# Report of the ESO OPC Working Group

Elias Brinks<sup>1</sup> Bruno Leibundgut<sup>2</sup> Gautier Mathys<sup>2, 3</sup>

- <sup>1</sup> Centre for Astrophysics, University of Hertfordshire, UK
- $^{2}$  ESO
- <sup>3</sup> Joint ALMA Observatory, Chile

With over 1000 proposals per semester for ESO telescopes, the community is facing the problem of fair and robust selection of observing programmes. We report here on a working group to investigate the current selection process and procedures and identify possible improvements. The working group report summarises the current process, based on peer review, that has been in use for many years at ESO and other comparable ground- and space-based observatories, and presents an inventory of the changes and approaches that may be considered to make the process more manageable and less timeconsuming, while at the same time preserving its integrity. The working group presented several recommendations, which have been discussed with ESO committees and analysed by the operational groups at ESO.

## Introduction: Identification of the problem

Over the past two decades, the number of proposals submitted every six months to ESO has been increasing almost linearly, reaching over 1000 proposals for the first time in 2009, followed by a slight decrease over the past three semesters. This large volume of proposals is due to a combination of additional ESO Member States joining, an increase in the number of active astronomers and the many facilities on offer. The increase is correlated with the number of Principal Investigators (Pls). For a description of the current Observing Programmes Committee (OPC) process see the accompanying article by Patat & Hussain (p. 17).

The process employed by ESO to select the best projects and award time on one of its telescopes is a classical peer review system. The ESO Director General (DG) allocates the observing time and customarily follows the scientific recommendations of the OPC and its panels. Very few official complaints, less than half a dozen, are lodged per semester once the outcome is made public, which is a testament to the dedication of the panel and OPC members and to the general acceptance of the process by the community.

The large number of proposals has started to make the review process increasingly onerous, both for astronomers serving on the OPC and its panels, as well as for the Observing Programmes Office (OPO). At the current level of proposals, 13 panels with six members each are needed in order to review the proposals and keep the number of proposals per panel member to a manageable limit. The time spent by panel members refereeing proposals, plus the time involved in the face-to-face meeting, represents a considerable inkind contribution. The current work load on panel and OPC members is very high and any further increase in the number of proposals will stress the current review process even more.

The situation is exacerbated by the substantial oversubscription rates of three to five (see Patat & Hussain, p. 17). Once applications on a telescope/instrument combination reach a tipping point, many good proposals will not make it to the telescope. As there is nothing fundamentally wrong with those proposals, their Principal Investigator (PI) will likely resubmit them with minor changes, adding to the already substantial oversubscription. The large oversubscription leads to frustrated users, as good proposals are not allocated time. Similarly, serving as a panel member can become an exasperating experience as many good proposals end up being marked as not good enough to be scheduled. Even when proposals are basically fine and nothing much can be mentioned to improve them, users expect to receive feedback.

In a sense, ESO is in danger of becoming the victim of its own success. Although OPO can still manage the number of proposals submitted in each semester, it is causing a workload that is becoming increasingly demanding for OPC panel members. This was the reason for setting up the OPC Working Group (OPC-WG). The members of the working group were Jacqueline Bergeron (IAP, France), Elias Brinks (Chair, University of Hertfordshire, UK), Fernando Comerón (ESO), Simon Garrington (Jodrell Bank Observatory, UK), Bruno Leibundgut (ESO), Gautier Mathys (ESO, now at the Joint ALMA Observatory), Michael R. Merrifield (University of Nottingham, UK), I. Neill Reid (Space Telescope Science Institute, USA) and Letizia Stanghellini (National Optical Astronomy Observatory [NOAO], USA). The OPC-WG was charged to examine the current processes and present recommendations for improvements.

# The current OPC process

The ESO process to rank proposals for telescope time is similar in philosophy to that employed at several other groundbased (e.g., NOAO) and space-based (e.g., Hubble Space Telescope [HST] and Spitzer) observatories. There are several proposal categories: Normal, Large, Guaranteed Time (GTO), Calibration and Target of Opportunity (ToO). The category of short proposals has been discontinued (see below). Director's Discretionary Time (DDT) proposals are handled in a separate process and were not discussed by the working group. For definitions of these categories the reader is referred to the ESO web pages<sup>1</sup> or one of the recent ESO Calls for Proposals. The process is divided into Phase 1 (proposal) and Phase 2 (detailed scheduling preparation). Technical feasibility is judged after Phase 1 and before Phase 2.

On average a panel deals with 70–80 proposals, to be reviewed usually within four weeks. Within each panel, all members are required to assign a pre-OPC grade to each of the proposals. Each panel member is first referee on 12–15 proposals which (s)he will need to present in the panel meeting and for which a comment will have to be provided after the meeting.

ESO applies a triage based on the pre-OPC grades, i.e., the lowest 30% of the ranked proposals on the basis of the pre-OPC grades are eliminated from discussion at the meeting. Panel members can still request to discuss a triaged proposal, if they think it merits it. Figure 1



Figure 1. To illustrate the effect that triage has on the selection process, the grades from before and after the panel meetings are shown for the proposals for ESO Period 86. The three horizontal lines show the top 10%, 15% and 25% of the final ranking (lower grades are assigned to better proposals) and are indicative of the oversubscription. The vertical lines

shows a comparison of the effect of triage for proposals from Period 86 with grades before and after the panel meetings. All other (viz. non-triaged) proposals are discussed during the panel meetings. Figure 2 shows plots of the grade and rank of proposals from Period 86, Panel B2 (Galaxies and galactic nuclei: Unresolved and resolved stellar populations) both pre-OPC and after panel discussion, showing generally good agreement, with a significant narrowing of the dispersion following the panel meeting. Scheduling is based on the grades assigned through the panel discussions. The OPC proper mostly focuses on the discussion of LPs and a sub-panel of the OPC deals with the ToO requests based on the grades given by the panels.

Large proposals (LPs) are discussed in joint panels, i.e., within the four broad science categories. A digest of this discussion and recommendation is fed to the OPC, which in turn votes which LPs are accepted. Calibration proposals are few and are only discussed in the OPC. See Patat & Hussain, p. 17 for more details.

## Validity of peer review

Peer review has become the gold standard for ranking proposals. It clearly has illustrate triage at the 30%, 40%, 65% and 80% level. At a 30% triage threshold, it is unlikely that any triaged proposal would have been unfairly excluded from the panel discussion. Proposals with a postmeeting grade of 5.0 were rejected; they include those that were triaged based on their pre-meetings grade, which did not qualify for scheduling.

the trust of the community. In a report on peer review by the Royal Society (1995) it is stated that:

"Peer review is to the running of the scientific enterprise what democracy is to the running of the country. It may not be the most efficient system but it is the least susceptible to corruption. The concept of peer review, in spite of all its difficulties, retains the confidence of most working scientists."

As mentioned, peer review enjoys the trust of the community. Those with experience of peer review, either as member of a time allocation committee or as PI or Co-I of a proposal, will gladly admit that there is a measure of randomness in the process. There is a consensus, largely unsubstantiated (see Figure 2), that the process is repeatable for the top 10-20% of proposals. In other words, the really outstanding proposals are recognised. Similarly, it is argued that there is broad agreement on the bottom guartile. Implicit in this view is that, for all remaining proposals, the outcome is determined more by external circumstances than by intrinsic merit of the proposal.

The WG has found just one paper, by Hodgson (1997, and references therein), that comes closest to a situation where two independent panels ranked the same set of proposals. The author concludes that the results of the two independent panels are correlated, but with a considerable random variation. There was no perceptible decrease in this variation near the top or bottom end of the distribution (see also the example in Figure 2).

Clearly, it is important for this or similar tests of the classical peer review process to be performed in order to quantify its reproducibility. The working group suggested that ESO could use the fact that each scientific category has several panels to test the reproducibility of the rankings by submitting a subset of the proposals to more than one panel. This would be done "blind" for the panel members and would therefore provide ESO with a means to measure to what extent the results are reproducible. The additional workload per referee was considered to be tolerable.

#### Evolutionary changes

The basic criteria for a better system are fairly straightforward: an ideal system produces a better quality outcome at a lower cost in terms of community investment. Many evolutionary changes were considered: to the way the review panels work, to the frequency at which the review is carried out, and to ways to restrict the number of proposals. In its deliberations, the OPC-WG was guided by the principle that for any proposed change, the process should:

- 1. be driven by scientific excellence;
- 2. be fair (as perceived by the community);
- 3. be robust against conflicts of interest;
- 4. be robust against abuse/cheating;
- 5. preserve confidentiality;
- result in an acceptable/manageable workload for the community;
- 7. provide useful feedback;
- 8. be adaptable to changing circumstances;
- 9. be as manageable for ESO as the current system.

The OPC-WG recommended a rationalisation of the large variety of proposal types. The OPC-WG supports the combination of the application forms for Normal and Short proposals and restricting the length of the scientific justification to one page. This has already been implemented from Period 87 (2010).



Figure 2. Average grade and standard deviation against run rank is shown based on the preliminary (left) and final (right) grades of the members of Panel B2 in Period 86.

Among the many changes considered, only those where the advantages clearly outweigh the disadvantages are presented here. A first suggestion was to split the selection according to telescope or groups of telescopes. For example, one could imagine a review process exclusively for the VLT Unit Telescopes (UTs) on Paranal (including VLTI), and others for the 4-metre telescopes on La Silla, and a separate one for the Atacama Pathfinder Experiment (APEX). Few projects request the use of telescopes on more than one site and the large majority are for Paranal (see Patat & Hussain, p. 17), and hence there is not much to be gained with such a procedure.

To decrease the load one could increase the panel size. To avoid even larger OPC meetings and contain the costs, one could imagine a system in which only half the panel members were assigned as primary referees and participated in the meeting while the other half provided only written comments and pre-OPC grades. A clear drawback is that this would require more or less doubling the number of referees.

Changing the frequency of OPC meetings was also discussed. A one-year cycle would not result in a doubling of the number of proposals, but rather result in an increase estimated at approximately 30%, because many proposals are resubmissions (as a result of the substantial oversubscription rate) and quite a few exploit the right ascension overlap regions between semesters. No obvious detrimental impact on the science is expected.

A reduction in the number of submitted proposals could also be achieved through the creation of a new category of monitoring proposals. These are for projects which in any given period lay only a limited claim on telescope resources, but run for many periods, if not years, and therefore guarantee the success of longterm monitoring campaigns. The total time would in general be less than for LPs and the number of semesters might go beyond the LP limit. Because LPs and monitoring proposals are less time critical, they could be discussed in alternating periods. The proposed change would thus be to schedule LPs, monitoring proposals, and proposals for La Silla on a yearly basis, the LPs for example in even-numbered periods, the other two sets in the odd-numbered periods.

The WG also looked into how the number of proposals submitted could be influenced. One way of doing this is to base future allocations on past performance. It should be remembered that there might be a good reason why data are not published (e.g., lack of adequate data, disappointing result, rejection by a journal, etc.). Sometimes at the Phase 2 stage, PIs realise that their original time estimate was wrong (usually too low and frequently due to underestimating the time required for overheads). This then forces them to either reduce the time on target or reduce the number of targets. This could render the original goals of the

proposal unattainable and, if the OPC knew this, could result in the proposal being downgraded.

## Revolutionary changes

The working group considered fundamental changes to the classical peer review system with potentially major effects for the community or ESO.

One could imagine setting a substantial lower limit on the number of hours or nights per semester for a proposal to be considered. This would force the community to form consortia and, depending on the lower limit set, would result in fewer proposals. There would then be no need for panels, the decision process being limited to the OPC. This would seriously reduce the workload. Also, proposals would likely attack "big questions" and propose coherent work packages. In such a scenario ESO would be outsourcing a large fraction of the decisionmaking process to the community. The OPC-WG saw many disadvantages to such an approach:

- 1. Powerful individuals or groups could monopolise certain areas.
- Instead of the decision on the merit of a proposal being taken in a nonpartisan way by referees who have no conflict of interest, decisions are effectively devolved to the consortia. Depending on how a consortium is organised and decisions taken, excellent science might lose out.
- 3. Small time requests for prime science will not be eligible unless repackaged as part of a larger application by a consortium.

Telescopes and Instrumentation

4. Unique and singular objects (e.g., Galactic Centre, SN 1987A, etc.) will be at a disadvantage.

Several observatories have opted for this mode of operation, usually for a substantial fraction of available observing time. Examples are e-MERLIN Legacy projects, Herschel Open Time Key projects, James Clerk Maxwell Telescope (JCMT) Legacy proposals, Hubble Space Telescope (HST) Legacy and Treasury proposals, Spitzer Legacy proposals and proposals for future instruments like ASKAP and MeerKAT, at least for the initial years of their operation. To some extent ESO has followed a similar route by implementing public surveys on the survey telescopes VISTA and the VST, and more recently also at the New Technology Telecope (NTT; PESSTO) and with the Gaia-ESO survey with UVES and FLAMES on UT2. ESO might consider extending such surveys with La Silla and Paranal telescopes when the European Extremely Large Telescope (E-ELT) is built. It was not clear to the OPC-WG how much is gained by having a mix of these megarequests and normal proposals.

A radical departure from classical peer review is the method of distributed peer review as described by Merrifield & Saari (2009). For reasons inherent to that method, it is not possible to extract what result distributed peer review would have produced, or how well it compares with classical peer review, by analysing the outcome of the latter. In order to do a proper assessment, a full trial would need to be designed.

#### **OPC Working Group conclusions**

The conclusions of the OPC-WG are as follows:

- There are small changes that can be applied to the current system that would reduce the workload without affecting in any major way the widely accepted peer review method. These are:
  - Implement a new category of monitoring proposals; these proposals can be long-term, but ask for modest amounts of time per semester. Open up the Call for this type of proposal

once a year, e.g., during oddnumbered Periods;

- Review La Silla proposals only once a year, also in odd-numbered Periods;
- Limit the Call for Proposals for Large Proposals to once per year as well, for even-numbered Periods.
- A more substantial change would be to change the frequency of the Call for Proposals to once per year. Although the number of proposals will likely increase, it will not double. The reduction in agility could, in cases where this is justified, be made up for by DDT.
- Most other considered changes, either in the way panels deal with the proposals, or by attempting to limit the number of proposals, carry disadvantages that outweigh their advantages.
- Peer review is broadly supported by the community. There is a preconception that top proposals will be recognised with little dispersion in their grades or ranking. Likewise with poor proposals. In other words, peer review is thought to deliver reproducible results. It is accepted that the grades or ranking of all remaining proposals is less clear cut and that some considerable dispersion might be expected. The OPC-WG tried to confirm this picture but came to recognise that precious little in terms of data exists that underpins the validity of this model of peer review. If anything, the few tests available to the WG showed that a large dispersion exists even for proposals in the top and bottom quartiles.
- The WG recommends that the reproducibility of peer review is further investigated. One method that was suggested would be that some 10% of the proposals in any Category (A, B, C and D), be seen by more than one panel. For those proposals that are seen by two panels, it is decided in advance which grade will be taken for the final ranking. This would be done "blind" for the panel members and would therefore provide ESO with a means to measure to what extent the results are reproducible.
- The WG is concerned about the fact that there are substantial differences between the Phase 1 and Phase 2 time requests. Although the amount of time on the telescope in the end is broadly compatible with that of the Phase 1

request, this implies that either fewer targets are observed, or that less time is spent per target. Either of these outcomes could cause a panel to give a highly ranked proposal a lower grade, which would mean that it should not have been scheduled in the first place. The WG recommends that if the time requests differ by more than what can be considered reasonable (to be judged alongside the technical feasibility), the proposal should not be scheduled. - Discussion on how the E-ELT can be implemented in the ESO proposal selection process was limited. The E-ELT will likely match or exceed the oversubscription rate of the most popular VLT UT. E-ELT proposals could simply be accommodated within the existing process provided that the science policies for the E-ELT and other ESO facilities remain aligned. Alternatively the E-ELT could be made available to consortia only, setting a lower limit to the number of nights that can be bid for and letting the community organise themselves into larger collaborations. In the opinion of the WG not all telescope time should be allocated in that way as there is a risk that the field becomes dominated by a small number of Pls, blocking access to individuals or small groups who, for whatever reason, lack access to these larger consortia. Also, small projects, in terms of telescope time, with a potentially high impact would never be scheduled on their own merit but would have to be incorporated within a larger time request.

- The OPC-WG ran a test with the distributed review method proposed by Merrifield and Saari. The test was useful in pointing out several issues that the user community would probably raise if it were to be introduced. Most obviously, the user community would need to be thoroughly educated about this method. Also, sufficient trust would need to be built up before this, or any other revolutionary approach, will be accepted. The OPC-WG encourages ESO to perform a larger scale experiment, more closely linked to the observing proposal process, to explore distributed peer review as an alternative method.

Given the current paucity of data on the effectiveness of peer review, the changing landscape, and the unpredictable effects of alterations to such a complex process, it is important to recognise that this report cannot represent a final conclusion on the right way forward.

#### Follow-up

The report by the working group was presented to the OPC, the Scientific Technical Committee (STC) and the Users Committee (UC). There was general support for the introduction of monitoring programmes and ESO has started developing this type of observing programme to begin in Period 92. The OPC discussed whether ESO should move to a one-year cycle for LPs and proposals for La Silla. There was a small majority for a one-year cycle for La Silla proposals, but clear discomfort about evaluating LPs only once per year. The UC also recommended ESO to stay with the half-yearly cadence for proposal submissions of all telescopes and types. ESO will hence not change the frequency of calls for observing proposals.

The ESO operations groups evaluated the impact of a move to a one-year cycle for all proposals in the current operation setup. They concluded that within the current support scheme a significant increase in effort would be needed. The work load would very strongly peak around the scheduling and Phase 2 as more proposals would have to be processed within the same time span to provide a full schedule at the beginning of the period. The impact for the execution of the observations is not as significant, but the cycle between proposal submission and data delivery could increase to more than one year, which was considered by the OPC-WG and the OPC itself as a possible disadvantage. Offsetting these delays for scientific programmes by an increased Director's Discretionary Time allocation was not considered sufficient. For the time being no change regarding the ESO periods is planned.

The OPC was strongly opposed to adding to the work load by increasing the evaluation of proposals. In particular, it did not like the idea of a parallel evaluation of some proposals to establish the validity of the peer review process. The position by ESO that the proposal selection process should not be used for a "social experiment" to investigate the effectiveness of peer review was supported by the OPC as well. The ESO database should allow an investigation into this question post facto and ESO should analyse the available data. The evaluation of the LPs is done in all subpanels in parallel and this could be used as a (limited) dataset for such an investigation. The OPC requested that no panel member should be asked to referee more than 70 proposals in a given selection round.

The OPC-WG was already very sceptical about forcing the community into large collaborations to reduce the number of submitted proposals. The ESO committees (OPC, STC and UC) concurred with this assessment and this will not be regarded beyond the current scheme of Public Surveys. ESO has implemented these surveys to make optimal use of the two survey telescopes and also to allow the community to establish a leading position in a specific subfield. The recent workshop on surveys (see the report by Rejkuba & Arnaboldi, p. 67) has demonstrated the success of this approach. Such surveys are followed by a special review panel to guarantee that the large investment of observing time results in a corresponding return to the community. ESO and its community will need to assess in a few years the balance between regular proposals, requesting a few nights, compared to surveys, with hundreds of nights. The OPC-WG will have a role in this assessment as well.

Changes between Phase 1 and Phase 2 for a given proposal will be handled more strictly in the future. The procedures are available and will be enforced. There is a clear imbalance between the numbers of proposals in the different scientific subcategories. A measure to better distribute the load between the different panels is to redefine the subcategories, with the goal of reaching a more even distribution across the OPC panels. This investigation has started and will be presented to the ESO committees in due course. In the mean time, the number of proposals has been decreasing over the past few periods to currently about 900 proposals per semester. This has the positive effect that the load on the OPC and its panels has decreased (by about 10%).

The OPC-WG suggested that it would be beneficial to re-visit the proposal selection process regularly, perhaps every few years. The composition of the OPC-WG was originally chosen to combine the expertise residing within several international observatories with high proposal pressures. The exchanges within the OPC-WG were very informative and allowed the members to have a fresh look at the processes involved in selecting the best science for a given facility. ALMA has just finished the proposal selection for Cycle 1 and already is dealing with more than 1000 proposals per cycle. Other observatories are faced with this problem as well and all working group members were interested in returning to these questions in a few years.

#### References

- Hodgson, C. 1997, Journal of Clinical Epidemiology, 50, 1189
- Merrifield, M. R. & Saari, D. G. 2009, Astron & Geophys, 50, 4.16 (arXiv:0906.1943)
- Patat, F. & Hussain, G. A. J. 2012, in *Organizations People and Strategies in Astronomy* — Vol. 2 (OPSA2), ed. A. Heck, in press
- Royal Society, 1995, Peer Review An assessment of recent developments.

#### Links

<sup>1</sup> ESO telescope time allocation:

http://www.eso.org/sci/observing/teles-alloc.html