

Operational metrics for the ESO Very Large Telescope. Lessons learned and future steps

F. Primas*^a, S. Marteau^a, L. E. Tacconi-Garman^a, V. Mainieri^a, S. Mysore^a, M. Rejkuba^a, M. Hilker^a, F. Patat^a, M. Sterzik^a, A. Kaufer^b, S. Mieske^b

^aEuropean Southern Observatory Headquarters, Karl-Schwarzschild-Str. 2, D-85748 Garching (Germany); ^bEuropean Southern Observatory Chile, A. de Córdoba 3107, Vitacura, Casilla 19001, Santiago de Chile (Chile)

ABSTRACT

When ESO's Very Large Telescope opened its first dome in April 1999 it was the first ground-based facility to offer to the scientific community access to an 8-10m class telescope with both classical and queue observing. The latter was considered to be the most promising way to ensure the observing flexibility necessary to execute the most demanding scientific programmes under the required, usually very well defined, conditions.

Since then new instruments have become operational and 1st generation ones replaced, filling the 12 VLT foci and feeding the VLT Interferometer and its four Auxiliary Telescopes. Operating efficiently such a broad range of instruments installed and available every night of the year on four 8-metre telescopes offers many challenges. Although it may appear that little has changed since 1999, the underlying VLT operational model has evolved in order to accommodate different requirements from the user community and features of new instruments.

Did it fulfil its original goal and, if so, how well? How did it evolve? What are the lessons learned after more than 15 years of operations? A careful analysis and monitoring of statistics and trends in Phase 1 and Phase 2 has been deployed under the DOME (Dashboard for Operational Metrics at ESO) project. The main goal of DOME is to provide robust metrics that can be followed with time in a user-friendly manner. Here, we summarize the main findings on the handling of service mode observations and present the most recent developments.

Keywords: Operations, metrics, performance indicators, efficiency, observing programmes, completion rates, database queries.

1. INTRODUCTION

Back in 1999, when operations of ESO Very Large Telescope started with the first Unit Telescope (UT1, Antu) and the first two instruments (FORS1 and ISAAC), ESO Paranal Observatory was the first ground-based facility to offer the astronomical community both Visitor (VM) and Service Mode (SM) observing¹. The goal was to allow for enough observing flexibility in order to accommodate the most demanding scientific programmes in terms of temporal and atmospheric constraints.

More than 15 years later, ESO Paranal Observatory operates four Unit Telescopes (Antu, Kueyen, Melipal and Yepun), the VLT Interferometer with its associated Auxiliary Telescopes, the VLT Survey Telescope (VST) and the Visible and Infrared Survey Telescope for Astronomy (VISTA). Furthermore, all 12 foci of the UTs are occupied by instruments. It has thus become more and more challenging to keep track of the overall progress of the several hundred observing runs²

¹ In ESO's terminology, Visitor Mode refers to "classical" observing (i.e. with the PI/astronomer observing at the telescope), whereas Service Mode refers to "queue" observing (i.e. with the PI/astronomer preparing their observations ahead of time and receiving the data as soon as observations were executed).

² A run is defined as the smallest coherent schedulable entity. A programme can include several runs.

approved every semester, disentangling their priorities and conflicts, in terms of similar Right Ascension bins and atmospheric conditions required, even among high-priority programmes.

One of the figures-of-merit most commonly reported by astronomical observatories is the so-called shutter-open time: this is a measure of the time a telescope is on-sky. Because its value gives us a hint of the quality and stability of the site atmospheric conditions as well as on the effectiveness of an observatory maintenance programme (i.e., how much technical downtime there is), it is indicative of the observatory functioning and operability robustness. Clearly, the larger the value, the higher the efficiency of the observatory; still, the shutter-open time does not tell us much about how effectively the time on-sky is actually used.

In fact, the ultimate success of an astronomical observatory relies on several factors: an efficient maintenance of the infrastructure, a realistic scheduling of the telescopes, an optimised usage of the available science time and finally on a healthy scientific productivity of the facility [1]. Here, we will focus mostly on how successfully the available science time is used.

2. THE VLT DATA FLOW SYSTEM

Operating an astronomical site like Paranal, with ten telescopes, one interferometer and 16 (17 as of Oct 2016) instruments requires a well-established infrastructure and operational model. The VLT Data Flow System incorporates all steps, from the formulation of a scientific idea (Call for Proposals, Phase 1) to its execution at the telescope (Phase 2, Quality Control) and final scientific exploitation (data archiving, Phase 3). The official reference is the VLT/VLTI Science Operations Policy³ document approved by ESO's Council (most recent update from 2004). More details can also be found in a number of SPIE and ESO Messenger articles (e.g., [2], [3], [4]).

The model is based on a mix of Service and Visitor Mode observing, which according to the VLT/I Science Operations Policy should allow for at least 50% SM in order to achieve an optimal scientific return. The available science time has different flavours, depending on the type of observing programme that was chosen during proposal submission: normal (standard case), Large (for time requests larger than 100h and using more than just one semester), Monitoring, Calibration, Guaranteed Time, Director's Discretionary Time, Target of Opportunity (ToO) programmes. According to the scientific ranking established by the Observing Programmes Committee (OPC) programmes approved in service mode are further divided in priority groups A, B and C. Rank-class A is reserved for the highest ranked/priority programmes, rank-class C is for filler programmes (those that can be executed under very relaxed constraints). Rank-class B includes all those programmes still favourably graded by the OPC but that did not make it to the first quartile. In comparison, programmes approved in VM do not have a specific priority assigned, but being scheduled at the telescope makes them equivalent to A-class SM programmes.

In more detail, the semester-based operational cycle of ESO's Data Flow System begins with the release of the Call for Proposals that triggers the submission of ~900-1000 observing proposals. Once the OPC has evaluated them based on their scientific merit, the final ranked lists per telescope are ingested in ESO's scheduling tool TaToo. The schedule of all available telescopes (cf. [5] for a detailed review of this process), along with the official notification of time awards to the individual Principal Investigators, represent the closing steps of the Phase 1 process. The Phase 2 process comes next, i.e., the preparation of the SM observations by the successful PIs (or their delegates), followed by the verification and optimization of each programme's observing strategy by the User Support Department. Observing queues are then prepared per instrument and made available to the Observatory. Night-time observations are carried out by the staff at the Observatory, with the aid of a complex ranking optimization which takes into account the scientific priority set by the OPC (and translated by ESO into rank classes A/B/C), as well as the observability of the programme and the relative priorities within each programme set by the users (cf. [6] for a detailed description). If problems arise the affected parties are informed via a ticketing system. After data are taken and their quality assessed, they are transferred in real time to the ESO Archive in Garching and become available to the PI (and their delegate(s)) as they reach the Archive. Smooth operations are guaranteed by real-time calibration completeness and health-checks for the instruments, as well as a careful monitoring of their performances ([7]).

³ Available here <http://www.eso.org/sci/observing/policies/Cou996-rev.pdf>

Despite the fact that this workflow and its underlying infrastructure has been in place since the very start of VLT operations ([2]), a variety of operational/instrumental enhancements have been added during the last 16 years, in order to offer our users' community better tools/interfaces and to keep pace with fast-evolving astronomical research community, techniques and other facilities (cf. [8] for a recent review).

From an operational point of view, the highest-impact change made to the system was probably the introduction of the User Portal in 2007, which offered a unified and user-controllable repository of personal information (via the set-up of personal accounts) for all science and observations related web-based applications and standalone software. Almost at the same time, ESO formalized the handling of programme change requests via a dedicated web-interface. More recently, the upgraded Night Log Tool allowed for the users to subscribe to a tailored distribution of observing logs for their own programmes. Last but not least, the upgrade of our help-desk ticketing system gathered all (operations-related) stakeholders under the same system, making the exchange of information easier, faster and more traceable.

From a more observational point of view, the importance of fast follow-ups of unexpected astronomical events was also recognized quite early on, via the introduction of Rapid Response Mode (RRM) observations. More recently longer term (spanning over multiple semesters) monitoring of sources was enabled through the Monitoring programme type. Finally, the arrival of two survey telescopes on Paranal triggered a deep review of ESO's observing tools in order to support the new operational challenges. This brought us a new version of the Phase 2 Proposal Preparation tool (known to users as P2PP) and of the Observing Tool, introducing the concept of scheduling containers of observations (to better follow time links, concatenations of observations, etc.), a more robust ranking algorithm and a modern Night Log Tool that reports on all operations aspects for each telescope. The upgrade of the tools was then also used to introduce in the system the figure of the delegate, the person entrusted by the official PI of the programme to follow different (or all) phases of the project (Phase 2 and the execution phase, data access via the Archive interface and/or Phase 3 data products submission).

3. KEY PERFORMANCE INDICATORS

In order to accurately answer the questions set up front (cf. the abstract) or even a more basic question like "*What is the best way to measure the overall efficiency of an observatory?*", one needs to carefully choose a set of robust and informative key performance indicators (KPI) and monitor them in time. Simply put, one could for instance correlate what gets scheduled with what gets executed. It seems a straightforward exercise, but one immediately realizes that there exist different references against which this could be computed, i.e. time (hours or nights) or number of programmes. The former is likely more informative for the observatory, the latter for the astronomers in the community. If one then factors in the variety of available instruments (some more challenging than others), of types of programmes and the fact that service mode runs are approved in three different priority groups for a total of 400–500 new runs per semester, it becomes immediately clear how complex this monitoring process is.

Ideally, our ultimate goal is to be able to follow-up what gets scheduled (based on the availability of science time released by the Observatory for a given semester) and then executed. By extracting this information per telescope/instrument/type of programme, both in terms of hours and number of runs, shall provide us with a thorough and detailed picture of our overall performance.

The Dashboard for Operational Metrics at ESO (DOME, [9], [10]) is a project conceived by the User Support Department in order to respond to the need for a systematic monitoring of how the time available for science observations is used. Its implementation relies on the identification of a set of robust benchmarks and the deployment of a system that allows for reproducible, reliable queries and their graphical representation. For the time being, the DOME project has looked only at Service Mode observations, which represent about 70% of the available time on the VLT/VLTI telescopes. Visitor Mode programmes provide completion feedback via observers' end-of-mission reports, making a database query of the selected KPIs more cumbersome. We also decided to restrict our first phase of the project to VLT/VLTI only, thus excluding all data from VST, VISTA and La Silla (VST and VISTA because they are survey telescopes; La Silla because it has been scheduled almost exclusively in VM for several years). Moreover, we further excluded from our analysis ToO/RRM runs and Large Programmes. Completion rates of the former strongly depend on having a suitable target for triggering an observation. In fact, only about 50% of the ToO/RMM time is effectively used, due to lack of suitable targets to trigger on. Completion rates of the Large Programmes are also different from all other

programmes, because some external factors (e.g., selection of targets depending on space mission observations) can affect their completion timescales. They will be looked at in more detail in the future.

We first tackled metrics of the very front-end area (Phase 1), limiting our efforts to the more operational indicators, like pressure, requested vs allocated time, number of unique Principal Investigators, etc, as reported in [9]. We then moved to the more complex area of Phase 2, i.e., the execution of Service Mode programmes. Here, the core question is about the number of runs and/or the amount of scheduled observing time that gets completed in each rank-class. But of course many correlated questions arise, like the completion fraction of A-class runs within their first Period, the total amount of time needed to complete A-class runs, the fraction of A-class observations that is carried over into subsequent periods.

4. DOME – SO FAR

DOME has so far implemented and is monitoring a significant number of metrics. Here we report on just a few examples. Figure 1 (next page) answers the core question, i.e., it represents the overall completion rates of A-/B-/C-class runs/hours over almost 10 years of VLT/I operations (from Period 78⁴ to Period 96). On average, ~80% of all A-class runs considered for this analysis (and time), 45% of B-class runs (34% in time) and 35% of C-class runs (25% in time) could be completed. By inspecting more closely the underlying data, we further note that i) approximately half of the 20% A-class terminated runs reaches a $\geq 50\%$ completion fraction; ii) there seems to be no correlation between the terminated A-class runs and the conditions they requested. The 35% completion fraction of the C-class runs is satisfactory as it refers to filler programmes. Although disappointing, the lower numbers for the B-class runs are not really surprising: B-class runs are usually strongly demanding in terms of constraints (comparable, in fact, to A-class runs) but clash with higher-priority (A-class) runs. Moreover, the schedule was constructed to use 100% of available science time for A and B-class runs, disregarding unpredictable losses such as technical downtime, need to repeat some observations, or unsuitable weather constraints when C-class fillers are executed.

Figure 2 provides one extra dimension, i.e., the time needed to complete A-class runs and shows that the bulk of each run is actually completed in the same period in which the run is first scheduled (P0, cf legend of Figure 2) or during the 1st carryover period (+1P).

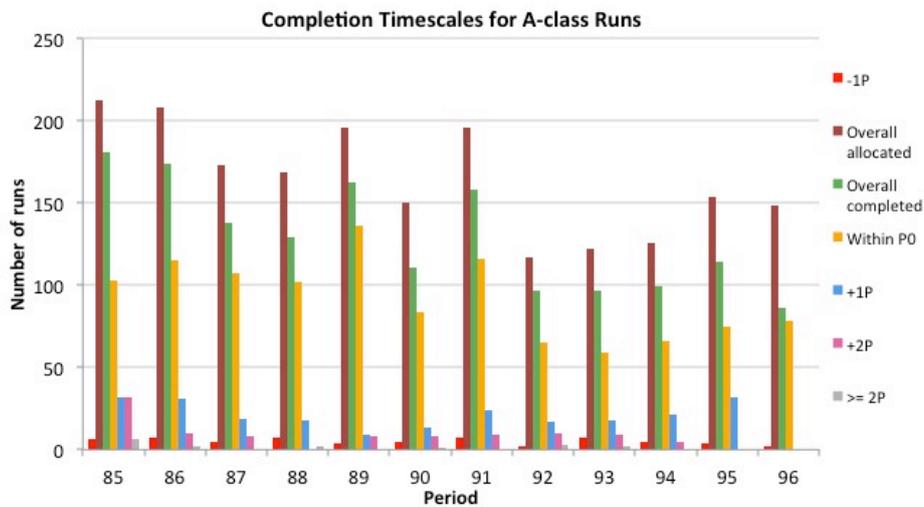


Figure 2 – Completion timescales for A-class runs. This plot shows the number of runs completed in the first period when they were scheduled (this is taken to be P0), in the first/second carryover period (+1P/+2P), longer than +2P and even before the official start of their P0 (referred to as ‘carryunders’ – marked as CU in Figure 4). Overall numbers of

⁴ Data from earlier periods are also available, but due to the introduction of new identifiers in operational databases around P78, earlier data cannot be presented at the same level of detail.

allocated (maroon) and completed (green) runs are shown as reference. Data are available since Period 78; here we show only the last 6 years for graphical quality reasons.

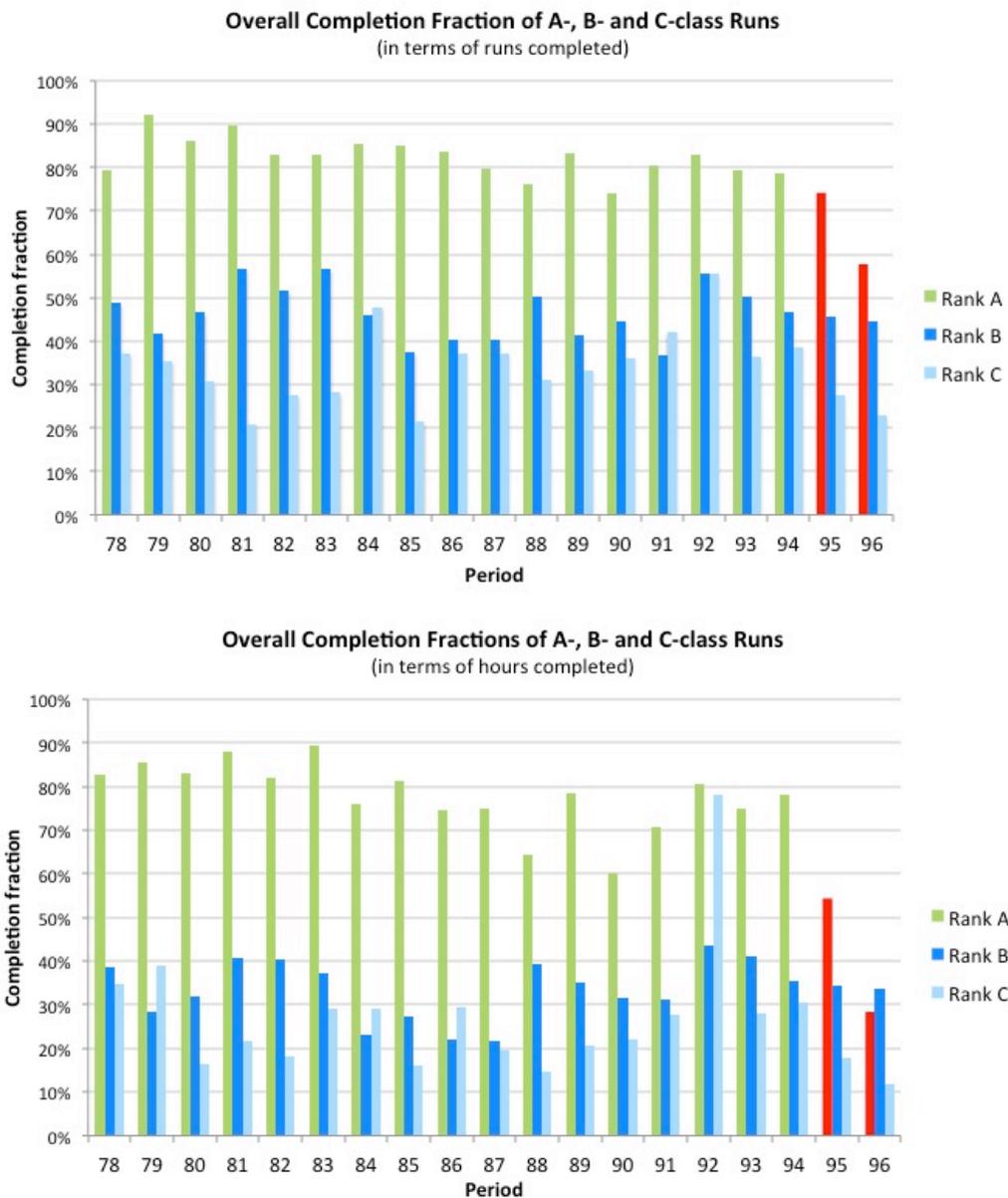


Figure 1 – Overall completion fractions in terms of number of runs (*top*) and number of hours (*bottom*) with respect to the total number of runs that were scheduled in A-/B-/C-classes (*top*) and to the total time that was allocated to A-/B-/C-class runs (*bottom*). For the A-class category, *overall* means that no attention is paid to when the corresponding hours/runs were completed, i.e., hours/runs executed as carryover are included. The different colour of the two rightmost Rank A bars indicates that their statistics are not final yet, because some A-class runs are still in the carryover queues.

One of the proclaimed goals for the implementation of a mixed (Visitor + Service Mode) operational model from the very start of VLT/I operations was to maximise the successful execution of those programmes requiring very demanding observing conditions. While the definition of “demanding observing conditions” may apply to different observing

factors, [10] took as example the seeing constraint and selected all SM runs that have requested, since Period 74, very good seeing conditions. Figure 3 summarises the results for approximately 10 years (from P74 to P93).

For the 0.4'' seeing case, we suffer from very small statistics because the demand for excellent seeing conditions has significantly reduced in recent years: we find only 37 runs in total, half of which have been completed. However, 72% of them were scheduled in B-class, thus impacting their chances of success from the very start. When runs are split between A and B rank-classes, one finds 67% completion rates for the A-class runs and 37% for the B-class runs. Number statistics for 0.6'' seeing runs is more robust with 559 runs in total. Although the overall completion fraction does not change much (54%), some rates increase when grouping the runs according to their rank-class or requested transparency. For A-class runs, we reach an overall completion fraction of 78%, which becomes 75%, 86% and 85% for runs requesting PHO/CLR/THN (photometric/clear/thin cloud) conditions, respectively. For B-class runs we remain instead around 32% for the overall completion fraction, with PHO/THN conditions scoring better (37% / 47% respectively) and CLR slightly worse (28%).

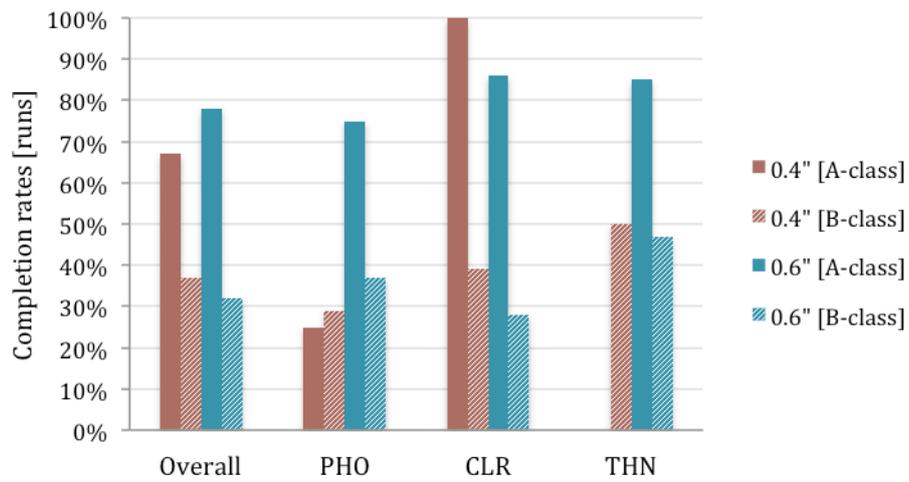


Figure 3 – Completion rates (in number of runs) of the most demanding (in terms of seeing and transparency) programmes scheduled at the VLT since Period 74.

Ideally, one would now like to see all this captured in one graph that takes all the numbers into account at the same time. [10] has attempted this (cf Figure 4, which is a reproduction of their original Figure 10), by connecting the time available for scheduling, the time that was scheduled and the time spent executing service mode observations.

The exercise was made by collapsing all VLT data (but not VLTI, because of the block scheduling constraints) over a not too old 3-year period of time (Periods 85–90) to see how we are doing globally. Such a plot has however one caveat, i.e. it must account also for the VM counterpart, for which we made the assumption that everything that got scheduled was indeed executed, except for weather and technical downtimes. These downtimes were distributed across periods, affecting the time assigned to Service and Visitor Mode according to their allocation percentages. This is also visible in Figure 4 (middle and rightmost columns, orange strips), where the VM execution strip is slightly thinner than the VM scheduled strip.

Before the OPC meets, the Observatory releases a so-called ‘technical time’ schedule (1st column). This includes the time reserved for future planned technical activities (here split into Engineering – magenta – and Commissioning – brown – slots) as well as the time to be likely invested for night-time calibrations (based on previous periods estimates; grey strip at the bottom). What is left is the time available for science observations (light blue) that needs to be scheduled (2nd column). Visitor Mode slots are usually reserved first in the schedule, leaving the remaining time for Service Mode observations (including carryovers from previous periods and a bad-weather statistical provision for SM, based on the actual amount of SM scheduled in the given semester). The actual amount available for scheduling newly approved programmes is usually split evenly between A- and B-class runs, whereas C-class runs (fillers) represent our protective

buffer against idle time and/or bad weather conditions (they are basically scheduled on top). When we reach execution (3rd column), granularity increases because now we are able to trace carryunders (CU), carryovers (CO), idle time, execution losses (time lost because of operational issues), ‘to be repeated’ (i.e., time invested in SM observations that did not fulfil the requested conditions when they were first attempted). At this stage, we are also finally able to use firm numbers for the times invested in technical activities (both engineering and commissioning) and the downtimes (both weather and technical).

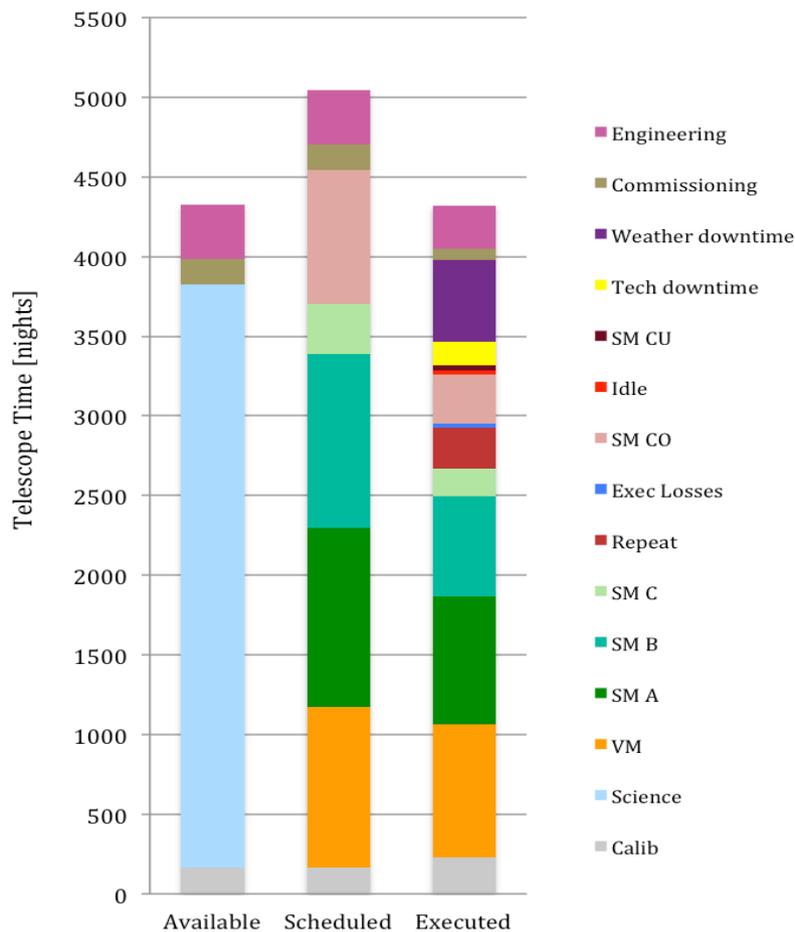


Figure 4 – A global comparison among *Available*, *Scheduled* and *Executed* times, over a 3-years period (April 2010 – March 2013), all VLT Unit Telescopes included.

What do we learn from this plot? Many aspects become apparent, especially by looking in more detail to the rightmost column – Executed – that refers to what happens at the telescope (note that ‘VM, SM_A/B/C/CO/CU’ boxes account only for successfully executed observations, whereas ‘Repeats’ account only for observations that failed and had to be re-queued). First, one can clearly see that the UTs suffer from quite negligible idle time and very little execution losses. Second, it confirms that B-class runs are the most affected ones in terms of completion rates, as already seen in Figure 4. Third, carryovers (CO) and ‘Repeat’ (observations outside users’ specified constraints, cf. legend of Figure 4) have a larger impact on the overall efficiency. They touch upon different aspects of ESO’s operational model: carryovers can be the result of unexpectedly large technical/weather downtimes, or of a too aggressive or sub-optimal scheduling; ‘Repeat’ could be a natural and acceptable consequence of attempting demanding observations. Investigating these further, we found that approximately 40% of the observations classified as ‘Repeat’ (i.e. observations that need to be repeated in order to fulfil the requested constraints) are due to varying seeing conditions and another 20% is due to other variable meteorological conditions. It is worth mentioning that these values strongly vary with the instruments, correlating with

operational complexity and sensitivity to good conditions (e.g. AO instruments tend to have larger amounts of repeated observations).

5. DOME – TODAY

DOME has so far demonstrated a systematic, process-driven approach to publishing a vital, predetermined set of Phase 1 and Phase 2 metrics as a function of period. The requirement of updating these numbers and creating the multitude of associated plots on a semester basis has proven to be rather demanding and time consuming, especially since at the beginning all data tables and plots were manually created in Microsoft Excel. To streamline the process, recently we have been able to improve the workflow, especially in terms of data curation and visualization. Today, the key steps involved in going from data extraction to data visualization to analysis and reporting can thus be summarised as follows:

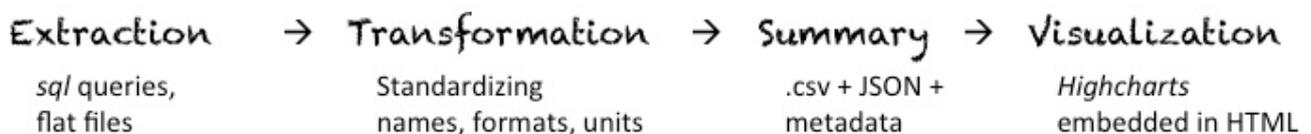
Defining Operational Metrics: In this first step, Key Performance Indicators (KPI) are defined along with a selection of measurable parameters that are required in order to derive the KPI or operational metrics, and to answer other relevant analytical questions. Most of our KPI (as described in previous sections) have not changed.

Extracting Data: This step involves extracting the operational metrics from varied data sources: queries are used to extract data from databases, datasets from flat files like Excel files are then carefully combined. Extraction procedures are recorded in a TWiki for repeatability.

Transforming and Summarizing Data: The transformation of data is necessary in some cases when their format is not uniform across sources, e.g. time units and parameter names. All transformation procedures, data mapping conventions, validation and correction steps, and data filtering rules are recorded in a TWiki. The transformed and cleaned data is stored in a 'Staging Area' in the form of .csv files containing summarized data – a data subset pertaining to a particular topic of analysis e.g. Oversubscription rates, Run Completion fractions, etc.

Visualization and Publishing: The .csv files from the above step are finally converted to *JSON-stat* files (a combination of the pertinent data values and some metadata) for easy parsing by a suite of scripts that generates the required plots, which are rendered using *Highcharts* plotting software. The *Highcharts* plots are interactive, downloadable and printable in different formats, and they expose the dataset directly behind each plot. The plots are then published in well annotated HTML ESO intranet pages within *Adobe Communicate 5*, or *CQ5*, our adopted web content management system.

The initial time investment to set up the extraction through visualization processes has been significant. The effort has however dramatically reduced for subsequent updates performed periodically because, with the following workflow in play, the data added is in increments at each update:



6. DOME – TOMORROW

No matter how streamlined the current workflow is, the central data store is still a set of .csv files, which represents a huge disadvantage; it is impossible to run a query against a repository of .csv files for ad hoc analysis. Also, it is not easy to combine data from other operational groups, for example, if one wants to see how weather downtime affected observation completion rates, there is no direct way to quickly determine this. And, it is nearly impossible to track KPIs across various operational phases e.g. Visitor Mode statistics from allocation through observation completion to publication. An important aspect that is missing from the current setup is the lack of a proper metadata association. This becomes especially important when we want to update the data that is retrieved from disparate sources or explain analysis results. To overcome these challenges, a different DOME infrastructure needs to be designed.

The future DOME project calls for the unification of these monitoring and reporting systems and for the standardization of the underlying data structures for analysis purposes. The immediate operational benefits of such an endeavour are: access to a single source of operations-wide accurate metrics that can guide analysis and strategy, a scalable architecture that provides both a historical perspective and a current snapshot of operational data, dynamic analysis of KPIs throughout the operations cycle, and flexibility to create new metrics on demand and on a short time-scale.

To realize these benefits and more, the most important task at hand in further developing DOME is building a robust database backbone, a central data repository with an extensible architecture so that it is straightforward to assimilate data from other groups in ESO Operations. The next task is to define a universal strategy for tagging all incoming analysis data with the appropriate metadata when inserting into the DOME data store; this makes it possible for traceability to be built into the metadata. The third step is to establish sound extraction and transformation procedures kept under strict version control; the transformation of the data together with the validation scripts make this set of steps a crucial checkpoint for data entering the DOME databases. With this store of high-quality data and with data access that is inherently optimized for ad hoc querying and custom charting scripts (e.g. current DOME webpages), the DOME repository is also highly optimized for the integration with powerful commercially available analytical or visualization software.

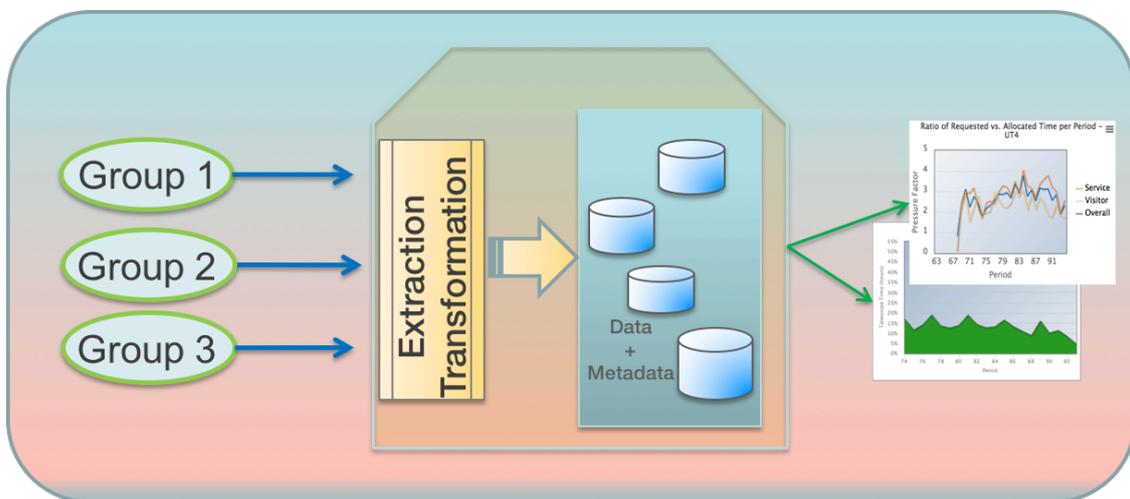


Figure 5 – A graphical representation of the workflow of the new DOME, i.e. relying on a dedicated database.

The ultimate goal of the project is to bring ESO-wide operational metrics into an environment that ensures data accuracy, has storage robustness with straightforward data access, and supports versatile visualization tools, thereby making DOME a true dashboard for analysis across ESO operations.

REFERENCES

- [1] Sterzik, M., Dumas, C., Grothkopf, U., Kaufer, A., Leibundgut, B., et al., “The Scientific Return of VLT Programmes”, ESO Messenger, No. 162, 2, (2015)
- [2] Quinn, P., Albrecht, M., Grosbol, P., Leibundgut, B., Peron, M., Silva, D., “ESO data flow system in operations: closing the data loop“, Proc. SPIE, Vol. 4010, p.56 (2000)
- [3] Silva, D., “Service mode scheduling: a primer for users”, The Messenger, No. 105, 18 (2001)
- [4] Comerón, F., Romaniello, M., Breysacher, J., Silva, D., Mathys, G., “Four years of Service Mode observing at the VLT: performance and user feedback”, The Messenger, No. 113, 32 (2003)
- [5] Patat, F. Hussain, G., “Selecting and scheduling observing programmes at ESO”, in Organizations, People and Strategies in Astronomy 2 (OPSA 2), Ed. A. Heck, p. 231 (2013)

- [6] Bierwirth, T., Szeifert, T., Dorigo, D., Nunes, P., Rejkuba, M., et al., “New Observing Concepts for ESO Survey Telescopes”, Proc. SPIE, Vol. 7737, 77370W (2010)
- [7] Hanuschik, R., Silva D., “VLT quality control and trending services“, ESO Messenger, No. 108, 4 (2002)
- [8] Marteau, S., Hainaut, O., Hau, G., Mieske, S., Patat, N., et al., “The evolution of observing modes at ESO telescopes”, Proc. SPIE, Manuscript 9910-59, in press (2016)
- [9] Primas F., Marteau, S., Tacconi-Garman, L. E., et al., “Operational metrics under one roof“, Proc. SPIE, Vol. 8448, 84480L (2012)
- [10] Primas, F., Tacconi-Garman, L. E., Marteau, S., Mainieri, V., Rejkuba, M., et al., “Fifteen Years of Service Mode Operations: Closing the Loop with the Community”, The Messenger, No. 158, 8 (2014)