also certain QC parameters. This system reads ‘QC1 ops logs’ (operations logs). The pipelines as run on-site write QC parameters into these logs, duplicated from the FITS header of pipeline product files.

Breakpoints – a date when instrument characteristics changed sharply, essentially breaking the then-current calibration validity chain.

Calibration Plan – instrument-specific description of what calibration data have to be acquired, why they are acquired, and how often they are acquired. Calibration Plans are maintained by Paranal Science Operations in co-ordination with QC.

Calibration Products – FITS files produced by the pipelines when raw calibration data are processed (also called master calibrations).

Calibration Database, Local – local database containing calibration products used by the pipelines and the Data Organizer on Paranal.

CalSelector – tool to create associations for the archive user between science and calibration files in a dynamic way, using OCA rules

CONDOR - public software used by the DFS for batch queueing and parallel or grid computing

Data Frame (frame) – a FITS file produced by executing a VLT(I) observation (raw) or by processing raw data into a product

Data Warehouse – database hosting all header information (also called: SAFIQ)

Data Products – FITS files produced when raw science or calibration data are processed by a pipeline recipe.

DFS – data flow system: software system to manage the whole data flow within the observatory

DTS - data transfer system: the software system to transfer all Paranal data over the fibre link and the internet to Garching

EDP – external data products, created by (mostly survey) PIs or data centres, and ingested by ASG into the archive

Extensions – FITS files often have several extensions; these are logically distinct sections with a header and a data part; typical cases for extensions are multi-detector files with one extension per detector.

FITS – flexible image transfer standard: the standard for file formats in astronomy and in particular at ESO; FITS files always have a metadata section (the header) describing the data, and at least one data section. They can have one or several extensions.

Headers (FITS Headers) – the standard FITS headers contained in all FITS files. FITS headers describe the data contained in the file (they contain 'metadata'). Apart from the main headers, there can be extension headers.

Health Checks – the set of calibrations taken regularly and independently of science data. They are intended to monitor critical instrument properties.

IDP – internal data products: science data processed by the QC group, using certified pipelines and master calibrations

JIRA – problem reporting system used by the pipeline group

Master Calibrations – QC certified calibration products

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	10 of 52

Observation Block (OB) – the fundamental scheduling unit of VLT science operations. An OB is a single pointing and contains the parameters necessary to configure the VLT system (telescope & instrument), and to execute a sequence of observations. OBs are usually executed only once. OBs can have one or several templates. Several OBs can be grouped together by containers.

Phase 1 – the observing proposal management process. Activities include: Call for Proposals revision and release; proposal submission & tracking; proposal scientific & technical review; time allocation; semester schedule creation & management. Phase 1 is managed by the Observing Programme Department OPO with assistance from the User Support Department.

Phase 2 – the Service Mode observing preparation process, managed by the User Support Department. During Phase 2, users create and submit Observation Blocks, as well as other supporting information.

Phase 3 – the process in which PIs provide their data products to ASG for archive ingestion; also: the process used by QCG to provide IDPs to ASG; the ingestion and the archive query interfaces together form the set of phase3 tools.

Pipeline infrastructure – the collection of DFS tools that are used to create Reduction/Association Blocks. The main components are tools for data association, reduction job scheduling, and processing. On Paranal, these tools are: the Data Organizer (DO), the Reduction Block Scheduler (RBS), and the Data Reduction System (DRS). For QC, these tools are: the ABbuilder, CONDOR (a batch queue system), esorex (as interface to the pipeline recipes), and the DRS (same as used at the VLT).

Pipeline – the collection of data processing and reduction recipes. There is one pipeline per instrument. The recipes have standard interfaces for input and output.

Paranal Science Operations – the department responsible for executing Service Mode observations and supporting Visiting Astronomers during Visitor Mode observations. PSO consists of staff astronomers, fellows, telescope-instrument operators (TIOs), Operations Specialists (OSs) and data handling administrators.

Quality Control – the process of assuring that VLT(I) data have a consistent and known quality.

QC parameters (often also QC1 parameters) – parameters computed by the pipeline used to assess data quality. We prefer the term ‘QC parameters’ since the QC process in which they are created does not really matter (QC0, QC1 or QC2/IDP).

QC1 database – the database where QC parameters are stored

Reflex – workflow software used by SDP to create data processing workflows

SDP – Science data products group, within the Back-end Operations Department, responsible for science grade pipelines and reflex workflows.

Science Archive Facility (SAF) – also called the Archive. This is a collection of databases that contain all the raw data produced by the VLT(I), all certified calibration products, and all QC parameters. QCG has privileged access to the full archive content via direct-access tools.

Scoring – grading of QC parameters and products. Scoring is the key concept for automatic certification of products and for monitoring instrument health. QC parameters of a product file are scored against configured thresholds. These can be static (e.g. from specifications) or dynamic (from statistics).

Service Mode (SM) – an operations mode where users with allocated observing time do not travel to the telescope, but submit Observation Blocks for ESO to execute on their behalf. Service Mode observations are flexibly scheduled, i.e. OBs are executed based on a variety of scheduling constraints including scientific priority, instrument configuration, and requested observing conditions.

Trending – the QC process of tracking the values of QC parameters as a function of time.

User Support Department (USD) – part of DMO, this Garching-based operations group provides front-line user support for Phase 1 and Phase 2 activity, as well as OB scheduling support during Service Mode operations.

User Portal – the main entry for external users to information about their proprietary data and to ESO services related to their run (e.g. access to OB execution status; download of raw and product frames etc.)

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	11 of 52

Visitor Mode (VM) – an operations mode where users with allocated observing time travel to the telescope and execute their Observation Blocks with direct on-site support of Paranal Science Operations.

1.4 Abbreviations & Acronyms

AB	Association Block
AOG	Archive Operations Group
BOD	Back-end operations department
DBCM	Database Content Management
DFO	Data Flow Operations
DMO	Data Management and Operations Division
DFS	Data Flow System
DPQC	Data Processing and Quality Control (Group)
EDP	External Data products
IDP	Internal Data Products
NGAS	next-generation archive system
OB	Observation Block
OCA	Organization, classification and association: ruleset for association of data
PPRS	Paranal Problem Reporting System
PSO	Paranal Science Operations
QC(G)	Quality Control, or Quality Control Group
SAF	Science Archive Facility
SCS	Science Operation Software Department
SOF	set of frames: the file read by the pipeline processing infrastructure <code>esorex</code> to execute a processing job
SM	Service Mode
ToO	Target of Opportunity: special observing mode for science data
USD	User Support Department
VM	Visitor Mode

1.5 Acknowledgements

The operations concepts and processes described in this document were developed and refined between 1998 and 2003 by the entire DFO/QC Scientists team. Interface definitions and support have been developed in collaboration with the ESO Archive and DFS development groups as well as Paranal Science Operations. The original Paranal specific framework was developed by Bruno Leibundgut, Paola Amico, and Reinhard Hanuschik in 1998 and 1999. Ferdinando Patat and Wolfgang Hummel made significant contributions to the framework described here. David Silva has written the first released version ('0.7C'), with major contributions from Reinhard Hanuschik. All later editions have been edited and updated by R.H.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	12 of 52

2. HIGH-LEVEL OPERATIONS CONCEPTS

2.1 QC Goals & Objectives

The QC group has the following high-level operational goals & objectives:

- Monitor short-term and long-term performance of VLT instruments via QC parameters to assure that ESO delivers astronomical data of known and controlled quality.
- Establish and maintain a closed QC loop to reliably ensure the proper calibrations of the recent science night, and the instrument health for the coming night.
- Follow the execution of instrument calibration plans; work with Paranal Science Operations to revise calibration plan as necessary.
- Certify and ingest calibration products into the archive, for later use by archive users.
- Create and maintain association rules for the processing of science data, and ingest them into the archive.
- For selected instrument (modes), process science data to science-grade data products; control their quality, and ingest them into the archive.

In the QC context, **pipeline supported** means a pipeline recipe exists to process a set of raw input data into a set of defined data products.

QC has the following operational process goals:

- work instrument independent whenever possible;
- work configuration driven – work with rules and properties;
- set up operations as stable as possible;
- store/read all relevant process information into/from a database or archive;
- implement and execute key workflows as automated as possible;
- document all workflows, processes and procedures;
- process calibration data as soon as possible;
- process data using certified and best-available calibrations;
- store all products in the archive;
- follow all QC Standards & Practices (see [AD01]).

2.2 Calibration Plan

Each instrument has a **Calibration Plan** that describes the types of calibration data to be acquired, their times and frequencies. Some calibration data are used to monitor specific aspects of instrument or system performance (e.g. readout-noise, bias level, response) while other calibration data are used for specific scientific calibration operations (e.g. flat-fielding, wavelength calibration). Whenever possible, raw calibration data are processed into calibration products (**master calibration frames**). The quality of calibration data is assessed and monitored using **QC parameters** computed by the pipeline or by procedures created and maintained by QCG.

QC plays an active role in defining and monitoring the instrument-specific Calibration Plans. In particular, it is important to tune calibration plans to make sure that calibration data are always available close enough in conditions to the related science data. Calibration frequency should ideally be driven by instrument stability and science processing needs.

QC also plays an active role in monitoring Calibration Plan execution. If data are taken too sparsely, QC works with Paranal to rectify the situation.

It is occasionally necessary to re-execute a science OB because QC has discovered that the proper calibration data were not taken. QC then interacts with USD to flag the affected OB correspondingly and trigger follow-up actions with SciOps.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	13 of 52

2.3 QC Scientists

The QC group consists of **QC Scientists** who are responsible for achieving the QC Goals & Objectives listed above. Each QC Scientist is responsible for several instruments. The specific responsibilities of the QC Scientists are given in *Guidelines for QC Scientists* [AD01; AD06, AD07].

In order to maintain their expertise and establish good working relationship to their Paranal colleagues, QC scientists are scheduled at least once per year to visit the VLT/VLTI.

2.4 Data Concepts

2.4.1 Data Frames: the fundamental QC data unit

The Data Flow System (DFS) has the following information hierarchy:

Observing programme: the overall description of a science project being executed by ESO. Programmes are relevant for the PI application process but not relevant for QC work. Often the term is used (incorrectly) as a synonym for a run.

Observing run: a specific set of scheduled observations for a specific instrument, with a user specified set of observing constraints. There are one or more runs for each programme. It is described by the FITS key OBS.PROG.ID¹. Mostly science (and acquisition) data carry a specific run ID. For formal reasons, daytime calibration data, as well as nighttime calibration data driven by the Calibration Plan (like standard stars), also carry a run ID, the 'technical run ID' which is fixed and unique per instrument.

Observation block (OB): a specific observation parameter set. Science OBs consist of one and only one telescope preset and a set of instrument configurations (usually limited to one instrument mode). Daytime calibration OBs have no telescope preset. There are one or more OBs for each run. All data processed by QC are produced via the execution of OBs. While science OBs carry a unique ID, standard calibration data usually carry formal IDs. An OB is identified by the hierarchical FITS key OBS.ID. All data belonging to the same OB execution carry the same OBS.START key.

Containers: An extended concept has been introduced a couple of years ago, originally for the survey instruments: containers linking OBs. This is a set of conditions which group OB execution conditions together, or mark OBs as belonging logically together. Various types of containers exist.

Observation templates: OB sub-units used to set parameters of observing sequences within an OB. Each OB contains one or more templates. Templates are often associated to a specific pipeline recipe. The hierarchical keyword TPL.START marks all raw frames belonging to the same template execution. A template is the fundamental unit for data processing.

Raw data frames: observation template output. Each template produces one or more raw data frames. Each frame is written in the FITS format. It is uniquely identified by its ARCFILE name². Confusingly, the term 'data product' is used at various places (e.g. in databases) for raw data. In the QC world, that term is reserved for pipeline-processed FITS files. They are uniquely identified by their PIPEFILE name.

Association block (AB): fundamental entity for QC processing. All information about a processing job is collected in an AB. In particular, it contains all affected raw files. Typically the set of raw files in an AB consists either of a single frame, or of all frames from a template.

The following table gives a few examples for this information hierarchy.

¹ This keyword name is obviously badly chosen but it's too late for a change.

² The ORIGFILE name is also present in each file header and is used by the Observatory e.g. in their night logs. Note that ORIGFILE names are **not** unique and will repeat after a year.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	14 of 52

[Programme:]	0104.B-0046		
Runs:	0104.B-0046A	0104.B-0046B	0104.B-0046C
Containers:	179240 (linking 2 or more OBs)
OBs:	179246 179248 179250
Template:	All files sharing the same <TPL.START> value		
ABs:
Raw frames:	179246:	ESPRE.2020-01-10T23:48:31.234.fits ESPRE.2020-01-10T23:49:14.071.fits etc.	...
	179248:	ESPRE.2020-01-13T02:48:91.004.fits etc.	...

Table 1: Information hierarchy for QC

Observing runs are executed in two fundamental operational modes. **Visitor Mode** implies that the observing run principal investigator (or a representative) is present at the telescope, making real-time decisions about the observations to execute. **Service Mode** implies that ESO staff execute Observation Blocks submitted to ESO by the PI. Other modes have been added (e.g. delegated VM mode).

A vaguely defined **Technical Mode** also exists. These are observing runs used to collect technical data. These runs can be associated with the official Calibration Plan, technical tests, science verification, or instrument commissioning periods. In general, QC only processes data related to the official Calibration Plan. However, commissioning data is often used to test QC systems for newly installed instruments, especially the pipeline recipes.

QCG primarily processes calibration data – individually or grouped by observation template. All pipeline-supported calibration data frames are processed. Science data are currently processed for selected instrument modes only, and de-coupled from the calibration streams.

OBs are usually only important for QC when reporting a data quality problem to USD or PSO.

Another aspect of hierarchy is the data acquisition pattern: the fundamental unit for QC data processing is the **day** which is defined as a 24 hour period which starts at 21:00 UT which guarantees that a new Paranal night always starts with a new date. That convention defines the so-called *DFO date*. With the important exception of incremental processing, QCG processes full sets of data per day, and this is what is called the *daily workflow*.

2.4.2 Data complexity and data volume

The QC workload is driven by the combination of data volume (number of pixels per unit time) and data complexity (number of required processing steps per pixel). Data volume can be further broken down into a combination of time and individual data frame size.

An operational rule-of-thumb is that data processing should take no longer than data acquisition. For QC this has become over the years an easy goal: with multi-core processors the compute power can be scaled and adapted to the needs. With the science processing of certain instruments with large volume and complexity (like MUSE), the data volume and data complexity challenge has shown up again. It is mastered by powerful multi-processor machines that can be scaled to the processing needs.

For quality control and certification, it is neither possible nor desirable to examine every single data product individually. QC has implemented a score-based automatic certification scheme in which only outliers are given significant attention.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	15 of 52

2.4.3 QC frame life-cycle

From the QC perspective, all data frames go through a standard process flow, illustrated schematically in Figure 1. The first step is to classify all frames, and associate them with other files as defined by the association rules. Incomplete associations are put aside and tried again later.

Next, the raw calibration frames are processed with the pipeline, QC parameters are extracted and QC reports created.

Then all calibration products undergo a quality control review, based on QC parameters, trending, scoring and inspection. This process is mostly automatic, based on automatic certification for green scores and investigation for red scores. Products that pass this review are called *certified*. They are stored in the archive. Processed frames that fail the QC are rejected and deleted.

Science data are currently not handled in the daily workflow. Selected science data are processed later off-line as IDP stream, in the PHOENIX workflow.

This QC process implies that:

- data in pipeline-unsupported modes are not QC checked;
- QC is always checking quality of data (not e.g. quality of night);
- problems in raw data which do not propagate into products are not detected.

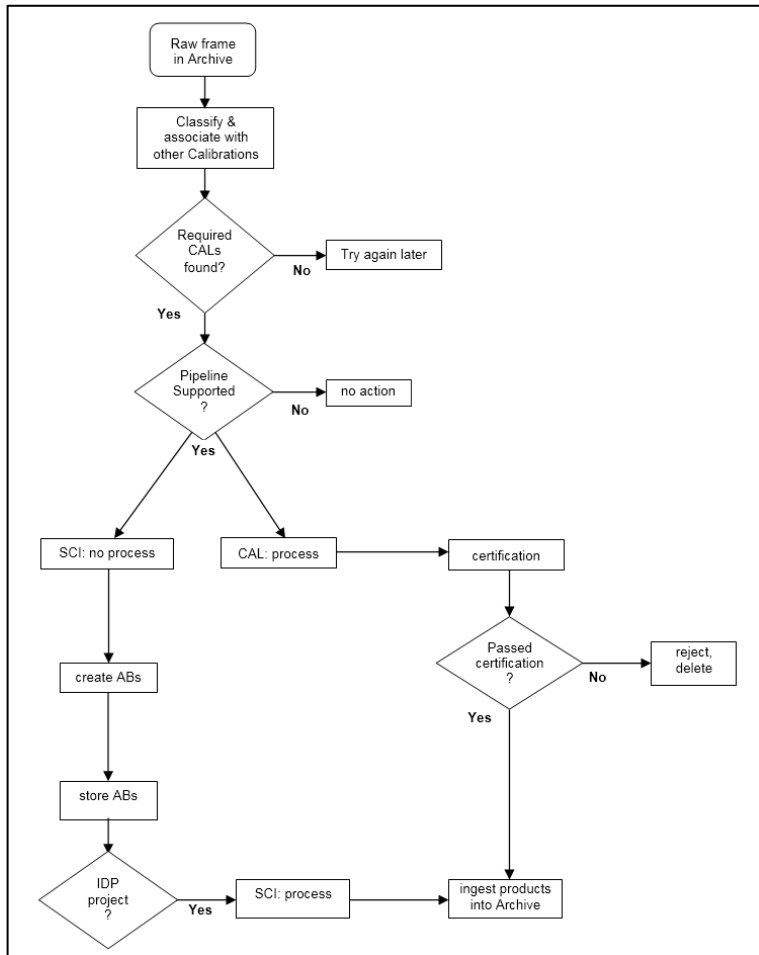


Figure 1. Frame flow diagram

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	16 of 52

2.4.4 QC comments

QC comments about processing success or failures are collected. Such comments provide non-obvious information about the frame, e.g. “Raw file has defective header” or “Frame could not be processed”. Comments about raw files or product files are stored in QC database tables (`raw_comments`, `prod_comments`).

2.5 QC operational models

2.5.1 Incremental processing

The fastest processing scheme for calibrations is *on-the-fly* processing, meaning they are processed as soon as a complete data set (either a single file or a set of files acquired by a single template) is available. This is the scheme for the Paranal automatic pipelines.

For QC the *on-the-fly* scheme is not optimal, for several reasons. First, there is no end-of-template signal in the QC environment, hence incomplete data sets would be processed. Second, *on-the-fly* processing requires a dedicated machine where no other processes execute. This is also not possible in the QC environment, where other automatic or interactive processes need to be executed.

Hence QC uses another processing scheme as a compromise, the *incremental* processing. A cronjob is called once per hour (`autoDaily`), searches for all new calibrations acquired since the last execution, creates and executes ABs for them, and then creates QC reports and scores.

With that approach QC can provide feedback about data quality to Paranal within one hour, at most, of data acquisition. This is fast enough for any practical purpose.

2.5.2 Daily workflow

Originally QC was committed to process all data, calibrations and science. With a rescoping in 2011, QC has developed a new operational model. It was then called *XXLight* because in comparison with the previous model it skips science processing and data packing. That model has become the standard model and is called the *daily workflow*. It has been complemented in 2013 by science processing to the science-grade level for selected instruments. In the future QC-v2 model, there will be a new version of the QC1+ (SCIENCE) workflow which will be a daily workflow following the successful execution of QC1+ (CALIB).

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	17 of 52

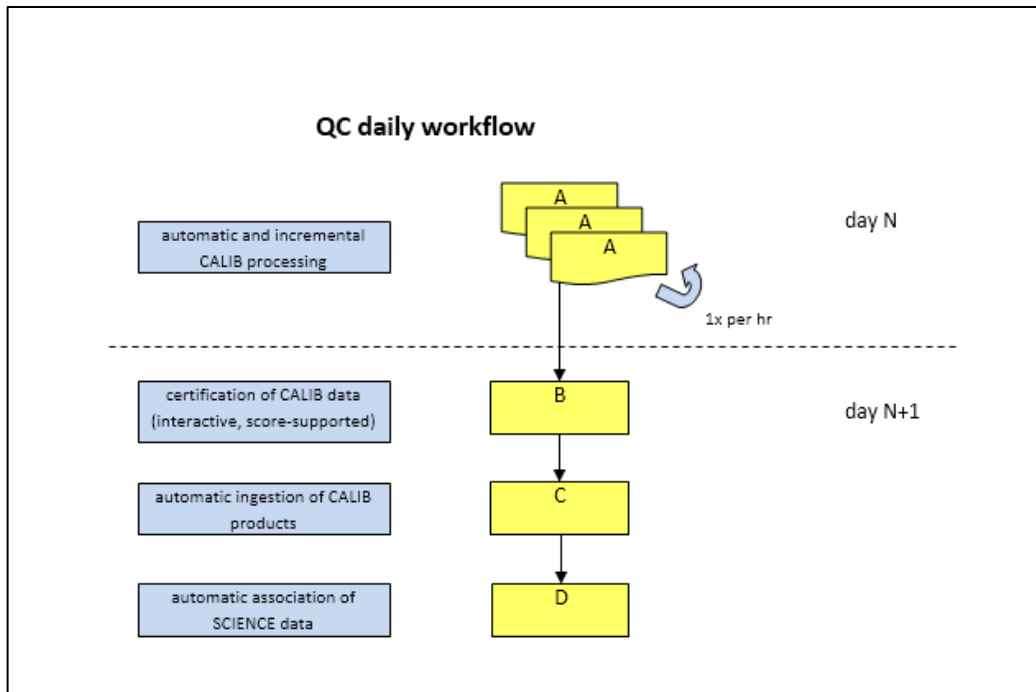


Figure 2. QC daily workflow

The daily workflow (Figure 2) is characterized by:

- Automatic, incremental processing of incoming calibration data;
- automatic certification of green scores, interactive investigation of red scores;
- handling of failures: provide analysis comments, no re-processing (one-attempt only);
- ingestion of master calibrations into the archive from where they can be accessed by external users through the *calSelector* service.

In addition, QC creates and stores (on the QC web server) associations for science data, with the purpose of

- verifying that the current OCA rules are still applicable;
- have reference ABs for follow-up IDP projects.

The OCA rules are maintained by QC as a service for *calSelector* and thereby for the community.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	18 of 52

3. INTERFACES

3.1 Interfaces: Operational

The QC operational mission can be summarized as information management, processing, and delivery. Therefore, the QC operational universe consists of information providers (“sources”) and information customers (“sinks”). The main information flow paths and interfaces are summarized in Figure 3.

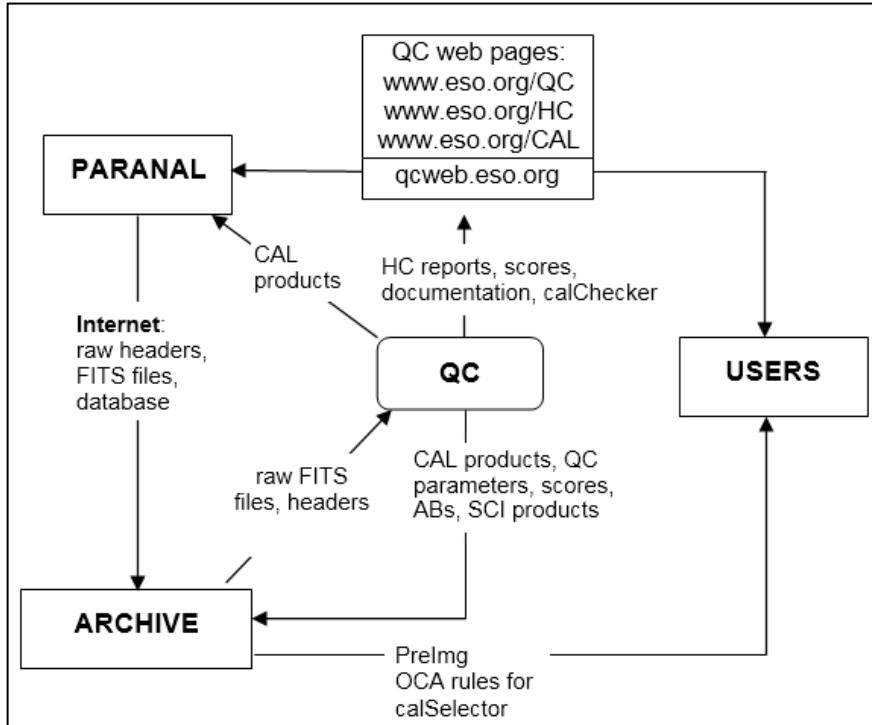


Figure 3: Information flow

3.1.1 Paranal Observatory

The Paranal **Science and calibration data** are the fundamental data produced by each operational instrument, created via the execution of **Observation Blocks (OBs)** by the **Paranal Science Operations (PSO)** staff. Each OB produces one or more **data frames**, formatted as FITS files. These FITS files are delivered over the Data Transfer System to Garching and stored in the ESO Science Archive. QC does not access the data on Paranal– it only accesses the data once they have been archived in Garching. More details about QC data access are provided in the Archive section below, including information about the special case of pre-imaging.

- **File headers:** the FITS header for each data file is extracted automatically by the on-site DFS, inserted into a database, and replicated over the Internet to the ESO Science Archive in Garching where they are immediately available for QC retrieval. Headers are dynamically updated when they are downloaded from the Archive.
- **Nightlogs:** the science operations activities at each telescope are logged using the Paranal Nightlog tool (NLT). Information logged includes personnel on duty, global observing conditions, time spend on each activity, and instruments used. Information about the execution of each OB is captured, including observing conditions during the execution period.

The Paranal Observatory complex (VLT, VLTI, VST, and VISTA) is the ultimate source of all raw (unprocessed) information. This information includes:

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	19 of 52

- **Problem reports:** the Paranal Problem reporting system PPRS is used to record and transmit operational and technical problem reports. QC scientists are notified via e-mail when a problem report relevant to their assigned instrument(s) is created.

PSO is also a customer of QC generated information, including:

- **Calibration completeness Monitor:** The calChecker (www.eso.org/CAL) monitors the execution of the calibration plan. It flags missing calibrations in order to trigger their acquisition in due time.
- **Health Check Monitor:** as the most important service to Paranal, QC maintains the Health Check monitor Web site (<http://www.eso.org/HC>) that gathers information useful for assessing current instrument performance. Its most important feature is a comparison of the latest QC parameters with previous trending. These QC parameters are scored (rated) and give feedback about compliance or alerts. Only a selected (IOT specified) set of parameters is displayed here. The Health Check Monitor also provides links to reference frames (products or raw files) and the QC1 database. See the *Quality Control* section for more details.
- **QC Web pages:** the QC Web site (<http://www.eso.org/QC>) exports a wide range of information, including documentation about data, QC products, and examples of pipeline products.
- **QC1 database:** all QC-generated QC parameters are available in a database maintained in the Archive. Various browsing and plotting tools are available from the QC Web pages for each instrument. See the *Quality Control* section for more details.
- **Reference frames:** QC maintains a web site with screenshots of all or a the most important data types per instrument mode (http://www.eso.org/observing/dfq/quality/ALL/ref_frames/ref_frames.html), to be used as reference for QC0 checks on the mountain.
- **Raw file displays:** Within the daily workflow, QC generates screenshots for all raw calibration files and displays the current last week worth of screenshots on the QC web server, for inspection by the Paranal daytime shift. These raw file displays also feed the reference frames web site.

3.1.2 ESO Science Archive

For QC, the ESO Science Archive is the source for all raw data frames. The Archive receives these data from Paranal over the DTS link. The DTS physically consists of: an ESO-owned fibre link Paranal-Antofagasta, the Chilean backbone and the transatlantic link.

In Garching the data are ingested into the Archive. They are then available to QC.

After information processing is completed, QC delivers the following items to the Archive:

- **Calibration products:** also called “master calibration files”, these are calibration products certified by QC to be technically correct and usable for science data processing. These frames are ingested into the Archive calibration database using DFS supported tools.
- **QC parameters:** all QC parameters are ingested into the QC1 database.

Also, **static calibration products** (the ones not derived from raw files, like line catalogues) are ingested by the QC group into the Archive, as delivered by the pipeline developers.

The following items are stored on the QC web server qcweb, for internal reference:

- **Logs and plots;**
- **raw file displays (dynamically, one week deep);**
- **scores, comments;**
- **ABs.**

3.1.3 User Support Department (USD)

The User Support Department (USD) provides QC with:

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	20 of 52

- **Operational problem reports:** using the PROP ticket system, questions or problems are collected from end-users, Paranal Science Operations, and USD. These questions and problems are forwarded to QC for response and/or action as appropriate.

QC provides USD with:

- **Data problem reports:** on rare occasions, QC determines that one or more observations were incorrectly executed. Such failures could be caused by operational problems (e.g. insufficient calibration data available) or technical problems (e.g. instrument failure occurred that was not detected by Paranal). In these cases, QC notifies USD which OBs produced the faulty data, describes the problem, and makes recommendations (if necessary) about how the OBs should be re-executed. It is then up to USD to negotiate the re-execution schedule with Paranal.

3.1.4 Instrument Operations Team (IOT)

Each instrument has an Instrument Operations Team (IOT). Each team is composed of representatives from the USD, PSO, and QC operations teams, as well as technical support engineers from Instrumentation, DFS, and OTS as needed. These teams meet several times per year to discuss all operational problems and potential changes for their instrument. The responsibilities of the IOTs are described in [RD06].

In the IOT context, the role of QC is to provide feedback about the effects on QC of operational problems or proposed changes. QC can also initiate problem reports or change requests. The QC Scientist attends the IOT meetings.

3.1.5 Science Data Products Group (SDP)

The Science Data Products Group has two main tasks: certify existing pipelines as science-ready, and develop Reflex workflows for interactive usage of these pipelines. The certification process includes checks of the algorithms and the processing strategy.

3.1.6 Archive Science Group (ASG)

QC interacts with ASG in the IDP projects. Once an IDP project is launched, the ASG scientist reviews the requirements for that project, coordinates change requests with PSD and creates an interface document.

3.1.7 End-users

ESO end-users are the scientists in the international community who are allocated time (both Service Mode and Visitor Mode), and all scientists downloading data from the archive. There is no direct interface of QC with the end users. QC creates and maintains the OCA rules for calSelector, ingests the certified master calibrations which can be downloaded by users from the archive, and creates IDPs which are offered on the archive interfaces.

Further products for the end users are:

- **Processed pre-imaging data:** using an automatic process, QC receives raw pre-imaging data from the Archive, processes it, and returns the data products to the Archive.
- **QC Web pages:** (<http://www.eso.org/qc/>) described in the QC/PSO interface section above, these pages with all their various sub-sections (Health Check, QC1 database interfaces, etc.) are available to the end-users.

3.1.8 ESO Management

Like all ESO groups, QCG submits **activity reports** several times a year. Each QC Scientist provides a summary of operational activity for the instruments for which they are responsible. Summary format has been specified by the group leader. These summaries are given to the group leader, who then revises and merges them into the final report which is submitted to the Department Head.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	21 of 52

3.1.9 QC Internal

QC maintains a loosely organized collection of internal operational Web pages. Every group member is encouraged to provide content. All internal documentation should be posted to this page, preferably in HTML or PDF formats. The top-level URL is <http://www.eso.org/~qc/groupIndex.html>.

3.2 Interfaces: Technical

3.2.1 Data Interface Control

Data structures produced by ESO telescopes and instruments are under configuration control. The overall syntax and semantic conventions are defined in the *Data Interface Control Document* [RD04]. Requests for configuration changes and extensions are reviewed by the Data Interface Control Board (DICB). This is a cross-ESO committee chaired by the Data Interface Control Scientist (part of the Database Content Management group, DBCM).

For QC, the most relevant data structures are FITS headers. The exact syntax of these headers is defined in data interface control dictionaries (DICD). These dictionaries are found on the DICB Web page: <http://archive.eso.org/DICB/>.

3.2.2 Science Operations Software SCS

The QC group has several technical interfaces with the Science Operation Software Department (SCS):

- **DFS installation:** the Software Engineering and Quality Group (SEQ) installs and configures the DFS on the QC operational machines.
- **Pipeline development:** the Pipeline Systems Group (PS) closely interacts with their equivalent QC Scientists to discuss algorithm development, software problem report resolution, QC parameter use, etc. The QC scientist works with the recipe development team to improve the usefulness of the pipeline products for quality control and science. This task is also supported by the Science Data Products Group.
- **DFS infrastructure development:** the Dataflow Infrastructure Group (DFI) interact with QC scientists during the development or revision of tools.

QC software problem reports (SPRs) and software change requests are reported via the JIRA ticketing system.

3.2.3 Data Products Board

The data products board reviews all proposals for improving the pipelines towards science readiness, and prioritizes submitted pipeline tickets. The Board is supposed to publish the results of its reviews.

3.2.4 Technical support groups

These groups and departments provide technical support services, as discussed in various other places in this document:

- **AOG = Archive Operations Group:** as the Archive “manager”, AOG is the source for all the raw frames (and the headers) processed by QC.
- **SOS = Systems Operations Support team:** they procure and install all QC operational hardware. SOS provides OS installation and administration, including regular backups of important directory spaces.
- **DBCM = Database Content Management:** this group provides the management of database content. They are responsible for database integrity, completeness and updates, as well as for tools to maintain header. In cases when QC discovers inconsistencies of headers, these are reported to the DBCM team who fixes them in the database. Also, hide requests (for useless files) are sent to DBCM.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	22 of 52

- **OTS** = Operations Technical Support: this group maintains the software and hardware to run the ESO databases, among them the main database operations server ESOECF which contains a variety of operational important databases accessed by QC.

4. OPERATIONAL WORKFLOWS

In the previous section, the QC information flow was described. In this section, the processes (i.e. **workflows**) developed to manage this information flow are described. These workflows have proven to be robust and reasonably instrument-independent. QC has developed tools to execute these workflows. They are described in the next section.

4.1 Daily Workflow

The idealized instrument-independent QC daily workflow is illustrated in Figure 4. Each step is explained in more detail below. Tools for each step are described in the next chapter.

The initial part of the daily workflow is automatic and picks incrementally any recent new calibrations (see also the sketch of the operational model in Figure 2). These are processed once per hour to make the QC parameters and scores available to the reporting tools and to provide up-to-date feedback to the Paranal daytime astronomers. Any later steps (“after 24 hours”) are interactive.

4.1.1 Header Processing and Report Creation

Daily workflow starts by retrieving the corresponding FITS headers from the Archive. These headers contain the so-called metadata (description of the data). They are used at many places in the QC workflows. Headers are ingested on the mountain and replicated to the Garching database server almost in real-time. Headers are updated dynamically when they are downloaded. Header files are stored in the data warehouse (SAFIQ) of the Archive for the complete history of the VLT.

If new headers are available, data reports are created. Each report contains one line per input data frame. Each line contains the raw data frame name (the ARCFILE name) and the values of an instrument-dependent set of keywords. The reports are available in HTML format.

The header reports have several purposes. They provide a quick overview of the night (the “data situation”). Then they provide a check for basic header integrity. The QC Scientist can quickly identify problems in quasi-real-time and report them to Paranal for resolution. Header reports are exported and are visible to Paranal staff so they can share the data overview.

4.1.2 Data File Retrieval and Calibration Processing

Incremental processing. An automatic workflow tool (*autoDaily*) scans once per hour the archive content for new *calibration* files. If found, the tool creates ABs and processes them. Any unsuccessful processing (because the AB is not yet complete, or required calibrations are not yet available) is attempted again in the next hour. This is called incremental processing.

Calling the workflow tool more often than 1x per hour is not reasonable since data processing usually needs some “data context”: raw files need to be combined, and calibrations may have dependencies on others.

The first step is to build **Association Blocks (ABs)**. ABs contain a list of (one or more) input raw files to be processed in one batch, the pipeline recipe to use, associated calibration products needed for processing, and the assigned (“predicted” or “virtual”) name of the calibration product. See [AD03], [AD04], and [AD05] for complete AB details.

In the process of AB creation, there is the important option to filter the input data stream. Based on configuration, this allows suppressing files from further processing in the QC system. Filtering is useful to prevent e.g. files with corrupt headers from entering the system.

All operations described so far (AB creation, reporting) are based on headers. With the execution of each AB, the first step is to download the raw files (and calibration product files if required). This dynamic download pattern minimizes data traffic and disk space requirements, implementing the important principle of “pixels on demand”.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	23 of 52

No further processing is done for calibration data without pipeline support. Pipeline supported calibration data are processed. ABs are extracted into the Sets of Frames (SOFs) required to call the DFS tool *esorex* being the interface to the pipelines. QC processes these files in parallel by using a batch queue system.

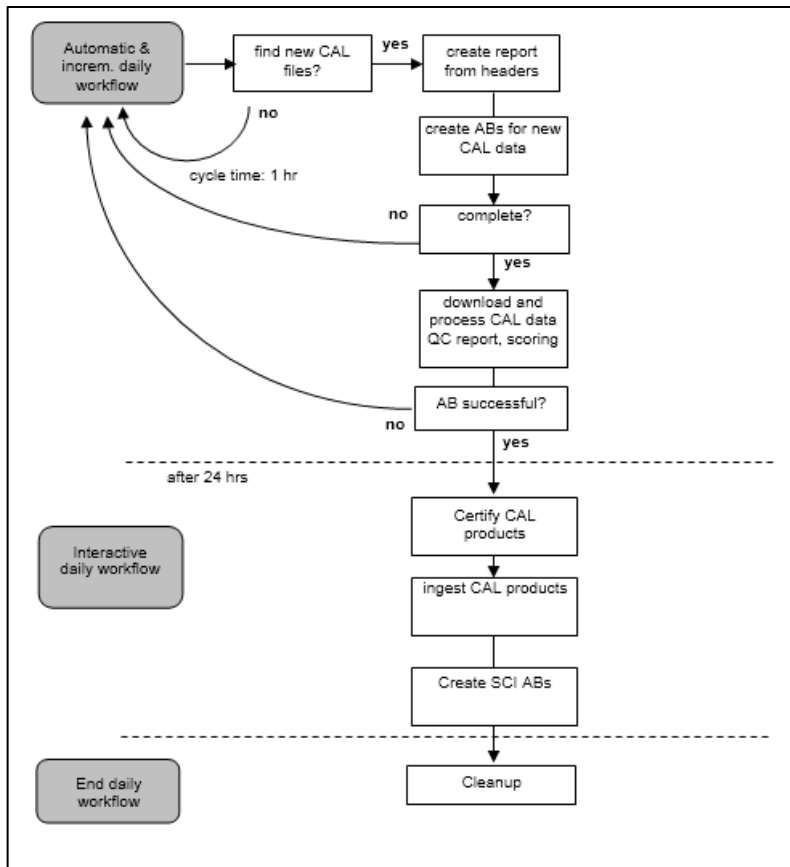


Figure 4: Daily workflow

AB execution order is governed by the **calibration cascade** [AD03], which is encoded in the AB configuration [AD04] and [AD05]. The calibration cascade specifies the types of calibration files needed and the sequence in which they must be processed to produce best possible calibration or science data product. For example, to produce a FORS2 image, it is first necessary to create a master bias frame, use that to produce a master flat frame, and then use both to produce the final FORS2 imaging data products. If there is a choice of two or more input calibration frames, usually the closest-in-time is chosen.

Calibration products (with pipeline-added QC parameters in the headers) are saved in a temporary workspace for later QC review. If necessary, additional QC parameters are computed using off-line (“QC”) procedures. All computed QC parameters are ingested into the QC1 database and used to generate **QC reports and score reports** (discussed in more detail in the *Quality Control* section). All calibration products, QC reports and scores are generated in incremental mode.

4.1.3 Calibration Data: Quality Control, scoring and certification

The principles of the quality control process are discussed in more detail in the Quality Control chapter of this document.

Quality control and certification is an inherently instrument-specific process. Nevertheless all instruments share the same basic QC review workflow.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	24 of 52

The process begins with *scoring*: QC parameters are compared to configured thresholds. Possible values are 0 (ok: compliance) and 1 (nok, outlier). Scores are graphically displayed and stored in a database. Scoring helps to reduce the complexity of quality assessment and data volume challenges [RD05].

Certification, as the second step, is based on score results, assisted by inspection of QC reports generated during calibration processing. The QC Health Check monitor can also be consulted. The number of reviewed reports and ABs can be drastically reduced for data types which are fully scored. Then, auto-certification can be configured which offers for review only those ABs with red scores plus a (configurable) randomly selected small fraction of ABs scoring green, or even complete auto-certification of green ABs.

Problems found during certification are investigated and (whenever possible) resolved by the QC Scientist. There are many classes of potential problems including: abnormal instrument behaviour, incorrect or incomplete execution of calibration plan, incorrect association leading to faulty AB, and pipeline recipe software bugs. Iterative interaction with Paranal Science Operations and the Pipeline development team is often needed to resolve problems.

Sometimes problems could be resolved, by fine-tuning recipe parameters or fixing a recipe bug. There is usually no re-execution of ABs, only exceptionally if e.g. improvements are tested.

In unrecoverable cases, science data might become affected: if they can't be calibrated with certified products in an acceptable time range, they must be repeated. Then USD has to be informed that all affected science OBs must be re-executed.

Calibration products that pass this review are called **certified**.

4.1.4 Calibration Data: Final organization

The next step involves final organization of Calibration products. Rejected calibration products are deleted, only their QC parameters are stored.

All other certified products are then distributed from the temporary certification area to the final product directories. They are renamed into a human-readable scheme which encodes the type of file, the acquisition date, and instrument setup parameters³. All certified calibration products are ingested into the Archive.

The static (or general) calibration products are also stored in the Archive. These are files not processed from raw files but generated by the pipeline developers.

Finally, the pipeline logs, the ABs, and the graphical QC reports are stored for future reference on the `qcweb` server.

4.1.5 Science ABs

The ABs for the science data are generated once all steps are executed for the calibrations. This is done to check consistency and validity of the current calSelector OCA rules, and to have the ABs ready for future IDP projects. Science ABs are stored on the `qcweb` server.

4.1.6 Final Clean-up

The daily workflow is finished by a number of clean-up actions. The `$DFS_PRODUCT` directories related to that date are purged, as are temporary listings. Various QC daily statistics are updated and ingested into a database. Finally the raw data directory is deleted. It is not required anymore for operational purpose. With all products ingested into the Archive, they can be replaced by their headers and then be deleted.

³ The calibration product naming convention is documented on the QC Web pages: <http://www.eso.org/qc/>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	25 of 52

As part of the clean-up, information about the workflow is stored for future reference (*histoMonitor*). Then the daily workflow is finished.

4.1.7 Possible ERROR conditions in the daily workflow

Several error conditions can arise during the daily workflow which must be handled by the QC Scientist:

Master calibration products fail quality control: scoring indicates when a calibration product deviates significantly from the nominal behaviour. The QC Scientist investigates the problem and reports it to Paranal as necessary. It may be necessary to re-schedule associated calibration or science OBs. Note that QC parameters from these non-compliant calibration products are still ingested into the QC1 database if their measurement was valid. If the problem was caused by a wrong recipe algorithm or bad parameter choice, the result is not valid.

Headers are incomplete, header keys are wrong: QC has the option to ask DBCM for header updates. If header keys are systematically wrong or incomplete, QC will try to work with Paranal to solve the problem upstream.

Raw data are bad: if raw files are found to be unusable (e.g. a fully saturated flat, or an arclamp exposure with no signal), they can be hidden in the archive, by an automatic DBCM request. This prevents these files, although they remain in the archive, from being associated by any automatic process.

4.1.8 The Association Block

The **Association Block (AB)** is a central entity in the QC daily workflow (for their specification see [AD03], [AD04], and [AD05]). The AB is a generalization of the concept of the Reduction Block (RB) and the Set-of-Frames (SOF) which are text file containers of association and data processing information. Extended beyond the needs of just data processing, the AB also contains all information required for the complete association (e.g. associated raw files or optional master calibrations). It is the complete description of the data processing and association information for a group of files.

ABs are available on-line on the `qcweb` QC server.

The following tools in the daily workflow interact with ABs:

- creation of ABs: **createAB**;
- processing of ABs: **processAB**;
- update of ABs: **updateAB** (plus flag updates by **scoreQC**, **certifyProducts**);
- verification of SCIENCE ABs: **verifyAB**.

4.1.9 Raw data lifetime

As part of the daily workflow, raw data are ingested into the QC system, processed, and deleted. The lifetime of raw data in the QC system starts when raw data are downloaded (incrementally, just before processing). Subsequent processing and certification (CALIB) typically occurs on a timescale of 1-2 working days.

The lifecycle can easily be minimized, essentially to the time required for processing. This is only useful when disk space is an issue. Preparation of ABs is entirely based on database information. Raw data pixels can be deleted any time after processing. In general, raw data are downloaded once and not moved: “pixels on demand”.

4.1.10 Product data lifetime

Product calibration data are ingested into the calibration archive, as soon as they have been certified. Afterwards they can be replaced in the QC system by their headers, making the calibration product file lifetime as short as a day or less.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	26 of 52

4.1.11 History record

It is important for QC to keep a record of all data processed. This is useful for tracking processing issues, but also to have a statistical overview and a record of unusual data situations, instrument breakpoints etc.

This historical record is kept in the local file system, plus on the QC web server `qcweb` where it is organized by the `dfos` tool *histoMonitor*.

4.2 PHOENIX workflow

For the creation of science-grade data products, there exists since 2013 the IDP workflow, also called PHOENIX workflow (after the *phoenix* tool supporting it). This workflow is *not* supporting all instruments. Within the QC-v2 context, it will evolve into the QC2 process.

Each instance of the process is based on a project that requires

- a pipeline certified science-grade by SDP,
- a data product standard defined by ASG, to be able to ingest the products into the archive.

The workflow definition and the project roadmap are defined in [RD-7] and [RD08].

The selection of an IDP project is then also taking into account the impact on the user's side: an IDP project should deliver data products that are widely used (from popular instrument modes), and the step towards science-grade data products should be non-trivial, i.e. not easily be done by the users.

The term science-grade means that the products are processed to the highest-possible level that does not require knowledge of the science case.

4.3 Special workflows

4.3.1 Pre-imaging workflow

A special workflow is handling pre-imaging data. These are data taken on Paranal for certain instruments which require user-interaction (to confirm that a certain OB is observable), or to yield high-precision coordinates which are not available from catalogues. Typically these data need to be processed and shipped to the users quickly in order not to miss a certain time window for the final observations.

Speed is therefore more important for this workflow than accuracy of the results. The workflow is set up such that the raw data are acquired on Paranal, sent into NGAS, transmitted by AOG to the QC operational account, where they are associated to a static set of calibrations, pipeline-processed and transmitted back to AOG. That group finally creates an ftp account where the data are available for download by the user for a certain window in time. Later the same data are also processed in the regular daily workflow. Currently the only remaining instrument with pre-imaging is FORS2.

4.3.2 Health Check Monitor Workflow

A special, automated workflow is feeding the Health Check Monitor⁴ which provides near-real time trending information about the most crucial QC parameters. These are the ones which have been agreed on by the IOTs and which are fed by dedicated daily calibrations. These data are processed on-line, the QC parameters are captured and sent to QC where an automatic process creates trending plots and publishes them to the web. The Paranal daytime SciOps staff, when they start their daily review, have the most recent data on their HC monitor.

⁴ <http://www.eso.org/HC>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	27 of 52

5. PROCESS INFRASTRUCTURE

In the previous sections, the QC information flow interfaces and the information processing workflows have been described. In this section, the process infrastructure developed by QC to manage these interfaces and workflows is described. Some information from previous sections is re-iterated for clarity of presentation.

From the QC perspective, the current DFS data management and processing modules are seen as a toolkit that must be embedded in a process infrastructure that supports the QC operational mission and workflow. Wherever needed functionality has been missing, QC has created software tools. Most of these tools are shell scripts, using other toolsets (e.g. python, isql) as necessary. There is also the python library *pyQC* developed for QC reports.

The common process infrastructure consists of:

- a suite of common tools which has been developed by QC to cover the daily workflow,
- a number of tools which have been developed by SEG and provide core functionalities of this workflow

The tool set is called *dfos = DFO system*. It is based on the workflows described in the last chapter. It combines tools which have the workflow processes hard-coded, and all instrument-specific information externalized in configuration files.

It has been developed largely by Reinhard Hanuschik, with contributions from other QC group members. The tool set is used by all QC accounts. It is maintained, with configuration control, on one QC account (currently giraffe@muc02) and distributed from there centrally (www.eso.org/~qc/dfos/).

This chapter describes the existing process infrastructure.

5.1 Compute Infrastructure

The current QC compute infrastructure has two components: one for the QC processing, and one for personal office work

5.1.1 QC processing infrastructure

The QC processing is done on multi-core blades⁵. Many but not all accounts run on virtual machines. Most VLT and VLTI instruments share one blade with 2 other instruments. These multi-core (muc) blades have three accounts each, with a 1 TB disk per account. The incremental processing is very efficiently distributing the load on each account over time.

For each of the instruments with higher data rates (survey cameras VIRCAM and OMEGACAM, the VLT multi-channel spectrograph MUSE, ESPRESSO and MATISSE), QC uses a dedicated muc blade. In addition there is are powerful multi-core blades reserved for the IDP processes.

5.1.2 Personal workstations

Each QC scientist has one workstation as a **personal workstation**, supported by the central ESO IT service, and used to manage e-mail, office applications, personal research, web browsing etc. These personal machines can run whatever OS the QC Scientist has chosen: Windows, MacOS, Linux. These machines are not considered to be operational critical – no attempt is made to make them operationally robust. They are managed by IT and follow their standards.

5.2 DFS Installation

The standard and central DFS installation is maintained by the OTS team. Here are all the tools and files needed to execute DFS commands.

⁵The details of the current architecture are documented under <http://www.eso.org/~qc/dfos/mucblades.html>.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	28 of 52

5.3 User Account and Directory Layout

For each instrument, an operational user account is created (e.g. `uves`). For each user, an instrument-independent directory structure exists. Part of the directories reside under the `$HOME` account, another part on the data disks.

5.3.1 Process Management Directory Tree

The top-level directory tree for each account has the following layout:

```
/home/<user>/
```

where `user` is usually the instrument name. It is further subdivided:

Sub-directory physical name	Logical name ⁶	Purpose
<code>bin</code>	<code>\$DFO_BIN_DIR</code>	common <code>dfos</code> tools
<code>config</code>	<code>\$DFO_CONFIG_DIR</code>	instrument-specific configuration files
<code>doc</code>	<code>\$DFO_DOC_DIR</code>	tool help
<code>qc or progs</code>	<code>\$DFO_PROC_DIR</code>	procedures to generate QC reports
<code>jobs</code>	<code>\$DFO_JOB_DIR</code>	jobs (executable files)
<code>monitor</code>	<code>\$DFO_MON_DIR</code>	monitor files (like status, logs, GUIs)
<code>gui</code>	<code>\$DFO_GUI_DIR</code>	directory for GUI information (small scripts executable from the monitor GUIs)
<code>install</code>	<code>\$DFO_INSTALL_DIR</code>	installation directory

Table 2: Structure of process management directory tree

These directories are logically defined in a resource file `$HOME/.dfosrc` which is part of the `dfos` installation. Relevant for the `dfos` software is their logical name. For easy cross-instrument orientation they also have common physical names.

5.3.2 Data Management Directory Tree

Day-to-day operations are managed within the `dfos` data management tree. This tree contains all the `dfos` data and working directories. The root is `/dataXX` where `XX` is a number⁷. It is located on the local data disk.

The next level is defined by the instrument or account name (e.g. `xshooter`, `fors2`). The next level contains the working directories for that instrument. Most are further subdivided by `<date>`. That level is managed by the `dfos` tools. `<date>` always has the standard 'DFOS' format, which is YYYY-MM-DD, e.g. 2017-12-31.

The top-level directory `/dataXX/<INS>/` is further subdivided into:

⁶ As defined in the `.dfosrc` resource file.

⁷ Having no specific meaning other that they have been ported from previous schemes and have kept their names for compatibility of historical content.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	29 of 52

Sub-directory	Logical Name ⁸	Use
raw/<date>	\$DFO_RAW_DIR	all input raw data
headers/<date>	\$DFO_HDR_DIR	all input headers
calib/<date>	\$DFO_CAL_DIR	certified calibration products
reduced/<date>	\$DFO_SCI_DIR	reviewed science products (for PHOENIX only)
logs/<date>	\$DFO_LOG_DIR	pipeline logs, ABs, etc.
listings	\$DFO_LST_DIR (/REPORT, /RAWCOM, /PRODCOM)	nightly reports of raw files; subdirectories for reports, comments
plots/<data>	\$DFO_PLT_DIR	graphical info (QC reports) per night
trending	\$DFO_TREND_DIR	information for the trending system

Table 3: Structure of data management directory tree

Note: <date> is defined to start at 21:00 UT and corresponds approximately to be Local Chilean date at the start of the Paranal night.

Each instrument has a \$DFS_PRODUCT directory, used as the output directory for the pipelines. This directory contains sub-directories by data types, further sub-divided by date:

\$DFS_PRODUCT/<TYPE>/<DATE>

This is used as *certification area* where all new pipeline product data accumulate until they get certified and distributed.

5.4 Backup scheme

The \$HOME directories are backed-up by IT. It is assumed that they contain software only (tools and configuration), no fits files. All fits files have to go to the data disks.

The data disks have a raid configuration and provide bulk storage for not yet archived information. Their backup scheme is more relaxed since it is assumed that all vital information is archived soon, so that a data loss never affects more than a few days worth of data and can be tolerated.

5.5 Daily Workflow Process

This is a sketch of the main tool set for the daily workflow. More technical details about the tools for this process, and about the operational background how to use them, can be found under the *dfos* web site <http://www.eso.org/~qc/dfos/>.

5.5.1 General pre-processing

The first step is to retrieve headers from the Archive, generate the header summary report, and review it for potential processing problems (e.g. missing calibration data, incorrect headers). This process is executed using the following tools:

- `dataLoader -t h -d <date>`: retrieve recentmost headers from the archive for <date> and write them to `$DFO_HDR_DIR/<date>`.

⁸ As defined in the .dfosrc resource file.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	30 of 52

- `createReport -d <date>`: produce header report for <date> based on these headers. Output is an HTML file stored in `$DFO_LST_DIR`. Configuration files are used to specify instrument-specific keywords to be extracted.

These initial steps are provided in an automatic way by the workflow tool **autoDaily**. It is called once per hour. It also provides the second step and the processing of the calibration data (see next section).

5.5.2 CALIB: create ABs

The second step is to create Association Blocks for calibration data. It is executed when new calibration data are found. See also the workflow sketch in Figure 4. The `dfos` tool **createAB** creates ABs for all input raw files of type CALIB from <date> in `$DFO_RAW_DIR`:

- `createAB -d <date> -m CALIB -i`

where the option `-i` indicates incremental processing (see below).

The AB creation process has three main steps:

1. classification of files
2. grouping (organization) of files
3. association of files

These three steps are fundamental and have been generalized into the OCA framework (`org./class./assoc.`) which is the DFS standard tool for all kinds of classification tasks.

The `createAB` tool has several association configuration files [AD05]:

`<instr>.classification.h`: definition of all `raw_types` (e.g. BIAS, FLAT, ...)

`<instr>.organisation.h`: definition of all grouping rules

`<instr>.association.h`: definition of all association rules

The tool scans the headers of the raw files, the calibration product files (as managed in the `$DFO_CAL_DIR/MCAL` directory by the tool), and the so-called virtual calibration products (under `$DFO_CAL_DIR/VCAL`).

Virtual calibration products are predicted, but not yet created products. They are created by executing predecessor (“parent”) ABs before they are needed by the current (“child”) AB. The content of VCAL is dynamic – files are added by `createAB` and deleted by `updateAB` (see below).

The MCAL directory contains links to certified calibration products. Its content is also dynamic – file names are added and removed by `updateAB`.

The general/static calibrations are also scanned by the tool.

When associating calibration products with raw frames, `createAB` first tries to find appropriate calibration products in MCAL, then in VCAL. Searches are made based on the applicable time-matching rule (in most cases: closest-in-time) and correct match key values. If no appropriate match is found in either file, an error is reported.

The following output is produced:

- one AB for each organized group of raw files, plus a corresponding association log;
- a sorted list of ABs ;
- updated VCAL and MCAL directories.

5.5.3 Processing CALIB ABs

Once all available calibration ABs have been created, a single batch job is created using the `dfos` tool

- `createJob -d <date> -m CALIB`

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	31 of 52

It sorts the ABs by dependencies into an executable job file under \$DFO_JOB_DIR. Depending on the configured pipeline infrastructure, the sorting is either linear (execution one-by-one), or comes in a self-organizing cascade with AB execution dependencies which can be executed by the batch queue system.

Once the batch job script is complete, it is executed automatically within *autoDaily*.

The central engine for calling a single AB execution is the *dfos* tool

- `processAB -a <ab>`

This tool is the interface between the ABs and the pipeline. It calls the pipeline recipes via the DFS tool *esorex*. It provides the following steps:

- download raw data,
- download calibration products if required,
- execute the pipeline recipe,
- put the products in the certification area under \$DFS_PRODUCT.

5.5.4 CALIB: QC reports and updating HC plots

After AB processing, the products are analyzed using instrument-dependent QC procedures. These execute one or more of the following actions:

- extract QC parameters from product headers,
- generate pipeline-unsupported QC parameters,
- ingest all created QC parameters into the QC1 database,
- create QC-defined and instrument-specific QC reports (see *Quality Control* section),
- call scoring tool for automatic assessment.

The QC procedures are instrument-specific and are maintained by the QC scientist⁹. They are invoked by the *dfos* tool

- `processQC -a <ab>`

All QC parameters are ingested into the QC1 database using the **qc1Ingest** tool provided by DFS.

The pipeline processing batch and the QC reporting batch are logically closely related and normally invoked sequentially in the job execution file.

As part of the *processQC* script, the tool for scoring (see next section) is called by *autoDaily*:

- `scoreQC -a <ab>`

At the end all affected reports of the HC monitor are updated, by calling

- `trendPlotter -r <report>`

in a batch with all report names affected by a new QC parameter value from calibration processing.

After finishing, the AB monitor is called to be fully up-to-date:

- `getStatusAB -d <date>`

It now has the processing status of all processed ABs, their QC reports and their scores. They are now ready for review by the QC scientist.

⁹ They are documented under <http://www.eso.org/~qc/docu/QC1/qc1docu.html>.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	32 of 52

This is the end of the automatic part of the processing workflow for calibrations, as controlled by *autoDaily*. It is called incrementally, once per hour, to provide a good load balancing and fast quality feedback to the mountain.

All of these steps can be called on the command-line but there is usually no need for this.

5.5.5 CALIB: Post-processing

Once the daily batch job is completed, the resulting products must be reviewed, as described in the *Daily Workflow* section. While the previous steps were completely automatic, managed by the workflow tool *autoDaily*, the following step is interactive, at least in principle. This is provided by the *dfos* tool *certifyProducts*:

- `certifyProducts -d <date> -m CALIB .`

The review process is normally automatic, where most ABs scoring green are certified AUTO and only a small fraction¹⁰ of the green scores, plus all red-scored ABs are interactively reviewed. If a product has been rejected, it is purged. Depending on the nature of the quality issue, only the products are deleted but the virtual calibrations are preserved (raw data are valid and have no quality problem, there is only a problem with the pipeline), or: the products and the virtual calibrations are deleted (raw data are not valid and should be hidden).

Certification is supported by scoring, provided by *scoreQC*. This tool compares the new QC parameters of a product with pre-defined thresholds which come either from specifications or statistics. Based on scores, products are auto-certified.

After the certification process, all products are re-organized and renamed. By calling

- `moveProducts -d <date> -m CALIB ,`

the products are moved from the certification area `$DFS_PRODUCT` to the final storage area `$DFO_CAL_DIR`. There is the option to configure, by `PRO.CATG`, product files which explicitly should **not** be moved (they will then be purged). This is required since many pipeline recipes produce intermediate-stage products which should not be delivered to the Archive.

Certified calibration products are re-named using

- `renameProducts -d <date> -m CALIB .`

The re-name convention uses a code for the instrument and `PRO.CATG`, a date, a version, and a set of instrument parameters.

The *moveProducts* tool also moves supporting log information (e.g. association logs, pipeline logs) to the date-specific `$DFO_LOG_DIR`, and graphical information, as configured, to `$DFO_PLT_DIR`.

The certified and renamed calibration products are now ingested into the Archive using

- `ingestProducts -d <date> -m CALIB`

which uses the DFS tool *cdbIngest*. If multiple calibration products of the same type were produced for a given day and all were found to be valid, typically only one in the series is ingested, except in the usual cases for one-to-one calibration products (e.g. MOS arc-lamps). The number and type of products to ingest are instrument-specific and is configurable.

The last step in the post-processing part is the updating part. Many ABs used to process data will initially contain associated calibration products with virtual (predicted) filenames. Once all products have been certified, renamed, and re-located, these virtual names are replaced by the real names using

- `updateAB -d <date> -m CALIB.`

The replaced virtual names are removed from `VCAL`.

¹⁰ Currently 20%.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	33 of 52

5.5.6 SCIENCE: AB creation

For SCIENCE data, ABs are created but not processed. Instead, the ABs are stored in `$DFO_LOG_DIR` for possible future reference in IDP projects.

With

- `createAB -d <date> -m SCIENCE`

the ABs are created (with master calibrations from `$DFO_CAL_DIR`).

The SCIENCE ABs are then verified using

- `verifyAB -d <date>`.

The purpose of this step is to check the AB content against the results of the `calSelector` association. This helps verifying that the `DFOS_OPS` OCA rules and the `calSelector` OCA rules keep aligned. If a problem occurs it needs to be investigated and may lead to a modification of the OCA rules.

This is the end of the daily workflow for SCIENCE data.

5.5.7 Final clean-up

The **finishNight** command purges the `$DFS_PRODUCT` directories, and optionally also the `$DFO_RAW_DIR` directories. It also triggers the daily statistics collection:

- `finishNight -d <date>`

and stores valuable information for future reference in a HISTORY area managed by the `dfos` tool *histoMonitor*.

5.5.8 Summary of Commands

Here is a typical sequence of commands for the daily workflow:

<i>autoDaily:</i> incremental and automatic processing of calibrations	<code>dataLoader -d <date> -t h</code>
	<code>createReport -d <date></code>
	for CALIB data:
	<code>createAB -d <date> -m CALIB</code>
	<code>createJob -d <date> -m CALIB</code>
	<code>execute batch processing job and processQC job</code>
interactive	<code>certifyProducts -d <date> -m CALIB</code>
	<code>moveProducts -d <date> -m CALIB</code>
	<code>ingestProducts -d <date> -m CALIB</code>
for SCIENCE data:	
interactive	<code>createAB -d <date> -m SCIENCE</code>
	<code>verifyAB -d <date></code>
for all data:	
interactive	<code>finishNight -d <date></code>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	34 of 52

Table 4: Steps of the daily workflow

5.5.9 Status flags

The dfos tools write status flags into a central status file (`$DFO_MON_DIR/DFO_STATUS`). This status file is process-oriented and date-oriented. Each date has a set of status entries which are evaluated by the DFO monitor to know about the current processing status in the daily workflow, and to offer the next reasonable step(s). The flags are written by the corresponding dfos tools.

5.6 PHOENIX workflow

5.6.1 Description of the workflow

The PHOENIX workflow is the workflow for the creation of “level 1” (or even higher) science-grade data products. These are science-grade data products with the instrument signature removed, processed in a standard way to a level that does not require knowledge of the science case. Since they are produced in-house, they are also called IDPs (Internal Data Products). They are created by the QC group, for selected and defined instruments and instrument modes.

This workflow is *project-driven*, in contrast to the daily calibration workflow which is *operations-driven*. It consists of a major historical batch that may cover the entire historical data of the project, or part thereof. It has also an operational part, namely after the historical batch, when new data from the current data stream are added to the IDP dataset. See more under RD07 and RD08.

5.6.2 The phoenix workflow tool

All necessary steps for the PHOENIX workflow are encoded in the *phoenix* workflow tool, in a similar way as *autoDaily* has all steps for the incremental processing of calibrations encoded. Actually many of their steps are very similar.

The following conditions need to be met for a PHOENIX project:

- The ABs must exist in a repository (in many cases this is `qcweb`), they are not created from scratch. Additional ABs might be created by a PGI, but all association information is contained in the AB repository.
- The master calibrations as listed in the ABs exist in the archive and don’t need to be created from scratch.

We describe here the major steps.

By calling

```
phoenix -d <date>
```

the following steps are executed:

- Download the SCIENCE ABs (the configured data source on `qcweb` is scanned for the ABs in the configured SCIENCE mode). Exit if none found.
- Modify ABs, to standardize their content for historical evolution or parameter choices.
- Create the processing jobs (for pipeline and for QC reports) (breakpoint C).
- Execute the jobs.
- Offer the products for review (breakpoint P).
- Call `moveProducts`, `renameProducts` (see section 5.5).

By calling the tool with parameters `-C/-P/-M`, one can stop at breakpoints C and P, or resume processing at breakpoint P.

For efficient mass production it is possible to call

- `phoenix -m <month>`

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	35 of 52

(with the above parameters). The only hard-coded exception is MUSE where only daily processing is enabled.

Often SCIENCE ABs have no dependencies. Then CONDOR can be fed very efficiently with the pipeline jobs. In order to be able to do the same with the QC jobs, they are made parallel.

Once configured and tested, the PHOENIX workflow can be very efficient. For instance, the whole UVES ECHELLE archive with about 100,000 raw files could be processed in less than a week on the 32-core blade `muc08`.

The PHOENIX workflow is executed on separate QC accounts, well separated from the operational ones, on dedicated servers. Many dfos tools are PHOENIX aware, meaning they can be used in both DFOS and PHOENIX environments, reacting in the proper way. For instance, the `dfoMonitor` tool can be used for managing PHOENIX projects. This makes the maintenance of these tools efficient and their usage easy.

The processing history and the workflow statistics are recorded by the tool

- `phoenixMonitor`

which has been derived from the DFOS tool *histoMonitor*.

5.6.3 PHOENIX DEEP

All ABs created in the dfos environment are OB based (or refer to single files). It turns out, however, that many programmes intend to go deep and design OBs for multiple target visits. This has become obvious in the MUSE project. As a pioneer project, MUSE_DEEP has been developed. It goes beyond the OB limit and attempts to combine all OBs in a single science product, if there are indications that this is the intention of the programme.

Currently there are no strict rules for PIs to design their phase-2 OBs such that the higher-level coupling can be safely and automatically read. The mechanism for this would be containers, but their use is optional. Also there are cases where targets could be combined even across programmes, for reaching ultimate sensitivity.

The current MUSE_DEEP project has developed several tools and concepts that can serve as a prototype for similar future projects. Part of these concepts are supported by `phoenix v3` (the processing of deep ABs, and the organization of deep products). But other steps are currently handled by customized scripts:

- a step to organize all existing products (previews) by target, programme and run, in order to identify possible candidates for deep combination
- a step to interactively select the suitable candidate OBs and combine them into deep ABs;

An important technical concept is the pseudo-date, which is a code for the `run_ID`. The natural data unit for DEEP processing is the `run_ID`, but almost all dfos tools expect the date parameter. With the pseudo-date, a `run_ID` is encoded in the date format, and then all processing steps can be called as if a certain date would be processed. For instance, `run_ID 094.A-0301B` would be transformed into the pseudo-date

`2094-03-01`

where the first number encodes the run B (1 would be A etc.); numbers 2-4 are the period, and 03-01 encodes the programme number 0301. The OPC category cannot be encoded but is irrelevant.

The phoenix tool, in its DEEP mode, is called like

- `phoenix -r <run_ID>`

(i.e. by user-readable `run_ID`) but uses the pseudo-date internally.

The phoenix tool, in DEEP_MODE, sorts the final products into `$DFO_SCI_DIR/<pseudo-date>` (the same for `$DFO_LOG_DIR` and `$DFO_PLT_DIR`). This setup is capable of hosting all DEEP products in specific directories, well

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	36 of 52

separated from any normal PHOENIX project. The *phoenixMonitor*, if called with parameter `-P`, will display all products properly sorted by `run_ID`¹¹.

5.6.4 IDP ingestion

IDPs need some special preparation of their headers (“conversion”) to become compliant with the phase3 standard. They also need their own ingestion tool. Both are embedded in the call

- `ingestProducts -d <date> -m SCIENCE.`

For UVES there exists a special DFS-provided conversion tool called `idp2sdp`. For the other projects there are customized scripts.

5.6.5 PHOENIX projects

Each IDP stream starts as a project, with a historical batch. Once this is processed and ingested, the IDP project then turns into an operational stream, with (often) monthly updates. Currently there are the following PHOENIX projects:

Name of PHOENIX project	Instrument, mode	Description
UVES_ECH	UVES Echelle, point source spectra, with or without image slicer, with or without Iodine cell	more than 100,000 IDPs, covering the entire UVES history since 2000-03
XSHOOTER_ECH	XSHOOTER slit mode (Echelle grating), any observing mode	more than 30,000 IDPs, covering the entire XSHOOTER history since 2009
GIRAFFE_MOS	FLAMES/GIRAFFE Medusa (MOS) mode, with or without SIMCAL lamp	more than 1,000,000 IDPs, covering the entire GIRAFFE history since 2003
MUSE_IFU	MUSE (currently wide-field mode, no AO only)	more than 5000 datacubes, covering the entire MUSE history (operations and science verification)
MUSE_DEEP	same as MUSE, with DEEP OBs combined	more than 200 deep datacubes
KMOS_IFU	KMOS IFU	upcoming 2020

Table 5: Current IDP projects

5.6.6 PHOENIX master calibrations

With the *phoenix* tool also master calibrations can be (re)processed. This makes sense if

- the pipeline has seen substantial improvements,
- there was originally no pipeline support at all,
- the pipeline has been modified such that old masters are not compatible with the current pipeline anymore.

The motivation for a CALIB project will usually come from a follow-up IDP project which benefits from the improvements of the new master calibrations.

¹¹ see http://qcweb.hq.eso.org/MUSE_DEEP/monitor/FINISHED/P94/P94.html as example.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	37 of 52

If *phoenix* is used for such a task, ABs must be created, in the standard *autoDaily* way. However a day-by-day processing is generally not adequate, due to dependencies and the calibration plan. Instead, a data pool is created which typically is 30 days. The tool is then called in monthly mode,

- `phoenix -m <month> -X`

where `-X` marks the master calibration mode for the tool. These are the principal workflow steps:

- Create the CALIB ABs for the configured modes and setups; employ the full data pool, using virtual calibrations.
- Create the CALIB processing jobs (for pipeline and for QC reports) (breakpoint C).
- Execute the jobs.
- Employ the *qualityChecker*, a module which combines current and historical QC information, to prepare for the *autoCertifier* (breakpoint P).
- Call *autoCertifier*, review the marked ABs.
- Call *moveProducts*, *renameProducts* (see section 5.5).
- The product ingestion is done in a separate step and includes a module for deleting any previous instance found.

The history of the CALIB project is stored with *phoenixMonitor*, in the same way as for IDP projects.

5.7 Special workflows

5.7.1 Pre-imaging workflow

The pre-imaging process is set up to run automatically. It is executed by AOG. They recognize and submit pre-imaging data to the respective QC machine, then trigger a processing tool, and finally capture the products. The QC part of the process is handled by the DFOS tool *processPreImg* which is generic and provides some monitor functionality. It is a workflow tool working with frozen versions of DFOS tools which are used on those particular accounts only. The execution status of a pre-imaging run is monitored on <http://www.eso.org/~qc/monitor/PREIMG/preimgMonitor.html>.

5.7.2 Health Check Monitor

The main QC tool to create Health Check (HC) plots is *trendPlotter*. It reads QC parameters from the QC1 database, prepares the graphics and puts them on the web. It is called like

- `trendPlotter -r <report> -f,`

usually as a cronjob.

A second tool interacts with the *trendPlotter* results to provide scores for the HC monitor¹². It is called by *trendPlotter*, as

- `scoreHC - r <report>`

and collects all scores for the report, as previously derived for the ABs by *scoreQC* and ingested into the score database.

5.7.3 calChecker workflow

This workflow is called as a cronjob twice per hour:

- `calChecker .`

¹² Under www.eso.org/HC

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	38 of 52

It creates the calChecker web site¹³, plus a local web site. It has two modules, one for checking the completeness of SCIENCE files (the “core” tool), and one for checking the completeness of calibrations for other calibrations (typically STD stars) called CAL4CAL. For some instruments like e.g. XSHOOTER, there is a third module, *tellTracker*, which monitors the completeness of telluric STD stars which are required within 2 hours of the science observations.

5.7.4 Workflow for management of *calSelector* rules

One task of the QC group is to create and maintain the OCA rules for CALSELECTOR, the archive service to find the calibration data for user-selected science data¹⁴. For each instrument, the current association rules used within DFOS are migrated into the current rules for CALSELECTOR, with some minor modifications to properly describe the validities of the calibrations. For the migration, the tool *calselManager* is used. This tool is a wrapper around the DFS tool *calSelector* and its various components, mainly the *rulesManager* component. In addition to options providing help and assistance for the initial migration, the tool offers support for

- ingesting the CALSELECTOR OCA rules into the database,
- download them from there,
- delete them in the database.

More about the complex functionalities is available with

- `calselManager -h`.

In many cases the CALSELECTOR OCA rules need to be versioned when looking back into the data history of an instrument. Therefore each rule in the database has a validity range which is defined upon ingestion.

Virtual products defining the raw input datasets are created upon ingestion of an OCA rule, and deleted when a rule is deleted.

In order to visualize the structure of a calibration cascade as defined by an OCA rule, the tool

- `createCalibMap`

is called. Its output, together with the rules, is collected under a common website¹⁵.

As part of the daily workflow, the SCIENCE ABs are checked against the CALSELECTOR-generated ABs, using the tool

- `verifyAB -d <date>`.

Effectively this is a check for consistency between the DFOS_OPS rules and the current CALSELECTOR rule.

5.8 DFS pipelines and infrastructure

In the core of the DFS process infrastructure are the data processing pipelines. That part of the infrastructure has three main components:

DFS Component	description	example
Batch queue system (BQS)	A system to support the call of a pipeline recipe. The BQS knows about the proper syntax of the recipe call. It provides the standard interface to the recipes. It also evaluates AB dependencies and	CONDOR → processAB → esorex → recipe

¹³ Under www.eso.org/CAL

¹⁴ For more information about CALSELECTOR from the QC perspective, see <http://www.eso.org/~qc/dfos/calSelector.html>.

¹⁵ http://www.eso.org/observing/dfo/quality/ALL/OCA/oca_rule_sets.html

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	39 of 52

	schedules ABs accordingly. The standard BQS system for QC is CONDOR.	
pipeline	Series of recipes which correspond to specific instrument modes and/or data types. To be pipeline-supported means a DRS recipe exists.	VISIR pipeline
Data reduction system	The processing engine, invoked by the recipe. Typically a recipe consists of several modules corresponding to different processing steps. Modules can be shared if they are common enough. The DRS is the incarnation of the algorithms (e.g. the C code).	all pipelines based on CPL (common C library)

Table 6: DFS pipelines and infrastructure

The *Batch queue system* is the interface to the pipeline. It has two main functionalities:

- invoke the reduction job from the command line,
- evaluate reduction job dependencies and schedule ABs accordingly.

In the QC world, reduction jobs are contained in an Association Block (AB) (see Sect. 4.1.8). Processing of each AB is invoked by the *dfos* tool *processAB* which extracts the relevant AB information into a set of frames (SOF) and then uses the DFS tool *esorex* to launch the pipeline recipe.

In this scenario, ABs are called one-by-one, and the proper execution sequence (imprinted by the dependencies of the ABs due to their position in the cascade) is controlled by the user, or by a job list prepared by *createJob*. A more powerful tool is the Batch Queue System CONDOR which is an external software package¹⁶ installed by DFS on the muc blades. This system evaluates the dependencies of cascaded ABs by using a WAITFOR mechanism: a dependent AB can only be scheduled if its 'parent' AB is properly executed. The system can use as many nodes as are available (currently configured between 8 and 56 depending on the muc blade in use).

Another system is used by the MUSE pipeline: internal parallelization. A MUSE recipe has certain parts that can split into up to 24 threads (depending on the available number of nodes). The number 24 comes from the number of spectrographs and corresponding fits file extensions. With careful orchestration, even two or three of such parallel streams can run on one blade, provided it is powerful enough¹⁷.

The pipelines are specific for each instrument. Each pipeline consists of a series of recipes. These are installed on the QC machines and maintained by DFS. The Data Reduction System layer is the lowermost one and is invisible for the QC scientist. The common pipeline library (CPL) has implemented all pipeline recipes with modules of a C library.

Two other components of the DFS are used on Paranal but not by QC:

- **Data Organizer (DO):** used to classify raw frames (“what am I?”), to associate them with calibration products necessary to process them, and to build Reduction Blocks (RBs) containing the information necessary to process the data. Classification and association are rule-based. Required calibration data reside in a **local calibration database**. For QC, this mechanism is completely replaced by *createAB*.
- **Reduction Block Scheduler (RBS):** executes a RB “first-in-first-out” queue. Translates RB into format needed to drive the Data Reduction System recipe.

These modules support a static, linear process philosophy appropriate for the quick-look, on-line environment that exists on-site on Paranal.

The main reasons for QC to implement different tools come from the operational differences between Paranal and QC:

¹⁶ Developed by the University of Wisconsin (find more information under <http://www.cs.wisc.edu/condor/>).

¹⁷ The current most powerful muc is muc11 with 56 cores and 2 TB memory. It can support up to 3 parallel recipe executions with internal splitting.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	40 of 52

- Data must be processed in specific order that may not correspond to sequence of data acquisition.
- The creation of calibration products is often dependent on other calibration data not present in a static local calibration database.
- QC recipe parameters may be different in different situations.

5.9 Infrastructure: monitors and GUIs

The standard daily workflow consists of a standard sequence of steps which can be called interactively or automatically.

The automatic approach is certainly preferable for those parts of the workflow which do not require decisions and follow a strictly periodic pattern, like header download and report creation. Other parts, mainly product certification, are interactive and require the QC scientist's expertise and decisions.

Generally speaking, the goal is to provide the QC scientist with all necessary information when decisions have to be taken, and to delegate all routine actions (the ones which do not require decisions) to background processes which are just monitored by the QC scientist.

A number of monitors and graphical user interfaces (GUIs) has been developed as part of the dfos system. These are:

- the DFO monitor (to monitor the daily workflow progress);
- the histoMonitor (to store relevant daily workflow information and present it for research);
- the phoenixMonitor (similar, for PHOENIX projects);
- the calChecker (to monitor proper execution of the calibration plan);
- the scoreManager (to list and control the scores for the HC monitor);
- the AB product monitor for the progress of AB processing.

These monitors are all web-based and use browser technology to offer active buttons which can be launched by the user to call a certain tool. That tool will usually update, after execution, the respective monitor so that the whole system is self-maintaining.

5.10 Infrastructure: workflow tools

The initial part of the daily workflow is supported by the workflow tool *autoDaily* which combines the automatic execution of AB creation, file download, pipeline processing and QC reporting for the CALIB files (section 5.5).

The PHOENIX process is supported by the *phoenix* workflow tool (section 5.6.2).

The management of CALSELECTOR rules is supported by the *calselManager* workflow tool (section 5.7.4).

The pre-imaging processing is supported by the workflow tool *processPreImg* (see section 5.7.1).

5.11 Supporting information: nightlogs, problem reports, history

Paranal nightlogs are a source of information in the daily workflow ("what kind of problems occurred?"). Nightlogs are available via the User Portal.

As part of the nightlogs, problem reports from the 'USG/QC' section of the nightlogs are extracted into the PPRS system which notifies the QC scientist automatically. This mechanism helps to track these problems and their solution.

The *histoMonitor* tool creates a local web site on the QC machine, plus a mirror on qcweb, and provides overview per day, comments, links to data reports, AB reports and nightlogs. In a similar way the *phoenixMonitor* collects the processing history for a PHOENIX project.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	41 of 52

5.12 QC web server *qcweb*

QC has its own web server *qcweb* to store and display information from the daily workflow:

- the AB product monitor pages (under http://qcweb.hq.eso.org/<instr>/logs/<date>/status_<date>.html)
- the data reports (under http://qcweb.hq.eso.org/<instr>/REPORT/list_GIRAFFE_<date>_data.html)
- the historical PI packages (under <http://qcweb.hq.eso.org/<instr>/FINISHED/>)
- the product monitor is fully linked with ABs, logs, score files, and QC reports, all under <http://qcweb.hq.eso.org/<instr>/logs/<date>> and <http://qcweb.hq.eso.org/<instr>/plots/<date>>.

These pages are linked to the *calChecker* main page, are visible ESO-internally (partly under password protection) and offer easy access for PSO daytime astronomers to QC information. They are also linked to the exported version of the *histoMonitor*, under <http://qcweb/<instr>/monitor/FINISHED/histoMonitor.html>.

The QC web server is the central memory of QC about historical processing. In particular, stored SCIENCE ABs may become valuable years later for follow-up PHOENIX projects.

5.13 Web server *stargate1*

This server is used to store the internal QC web pages, like the *dfos* documentation (<http://www.eso.org/~qc/dfos/>).

6. QUALITY CONTROL

6.1 Basic Principles

ESO has implemented a multi-layer quality control process. The basic principles of this process have been presented in [RD03]. They are briefly repeated here.

Fundamentally, ESO controls instrument performance, not science quality. The quality control process focuses on assessing calibration data to determine and track instrument performance via the measurement of QC parameters. Science data are mainly checked on site, in real time, their quality control is done with respect to the ambient parameters but not to instrument properties. It is a fundamental operational assumption that the science data are technically correct

- if the provided Exposure Time Calculator was accurate (against which the PI planned the observations),
- if the instrument was performing correctly (to be checked on the calibration data: QC1),
- if the calibration data reflect the instrument properties correctly,
- and if observations were obtained under user-specified observing conditions (as checked by Paranal staff during the night: QC0).

6.1.1 QC0

Paranal Science Operations (PSO) is responsible for the level 0 of quality control, called QC0. This process has several components. In the first component, the Paranal night-time staff assess whether or not an OB has been executed under the same or better observing conditions specified by the user. This check is performed manually by the Paranal night-time staff, sometimes using output from the on-line pipeline. Based on this assessment, the OB is classified and a decision is made whether or not it should be repeated in the future.

The second QC0 component is an automatic data integrity check. A small number of properties are checked, such as the basic validity of the FITS header and pixels, the presence of key header parameters, etc. This check is performed by the DFS Data Handling Sub-system (DHS).

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	42 of 52

The third QC0 component is the daily review of calibration data. Before SciOps 2.0 (the operational model on Paranal being in place since 2011), this was a series of manual checks on all acquired raw calibration data, to assess their basic integrity and to confirm that all necessary calibration data have been acquired. With the closed QC loop and SciOps 2.0, this has been replaced by the automatic *calChecker* (completeness checks) and score-based Health Check monitor (scored calibration quality checks). The task of the Paranal daytime QC is now basically to review these pages, check if there are issues and follow them up.

The QC *calChecker* tool, the raw file previews, and the Health Check Web pages provide information to support this review process.

6.1.2 QC1

The QC group is responsible for level 1 quality control, generally known as QC1. The QC1 process has several components. These components are directly related to measurable quantities.

The most important component is the measurement and analysis of QC parameters. Examples of such parameters are median bias level, readout noise, flat-field level, photometric zeropoint, spectrographic dispersion, etc. They might refer to raw or processed data. In most cases, parameters are computed by pipeline recipes and stored as header keywords in pipeline products. In some cases, they are computed using QC maintained scripts. The QC parameters are extracted and ingested into the QC1 database. As part of the daily workflow described above, QC assesses the value of each parameter, as well as how parameters values are evolving with time. This former assessment allows QC to detect sudden instrument performance changes, while the latter assessment (known as **trend analysis** or **trending**) facilitates the detection of more gradual changes.

The core of the QC process is the **Health Check (HC)** – a comparison of the QC parameters from the latest calibration data with the historical QC1 values computed earlier. The HC monitor¹⁸ collects the trending of the most important instrument properties, together with an evaluation about compliance with thresholds (scoring). Paranal staff as well as QC scientists are in charge of inspecting the Health Check Monitor (see below). The QC Scientist making this check is often more experienced with the calibration data and can sometimes detect problems not obvious to the Paranal staff member. On the other hand, Paranal staff is looking at the current, most recent data (while QC certifies previous dates), and can “touch the instruments”, i.e. initiate actions directly.

6.1.3 Scoring

The concept of scoring is part of the QC process to manage data complexity and high data volume. QC parameters are evaluated against configured thresholds and scored as 0 (ok) or 1 (nok). This approach essentially searches for outliers. Thresholds can be static (configured from specifications or experience) or dynamic (statistical thresholds).

Scoring is implemented in two stages:

- the QC reports contain scoring per QC parameter,
- the Health Check reports read scores and evaluate them.

Scoring is a key concept since it

- applies objective criteria in the certification process,
- applies the same criteria to all products (“it never gets tired”),
- can support complex situations (N parameters from M detectors),
- is flexible: it can adapt to the evolution of a parameter.

¹⁸ <http://www.eso.org/HC>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	43 of 52

6.2 QC Reports

6.2.1 General Overview

The basic daily QC process workflow is:

- **QC parameter creation:** generate QC parameters using pipeline recipes or QC scripts;
- **QC parameter ingestion:** capture them from the file header or local files and store them in the central QC1 database;
- **Scoring:** evaluate QC1 results against thresholds;
- **QC report creation:** using QC scripts, generate a graphical report of the products;
- **QC report review:** review the reports and use them for product certification;
- **Update HC reports:** update standard HC reports to include the new QC parameters (these reports form the Health Check monitor).

The three main components are: QC report creation, scoring, and review. They are provided by the *dfos* tools *processQC*, *scoreQC* and *certifyProducts*.

The QC scripts are called by the wrapper tool *processQC*. They fulfill the tasks of capturing and ingestion (and in some cases computation) of QC parameters, and of QC reports creation.

Collection of QC parameters from the headers is supported in a standard way by the tool *writeQC*.

Ingestion of QC parameters is supported by the DFS tool *qc1Ingest*. Once in this database, QC parameters can be browsed or plotted.

Most of the QC reports are complex multi-panel figures designed for a quick analysis of a specific area of instrument performance by the QC Scientist. These reports are created off-line, and then reviewed as part of the daily workflow. QC reports get stored on the QC web server *qcweb*. Their content is documented on <http://www.eso.org/~qc/docu/QC1/qc1docu.html> (created by the *dfos* tool *qcDocu*).

6.2.2 QC1 Database and Tools

The **QC1 database** with the associated **QC1 browser** and **QC1 plotter** tools stores all QC parameters generated by QCG. The browser tool allows users to search for QC parameters within a date range, while the plotter allows users construct simple XY plots with a variety of filter, statistics, and plotting options.

The QC1 database and its tools are available on-line from the Archive under the URL <http://archive.eso.org/bin/qc1.cgi>.

User documentation is available on-line from http://www.eso.org/qc/QC1_database/.

6.2.3 QC1 plots

As part of the daily workflow, each instrument has a Quality Control and certification process (see Sect. 5.5.3 and 5.5.5). To support this process, the QC group creates special QC reports that not only visualize QC1 data but also other pipeline products, like previews of spectra. These plots are generated after pipeline processing. They are created by specialized procedures created and maintained by the QC scientist. As a common codeshare platform, the python library *pyQC*¹⁹ has been developed. Many reports are also written in MIDAS.

The reports are reviewed during the certification part of the daily workflow. They are saved in the DFO directory structure and also on the QC web server *qcweb* for later reference. They are accessible through the *histoMonitor*.

¹⁹ <http://www.eso.org/~qc/tqs/pyqc/pyqc.html>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	44 of 52

6.2.4 Health Check Monitor and tools

The Health Check Monitor provides an overview of the current instrument performance in near-real time. It also provides links to supplementary information (e.g. example master calibration frames) useful for calibration QC.

The main feature of this monitor is a set of specialized trending plots that combine historical and current QC data. See Sect. 4.3.2 for a process description. For further details, see <http://www.eso.org/qc/HEALTH/>.

The HC monitor has a link to history data. These are historical trending plots in the same format as the current HC plot. The full set of history plots constitutes the instrument performance record.

The HC monitor and the historical plots are created by the `dfos` tool *trendPlotter*. This is a tool which reads various data sources (QC1 database, local data files), calls a helper `python` tool (called `pet.py`) to create plots, creates an HTML page and exports all results to the HC web site. Each report has its own configuration file. The result page includes the plot, close-ups, a result table and some basic documentation.

The plots and pages generated by *trendPlotter* should always be up-to-date within one hour.

The HC reports have a quick-look version which is based on the evaluation of the scores. These are read from the score database, re-arranged according to the QC parameters trended in a specific report, and aggregated to higher levels, like report score, group score, and instrument score.

6.3 Certification

The certification step in the daily workflow is supported by the results of the scoring. For each AB, the call of

- `scoreQC -a <ab>`

creates a score report. This report is linked to the AB monitor. It details, per QC parameter, the score result and provides a link to a dynamic query to the QC1 database (Figure 5). Multi-detector instruments also have a detector report, with scores re-arranged per detector (Figure 6). The total score is displayed on the AB monitor.

The certification process is configured to become semi-automatic by passing all products related to ABs which score OK. Only the NOK ABs are reviewed for problem analysis.


Parameter score report

GIRAF.2008-03-02T12:07:10.339_tp1.ab
 RAW_TYPE: FFLAT
 setup: Medusa1_H525.8B
 time range: 2007-12-12 ... 2008-03-11

[AB](#) | [ALOG](#) | [PLOG](#) | [QC1_plotter](#)

num_fib	num_out	mean_width	rms_width	flux
■	■	■	■	■

HC plot(s): [HC_Nfibers](#) | [HC_flux](#)

coversheet: 

Score data: [details ...](#)

✘ score result: 2/5 best: 0/5

← scores per QC1 parameter

← total score for this product

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	45 of 52

Figure 5. Score report

GIRAF_2008-03-02T05:56:16.720.ab	compl.	OK	giscience	SCI_MED	Medusa2_H525.8B	OK	P_LOG	0.9	DONE			OK
GIRAF_2008-03-02T12:07:10.339_tpl.ab	compl.	OK	gimasterflat	FFLAT	Medusa1_H525.8B	OK	P_LOG	2.7	DONE	✗	(2/5)	REJ
GIRAF_2008-03-02T12:18:52.880.ab	compl.	OK	giwavecalibration	WAVE	Medusa1_H525.8B	OK	P_LOG	0.7	DONE	✗	(4/4)	REJ
GIRAF_2008-03-02T12:51:32.483_tpl.ab	compl.	OK	gimasterflat	FFLAT	Medusa1_H525.8B	OK	P_LOG	3.0	DONE	✓	(0/5)	OK
GIRAF_2008-03-02T13:02:40.932.ab	compl.	OK	giwavecalibration	WAVE	Medusa2_H525.8B	OK	P_LOG	0.9	DONE	✓	(0/4)	OK
GIRAF_2008-03-02T13:15:29.979_tpl.ab	compl.	OK	gimasterbias	BIAS	1	OK	P_LOG	0.5	DONE	✓	(0/6)	OK
GIRAF_2008-03-02T13:26:16.207.ab	compl.	OK	giwavecalibration	WAVE	Argus_H525.8B	OK	P_LOG	1.5	DONE	✓	(0/4)	OK

Figure 6: AB monitor with score flagging.

6.4 QC parameters not computed by pipelines

There are cases when QC parameters are not computed by the pipelines but by the QC procedures. Possible reasons can be:

- The pipeline cannot compute them, since e.g. a comparison between different products is required, or comparison to earlier reference frames.
- For some QC parameters (like photometric transformation coefficients for extinction and color terms) the accumulation and critical selection of many input data points is required and cannot be handled by an automatic procedure.
- There are other cases when the situation is too complex to be treated properly by the pipeline.
- A new parameter might be experimental and not yet available in the pipeline.

It should be the goal of the QC scientist to have all QC parameters processed by the pipeline recipes. Obvious advantages of that approach are standardization and reduction of the processing burden. Furthermore all recipe-provided QC parameters go into headers, from where they can be retrieved e.g. for the HC process. QC parameters not produced by the pipeline are stored in the QC1 database but not in the headers.

7. QC WEB PAGES AND DOCUMENTATION

7.1 Basic principles

As part of their general tasks and responsibilities, the QC scientists maintain a set of instrument-specific web pages with documentation about the QC process of their instrument. These web pages are known as the 'QC web pages' and are found under <http://www.eso.org/qc>. Their high-level purpose is to provide documentation about the QC process, the data and the data packages.

The documentation is public. The main customers of these QC pages are:

- Paranal Science Operations, who find here the prime source of information about all aspects of the *data* of their instrument;
- external users, who find here a documentation of the data processing and association, of pipelines, and of instrument trending information;
- internal users, like instrument and detector groups and pipeline developers;
- QC group members, who find a complete historical record of the evolution of their instrument, which becomes especially important in case of assignment changes and new group members.

In addition there is the web site of the Health Check monitor, (<http://www.eso.org/HC>), and the web site of the *calChecker*, (<http://www.eso.org/CAL>). The HC monitor is maintained by configuration files and the tool *trendPlotter*.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	46 of 52

7.2 Structure

The QC web pages should follow certain standards and have a common structure. For each instrument, there is always an index page, serving as main entry point²⁰, and two main sections:

- Quality Control: the process as such, with a documentation of QC0, QC1, and trending; tutorial pages with additional information about the trending reports, instrument evolution, unusual effects etc.
- Pipeline: that section collects information about both raw and processed data, pipeline recipes and algorithms, calibration plan etc.

The tutorial pages are also linked to the HC monitor.

7.3 Maintenance and issues

The creation and maintenance of the QC web pages is task of the responsible QC scientist. There are certain policy statements²¹, as well as general ESO rules about web pages. Generally, these pages should be kept up-to-date and complete. A new instrument will see its core set of QC web pages developed soon after (or even before) start of operations. The goal is to be essentially complete in the web documentation within a year after that date.

There is a dfos tool for managing the QC documentation pages, called *webCMS*. It maps the existing QC web pages into a local directory tree in the operational account. The web page maintenance takes place in this directory tree. The QC scientist can use any editor for updating the content. Then *webCMS* is called on the specified directory, and the corresponding web page is automatically updated. The entire internal structure of the web pages, and the operational rules about content, are encoded in the *webCMS* tool. This provides ease of maintenance and common look and feel.

Some other tools exist for the management of web pages, and also some dfos tools create HTML formatted output. In particular the following ones:

- *webNavbar*: a tool to create navigation bars in a standard and easy way;
- *webCMS*: a tool to create the QC web pages in a standard way;
- *trendPlotter*: creates not only the content but also the structure of the HC pages;
- *calChecker*: the same for the calChecker pages;
- *qcDocu*: creates the documentation pages for the QC reports;
- *rawdisp2reference*: created and manages the raw file displays.

These tools are listed on the dfos help page under “TQS” (trending and QC).

All public QC pages are signed by a mail address as to provide a contact point. The functional address from the list of QC mailing addresses is used which can be easily maintained.

²⁰ See for example http://www.eso.org/observing/dfos/quality/index_giraffe.html

²¹ Collected under <http://www.eso.org/~qc/procs/>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	47 of 52

8. PROCESS MANAGEMENT & CONTROL

8.1 Goals

QC has the following cross-instrument (“common”) process management goals:

- common daily workflow;
- common process and data management directory structures;
- common QC and certification processes;
- common cronjob pattern;
- common usage of tools, instrument-specific information externalized in configuration files;
- common high-level interfaces to control and track the workflows;
- common development and maintenance standards;
- common nomenclature for tasks and products;
- common process control statistics & reports;
- common look and feel of all public and operational web pages.

8.2 Current status

8.2.1 Process metrics

Metrics for the daily workflow are volume and performance of the process. The *dfos* tool *extractStat* measures the parameters of the daily workflow as listed in Table 7. These parameters are stored in a QC1 database table.

Statistics parameter per date	monitoring goal
number and volume of raw and product CALIB frames	required bandwidth and disk space
number of processed ABs	processing complexity
number of certified ABs	QC scientist’s workload
number of created SCIENCE ABs	association statistics and complexity
AB and QC report execution times	muc blade performance
Processing delays (between file availability and QC parameter)	QC process efficiency

Table 7: Statistics parameters for daily workflow

Graphical reports are available under the WISQ site (see 8.2.2).

8.2.2 Workflow information system WISQ

The database system storing these metrics is called WISQ: workflow information system for QC²². It uses the QC1 database technology and interfaces. There are also reports in the *trendPlotter* style.

²² <http://www.eso.org/QC/WISQ/overview.html>

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	48 of 52

Metrics in WISQ	workflow part
(compressed) data production on Paranal	[read from database]
raw data downloads	daily workflow
product data uploads	daily workflow
AB processed	daily workflow
package statistics (until 2011-10-01)	package workflow (not maintained anymore)
delay statistics	daily workflow
workload per QC scientist	daily workflow
score significance	score configuration
IDP downloads	[read from database]
IDP quality control	phoenix processing

Table 8: QC metrics in the workflow information system

8.2.3 Process management tools

A number of common monitors and GUIs have been developed. For details see Sect. 5.9. The interfaces have a stripped-off version which is a passive monitor and which is, for each instrument, collected under the central URL <http://www.eso.org/~qc/monitor/monitors.html>. The version on the local QC system is an active web-based GUI which is maintained and updated by the process itself.

A number of status flags for the daily workflow are written into local status files under \$DFO_MON_DIR (DFO_STATUS). This is a simple mechanism to follow the status of these workflows. The *dfoMonitor* reads the status flags and evaluates them in order to decide about the next steps to offer to the user.

8.2.4 Process development

The processes described here have been shaped in the first year of VLT operations (1999) and have evolved over the years. While the process was essentially common to all instruments, there was originally one 'proprietary' solution per instrument implemented.

By 2006, the QC process for all instruments had been aligned the common path. Since then, the common process has been extended, improved and adapted to ever changing conditions. The major milestones have been:

- archive ingestion of science and log product files (2007);
- implementation of calChecker and scoring (2008);
- implementation of the fast data transfer (2008);
- implementation of database storage of package structure (2009); extension of packing to Visitor Mode;
- de-commissioning of science processing and packing (2011);
- implementation of *calSelector* associations for science data (2011);
- implementation of PHOENIX process for IDPs (2013);
- migration to raw-to-master mode for calSelector (2015);
- extension of PHOENIX to generation of historical master calibrations (2015);
- implementation of raw display service (2015);
- implementation of PHOENIX DEEP (2016).

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	49 of 52

Because even stable processes slowly drift and evolve, tools to implement them also need to evolve. The experience with the dfos tools clearly demonstrates that a script-based approach provides all the flexibility and transparency required for an easy adaption. Shell scripting skills are expected from each QC scientist, so that adaption of an existing tool to a modified process can be implemented by one QC person without creating a single-failure point.

It has therefore been decided to delegate the development of some core tools to DFS while all other tools are kept under direct control of the QC group. Examples for core DFS tools are: *fitsreport*, *ABbuilder*, *HdrDownloader*.

All dfos tools are documented under <http://www.eso.org/~qc/dfos/>.

8.2.5 Process Implementation

For every new instrument mode, the following integration scope of work has to be followed:

- create pipeline products & related QC parameters;
- create QC1 database configuration files;
- configure scoring system;
- create QC reports for pipeline products and QC parameters;
- ingest products and QC parameters;
- create QC Web pages: main (index) page, sections “Pipeline”, “Trending & QC”;
- create initial CALSELECTOR OCA rule set.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	50 of 52

9. FUTURE ISSUES, FUTURE DIRECTIONS

9.1 Data processing

9.1.1 Data rates

Data volume and complexity has grown since 1999 and will continue to grow over the next years.

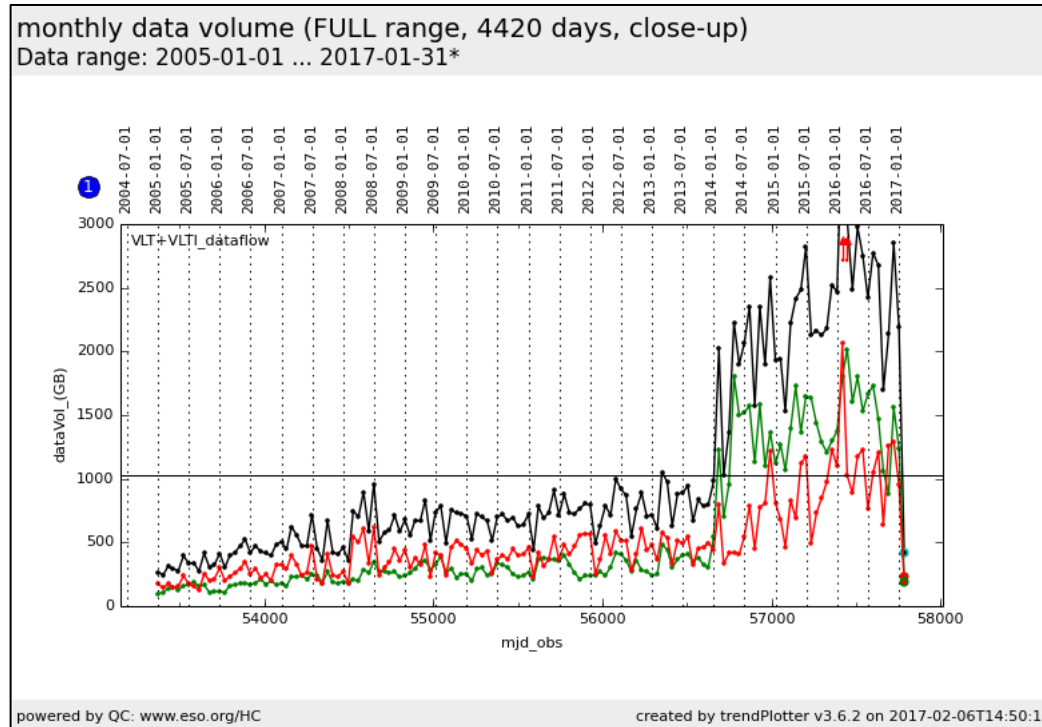


Figure 7: Evolution of VLT(I) data rates (compressed, relevant for data transfer). Red: science data; green: calibration data; black: sum of both

In Figure 7, the data rates for the VLT(I) instruments (from telescopes UT1-4) are collected. The evolution until about 2008, with an average increase by 90GB/month per year was characterized by adding more and more instruments (like e.g. X-Shooter and HAWKI). Until the end of 2013 the data rate was increasing more slowly (by 40 GB/month per year). The commissioning of MUSE then gave rise to a sudden increase to the current level.

The VLT/VLTI data rates are currently in the range of 2.3 TB/month. The VIRCAM data rate is between 1 and 2 TB/month, with seasonal variations. The OMEGACAM@VST data rate is slightly more than 1 TB/month. Together, the Paranal instruments currently (2017) deliver about 5 TB of compressed data per month²³. Not surprisingly, the various regimes of the trend over the years are clearly identified by the start of operations for the survey telescopes (with VIRCAM and OMEGACAM) and the MUSE instrument. All three instruments have arrays of detectors (VIRCAM: 16; MUSE: 24; OMEGACAM: 32) and produce together about 80% of all Paranal data. In comparison, the “burst” modes of certain instruments, originally thought to become a major source of data volume as well, currently do not produce a noticeable fraction of data.

²³ Details in http://www.eso.org/observing/dfo/quality/WISQ/FULL/trend_report_dataflow_MONTH_HC.html.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	51 of 52

9.1.2 Processing strategies

The QC group is currently processing about 5.5 TB/month (CALIBs only) in the daily workflow, 2.5 TB of this is MUSE only. The IDP projects add another 0.5 TB/month, almost exclusively MUSE. While the other IDP projects are also significant in terms of number of ABs and products, their volume is negligible in comparison. During history processing, there are occasionally additional peaks of terabytes per month.

While MUSE seems to remain dominant over the next years in terms of data volume and processing load, there might be a new evolution with the AO-supported instruments on UT4. It needs to be seen how this will affect the design and execution of science OBs.

9.1.3 Compute server strategies

QC has adapted to the historical changes by moving from four single-processor HP machines (1999-2002), over a system with 13 dual-core blade servers (one per VLT instrument) plus a cluster of 20 similar blades (for the survey instruments), as used until 2012, to the current system of 8 multi-core blades for the daily processing of the Paranal calibrations and 3 multi-core blades for the PHOENIX projects.

While the number of detectors and hence the number of pixels to be processed is relatively stable over the years, once an instrument is put into operations, the compute platform can be scaled and adapted relatively easily. Therefore, neither disk space nor compute resources are currently a bottleneck²⁴.

An optimized data flow operations model leads to several general requirements for the QC compute servers:

- the number of times a byte is moved should be minimized;
- the number of co-existing data copies should be minimized;
- the life-time of any given data copy should be minimized.

The first requirement is driven by network bandwidth while the latter two are driven by data storage capacity. Obviously, the exact implementation of these requirements depends on bandwidth and capacity availability and cost. For current instruments, these requirements led to a data management process where raw data are retrieved from the Archive on demand, when they are processed, and then deleted immediately.

Data complexity can be characterized as number of floating point operations per byte. Using this definition, spectroscopic data is more complex than imaging data, and spectroscopic data complexity varies according to mode (i.e. long-slit, multi-object, integral field). Naturally, survey support increases imaging complexity because more operations per byte are needed. From this viewpoint, it is clear that MUSE is currently the ESO instrument with highest data complexity and compute requirements. Future ESO instruments will likely not have higher data complexity than current ESO instruments.

In other words, the number of operations per pixel is unlikely to increase by a large factor in the future²⁵ – but there will be more pixels to process.

Another uncertain aspect in these forecasts is the fact that some future or current instruments (like VISIR or GRAVITY) have operational “burst” modes, triggered e.g. by rapid atmospheric variability which is tackled by the attempt to freeze it by rapid read-outs.

The current QC compute model is very flexible and has been configured as follows²⁶:

²⁴ A total of 252 cores (7*12 cores plus 2*32 cores plus 1*48 cores plus 1*56 cores = 252 cores) is available for processing on the current 11 muc servers.

²⁵ This statement assumes that pixels are in general only processed once. Data complexity can also increase if pixels are processed many times, due to (e.g.) calibration problems.

²⁶ More and up-to-date information available under the URL <http://www.eso.org/~qc/dfos/mucblades.html>.

ESO	QC Handbook	Doc:	VLT-SPE-ESO-19600-3986
		Issue:	5.2
		Date:	2020-01-24
		Page:	52 of 52

Server	Usage
muc01...muc04:	three active accounts per UT (low-data volume instruments)
muc05:	three active VLTI accounts, plus pre-imaging (low-data volume instruments)
muc06, muc07:	one account each, survey cameras (muc06: ocam, muc07: vircam)
muc08:	phoenix processing
muc09:	operational account for MUSE
muc10:	phoenix processing for MUSE
muc11:	phoenix processing for MUSE; operational account ESPRESSO
per account:	1 internal disk, with dfos software, pipelines etc. 1 external disk (1 or more TB), with data and long-term memory

Table 9: Current QC compute server model

The shared account servers (muc01...muc04 and also muc08) benefit a lot from the automatic resource sharing provided by condor. If no concurrent processes from the other accounts take place, an account can use up to 6 core at once. In case of competition, condor provides a “first-come, first-served” service which is self-managing and on average still very efficient. On the single-user machines muc06/07/09, up to 32 cores can be used for parallel processing tasks which is a perfect match even for OMEGACAM with its 32 detectors. The MUSE machines muc09/10/11 are not condor-enabled.

9.2 Re-processing

The PHOENIX process (as described in section 5.6) is currently being executed on the multi-core blades muc08, muc10, and muc11. The strict separation of a PHOENIX account and a DFOS account is important since they have very different compute and priority profiles: the PHOENIX account can saturate even a 48-core blade, for a long time, but has no high processing priority. A DFOS account has rather moderate processing requirements (in terms of number of jobs, disk space, memory) but has high urgency, due to the closed QC loop with Paranal.

Total processing times for a complete PHOENIX project (pipeline and QC) are in the range of days (for UVES), weeks (XSHOOTER) or months (GIRAFFE). The MUSE tasks can only run in daily slices.