Issue 6, February 2019

BOLD THINKERS DRIVING REAL-WORLD IMPACT

GOVERNANCE WORKING PAPER SERIES

Monitoring, Evaluating and Learning for Complex Programs in Complex Contexts: Three Facility Case Studies

Tara Davda and Lavinia Tyrrel

Key Words: Monitoring, Evaluation and Learning, Complexity, Thinking and Working Politically, Doing Development Differently, Facilities, Australian aid, Papua New Guinea, Timor-Leste, Indonesia.

Acknowledgements: The authors would like to sincerely thank the KOMPAK, ATLPHD and PNG Governance Partnership staff who contributed their time for interviews, and the reviewers for their thoughtful comments and efforts towards improving this manuscript; in particular, Mark Fiorello and Graham Teskey.

Table of Contents

Foreword

Since its founding in 1965, Abt Associates has placed great emphasis on learning from its experience in all that it does – domestically in the United States and Australia, and internationally in South America. Africa, Asia and the Pacific. We are constantly seeking to pursue new and better ways of delivering community and development assistance. Initially much of Abt Associates work was focused in the fields of health and social policy. Today, the company also applies its technical and program capabilities into the governance arena – including on issues of front-line service delivery, community driven development, local governance, economic and public sector management and leadership and coalitions.

We see 'governance' not merely as a sector: it is a way of thinking about how development occurs: whose interests are being promoted? What is the pattern and structuring of incentives that influence the pace and direction of change? And perhaps most critically of all, how does the mosaic of formal and informal *institutions* interact to determine the possibilities and parameters of the change process? We apply 'governance' as a way of working across all sectoral and governance-specific investments.

This paper addresses an issue of real strategic importance to both the Australian government and to Abt Associates: how do we judge the overall performance of each Facility? By their nature, Facilities create intellectual and management challenges not present in less complex and ambitious development initiatives: is it possible meaningfully to aggregate results arising from different programs? Just how much 'contribution' to a high-level development goal is required and how can such a convincing argument be constructed? Do we possess the skills to monitor progress in real-time and adapt our programs accordingly? And most importantly, to what extent does the use of the project framework hinder or help monitoring and learning about Facility-wide performance?

The authors here review Abt's experience in managing three high-value, high-profile governance Facilities – in PNG, Timor-Leste and Indonesia. The findings are sobering. The development community has much to learn from these three initiatives if it wishes to demonstrate that argument of development effectiveness: that 'the whole is greater than the sum of the parts.' Identifying the problem is the first step to considering a solution.

Graham Teskey

Principal Technical Lead – Governance
Abt Associates

Monitoring, Evaluating and Learning for Complex Programs in Complex Contexts: Three Facility Case Studies

Tara Davda and Lavinia Tyrrel

1. Executive Summary

1.1 Surprisingly little has been written about the experience of monitoring, evaluation and learning for complex programs in complex contexts (see box 1). While there is a growing body of evidence about how to establish teams, budgets and partnerships to 'think and work politically' (TWP) (see Ladner 2015; Cole et al 2016; Faustino and Booth 2014), there is limited guidance on what it takes to establish Monitoring Evaluation and Learning Frameworks (MELFs) for these sorts of programs. Further, where evidence does exist, it focuses on single sector projects – not MEL for complex portfolios using a range of modalities, targeting a variety of development problems. This knowledge gap has specific implications for the high-value, multi-sector 'Facilities' that the Australian Department of Foreign Affairs and Trade (DFAT) funds in Indonesia, Papua New Guinea and Timor-Leste, and Abt Associates manages.

1.2 The findings in this paper are based on a study of these three Facilities. The paper identifies seven areas where lessons have emerged, or deviation from more conventionally designed and implemented MELFs has been required. These differences emerge at all stages of implementation. They include:

i. clarifying the Facility's strategic intent, overall-theory of change (i.e. not just a theory of action) and an agreed strategic plan which would then guide the development of a MELF;

ii. designing MELFs to meet multiple, sometimes competing demands regarding accountability, public diplomacy/ communication, evaluation and internal learning;

iii. finding ways to explain what the Facility is achieving, without simply aggregating results up from one level of the project frame to the next;

iv. the challenge of setting facility-wide indicators and telling a persuasive contribution story;

v. the challenge of setting baselines across such large, constantly changing, portfolios of work in data-poor country contexts;

vi. developing systems to embed learning into programming, and;

vii. finding staff who can apply MEL to projects working in adaptive and politically-informed ways, and quarantining budgets for this.

1.3 The high-level lesson to emerge is that traditional forms of monitoring and evaluation – where the primary focus is on accountability, ex-post learning and evaluation, linear change and deliberate (rather than emergent) strategies (see Table 1) – do not lend themselves well to the Facility model. There are a number of reasons for this: Facilities are designed to be flexible and adapt during implementation; results cannot simply be 'aggregated up' across the plethora of (often unrelated) activities and projects a Facility manages; Facilities are trying to influence institutional change (which is by its very nature, is hard to measure); MEL is often deprioritised and kept at arm's reach to programming, and; it is difficult to find staff capable of adapting and applying MEL methods to the Facility model.

1.4 In concluding, we suggest that – if the international community is serious about transforming how complex programs are measured and adapted – then the place to start is not MEL methods, but the project framework itself. A fundamental impediment to effectively applying MEL in Facilities is the application of linear change models to complex contexts. Although project frameworks serve a

critically important accountability purpose (to give donors confidence about what they are 'buying' with their aid funds), in some cases, it has been mistaken for a hard performance benchmark – working actively against more flexible and adaptive forms of program management. It is also extremely difficult to describe, summarise and plan a (often expermential) portfolio of investments using a linear change approach. In these cases, the project framework incentivises donors and implementers to separate MEL from programming and lock in activities, outputs and outcomes up-front. MEL thus become a compliance tool to 'track' whether the project achieves the pre-determined outputs it set itself, rather than a way of learning about the context, what's working, what's not and why – and adapting the program in response.

2. Introduction

2.1 Abt Associates (Abt) manages the implementation of three, highly visible, high-value, multi-sector Australian Government Facilities: *the Australia-Timor-Leste Partnership for Human Development (ATLPHD*, known as PHD) in Timor-Leste, *KOMPAK* in Indonesia and *the PNG Governance Facility* (formerly the PGF) in Papua New Guinea. Together, they have a combined value of over A$500m over four years, which represents approximately 20% of Australia's total bilateral aid to these countries[1]; as well as a significant share of partner government spend (e.g. 8% of sector spend in Timor-Leste).

2.2 The experience of managing these investments has afforded Abt a unique perspective on the limitations of current MEL theory and practice. Based on data generated by a qualitative study of select investments in the portfolio, this paper explores the contractor's experience of developing MEL approaches for large multisector Facilities.

2.3 The purpose of this paper is to contribute to the argument for more sophisticated, adaptive and mixed methods approaches to MEL in large, multi-sector development Facilities. It is also anticipated that the paper will support practitioners, program staff, managing contractors, policy makers and donor staff who work with M&E in complex, large scale multi-sector Facilities.

2.4 The following section (section 3) situates the study within current research on the Facility model and the application of MEL within complex programs and operating contexts. Section 4 summarises the research method and limitations; next, section 5 describes each Facility, the context and their respective approach to MEL. The findings, as they relate to international theory and practice, are outlined in section 6. Lastly, the paper concludes with suggestions on the design and management of future Facilities; of relevance for donors, practitioners and policymakers.

3. Situating the Study – the literature on the Facility model and complex change

3.1 Facilities or 'large programmes'[2], in which "multiple projects are grouped together under a wider umbrella, have become an increasingly popular model of development delivery among donors" (Buffardi & Hearn, 2015:4). While the justification for this modality varies across the literature, common reasons cited include: improved opportunities for wider learning and integration (Buffardi & Hearn, 2015); economies of scale and transaction efficiencies (where a single programme management structure generates improved value for money through the delivery of programmes more cost-effectively) (DFAT, N.D.a); and improvements in donor organisational capability (DFAT,

[1] Based on available 2015-16 and 2016-17 Australian Government budget estimates for bilateral aid spend, KOMPAK represents about 8% of total bilateral aid to Indonesia; PHD represents about 35% of total bilateral aid to Timor Leste and PGF represents about 18% of total bilateral assistance to PNG.

[2] Buffardi & Hearn (2015) also note that such arrangements are also known as 'consortia' or 'schemes'

N.D.c). Facilities are believed to "better position [donor] staff to focus on their core responsibilities: policy, Facility oversight and strategic stakeholder management and reporting" (DFAT, 2015, in De Lacy, 2017).

3.2 However, despite the ongoing use of Facilities to deliver aid globally – extending as far back as the 1970s – there is no single agreed definition of what constitutes a Facility. As such, this paper refers to two definitions: Bayley (2019) and De Lacy et al (2017). Bayley[3] argues that there are four categories of Facilities:

> "(1) Supporting and administrative Facilities: do not seek to achieve development outcomes, nor are they accountable for doing so. Their primary purpose is to provide administrative, operational and/or logistic support to other donors or implementing partners who are delivering aid projects. An example of this type of Facility is the Solomon Islands Resource Facility;
>
> (2) Grants Facilities: primarily focused on the delivery of discrete pieces of work that have clear timeframes and outputs. For example, infrastructure Facilities;
>
> (3) Adaptive Facilities: by virtue of the type of problems these investments work on (i.e. complex, politically charged) these Facilities purposively seek to work in flexible, iterative and responsive ways. KOMPAK in Indonesia, is an example.
>
> (4) Consolidated Facilities: an amalgamation of programs, not necessarily sharing a common outcome, but hopefully providing a more efficient delivery model than individual partnerships or streams" (Bayley, 2019, pers. comm., 10 Jan 2019).

Similarly, De Lacy et al (2017) expand upon the definition of adaptive Facilities, noting that they are all "*multi-program in scope* (with the aim of achieving efficacy and/or effectiveness goals); *comprehensive* (can undertake a range of aid management or development functions, and administer a range of financing instruments); *adaptive and responsive* (in terms of managing budgets, designing and implementing projects and selecting and working with partners); and *strategic* (in terms of translating high level goals into a set of aid programs that can deliver on these aims, perhaps with the exception of Facilities design to provide purely operational services" (De Lacy et al, 2017:5)[4]. This diversity makes developing MEL frameworks particularly challenging and is the focus of the latter part of this paper.

3.3 In the three Facilities reviewed in this paper, the MEL agenda was more complex than much of that discussed in the literature and donor practice notes to date. While the complexity literature has much to add to the specific methods Facilities can employ to understand and 'measure' change for discrete projects in complex contexts, there is little written on how this applies to performance monitoring and learning at the Facility-wide level. In short, the literature deals well with complexity-aware monitoring, evaluation and learning, but less well with complex aid investments working in complex aid contexts that require a range of MEL methods and approaches.

[3] Scott Bayley was previously the Principal Sector Specialist for Monitoring, Evaluation and Learning at DFAT; he currently works for Oxford Policy Management.

[4] For a more detailed explanation of facilities, please see De Lacy et al, 2017, *Managing Facilities: a stock-take from the first 12 months,* Governance Working Paper Series, Issue 3, December 2017

3.4 Donors have increasingly acknowledged that aid is more likely to achieve results when it is politically informed and able to respond to the local context in which it is operating. Yet practice has been slow to catch up with rhetoric. While there is a growing body of evidence about how to establish teams, budgets, projects and partnerships, to 'think and work politically' (TWP) (Ladner 2015; Cole et al 2016; Faustino and Booth 2014), there is limited guidance on what it takes to establish MEL frameworks for adaptive and politically informed programs. Further, where evidence does exist (Ladner 2015), it focuses on single sector programs. This research gap has the following implications for Facilities:

> **Box 1: Simple, Complicated and Complex Change**
>
> There is a growing body of theory regarding the application of MEL for adaptive and politically informed programs (Roche and Kelly 2012; Ladner 2016; USAID 2018). Much of this draws from complexity and systems thinking: recognising that most aid programs are operating in 'complex' contexts. In these instances, change is very hard to predict at the start of the program. Indeed, it may not even be clear come program end. *"While experience and principles from other situations may guide the design and implementation of such work, it is often the case that it is only by probing and acting that understanding is developed. In these situations, regular monitoring and feedback provide the information to enable the program to assess its progress, or not, towards its objectives, and adapt as experience and learning develops"* (Roche and Kelly 2012: 8-9). This is in contrast to changes which may be 'simple' or 'complicated': whereby the relationship between cause and effect are much easier to predict or uncover with a bit of analysis, consultation or reflection on past experience or lessons learnt from other contexts.

i. Most donor issued MEL guidance is suited for programs working in 'simple' change contexts (see Box. 1, for more detail on *Simple* Vs. *Complex* change) or in other words, in contexts where it is considered possible to anticipate with some degree of certainty how change will happen. Features of simple change contexts include: an environment where few changes are expected during implementation, and most (if not all) variables are known up-front. In this instance, the primary purpose of MEL is to track progress towards a predetermined set of indicators and outcomes, thus donors and implementers can be fairly confident of issuing a design which contains a static program logic. This means the development of a TOC, and the specifying of how inputs lead to outcomes, is relatively (more) straightforward. However, because of the complex change contexts Abt's Facilities operate in, this type of logic is poorly suited.

   A definitive feature of complex change is that the relationship between cause and effect (and hence inputs and outcomes) is hard to predict. As such, monitoring and evaluation methods are important ways for the program to uncover new information, understand what is changing in the operating context, and adapt in real-time. Such an approach favours flexibility: the ability to adapt strategy, inputs, outputs and even outcomes, and stands in sharp contrast to the traditional, planned approach to aid, and the basis on which most practical MEL guidance has been developed.

ii. The evidence compiled on MEL for programs seeking to think and work politically is largely drawn from single issue, single modality case studies. In the case of Facilities, however, implementers are managing a portfolio of projects, grants and contracts, some of which are working in adaptable and politically informed ways and others which are planned in nature. A mix of monitoring and evaluative techniques are required to suit the range of program modalities used by the Facility. There is a need to focus research efforts on this area.

iii. A Facility must also explain its achievements at both the individual program and whole-of-portfolio level. Such features require a more sophisticated approach to MEL than both

traditional or experimental theory and practice currently offers. A mix of monitoring and evaluative techniques are required to suit the range of program modalities used by the Facility.

4. Method and Limitations

4.1 The findings in this paper are based on a qualitative study, which was undertaken over four months (September to December) in 2018. The study was undertaken in three stages:

i. *Document review:* first, the team reviewed key source documents and relevant literature from each program to refine the scope of the study and identify research questions. These included: original investment concept notes or design documents, strategic plans, annual plans, MEL frameworks or performance assessment frameworks, and other relevant documents.

ii. *Semi-structured interviews and thematic analysis*: Second, semi-structured telephone interviews were conducted with six MEL specialists from the Facilities; two each from KOMPAK, PHD and the PNG Governance Partnership. In each case, a set of questions which explored the Facility context, MELF development, modality and methods, and what worked, what did not and why, were provided to interviewees beforehand. Although the questions were used to guide the interviews, discussion was mostly unstructured to allow themes to emerge. The interviews were recorded, transcribed and coded thematically.

iii. *Triangulation and 'sense checking'*: Lastly, the findings were analysed and presented in the form of a consultation brief, with comments and feedback solicited from the interviewees, the Abt Associates senior executive team and team leaders. Five (out of a possible six) responses were received from program staff and seven from the senior executive and team leaders. This allowed emerging findings and themes to be refined with the views of the senior executive involved in each program. Once the initial findings had been analysed, a number of the original interviewees were further surveyed; this time to allow interviewees and enumerators to explore topics in greater detail.

4.2 There are limitations to this study which should be noted. The sample size of Facility case studies is small, by virtue of the fact that Abt Associates manages only three Australian-funded Facilities working on institutional and sectoral outcomes. While there are numerous other Australian-Government funded Facilities operating in the Pacific, it was not appropriate for this study to investigate their experience given they are managed by other contractors. This meant the pool of staff with relevant experience from which to draw on during the interview process, was necessarily small.

4.3 Additionally, this study focused primarily on the *Facility-wide* monitoring, evaluation and learning frameworks (MELFs) or performance assessment (PAF) systems each Facility established. Data was not collected at the project or activity level. Given that each of the three MELFs are currently being implemented (and in two cases, finalised) the review also focuses on the design and conceptualisation of the MELFs, and to a lesser extent, their implementation.

5. Discussion

5.1 The following discussion serves two purposes. First, to summarise the standards and expectations which donors 'conventionally' apply to aid investments – and against which the three Facilities in this paper have been analysed; and second, to discuss the unique ways each Facility went about establishing their MELFs and adapting 'conventional' donor wisdom to suit their needs.

Common Donor Expectations

5.2 There is broad agreement in the literature that MEL refers to the use of management systems and processes to promote and assess the effectiveness of aid investments in achieving development goals. Most donor-issued guidance on Monitoring and Evaluation thus refer to a similar set of ideas: how to establish systems that generate credible information to inform program and budget decisions, support learning, and systems for communicating results to the public and funding bodies. But each emphasises a slightly different set of features, as shown in table 1 below. This table draws on guidance from three major donors, including the donor of focus in this study.

Table 1: DFAT, USAID and DFID MEL Guidance - commonalities and differences.[5]

	DFAT	USAID	OECD-DAC
Features emphasised	Learning and program decision making focus; MELF all starts with design and a clear logic; Emphasis is on investment level Goal, outcome, output hierarchy; Focus on performance indicators and evaluative questions; Baselines wherever possible; Responsibilities allocated for MEL; MEL plan costed and resourced;	Learning and accountability focus; MEL priorities identified at each part of project cycle; Used at country, project and activity level; Goal, objective, intermediate result, sub-intermediate result hierarchy; Focus on performance indicators and results; Baselines and targets required;	Learning and accountability focus; Follows standard project cycle; Emphasis on project level; Output, outcome goal hierarchy; Focus at evaluation on sustainability, relevance, efficiency, effectiveness and impact;
Common features	1. Accountability is primary focus for MEL efforts 2. MEL is based off an investment design and associated project logic 3. MEL approaches reflect a simple, linear change model (input → goal) vs emergent strategies 4. Performance indicators key mechanism for measuring performance 5. Baselines preferred for tracking progress 6. Generally an *ex-post* focus for evaluation		

The Three Facilities at a Glance

5.3 Although all three Facilities were established under a common proviso – to provide a more effective and efficient way of delivering Australian aid – their structure and strategic intent differs according to the operating context. As a result, each Facility has approached MEL in very different ways and adapted the 'conventional' wisdom as prescribed by major donors (Table 1) to meet their needs. The following summarises some of the core features and peculiarities of each Facility.

[5] For further information see:
https://www.usaid.gov/sites/default/files/documents/2151/USAIDEvaluationPolicy.pdf
https://usaidlearninglab.org/monitoring-toolkit?tab=2
https://usaidlearninglab.org/evaluation-toolkit
https://dfat.gov.au/about-us/publications/Documents/monitoring-evaluation-standards.pdf
http://www.oecd.org/development/evaluation/dcdndep/41612905.pdf

Table 2: The Three Facilities at a Glance

	KOMPAK	PHD / ALTPHD	Governance Partnership
Timeline	Commenced January 2015. Facility end-date anticipated at June 2022.	Commenced June 2016. Facility end-date (of first phase) June 2021.	Commenced December 2015. Facility end-date (of first phase) November 2020
Geographic scope	Works in seven provinces, across 26 districts.	Works at national, sub-national and community levels with activities in all 13 municipalities (activity levels vary substantially by sector).	Works at national, sub-national (up to 12 provinces and districts) and community levels, including the Autonomous Region of Bougainville
Value	AUD 177 million over 7.5 years	AUD 120 million over five years	AUD 450 million over five years
Facility 'type' (see 3.2)	Adaptive	Adaptive and consolidated	Adaptive and consolidated
Sectoral focus	Local level service delivery, governance and economic development	Human development (health, water and sanitation, education, nutrition, gender equality, disability and social protection)	Governance including community development (including along the Kokoda Track and in the Autonomous Region of Bougainville, decentralisation, economic growth and reform, public sector leadership, gender and inclusion. Also contains provision for grant making in health and education.
Strategic intent as relevant to the MELF	KOMPAK's goal is 'poor and vulnerable Indonesians benefit from improved delivery of basic services and greater economic opportunities.' It is an "adaptive and problem-driven Facility" (KOMPAK, 2018:11), with a focus on governance, in particular, improving policy and regulatory frameworks, public financial management, service delivery capacity, and citizen engagement and social accountability.	At design, ATLPHD was described as a 'quasi-Facility', with a portion of the budget allocated to predefined activities and the remainder for programming opportunistically (ALTPHD, 2018). Prior to roll-out[6] (and similar to the other Facilities examined in this paper), Australia's investments in human development were delivered by a variety of contractors. ATLPHD collates them under a single managing contractor and goal of promoting 'social capital for all'.	The goal of the PNG Governance Partnership is to contribute to security, stability and inclusive prosperity in PNG, and improve programming to promote governance processes and institutions for stability and inclusive growth. However, as it is the highest value and highest profile of the three Facilities, the MELF not only measures development impact (i.e. on governance institutions) but also the validity of the facility model itself (i.e. was it 'worth' the investment?).
Partners (selected listing)	Key GoI partners: BAPPENAS, Ministry of Villages, Ministry of Finance, Ministry of Human Development and Culture, and Ministry of Home Affairs, NGOs,	Key GoTL partners including: Ministry of Health, Ministry of Education, Youth and Sport, Ministry of Social Solidarity and Inclusion; International NGOs including: Marie Stopes Timor-Leste, Catalpa International, Plan, Health Alliance International; CBM Australia; Local NGOs including: Alola Foundation, Ba Futuru, Rede Feto	Key GoPNG partners including Treasury, Department of Finance, DPLGA, DPA; Church groups; provincial and district authorities; NGO and community groups (e.g. the Voice); think-tanks and academic institutes (e.g. NRI, UPNG, ANU, UQ).
Instruments and/or methods (sample only)	Policy advocacy and dialogue, research, pilots, and capacity development and institutional strengthening.	Grants, advisory support and capacity development, community engagement and empowerment, policy advocacy and dialogue, institutional strengthening.	Policy advocacy and dialogue, capacity development and advisory support, grants, training, core support for select institutes and institutional strengthening.

[6] Prior to PHD, Australia's investments in human development focused on eight areas: education, health, water, sanitation and hygiene (WASH), gender, nutrition, disability and social protection (ALTPHD, 2018).

Sources: KOMPAK (2018), *Performance Management Framework of KOMPAK 2018-2022*, August 2018; Department Foreign Affairs and Trade (DFAT), *Australia Timor-Leste Partnership for Human Development Investment Design Document*, N.D.b. (Authors' own table).

The Three Facility-wide MELFs: Explained

5.4 What follows is an analysis of the three Facility MELFs as they were at end 2018. This analysis focuses on the MELFs as they were 'designed to be', versus what they actually 'are' or 'are not' in practice. Given that each of the three MELFs are still being tested and evolved – as the Facilities themselves change – it is exceedingly difficult to predict which elements are going to operate as envisaged at design, and which are not. As such, these descriptions ought to be understood for what they are: a secondary source document review and framing analysis – not the final word on how each Facility has approached MELF in practice.

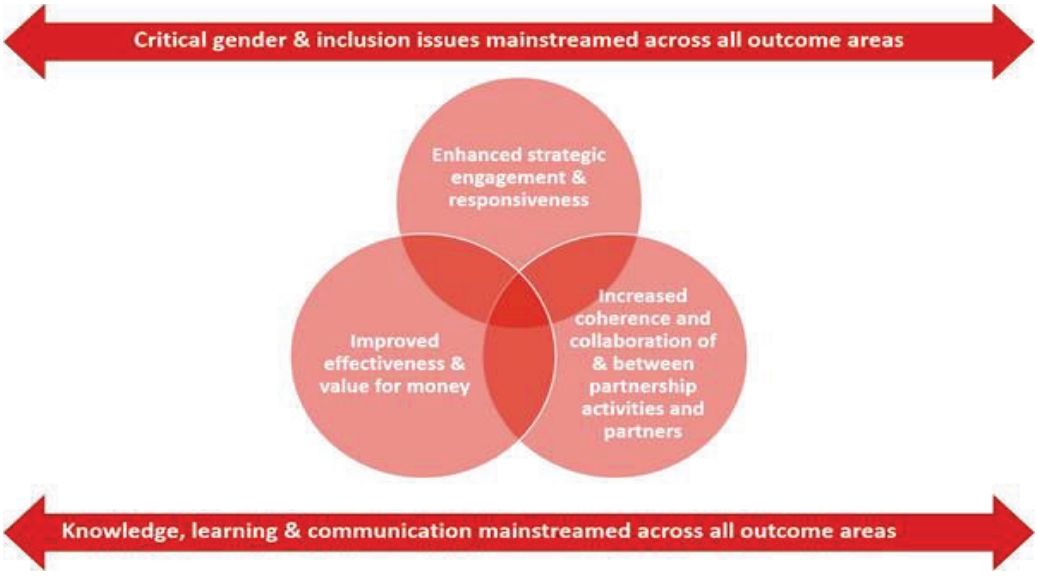
PNG

5.5 The PNG Governance Partnership whole of Facility MELF has had several iterations. Early versions (2017, 2018) had a strong emphasis on facility-wide outcomes: i.e. change in governance institutions; change in the quality of relationships (in particular between Australia and PNG) and change in how the facility was working more effectively and efficiently to deliver its aid programs. However, in mid- 2018, the Partnership underwent a strategic shift. Authority and decision making were delegated to workstream areas (e.g. leadership and decentralisation) and there was a shift away from a single facility-wide goal, theory of change and guiding strategy to inform programming. Instead, workstreams assumed greater responsibility for setting their own outcomes and working towards them – using the services offered by the single Facility operating and program platform. As such, the MELF itself was revised.

5.6 Reflecting this changed strategic intent, the 2019 PNG Partnership MELF measures performance via: (a) outcomes relating to the effectiveness of the Facility model itself, and (b) outcomes relating to the collective contribution of its governance programs in PNG. This means the Facility is assessed on *both* its contribution to governance improvements, and the Facility as a modality for delivering aid.

i. To gather evidence against the first (i.e. 'a') outcome area, the MELF focuses on three sets of process-level outcomes: enhanced strategic engagement and responsiveness, improved effectiveness and VfM; and increased coherence. These three areas are shown in Figure 1. The underlying assumption of outcome area (a) is that if the Facility model is effective, it should be possible to demonstrate collaboration across program teams on shared problems; better alignment between programs, GoPNG priorities and governance issues; collaboration between operations and program teams; and efficiency dividends from using a single management platform.

Figure 1: PNG Governance Partnership outcome areas



ii. To monitor performance on the second outcome area, (b), each partnership area has its own TOC and MELF which detail program specific outcomes and indicators. This enables the Facility to report six separate results stories (one for each partnership).

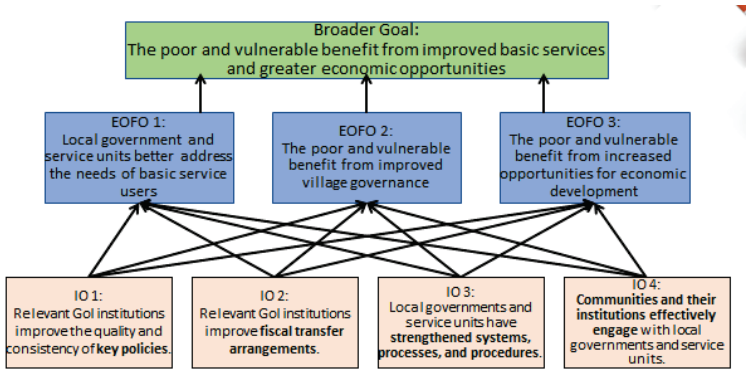
5.7 To assess how the Facility (both the modality itself and the six partnership areas) have impacted governance in PNG, the Facility plans to undertake an annual review and reflection exercise using data collated from the above two outcome areas. This process (termed 'Strategy Testing') requires program teams to schedule "periodic, structured breaks from day-to-day program implementation to collectively reflect on what they have learned and to ask whether the assumptions underpinning their program strategies are still valid in light of new information, insights, and shifts in local context." (Ladner, 2015:5). Programs are adjusted based on the learning from this reflective process. Using this process, and the PNG Governance Update (a national and sub-national dataset compiled by the Facility, using international governance metrics and citizen perception surveys), staff and partners will assess what they believe the Facility's contribution to governance outcomes has been over the past year.

5.8 Data sources for the PGF MELF include the six program MELFs, evidence from the whole of Facility outcome areas, and assessment of the state of governance in PNG using national level quantitative data sources combined with an annual sub-national governance perceptions survey. Key means of verification include stakeholder interviews, analysis of policy and program documents, minutes from internal meetings, VfM analysis and case studies.

INDONESIA

5.9 As an adaptive and flexible Facility at design, KOMPAK's projects were devised to respond to emerging opportunities and risks, changes in circumstances, government priorities and new information generated by learning and reflection. In practice however, while some of this flexibility (as envisaged at design) has remained, some of it has not. Conceptually, KOMPAK's MELF describes a *nested program logic,* with three cascading levels: A higher level goal defines KOMPAK's development contribution in Indonesia (see Fig. 2, below) while three End of Facility Outcomes (EOFOs) capture the changes it is hoping to achieve, and the outcomes it will be assessed against (KOMPAK, 2018c).

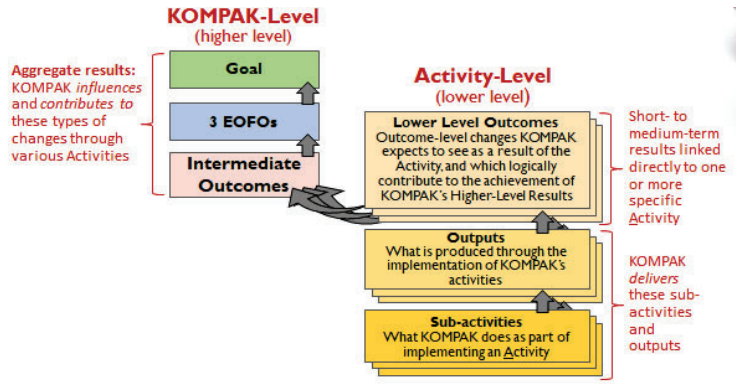
Figure 2: This figure illustrates KOMPAK's Facility-wide MELF, in particular the relationship between the EOPOs and goal.



KOMPAK Internal Presentation, 2018.

5.10 Intermediate outcomes sit below the EOPOs and describe *"the medium-term changes in behaviour, practice, and decisions"* (KOMPAK, 2018c) resulting from KOMPAK's intervention. These 'stepping stones' contribute to the higher (EOFO) level outcomes. The next layer in the program logic is the lower order or 'activity level'. (See Fig. 3 below). These 'outputs' generate certain 'outcomes' and describe the short-to-medium term changes KOMPAK is working towards. They are achieved through implementing specific activities (such as technical assistance (TA), pilot and training activities, advocacy efforts, and research) (KOMPAK, 2018b).

Figure 3: KOMPAK's Nested Program Logic

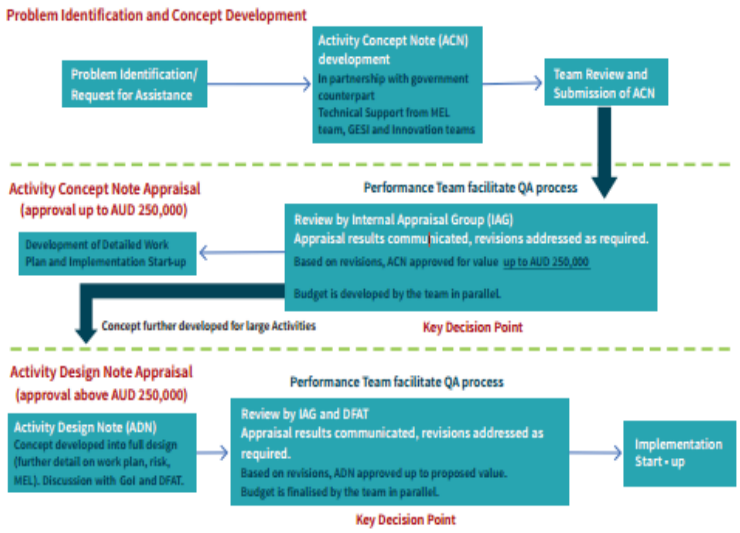


KOMPAK Internal Presentation, 2018.

5.11 Wherever possible, KOMPAK attempts to use a problem driven-approach to program management – and thus MEL (KOMPAK, 2018). In theory, this means that some activities in the portfolio are not prescribed at design but developed in response to emergent opportunities and priorities. These, more flexible activities, are monitored on their performance: KOMPAK uses a performance criterion to evaluate and appraise their design, a process informed by (among other things) Indonesian Government commitment, and the extent to which each activity contributes to

Indonesian government development priorities[7] (Fig. 4, below, illustrates this). In reality however, it has been difficult to maintain this level of sophistication for all of the activities KOMPAK implements.

Figure 4: KOMPAK Problem Driven Activity Development Process



Source: KOMPAK Internal Presentation, 2018.

5.12 KOMPAK has also adopted a twice-yearly review and reflection process with partners (KOMPAK 2018: 11). This appraisal process is underpinned by the collection of activity 'progress markers.' These monitor change at the outcome level and describe what the Facility is hoping to witness within one year. They are considered "a step along the way to realising (..) larger scale change" (KOMPAK, 2018b), and provide "early indications that a system may be changing; for example: changes in the availability or flow of information within the system; changes in rules or regulations that govern the system; the involvement of new actors or changes in interaction between existing actors; or changes in the perspectives of parties who have influence over the system (e.g. formal or informal leaders)" (KOMPAK, 2018b:11).

5.13 Notwithstanding the aforementioned strengths of the KOMPAK MELF, it is important to acknowledge that the Facility is still in the early days of implementing this approach. Its MEL framework has been redesigned on several occasions, and early indications suggest that many aspects will continue to require modification. As has also been the case in PNG, this is symptomatic of the mixed (MEL) capacity of the team (on both the management and donor side), and the changing demands made by DFAT and GoI throughout the Facility's life (M&E adviser, Pers. Conversation; Feb 2019).

5.14 However, there is evidence of progress. During KOMPAK's pilot phase, the focus was necessarily on program delivery and implementation, but presently — some two years into its life cycle — interviewees describe being aware of the "need to focus less on the transactional 'doing' of delivering the programs (e.g., improving how clinics are run etc), and more on the consolidation of knowledge management (e.g., testing models and analysing how effective they have been)"

[7] Activities are reviewed at different points throughout the year, and at times that tie in with stakeholder reporting requirements. Example criteria for KOMPAK ceasing an activity include high cost, resource intensive, declining GoI commitment, and a failure to meet performance milestones. Activities valued above AUD 250,000 require more rigorous approach to scheduling, program logic and M&E.

(Interview with Program Staff, Nov, 2018). As one program staff member commented, "We need to get better at looking across the entire portfolio, and understanding how we have influenced change, package that knowledge and be able to pass it on." (Interview with Program staff, Nov. 2018). These insights are positive. They suggest a sophisticated understanding of adaptive programming in Facilities, with teams able to describe how KOMPAK tests, learns and iterates, and identify where it needs to do better. Given how little the international literature has to say on MELF for Facilities and complex portfolios of investments, there is certainly scope for a margin of error.

Timor-Leste

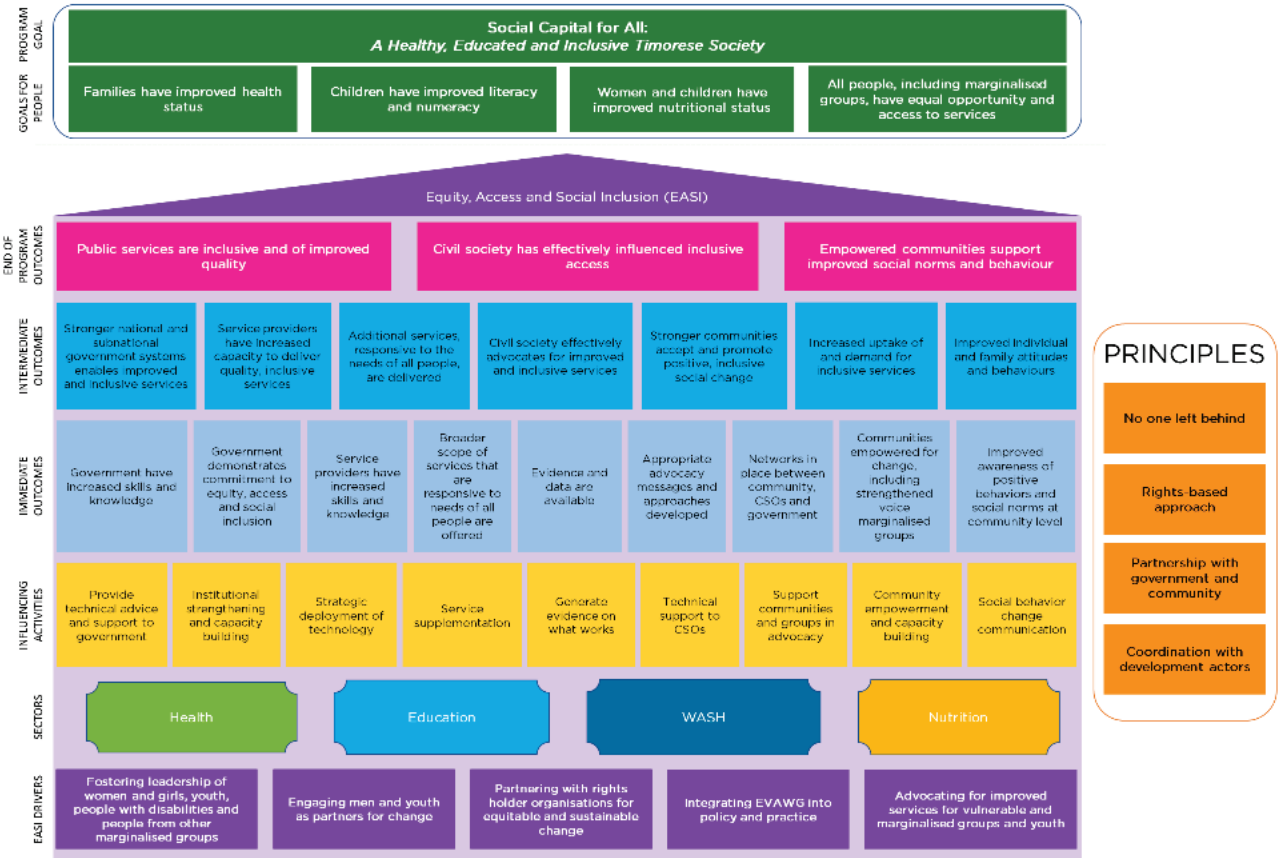
5.15 The Timor-Leste MELF has had a rather inauspicious beginning. Although external MEL support was specified at design stage in the form of external M&E support (M&E House) (DFAT, N.D.c), the Australia-Timor-Leste Partnership for Human Development (ATLPHD or PHD) began programming without this. As a result the development of the MELF was delayed[8]. Stakeholder interviews provide some insight into possible explanations: lower than expected M&E capacity (both in-house and donor teams), the challenge of designing a flexible MELF while supporting legacy grants and grantees, and the lack of clarity on Facility strategic intent. In many ways, it was a perfect storm.

5.16 Moreover, early iterations of the MELF (2017, 2018) struggled to gain traction as they attempted to set a 'vision' of the PHD that departed from the original design. For example; the 2017 MELF specified three high level changes: better human development outcomes for Timorese people; improved government commitment to service delivery; and change brought about by different sectors working together (e.g. health and education). Such outcomes relied on eight sectors within PHD being able to coordinate internally, and across the Timorese Government (sub-nationally, nationally and across different agencies) to improve services at the front line. With high levels of the PHD budget already committed at design (>70%), weak incentives for government agencies to work cross-sectorally, and a national election looming, it became apparent that PHD would find it difficult to trigger the sort of change needed to achieve these goals. Thus, the Facility strategy, and MELF was once again revised.

5.17 Similar to KOMPAK, the current PHD MELF uses a nested program logic to capture performance related information. Three cascading TOCs (one for each EOPO) sit under a single overarching, Facility-wide TOC. Figure 5 illustrates this process. Each of the EOPOs are linked to key evaluation questions, which are used to evaluate Facility level performance. Below this sit the Intermediate Outcomes, which can be described as sub-questions and which in turn, link higher level questions to EOPOs. Attached to each sub-question are TOCs for each of the main sectors (health, education and WASH). These TOCs relate to the overall TOC but are tailored to the interventions and modalities of each sector. For example, for Water Sanitation and Hygiene (WASH), the TOC looks at *"…'to what extent have we [ATLPHD] contributed to making rural water services more inclusive and reliable' and 'to what extent are communities demonstrating improved hygienic behaviours' with the focus on target communities where we [ATLPHD] have WASH programming"* (Interview with Program staff, Nov, 2018).

[8] Indeed, it wasn't until June 2018 – two years into implementation – that an M&E advisor was brought on for the facility.

Figure 5: ATLPHD Program Theory of Change (revised September 2018)



Source: Partnership for Human Development - Australia Timor-Leste, 2018, *PHD Guiding Strategy (DRAFT),* Version 1.0

5.18 At the Facility level, the team have yet to finalise a set of indicators that can collect information on program effectiveness (i.e. impact) and process related outcomes, such as efficiency and management. At the sector level, indicators focus either on the (a) effectiveness of interdisciplinary programming (for example, how well PHD worked across its four sectors, or how effectively gender and social inclusion was mainstreamed and applied across programs), or (b) on achievements made by a specific sector[9].

5.19 The sector level indicators (60 in total, with approx. 20 per program) are both quantitative (e.g. the number of service delivery improvements in a given area, or the number of people reached by particular services) and qualitative (e.g. instances of significant or behavioural change in particular area, or instances of people changing their behaviour around women and girls in leadership scenarios) (Interview with Program staff, Nov. 2018).

[9] It is worth noting that the program also tracks indicators in the Australian Government's country-wide Performance Assessment Framework. These indicators help Australia monitor performance across its entire portfolio of aid investments (i.e. beyond ATLPHD). The indicators were set and are managed by M&E House.

6. Key Findings: Implications for International Theory and Practice

6.1 When Abt's Facility experience of MELF design was compared to the international literature and donor practice notes on traditional MEL frameworks, seven lessons and areas of difference were identified. These are summarised in Table 2 and expanded in text below.

Table 2: Comparing the Facility Experience to Common Donor Guidance (Table 1)

Issue		Findings
Strategy	(i) strategic intent	• One of the most challenging objectives - to clarify the Facility's strategic intent, overall Theory of Change and an agreed strategic plan which would guide the development of the MELF • Related to this: distinguishing between Theories of Change and Theories of Action, and ensuring the former is given sufficient attention in 'M' and 'L'.
	(ii) purpose of MELF	• Expectations of MELFs were far more complicated than anticipated. • MELFs had to serve multiple purposes - accountability, public diplomacy/ communication, evaluation and internal learning.
MELF structure	(i) program logics and theories of change	• Nested frameworks and aggregation of results generally do not suit the Facilities studied - more complex, mixed methods approaches are required. • Have to be able to not only understand what development impact the Facility has had - but also test the validity of the modality itself

	(ii) indicators and defining 'success'	• The higher up the program hierarchy we went - the harder it became to understand and measure change
		• Relatively good at setting and tracking output level change - but much harder to understand and set indicators at the outcome and goal level: why?
		  ○ Success is hard to measure (i.e. complex institutional change)
		  ○ Attribution impossible at outcome level but contribution possible

	(iii) Baselines and data quality	• Traditional approaches to baselining hard to apply to constantly changing portfolio • Qualitative, mixed methods (e.g. strategy testing) combined with some form of external review/ verification much easier to apply at Facility-wide level • The quality of the country's own data institutions and capacity also has significant impact on ability to set baselines and undertake constitution analysis
MELF Implementation	(i) learning, reflection and program adaptation	• The second most challenging objective - to actually use information, learning and MELF data to adapt programming (budgets, activities etc) in real time • MELF teams and systems usually separate to implementation • Incentives meant output level reporting often trumped program learning, reflection and adaptation activities - at least in the early stages of implementation
	(ii) resourcing and legacy investments	• Very difficult to find appropriately skilled staff who are both technically strong on MEL approaches and understand adaptive programming • Tendency to rely on international experts which reduces capacity in-

		house • Donor staff move away from aid specialist to policy generalist skills • Funding for MEL activities generally not allocated at design and sometimes a lower priority at implementation

Program Strategy

Finding #1: Strategic Intent

6.1 Identifying each Facility's strategic intent has proved to be one of the most challenging aspects of MELF development. All three Facility investment design documents (IDD) set high-level goals but did not articulate a medium to long-term implementation strategy or theory of change. Although this was a deliberate decision – each of the Facilities were designed to have high levels of flexibility during implementation[10] – it meant that, unlike more traditional projects (where inputs, outcomes and outputs are tightly defined at design), much of the work a donor would normally do before tender had to be done during the first 12 months of implementation. Ongoing projects needed to continue operating seamlessly, yet operational aspects were still pending or being developed (staff recruitment, development of program platforms, and meeting expenditure targets, for example). For the PNG Governance Partnership (formerly PGF), the challenges began before novation: "the protracted contract negotiation process between DFAT and the Contractor reduced the scheduled transition period available to set-up the PGF's operations and undertake strategic planning." (QTAG, 2018:22). Further, the "complete turnover of relevant A-based staff at [DFAT] PNG post during transition added to the challenge during this process. This introduced new perspectives and views with little reference to of knowledge of the extended design phase." (QTAG, 2018:22). For these reasons, during the first 12 months, opportunities to define strategic intent were put on the back-burner.

Box 2: Theory of Action Vs. Theory of Change

"A Theory of Action is the delivery model for a Theory of Change. A Theory of Change describes the processes through which change comes about for individuals, groups or communities. A Theory of Action articulates the mechanisms through which the activities are being delivered, e.g. through which type of actors (for example, NGOs, government or markets) and following what kind of processes (for example, grants to NGOs disbursed from a challenge fund, provision of technical assistance, advocacy activities, facilitation of or the establishment of partnerships)." – Source: What is a Theory of Action, Coffey

6.2 Under typical circumstances, an IDD will detail a single, clearly articulated program logic and 'Theory of Change' (ToC) – assumptions about how change occurs in the environment the program is operating (how inputs / activities lead to outputs, how outputs lead to End of Program Outcomes (EOPOs), and how said outcomes contribute to overall investment goals). Next, beneficiaries and delivery mechanisms are identified, a process which forms the program's 'Theory of Action' (ToA) – or what it is the program plans to do to affect change. Typically ToAs are supported by TOCs, and together they form the backbone of a program (see Box 2).

6.3 However initially, in the three Facilities Abt manages, the ToCs were closer to ToAs: describing what the Facility would do. On one hand this is not problematic, however, in practice, the absence of ToCs meant that each Facility focused on monitoring what their *activities and projects* were seeking

[10] In the case of KOMPAK, this was a deliberate decision to allow the managing contractor the flexibility to 'design and implement' simultaneously.

to do (i.e. their technical appropriateness) – not the *underlying drivers of change* which a ToC would articulate. This meant that each Facility was not systematically monitoring and reflecting on the political context (i.e. local actors, institutions and coalitions), how it was changing, and thus what opportunities, risks or challenges this presented the Facility's work and the feasibility of achieving Facility outcomes and goals. In each case, it was the process of developing ToCs (at the project or Facility-wide level) which brought key political assumptions to the fore and helped teams become more politically-aware in how they were programming.

6.4 Each of the three Facility outcomes were also expressed as open-ended statements such as 'better education systems' or 'institutional change'. The PHD IDD for example, describes the Facility structure as four pillars and draft goals, linking to eleven EOPOs, and writes that the Facility *"will work towards the Goals and EOPOs (…) which will form the basis of a more detailed M&E Plan that will be developed during the inception phase of the program"* (DFAT, N.D.c). In the PNG Governance Partnership IDD, the end-of-Facility objective is described as "*improved programming to promote processes and institutions for stability and inclusive growth in Papua New Guinea.*"[11] This was to be accompanied by a Governance Strategy (which would outline the Facility's strategic intent), provided by DFAT to the managing contractor after the tender was awarded. However, this did not eventuate and left a critical gap in the Facility's strategy development – and thus the MELF.

> During interviews, one facility staffer commented that although there was: *"reasonable clarity at the contracting stage (…), that was subverted in the early implementation stage. Ideally, we need consistency in strategy while acknowledging [that the facility is] responding to lessons learnt during implementation. A major issue has been, is [the facility] here to strengthen government systems or not?"* (Survey interview, program staff member, 21 November 2018).

6.5 Each IDD also failed to clearly identify who (donor, implementer or partner government) would be responsible for creating and setting the strategic direction for each Facility – and where responsibility lay for adapting it over time. For example, the PHD IDD specified the managing contractor as responsible for providing advice on strategic direction and analysis, with the Australian Embassy in Timor-Leste responsible for consulting with GoTL on decision making, and strategic oversight (ATLPHD, 2018). However, another PHD document acutely observes that there is no "further detail on the conceptual or practical difference between these two terms [strategic oversight and advising on strategic direction]." (ATLPHD, 2018:2).

6.6 This finding is critical for one reason. In order to know *how to measure* whether the Facility is on track to achieve its overall aims (i.e. the MELF), Facilities must first know what it is they are *trying to achieve and why* (i.e. the strategic intent). In each case, the lack of clarity regarding strategic direction and who is responsible for it, crippled the development of the Facility MELFs. At the time of writing, only one Facility, KOMPAK, had a revised and agreed MELF in place.

Finding #2: Purpose of MELFs

6.7 The experience described above suggests that the demands which large-scale, complex and multi-sector investments place on MELFs are greater than is addressed in current donor guidance notes.

i. Accountability: Most donors are, first and foremost, interested in accountability. There are two components to this: (i) public accountability; being able to justify to domestic constituents and partners what the investment is achieving and (ii) performance accountability; data that allows the donor to determine whether the investment is an effective and efficient use of their funds.

[11] See p.5 of https://dfat.gov.au/about-us/publications/Documents/png-governance-facility-design-document.pdf

Meeting these two purposes is an ambitious task. For example; the data required by domestic constituents (e.g. personal anecdotes of change) is different to that required by the partner government (e.g. reporting on expenditure and achievements at the agency level).

One way KOMPAK is attempting to manage this demand for reporting is through the development of 'snapshot' indicators (Fig. 6, below). These "*relatively simple performance measures (…) help the Facility to communicate aggregate results (from across multiple Activities) to key stakeholders"* (KOMPAK, 2018b: 20). Assembled using a traffic light system, they present a brief overview of implementation against annual workplan, and document "whether individual sub-activities are on track, delayed, or have been cancelled" (KOMPAK, 2018b:10). Although not formal measures of performance, these indicators allow the Facility to partially meet stakeholder demand for information. Crucially, these indicators are easily accessed by internal team monitoring and learning processes. This ensures that instead of adding an additional reporting layer, they support teams by collecting data which is directly relevant to activity appraisal processes.

Fig. 6: KOMPAK's snapshot indicators

#	Indicator
1	Additional GoI (national/province/district/village) financial resources (in AUD) committed to KOMPAK-supported initiatives.
2	Number of pilots supported by KOMPAK.
3	Number of women/men/persons with disabilities reached through KOMPAK capacity building efforts.
4	Estimated number of women/men/persons with disabilities who obtain legal identity documents with direct/indirect support from KOMPAK.
5	Number of districts that implement improvements to service delivery systems.
6	Number of villages that implement improvements to service delivery systems.
7	Percentage of KOMPAK Activity Budget allocated to: • GESI-targeted activities • GESI-mainstreamed activities.

Source: KOMPAK, 2018b, Performance Management framework 2018 – 2022

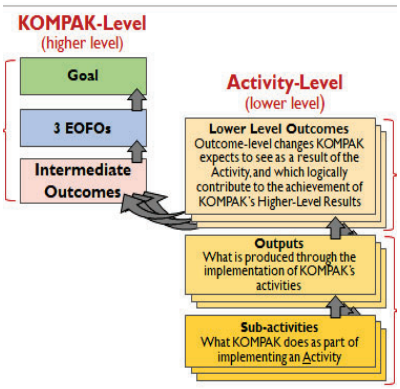
ii. Learning: Second, and perhaps the most difficult purpose to for the MELFs to achieve, has been to promote program learning and adaptation. Putting in place the systems for monitoring and learning - while often under-resourced - is the easy bit; but understanding the information these systems generate and systematically applying them in programming has been harder. The reasons for this are examined in Finding #6 below.

MELF Structure

Finding #3: Project level logics and theories of change

6.8 All three Facilities have found that 'traditional' approaches to program level MEL do not suit the Facility model. Historically, MEL frameworks were designed to apply to individual programs not portfolios of investments. Thus more 'traditional' MEL approaches, particularly aggregating results from one level of the project frame to the next, have tended not to work. As such, each Facility has had to develop its own approach to explaining how its constituent parts are contributing to an overall Facility goal.

6.9 The development of KOMPAK's MELF has been iterative. Initially, the Facility developed multiple theories of change for each area of work, but although this enabled flexibility, it was hard for staff to trace the contribution of their activity or project to the overall Facility goal. *"The many different 'Theories of Change' at the activity level and the lack of adequate baseline data (..) made it difficult to consolidate information on KOMPAK's overall achievements."* (KOMPAK Performance Management framework 2018 – 2022, p.4). After a number of attempts, the Facility shifted to a single overarching program logic, an approach which links *lower level* outcomes and outputs to its *higher-level* outcomes and goal [emphasised in figure on right] – and (at least in theory) enables all activities to draw a line of sight between what they are doing, and improved services and economic opportunities for the poor (the goal). In practice, not all activities have yet been able to develop this logic and link their work as clearly as hoped to higher order outcomes.



6.10 Similarly, the PNG Governance Partnership found there was no single ToC which was able to unite the various projects and grants to explain how they all worked together to achieve a common governance outcome. Instead, the team has adopted process-level indicators at the Facility-wide level to measure how well each work-area is responding to change, collaborating and coordinating its efforts, and achieving value for money. The overall development impact of the Facility is then measured qualitatively, through an annual strategy testing and review and reflection process.

Finding #4: Indicators and showing 'success'

6.11 In all three case studies, the Facility level MELFs were - at least initially - better at explaining lower-level change (at the input/output level) and weaker at explaining higher-order change (the outcome/goal level). Staff were proficient at collecting output level data, but the higher up the project hierarchy they went, the harder it was to understand change and tell a convincing story about how outputs led to Facility outcomes, and how outcomes contributed to the Facility goal.

ATLPHD described the challenges they had experienced: "*For [our Facility] this is a particular struggle in health, where our investments were never designed coherently to lead to a specific goal, which is why the TOC process is particularly challenging and more of a retrofit. Contrast to education, where the investments are more targeted and coherent.*" (Survey interview with program staff, 21 November 2018).

6.12 This was largely a result of two factors. The first well understood in complexity literature, but the second less so.

i. Success - when understood as complex institutional change – takes time and is very hard to measure. In the case studies here, there was a desire from DFAT to demonstrate tangible results and early on in implementation. This was, in part, driven by overambitious expectations about what the new Facilities could deliver, and the need to justify why the modality was effective. However, the reality was that all three Facilities struggled to make their initial achievements visible in their first year: in particular, demonstrating the value of the Facility to partner governments and donor headquarters.

There are a number of reasons for this. In their first few months of operation, the volume of urgent work confronting the Facilities was overwhelming, and partner engagement and upwards communication of achievements was slow. Additionally, each Facility was trying to affect long-term, non-linear political change, processes which do not lend themselves to easily communicable results. In the PNG Governance Partnership for example, indicators of the Facility's success include long term advisers building relationships in government, convening

discussions and producing analysis over several years, before there is anything 'substantial' to show from this work. Indeed (and similarly in the case of KOMPAK), the real developmental changes may not be apparent until well beyond the life of the Facility. Furthermore, in the case of PNG, the very act of deepening the relationship between Australia and the host government – through its projects and advisers – is itself a legitimate long-term goal for both governments: but one which is not commonly communicated by aid programs.

ii. The tension between contribution and attribution: *"...in the context of the current focus on results, donors are very interested in being able to draw a direct link between their inputs and the results identified"* (Roche and Kelly 2012). However, for programs seeking to work in adaptive, flexible and politically-informed ways, the contribution of the donor-funded project may be small or hard to trace. For example; supporting a consultant to advise a foreign government on a policy reform matter. In these cases, it is hard conclusively to prove that the input (or the donor's investment) caused a complex policy or institutional change. The reality is that factors outside donor control will conspire to bring about change. The Facility experience does, however, demonstrate that alongside the actions of local actors and local events, it is possible to demonstrate the *contribution* of a donor-funded project to institutional change. In addition, it is also possible to demonstrate results as implementation occurs - although almost never at goal or outcome level until the Facility ends.

6.13 Donor practice notes emphasise the importance of setting clear, goal and outcome level indicators. However, portfolios of investment or Facilities are more likely to require a range of evaluative questions and methods to explain how change is occurring at the outcome and goal level. Given the open-ended nature of each Facility's outcomes, it is often more appropriate to define a mix of evaluative questions and qualitative and quantitative indicators to explain what changes and results are being witnessed. This mix of questions and indicators allows for a messier change story to be told. Rather than dictating a single, linear path to 'success', evaluative questions can capture many different ways in which the Facility's activities contribute to change.

This approach also allows for more flexibility; activities can be adjusted during implementation and still find ways to show how they are relevant to the Facility's overall markers of success through one or more pathways to change. Based on M&E house/Clear Horizons' recommendation, this is the approach adopted in Timor by ALTPHD. However, program staff commented that although this approach has been effective, "*it takes time to establish, and the lack of clarity on strategy is a huge bottleneck."* (Survey interview, ALTPHD program staff member, 21 November 2018)

Finding #5: Baselines and data quality

6.14 In most donor monitoring systems, it is considered good practice for an investment to provide information on its starting point. This can inform future investment decisions, activity planning and provide useful feedback on progress towards outcomes. However, progress on the collection of baseline data has varied across Facilities. Some Facilities (e.g. KOMPAK) and certain work areas (e.g. the Bougainville program in PNG) have found it easier than others. This has had a significant impact on the ease with which each Facility has been able to demonstrate advancement towards its Facility level goals and outcomes.

6.15 Because each Facility was constantly changing in response to the donor and/ or host government's policies and preferences, it was difficult to set baselines during the first year, if not longer. When baselines were set (as in the case of KOMPAK) these quickly became redundant as the program strategy, budget and activities were revised. Furthermore, it became apparent that quantitative means of verification were often insufficient or inappropriate for measuring the types of change the Facilities were trying to affect. As a work around, some of the Facilities adopted mixed methods approaches - such as stakeholder interviews, combined with policy document review and training output data - to try and paint a more nuanced picture of what was changing and why. In other cases, baselines were able to be constructed for discrete parts of the Facility (e.g. activities

in a single province in Indonesia), rather than at a work-stream or Facility-wide level. In short, some form of baseline was better than no baseline.

6.16 In the PNG Governance Partnership, it was a struggle to generate baseline data, beyond one or two discrete projects, for a number of reasons.

i. The lack of a Facility strategy inhibited the early development of a MELF, and immediately after novation, the focus was on establishing and implementing the necessary systems and processes to get the Facility up and running. Constructing baselines was simply a lower priority. The following passage describes the early period: "*at inception, the Facility focused on creating a skilled team to support the Australia–PNG partnership. This was followed by the development of an interim and the Long-Term Development Strategy, which provided clarity and direction as to how and where programs would operate. After the elections in mid-2017, and the launch of the Government's Alotau Accord II policy and the 100 Day Plan to which the Facility EGIG partnership contributed to, GoPNG priorities have become much clearer.*" (Abt Associates, 2017:3).

ii. The Governance Partnership also found it easier to construct baselines for projects that were heavily planned and well-established – rather than projects trying to adopt TWP and adaptive management approaches. For example, the Bougainville program, which features a traditional M&E framework, collected early data on a number of indicators in the Commodity Support Facility (CSF)[12]. Thus, once the grants were novated into the Facility, the Bougainville program was able to report on objectives such as increasing the volume and quality of cocoa production, and cash flow to cocoa farming families and communities; progress made towards controlling the spread of pests such as cocoa pod borer, and improvements in plot management (Bougainville Program, 2018:25). However, in the Decentralisation and Citizen Participation (DCP) workstream, their adaptive and politically informed approach saw them waiting some time before deciding which districts they would work in; to this day, the DCP workstream continues to adapt their program approach based on where they will have greatest impact.

   Overall, this discrepancies between projects has meant that baseline data exists for some parts of the Governance Partnership, but not others, making it challenging to describe the overall contribution of the Facility towards its outcomes.

iii. KOMPAK by contrast, features operational differences which have allowed it more easily to capture some baseline data. KOMPAK started life piloting interventions in a small number of districts. When these pilots performed well, they were taken to scale and replicated across other provinces and districts. This means that data from small, localised pockets, on specific indicators that are targeted by the pilots, has in some instances, been collected. It has also performed an analysis of 2015 village and district level budget-related data (line items etc) and data from village plans, and in 2017, conducted a survey at household, service unit, village, district and subdistrict levels.

6.17 The challenge of using baselines to describe the Facility contribution has also varied according to the availability, accuracy and sophistication of secondary data sources. In countries with sophisticated data institutions and capacity, it has been comparatively simpler to describe the Facility contribution by using external data sources to construct baselines. In countries where these conditions are not met (for example, PNG), it has been much harder to demonstrate progress by the investment towards goals. In Indonesia, robust secondary governance related data sources exist in the form of the National socio-economic survey (SUSENAS), the National Workforce Survey (SAKERNAS), and the Village Potential Survey (PODES). These were used by KOMPAK to describe the state of governance in Indonesia at the start of the investment. In PNG, however, the conditions under which data is collected, stored, analysed and managed are very different; a context which

[12] The CSF is an economic development initiative intended to provide support to Bougainville primary industries.

affected the Governance Partnership's ability to describe the state of governance at the start of the investment.

MELF Implementation

Finding #6: Learning, reflection and program adaptation

6.18 'Learning by doing[13]' lies at the heart of the TWP and adaptive management agenda: yet it was undoubtedly one of the most challenging parts of applying MEL to the Facility context. In order to apply a 'learning by doing' approach programs have to collapse design, implementation and monitoring and undertake these tasks simultaneously, not sequentially as in the case of more mainstream projects. The argument is that projects will be more effective at achieving development outcomes when they are able to understand, learn from and respond to new information and the political context in real time; rather than having to wait until mid-term review or end of program evaluation. All three Facilities studied here were designed with the specific aim of being able to apply the 'learning by doing' principles across some, or all, of their portfolio. Yet in each case, teams focused on the 'M' (monitoring) at the expense of the 'E' (evaluation) and 'L' (learning) – at least in the initial stages of implementation.

6.19 From launch, each Facility was immediately able to start generating output level monitoring data, however, it has taken much longer for evaluative and learning processes to emerge - and more importantly, become a tool for adapting programming in real-time. There are five reasons for this:

i. the time and scale of input required to establish Facility operating systems and novate existing projects was underestimated. This distracted from the less 'urgent', but fundamentally critical tasks, of setting Facility wide strategies, ToCs and MELFs;

ii. output reporting to stakeholders necessarily superseded reflection and learning. One Facility noted the *"inordinate amount of time the project has had to spend on reporting, which takes away resources from efforts to develop and implement more in-depth documentation, evaluation, analysis and strategic thinking. The six-monthly reporting cycles and output focused [donor performance assessment] indicators reinforce 'monitoring' rather than allowing for longer term processes to be established."* (Survey interview with program staff, 21 November 2018);

iii. MEL functions and systems were initially kept separate from implementation teams. Too often MEL units were created to fulfil corporate reporting requirements, rather than project-based learning requirements. Program teams often did not see learning or monitoring as part of 'their job';

iv. MEL and implementation skills were *"...often difficult to find in the same person. Staff tend to specialist in implementation or M&E - largely because of the way the project cycle has always been depicted"* (Teskey and Tyrrel, 2017).

v. Lastly, learning processes were rarely linked explicitly to budget and activity/design decision making processes. In all three cases, the appetite for internal learning grew dramatically as each Facility matured, yet it was a struggle for teams to make direct changes to their strategies, budgets, activities and partner arrangements as a direct result of learning processes. Sometimes this was due to the view that 'learning is for learning sake'. It was not deemed a legitimate or valuable reason to change program direction. Whereas other times the desire to adapt in response to learning was thwarted by disincentives from the donor. In program interviews, respondents repeatedly referred to the challenge of risk aversion combined with overly onerous approval processes for micro-level activity changes. *"Sometimes the hassle of*

[13] Also known as 'iterative step-wise learning' or adopting 'rapid feedback loops'.

changing the program outweighed the value we saw in doing it!" (Program Interviewee, Anonymous, December 2018).

6.20 It is clear that the larger and more complex the investment, the harder it is to operationalise a 'learning by doing' and 'adaptive management' approach to MEL. Despite the rhetoric, both donors and implementers alike have tended to prefer the more 'traditional' features of each Facility: aspects which separate design, implementation and MEL are still found in all three facilities. These are perceived as more easily understandable and an effective way to generate quick development impact stories. This suggests there is an institutional shift which needs to occur within donor agencies and implementers alike to collapse the project cycle and demand that implementation, monitoring and learning be combined. Teskey (2017) has termed this the 'TWP project framework', an alternative to the standard project frame. In the absence of such a fundamental shift, the disjuncture will probably continue.

Finding #7: Resourcing and Legacy Investments

6.21 (i) Resourcing: the final point to emerge from the three Facilities in this study is that MEL resourcing has been problematic – especially when it comes to 'Learning', which is often the first 'non-essential' activity to be cut when funding is tight. Each Facility struggled to attract, retain, fund and equip teams with the appropriate data collection, analysis and MEL skills. There are three aspects to this. (i) In the early days of project mobilisation and implementation, the urgent supersedes the important, and MEL considerations drop off the radar. It is particularly important therefore, to ensure budget is quarantined for MEL activities and for MEL designated staff. (ii) As one Facility commented: *"We tend to under-resource for learning. The skillset is usually geared more toward M/E and less so on learning."* Often MEL is viewed as a highly specialised set of skills, separate to program management capabilities. As such, each Facility tended to rely on the help of technical advisers - at least initially - rather than building and embedding "M" and "L" capabilities within program teams themselves. (iii) The three Facilities were mobilised at the same time as the donor moved away from aid specialist, to policy generalist skills. This as DFAT themselves conceded (2018), reduced the donor's ability to provide technical oversight of MEL processes. As such, the managing contractor needed to recalibrate its own workforce to be able to supply these skills in-house. In the case of PNG, for example, this was achieved through a combination of in-house program specialists, governance and development expertise from the Managing Contractors head-quarters, as well as oversight from MEL specialists in the external Quality and Technical Assurance Group (QTAG)

6.22 (ii) Legacy investments: to varying degrees, the Facilities reviewed here inherited historical areas of program focus, programming styles, objectives, and stakeholders. Often this has meant that it has been challenging for them to (a) program flexibly, and (b) program strategically. As one program staff member commented, *"…the politics and history of the programs have affected their impact - 87% of operating budget has gone to legacy programs. It's hard to deliver innovative programs (under the Facility) if majority of budget is funding legacy programs and partners"* (interview with ALTPHD program staff, November 2018). KOMPAK too, began as an amalgamation of existing projects, having inherited previous DFAT funded investments (AIPD, PSF, and CSO). This made the development of MELFs difficult; they not only had to reflect the aspirations of the Facility but find a way to retrofit indicators and the means of verification to existing projects which had existing performance frameworks in place. As program staff from ATLPHD commented, *"much of PHD is still founded on long running legacy programs thus the strategic intent is being uncomfortably retrofitted"*.

7 Conclusion

7.1 This paper has focused on three critical aspects of Facility management: monitoring evaluation and learning. Based on a review of three Australian-funded 'governance' Facilities which Abt is managing (ATLPHD, KOMPAK and the PNG Governance Partnership) – this paper has found seven areas where serious deviation from more 'traditionally' designed and implemented MELFs has been required. These differences have emerged at all points in the programming and management cycle: from clarifying strategic intent; to developing clear, adaptable theories of change; to defining baselines and indicators for constantly changing investments in data poor environments; or the difficulty of embedding learning in programming or recruiting staff who know how to apply MEL to projects working in adaptive and politically-informed ways.

7.2 The critical lesson to emerge is that traditional forms of MEL do not lend themselves well to the Facility model. This stems from one simple fact. Traditional forms of MEL are based on a planned and largely linear project model. This model may work in simple change contexts, where there is a clear line of sight between activities, inputs, outputs and outcomes. Not complex projects working in complex political contexts, where institutional change is often the underlying goal. Thus, in each case examined, this paper has found that teams have needed to develop their own unique a mix of conventional and experimental approaches to MEL to try and overcome these challenges.

7.3 In concluding, and if the international community is serious about transforming how complex programs and complex change are measured – then the place to start is not MEL methods, but the logic of the project framework itself. An impediment to effectively applying MEL to Facilities is the dogmatic use of linear change models across the aid industry. Although the project framework serves a critically important accountability purpose (to give donors confidence about what they are 'buying' with their aid funds), in some cases, it has been mistaken for a hard performance benchmark – working actively against more flexible and adaptive forms of program management. It is also extremely difficult to describe, summarise and plan a (often expermential) portfolio of investments using a linear change approach. In these instances, the project frame incentivises donors and implementers to separate MEL from programming and lock in activities, outputs and outcomes up-front. MEL thus become a compliance tool to 'track' whether the project will achieve the pre-determined outputs it set itself, rather than a way of learning about the context, what is and is not working, and why – and adapting the program in response.

References

Abt Associates, 2017, *Six-monthly Implementation Report on Abt PNG Management Services Support to the PNG-Australia Governance Partnership*, July–December 2017

Bayley, S, 2019, *Personal communication*, Canberra, 10 Jan 2019

Buffardi, A., and Hearn, S., 2015, *Multi-project programmes: functions, forms and implications for evaluation and learning. Methods Lab Working Paper*; Overseas Development Institute: London, UK., December 2015

Cole, W. Ladner, D. Koenig, M and Tyrrel, L., 2016, *Reflections on Implementing Politically Informed, Searching Programs: Lessons for Aid Practioners and Policy Makers*: Case Study 5, The Asia Foundation: Thailand.

De Lacy, J., 2017, *Facilities deserve a place in development,* The Governance Soapbox Blog, November 8, 2017

Department Foreign Affairs and Trade (DFAT), N.D.a, *DFAT internal practice note on Facilities.*

Department Foreign Affairs and Trade (DFAT), N.D.b, *Australia Timor-Leste Partnership for Human Development Investment*

Department Foreign Affairs and Trade (DFAT), 2015, *Design Document Investment Design Document For Papua New Guinea Governance Facility (PGF),* Final version (9 July 2015), accessed 30/01/2019 <https://dfat.gov.au/about-us/publications/Documents/png-governance-Facility-design-document.pdf>

Department Foreign Affairs and Trade (DFAT), N.D,c, *DFAT Strategic Framework 2015-2019*

Engen, J. et Al, 2017, *KOMPAK Refresh, December 18th, 2017, Jakarta.*

Faustino, J. and Booth, D., 2014, *Development Entrepreneurship: How Donors and Leaders can Foster Institutional Change,* Working Politically in Practice Series: Case Study 2, The Asia Foundation: Thailand.

KOMPAK, 2018, *KOMPAK Program Logic and Ways of Working 2018-2022,* August 2018

KOMPAK, 2018b, Performance Management framework 2018 – 2022

KOMPAK, 2018c, *Program Logic and Ways of Working Strategic Planning,* ISA Mission, 29 - 30 Oct

Ladner, D, 2015, *Strategy Testing: An Innovative Approach to Monitoring Highly Flexible Aid Programs*, Working Politically in Practice Series: Case Study 3, The Asia Foundation: Thailand.

Partnership for Human Development - Australia Timor-Leste, 2018, *PHD Guiding Strategy (DRAFT),* Version 1.0

Pasanen, T., et Al., 2018, *Outcome monitoring and learning in large multi-stakeholder research programmes: lessons from the PRISE consortium*, Overseas Development Institute: London, UK, December 2018

Quality and Technical Assurance Group (QTAG), 2018, *DRAFT Annual Review Report Papua New Guinea–Australia Governance Partnership,* June 2018

Roche, C. and Kelly, L., 2012. *The Evaluation of Politics and the Politics of Evaluation.* Background Paper 11. Developmental Leadership Program: UK.

Teskey, G. 2017. *Thinking and Working Politically: Are we Seeing the Emergence of a Second Orthodoxy?* Governance Working Paper Series: Issue 1. Abt Associates: Canberra.

Teskey, G. and Tyrrel, L. 2017. *Thinking and Working Politically in Large Multi-sector Facilities: Lessons to date.* Governance Working Paper Series: Issue 2. Abt Associates: Canberra.

The Bougainville Program, 2018, *The Bougainville Program (BP) Eighteen-Month Report July 2016-December 2017 – Draft*, February 20th, 2018.

USAID, 2018. *Discussion Note: Complexity Aware Monitoring. Discussion Note Version 1*, July 2018. USAID: United States.

Abt Associates

Abt Associates is an engine for social impact, dedicated to moving people from vulnerability to security. Harnessing the power of data and our experts' grounded insights, we provide research, consulting and technical services globally in the areas of health, environmental and social policy, technology and international development.

As the Australian subsidiary of Abt Associates, we work with our clients to implement bold, innovative solutions to improve the lives of the community and deliver valued outcomes for our clients. We provide a comprehensive range of services from policy to service delivery in the public and private sectors contributing to long-term benefits for clients and communities. Operating in remote and challenging environments, we offer extensive experience in the region, strong technical capacity, and a proven project management track record.

To learn more visit **www.abtassociates.com**

To follow the latest thinking on governance, visit the Governance Soapbox: **www.abtassocgovernancesoapbox.wordpress.com**

Abt

BOLD THINKERS DRIVING REAL-WORLD IMPACT

Printed June 2019