The Marine National Facility Proposal Assessment Criteria Review

February 2020

Bruce Mapstone

Executive Summary

Australia's Marine National Facility (MNF) in 2018 received funding to support full-time at-sea operations of the national blue-water research vessel *Investigator*. *Investigator* now is funded for 300 days at sea, a significant increase from the 180 sea-days funded previously. The MNF has reviewed its strategy and the processes by which use of *Investigator* is allocated in response to that change. The MNF has moved to an approach of allocating support in 5 Streams to service research arising from (1) recognised national priorities, (2) key questions of science, (3) technology innovation needs, (4) user-funded demand, and (5) publicly funded programs that require MNF facilities.

The new operating environment for the MNF precipitated questions about the existing proposal assessment processes by which proposals for MNF support were judged. Existing assessment processes had served the MNF and research communities well for several decades but they might not remain fit-for-purpose in the new full-funded, full-time, Streamed operations of the MNF. This review is to inform the framing of assessment processes appropriate for the new MNF Strategy 2030.

The review was to recommend appropriate bases of 'merit' by which access to, and support from, the MNF should be awarded in each Stream. Key considerations were that the MNF is a publicly funded Landmark National Research Facility and its use should be both in the national interest and for support of high-quality 'research' (broadly defined).

Research Quality and (delivery of) National Benefit were two principal assessment criteria for MNF support under the existing MNF access allocation framework. Those concepts remain appropriate and central to the future operations of the MNF and have been recommended as the two primary pillars, or principles, of 'merit' by which access to the MNF, particularly *Investigator*, should be decided. The term 'National Benefit', however, often caused uncertainty with applicants and reviewers of proposals and arguably might be too narrowly focussed for all Streams in the new MNF Strategy. It is recommended here that a more general principle of Research Benefit be adopted that allows for benefit to be claimed from research across a broader spectrum, as appropriate to the focus of the five new Streams of allocation. It remains important that those Research Benefits demonstrably should be in Australia's national interest. It is recommended that, with this change, the same merit principles of Research Quality and Research Benefit should apply to all 5 Streams. Weighting these merit principles differently for different Streams is not recommended, provided there is adequate Stream-specific focus in the assessment of proposals' claims against relevant criteria.

The review also was to draft specific criteria for assessment of proposals for MNF support, with consideration to the requirements of the five access Streams. A preference was clear for relatively few criteria and for criteria that could be applied to multiple streams, if possible. It is recommended that a common set of criteria be used to assess proposals for support from all Streams but that Stream-specific guidance for preparation and assessment of proposals be used to tailor appropriately the ways in which criteria are addressed in each Stream. Four assessment criteria are recommended for each of the merit principles (Research Quality and Research Benefit). Stream-specific guidance is provided to assist applicants address the criteria appropriately in the context of their nominated Stream. Stream-specific guidance to assessors of proposals, together with directions for scoring proposals, also is provided to ensure applications are judged appropriately in each Stream but within a framework that is consistent across Streams. Such an approach will be conceptually, operationally, and administratively simpler and more equitable than applying disparate, Stream-specific principles of merit or proposal assessment criteria. Equitable and comparable analysis and presentation of assessment results also would be more difficult with Stream-specific merit principles or assessment criteria.

The final task for this review was to *recommend a method or methods of analysing and presenting* [proposal assessment] *scores to provide transparency of process and focus discussion of results, both within assessment subcommittees and at the Steering Committee, within Streams and across all Streams.* An analytical method is recommended for all Streams that provides for staged, systematic, and internally consistent consideration of proposals' strengths or weaknesses against the merit principles and their assessment criteria at all stages of the assessment process, from initial scoring to consideration by the MNF Steering Committee. Criteria for rejection of proposals as 'unsupportable' or eligible for only 'conditional support' also are recommended, based on the recommended scoring and analysis approach.

Contents

Introduction

The Marine National Facility (MNF) has been funded by the Australian Government and operated by the CSIRO since 1984. A blue-water research vessel has been, and remains, the primary capability of the MNF. A range of additional observation and sampling tools are provided for use on the vessel, usually at no additional cost to researchers. The MNF also provides extensive technical support services for research on the MNF vessel and, with the CSIRO, data management and delivery services for data collected during voyages. The MNF since 2014 has operated the Research Vessel (RV) *Investigator*, a vessel purpose-built for multi-disciplinary marine and atmospheric research with accommodation for 40 researchers on voyages of up to 60 days and a steaming range of 10,000 nautical miles .

Investigator was designed for 300 days of operational sea-time per year and the Federal Government approved funding for such full-time operation in its 2018 budget. That decision initiated a material increase in the operational capacity of the MNF, which previously had been supported only for up to 180 sea-days of full-funded 'granted' voyages. Another 120 days of potential sea-time was available for 'user funded voyages', though that time was never fully subscribed by part-paying or full-charter users.

The support for full-time operation of *Investigator* precipitated questions about the appropriateness of existing approaches to allocating sea-time that was fully supported by the MNF. The MNF in 2018–19 therefore commissioned a review of its MNF Access Framework. The frameworks for operation of the MNF and allocation of MNF facilities have been revised in response to that review and the new approach is explained in *MNF 2030: A 10-year Strategy for the Marine National Facility to guide the use of Australia's dedicated blue-water research facility*. The updated mission of the MNF is:

> *We facilitate safe, efficient and excellent ocean and atmospheric research that is well aligned with national priorities and addresses Australia's grand challenges for society, the economy and the environment.*

A key change to the way the MNF allocates sea-time and technical support is the introduction of "Streams" of allocation that provide more targeted and strategically driven foci for use of the facilities. The five Streams introduced are:

Stream 1 — Priorities-driven research, for proposals explicitly addressing national priorities for blue-water research;

Stream 2 — Science-driven research, for proposals with the primary purpose of advancing scientific knowledge that do not directly address national priorities for blue-water research;

Stream 3 — Technology and Innovation projects, for proposals to undertake high quality development and testing of innovative technology;

Stream 4 — User-funded research, for proposals that are in the national interest and rely on RV Investigator's specific capabilities; and

Stream 5 — Strategic Partnerships with national publicly funded programs/institutions that rely on regular access to MNF capabilities to support data and sample collection.

Revision of the MNF strategy and the move to Streamed allocation has generated reconsideration of the appropriateness of the existing criteria by which projects are selected for use of MNF facilities, especially *Investigator.* Streamed allocation of *Investigator* sea time also has stimulated review of the balance of weight given to research excellence and national benefit that had been applied to date and whether Stream-specific proposal assessment criteria are required.

This review was commissioned to seek criteria that are fit-for-purpose for fair assessment of proposals within and across the five Streams of MNF future operations. The *Scope of Work* for the review is:

For each of the five (5) streams of access to the Marine National Facility (MNF) ...

a. Review definitions of the basis of 'merit' for each of the Streams, potentially resulting in one or a couple of definitions that apply across multiple Streams (the fewer the better);

b. Draft clear and unambiguous criteria by which proposals in each Stream would be assessed against the relevant merit principles (as few criteria as possible to meet the need for each Stream and, preferably, apply across multiple Streams);

c. Craft guidance for assessors to use when scoring proposals against the criteria (that is unambiguous and encourages critical scoring); and

d. Recommend a method or methods of analysing and presenting scores to provide transparency of process and focus discussion of results, both within assessment subcommittees and at the Steering Committee, within Streams and across all Streams.

List of Recommendations

Listing recommendations alone, excised from the narrative that gives each its context, risks coarse interpretation of them, with neither nuance nor perspective. I nevertheless provide such a list here for those who want to 'see the punch-lines' before wading through the detail. Recommendations are listed under the high-level topic to which they relate. I also provide the page number where the most relevant narrative for each recommendation begins to help readers delve into the background selectively.

Recommendation	Page
The MNF weight assessment criteria equally within each merit principle for analysis of assessment results.	35
The MNF adopt the recommended methods for analysis of assessment results and presentation to the Research Advisory Committee, National Benefit Assessment Panel, and MNF Steering Committee.	44
The MNF Steering Committee allocate ship-time with explicit reference to proposals' assessments against both Research Quality and Research Benefit rather than based on the ranks of a combined (RQ+RB) score.	44
The MNF rejects outright any proposal that fails to meet minimum standards judged against individual assessment criteria rather than higher-level, aggregated ratings.	45
8. The minimum standard of assessment scores for potential MNF support be that a proposal has no score of less than 2 out of 10 against any criterion and no more than one criterion scored 2 or 3 out of 10.	45
The MNF rejects outright any proposal that would fail the minimum standard of criterion-level scoring for half or more of the assessors and review carefully any proposal that would be considered a 'fail' by fewer than half of assessors.	46
Any projects with scores of 4 or less, averaged over assessors, for any criterion (criteria) be granted access, if available, conditional on satisfactory resolution of the issues that precipitated the low score(s).	46

Findings and Recommendations

Preamble

Reframing the MNF allocation framework into five Streams of access that align with different 'types' of work clarifies the scope of activities that the MNF Steering Committee considers appropriate for the MNF. An underlying principle of all such work is that the MNF is a national *research* facility, and it is reasonable to infer that all activities envisaged for usual MNF support fit with some definition of research. I use 'research' hereafter in a broad sense to mean data-gathering activities intended to generate new information, including (but not limited to) investigating fundamental processes, discovery, exploration, monitoring, mapping, or technology development. My use of 'research' therefore should not be construed to discriminate among the activities envisaged for each Stream of MNF access but as a convenient catch-all term for all such activities considered acceptable for deployment of a national research facility.

An issue related to proposal assessment and support allocation is the justification for a project to be considered at all for MNF support. It is stated clearly for Streams 4 and 5, for example, that proposals will be considered where projects "... *rely on RV Investigator's specific capabilities ...*" (Stream 4) or " *... rely on regular access to MNF capabilities ...".* It might be argued that these sorts of 'entry criteria' should apply for proposals in all Streams because it would be inefficient to deploy *Investigator* and MNF support to projects that could be done from alternative platforms available in Australia. Such an approach is obstructed substantially by the absence of any mechanisms for coordinating, or harmonising, allocation of Australian marine research facilities, however, meaning that proponents are required to apply independently to multiple suitable facilities to access any one of them. A proposal for work that could be done on either *Investigator* or *Nuyina*, for example, technically could be excluded from time on each vessel because it could be done on another platform, and so excluded from both assessment processes.

An appropriate compromise at this stage, therefore, would be to require a brief statement from each proponent explaining why MNF support, and time on *Investigator* in particular, is essential for the proposed work but without applying an exclusion rule if the work might be able to be done elsewhere. The MNF has anticipated such an approach by making it clear that proposals are for MNF support on a suitable fit-for-purpose platform, not necessarily *Investigator*, and reserving the option to deploy other ships if available.

The circumstances under which such alternative deployments, or *a priori* rejection of a proposal, might occur are likely to be Stream-specific and based primarily on MNF policy and operational considerations. That judgment would seem most appropriately made by the MNF Executive, therefore, and recommended to the Steering Committee, rather than being sought from proposal reviewers or assessment panels. It seems reasonable to ask reviewers to comment on whether ship-time has been requested appropriately for a project but it is not really a reviewer's role to assess whether other facilities might be available in Australia and logistically could or should be accessed for the work. I therefore have not emphasised justifying the use of *Investigator* in assessment criteria for any Streams but recommend a transparent process for judging the appropriateness of *Investigator* for each proposal be developed and publicised by the MNF.

The following material has not been drafted specifically to fit with the existing MNF application (proposal) forms. Doing so might have imposed unnecessary or perverse constraints on framing the sort of information required to address Stream strategies or robust assessment processes. It would be useful to review the application documents once a final set of Stream-specific guidelines and assessment processes are adopted to ensure an efficient match between the information that is sought by the MNF and the form in which it is to be provided by applicants. Close alignment between application forms and assessment criteria should be considered. Such an alignment will help focus applicants on necessary information and facilitate efficient review of proposals by assessors.

Interpretation of Merit

a. Review definitions of the basis of 'merit' for each of the Streams, potentially resulting in one or a couple of definitions that apply across multiple Streams (the fewer the better).

The bulk of this review is about the criteria by which proposals for MNF support would be assessed. That assessment will be done according to what constitutes merit. It is important, therefore, to recognise a distinction between the qualitative principles that define 'merit' and the specific criteria that then are appropriate to assessing how well those principles of merit are met. The discussion of whether different definitions of merit should apply to different Streams is one of principle, therefore, rather than detail, and is approached at this stage without concern for the details of criteria by which merit will be assessed.

What should be considered in defining 'merit' for support from the MNF, and sea-time on *Investigator* in particular, was considered at the outset of the Review of the MNF Access Framework[1]. The review included a recommendation that "*The MNF formally adopt an application of 'merit' comprising both research quality and delivery of national benefit for determining access to the Marine National Facility*", which seems to have been accepted in-principle by the MNF Steering Committee (SC). The focus of this section of this review, therefore, is on whether that position remains appropriate to each of the new access Streams and whether Stream-specific interpretations of 'merit' are necessary.

There is little likelihood that any definition of 'merit' would ignore the technical quality of research proposed for MNF support. The previous review reported that such a position would be at odds with expectations of both MNF stakeholders and government policy. Another key consideration in advising how 'merit' should be defined is that the MNF is one of Australia's Landmark National Research Facilities and, since 2018, is fully funded by the public, through the Federal Government. Those circumstances suggest that any allocation of MNF resources should have a reasonable prospect of generating some 'return on investment' to the nation. It is important that use of the MNF can be justified credibly as being in the national interest or providing some identifiable benefit to Australia or Australians flowing from the research. That position also was supported by most stakeholders consulted for the previous review and judged to be consistent with government policy guidelines, either explicitly or implicitly. It is a reasonable conclusion, therefore, that the 'merit' of all applications to use the MNF, through any of the five access Streams, should consider both the quality of the research proposed and the prospect that it will deliver some form of [national] benefit, notwithstanding that the form of benefit might differ among Streams.

Recommendation

1. The MNF retain an overarching principle that merit for MNF support should be based on the two primary principles that research (a) should be of high quality and (b) have a reasonable prospect of delivering some form of benefit to or within Australia, directly or indirectly[2].

Merit Assessment for All Streams?

An implicit feature of 'merit' assessment is verification that proposed research is 'fit-for-purpose'. The foci of the five MNF access Streams clearly stipulate different contexts for which different 'end-user' constituencies are targeted. Streams 1, 2, and 3 involve 'granting' MNF resources to support applicants who claim to be proposing research relevant to the relevant Stream strategy, but otherwise have no evidence *prima facie* that their work is fit-for-purpose. Consideration for MNF support therefore would be expected to include some form of assessment by the MNF of 'fitness for purpose', at least.

Streams 4 and 5 will involve either material investment by an external entity, possibly covering the full cost of a voyage by *Investigator*, or programs of work that already have been approved as sufficient priority to warrant public funding. Some might question, therefore, whether projects supported through Streams 4 or 5 should be subject to the same or similar assessment processes as those supported through Streams 1,2, or 3.

It could be argued that work proposed by users who want to fund the use of *Investigator*, for example, will be well-considered by those underwriting it and its fitness-for-purpose in general, and research quality or

[1] Mapstone, B.D. 2019. *The Marine National Facility Access Framework Review*, 57pp.

[2] These 'principles' of merit were cast as selection criteria in the former MNF Access Allocation Framework. They are 'elevated' to be overarching principles here, for which assessment criteria will be drafted and used to judge how well proposals for access meet each principle.

research benefit in particular, should not be the business of the MNF or its assessment panels. Proponents of proposals for partnership with the MNF through Stream 5 similarly might argue that the merit of their work has been endorsed already by virtue of its approved public funding and so shouldn't be subject to (re)assessment by the MNF. It might be argued, therefore, that the role of the MNF in these Streams should be focussed principally on whether the proposed work *depends* on MNF facilities or could be done with other resources and then negotiating whether and, if so, under what conditions *Investigator* and other facilities will be available to Stream 4 or 5 proponents.

An alternative argument would be that the MNF is responsible for the *appropriate* use of *Investigator* and associated publicly-funded facilities and, accordingly, should verify that Stream 4 or 5 proposals fit with the MNF mission and strategy. Either part-funding in Stream 4 or partnership in Stream 5 still involves some, probably considerable, investment by the MNF and, potentially, opportunity costs to other Streams. Those circumstances imply a responsibility for the MNF to verify transparently and defensibly that its commitments are justified. There also is a clear reputation risk of the MNF supporting work that is seen to be of poor quality or not clearly in the national interest, notwithstanding who has funded the work. There accordingly should be some mechanism for judging whether Stream 4 or 5 work is of sufficient merit, in both quality and benefit, to justify MNF support, whether by 'charter' or partnership. There likely will be contractual and logistic arrangements for MNF support to Streams 4 and 5 that will differ from those for work in Streams 1, 2, and, probably, 3 but those other Stream-specific negotiations do not obviate the need for proposal assessment.

Public accountability for spending the money allocated to the MNF is relevant here. Justifying future funding of the Marine National Facility is likely to hinge at least partly on a convincing case that public money spent on the MNF is in the national interest, supported by evidence that that money has been and will be used to support research that benefits Australia. That accountability will be more robust and convincing if it can be demonstrated by access allocation processes that the national interest is central to judgments about all the research the MNF supports, through whichever Stream.

I therefore conclude that it is most appropriate that all applications for support by the MNF, through any Stream, should be assessed formally against MNF merit principles. All supported work should be expected to be, and demonstrated through rigorous assessment processes to be, of sufficient quality and make sufficient contribution to Australia's national interest, either directly or indirectly, to justify investment of MNF resources.

Recommendation

2. All applications for MNF support, through any of the 5 Streams, should be subject to formal and transparent Merit Assessment.

Stream-specific Merit

Quality of Research

Fitness-for-purpose applied to research quality arguably should be focussed on the research *per se*, as opposed to what it is intended to deliver beyond new disciplinary knowledge. It seems likely from that perspective that fitness-for-purpose would have common core principles irrespective of the nature of the research. The research should have a robust disciplinary rationale, clear and appropriate objectives, sound methods, demonstrable feasibility, the prospect of informative results, and be being done by people with appropriate expertise. These components of *research quality* should be satisfied for work in each Stream if that work is to justify use of a national research facility. There clearly will be some variation or nuance in research context among Streams but those Stream-specific contexts could, and should, be articulated in the research rationale without necessitating fundamentally different principles of research quality. It therefore is appropriate that research quality assessment should be a common and important component of merit across all Streams. There seems little or no reason to apply qualitatively different principles for research quality to different Streams, even if the nature of research differs among Streams.

Benefit from Research

The new access Streams go some way toward better defining the intended beneficiaries and context of benefits that should flow from research in some Streams but also raise a question of whether '*national* benefit' is a required or appropriate frame through which to assess the prospective utility of all work supported by the MNF.

The CSIRO has defined [research] impact as " *An effect on, change or benefit to the economy, society*

and environment, beyond contributions to academic knowledge[3]*"* and asserted that "... *impact is the effect of [CSIRO] work that is generated after this work has been adopted*." The 'end-user beneficiaries' in such a context might be in national or State agencies, the Australian public represented by the Federal Government, small or large businesses operating in Australia (or even elsewhere), or even other researchers, for example involved in a multi-national research campaign. Australian research contributions to Antarctic science, for example, might be in Australia's national interest because contribution to science is a key currency of discourse under the Antarctic Treaty. The key feature of benefit (or impact for CSIRO), however, is that it reflects the application of research outputs to uses beyond just the provision of new knowledge. Fitness-for-purpose in benefit, therefore, should entail ensuring alignment of expected outputs from research with the needs of the relevant end-user constituency.

Characterisations of Streams 1, 4, and 5 clearly and unequivocally require that research will be to address (known) "... *national priorities* ...", "... *are in the national interest* ...", or are aligned "... *with national publicly funded programs/institutions* ...". The anticipated beneficiaries from research in each of these Streams seem most likely to be in the policy or management domains, or the Australian public, since each hinges on prior identification of research needs that presumably were approved on the basis of their importance to the nation. A 'national interest' or 'national benefit' dimension to merit for MNF support seems unequivocal for those Streams.

Access Streams 2 and 3 in the MNF 2030 Strategy make no mention of intended beneficiaries of the research done through them, with requirements only to "... *advance scientific* [sic][4] *knowledge* ..." and for "... *development and testing of innovative technology*" respectively. It might be argued that MNF support for Stream 2 research could be justified entirely on the basis that it will deliver new knowledge to a research discipline. Filling the knowledge gap being addressed might not have been recognised formally before as a national priority or even as having 'national interest'. It alternatively might be argued that links to national interest here might be more tenuous or indirect but contributions to knowledge nevertheless should be able to be shown credibly by the proponent to be relevant to Australia to justify support from a publicly-funded Landmark National Research Facility. The latter would be very similar to what was required for awarding Type 1 granted voyages under the previous access allocation framework.

Research in Stream 3 is expected to lead to technology improvements, presumably for applications in marine or atmospheric research or marine industries. Benefit from research in this Stream perhaps will have greatest potential to flow to individuals or the private sector since no stipulation is provided that the "... *development and testing of innovative technology* ..." must be in the national interest. It arguably will be inappropriate here too to support such research if the innovations or technology improvements bear no benefit to Australia, for example through Australian industry development, enhanced research capability, or more efficient data collection for surveillance or other activities. A requirement to demonstrate benefits of national interest would seem appropriate for Stream 3 activities for the same underlying reasons as for the other Streams, though the vector for flow of benefits might be quite different.

The operational definition of 'national benefit' in the previous MNF Access Framework was obscure to many stakeholders. Some applicants for MNF support found it difficult to address the (then) criterion and some assessors found it difficult to judge claimed national benefit. The label to some degree connotes a focus on benefits that are national in scale and focus and, arguably, expected to be effected through national policy or regulatory outcomes. I recommend, therefore, use of the less-loaded term 'research benefit' rather than 'national benefit' to connote benefits flowing from the research to some specific end-user constituency. My intention in reframing 'benefit' is to allow for multiple 'paths to benefit' through which research outputs from any Stream will be used *beyond* the core disciplinary research community within which a project sits. Research benefit should be interpreted to mean 'benefits to end-users arising from the research', similar to CSIRO's definition of impact above.

Recommendation

3. Research Quality and Research Benefit should be the core principles of merit assessment for all five Streams with relevant Stream-specific considerations, if required,

[3] CSIRO 2015, *Impact Evaluation Guide 2015*, 48pp.

[4] Stream 2 is framed with respect only to "science", which for many will be a loaded framing connoting a specific (probably narrow) set of disciplines considered eligible for MNF support. Mapstone (2019) reported, based on stakeholder commentary, that such terminology is likely to deter some prospective applicants from disciplines outside the traditional marine or atmospheric science disciplines. I suggest therefore, and use in this review, 'discipline-driven' rather than 'science-driven' descriptors for Stream 2, to obviate that risk.

applied in the assessment process rather than by different definitions of merit.

Assessment Criteria

b. Draft clear and unambiguous criteria by which proposals in each Stream would be assessed against the relevant merit principles (as few criteria as possible to meet the need for each Stream and, preferably, apply across multiple Streams).

c. Craft guidance for assessors to use when scoring proposals against the criteria (that is unambiguous and encourages critical scoring).

The above two tasks (drafting criteria and assessor guidance) are treated together in this section because they largely go hand-in-glove. Specific guidance to assessors is presented here because that guidance informs the interpretation of criteria and judgments about the suitability of the criteria for measuring how well proposals measure-up against merit principles. Advice to assessors is linked tightly to a suggested approach to scoring proposals against the criteria, so I also include scoring directions here. I discuss the general scoring framework first because I recommend a common approach to scoring against criteria for both merit principles. Articulating that framework first will simplify presentation, and interpretation, of the recommended assessment criteria and associated assessor guidance. Suggested analytical methods and weighting merit principles or criteria differently for different Streams are discussed later.

Comparing Assessments among Streams

Streamed access allocation could be approached in at least two ways. One approach would be to set *a priori* available *Investigator* sea-time for each Stream. This would be a structured and deliberate strategy but potentially, if applied immutably, would result in unused sea-time if a Stream allowance was undersubscribed. There seems little prospect that available ship-time would go unallocated because of strict adherence to strategically set Stream allowances, so this approach likely would require some reassessment of initial Stream allowances and comparison of proposal among Streams to resolve which proposals from over-subscribed Streams were most deserving of surplus time from another Stream.

A second approach would be to distribute sea-time among Streams in response to the balance of proposals received in an allocation cycle. This approach arguably is more flexible but inherently more reactive to demand. Decisions of priority and distribution of ship-time among Streams are deferred to the end of an allocation round, and so almost certainly would require some form of comparison of proposal assessments across Streams.

It is likely that the actual mechanism used to allocate ship-time among Streams will be a combination of both of the above approaches. It seems unlikely that assessment and ranking of proposals always will be constrained within Streams, unless all Streams are always over-subscribed with acceptable proposals. It is important, therefore, that comparison of assessments across Streams is facilitated by both initial assessments and analyses of results.

The following is based on an assumption that the two merit principles of Research Quality and Research Benefit are retained for all five Streams. The focus here, accordingly, is on criteria by which proposals might be assessed against those merit principles and whether a single set of criteria can be applied usefully to all access Streams or different criteria are required in some or all Streams. There is a preference stated clearly in 'b.' (above) toward a common set of criteria that can be applied meaningfully to all Streams, if possible. Such a framework will be conceptually, administratively, and operationally simpler than crafting Stream-specific criteria and then looking for ways *post-hoc* to reconcile assessments from different Streams. It also is consistent with the approach preferred by most stakeholders consulted by Mapstone (2019): "*... Most considered that the same criteria (whatever they ended up being) should be applied to assessments of all proposals ...*".

A common assessment framework will make comparisons of assessment results among Streams straightforward and minimise uncertainties in ranking proposals across Streams, if necessary, provided assessment results are analysed and presented in a consistent way for all Streams. I therefore weighted favourably the feasibility of cross-Stream comparisons when considering assessment criteria and analysis and presentation of assessment results. I accordingly first drafted principle-based criteria that I thought might have general application and considered where they might fail, rather than drafting Stream-specific criteria semi-independently and looking for similarities among them.

Scoring Framework for Merit Assessment

An important aspect of this review is to recommend clear guidance for assessors of proposals that encourages critical scoring using a broad range of available scores, and reduces the prospect that all proposals would end up with scores within a tight range. Two implicit assumptions of the preference for disparate scores are that: (i) demand for MNF resources, especially sea-time on *Investigator*, is likely to exceed availability; and, therefore, (ii) the assessment process is necessarily comparative and competitive, requiring decisive discrimination among competing proposals.

Proposal assessments, however, generally are not explicitly, or even implicitly, comparative at the initial steps because most proposals are reviewed in isolation. Reviewers often only see one or a small subset of the [competing] proposals and their primary responsibility is to comment on the degree to which a proposal meets specific criteria, not to judge whether one proposal is better than another. It clearly is important to evaluate whether proposals meet some minimum requirements to avoid supporting low-quality or irrelevant research. Guidance to reviewers therefore necessarily is targeted specifically toward assessing the research qualities or research benefits of each proposal *per se*, as is initial scoring.

There would be little need, if any, to rank proposals if MNF resources were sufficient to support all acceptable requests. That cannot be assumed, however, so I will recommend later an approach to analysis and reporting that focuses primarily on turning the initial merit scores into a comparative assessment to help decide which acceptable proposals have the strongest cases for MNF support. I am not focussed on comparative assessment in this section other than to note that the suggested initial scoring framework has been designed to facilitate the later comparative assessments.

The Australian Antarctic Science Program proposal assessment framework provides relatively specific direction to reviewers about scoring ranges for particular proposal qualities. I have adopted a similar approach here to endeavour to prescribe scoring that will be consistent among reviewers and encourage reviewers to score across the available range.

Scoring guidance should lead assessors to consider critically how well a project meets desirable criteria and provide clear guidance about what standard of meeting a criterion would result in a particular score or range of scores. It is difficult to ensure that reviewers score over a wide range and it is almost inevitable that most proposals will be neither dreadful nor astonishing, resulting in a tendency for scores to be 'bunched' in the mid-range. The following general scoring framework, however, caters for the dreadful and astonishing but also provides guidance for some discrimination among the intervening proposals. It lays out the sort of qualities that I suggest can be used to prescribe specific scores against multiple criteria, given appropriate criterion-specific guidance. The framework has scores assigned at a relatively coarse scale (0–5) according to a range of assessments from 'Unsupportable' to 'Compelling'. A more granular scale could be used, but I suggest that little additional information will be gained from more granular scoring within each category.

- ***Unsupportable.*** The proposal is fatally flawed or materially deficient in respect to the criterion[5]. The assessor is not provided with sufficient information to judge whether the project could meet the criterion or the information provided indicates clearly that the project as described cannot satisfy the criterion. The project should not be supported [on this criterion]. ***Score***: 0
- ***Poor.*** The project is unlikely to be fatally flawed but the claim against the criterion is weak or unclear. The assessor has some difficulty in judging whether the project will satisfy the criterion. Significant clarification of the case addressing this criterion would be required to support the project [on this criterion]. ***Score***: 1
- ***Adequate to Good.*** Fit with this criterion is reasonably clearly stated, with acceptable justification. Some non-fatal shortcomings in the case or weaker aspects of the case would result in an 'adequate' assessment whilst a case that was well explained and reasonably convincing would result in an assessment of 'good'. ***Score***: 2 or 3.
- ***Strong***. The project has a thorough, clear, and robust claim against the criterion in all respects. There is very little or no doubt for the assessor that the criterion is well met by the proposal. The work is well presented and demonstrably consistent with contemporary practice [as relates to the criterion]. ***Score***: 4
- ***Compelling***. The project represents a compelling exemplar of good practice by the criterion, with an outstanding case being made of how the work will satisfy the criterion. The assessor has no doubt that the project meets the criterion to an exceptional level. ***Score***: 5

[5] Criterion (criteria) here has a generic meaning, noting that scoring recommended below will be against each or two specific questions (or sub-criteria) within each criterion.

The above framework is presented throughout the following sections with suggested guidance tailored to the criterion or assessment question being asked. I include separately some guidance about additional narrative comments that might be invited from reviewers but I suggest that guidance can be generic rather than Stream-specific.

Merit Principle 1: *Research Quality*

I outlined above what features should be considered in assessing the quality of research, in general: *Research should have a robust disciplinary rationale, clear and appropriate objectives, sound methods, demonstrable feasibility, the prospect of informative results, and be being done by people with appropriate expertise.* These features are distilled here to four assessable primary facets of a proposal that constitute specific assessment criteria:

1. Research rationale & objectives;
2. Research design and methods;
3. Project feasibility and delivery, given available resources; and
4. Capability of the research Team, including research leadership.

Criteria 1 and 2 are about the justification of the project as a contribution to research relevant to the nominated Stream and its standing compared to contemporary best-practice in relevant research disciplines. Relevance to need here relates only to whether a case has been made that the research will fill important information needs for the nominated Stream in the current allocation round. The focus is on how well the case for doing the research has been made, not about the strength of benefits that are expected to flow from the work, which is assessed separately. Criteria 3 and 4 go to the prospect that the research can be done as proposed and the likelihood that the research team will deliver research outputs as promised.

The above four criteria should be applicable to assess any category of research proposal, whether, for example, testing fundamental hypotheses, exploration, monitoring, discovery, technology development, or sampling to fill important information gaps, such as through seabed mapping. It therefore should be straightforward to apply the criteria to proposals in any Stream. The criteria clearly need to be answered in context, but that context should be provided by guidance to applicants for each Stream — especially where Stream priorities will change with deployment period (e.g., Stream 1). Tailoring assessments to the focus or priorities of a Stream for any deployment period should not require redrafting assessment criteria but be done by articulating clearly the type of research needed to meet Stream priorities. Such an approach will be more directed and provide more focus for assessments than previously, when proposals from disparate disciplines competed for sea-time against common criteria. Stream-specific nuances could be provided if necessary in guidance to assessors for scoring against these criteria. There seems no reason, therefore, to define different Stream-specific criteria by which to judge Research Quality.

Each of these criteria involves assessment of two related aspects of a proposal (e.g., rationale and objectives) that could result in ambiguity in final assessments if not managed appropriately. There are at least two ways that these double-valued criteria could be evaluated to minimise such ambiguity: split each criterion into 2 separate, free-standing criteria; or require each component to be evaluated and ‘scored’ explicitly. I recommend the latter because the two components of each criterion clearly are related properties of the intent of the respective criterion, rather than largely separate aspects of a research proposal. Keeping the number of assessment criteria small also should simplify analysis and presentation of the final results.

The two aspects of each criterion can be addressed by asking, and scoring against, just two relevant questions, with appropriate scoring guidance to reviewers. Tracking the derivation of scores against each criterion thus will be available to interrogate nuances of assessments when necessary to make decisions between closely competing proposals, resolve issues for proposals that are ‘line-ball’, or to provide useful feedback to unsuccessful applicants.

Table 1 provides general guidance to reviewers of Research Quality against each of the four criteria, for all Streams. Table 2 provides detailed Stream-specific guidance for scoring against the criteria, including the questions posed to guide consideration of the two facets of each criterion, and scoring instructions. I also indicate in the table that the final score against each criterion should be the simple sum of the two scores against the specific questions (or sub-criteria) posed for that criterion.

Recommendation

4. Research Quality assessment for all Streams be based on explicit consideration of four criteria, scored separately according to specific guidelines for assessors:
    1. ***Research Rationale***: The reason and context for the research, including the research objectives, given Stream strategy;
    2. ***Research Rigour***: Robustness of research design and soundness of the proposed methods;
    3. ***Research Feasibility***: Feasibility of the research proposed and likelihood of promised research outputs being delivered, given available resources; and
    4. ***Research Capability***: Capabilities and capacity of the team to complete the research, including research leadership.

Guidance for reviewers of Research Quality.

Table 1: General guidance to reviewers of Research Quality, including for narrative commentary.

Merit Principle 1: Research Quality - General Guidance to Reviewers	
Reviewers are asked to assess the extent to which a research* proposal meets four criteria (below) and score the proposal against two specific questions for each criterion according to the guidance provided in the attached Table. You are asked to score the proposals fairly but critically against the criteria, based on the scoring guidance in the Table. Your assessment should be based entirely on the content of the proposal, given the MNF background information you have received. You should avoid making allowances for known researchers when assessing proposals against criteria 1–3. Avoid, for example, inferring that poorly or incompletely described research will be soundly designed and executed because it is proposed by well-known, high-performing researchers. It is important for equity of assessments across the range or proponents, from early career to very experienced researchers, that standing of the research against criteria 1–3 is assessed on the merits of the proposal, not on what you know of the proposers. Your knowledge of proponents is expected to inform your judgments against criterion 4, though insufficient or poor justification of research track record or experience should be scored accordingly. Explanatory comments are welcome about merits or deficiencies against any criterion. Comments should be framed to help the MNF interpret your scores or provide constructive feedback to applicants.	
Criterion	**Assessment Guidance**
1. ***Research Rationale:*** *Reason for the research, including the research objectives, given Stream strategy.*	Evaluation of the proposal against this criterion should be focussed on whether the applicant has demonstrated that the research is relevant to MNF Stream objectives (provided separately). The rationale for the work should establish that the proposed research activities are: • important to address the information needs for the Stream end-users; and • Well-grounded in the theory and practice of relevant disciplinary or multi-disciplinary fields. Relevance to Stream objectives here relates only to whether a case has been made that the research will lead to important information for the nominated Stream. The focus is on how well the case for doing this research has been made, not about the strength of benefits or end-user outputs that are expected to flow from the work, which is assessed separately. Research objectives should: • clearly and directly arise from the rationale for the project; and • state specifically and unambiguously what research or technological questions or hypotheses are being addressed.
Criterion	**Assessment Guidance**

2. Research Rigour: *Robustness of research design and soundness of the proposed methods.*	Evaluation of the proposal against this criterion should be focussed on whether the project is well-designed and based on methods that are at least consistent with current best-practice in the relevant discipline(s). You should be provided with sufficient information for you to duplicate the work, in principle. The design of the field and laboratory work and data analyses should provide a sound basis for robust inferences to answer the research objectives. Poorly or incompletely described research design should be marked down. Research methods are not required to be cutting-edge or innovative but should be consistent with current best practice. Innovative or cutting-edge methods may be a strength of the proposal but we ask that you also consider the degree to which risks associated with new methods might jeopardise the research and assess whether the proponents have provided adequate mitigation for those risks.
3. Research Feasibility: *Feasibility of the research proposed and likelihood of promised research outputs, given available resources.*	Evaluation of the proposal against this criterion should be focussed on whether the applicant has convinced you that the work they propose can be completed with the resources they bring to the project and those requested from the MNF. It should be demonstrated that the applicant has planned the field schedule carefully and in detail, including allowing for contingencies such as bad weather. The plans for the voyage should be sufficient but also efficient — please consider whether the applicant is asking for more vessel time than is required (including reasonable contingencies). Please also consider whether the time allowed for post-voyage processing of samples and data is sufficient and realistic. Has the applicant sufficient funding to complete the project, including allowing for preparation and submission of relevant research outputs such as publications or reports.
4. Research Capability: *Capabilities and capacity of the team to complete the research, including research leadership.*	Evaluation of the proposal against this criterion should be focussed on whether the proposal contains sufficient evidence that the research team has the capabilities and capacity to do the research proposed. The application should articulate the expertise and experience of key team members, how much time they will commit to this project, and who will have leadership of the main components of the project, both at sea and in subsequent analyses. It should be clear, probably from attached *curricula vitae*, that the team has sufficient depth and track record of completing projects and reporting results for you to be comfortable that promised research outputs will be delivered within the project timeline. The project will need a capable voyage leader and the proposal should make clear how that leadership will be provided. It is not required that the nominated voyage leader has prior experience and it is desirable that new leaders are developed. It is essential, however, that first-time leaders have sound mentoring or back-up at sea. The application should make clear how such support will be provided if the voyage leader is new to that role. Applications that have clearly articulated plans for mentoring new leaders, whether as primary or deputy leaders, should be scored favourably.

* Research here means the information gathering activities in the proposal, including, *inter alia*, genres of testing hypotheses about fundamental processes or theory, monitoring environmental conditions or biota, mapping ocean features, or developing new or enhancing existing technologies. Your assessment should be with regard to current practice in the relevant genre.

Table 2: Proposed assessment criteria, specific assessment questions, and scoring guidance for the Research Quality principle of merit.

Merit Principle 1: Research Quality — Guidance for Scoring by Reviewers		
Primary Assessment Criterion	**Assessment Questions**	**Scoring Guidance**
1. ***Research Rationale:*** *Reason for the research, including the research objectives, given Stream strategy.* [Score 0–10 = 1a+1b]	1.a. Are the reasons for needing to do this research clearly explained with reference to the relevant MNF Stream strategy? [Score 0–5]	i. ***Unsupportable:*** Need for the research not stated and relevance to Stream strategy not established (Score 0) ii. ***Poor:*** Need for the research not clear and relevance to Stream strategy weakly or vaguely argued (Score 1). iii. ***Adequate–Good:*** Need for the research reasonably clearly articulated with moderate justification of relevance to Stream strategy (Score 2–3). iv. ***Strong:*** Need for the research well argued with clearly established importance to Stream strategy (Score 4). v. ***Compelling:*** A compelling case has been made that the research will provide key knowledge essential to the Stream strategy (Score 5).
	1.b. Are the research objectives or research questions clearly stated and appropriate for the articulated need for the work? [Score 0–5]	i. ***Unsupportable:*** Research objectives very obscure or absent (Score 0). ii. ***Poor:*** Research objectives vague or poorly linked to need, or both (Score 1). iii. ***Adequate–Good:*** Research objectives stated with moderate or good specificity and moderate to clear link to need(Score 2–3). iv. ***Strong:*** Research objectives clear and specific, with good link to need (Score 4). v. ***Compelling:*** Research objectives specific and very clear and convincingly presented as essential to addressing articulated need (Score 5).
2. ***Research Rigour:*** *Robustness of research design and soundness of the proposed methods.* [Score 0–10 = 2a+2b]	2.a. Is the design of the research, including design of sampling or experimental programmes and proposed analyses, robust? Is it sufficient to meet project objectives? [Score 0–5]	i. ***Unsupportable:*** Sampling or experimental design(s) during voyage not articulated or fundamentally flawed; proposed analyses not provided or highly inappropriate (Score 0). ii. ***Poor:*** Details of sampling or experimental design(s) during voyage poorly explained or potentially flawed; proposed analyses poorly explained or deficient (Score 1). iii. ***Adequate–Good:*** Sampling and experimental design(s) during voyage outlined and apparently sound, but with some gaps; proposed analyses only incompletely explained or of questionable appropriateness (Score 2–3). iv. ***Strong:*** Sampling or experimental design(s) during voyage clear and inferentially robust; proposed analyses well explained and appropriate (Score 4). v. ***Compelling:*** Sampling or experimental design(s) during voyage state-of-the art and inferentially difficult to fault; proposed analyses thoroughly explained and clearly best-practice (Score 5).

	2.b. Are the methods, whether well-established or innovative, fit-for-purpose and appropriate for the objectives of the work? [Score 0–5]	i. ***Unsupportable:*** Field methods not explained or demonstrably ill-suited to delivering data needed to meet project objectives (Score 0). ii. ***Poor:*** Field methods poorly explained or poorly aligned with data needed to meet project objectives (Score 1). iii. ***Adequate–Good:*** Field methods reasonably clearly explained and likely to be appropriate to project objectives, though perhaps not always current with contemporary best practice in the relevant fields. Likely to deliver data adequate to meet project objectives (Score 2–3). iv. ***Strong:*** Field methods well justified, robust, and highly likely to deliver high-quality data to meet project objectives (Score 4). v. ***Compelling:*** Field methods clearly at contemporary best-practice or innovative and cutting-edge and highly likely to deliver outstanding and novel data to meet project objectives (Score 5).
3. Research Feasibility: *Feasibility of the research proposed and likelihood of promised research outputs, given available resources.* [Score 0–10 = 3a+3b]	3.a. Is the work well-planned, efficient, and feasible with the facilities and funding (during voyage and afterwards) and sea-time available?	i. ***Unsupportable:*** Voyage plan unclear or available facilities and funding clearly insufficient to ensure completion of project, or both (Score 0). ii. ***Poor:*** Voyage plan needs refinement or clarity and appears either inappropriate or excessive. Available facilities and funding potentially insufficient to secure completion of project, or both (Score 1). iii. ***Adequate–Good:*** Voyage plan reasonable but with some limitations (e.g., insufficient allowance for bad weather, or more time sought than justified); facilities and funding appear adequate to complete the project (Score 2–3). iv. ***Strong:*** Voyage plan well justified, realistic, and efficient and clear evidence that facilities and funding are adequate to complete the project (Score 4). v. ***Compelling:*** Voyage plan very thorough, well justified, and appropriate for the project taking into account contingencies. Realistic and demonstrated abundance of facilities and funding to complete the project (Score 5).
	3.b. Are the promised research outputs likely to be delivered within the proposed schedule, given available resources?	i. ***Unsupportable:*** Research outputs very unclear, not linked to project objectives, or not deliverable within project timeline (Score 0). ii. ***Poor:*** Research outputs unclear or poorly linked to objectives, or both (Score 1). iii. ***Adequate–Good:*** Proposed outputs clear but somewhat vague about how they will meet project objectives or some uncertainty about delivery within project timeline (Score 2–3). iv. ***Strong:*** Research outputs clearly articulated, linked very well to project objectives, and expected to be deliverable within project timeline (Score 4). v. ***Compelling:*** Research outputs excellently detailed with compelling case they will meet project objectives and be delivered within project timeline (Score 5).

4. Research Capability: *Capabilities and capacity of the team to complete the research, including research leadership.* [Score 0–10 = 4a+4b]	4.a. Does the research team have the essential capability and experience to complete the proposed work, both during and following a voyage?	i. ***Unsupportable:*** Project team only partly secured or very inexperienced with significant gaps in the expertise necessary to complete the project (Score 0). ii. ***Poor:*** Project team is weak or lacking experience or has important gaps in the expertise necessary to complete the project (Score 0–1). iii. ***Adequate–Good:*** Project team has good depth or relevant experience and has most or all of the expertise necessary to complete the project satisfactorily (Score 2–3). iv. ***Strong:*** Project team is very strong and experienced with all the expertise necessary to complete the project to a high standard (Score 4). v. ***Compelling:*** Project team is outstanding with field leaders and outstanding early career researchers participating in voyage and has all the expertise necessary to complete the project to a very high standard (Score 5).
	4.b. Does the research team include sufficient research leadership for the team, both on land and at sea?	i. ***Unsupportable:*** Project team has no people with previous voyage experience and does not explain who will provide overall project leadership (Score 0). ii. ***Poor:*** Project team has few or no people with demonstrated voyage leadership experience (though perhaps with voyage experience) or unclear explanation of overall project leadership to deliver promised outputs, or both (Score 1). iii. ***Adequate–Good:*** Project team has at least one person with previous voyage leadership experience (who will be on the voyage) and apparent overall project leadership to deliver promised outputs (Score 2–3). iv. ***Strong:*** Project team has strong demonstrated voyage leadership experience that will be on the voyage and clear project leadership for delivery of promised outputs (Score 4). v. ***Compelling:*** Project team has compelling plan for voyage leadership, including mentoring new or future voyage leaders, and clearly articulated project leadership plan that will ensure timely delivery of promised outputs (Score 5).

Merit Principle 2: *Research **Benefit***

Mapstone (2019) reported widespread uncertainty about how the prospect of national benefit should be demonstrated in research proposals or assessed against the existing 'national benefit' criterion. There was a clear need for greater clarity about addressing the national benefit expectation, including clearer criteria and education and guidance to applicants about how to address them in applications. The 'elevation' of "delivering [national] benefit" to principles of merit allows for development of more specific and detailed assessment criteria that might address these issues.

Previous guidance for scoring against the national benefit criterion did provide some clarity that should have helped applicants frame their proposals, but that advice was 'lost' in the guidance to reviewers. The following statements from the previous assessment guidelines remain relevant as a starting point from which to draft advice to applicants through any Stream of the revised access strategy.

[Applications should:]

- *contain demonstrable policy, industry or other stakeholder link;*
- *offer clear, realistic and appropriate objectives, not just for the science delivery, but for its application and uptake by end users;*
- *provide very high levels of supporting documentation including letters of support from end users in which they:*
  - *specify the path to application of the science being undertaken;*
  - *attest that the science to be conducted is essential to their business or area of responsibility; and*
  - *express confidence in the applicants ability to deliver;*
- *incorporate a strong plan for engaging the end users and demonstrate a history of previous engagement;*
- *demonstrate significant potential for impact, seen as essential by the end users ...; and*
- *provide strong evidence of previous translation of scientific results into demonstrable, significant impact.*

The above advice goes to the sort of evidence that is expected to demonstrate the prospect of benefit beyond advancing research knowledge and would be useful as guidance to applicants provided in the introductory narrative in addition to scoring guidance. Such guidance, framed appropriately for each access Stream, should provide sufficient clarity and context for applicants to articulate specific claims of benefit for that Stream. That approach also should facilitate a common assessment framework, including criteria that are consistent in form across Streams, and so provide administrative and procedural consistency, and arguably fairness, among Streams when awarding MNF support.

One criticism of previous guidance to assessors of national benefit was that time frames for delivery were confounded with the judged importance to end users. Earlier delivery tended to be equated with greater benefit. 'Time to benefit' will be important in some contexts but it is unlikely that short-term benefits always will be considered more important than longer-term benefits by all end-users. I therefore have not included a preferred 'time to benefit' explicitly in assessment criteria. I instead base assessment criteria on an expectation that either the desired time for delivery of outputs to end-users will be captured in the relevant priority settings (Streams 1, 4, or 5) or through applicant engagement with expected research beneficiaries (Streams 2 and 3) and articulated clearly in proposals. It follows that failure to articulate context-specific times to benefit should be seen as a proposal weakness.

Mapstone (2019) suggested that clarity of claims of [national] benefit might be helped by flexibility in framing the specific national benefit assessment criterion, especially for assessment of priority- or campaign-targeted proposals ('applications' below).

> "*It might be appropriate, for example, to tailor assessment* [*sub-*][6] *criteria for different assessment cycles to optimise selection of applications against different priorities in different regions or deployment cycles. Such flexibility would not be expected to extend to wholesale recasting of assessment criteria but should be countenanced as a means of applying specific focus on cycle-specific priorities at* [*sub-*] *criterion level within the primary criterion* [= merit principle]. *It does not seem appropriate to narrow the focus of the national benefit criteria for assessment of research-driven applications or user-funded applications, but being more specific about particular foci for policy-driven voyages in different assessment cycles would help guide applicants when preparing policy-targeted or campaign-targeted*

[6] The 'assessment sub-criteria' described by Mapstone (2019) effectively equate to 'assessment criteria' in this document, whilst his 'primary criteria' equate to merit principles here.

applications and assessors in evaluating those proposals. "

The above suggestions will be captured most usefully if the tailored information to inform proposal preparation and assessment is provided for each Stream when each call for applications is prepared, *without* modifying the assessment criteria *per se*. This approach will allow for constancy and consistency of the assessment framework both among and within Streams whilst also recognising that benefit focus regarding end-user constituency, priority information needs, time to delivery, forms of outputs, and so on will differ among Streams and, for some Streams, among deployment cycles. This is the approach I've adopted here in drafting common assessment criteria by which to measure claims of research benefit.

The key features of a claim that the proposed research would deliver end-user benefits of national interest can be captured by the following broad criteria, given clear definition of a Stream's requirements:

1. Rationale for the work in terms of identified end-user need or national interest;
2. Outputs aligned with end-user need and deliverable in appropriate timeframes;
3. Convincing path to benefit, including involvement with expected beneficiaries or end-users; and
4. Evidence of capability to deliver benefit to end-users, including responsibility for doing so.

Criteria 1 and 2 address relevance of the project, including objectives, and proposed outputs to end-user needs. Relevance to need here relates only to whether a case has been made that the project will supply end-users with needed information relevant to the nominated Stream in the current allocation round. It is analogous to criterion 1 for research quality but the focus here is on justifying end-user relevance whereas the other case is about the rationale for the research that will underpin the end-user outputs.

Criteria 3 and 4 go to verifying that outputs will be appropriate to end-users and delivered in ways that mean they credibly have prospects of being used. Important considerations are whether the project team has relevant experience to work effectively with end-users and has established mechanisms to deliver outputs to end-users in a timeframe and form that will facilitate application. These criteria are not expected to be about delivery of research outputs (e.g., publications in research journals) unless those are justified as the most appropriate outputs for the nominated end-user beneficiaries of the project (e.g., which might be the case for some Stream 2 proposals).

The intent here, as with Research Quality criteria, is that the four criteria can be used to assess the prospect of benefit from any category of research proposal, including those intended to inform public policy, support regulatory monitoring, develop new technology, or advance Australia's international standing in a research discipline, including in the interests of soft diplomacy (e.g., within the Antarctic Treaty System). It should be straightforward to apply the above criteria to benefit assessment for any Stream-specific application, provided sufficient Stream-specific context is provided for applicants to identify relevant end-users and identify their research and output needs.

At least some of these criteria also have two related components for assessment (e.g., outputs and time to delivery) that could result in ambiguity in assessments. I recommend the same approach to avoiding ambiguity here as for application of Research Quality criteria: address the two aspects of each criterion by asking, and scoring against, two relevant questions, with appropriate scoring guidance to reviewers (Table 4). I also indicate in the table that the final score against each criterion should be the simple sum of the two scores against the specific questions (or sub-criteria) posed for that criterion, similar to the process for scoring against Research Quality criteria.

I provide here (Table 3) some draft overarching explanatory advise to reviewers of Research Benefit in all Streams, analogous to that for reviewers of Research Quality in Table 1.

Recommendation

5. Research Benefit assessment for all Streams be based on explicit consideration of four main criteria, scored separately according to specific guidelines for assessors:
    1. ***Benefit Rationale***: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream;
    2. ***Benefit Outputs***: Alignment and utility of project outputs to articulated end-user needs and national interests;
    3. ***Path to Benefit***: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs;

4. ***Capacity to Deliver Benefit***: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.

Guidance for reviewers of Research Benefit.

Table 3: General guidance to reviewers of Research Benefit, including for narrative commentary.

Merit Principle 2: Research Benefit - General Guidance to Reviewers	
Reviewers are asked to assess the extent to which a proposal is likely to deliver research benefit. Research benefit here means the application of research outputs to uses beyond the provision of new academic knowledge in the discipline of the research being proposed. Research benefit therefore implies applications in policy, regulation, the economy, society, to understand the environment, for contribution to other disciplines or major research initiatives, etc., usually effected after the proposed research is finished. The 'end-user beneficiaries' in such a context are those who will apply the project outputs and might be in Federal or State agencies, small or large businesses operating in Australia, or even other researchers, for example involved in a multi-national research campaign in which Australia is a participant and from which Australia will gain important information. Proposals should demonstrate that benefits from the projects will be in Australia's national interest. You are asked to judge how well the proposal meets four criteria (below) and score the proposal against two specific questions for each criterion according to the guidance provided in the attached Table. You are asked to score the proposals fairly but critically against the criteria, based on the scoring guidance in the Table. Your assessment should be based entirely on the content of the proposal, given the MNF background information you have received. It is important for equity of assessments across the range or proponents that likelihood of benefit against criteria 1–3 is assessed on the merits of the proposal, not on what you know of the proposers or beneficiaries. Avoid, for example, inferring that poorly or incompletely described end-user benefits or path to benefit will be OK because it is proposed by well-known, high-performing researchers or you have confidence that a nominated end-users will 'join the dots'. Your knowledge of proponents is expected to inform your judgments against criterion 4, though insufficient or poor justification of delivery track record or experience should be scored accordingly. Delivery track record clearly is developed over substantial time, meaning there is potential for Criterion 4 to discriminate against early career researchers. You are asked, therefore, to assess the depth of capability and experience in the team, not just the principle investigator. Proposals that include a specific and credible plan for senior researchers with strong delivery track records to mentor early- or mid-career researchers, who might be the principle investigators, should be scored favourably. Explanatory comments are welcome about merits or deficiencies against any criterion. Comments should be framed to help the MNF interpret your scores or provide constructive feedback to applicants.	
Criterion	**Assessment Guidance**
1. ***Benefit Rationale:*** *Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.*	Evaluation of the proposal against this criterion should be focussed on whether the applicant has demonstrated that the project is responding to information needs or gaps of relevant MNF Stream end-users. The rationale for the work should establish that the proposed activities will result in outputs that are: • important to for the information needs for the Stream end-users or Stream objectives; and • in Australia's national interest by virtue of the expected end-user contributions to national or State policy, regulatory, or other priorities, Australian industry or society, or participation in international activities to which Australia is a signatory, including multi-national research campaigns where appropriate to the Stream. Relevance to Stream objectives here relates to whether a case has been made that the project addresses explicitly important information needs for the nominated Stream. The focus is on how project outputs are aligned with end-user needs, not about justification of the research as a contribution to a research discipline, which is assessed separately. Project objectives (for outputs) should: • clearly and directly arise from the articulated end-user needs; and

	• state specifically and unambiguously what is intended to be delivered to end-users to address those needs.
Criterion	**Assessment Guidance**
2. ***Benefit Outputs***: *Alignment and utility of project outputs to articulated end-user needs and national interests.*	Evaluation of the proposal against this criterion should be focussed on whether the promised outputs for end-users are well-matched with the stated information needs of Stream end-users. There should be evidence that the expected end-user beneficiaries have either contributed to the design of outputs intended for them or, at minimum, endorsed the proposed outputs. It is desirable that end-users will have clear input to preparation or review of planned outputs. End-user outputs might take many forms, including reports, briefings, advice, technology designs, inputs to other research activities (e.g., data inputs to ecosystem or climate models). It is important that the proposal establishes that the outputs proposed will be fit-for-purpose specifically for adoption by the expected end-users.
3. ***Path to Benefit***: Path to benefit for end-user outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key user-related outputs.	Evaluation of the proposal against this criterion should be focussed on whether the applicant has convinced you that they have established a clear strategy for engaging with end-users to transfer effectively outputs or products intended for use by end-users. There should be a clear plan for the format, mode, and timetable for delivery of end-user outputs. It should be demonstrated that the applicant has liaised with end-users in the design of a path-to-benefit and that the end-users endorse the plan. The case for an effective path to benefit will be strengthened by evidence, usually from the end user(s), of how the planned outputs will be applied to effect change in the relevant operational domain.
4. Capacity to Deliver Benefit: *Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.*	Evaluation of the proposal against this criterion should be focussed on whether the proposal contains sufficient evidence that the project team has demonstrated capabilities and capacity to engage effectively with end-users and deliver outputs to them in appropriate forms. The proposal should articulate the end-user engagement experience of team members and stipulate who will have leadership of delivering outputs or products to end-users. Output delivery can be led by researchers with little relevant experience provided there is a robust strategy for them to be mentored by others with a sound track record of end-user engagement. It should be clear, preferably from previous end-user testimonials, that at least some team members have strong track records of delivering end-user outputs from research. There should be sufficient evidence provided for you to be comfortable that promised end-user outputs will be delivered within the promised timeline, which should be explicit.

Table 4: Proposed assessment criteria, specific assessment questions, and scoring guidance for the Research Benefit principle of merit.

Merit Principle 2: Research Benefit		
Primary Assessment Criterion	**Assessment Questions**	**Scoring Guidance**
1. ***Benefit Rationale:*** *Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream* [Score 0–10 = 1a+1b]	1.a. Is a clear rationale for the project provided that establishes the relevance of the work to research needs of Stream end-users and is in Australia's national interest? [Score 0–5]	i. ***Unsupportable***: Project not justified with reference to end-user needs or national interests with no evidence of end-user input or support. (Score 0). ii. ***Poor***: National or end-user need for research articulated vaguely or end-user support for project unclear, or both. (Score 1). iii. ***Adequate–Good***: End-user or national needs for research reasonably clear and project objectives credibly justified by reference to those needs, including some supporting evidence from end-users (Score 2–3). iv. ***Strong***: Clear articulation of end-user needs for research and its national interest, with strong end-user support demonstrating robust justification that project objectives will address those needs (Score 4). v. ***Compelling***: Very strong justification that project objectives are framed specifically to deliver important end-user needs that are in the national interest, with a compelling end-user supported case that the research is targeted to fill relevant end-user information gaps (Score 5).
	1.b. Are project objectives appropriate to addressing the end-user needs or national interests used to justify the research (a above) [Score 0–5]	i. ***Unsupportable***: Project objectives not relevant to project rationale and will not address national interests or end-user needs. (Score 0). ii. ***Poor***: Relevance of project objectives only indirectly or vaguely relevant to national interests or end-user needs for research. (Score 1). iii. ***Adequate–Good***: Project objectives credibly- to well-justified by reference to articulated end-user needs or national interests (Score 2–3). iv. ***Strong***: Robust, documented justification that project objectives will address identified end-user needs and national interests, with sound evidence of support from end-users (Score 4). v. ***Compelling***: Very strong justification that project objectives are framed specifically to deliver important end-user needs and is in the national interest, with a compelling case, supported by end-users, that the objectives specifically address key end-user needs of national interest (Score 5).
2. ***Benefit Outputs***: *Alignment and utility of project outputs to articulated end-user needs and national interests* [Score 0–10 = 2a+2b]	2.a. Are the proposed project outputs well matched to articulated national interest or end-user demand for products from research? [Score 0–5]	i. ***Unsupportable***: No evidence provided of end-user demand for proposed project outputs or outputs not at all relevant to stated demand, or both (Score 0). ii. ***Poor***: Some evidence of end-user demand for project outputs provided but with poorly framed expectations of what is sought or proposed project outputs not credibly linked to end-user demand or national interest, or both (Score 1). iii. ***Adequate–Good***: Supporting evidence of end-user demand provided with reasonably clear articulation of why research is needed and in the national interest and outputs well-targeted to meet stated demand (Score 2–3).

		iv. ***Strong***: Robust supporting documentation from end-users that articulates clearly and specifically what outputs sought from the project and outputs strongly aligned with that demand (Score 4). v. ***Compelling***: Project outputs or products match directly those sought by end-users and demonstrably fit with end-user specifications and of national interest, which is demonstrated unequivocally by detailed supporting documentation (Score 5).
	2.b. Is there evidence of collaboration with end-users in project development, especially the identification and framing of end-user targeted outputs? [Score 0–5]	i. ***Unsupportable***: No evidence is provided that project outputs have been discussed with end-users or that end-users want the outputs, or both (Score 0). ii. ***Poor***: Apparently some discussion of outputs with end-users but no evidence of support from end-users or end-user endorsement of project outputs (Score 1). iii. ***Adequate–Good***: End-users clearly involved to some degree in project formulation and articulation of outputs but depth of collaboration either unclear or tentative (Score 2–3). iv. ***Strong***: Well demonstrated evidence of end-user involvement in project development with clear contributions from end-users to formulation of project outputs (Score 4). v. ***Compelling***: End-users integral to project development, including as members of project team, with convincing evidence that end-users helped frame the proposed outputs and endorse them as fit-for-purpose (Score 5).
3. ***Path to Benefit***: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs. [Score 0–10 = 3a+3b]	3.a. Is there a specific plan for post-voyage engagement with expected beneficiaries of the project, including a clear timetable for delivery of outputs to end-users?	i. ***Unsupportable***: No plans presented for post-voyage engagement with end-users and not at all clear how end-user outputs will be delivered (Score 0). ii. ***Poor***: Vague or cursory reference to post-voyage engagement with end-users with mechanisms and timetable for transmission of user outputs vague (Score 1). iii. ***Adequate–Good***: Exposition of some post-voyage engagement with end users and reasonable, if somewhat general, mechanisms and timetable for delivering outputs to end-users (Score 2–3). iv. ***Strong***: Detailed engagement plan for delivering specific outputs to end-users with clear and credible timetable (Score 4). v. ***Compelling***: Well reasoned and strongly substantiated actions to engage with end-users that have demonstrated end-user support and are scheduled with a specific timetable for delivery explicitly related to end-user processes for receipt and application of outputs (Score 5).
	3.b. Will end-users be involved in preparation or delivery of project outputs?	i. ***Unsupportable***: End-users either not identified or apparently not involved in project delivery (Score 0). ii. ***Poor***: End-users identified but apparently unlikely to be involved in either preparation or delivery of end-user targeted outputs (Score 1). iii. ***Adequate–Good***: End-users clearly identified with opportunities articulated for end-user involvement in output preparation or review (Score 2–3).

		iv. ***Strong***: Specific plans articulated for end-user involvement in output preparation or review with clear delivery mechanisms and credible prospect of end-user uptake (Score 4). v. ***Compelling***: Clear and convincing plans for end-user engagement in project delivery with delivery mechanisms agreed with end-users and convincing evidence of probable end-user uptake and application of outputs (Score 5).
4. Capacity to Deliver Benefit: *Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.* [Score 0–10 = 4a+4b]	4.a. Does the project team include people with demonstrated capability to deliver outputs effectively to end-users?	i. ***Unsupportable***: No evidence presented of any project personnel having experience or track record of working effectively with end-users (Score 0). ii. ***Poor***: Indicative but non-specific assertions of experience working with research end-users but without clear evidence (Score 1). iii. ***Adequate–Good***: Moderate to solid evidence that one or more project personnel have worked effectively with research end-users before (Score 2–3). iv. ***Strong***: Clearly stated evidence of prior effective experience in working with research end-users to deliver end-user benefit (Score 4). v. ***Compelling:*** Outstanding and well-documented track record(s) of project personnel working with end-users to deliver research benefits, supported by end-user testimonial(s) (Score 5).
	4.b. Are the responsibilities for ensuring output delivery to end-users clear and likely to be effective?	i. ***Unsupportable***: Responsibility for ensuring end-user output delivery not specified (Score 0). ii. ***Poor***: Only general statements about responsibilities for or engagement in delivering outputs to end-users (Score 1). iii. ***Adequate–Good***: It is clear who on the project team will have carriage of delivering outputs for end-users but with few or incomplete operational details (Score 2–3). iv. ***Strong***: Leadership of end-user output delivery is explicit with clear responsibilities and capabilities for effecting delivery (Score 4). v. ***Compelling***: A detailed plan for output delivery is presented with clear and explicit leadership and operational responsibilities against a specific delivery timetable (Score 5).

Stream-specific Advice to Applicants

The approach recommended above, specifically to apply a common assessment framework with consistent assessment criteria across all Streams, has several advantages, including:

- The assessment framework, including criteria and scoring guidance, remain static once resolved to be appropriate and robust;
- There will be consistency, and arguably fairness, over time (among allocation cycles) in the assessment process, obviating the risk that redrafting assessment criteria for different allocation cycles introduces unintended variation in assessment 'standards' or biases;
- Assessment results will be readily comparable among Streams, based on a common set of criteria with a common reflection of MNF strategy and national policy guidelines for use of national research facilities;
- Assessment panels will work with a single framework of common criteria rather than having to interpret different proposals, and results of reviews, against different criteria;
- The prospect of comparing recommended support among Streams on inconsistent, or at least different, bases will be diminished when resolving the appropriate balance of allocation across Streams in each deployment cycle.

There also clearly are some risks with applying standard assessment criteria:

- The assessment framework might fail to align with specific Stream strategies and so fail to provide a sufficiently robust filter through which to assess Stream proposals;
- The common assessment framework, or specific criteria, might not be applicable easily to one or more Streams, making preparation or assessment of proposals unnecessarily difficult;
- It might be considered that a common framework embeds some unfairness, for example by focussing too much on tangible, short-term, or user-articulated benefits and thereby effectively devaluing longer-term, strategic research benefits, despite the latter being of national interest;
- The adoption of a static framework might result in the assessment criteria becoming dated or having variable relevance to different allocation cycles, at least for those Streams with dynamic priorities (Streams 1, 4, and 5).

I have argued that the utility of a common assessment framework and avoiding potential shortcomings of same rely on the provision of Stream-specific statements to articulate each Stream's strategic focus and clear guidance for preparing applications. The first three of the above risks will be mitigated largely by appropriately tailoring Stream-specific advice to applicants and assessors. That advice should ensure that the proposals in each Stream, and their assessments, can be focussed on the particular requirements for that Stream, without requiring the wording of assessment criteria for Research Quality or Research Benefit. Poor or ambiguous advice will heighten these risks. I recommend careful consideration, therefore, of the draft advice provided here, including some 'road testing' to test whether they create or embed ambiguities that will undermined the use of a common assessment framework.

The Stream strategic intents largely are stated in the MNF Strategy 2030, though it will help there if the expected constituencies for Streams 2 and 3 are stated in the same way they are for Streams 1, 4, and 5. I suggest below some rewording of statements of Stream strategy and intent for Streams 2 and 3 in the interests of providing greater clarity of potential beneficiaries and consistency with Streams 1, 4, and 5.

Tables 5–9 provide draft Stream-specific guidance to applicants for preparing proposals for each Stream to set-up the use of consistent assessment criteria for Research Quality and Research Benefit.

Recommendation

6. Stream-specific guidance be provided to applicants about framing proposals to address Research Quality and Research Benefit assessment criteria with reference to each Stream's strategy.

Table 5: Guidance for preparation of proposals to address specified National Priorities in Stream 1.

Stream 1: Priorities-driven research, for proposals ***explicitly addressing national priorities*** for deployment	
Research in Stream 1 is expected to address directly the national priorities stated for the current allocation cycle and have demonstrably credible prospects of delivering benefits to the nominated end-users who have carriage of those priorities. You should discuss your research ideas with those end-users and verify that the research you have in mind is relevant to their priorities, and likely to fill gaps in the knowledge-base relevant to those priorities. You also should discuss what they need, what it is possible (and not possible) for you to deliver, what outputs from your research will be useful to them, in what form they will be useful, and when they will need those outputs.	
Addressing Research Quality Criteria	
Assessment Criterion	**Guidance to Applicants**
Research Rationale: The reason and context for the research, including the research objectives, given Stream strategy.	Your research should be justified both with reference to its alignment with Stream 1 priorities in the current allocation cycle and with reference to *relevant* disciplinary knowledge and theory, including current research questions. The need for multiple voyages, if sought, should be justified explicitly, including for recovery of deployed equipment. Research objectives should address clearly the key disciplinary research questions relevant to Stream 1 priorities. Research objectives should be specific and unambiguously linked to the objectives for delivering end-user outputs so that assessors can see how addressing the research objectives will underpin delivery of benefit to the national priorities for the allocation cycle (see guidelines for addressing Research Benefit).
Research Rigour: Robustness of research design and soundness of the proposed methods.	You should explain the design (sampling, experimental, analyses) of your research and the methods you will use to gather data sufficiently clearly for a reviewer to be able to repeat the work, in principle. Research methods are expected to be consistent with current best-practice. Clearly indicate if any of the methods you will use are developmental or innovative and describe how you will mitigate potential risks of using such cutting-edge methods.
Research Feasibility: Feasibility of the research proposed and likelihood of promised research outputs, given available resources.	You are requesting granted use of MNF facilities, especially the high-valued RV *Invesigator*. It is important that you demonstrate you have a credible voyage plan that will enable the field work to be completed, including allowing for bad weather and other contingencies, and a post-voyage plan for analysis and write-up that will see the research results published, samples processed, and data archived within a reasonable (specified) time after the voyage (usually not more than 3 years). Your use of Investigator must be efficient — do not request more sea-time than needed.
Research Capability: Capabilities and capacity of the team to complete the research, including research leadership.	You will need to demonstrate that your team includes all the expertise necessary to complete the research and that you have a solid base of experience, including on research vessels, and people who have led sea-going research before. Projects that include a mix of early career and experienced researchers will be viewed favourably, especially those with clear plans for mentoring early career researchers in sea-going research leadership. Voyage leadership need not be by established researchers but you should demonstrate that you will have on-board sufficient voyage leadership experience to mentor a first-time leader. Details of research track record and publication output generally will be demonstrated by provision of CVs.

Stream 1: Priorities-driven research, for proposals ***explicitly addressing national priorities*** for deployment	
Addressing Research Benefit Criteria	
Assessment Criterion	**Guidance to Applicants**
Benefit Rationale: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.	Proposals should establish clearly that the research proposed directly addresses key questions or needs stated in the national priorities for deployment in the current allocation round. Ensure your proposal: • contains explicit links to the priorities identified for this allocation round, such that a reviewer can see easily that the research will fill one or more important information gaps needed in the relevant priority area; • Provides clear, realistic, and appropriate objectives that address the Stream national priorities — not just for the research outputs, but for their application and uptake by the end-users for whom the identified national priorities are important (e.g., the relevant government agency or department).
Benefit Outputs: Alignment and utility of project outputs to articulated end-user needs and national interests.	You should describe expected end-user-targeted outputs that are explicitly framed for adoption and application by those end-users. These outputs might include research reports but in Stream 1 also likely will need to include more specific products that fit with end-users' expertise and operational needs. You should discuss outputs with expected end-users during proposal preparation and preferably provide supporting statements from end-users that demonstrate that the outputs are seen as essential by the end-users and have significant potential for adoption.
Path to Benefit: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs.	You should lay-out clearly a strong plan for engaging with the Stream's nominated end-users, being clear about the mechanisms and timetable by which you propose to engage with them. Provide convincing evidence that your delivery strategy is endorsed by targeted end-users. You should aim to provide high levels of supporting documentation, including letters of support from end-users in which they: • specify the path to adoption of the project outputs; • attest that the research and planned outputs are essential to their business or area of responsibility; and • express confidence in your ability to deliver the expected outputs of benefit.
Capacity to Deliver Benefit: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.	You should provide evidence to demonstrate that your team includes people with a track-record of effective engagement with end-users, including such things as; • Strong evidence of previous translation of scientific results into outputs that had demonstrable utility; • Supporting testimonials from previous research end-users with whom members of your team have worked.

Table 6: Guidance for preparation of discipline-driven[7] Proposals in Stream 2.

Stream 2: Discipline-driven proposals with the ***primary purpose of advancing knowledge*** and that do not directly address priorities specified in other Streams but are directly or indirectly in Australia's national interest.	
Research in Stream 2 should address important disciplinary, or preferably multi-disciplinary, research questions of national or international standing and advance significantly the understanding of key features or processes in the ocean or atmosphere or Earth system, including humanity. The research should be relevant to Australia or address key questions that it is in Australia's national interest directly or indirectly to have answered. You should identify the national information needs, policy area, or public interests that your research would be expected to inform, whether directly or through contributions to global knowledge on which Australia draws for its national interests. You should identify agencies or sectors, potentially including relevant international research programs in which Australia participates, which you expect to benefit from the information you will produce. Specific links between your research and national priorities will be an advantage.	
Addressing Research Quality Criteria	
Assessment Criterion	**Guidance to Applicants**
Research Rationale: The reason and context for the research, including the research objectives, given Stream strategy.	Your research should be justified in the context of disciplinary or multi-disciplinary knowledge and theory, with particular reference to current research questions of recognised importance to the relevant discipline(s). You should articulate how your research will contribute significantly to advancing knowledge in your field or test specific hypotheses to advance theory. The need for research over multiple years should be justified if multiple voyages are being sought. Research objectives should address clearly the key disciplinary questions or knowledge gaps you have articulated.
Research Rigour: Robustness of research design and soundness of the proposed methods.	You should explain the design (sampling, experimental, analyses) of your research and the methods you will use to gather data sufficiently clearly for a reviewer to be able to repeat the work, in principle. Research methods are expected to be consistent with current best-practice. Clearly indicate if any of the methods you will use are developmental or innovative and describe how you will mitigate potential risks of using such cutting-edge methods.
Research Feasibility: Feasibility of the research proposed and likelihood of promised research outputs, given available resources.	You are requesting granted use of MNF facilities, especially the highly-valued RV *Invesigator*. It is important that you demonstrate you have a credible and efficient voyage plan that will enable completion of field work, including allowing for bad weather and other contingencies, and a post-voyage plan for analysis and write-up that will see research results published, samples processed, and data archived within a reasonable time (usually not more than 3 years).
Research Capability: Capabilities and capacity of the team to complete the research, including research leadership.	You will need to demonstrate that your team includes all the expertise necessary to complete the research and that you have a solid base of experience, including on research vessels, and people who have led sea-going research before. Applications that include a mix of early career and experienced researchers will be viewed favourably, especially those with clear plans for mentoring early career researchers in sea-going research leadership. Voyage leadership need not be by established researchers but you should demonstrate that you will have on-board sufficient voyage leadership experience to mentor a first-time leader.

[7] I commented earlier (footnote, page 7) about potential negative consequences of restricting Stream 2 to 'Science' driven proposals. I suggest using 'discipline-driven' to indicate clearly that proposals arising from any (relevant) disciplines could be considered for allocation in Stream 2. 'Open-call' also could be used to label Stream 2 more generally.

Stream 2: Discipline-driven proposals with the ***primary purpose of advancing knowledge*** and that do not directly address priorities specified in other Streams but are directly or indirectly in Australia's national interest.	
	Details of research track record and publication output generally will be demonstrated by provision of CVs.
Addressing Research Benefit Criteria	
Assessment Criterion	**Guidance to Applicants**
Benefit Rationale: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.	Proposals should establish clearly that the research proposed will provide knowledge important directly or indirectly for improving Australian interests in policy, governance, industry, or understanding of national environmental, social, or economic circumstances. Your justification here should: • Identify the Australian interest(s) that stand to benefit from the new knowledge you will deliver and who (the end-users) will be the recipients of relevant outputs that will deliver that new knowledge; • Explain why your research is relevant to Australia's national policy, industry. or public interests, or international standing, such that reviewers can see that the research will fill important information gaps in the relevant area; • Provide clear, realistic, and appropriate objectives that address the national interests you identify — not just for the research outputs, but for their prospective uptake by the end-users for whom the identified national interests are important (e.,g., the relevant government agency or department, national research campaign, etc.).
Benefit Outputs: Alignment and utility of project outputs to articulated end-user needs and national interests.	You should describe expected outputs for end-users that are explicitly framed for use by those end-users. End-users of Stream 2 research might include other researchers who will apply your results to national or international research initiatives, such as improving models of climate, the ocean, biodiversity or informing global research campaigns. Outputs might include research publications but also likely will need to include products or process that fit end-users' needs. You should explain how your results will be presented in ways that enable their application to other fields, models, etc. You should discuss outputs with expected end-users (including other researchers) and preferably provide supporting statements that demonstrate outputs are seen as valuable, accessible, and relevant to those end- users.
Path to Benefit: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs.	You should lay-out clearly a plan for delivering the knowledge targeted at your nominated end-users, being clear about the mechanisms and timetable by which you will engage with them. Provide convincing evidence that your delivery strategy is likely to align with end-user needs. Your case for benefit will be helped considerably by supporting documentation, including letters of support from end users, including other researchers where relevant, in which they: • Attest that the new knowledge you will deliver will be useful to them; and Express confidence in your ability to deliver the expected outputs.
Capacity to Deliver Benefit: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.	You should provide evidence to demonstrate that your team includes people with a track-record of effective engagement with end-users, including researchers in other disciplines if relevant, including such things as; • Strong evidence of previous translation of scientific results into outputs of demonstrable utility to others; and

Stream 2: Discipline-driven proposals with the ***primary purpose of advancing knowledge*** and that do not directly address priorities specified in other Streams but are directly or indirectly in Australia's national interest.	
	• Supporting testimonials from previous research end-users with whom members of your team have worked.

Table 7: Guidance for preparation of technology development proposals in Stream 3.

Stream 3: Proposals to undertake ***development and testing of innovative technology*** that will advance Australia's capability or capacity for research and understanding or use of it's marine or atmospheric domains.	
Research in Stream 3 is expected to address directly areas in which development of new technologies or improvement of existing technologies will enhance the collection or delivery of data from marine or atmospheric environments or the Earths System, including humanity, or contribute to marine industries of national interest. Your proposal should identify clearly the technology opportunity that is being addressed and demonstrate how delivering the proposed technology innovations or improvements will be in Australia's national interest. You should discuss your technology research with those who you expect to use the resulting technology improvements or innovations (your end-users) and verify that the research is likely to fill important technology gaps in their operations, whether in research, governance, or industry. You also should discuss with end-users how your results will be applied to develop new products or in future activities.	
Addressing Research Quality Criteria	
Assessment Criterion	**Guidance to Applicants**
Research Rationale: The reason and context for the research, including the research objectives, given Stream strategy.	Your research should be justified both with reference to its alignment with a clear opportunity for technology innovation to make a difference in some activities and also with reference to current status of technologies in *relevant* operational contexts. The need for research over multiple voyages should be justified if multiple voyages are being sought, including to recover previously deployed equipment. Research objectives should address clearly the key technological questions that need to be addressed and be linked transparently to the objectives for delivering end-user outputs (see guidelines for addressing Research Benefit).
Research Rigour: Robustness of research design and soundness of the proposed methods.	You should explain the design (technology, testing regime, performance metrics, analyses) of your research and the methods you will use to trial your technologies to gather performance data sufficiently clearly for a reviewer to be able to repeat the work, in principle. Research methods are expected to be consistent with current best-practice in technology development and testing.
Research Feasibility: Feasibility of the research proposed and likelihood of promised research outputs, given available resources.	You are requesting granted use of MNF facilities, especially the highly-valued RV *Invesigator*. It is important that you demonstrate you have a credible and efficient voyage plan that will enable the field work to be completed, including allowing for bad weather and other contingencies, and a post-voyage plan for analysis, refinement, and write-up that will see the research results documented appropriately and technology concepts secured, including through patents if appropriate, within a reasonable (specified) time after the voyage (usually not more than 3 years).
Research Capability: Capabilities and capacity of the team to complete the research, including research leadership.	You will need to demonstrate that your team includes all the expertise necessary to complete the research and you have a solid base of experience, including with deployments at sea, and people who have led sea-going technology research before. Applications that include a mix of early career and experienced researchers or technologists will be viewed favourably, especially those with clear plans for mentoring early career individuals in sea-going leadership of technology development or operation. The voyage leader need not be an experienced leader but you should demonstrate that you will have on-board sufficient voyage leadership experience to mentor a first-time leader. Details of technology development track record and application generally will be demonstrated by provision of CVs.

Stream 3: Proposals to undertake ***development and testing of innovative technology*** that will advance Australia's capability or capacity for research and understanding or use of it's marine or atmospheric domains.	
Addressing Research Benefit Criteria	
Assessment Criterion	**Guidance to Applicants**
Benefit Rationale: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.	Proposals should establish clearly that the research proposed will either provide new insights for technology development or innovation or field-test recent technology innovations. Your justification here should: • Identify the end-users of the technology that you are researching and the application(s) for which the technology is being developed, whether in research, industry, government, or production sectors; • Explain clearly why your research is important to fill a technology gap or opportunity or address a technology deficiency in need of resolution; • Establish that the technology improvement or innovation you seek will be in Australia's national interest, for example through improved research, monitoring, or management capability or efficiency, industry development, or international standing; and • Provide clear, realistic, and appropriate objectives that address the technological needs you identify — not just for the research outputs, but for uptake and use or further development or production by the nominated end-users.
Benefit Outputs: Alignment and utility of project outputs to articulated end-user needs and national interests.	You should describe expected outputs for delivery to end-users that are explicitly targeted at technology applications for use in Australia's interests. These outputs might include research reports but also likely will need to include more specific products for technology transfer, such as design and operation details, patents, or capability development materials or training. You should discuss these outputs with expected end-users during proposal preparation and provide supporting statements from end-users that demonstrate your outputs are seen as fit-for-purpose and readily applicable to further development or adoption of the new or enhanced technologies.
Path to Benefit: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs.	You should lay-out clearly a plan for delivering the technology innovations to your nominated end-users, being clear about the mechanisms and timetable by which you propose to engage with them to transfer technological information or products. Provide convincing evidence that your delivery strategy is endorsed by targeted end-users and aligns with their operational needs. You should aim to provide high levels of supporting documentation, including letters of support from end users in which they: • Attest that the technology research to be done is addressing an important technology gap or deficiency; • Specify how they will apply the results of your technology research; and • Express confidence in your ability to deliver the expected outputs in ways that will be useful to them.
Capacity to Deliver Benefit: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.	You should provide evidence to demonstrate that your team includes people with a track-record of effective engagement with end-users, including such things as; • Strong evidence of previous application of new technologies or transfer of technologies to end-users; and • Supporting testimonials from end-users with whom you have worked before to implement new technologies.

Table 8: Guidance for preparation of proposals for user-funded research in Stream 4.

Stream 4: ***User-funded research***, for proposals that are ***in the national interest and rely on RV Investigator's specific capabilities***.	
Research in Stream 4 will be funded in part or full by the applicant or supporting end-user(s) and will be considered for MNF support only if it is demonstrated in advance that the proposed work relies on specific capabilities of the MNF that are not reasonably available otherwise in Australia. You will need to demonstrate in initial discussions with the MNF Executive that the research *depends on* of use of RV *Investigator* and could not be done from a reasonably available alternative platform. You also will need to demonstrate that you have the money to pay for the use of *Investigator* to the extent agreed with the MNF. The work will need to be of a research nature, meet MNF research quality expectations, and be demonstrably in Australia's national interest, directly or indirectly. You should discuss your research with the MNF and relevant national agencies, as national interest end-users, to verify that your research will satisfy these requirements.	
Addressing Research Quality Criteria	
Assessment Criterion	**Guidance to Applicants**
Research Rationale: The reason and context for the research, including the research objectives, given Stream strategy.	Your research should be justified with reference to its alignment to *relevant* disciplinary knowledge and theory, including current research questions. The specific knowledge gaps that will be addressed and contribution to publicly available research literature, whether as journal articles or public reports, should be spelled-out. The need for multiple voyages, if sought, should be justified, including for recovery of equipment deployed previously. Research objectives should address clearly the key disciplinary questions you identify and be linked transparently to the delivery of national benefit, where relevant (see guidelines for addressing Research Benefit).
Research Rigour: Robustness of research design and soundness of the proposed methods.	You should explain the design (sampling, experimental, analyses) of your research and the methods you will use to gather data sufficiently clearly for a reviewer to be able to repeat the work, in principle. Research methods are expected to be consistent with current best-practice. Clearly indicate if any of the methods you will use are developmental or innovative and describe how you will mitigate potential risks of using such cutting-edge methods.
Research Feasibility: Feasibility of the research proposed and likelihood of promised research outputs, given available resources.	You must present a credible and efficient voyage plan that will see the field work completed, including allowing for bad weather and other contingencies, and a post-voyage plan for analysis and write-up that will see the research results published or otherwise distributed, samples processed, and data archived within a reasonable time after the voyage (usually not more than 3 years). It is expected that the bulk of data and samples gathered will be publicly available, consistent with its genesis from a national research facility, after an agreed period of confidentiality, if necessary.
Research Capability: Capabilities and capacity of the team to complete the research, including research leadership.	You will need to demonstrate that your team includes all the expertise necessary to complete the research and that you have a solid base of experience, including on research vessels, and people who have led sea-going research before. Applications that include a mix of early career and experienced researchers will be viewed favourably, especially those with clear plans for mentoring early career researchers in sea-going research leadership. The voyage leader need not be an experienced leader but you should demonstrate that you will have on-board sufficient voyage leadership experience to mentor a first-time leader. Details of research track record and publication output generally will be demonstrated by provision of CVs.
Addressing Research Benefit Criteria	

Stream 4: ***User-funded research***, for proposals that are ***in the national interest and rely on RV Investigator's specific capabilities***.	
Assessment Criterion	**Guidance to Applicants**
Benefit Rationale: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.	Proposals should establish clearly that the research proposed addresses key issues that have implications for Australia's national interests, as well as being useful to the commissioning user. The national interest might be served, for example, by contributions of knowledge relevant to policy, regulation, environmentally, socially, or economically. Ensure your proposal: • Explains why the project is in Australia's national interest, preferably with reference to national priorities, such that a reviewer can see easily that the research will fill one or more important information gaps of national interest and so justify deployment of a national research facility; • Articulates clearly your project objectives and outputs and explains how the national interests you have nominated are served by meeting those objectives.
Benefit Outputs: Alignment and utility of project outputs to articulated end-user needs and national interests.	Project outputs likely will include reports or products tailored for the commissioning user but in Stream 4 there also will need to be some outputs that fit with national interest end-users' (information) needs. You should describe expected outputs that you expect to be of national interest, preferably with at least some that are targeted specifically for use by your nominated national interest end-users. You should discuss these outputs with expected national interest end-users during proposal preparation and preferably provide supporting statements from those end-users that demonstrate that the outputs will be in the national interest.
Path to Benefit: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs.	You should explain clearly how you will deliver outputs to the commissioning user and also provide a plan for engaging with the nominated national interest end-users, being clear about the mechanisms and timetable by which you propose to engage with them. Provide convincing evidence that your delivery strategies, for both contracting user and national interest end-user, are endorsed by those users. You should aim to provide supporting documentation, including letters of support from commissioning and national interest end-users in which they: • specify the path to application of the research to be done; • attest that the research to be done is essential to their business or in the national interest (as appropriate); and • express confidence in your ability to deliver the expected outputs.
Capacity to Deliver Benefit: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.	You should provide evidence to demonstrate that your team includes people with a track-record of effective engagement with end-users, either in the sector of the commissioning user or in the national interest, including such things as; • Strong evidence of previous translation of scientific results into outputs that had demonstrable utility; • Supporting testimonials from previous research end-users with whom members of your team have worked.

Table 9: Guidance for preparation of proposals for research partnerships with the MNF in Stream 5.

Stream 5: ***Strategic Partnerships with national publicly funded programmes or institutions*** that rely on regular access to MNF capabilities ***to support data and sample collection in the national interest***.	
Research in Stream 5 is expected to address directly national priorities established through prior approval of publicly funded programs or institutions where MNF facilities are needed and for which deployment of a national research facility is appropriate. It is expected that Stream 5 research will involve recurrent deployments over multiple years through formal partnership arrangements between the MNF and other publicly funded agencies. The research will need to have demonstrably credible prospects of delivering the outcomes nominated by partner programs or organisations (end-users) as reflecting directly the national interests for which they were funded. Prospective partners should discuss their research requirements with the MNF at an early stage to verify that the research is appropriate for a partnership arrangement with the MNF. Specific arrangements for MNF support or voyage scheduling over multiple years will be resolved in the partnership agreement.	
Addressing Research Quality Criteria	
Assessment Criterion	**Guidance to Applicants**
Research Rationale: The reason and context for the research, including the research objectives, given Stream strategy.	Your research should be justified both with reference to its alignment with Stream 5 partner priorities and also with reference to relevant disciplinary knowledge and theory, including current research priorities. The need for research over multiple years should be explained thoroughly. Research objectives should address clearly the key disciplinary questions relevant to the Stream 5 partners' priorities and be linked transparently to the objectives for delivering end-user outputs (also see Research Benefit guidelines).
Research Rigour: Robustness of research design and soundness of the proposed methods.	You should explain the design (sampling, experimental, analyses) of your research and the methods you will use to gather data sufficiently clearly for a reviewer to be able to repeat the work, in principle. Research methods are expected to be consistent with current best-practice. Clearly indicate if any of the methods you will use are developmental or innovative and describe how you will mitigate potential risks of using such cutting-edge methods.
Research Feasibility: Feasibility of the research proposed and likelihood of promised research outputs, given available resources.	You are requesting granted use of MNF facilities, especially the high-valued RV *Invesigator*. It is important that you demonstrate you have credible and efficient voyage plans that will enable the field work to be completed, including allowing for bad weather and other contingencies, and a post-voyage plan for analysis and write-up that will see the research results appropriately documented, samples processed, and data archived within a reasonable time after each partnered voyage (usually not more than 3 years after each voyage).
Research Capability: Capabilities and capacity of the team to complete the research, including research leadership.	You will need to demonstrate that your team includes all the expertise necessary to complete the research and that you have a solid base of experience, including on research vessels, and people who have led sea-going research before. Applications that include a mix of early career and experienced researchers will be viewed favourably, especially those with clear plans for mentoring early career researchers in sea-going research leadership. Voyage leadership need not be by established researchers but you should demonstrate that you will have on-board sufficient voyage leadership experience to mentor a first-time leader. Details of research track record and publication output generally will be demonstrated by provision of CVs.

Stream 5: ***Strategic Partnerships with national publicly funded programmes or institutions*** that rely on regular access to MNF capabilities ***to support data and sample collection in the national interest***.	
Addressing Research Benefit Criteria	
Assessment Criterion	**Guidance to Applicants**
Benefit Rationale: Justification for the project against national policies, management requirements, national or international research priorities, or specific end-user or industry-relevant needs, with demonstrable interest to Australia and alignment of project objectives with specific information needs for the Stream.	Proposals should establish clearly that the research proposed addresses directly key questions or needs of national interest reflected in funding of the national program or institution. Ensure your proposal: • Provides a clear link to the objectives of the partnering national program or institution that reflect the national priorities for which it was funded, such that a reviewer can see easily that the research will fill one or more important information gaps of national interest; • Provides clear, realistic, and appropriate objectives that address the funded program objectives — not just for the research outputs, but for their application and uptake by the end-users for whom the identified national priorities are important (e.g., the partnering program or institution, or their funding agencies).
Benefit Outputs: Alignment and utility of project outputs to articulated end-user needs and national interests.	You should describe expected outputs for use by the partnering, or other, end-users that are explicitly targeted to adoption and application to meet the funded program objectives. These outputs might include research reports or publications but in Stream 5 will need to include more specific products that fit with partner operational needs. Project leaders should discuss these outputs with the partner end-users during proposal preparation and provide supporting statements from the partner that demonstrate that the outputs are central to addressing the needs for which they were publicly funded.
Path to Benefit: Path to benefit for end-user focussed outputs in Stream context, including post-voyage engagement with beneficiaries, mechanisms for uptake or adoption of end-user relevant products, and timetable to deliver key outputs.	You should lay out a clear plan by which the partner end-user intends to apply project outputs, including the mechanisms and timetable by which those outputs will be provided for use. Supply convincing evidence that your delivery strategy is endorsed by the partner end-user. You should aim to provide high levels of supporting documentation, including letters of support from the partner in which they: • specify the path to application of the research to be done; • attest that the research to be done is essential to the purpose for which they were funded; and • express confidence in your ability to deliver the expected outputs of benefit.
Capacity to Deliver Benefit: Demonstrated capability within the project team to deliver outputs for the benefit of Stream end-users.	You should provide evidence to demonstrate that your team includes people with a track-record of effective engagement with the partner or similar end-user, including such things as; • Strong evidence of previous translation of research into outputs that had demonstrable utility; • Supporting testimonials from previous similar end-users with whom members of your team have worked.

Analysis and Presentation of Application Assessments

d. Recommend a method or methods of analysing and presenting scores to provide transparency of process and focus discussion of results, both within assessment subcommittees and at the Steering Committee, within Streams and across all Streams.

The need to compare proposal assessments within and among Streams mainly is driven by a need to rank proposals for allocation of ship-time that is less than required for all proposals. An assumption here is that proposals deemed unsupportable will be culled on the basis of their individual merit and there remain more 'supportable' proposals than can be supported. The emphasis for analysis and presentation of assessment results, therefore, is on highlighting differences among [acceptable] proposals to inform decision-making about support, rather than to judge proposal merit. I will address the culling of unsupportable proposals later.

A corollary of an effective competitive assessment framework is that acceptable proposals receiving low final scores might not be bad proposals *per se* but just not as good as other proposals being considered. This sentiment has informed my recommendation of the following approach to analysing and presenting assessment results. The emphasis in the following is on 'spreading out' the acceptable proposals across a wide range of derived scores for merit principles (Research Quality and Research Benefit) to discriminate amongst proposals of different strengths and weaknesses. The results should not be used to infer the absolute merit of the proposals, though clearly there the two will be correlated.

Weighting Merit Principles or Assessment Criteria

The introduction of Streamed access arrangements, including identifying different beneficiaries of research in different Streams, raises a question of whether merit principles should be weighted differently for different Streams (assuming the same principles are applied to all Streams). A related question is whether the above assessment criteria should be weighted differently, either for use in all Streams or with Stream-specific weights. I discuss each question below before describing suggested analyses.

Merit Principles

Mapstone (2019) reported diverse stakeholder views about weighting MNF proposal assessment criteria,

> *... covering the full range of options from heavily favouring [national] benefit to heavily favouring 'pure science excellence'*,

though

> *... many considered that the weighting of criteria should be adjusted to fit the primary purpose of application categories, presuming that delineation of research categories is articulated in an MNF strategy (e.g., research- or technology-driven applications, strategic priority driven applications, ...)*.

He went on to say that ...

> *Resolving the diverse views is difficult but one option for doing so would be to have one score for research quality (RQ, including team research capability) ... and one score for national benefit (NB, including team capability to deliver to end users) ... and weight those scores differently for research-driven open-call applications (e.g., 60:40% or 70:30% RQ:NB), policy-driven applications (e.g., 50:50% or 40:60%), or user-funded proposals (e.g., 40:60% or 30:70%). The absolute weighting inevitably is a, somewhat arbitrary, judgment call.*"

Weighting the (now) primary merit principles of Research Quality and Research Benefit implies a desire for a final single score for ranking proposals. That might be useful ultimately but it will be more informative to present the principles initially together, not combined, as discussed below, so that the balance of proposal strengths or weaknesses against each principle can be seen clearly. Tailoring advice for proposal preparation and assessment carefully to the context of each Stream should obviate the need to weight final scores because claims for research benefit, in particular, already will have been aligned with the the purpose of each Stream. I suggest, therefore, that weighting scores for merit principles is unnecessary and might obfuscate comparisons of proposals across Streams, when necessary. I provide suggested Stream-specific weightings for merit principles below (Table 10) in case they are preferred. I recommend they not be applied, however, or, if they are, they be applied only as a final step to collate overall rankings of proposals.

Recommendation

7. The MNF weight equally the overarching merit principles when considering proposals in all Streams, given appropriately targeted proposal and assessment guidelines for each Stream.

Table 10: Suggested weighting, if required, for merit principles in final MNF proposal assessment.

	Weighting	
Stream	**Research Quality**	**Research Benefit**
1: Strategic priority research	0.3	0.7
2: Discipline-driven research	0.7	0.3
3: Technology-driven research	0.4	0.6
4. User-funded research	0.5	0.5
5: Publicly-funded partnerships	0.3	0.7

Assessment Criteria

Weighting amongst the assessment criteria for each merit principle also might be favoured, and was recommended by Mapstone (2019), but also would be fairly arbitrary and likely to complicate unnecessarily the assessment process. It might be argued that failure against any criterion would be equally concerning for the successful completion of a project. A well-justified research proposal that was based on flawed design and methods, for example, would be unlikely to deliver robust research results or, probably, the expected end-user benefits. Research that was well conceived but proposed by a team that lacked appropriate expertise would carry a high risk of failing to meet expectations. A project with well-framed outputs but very poor capacity to deliver them or poorly conceived mechanisms to do so likely would struggle to realise expected benefits. The rationale for setting different weightings for different criteria accordingly is unclear.

I provide some suggested weights for Research Quality and Research Benefit criteria below (Table 11) in case they are desired by the MNF but recommend equal weighting of all criteria, given appropriate Stream-specific advice to applicants and assessors. The weights in Table 11 inevitably are arbitrary but are based in my judgment of relative risks to satisfactory project completion and delivery of benefit arising from poor performance against each criterion, and the prospect that deficiencies can be resolved in response to review without wholesale reworking of the project. The weights would be the same for assessments in each Stream.

Recommendation

8. The MNF weight assessment criteria equally within each merit principle for analysis of assessment results.

Table 11: Suggested weighting, if required, for assessment criteria for research quality and research benefit.

Merit Principle	Criterion	Weighting
Research Quality	1: Research Rationale & Objectives	0.25
	2: Research Design & Methods	0.35
	3: Research Feasibility	0.2
	4. Research Team	0.2
Research Benefit	1: Benefit Rational & Objectives	0.2
	2: Benefit Outputs	0.3
	3: Path to Benefit	0.3
	4. Capacity to Deliver Benefit	0.2

Analysing Scores

The following analyses are adapted from an approach developed in the 1990s by CSIRO[8] and emphasise graphical presentation of results to illustrate standing of projects along two axes at each step of evaluation. I include illustrative graphics from a set of scores I invented for the purpose. The approach is similar for the Research Advisory Committee (RAC) and National Benefit Assessment Panel (NBAP) and so I provide a single general description where possible to avoid simply repeating material for each panel. Analyses of scores against Research Quality (RQ) criteria are prepared for the RAC and those against Research Benefit (RB) criteria are prepared for the NBAP. Generic references to 'four criteria' and 'assessment scores' therefore should be read as applying to either RQ criteria or RB criteria unless otherwise specified. Analysis and presentation to the Steering Committee (SC) , after panel assessments, is described separately.

[8] CSIRO (1991). CSIRO Priority Determination 1990: Methodology & Results Overview. Canberra.

Analysis and Presentation for Assessment Panels

The analysis for the assessment panels has the following steps, elaborated below, once assessors scores have been received:

1. Sum the two sub-criterion scores (out of 5) for each criterion to produce a score out of 10 for the relevant criterion for each assessor of each proposal;
2. Combine the scores for RQ or RB criteria 1 and 2 into a single score and combine scores for RQ or RB criteria 3 and 4 similarly, again for each assessor of each proposal;
3. Average each of the two sets of combined scores for RQ or RB (considered separately) across assessors of each proposal, identify maxima and minima of each set, and calculate score standard errors (SE), or other measure of uncertainty, for that proposal;
4. Plot the proposals' mean scores and either ranges or SEs on two axes representing the combined criteria scores (criteria 1 & 2 on one axis, criteria 3 & 4 on the other);
5. Use the plots as a guide to focus discussion by the relevant assessment panel (RAC, NBAP) of the strength of support for each proposal, including exploration of the underlying assessments as appropriate.

These steps are described in more detail below.

Step 1: Calculating Criterion Scores

I recommend step 1 involves summing the sub-criterion scores but they could be combined by multiplication and scaling to also deliver a score out of 10. Multiplication would mean that any score of zero for a sub-criterion (question) would result in that criterion ending up with a score of zero, negating any credit against the other sub-criterion question (presuming it was scored one or more). That approach implicitly infers that the sub-criteria are essential 'bars to clear' for any proposal. Sub-criteria questions are intended to focus assessments rather than prescribe essential elements of 'success' and so such an emphasis on sub-criteria is not what was intended for them. Hence, I recommend that the sub-criteria scores be summed, not multiplied and scaled, to produce criterion-level scores.

Step 2: Combining Pairs of Criterion Scores

The second step is a key step in preparing scores for presentation and discussion by the RAC and NBAP.

The four RQ criteria address two qualitatively different aspects of RQ for each proposal. The Research Rationale and Objectives (RQ criterion 1) together with Research Design and Methods (RQ criterion 2) go to the heart of the conceptual and technical strength of the proposed research, irrespective of who is doing it or whether it is feasible on the proposed voyage. These two criteria together can be used to derive a score for what I'll term 'Research Concept'. The remaining two RQ criteria (Research Feasibility and Research Team) deal with whether the proposed resources and team capability are sufficient to provide confidence that the research will be executed successfully. They can be used together to score 'Research Execution'. Research Concept and Research Execution form the two main dimensions by which the RAC might consider RQ, They will be referred to generically as 'Axis 1' and 'Axis 2' respectively in the following discussion.

Similarly, the four RB criteria address two qualitatively different aspects of RB for each proposal. The Benefit Rationale and Objectives (RB criterion 1) together with End-user Outputs (RB criterion 2) go to the heart of the alignment of the proposed project with end-user needs. These two criteria together can be used to derive a score for what I'll term 'Benefit Need'. The remaining two RB criteria deal with whether there is a credible path to benefit and the relevant experience in the project team to deliver outputs to end-users to enable the expected benefit. These criteria can be used to score 'Benefit Delivery'. Benefit Need and Benefit Delivery form the two main dimensions by which the NBAP might consider RB. They will be referred to generically as 'Axis 1' and 'Axis 2' respectively in the following discussion.

This step involves combining the relevant pairs of criteria into a single score for Axis 1 and Axis 2 for RQ and RB evaluation. The criterion scores could be combined as a simple sum but I recommend multiplying them to accentuate differences among projects when plotted. Any adjustments to scores to account for weighting criteria differently (e.g. Table 11) should be applied before the multiplications (or additions, if preferred). The derived scores for the RAC axes, therefore, will be:

$$\text{Research Concept} = C1_{RQ} * W_{RQ1} * C2_{RQ} * W_{RQ2};$$

$$\text{Research Execution} = C3_{RQ} * W_{RQ3} * C4_{RQ} * W_{RQ4};$$

Where

$C1_{RQ}$, $C2_{RQ}$, $C3_{RQ}$, and $C4_{RQ}$ are the scores for Research Quality assessment criteria 1, 2, 3, and 4 respectively; and

W_{RQ1}, W_{RQ1}, W_{RQ1}, W_{RQ1}, are the weightings applied to RQ criteria 1, 2, 3, and 4 respectively, if criteria are weighted differently[9].

The derived scores for the NBAP axes, analogously, will be:

Benefit Need = $C1_{RB}$*W_{RB1} * $C2_{RB}$*W_{RB2};

Benefit Delivery = $C3_{RB}$*W_{RB3} * $C4_{RB}$*W_{RB4};

Where

$C1_{RB}$, $C2_{RB}$, $C3_{RB}$, and $C4_{RB}$ are the scores for Research Benefit assessment criteria 1, 2, 3, and 4 respectively; and

W_{RB1}, W_{RB1}, W_{RB1}, W_{RB1}, are the weightings applied to RB criteria 1, 2, 3, and 4 respectively, if criteria are weighted differently.

The criterion-level scores from each reviewer, including members of the relevant assessment panel, are treated independently at this step. The results of the multiplications are scores out of 100 on each axis from each assessor, presuming the multiplication method above is used[10].

The role of zero scores for criteria should be considered in this step in which I recommend multiplying pairs of criterion scores. Retaining any criterion-level score of zero (meaning that proposal scored zero against both sub-criterion questions) will 'zero-out' any value attributed to the other criterion in the pair. That would be appropriate if it is decided (as recommended later, refer '*Unsupportable Proposal*s') that a zero score against any criterion would render a proposal unsupportable. It might be preferred, alternatively, that the value attached to the non-zero criterion in a pair is retained for later consideration. That result can be achieved in two ways: either sum (and scale) the criterion-level scores for each axis, or assign a value of 1 to zero-scored criteria at input to this step, before multiplication with the other score in the pair. Either option will tend to diminish the 'spreading' effect of multiplying criterion scores but the latter will affect only proposals with low (zero) scores against any criterion. I recommend that the raw scores not be changed in such an adjustment, but retained as a clear record that a project had been considered unsupportable against that criterion, notwithstanding the score adjustment for analysis[11].

Step 3: Calculate Axis-score Means and Variation

The axis 1 and axis 2 derived scores are averaged across assessors, or subgroups of assessors if required (e.g., to consider peer reviews and assessment panel reviews separately). The result is the mean of all assessors' derived 'axis scores' for each proposal.

It also will be useful to identify the lowest (harshest assessor) and highest (most generous assessor) score on each axis for each proposal, and calculate a standard measure of uncertainty for the mean (I suggest the Standard Error, SE). These range and uncertainty measures will inform discussion of the proposal assessments at the RAC or NBAP.

Step 4: Plot Projects by their Axis Scores.

The mean axis 1 and axis 2 scores are plotted against the relevant axes for consideration by the relevant panel. It is useful to plot some measure of variation in the scores, whether range or SE, to illustrate the degree to which assessors had coherent or disparate views of a proposal and to inform panel discussion. Results can be plotted by Stream or for multiple Streams to compare assessments among Streams.

These plots should be prepared by the MNF prior to panel meetings and it is assumed, therefore, that all reviewers, including panel members, will have provided their scores well before the meeting. It is important that individuals' scores are not influenced by panel discussions, notwithstanding that the final standing of any project might be modified by the panel at its meeting (see below). Figure 1 shows example plots for Research Quality or Research Benefit criteria for individual streams and Figure 2 shows combined plots of Research Benefit criteria for all proposals from multiple streams.

---

[9] The formulation of the weight-factor will depend on how weights are framed. The weight-factors for the proportional weightings shown in Table 11, where the weights sum to 1 across the 4 related criteria, is

$WF_{Cn} = PW_{Cn} * 4$,

where

WF_{Cn} is the weight factor applied to the raw scores for criterion *n*; and

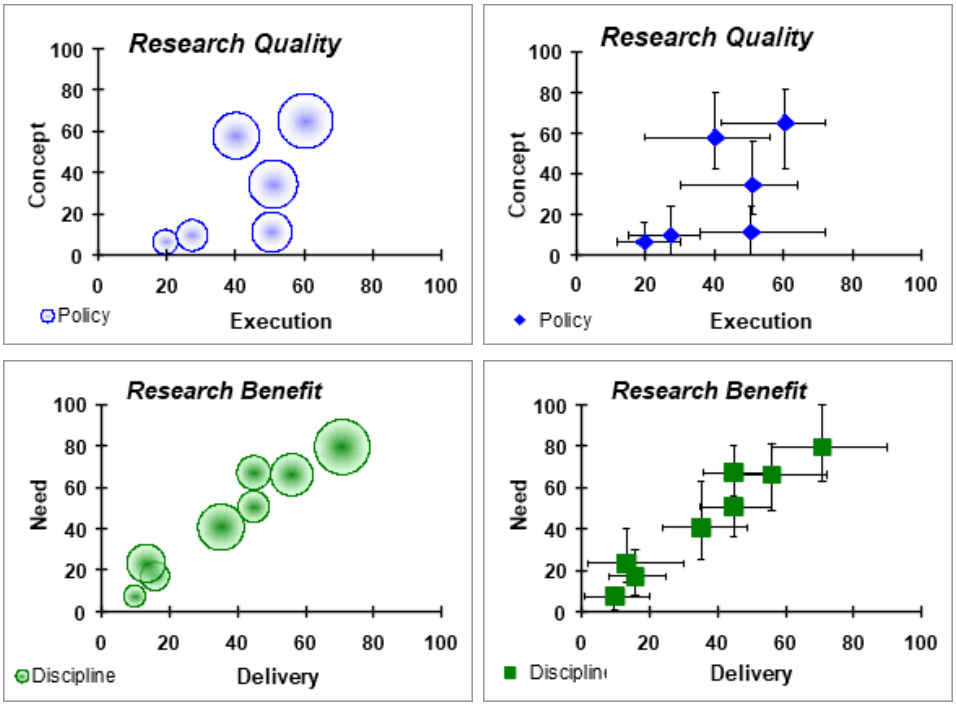
PW_{Cn} is the proportional weighting of Criterion *n*, such as those indicated in Table 11.

All criteria have been weighted equally (All WE = 1) in the example analyses presented here.

[10] The criterion scores can be summed instead of multiplied if preferred, but doing so will diminish the 'spreading' effect of this step and result in more 'clumped' axis scores. Summed scores will have a potential range of 0–20 on each axis. Scores can be scaled to a range of 0–100 to provide a consistent view across this a later steps, if desired.

[11] I have not applied any adjustments to remove zero scores in the examples presented in this report.

a. Criterion scores multiplied to produce axis scores.



b. Criterion scores summed & scaled to produce axis scores (out of 100).

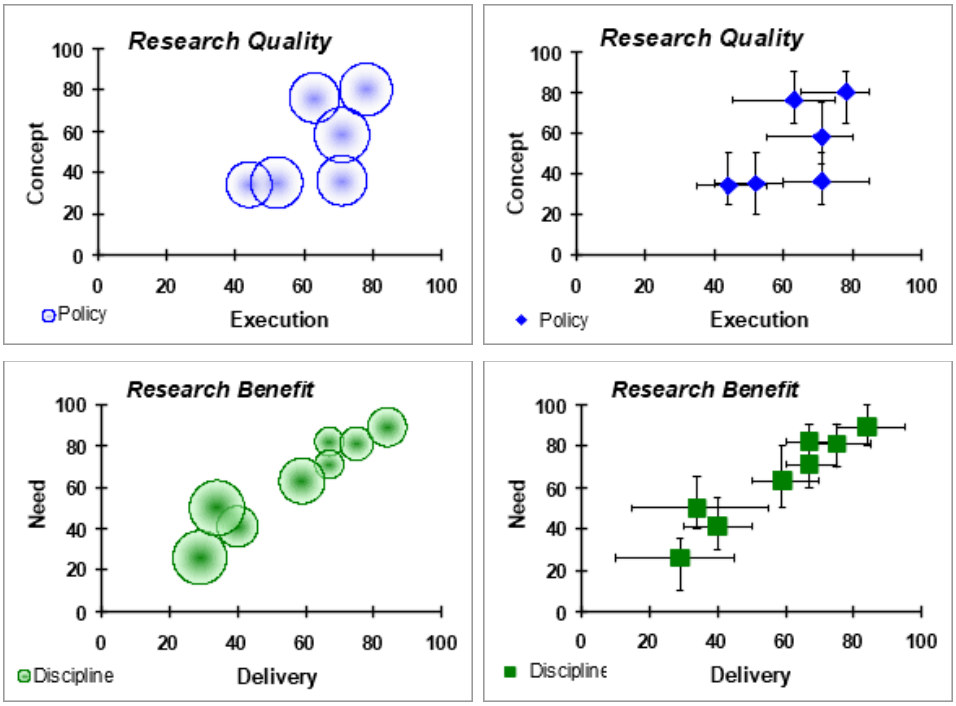


Figure 1: Plots of average combined criteria on Axis 1 (Research Concept or Benefit Need) and Axis 2 (Research Execution or Benefit Delivery) based on scoring against Research Quality or Research Benefit criteria 1 and 2 or 3 and 4 respectively. Each point represents results for a single proposal. Bubble plots (left) show approximate standard errors (mean of SE on both axes) of assessments around each point and error bars on right plots indicate maximum and minimum scores for each proposal on each axis. Examples are shown for Research Quality criteria (from some hypothetical Stream 1 — policy-driven proposals, blue plots) and Research Benefit (from some hypothetical Stream 2 — Discipline-driven proposals, green). The top 4 plots (a) show results calculated by multiplying relevant criterion scores to derive axis scores and the lower 4 plots (b) illustrate results of summing and scaling the same criterion scores. Multiplications tend to separate proposals more than summing on each axis, potentially reducing ambiguity in comparisons.

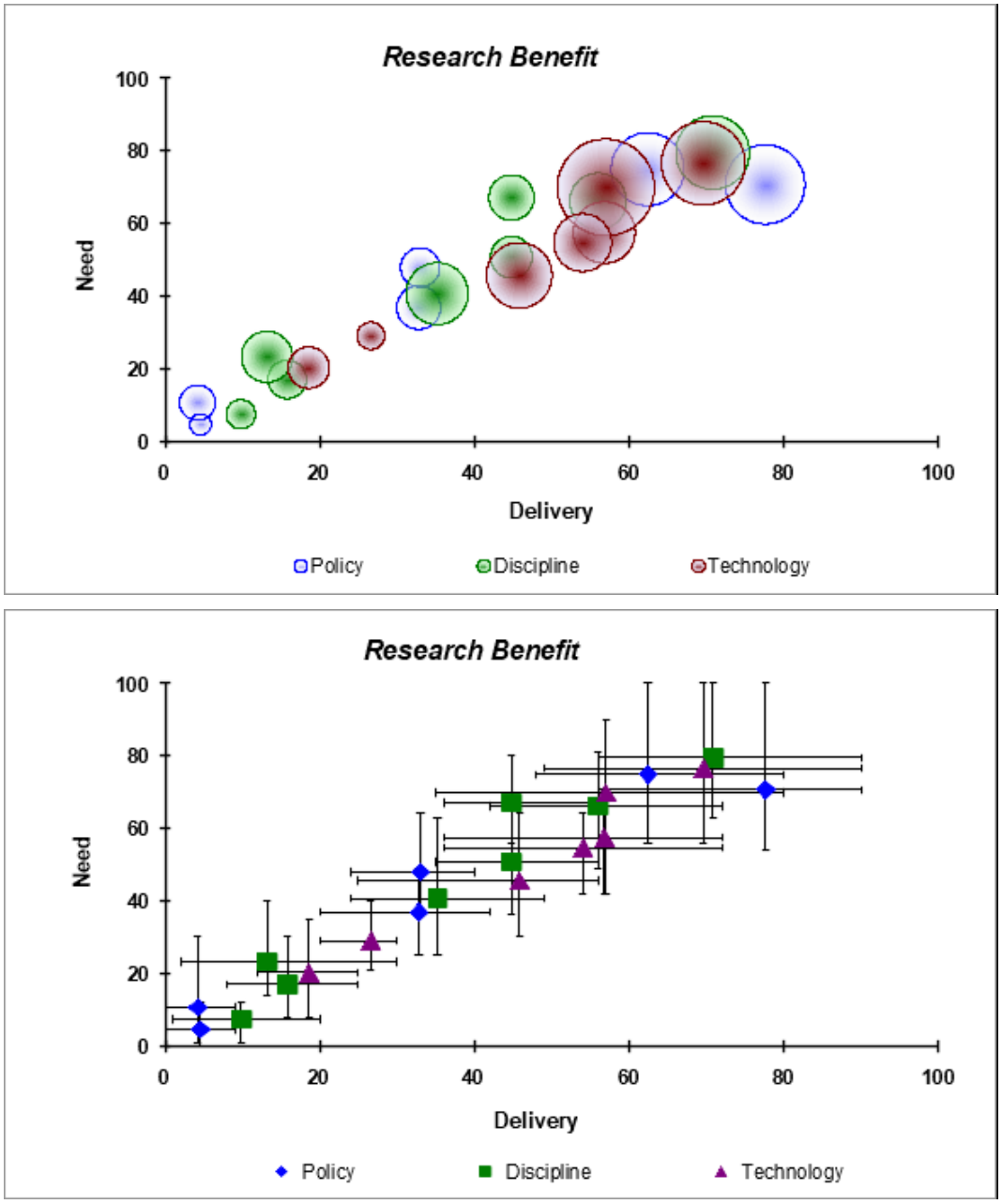


Figure 2: Plots of Benefit Need against Benefit Delivery (potential) for multiple projects from three MNF allocation Streams: Stream 1 (Policy-driven proposal, Blue), Stream 2 (Discipline-driven proposals, green), and Stream 3 (Technology-driven proposals, red). Each point represents results for a single proposal. Bubble plots (left) show the approximate standard error (average of SE on both axes) of assessments around each point and error bars on right plots indicate maximum and minimum scores for each proposal on each axis.

Step 5: Use of Plots by Assessment Panels.

The intention of the plots (e.g., Figs. 1, 2) is to highlight the relative strength of proposals along each axis. Proposals in the top-right of the plot generally would be inferred to be very strong and the first candidates for support, based on either RQ or RB, depending on for which panel the plot is prepared. Those in the bottom-left would be those least deserving of support. Projects toward the top-left or bottom-right of the plot normally would be those where most discussion might be focussed because they were strong on one axis and weak on the other, perhaps indicating a need to delve into the underlying scores and reviewers' comments to assess the seriousness of the deficient area. Proposals falling in the centre region of the plot also might warrant further interrogation, though they likely are there simply because they were assessed as solid, but probably not outstanding, on both axes. Discussion also might be guided by the spread of scores around the means. Projects for which scoring clearly was inconsistent among reviewers probably would warrant further investigation of the underlying differences among reviews. It should not be inferred that

projects scored highly or poorly on both axes need no discussion, but the arrangements of projects on the plot should be used to guide where discussion is most likely to be necessary to resolve marginal or uncertain cases or to review proposals flagged as 'unsupportable' by the MNF secretariat (refer Unsupportable Proposals, below).

The above analysis and presentation can be done for each Stream separately but the plot also can capture results from multiple streams (per Fig. 2). Use of the same underlying scoring framework provides a coherent and consistent basis on which to compare assessments across Streams, notwithstanding the Stream-specific guidance to scoring. Comparison among Streams might be required for several reasons, including to review the distribution of scores from different Streams to assess the likelihood of Stream-specific scoring biases or whether proposals are consistently stronger in some Streams than others (if Stream-specific biases are small). Applying Stream-specific weightings for criteria will complicate cross-Stream comparisons because those Stream-specific weightings will move projects' raw scores across the plot in different ways for different Streams. I therefore recommend that any criterion-specific weighting is applied to all Streams and Stream-specific weightings, if desired, are reserved for later application.

The RAC or NBAP might resolve in discussion, and with reference to the underlying assessment scores, that some proposals are misplaced on the plot. That judgment might arise, for example because of one or more aberrant score(s) or because panel discussions identified strengths or weaknesses that apparently had been missed or misinterpreted by some reviewers. The analyses I suggest are not intended to preclude such panel judgments but to inform the discussion that precedes them in a structured and consistent way, both within panels over time and between panels. I recommend, however, that judgment calls by either panel do not result in changes to the original scoring but are instead added to a proposals portfolio of scores as a set of separate 'panel scores' with annotations against project assessments to record what adjustments were recommended and why. It is important that the original assessments are retained in-tact in case further exploration is required and for comparisons with assessments in other application cycles to inform improvements in assessment processes.

Analysis and Presentation for the MNF Steering Committee

The above steps are about guiding discussions by the RAC and NBAP in forming recommendations to the MNF Steering Committee (SC). The next step in analysing assessment scores is to summarise scores from the panels to accompany their recommendations for presentation and discussion by the SC. This step involves distilling scores for each proposal from each assessor and the two assessment panels (if required, per above) for the primary merit principles (Research Quality, RQ, and Research Benefit, RB). The analyses and presentations here are analogous to those for the assessment panels but applied at a 'higher level'.

Derived scores for each proposals and assessor plotted on the two axes for the RAC or NBAP (Figs. 1, 2) are combined into a single score for RQ or RB respectively. There again are multiple ways in which that combination can be done (e.g., additively, averaging, multiplication) but I recommend that the position of each project on the respective RAC or NBAP plots be represented by its distance from origin of each plot. That is, the RQ or RB scores will be calculated as

$$RQ|RB = \sqrt{\left(Axis1^2 + Axis2^2\right)}$$

Where:

RQ|RB are either the overall score for Research Quality or Research Benefit derived from each assessor and RAC or NBAP respectively (if required),

Axis 1 is the proposal's derived score for the relevant Axis 1 (Research Concept or Benefit Need), and

Axis 2 is the proposal's derived score for the relevant Axis 2 (Research Execution or Benefit Delivery).

The results of this step will be scores in the range 0–141.4 for both RQ and RB from each assessor. These scores can be scaled to the range 1–100 on each axis for consistency with other steps but there is no other reason to do so. The proposals are then plotted against two axes representing Research Quality and Research Benefit with the derived RQ and RB scores positioning the proposals in the plot. I recommend that any Stream-specific weighting of Research Quality or Research Benefit not be applied prior to plotting the results at this stage but, if required, later when deriving an overall score for each proposal (below).

Example plots for some hypothetical proposal scores are shown in Figure 3 for individual Streams, and Figure 4 for all Streams together.

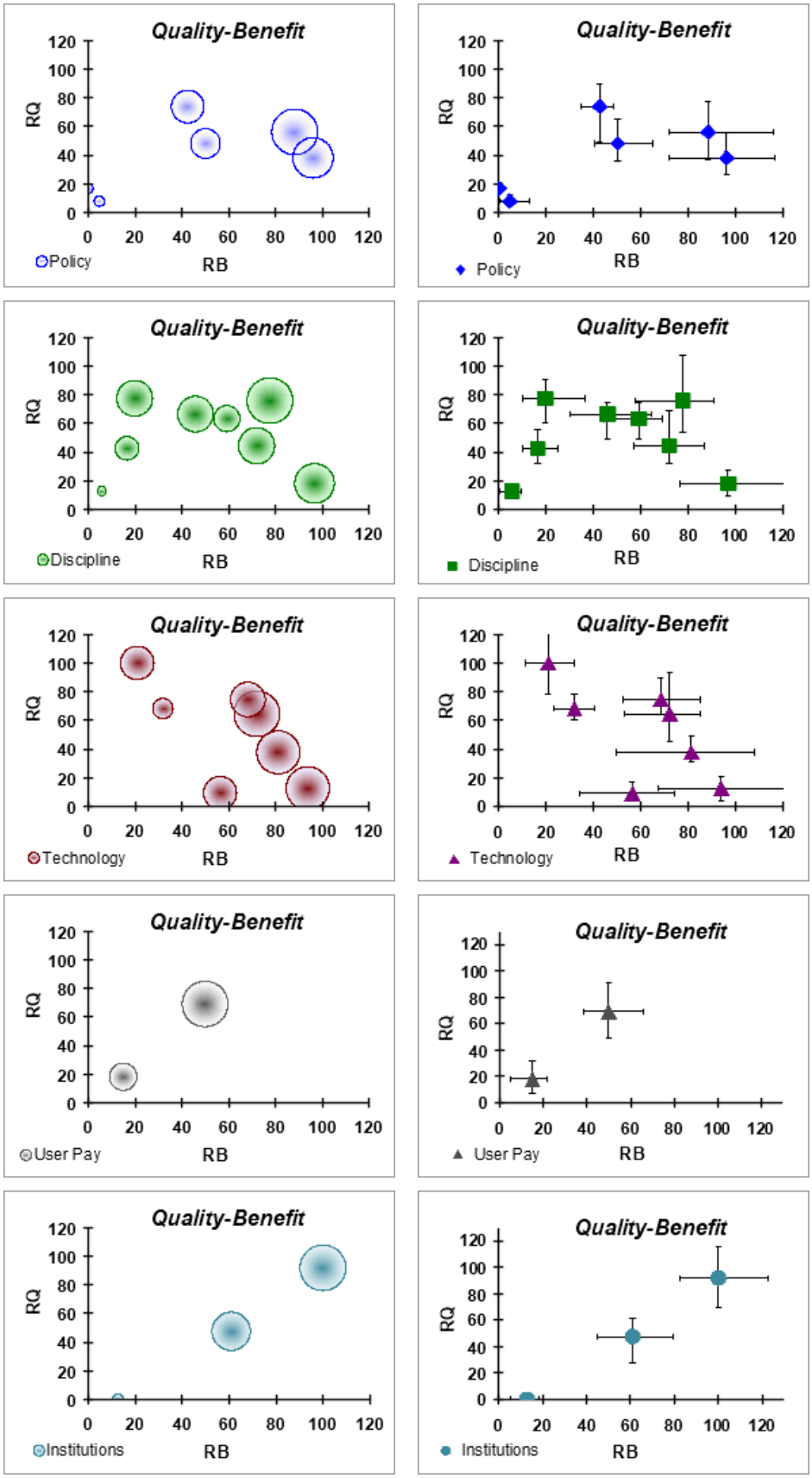


Figure 3: Plots of derived overall scores for Research Quality (RQ) and Research Benefit (RB) for each proposal, scaled to 0–100 and plotted by Stream (Streams 1–5 top–bottom). Bubbles (left) show approximate Standard Errors around mean scores (centres of bubbles). Right plots show mean scores for each proposal with assessors' maximum and minimum derived scores on each axis.

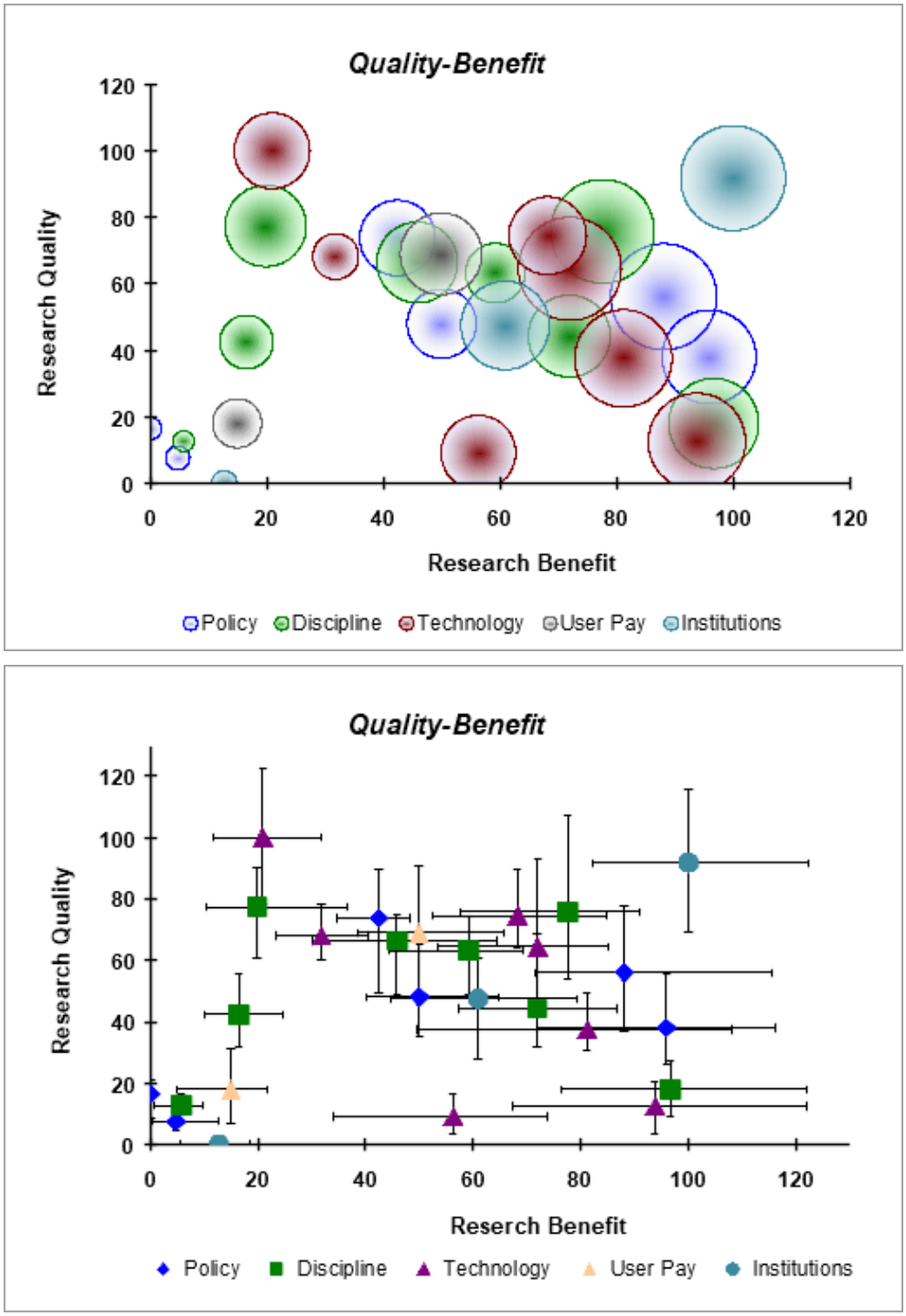


Figure 4: Plots of derived overall scores for Research Quality (RQ) and Research Benefit (RB) for each proposal, in all Streams. (Stream 1 — Blue, Stream 2 — Green, Stream 3 — Red, Stream 4 — Grey, Steam 5 — Teal). Bubbles in the top plot indicate approximate Standard Errors (averaged over both axes) around scores, with mean scores at the centre of each bubble. The lower plot shows the mean scores of each proposal on each axis (symbols) with error bars showing assessors' maximum and minimum derived scores for each proposal.

Criterion-level scores from each reviewer, including members of the relevant assessment panel, are treated independently to reach this stage. Scores are averaged across assessors and the means plotted for consideration by the SC. Scoring variation among assessors is plotted to inform discussion by the SC. It should be noted that reviewers of proposals for RQ will not be the same people who reviewed proposals for RB, so plotted values will not be traceable to the same individuals on both axes.

The intention of the plots in Figures 3 and 4 is to highlight the relative strength of proposals along each RQ

and RB axis. Proposals in the top-right of the plots would be inferred to be strong against both merit principles, and so the first candidates for support. Those in the bottom-left would be those least deserving of support, *prima facie*. Proposals toward the top-left or bottom-right of the plot normally would be those where most discussion might be focussed because they were strong on one axis and weak on the other, indicating that some additional consideration might be warranted if they were candidates for support if ship-time was available in the corresponding Stream. Proposals falling in the centre region of the plot also might warrant further interrogation, though they likely are there simply because they were assessed as solid, but not outstanding, on both axes. Discussion also might be guided by the spread of scores around the means. Proposals with disparate assessments might prompt further investigation before committing MNF support to them. It should not be inferred that projects scored highly or poorly on both axes need no discussion, but the arrangements of projects on the plot should guide where SC discussion is most likely to be necessary.

A companion plot with requested ship-time listed at the position of each proposal on the plot (Figure 5) can be used to inform scheduling implications of alternative approvals of support. Allocation of sea-days might be expected to work from the top-right back toward the bottom-left within a Stream. Seeing the requested sea-time in in the context of each proposals position on the RQ-RB plot might be useful especially at the margins where available sea-time for a Stream is being exhausted. The RQ-RB context also might be useful where a Stream is under-subscribed and surplus sea-time is being allocated to proposals in other Streams.

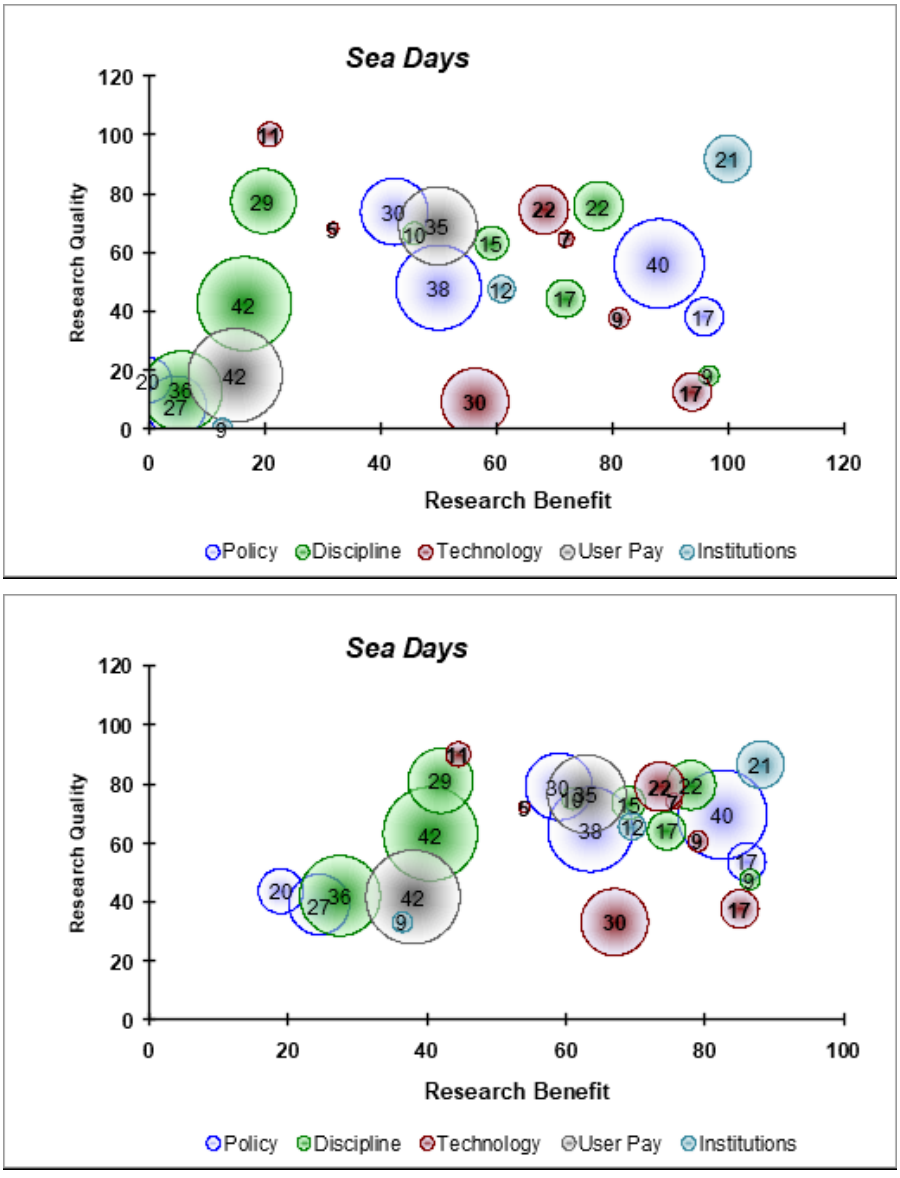


Figure 4: Plots of requested sea-days for each proposal in each Stream. The position at the centre of each bubble reflects the score of a proposal on the RQ and RB axes whilst bubble diameters represent the ship-times requested by the proposal. Requested ship-time (in days) also is listed at the centre of each bubble. The top plot is for RQ and RB results calculated per text, whilst the bottom plot is of results calculated by simply summing and scaling the scores at each step.

The application of Stream-specific weighing to RQ and RB scores has been discussed above. Applying Stream-specific weightings prior to presenting RQ–RB plots will complicate cross-Stream comparisons on plots because those Stream-specific weightings will move proposals' derived scores across the plot in different ways for different Streams. I therefore recommend that any Stream-specific weightings of RQ and RB, if desired, be applied after consideration of the above plots.

Recommendation

9. The MNF adopt the recommended methods for analysis of assessment results and presentation to the Research Advisory Committee, National Benefit Assessment Panel, and MNF Steering Committee.

It might be desirable for the SC to see a set of single scores for the proposals that facilitates easier ranking of proposals within or across Streams. A single final score could be calculated in various ways but the two most obvious likely are to either sum, or average, the derived RQ and RB scores or calculate the distance of each proposal from the origin on the RQ-RB plots:

$$FinalScore = \sqrt{\left(RQ^2 + RB^2\right)}$$

Neither approach conspicuously is better than the other and both will result in ambiguity in ranking proposals. Proposals with RQ:RB scores of 20:70 or 70:20 will have the same final score by either method but have conspicuously different strengths that might affect their allocation of ship-time, especially if from different Streams. Each method could deliver the same single result from numerous other RQ–RB pairs. I suggest, therefore, that the allocation of ship-time be with reference to the standing of proposals on the RQ–RB plots rather than defaulting to ranks based simply on calculating a final composite score.

I don't anticipate that the SC will review or revise the final RQ or RB scores of individual proposals but rather concentrate on which proposals are allocated sea-time given their assessments, recommendations from the RAC and NBAP, and advice from the MNF Executive about scheduling constraints within and among Streams. It will be valuable over successive allocation cycles, however, to capture some commentary from the SC about the utility of approach and its performance across Streams. Periodic review by the MNF of the approach also should be considered to assess whether it is or remains appropriate across Streams and, if so, whether there is evidence of shifts or issues in assessment characteristics that might warrant changes to either criteria or scoring guidance.

Recommendation

10. . The MNF Steering Committee allocate ship-time with explicit reference to proposals' assessments against both Research Quality and Research Benefit rather than based on the ranks of a combined (RQ+RB) score.

Unsupportable Proposals

Setting the minimum standard that must be met by proposals inevitably is somewhat arbitrary but should be based on some transparent and defensible 'standards' that are applied relatively impartially and consistently. It seems unlikely that the standing of proposals relative to other proposals would be a reasonable basis on which to deem a proposal unsupportable against the MNF merit principles, notwithstanding that comparative standing will influence whether a proposal is supportable operationally.

The scoring guidance for initial merit assessment clearly has been drafted with 'setting a minimum standard' in mind, given inclusion of a scoring category of 'unsupportable'. The intent of scoring a proposal as unsupportable (score 0) is to state clearly that the proposal could not fulfil, even to a low or mediocre standard, the expectations of a component of a selection criterion. The accumulation of one or more zero scores, therefore, provides a foundation for ruling a proposal out of further consideration for MNF support.

There are several issues of granularity related to using scores to set an exclusion rule:

1. The level (sub-criterion questions, criteria, aggregated scores) at which low scores would be considered an appropriate basis for proposal disqualification;
2. How many very low scores at the appropriate level would justify disqualification;
3. How many of the multiple assessors would need to give a proposal very low scores to justify disqualification;
4. Where in the assessment process should a decision to reject a poor-quality proposal sit — with the MNF Executive after initial reviews, with assessment panels, or with the Steering Committee.

Assessment level for proposal rejection

Rating a proposal as 'unsupportable' (score = 0) against any sub-criterion question might be considered sufficient to indicate rejection of that proposal. It might be argued, however, that a proposal that had, for example, compelling design (5) but unsupportable methods (0) was redeemable by changing the methods. Alternative rules might stipulate that scores of zero (out of 5) against two, three, or more sub-criterion components should be the basis of rejecting a proposal outright. Any 'rejection rule' based on sub-criterion scores, however, effectively treats the sub-criteria as essential 'criteria' for support, which is not the intent of posing specific questions within each criterion to guide scoring.

A proposal that fails completely to meet any one criterion for either Research Quality or Research Benefit likely would be more flawed fundamentally. Such a flaw would require significant redrafting or supplementary provision of essential information and, probably, formal reassessment. Providing opportunity for extensive revision of such poorly scored proposals raises issues of equity amongst proponents, unless all proponents were allowed to revise (substantially) their proposals, which in turn might require reassessment. Redrafting and reassessment goes beyond clarification of questions by way of rejoinder to reviewers' comments.

Accumulating criteria scores to overall research quality or research benefit scores inevitably will obscure the detail of assessments and 'hide' poor or exceptional performance against individual criteria. The net effect, depending how the aggregation is done, might be that only proposals that score very poorly against multiple, perhaps most, criteria will end-up with very low scores overall. That effect might be a reasonable argument to use some level of aggregated score as the basis for proposal rejection because only systemically flawed proposals will fail outright. The opposite argument is that using aggregated scores to set exclusion rules inevitably will mean accepting proposals that have serious flaws in one or more aspects (criteria) considered important, so representing a reputation and investment risk for the MNF.

There is no obviously right answer here, but I suggest that the criterion level, rather than sub-criterion or aggregate levels, is the appropriate granularity for rejecting proposals outright, if at all.

Recommendation

11. The MNF rejects outright any proposal that fails to meet minimum standards judged against individual assessment criteria rather than higher-level, aggregated ratings.

Minimum scoring standards for acceptance of proposals

I suggested above that even a single criterion score of zero (proposal considered 'unsupportable' against both sub-criteria) should be sufficient to exclude a proposal because that rating would indicate a material flaw. A similar argument might apply for proposals that scored 1 against a criterion (0+1 for sub-criteria), because at best only one of the sub-criteria was considered 'poor' and the other 'unsupportable'. Criterion-level scores of 2 or 3 could arise only from sub-criterion scores of 0+2, 0+3, 1+1, or 1+2, meaning that the proposal at best had sub-criteria pairs assessed as 'acceptable' or 'good' and 'unacceptable', 'acceptable' and 'poor', or 'poor' for both. It might be argued that a proposal so assessed against only one criterion but that scored higher against all others (>=4 out of 10) could be substantiated by rejoinder comments but that such scores for multiple criteria would indicate more systemic weakness.

I recommend, therefore, that any proposal that is scored overall zero or one for any criterion or that has more than one criterion scored 3 or less (out of 10) should be rejected. These rules would exclude the prospect of supporting any proposal that was considered 'poor' against any assessment criterion. It follows that proposals with at least a score of 2 or 3 (out of 10) against just one criterion and 4–10 against the remaining seven could, in principle, gain MNF support.

Recommendation

12. The minimum standard of assessment scores for potential MNF support be that a proposal has no score of less than 2 against any criterion and no more than one criterion scored 2 or 3 out of 10.

Multiple assessors and minimum standards

The use of multiple assessors, either within assessment panels or externally, is a desirable feature of the MNF application and assessment process. It might be expected that there would be some consistency among assessors in scoring proposals, especially perhaps for proposals that were conspicuously weak. It is equally unlikely, however, that there would be scoring unanimity across assessors, even for very poor

proposals. It will be important, therefore, to be explicit about what frequency of very low scores from multiple assessors would be sufficient to determine a proposal 'unsupportable'.

The number of assessors who complete reviews of each proposal will vary, so it probably makes most sense to base any inference based on collective assessments on a proportion of assessors that rate a proposal below some level, rather than on some absolute (arbitrary) number of low scores. There clearly, again, is no 'magic proportion' here but I suggest that if a proposal would fail by the above criterion-level rules for half or more of assessors then that project would represent a high support-risk for the MNF, and should be rejected. A proposal that would be 'failed' by fewer than half of its assessors' ratings also ultimately might be considered unsupportable but that conclusion perhaps should be based on some further consideration by the relevant assessment panel or SC.

Recommendation

13. The MNF rejects outright any proposal that would fail the minimum standard of criterion-level scoring for half or more of the assessors and review carefully any proposal that would be considered a 'fail' by fewer than half of assessors.

There is a risk in the above rules that projects that were considered supportable but were weak in some aspects could be allocated ship-time if available sea-time was under-subscribed. That result might represent a reputation risk for the MNF. I recommend, therefore, that any projects with one or more criterion scores of 4 or less following consideration of rejoinders be granted access, if available after awarding higher-ranked proposals, conditional on satisfactory resolution of the issues that precipitated the low score(s).

Recommendation

14. Any projects with scores of 4 or less, averaged over assessors, for any criterion (criteria) be granted access, if available, conditional on satisfactory resolution of the issues that precipitated the low score(s).

Process for determining proposals 'unsupportable'

Ultimate responsibility for endorsing or declining MNF support for a project vests with the Steering Committee (SC) or its delegate, typically the Chair. It follows that exclusion of proposals during the assessment process would inappropriately circumvent SC decisions.

The RAC and NBAP have primary responsibility for evaluating proposals' standings technically and operationally and so are probably best-placed to judge whether a proposal has been assessed fairly against the relevant criteria. It is important, however, that conspicuously weak proposals do not consume panel time that would be spent more productively considering other proposals. I suggest therefore, that the MNF secretariat provide a triage role and flag proposals that *prima facie* would fail the minimum standards for MNF support (above), prior to RAC or NBAP meetings. Those proposals then should be reviewed by the panels with particular attention to possible assessment inconsistencies or inequities that had unfairly rendered them unsupportable. The panels then would recommend to the SC whether a project's failure to meet minimum standards should hold or, alternatively, that a proposal should be considered further for allocation. Projects that were considered unsupportable by the RAC or NBAP nevertheless should be included in subsequent analyses and presentation to the SC for a final decision, but clearly flagged as 'considered unsupportable'.

Acknowledgements

I thank Barbara Musso and Ben Rae for early discussions about the project and Barbara for her insightful comments on an early draft of the report. I especially thank Gina Mercer for her willingness to wade through the report and provide helpful comments from a completely MNF-independent perspective.