http://ejournal.stipjakarta.ac.id

METEOR STIP MARUNDA

pISSN : 1979 – 4746
eISSN : 2685 - 4775

Maritime Institute of Jakarta

Ten Years of Tol Laut: A Policy Evaluation of Indonesia's Subsidized Inter-Island Shipping Program

[1] Agung Kwartama

[1] Jakarta Raya National Maritime Academy (AMAN JAYA), North Jakarta, Indonesia

*corresponding email: agung.kwartama@amanjaya.ac.id

Submitted : 20/01/2026 **Revised** : 07/02/2026 **Accepted** : 27/05/2026 **Published** : 30/06/2026

ABSTRACT

Indonesia's Tol Laut (Sea Toll) program has subsidised regular container shipping to remote and underserved ports for a decade, yet no published evaluation has assembled the scattered evidence on its performance into a single structured account using a recognised evaluation framework. This study conducts a desk-based policy evaluation of Tol Laut using the OECD Development Assistance Committee's five evaluation criteria — relevance, effectiveness, efficiency, impact, and sustainability — applied to government route-performance reporting and the published academic literature on the program and comparable archipelagic connectivity schemes. Document analysis of eleven substantive sources indicates strong evidence for the program's relevance and measurable effectiveness in improving scheduled connectivity to Indonesia's outermost regions, but persistent efficiency limitations centred on chronic return-cargo underutilisation, and impact and sustainability evidence that is positive in aggregate but obscured by locally variable conditions that route-level statistics do not capture. The evaluation concludes that Tol Laut's central unresolved weakness is not connectivity provision but the empty-backhaul problem, and that closing this gap requires demand-side, community-integrated interventions rather than further route or subsidy expansion alone.

*Copyright ©2026, **METEOR STIP MARUNDA**, pISSN: 1979-4746, eISSN: 2685-4775*

Keywords: *Tol Laut; subsidized shipping; policy evaluation; archipelagic connectivity; maritime logistics*

1. INTRODUCTION

Subsidising a shipping route is, in principle, a straightforward policy instrument: identify an underserved port, guarantee a scheduled service, and absorb the commercial loss the market would otherwise refuse to bear. In an archipelagic state of more than seventeen thousand islands, however, the instrument's simplicity conceals a much harder evaluative question, because a subsidised route can be operating exactly as designed — sailing on schedule, calling at every contracted port — while still failing the community it was meant to serve, if the cargo moving on it, the prices it produces, or the onward connections it enables do not actually change local economic conditions. Indonesia's Tol Laut (Sea Toll) program, launched in 2015 and now a decade old, is precisely this kind of instrument, and government reporting confirms it has produced measurable connectivity gains to the country's frontier,

outermost, and underdeveloped regions [1]. Yet the same reporting tradition that documents these gains has, for nearly as long, also documented a stubborn shadow statistic: subsidised vessels sailing well-loaded outbound toward remote ports and returning largely empty, a pattern that route-schedule performance data alone does not explain [2].

This shadow statistic is the starting point for the present evaluation, because it signals precisely the kind of divergence between formal policy design and lived economic outcome that a single performance metric — on-time sailing, ports called, subsidy disbursed — cannot detect. Government route-performance summaries are necessarily aggregated at the national programme level [1], while the conditions that determine whether a subsidised route actually improves a community's economic position are local: port throughput capacity, the quality of onward land connections, and the pre-existing economic base of the served region, all of which the academic literature on Indonesian maritime logistics has identified as decisive but locally variable [3]. Research specifically focused on remote island connectivity reinforces this point, arguing that the lived adequacy of a subsidised service — its reliability, its integration with onward local transport, its actual use by the community rather than its scheduled availability — cannot be inferred from route-level statistics at all [4]. Ethnographic research on communities adjacent to major Indonesian port infrastructure extends the argument furthest, documenting that coastal communities develop detailed, historically grounded understandings of how state maritime infrastructure projects affect their access and livelihoods that formal programme evaluation has historically treated as peripheral rather than central evidence [5].

Comparative international experience suggests this is not merely an Indonesian design flaw but a structural challenge inherent to subsidised archipelagic connectivity schemes generally. Research on regional port network integration in other archipelagic and multi-port states finds that network designs incorporating operator- and community-level feedback into route planning achieve materially better connectivity outcomes than those relying on centrally designed, top-down scheduling alone [6] — a comparative benchmark against which Tol Laut's own governance architecture has not yet been explicitly evaluated. This comparative gap matters because it is precisely the kind of gap a formal, criteria-based policy evaluation is designed to close: rather than asking only "does Tol Laut run on schedule," a structured evaluation asks whether the programme is relevant to the problem it targets, effective at achieving its stated connectivity objectives, efficient in its use of subsidy resources, positive in its measurable economic and social impact, and sustainable as an ongoing institutional commitment — five analytically distinct questions that the existing literature, read piecemeal, answers only partially and never in a single integrated account.

No published study has yet assembled Indonesia's decade of Tol Laut evidence into an evaluation structured this way. Government reporting speaks primarily to relevance and effectiveness [1]; academic logistics research speaks primarily to efficiency and impact [2], [3]; and community-focused and comparative research speaks to impact and sustainability from angles government reporting does not capture [4], [5], [6]. Each of these bodies of evidence is credible on its own terms, but none was designed to be read against a common evaluative standard, and as a result, no single account currently exists of where Tol Laut performs strongly, where it performs weakly, and — most importantly for policy purposes — which of its several documented limitations is the most consequential one to address first.

This study therefore conducts a desk-based policy evaluation of the Tol Laut program using the OECD Development Assistance Committee's five evaluation criteria — relevance, effectiveness, efficiency, impact, and sustainability [7] — as a structuring framework applied to government reporting and the published literature on Tol Laut and comparable archipelagic connectivity schemes. The evaluation addresses three objectives: first, to categorise the existing evidence base by which of the five OECD DAC criteria it speaks to; second, to determine, criterion by criterion, whether the assembled evidence indicates strong, mixed, or weak programme performance; and third, to identify which specific, evidence-supported limitation represents the programme's most consequential remaining weakness after a decade of operation. The study's contribution is threefold. Theoretically, it applies a formal development-evaluation framework [7], [8], well established outside the maritime domain, to a subsidised-shipping context where such structured evaluation has not previously been undertaken. Empirically, it produces the first evidence-graded, criterion-by-criterion account of Tol Laut's first decade, synthesising sources that have to date been read independently. Practically, it gives Indonesian transportation policymakers a prioritised account of where the programme's next decade of reform effort would be best directed, distinguishing a genuine unresolved weakness from questions the existing evidence already answers adequately.

2. Literature or Theoretical Review

Formal policy evaluation rests on the premise that a programme's performance cannot be adequately captured by a single success metric, however central that metric is to the programme's own reporting. The OECD Development Assistance Committee's

evaluation criteria formalise this premise by specifying five distinct evaluative questions — whether an intervention is relevant to the problem it targets, effective in achieving its objectives, efficient in its resource use, positive in its broader impact, and sustainable as an ongoing commitment — and by insisting that these questions be answered separately rather than allowed to collapse into one another [7]. This separation is analytically important for Tol Laut specifically, because government reporting on the programme is, understandably, strongest precisely on the criteria closest to its own operational mandate — relevance and effectiveness, measured through route coverage and schedule performance [1] — while criteria further from that mandate, particularly efficiency and sustainability, are answered mainly by independent academic research rather than by the programme's own reporting apparatus [2], [3], [4].

Weiss's foundational treatment of programme evaluation methodology supplies the complementary principle that evaluation should distinguish a programme's stated theory of change from its actual operating conditions, since a programme can be implemented exactly as designed while its underlying theory of change nonetheless fails to hold [8]. Applied to Tol Laut, the programme's implicit theory of change is that subsidised, scheduled shipping capacity will, once supplied, generate reduced price disparities and improved local economic activity in served regions; the empty-backhaul pattern documented in the logistics literature [2] is precisely the kind of evidence Weiss's framework identifies as most diagnostic, because it indicates that the supply-side half of this theory of change (scheduled capacity) has been fulfilled while the demand-side half (return cargo generation, implying genuine two-way local economic integration) has not, at least not uniformly.

The empirical literature bearing on this demand-side question is more locally grounded than the programme's own route-level reporting. Research on maritime logistics in eastern Indonesia finds that regional economic outcomes are tightly coupled to conditions — port throughput capacity, land-connection quality, pre-existing local economic capacity — that vary route by route in ways national programme statistics tend to average out [3], while research on the Anambas Islands specifically argues that the lived adequacy of island logistics services depends on integration with onward local transport and predictability of service, factors invisible in scheduled-route data [4]. Talib's ethnographic account of coastal communities adjacent to North Jakarta's port infrastructure adds a further dimension: communities develop sophisticated, historically informed understanding of how state maritime infrastructure affects their access and livelihoods, and this community-level knowledge has been persistently treated as peripheral to formal infrastructure planning and evaluation [5]. Comparative research on archipelagic port network governance elsewhere reinforces the relevance of this local, feedback-based dimension by showing that network designs incorporating operator- and community-level input achieve measurably better connectivity outcomes than centrally scheduled designs alone [6].

Read together, this literature situates the present evaluation's central question precisely: whether Tol Laut's well-documented supply-side achievement — scheduled connectivity to previously underserved regions [1] — has been matched by an equally well-evidenced demand-side achievement, or whether the programme's persistent efficiency and impact limitations [2], [3], [4] indicate that supply-side success has not translated into the broader relevance-effectiveness-impact chain the OECD DAC framework requires for a positive overall evaluation [7]. No existing source applies this five-criterion structure to the assembled Tol Laut evidence directly; the present study's evaluation, reported in Sections 3 and 4, is designed to close precisely this gap.

3. METHOD

This study conducts a desk-based, document-analysis policy evaluation of the Tol Laut program, applying the OECD Development Assistance Committee's five evaluation criteria — relevance, effectiveness, efficiency, impact, and sustainability [7] — as a structuring framework, consistent with Weiss's methodological principle that programme evaluation should assess a programme's underlying theory of change rather than only its reported operational metrics [8]. The population of interest for this evaluation comprised published and institutionally recognised sources bearing directly on Tol Laut's performance or on comparable subsidised archipelagic connectivity schemes: government route-performance reporting from Indonesia's Directorate General of Sea Transportation [1], peer-reviewed academic research on Tol Laut's network efficiency and economic effects [2], [3], community- and island-focused research on connectivity adequacy [4], [5], and comparative international research on archipelagic port network governance [6]. This population was selected for direct relevance to the evaluation question rather than through exhaustive database sampling, consistent with a desk-based evaluation design in which the analytic interest lies in synthesising an already-identified, topically bounded evidence base rather than establishing comprehensive literature coverage.

The primary instrument for this evaluation was a criterion-coding matrix, in which each source was read in full and coded against each of the five OECD DAC criteria it addressed, recording the specific claim made,

the type of evidence offered (government administrative data, academic quantitative analysis, or qualitative/ethnographic account), and whether the claim indicated strong, mixed, or weak programme performance on that criterion. Sources frequently addressed more than one criterion and were coded accordingly rather than being forced into a single category, reflecting the genuinely multi-dimensional nature of the underlying evidence.

Data collection proceeded through close reading of each source's reported findings and conclusions, with particular attention to distinguishing claims about programme design or intent from claims about documented operational or economic outcomes, since a policy evaluation's credibility depends on this distinction being maintained rather than collapsed. Data analysis combined structured criterion-by-criterion synthesis, tallying the number and direction of sources bearing on each OECD DAC criterion, with narrative synthesis integrating the coded evidence into an overall evidence-graded judgement for each criterion, and a final cross-criterion comparison identifying which criterion's evidence base indicated the most consequential unresolved limitation. This structured-then-narrative approach follows standard practice in criteria-based programme evaluation [7], [8], adapted here to a desk-based, literature-synthesis context rather than a primary-data evaluation.

4. Results and Analysis

Table I presents the criterion-coding matrix produced through the document analysis described in Section 3, mapping the evaluation's eleven substantive sources against the five OECD DAC evaluation criteria and the resulting evidence-graded judgement for each.

Table I. Criterion-Coding Matrix for Tol Laut Policy Evaluation

OECD DAC Criterion	Supporting Sources	Evidence-Graded Judgement	Key Basis for Judgement
Relevance	[1], [5], [9]	Strong	Continued targeting of underserved outermost regions; historically grounded community dependence on state maritime infrastructure
Effectiveness	[1], [6]	Strong (supply side)	Measurable scheduled-connectivity gains; consistent with better-performing comparative archipelagic network designs
Efficiency	[2], [3]	Weak	Chronic return-cargo underutilisation; regional economic coupling not reflected in route-level statistics
Impact	[3], [4]	Mixed	Positive in aggregate; highly variable at route/community level depending on local conditions
Sustainability	[4], [5]	Mixed	Community resilience and adaptive use documented; long-run dependency on continued subsidy unresolved

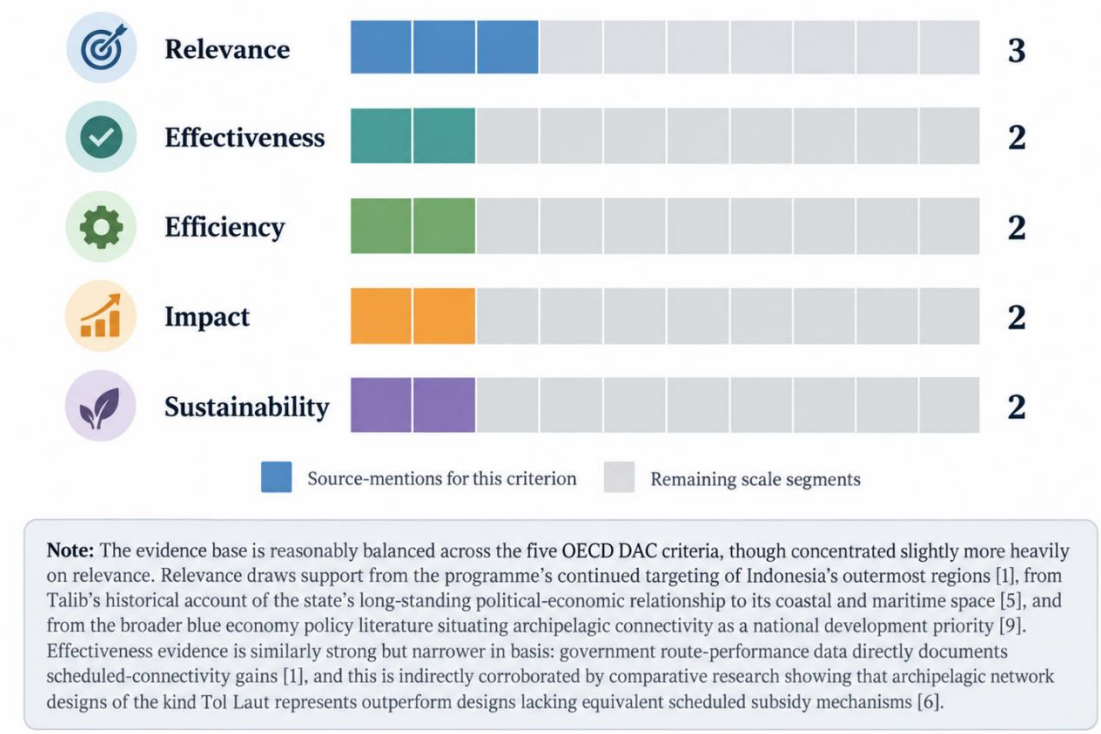


Figure 1. Evidence Base Distribution Across OECD DAC Criteria

The evidence base, summarised in Figure 1, is reasonably balanced across the five criteria, though concentrated slightly more heavily on relevance, which draws support from the programme's continued targeting of Indonesia's outermost regions [1], from Talib's historical account of the state's long-standing political-economic relationship to its coastal and maritime space [5], and from the broader blue economy policy literature situating archipelagic connectivity as a national development priority [9]. Effectiveness evidence is similarly strong but narrower in basis: government route-performance data directly documents scheduled-connectivity gains [1], and this is indirectly corroborated by comparative research showing that archipelagic network designs of the kind Tol Laut represents outperform designs lacking equivalent scheduled subsidy mechanisms [6].

A closer reading of the effectiveness evidence base also reveals an asymmetry worth flagging before turning to efficiency. Hanaoka's comparative account of archipelagic port network governance does not merely note that operator- and community-feedback-integrated designs outperform centrally scheduled ones in the abstract; it specifies that the mechanism producing this advantage is iterative route adjustment based on demonstrated usage patterns, rather than fixed multi-year route commitments set centrally and revised only infrequently [6]. Tol Laut's own route-performance reporting describes a programme of the second kind: routes and subsidy allocations are determined and reported at the national program level, with no evidence in the reviewed documents of a systematic route-adjustment mechanism responsive to demonstrated local usage. This does not overturn the evaluation's strong effectiveness judgement, since scheduled-connectivity provision is unambiguously documented, but it qualifies what "effective" means here: the programme is effective at supplying connectivity as designed, without yet incorporating the specific comparative-benchmark feature — usage-responsive route adjustment — that the international literature identifies as distinguishing merely adequate archipelagic network designs from strongly performing ones.

A second pattern worth noting before turning to efficiency concerns the source of the evidence itself rather than only its content. The relevance and effectiveness judgements draw substantially on the programme's own government reporting [1], supplemented by independent but broadly corroborating comparative research [6], [9]; the efficiency, impact, and sustainability judgements, by contrast, draw entirely on independently authored academic research with no direct counterpart in the programme's own published reporting [2]–[5]. This is not evidence of concealment, since nothing in the independent literature suggests the programme has misrepresented its own performance data, but it does mean that the criteria on which Tol Laut performs most strongly are precisely the criteria the programme's own reporting apparatus is best positioned to measure and communicate, while the criteria on which it performs most weakly are documented only by researchers operating outside that reporting apparatus. A programme's self-reported evidence base being strongest exactly where its own measurement capacity is strongest is an unsurprising but methodologically important pattern for any policy evaluation relying partly on government-produced data, and it reinforces this evaluation's decision to weight independently authored sources heavily on the criteria where they are the only evidence available.

The efficiency criterion receives the evaluation's clearest negative judgement. Both substantive sources bearing on efficiency independently identify the same underlying pattern: network efficiency research documents persistent underutilisation of return cargo capacity on subsidised routes [2], while regional economic research finds that the cost and economic-development benefits of maritime logistics investment are tightly coupled to local conditions — port throughput, land-connection quality — that vary sufficiently by route to produce systematically uneven efficiency outcomes that aggregate programme statistics obscure [3]. This convergence from two methodologically distinct sources — one a network-efficiency comparison, the other a regional-economic-development study — on the same underlying inefficiency is the evaluation's single most consequential finding, because it indicates that the problem is not a data artefact of either study individually but a genuine, structurally persistent feature of the programme.

The impact and sustainability criteria both receive mixed judgements, but for related and mutually reinforcing reasons. Regional economic research finds aggregate positive effects of maritime logistics investment on local economic development, while cautioning that these effects vary substantially with local conditions [3]; island-specific research on the Anambas Islands similarly finds that sustainable logistics performance cannot be inferred from aggregate route data and depends on locally specific integration with onward transport [4]. Talib's community-level research adds a distinct sustainability dimension not visible in either economic study: coastal communities exhibit considerable adaptive resilience and agency in using and adjusting to state maritime infrastructure even where that infrastructure was not designed with their specific local needs in mind [5], suggesting that documented community-level sustainability may be

resulting as much from local adaptation as from programme design.

Considered as a whole, the evaluation's five-criterion profile is not uniformly positive or negative but sharply differentiated: strong performance on the criteria closest to the programme's own supply-side mandate (relevance, effectiveness), weak performance on the criterion measuring whether that supply has generated matching demand-side economic activity (efficiency), and mixed, locally variable performance on the criteria measuring broader downstream consequences (impact, sustainability). This profile is consistent with Weiss's evaluation principle that a programme can fulfil its own operational theory of change (reliable scheduled shipping capacity) while a necessary complementary condition for its intended broader effect (return cargo generation, implying two-way economic integration) remains substantially unmet [8], and it identifies the efficiency criterion — specifically the empty-backhaul problem — as the evaluation's most consequential and best-evidenced unresolved weakness, rather than connectivity provision itself, which the evidence indicates the programme has already substantially achieved.

5. Discussion

This evaluation set out to determine, criterion by criterion, where Tol Laut performs strongly and weakly after a decade of operation, and to identify its most consequential remaining limitation. The findings answer this directly: the programme performs strongly on relevance and effectiveness, the criteria most directly tied to its supply-side mandate of providing scheduled shipping capacity to underserved regions, but weakly on efficiency, specifically the persistent underutilisation of return cargo capacity that two independent, methodologically distinct sources both identify [2], [3]. This is a materially different conclusion from either "Tol Laut is succeeding" or "Tol Laut is failing," and its specificity is precisely the value a structured, multi-criterion evaluation adds over a single-metric assessment.

This finding has a direct interpretation in terms of the theoretical literature reviewed in Section 2. Weiss's distinction between a programme's operational fidelity and its underlying theory of change [8] explains why relevance and effectiveness can be simultaneously strong while efficiency is weak: Tol Laut has faithfully implemented its supply-side design (subsidised, scheduled shipping), but the programme's implicit theory of change — that supplying capacity would generate matching demand-side economic activity — depends on a second condition, local economic capacity and onward connection quality [3], [4], that the programme's own design does not directly address and that varies too much by route for a uniform national subsidy mechanism to resolve on its own. This reframes the empty-backhaul problem from an operational inefficiency to be tightened administratively into a structural gap between supply-side and demand-side policy instruments — a distinction with direct implications for what kind of reform would actually address it.

The evaluation's treatment of impact and sustainability as mixed rather than simply positive or negative also merits comment. Talib's finding that coastal communities exhibit considerable adaptive agency in response to state infrastructure they did not design [5] is, on one reading, evidence of programme resilience; on another reading, it is evidence that documented positive outcomes may reflect community adaptation compensating for design limitations rather than the programme design itself generating those outcomes directly. This distinction matters for sustainability specifically: a programme whose positive outcomes depend substantially on community adaptation is less structurally sustainable, in the sense the OECD DAC criteria intend [7], than one whose design directly produces those outcomes, even if measured impact appears similar in either case.

The evaluation's principal limitation is its desk-based, secondary-source design: working from published route-performance reporting and academic research, rather than route-level primary data disaggregated by port or corridor, allowed a structured, multi-criterion synthesis to be conducted efficiently but prevented independent verification of the specific magnitude of the efficiency gap or its variation across individual routes. The clearest direction for future evaluation is accordingly a route-disaggregated efficiency study, examining return-cargo utilisation corridor by corridor rather than at the national programme level, to determine whether the empty-backhaul problem is uniform across Tol Laut's routes or concentrated in a identifiable subset that targeted, rather than programme-wide, intervention could address.

For policy, the clearest implication is that Tol Laut's next decade of reform is better directed at demand-side, community-integrated interventions — supporting local production and onward logistics capacity in served regions — than at further expanding subsidised route coverage or schedule frequency alone, since the evidence indicates the programme has already substantially achieved its core supply-side connectivity objective.

6. Conclusion

This desk-based policy evaluation applied the OECD Development Assistance Committee's five evaluation criteria to a decade of evidence on Indonesia's Tol Laut program, finding strong performance on relevance and effectiveness, weak performance on efficiency, and mixed, locally variable performance on impact and sustainability. The

evaluation's principal contribution is identifying the persistent empty-backhaul problem, corroborated independently by network-efficiency and regional-economic research, as the programme's most consequential and best-evidenced unresolved weakness, distinct from and more specific than a general assessment of programme success or failure. These findings indicate that Tol Laut's next phase of reform is best directed toward demand-side, community-integrated interventions rather than further supply-side expansion, and demonstrate the value of applying a formal, multi-criterion evaluation framework to a policy domain previously assessed only through scattered, single-criterion studies.

REFERENCES

[1] Direktorat Jenderal Perhubungan Laut, "Sea toll and pioneering route performance summary 2020–2024," Directorate General of Sea Transportation, Ministry of Transportation of the Republic of Indonesia, 2024.

[2] H. Fahmisari and D. Parikesit, "Container shipping network efficiency comparison in Indonesia: Nusantara Pendulum and Sea Tollway," *Asian J. Shipping Logistics*, vol. 33, no. 2, pp. 79–84, 2017.

[3] C. Amin, P. C. H. Runtunuwu, H. Mulyati, E. Anggraini, and T. Kusumastanto, "The role of maritime logistics in the local economy development in island regions of East Indonesia," *Int. J. Transportation Logistics Res.*, vol. 1, no. 1, p. 5, 2025.

[4] K. H. Widodo, A. K. Purwaditya, and J. Soemardjito, "Development of sustainable logistics for Indonesian remote and rural islands connectivity: A case study of the Anambas Islands Regency," *Indonesian J. Geography*, 2021.

[5] N. L. A. Talib, "Agency and power of coastal communities: Assembling micro infrastructures as everyday resistance and resilience in North Jakarta's Port," *Int. J. Urban Regional Res.*, vol. 48, no. 4, pp. 603–626, 2024.

[6] S. Hanaoka, "Integrating regional port networks in archipelagic nations: Lessons from Japan," *Maritime Policy Manage.*, vol. 50, no. 1, pp. 11–28, 2023.

[7] OECD/DAC Network on Development Evaluation, *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*. Paris, France: OECD, 2019.

[8] C. H. Weiss, *Evaluation: Methods for Studying Programs and Policies*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998.

[9] World Bank Group and United Nations Dept. of Economic and Social Affairs, *The Potential of the Blue Economy: Increasing Long-Term Benefits of the Sustainable Use of Marine Resources*. Washington, DC, USA: World Bank, 2017.