

aleph strategies

Independent Review  
ARTF Third-Party Monitoring Agent  
Performance

Final Report

January 2021



**THE WORLD BANK**  
IBRD • IDA | WORLD BANK GROUP





© Aleph Strategies Limited

Address: Bourne House, 475 Godstone Road, Whyteleafe, Surrey, United Kingdom, CR3 0BL

Company Number: 11198794

Authors: Jonathan Rider and Daniel Skillings

Contact: Jonathan Rider, Co-Founder and Director, [j.rider@alephstrategies.org](mailto:j.rider@alephstrategies.org)



**aleph strategies**

Impact Evaluation :: Policy Support :: Market Insight



## Acronyms

AAIP	Afghanistan Agricultural Inputs Project
ADB	Asian Development Bank
AFMIS	Afghanistan Financial Management Information System
AITF	Afghanistan Infrastructure Trust Fund
AMELA	Afghanistan Monitoring, Evaluation, and Learning Activity
APSDP	Afghanistan Power System Development Project
ARAP	Afghanistan Rural Access Project
ARTF	Afghanistan Reconstruction Trust Fund
CBKT	Capacity Building and Knowledge Transfer
CCAP	Citizens' Charter Afghanistan Project
CDC	Community Development Council
CM	Citizen Monitors
CMU	Country Management Unit
EQRA	Education Quality Reform in Afghanistan
EQUIP	Education Quality Improvement Program
ESF	Environmental and Social Framework
ESS	Environmental and Social Standards
ESMP	Environmental and Social Management Plan
FCV	Fragility, Conflict and Violence
GoIRA	Government of the Islamic Republic of Afghanistan
GSDRC	Governance and Social Development Resource Centre
HEDP	Higher Education Development Project
IP	Implementing Partner
IPF	Investment Project Financing
IRDP	Irrigation Restoration and Development Project
IW	Investment Window



KII	Key Informant Interview
KPI	Key Performance Indicator
MA	Monitoring Agent
MAIL	Ministry of Agriculture, Irrigation, and Livestock
MoE	Ministry of Education
MoF	Ministry of Finance
MoPW	Ministry of Public Works
MPF	Multi-Partner Trust Fund
MRRD	Ministry of Rural Rehabilitation and Development
MSI	Management Systems International
NHLP	National Horticulture and Livestock Project
NHRP	Naghlu Hydropower Rehabilitation Project
NPBSE	Non-Payroll-Based Salary Expenditure
NSP	National Solidarity Program
O&M	Operations and Maintenance
OFWMP	On-Farm Water Management Project
PAISA	Payments Automation and Integration of Salaries in Afghanistan
PBSE	Payroll-Based Salary Expenditure
PFEML	Public Finance and Expenditure Management Law
PFM	Public Financial Management
PMU	Provincial Management Unit
PPE	Personal Protective Equipment
RCW	Recurrent Cost Widow
RR	Reporting Requirements
SA	Supervisory Agent
SIGAR	Special Inspector General for Afghanistan Reconstruction
SoE	Statement of Expenditure



SoS	Scope of Services
THRCP	Trans Hindukush Connectivity Project
ToR	Terms of Reference
TPM	Third Party Monitoring
TTL	Team Task Leader
USD	United States Dollar
WBG	World Bank Group
WEE-RDP	Women's Economic Empowerment-Rural Development Program



## Table of Contents

1. Executive Summary.....	1
1.1. Topline .....	1
1.2. Monitoring Agent .....	1
1.3. Supervisory Agent .....	2
1.4. Observations and Recommendations .....	3
2. Introduction .....	4
2.1. Purpose and Scope.....	4
2.2. Methodology Note .....	4
2.3. Report Structure .....	5
2.4. Limitations to the Methodology.....	7
3. Background and Context.....	7
3.1. What is the ARTF? .....	7
3.2. Why does TPM exist for the ARTF? .....	8
3.3. Who were the contractors? What were they responsible for? .....	8
4. Monitoring Agent Review .....	9
4.1. Summary .....	9
4.2. Introduction: What was the MA contracted to deliver? .....	9
4.3. Design.....	10
4.4. Deliver .....	14
4.5. Results.....	18
5. Supervisory Agent .....	22
5.1. Summary .....	22
5.2. Introduction: What was the SA contracted to deliver? .....	22
5.3. Design.....	23
5.4. Deliver .....	28
5.5. Results.....	35
6. Literature Review of Third-Party Monitoring in FCV Contexts .....	38
7. Observations and Recommendations .....	41
8. Appendices.....	44
8.1. Appendix A: Detailed Methodology .....	44
8.2. Appendix B: Status Quo of MA Recommendations.....	50
8.3. Appendix C: Select Bibliography.....	55



## Table of Boxes, Figures and Tables

Focus Box 1: Summary of MA Headline Findings .....	9
Focus Box 2: Summary Eligibility Criteria .....	11
Focus Box 3: Summary of SA Headline Findings.....	22
Focus Box 4: Environmental, Social and Gender Safeguarding Standards .....	30
Figure 1: MA Monitoring Process, SoE (Statement of Expenditure) .....	11
Figure 2: Evolution of Eligibility Rate 2002 – 2019.....	21
Figure 3: Growth of the TPM portfolio.....	24
Figure 4: Example of outcome-level performance indicators .....	43
Figure 5: Volume of available information for the Review.....	48
Table 1: ToR Questions .....	5
Table 2: Analysis of MA Contractual Obligations .....	15
Table 3: Expenditure Claimed and Disbursed (USD millions). Source MA ARTF Completion Report ....	19
Table 4: Examples of tools employed on each program .....	25
Table 5: Summary of site visits by project.....	28
Table 6: TPM Cost Benchmarking .....	40
Table 7: Research Framework.....	45
Table 8: Key Informant Interviews .....	49
Table 9: CBKT Recommendations by MA .....	50



## 1. Executive Summary

---

### 1.1. Topline

Aleph Strategies conducted a Review of the Third-Party Monitoring (TPM) Agents for the Afghanistan Reconstruction Trust Fund (ARTF) engaged by the World Bank Group (WBG) from 2015 to 2019. A Monitoring Agent (MA) was contracted to support the ARTF Recurrent Cost Window (RCW), and a Supervisory Agent (SA) was contracted to support the ARTF Investment Window (IW). Contract extensions and amendments later during the contract period added responsibility to the MA for conducting internal control assessments and transaction reviews of IW-financed development projects, in coordination with the SA.

The Review Team asked three overarching questions:

- Did the TPM agents **design** an appropriate methodology to provide the contracted services?
- Did the TPM agents **deliver** these services in line with contractual obligations and quality standards?
- Were the **results** of the TPM agent services useful to key stakeholder groups?

**Overall, the Review Team finds that both the MA and the SA performed well on the delivery of their services. The MA and the SA delivered activities in line with contractual requirements and in line with WBG operational standards.** The Review Team finds that the MA and the SA provided an invaluable instrument for WBG Task Teams to monitor their programs and to provide assurances that government partners were implementing activities correctly. The MA and the SA were flexible in responding to the needs of donors, the WBG, and the Government of the Islamic Republic of Afghanistan (GoIRA), adapting approaches, tools, and reports to reflect new requirements throughout the duration of the review period. The MA and SA employed highly participatory approaches with GoIRA partners, ensuring strong collaboration and knowledge sharing.

**WBG stakeholders were generally satisfied with the utility of the reports produced.** By-and-large, the results allowed the WBG Country Management Unit (CMU) and WBG Task Teams to perform their jobs better in a volatile and fragile environment. **GoIRA stakeholders interviewed as part of this Review were generally positive about the function of both of the MA and SA**, and there is some tentative evidence to suggest that government capacity in executing and monitoring projects has been improved. **There seems to be a broad consensus among donors interviewed that the MA and SA are useful instruments.**

A summary of findings for the MA and SA is presented below.

### 1.2. Monitoring Agent

Did the MA **design** an appropriate methodology to provide the contracted services?

**The MA designed a strong approach for the RCW.** The MA designed its RCW approach in line with contractual obligations. It consulted with WBG and GoIRA counterparts to inform its methodology and sampling plan, and key informants agree that the approach employed was robust and transparent.

The MA's expanded scope in 2018 to conduct internal control assessments of IW projects filled a significant gap and stakeholders interviewed applauded the MA's flexibility and adaptability to take on this new role. The MA piloted a comprehensive methodology but as the approach was still in its nascent stage of development by the end of the contracting period, some WBG Task Teams interviewed had some suggestions on how the methodology could be refined moving forward. Suggestions included that the MA should conduct in-person site visits to better include local official views in the design process; sampling should be more representative; and reviewing how the methodology could be further tailored to meet the unique nature of each project. In regards to further tailoring the approach,





it is likely that WBG Task Teams just require further clarity as each internal control assessment is already adapted to assess unique project activities.

### Did the MA *deliver* these services in line with contractual obligations and quality standards?

**The Review Team finds that the MA delivered well on its contractual obligations.** The WBG former and current CMU remarked that the MA was professional, working to the letter of the brief quite well. GoIRA counterparts found the MA to be more of a ‘supportive partner’ rather than simply an ‘auditor’. Furthermore, the MA proved to be flexible; a welcome attribute given the volatile nature of operating in Afghanistan and the changing priorities and demands of ARTF stakeholders.

**Reporting requirements were delivered on schedule in line with objectives and timelines set out in the contracts.** The MA produced monthly, quarterly, and annual reports as well as withdrawal applications, the Capacity Building and Knowledge Transfer (CBKT) plan, and special reports when required, among others.

### Were the *results* of the MA’s services useful to key stakeholder groups?

In line with the ARTF administration agreement, the MA contract establishes the WBG and GoIRA as the MA’s direct audiences, but ARTF donors are a crucial additional audience.

**The results were useful for WBG stakeholders.** The WBG derives value from assurances to perform fund disbursement. The MA services fulfilled this function and came at a cost of 0.94% of total RCW gross disbursement over the contracting period. WBG stakeholders have noted that there may be opportunities to better integrate MA findings into WBG supported reform and capacity programming.

**GoIRA stakeholders found the reports, findings, and, in general, the partnership with the MA to be invaluable.** GoIRA derives value from receiving funds to cover civilian recurrent expenditures. As such, GoIRA stakeholders said they welcomed MA scrutiny of reporting in order to access funds. GoIRA was able to leverage further value from the Ineligibilities Working Group which was found to be an effective strategy setting and information sharing forum. Finally, there is much evidence to suggest line ministries found the IW internal control assessments to be a valuable addition later in the contract period, as ministries accepted and implemented most of the recommendations found in the reports.

While donors appreciated access to the full MA reports from 2018 onwards and therefore perceived more utility of MA services near the end of the contract period, there is demand for even further information (in particular on GoIRA’s public financial management capabilities). The WBG, donors, and the current TPM agent should clarify together what is and is not possible given TPM agent’s role and responsibilities. In addition, a common concern raised was the utility of the eligibility rate as the headline metric tracked over time.

## 1.3. Supervisory Agent

### Did the SA *design* an appropriate methodology to provide the contracted services?

**The SA designed a strong methodology to deliver the services outlined in the contracts.** The SA developed a broad range of general and bespoke tools to meet the requirements of data collection for each of the ARTF IW projects it monitored. The SA consulted widely with WBG teams on the design of individual scopes of work for each project. GoIRA counterparts in relevant line ministries were also engaged in the design of monitoring protocols for each project the SA monitored.

### Did the SA *deliver* these services in line with contractual obligations and quality standards?

**The Review Team finds that the SA delivered services in line with the objectives and timelines set out in the relevant contracts.** The SA exceeded the annual targets for site visits. The SA adopted inclusive approaches and provided flexible service delivery. The SA employed innovative approaches to data collection and made good use of open-source and bespoke technologies. Approaches were delivered in a manner sensitive to fragile and conflict affected contexts. **Reports were produced on time, on a**



**monthly, quarterly, and annual basis.** A review of these reports contained on the WBG shared drive confirms this to be the case. The monthly reports contained detailed descriptions of the number of sites visited and the number of deviations uncovered. They contained analysis of the nature of deviations and a breakdown of deviations and rectification by ministry. They also contained sections on 'trend analysis', 'challenges' and 'recommendations' - all of which are well written and well presented.

Were the **results** of the SA's services useful to key stakeholder groups?

**Overall, WBG Task Teams felt that the SA's services were sufficient to provide assurances that programs were being implemented and monitored effectively by the government.** WBG Task Teams are rarely permitted to visit project locations to carry out site verification visits due to insecurity. Therefore, the ability to gather data through an external service provider is an invaluable asset.

**GoIRA stakeholders were generally highly complimentary of the SA,** and the Review Team tentatively concludes that the TPM services did contribute to an improvement in the government's capacity to implement and monitor projects. The shift, during the contract implementation period, to an online deviation tracker – Ardea – resulted in improved rectification speeds, and replaced an increasingly cumbersome and outdated tracking instrument.

## 1.4.Observations and Recommendations

Based on an evidentiary review, interviews with ARTF stakeholders, and Aleph Strategies' own experience delivering research programs in Afghanistan and other FCV countries, the Review Team presents observations and recommendations to help guide the existing TPM agent contract and longer-term evolution of TPM engagements.

1. There are opportunities to share lessons learned and best practices with other large-scale TPM engagements in Afghanistan. The WBG should encourage increased dialogue with GoIRA and other funders of TPM to ensure consistency where possible and complementarity where desirable.
2. The ARTF TPM arrangement employs a wide range of methodologies and generates data that has applicability beyond the individual projects monitored. The WBG should ensure that learning from TPM monitoring informs program design and portfolio management.
3. WBG Task Team Leaders operated in silo, managing TPM activities for their respective programs in isolation from one another. Strengthening coordination and collaboration between WBG Task Team Leaders should be a priority for the next phase of contracting.
4. The ARTF TPM engagement has continued to grow in scale each year with no clear exit strategy. The WBG may wish to consider building an exit strategy for TPM highlighting the conditions (security, government capacity, etc.) or targets that could trigger a gradual scaling down of TPM operations.
5. The WBG does not have a TPM agent performance measurement framework or consolidated list of operational and delivery standards by which to measure TPM agent performance. Such a framework would strengthen mutual accountability between the WBG and the TPM agent, and would provide greater clarity to government partners and donors as to the roles, responsibilities (and indeed limitations) of the TPM agent.
6. Neither the SA nor the MA contracts contained outcome-level performance indicators. Outcome-level KPIs should be co-created with the WBG and the current TPM agent to: a) hold the TPM agent to outcome level results; b) set a path towards impact contributions; and c) have evidence to better demonstrate and communicate the value of TPM to donors.



## 2. Introduction

### 2.1. Purpose and Scope

The WBG contracted Aleph Strategies to independently review the Third-Party Monitoring<sup>1</sup> (TPM) of the Afghanistan Reconstruction Trust Fund (ARTF) financed operations for the period 2015 to 2019. The Review took place from August to October 2020.

The objectives of the Review were to:

1. Evaluate the performance of the Monitoring Agent (MA) and Supervisory Agent (SA) in achieving the purposes and objectives of their contracted assignments;
2. Make recommendations for maximizing the impact of TPM in the Afghanistan program, including on governance of TPM services and procedures for following up on TPM findings;
3. Draw lessons from the Afghanistan experience that may be relevant to the implementation of TPM globally and the WBG Strategy for Fragility, Conflict and Violence (FCV) 2020 – 2025.<sup>2</sup>

This Review follows and draws on the TPM findings from previous evaluations and reviews of the ARTF, published in 2005, 2008, 2012, and 2017, as well as the supervisory monitoring-focused assessment conducted at the end of the first phase of the SA contracting period in 2014. The Review covers both ‘downstream’ factors (i.e. fulfillment of contractual requirements and results) and ‘upstream’ factors (i.e. WBG TPM management and governance). This is not a Review of the World Bank Group’s (WBG) use of TPM in general, nor a Review of TPM per se. Instead, it provides an evidence-based account of how the contracted TPM agents performed on service delivery. Moreover, the Review provides insights on how the WBG performed in managing the contracts and crucially, extracting value from findings.

### 2.2. Methodology Note

This Review consisted of:

- a. The creation of an evaluation framework to deliver a clear verdict of MA and SA performance from 2015 to 2019. The framework was built around three overarching questions:
  - Did the TPM **design** an appropriate methodology to provide the contracted services?
  - Did the TPM **deliver** these services in line with contractual obligations and quality standards?
  - Were the **results** of the TPM services useful to key stakeholder groups?

Each overarching question has several sub-questions based on those found in the terms of reference (ToR) for this assignment (see ToR Index below), a reading of available WBG documents on TPM (including the World Bank’s *Demystifying Third-Party Monitoring*), a wider reading of the growing body of TPM literature in FCV contexts (see also the literature review at the end of this report), and the Review Team’s own experience measuring the quality of development programs and service contracts. Please refer to Appendix A Methodology for further details on the research framework.

- b. An **extensive document and data review**. Please refer to Annex C Bibliography which illustrates the selected range of documentation analyzed as part of this Review.

---

<sup>1</sup> The WBG refers to TPM as the process of contracting external agents to assist governments and the World Bank in performing essential monitoring and compliance verification activities within the framework of a given project or portfolio.

<sup>2</sup> The WBG Strategy for FCV 2020 – 2025 recognizes TPM as one means of enabling supervision in highly insecure environments, and highlights key lessons learned from WB experience with TPM operations to date. This Review builds on those findings.



- c. **Key informant interviews (KIIs)** with several key stakeholder groups to triangulate information from several different perspectives: WBG Country Management Unit (CMU); WBG project task teams; donor representatives who engaged with the TPM service providers; GoIRA counterparts; the MA; the SA; and external experts on TPM. A total of 58 individuals were interviewed. Please refer to Appendix A Methodology for further details on the number of individuals interviewed by stakeholder group.

From 2015 to 2019, the MA and SA had separate contracts, had different mandates, and largely worked independently from each other, although the WBG made efforts in the later phases of contract implementation to integrate aspects of their work. As such, the Review comprises two self-contained sub-reviews, one focusing on the MA and the other on the SA. In the recommendations section, the Review brings together key findings from both sub-reviews to provide an overarching assessment of performance.

## 2.3. Report Structure

The report is structured around the three overarching questions above, focusing on **design, deliver** and **results**. For ease of navigation, the Review Team presents an index of questions taken from the ToR for this Review exercise. Readers can quickly find sections of the report that cover each question. Not all questions from the ToR were relevant for both the MA and the SA. The Review Team was careful to ensure that the questions were answered and when they were not, provide reasons for their absence.

Table 1: ToR Questions

Question	Link
Did the contractors deliver on services and reporting in line with the objectives and timelines set out in the relevant contracts?	<a href="#">MA: p. 14</a> <a href="#">SA: p. 28</a>
Did the contractors employ appropriate and adequate methodology in delivering contract services?	<a href="#">MA: p. 10</a> <a href="#">SA: p. 23</a>
Did the contractors employ appropriate, gender-sensitive strategies for reaching and engaging communities, including those in remote or hard-to-reach areas?	<a href="#">SA: p. 29</a>
Did the contractors respond to evolving WBG/GoIRA needs, including for adjusting TPM approaches to respond to earlier findings, in line with contract requirements?	<a href="#">MA: p. 17</a> <a href="#">SA: p. 30</a>
Did the contractors introduce innovations that may be relevant in the current phase of Afghanistan TPM or for TPM programs elsewhere?	<a href="#">SA: p. 32</a>
Did the contractors effectively collaborate with relevant government stakeholders in monitoring processes, including follow-up, while maintaining appropriate independence?	<a href="#">MA: p. 13</a> <a href="#">SA: p. 32</a>
Did the contractors engage relevant community stakeholders in monitoring processes effectively?	<a href="#">SA: p. 29</a>
Were mechanisms established to follow-up on findings and recommendations effective?	<a href="#">MA: p. 16</a> <a href="#">SA: p. 33</a>



<p>While contractual emphasis on capacity development evolved over time, did the contractors' interaction with government, communities, and other stakeholders have impact on these stakeholders' capacity for:</p> <ul style="list-style-type: none"> <li>A) Monitoring the implementation of WBG-financed projects?</li> <li>B) Adhering to fiduciary standards, including without limitation expenditure eligibility?</li> <li>C) Adhering to environmental and social safeguards standards, including without limitation gender requirements?</li> </ul>	<p><a href="#">MA: p. 16</a></p> <p><a href="#">SA: p. 29</a></p> <p><a href="#">SA: p. 32</a></p> <p><a href="#">SA: p. 36</a></p>
<p>Were TPM data and findings of adequate quality, reliability, and relevance to the implementation support, fiduciary, and safeguards objectives of the contracts?</p>	<p><a href="#">MA: p. 14 - 17</a></p> <p><a href="#">SA: p. 28 - 35</a></p>
<p>Were TPM findings and recommendations presented in a format that was usable to relevant WBG and GoIRA teams?</p>	<p><a href="#">MA: p. 14 - 17</a></p> <p><a href="#">SA: p. 33</a></p>
<p>Were TPM findings and recommendations accessible/understandable to ARTF stakeholders?</p>	<p><a href="#">MA: p. 18 - 21</a></p> <p><a href="#">SA: p. 33</a></p>
<p>Were TPM findings communicated in a manner that provided assurance to ARTF contributors regarding 'value for money' in ARTF programs?</p>	<p><a href="#">SA: P. 34</a></p> <p><a href="#">MA: p. 18 - 21</a></p>
<p>Did TPM services, findings, and recommendations support enhanced 'value for money' in the ARTF-financed portfolio in line with contract objectives?</p>	<p><a href="#">MA: p. 18 - 21</a></p> <p><a href="#">SA: p. 34</a></p>
<p>Were TPM findings gender disaggregated, where applicable, to help ARTF stakeholders understand differential impact on men and women?</p>	<p><a href="#">SA: p. 29</a></p>
<p>How were the contractors' approaches similar to/distinct from other experiences of large-scale TPM described in research and literature on monitoring and implementation support in other fragile contexts?</p>	<p><a href="#">p. 38</a></p>
<p>Are there lessons learned on the design and implementation of TPM from global experience that are relevant to evaluating the contractors' performance?</p>	<p><a href="#">p. 38</a></p>
<p>What lessons from the implementation of TPM during the 2015-2019 contract periods should be considered for:</p> <ul style="list-style-type: none"> <li>A) Implementation and management of TPM under the consolidated contract modality introduced in January 2020 for the current phase of TPM contracting arrangements?</li> <li>B) The future of monitoring and implementation support as the WBG plans beyond the current contract?</li> </ul>	<p><a href="#">p. 41</a></p>



## 2.4. Limitations to the Methodology

There are a number of inherent limitations with a Review of this nature that balances an ambitious scope with a fixed budget and timeframe.

1. This Review covers TPM activities conducted between 2015 and 2019. It does not include activities conducted before this time;
2. Some targeted interviewees did not respond to a request for interview or were unavailable during the timeframe of this Review: August – October 2020. In a context such as Afghanistan, turnover can be high and therefore many targeted interviewees were no longer working in the positions they held at the time of project implementation. Interviews were conducted on a best-effort basis within the timeframe of the Review, which required Aleph to complete research in a three-month period;
3. The Review Team was only able to interview a small proportion of stakeholders involved in the ARTF both currently and historically. Views expressed in this report should not be treated as representative. Where possible the Review Team has balanced feedback from stakeholders with documentary evidence to provide nuanced analysis. The Review Team highlights where this has not been possible;
4. The Review does not provide an individual analysis of TPM performance under each of the 15 ARTF-funded projects as this would require a significantly greater investment of time and resources. Instead, the Review provides a summary performance analysis of the SA's activities in general. Specific examples were drawn from individual projects for illustrative purposes.

## 3. Background and Context

---

### 3.1. What is the ARTF?

The Afghanistan Reconstruction Trust Fund (ARTF) is a WBG administered multi-donor trust fund established in 2002 to provide coordinated financing in support of the Government of the Islamic Republic of Afghanistan's (GoIRA) operating costs and priority development programs. The ARTF creates shared responsibility and accountability between GoIRA and the international development community, and ensures that the Afghan Government retains strategic agency on project implementation. The ARTF provides an important platform for aligning donor priorities and channeling resources into prioritized areas. It helps mitigate the challenges associated with off-budget, un-coordinated funding, which can create parallel (and at times competing) priorities set by independent donors. As of February 2020, 34 donors have contributed over USD 12 billion making the ARTF the largest contributor to the Afghan budget.<sup>3</sup>

During the period under review, the ARTF comprised two primary recipient-executed funding windows:

- i) The Recurrent Cost Window (RCW): this provided predictable and transparent operating (non-security) budget support for costs such as civil servant salaries, operations, and maintenance. Since 2002 approximately USD 5 billion has been disbursed through three channels: a) baseline support; b) an incentive program; and c) ad-hoc payments.
- ii) The Investment Window (IW): this provided funding for programs to develop five sectors: agriculture; rural development; infrastructure; human development; and governance/public sector capacity. Since 2002 more than USD 5.7 billion has been made available for closed and active on-going development projects.

---

<sup>3</sup> The World Bank Group in Afghanistan – Country Update, p. 4.



The ARTF administration agreement authorized the WBG to finance third-party monitoring costs directly from ARTF contributions, but under the 2018-2020 Partnership Framework and Financing Program, the Bank-executed Anti-Corruption and Results Monitoring Action program window was established in the fund, and financing of TPM expenses was moved under this window for improved clarity.

### 3.2. Why does TPM exist for the ARTF?

TPM exists for two primary reasons. First, contributing ARTF donors expect independent monitoring to ensure that ARTF funds are disbursed for their intended purposes. In a context of low government capacity, high levels of corruption, and the potential for fund mismanagement, GoIRA line ministry internal monitoring is perceived to be insufficient. Second, volatile security conditions in Afghanistan prohibit the WBG's task teams from conducting field monitoring activities, necessitating the engagement of external contractors to conduct this work.

### 3.3. Who were the contractors? What were they responsible for?

During the period under review, two separate contractors provided distinct third-party monitoring services. [BDO LLP London](#) served as the independent MA from 2016 to 2019<sup>4</sup>, with a primary responsibility for monitoring the eligibility of GoIRA expenditure for reimbursement through the RCW. Contract extensions and amendments later during this period added responsibility for conducting internal control assessments and transaction reviews of ARTF-financed development projects, with the latter to be coordinated with the SA. The total value of the MA's contracts from 2016 through 2019 was USD 12.78 million.<sup>5</sup> [Management Systems International \(MSI\), Inc.](#) served as the independent SA from late 2015 to 2019, responsible for monitoring the implementation of, and reporting on IW-financed projects. In contract extension amendments, the SA was also assigned responsibility for coordinating with the MA to support physical verification of civil servants (a component of MA expenditure eligibility monitoring) in hard-to-reach areas. The total value of the SA's contracts from 2015 to 2019 was USD 25.28 million.<sup>6</sup>

---

<sup>4</sup> BDO has served as the MA since early 2013 but only the contract period from mid-2016 until the end of 2019 is under review.

<sup>5</sup> BDO contract 8005616 and extensions.

<sup>6</sup> ARTF Financial Status Report, July 2020.





## 4. Monitoring Agent Review

### 4.1. Summary

*Focus Box 1: Summary of MA Headline Findings*

Review Question	Headline Findings
Did the MA <b>design</b> an appropriate methodology to provide the contracted services?	<p>The MA designed a robust and appropriate methodology for its RCW monitoring activities.</p> <p>The MA piloted an innovative and comprehensive methodology for IW-financed project internal control assessments in 2018 and 2019. As the approach was still in its nascent stage of development, there was some confusion among WBG Task Teams interviewed in regards to the appropriateness of the methodology for the Afghan (high-risk) context. In the next phase of contracting, the third-party monitors should further clarify how the assessments are made unique for their respective projects.</p>
Did the MA <b>deliver</b> these services in line with contractual obligations and quality standards?	<p>The MA delivered services in line with the contractual obligations and to a high-quality standard, and in many cases exceeded expectations.</p>
Were the <b>results</b> of the MA's services useful to key stakeholder groups?	<p>The reports and findings were found to be invaluable to the WBG and GoIRA audiences. While donors appreciated access to the full MA reports from 2018 and therefore perceived more utility of MA services, there is demand for even further information (in particular on the GoIRA's public financial management capabilities). The WBG, donors, and the current TPM agent should clarify together what is and is not possible given TPM agent's role and responsibilities.</p>

### 4.2. Introduction: What was the MA contracted to deliver?

The WBG issued three contracts to the MA from June 2016 to December 2019: (1) June 2016 – August 2018; (2) September 2018 – April 2019; and (3) May 2019 – December 2019. Each contract contained a detailed scope of services (SoS), reporting requirements (RR) and key performance indicators (KPIs).

The SoS for the MA can be summarized as follows:<sup>7</sup>

- Provide independent monitoring and scrutiny of the government civilian budget;
- Verify payroll and operations and maintenance (O&M) expenditure incurred by line ministries;
- Maintain civil service headcount database with data entry and consistency checks;
- Monitor RCW withdrawal applications and disbursements between GoIRA and the WBG;
- Provide information to the Ministry of Finance (MoF) regarding high-value ineligible items;
- Keep the MoF and the WBG informed on emerging issues and progress;

<sup>7</sup> ARTF Partnership Framework and Financing Program (2018 – 2020), p. 49.





- g. Work with the MoF to improve compliance with the RCW eligibility criteria and strengthen its capacity;<sup>8</sup>
- h. Monitor IW project internal control and transactions in coordination with the SA.<sup>9</sup>

Major reporting requirements to the WBG included:

- a. **Monthly reports:** a topline summary of the previous month's monitoring findings and support activities (four pages maximum);
- b. **Quarterly reports:** a detailed analysis of trends and recurring challenges found over the quarter and actionable recommendations for corrective actions;
- c. **Annual reports:** a draft annual statement of uses and application of the ARTF RCW funds for submission to the MoF and auditors; and a consolidated report on progress made, major risks identified and resolved, outstanding action items, and reasons for nonresolution.
- d. From September 2018 to December 2019, **monthly and quarterly reports on IW project internal controls and transactions.**

The MA was also tasked with producing **special reports** for the WBG on the basis of emerging issues. For example, in 2016 the MA authored a detailed report on potential unauthorized salary deductions in Baghlan province.

The WBG included two KPIs in its contracts to assess the performance of the MA in relation to possible contract renewals:

- a. **KPI1: Quality and timeliness of periodic reports/deliverables** - Withdrawal applications submitted to the World Bank controllers unit within 10 working days of receipt of the set of applications from the government; monthly and quarterly operations reports submitted to the Bank within four weeks of the month/quarter end; and an annual report submitted to the Bank within six months of the end of the period;<sup>10</sup>
- b. **KPI2: Transfer of knowledge and capacity building** - Submit to the WBG and GoIRA a CBKT plan within the first two months of the contract award.<sup>11</sup>

Finally, the contracts stipulated GoIRA contributions, WBG undertakings, responsibility for security arrangements, and, crucially, a detailed breakdown of MA Team composition expectations (including balancing international and local staff as well as resident and non-resident in Kabul).

### 4.3.Design

Did the MA design an appropriate and adequate methodology in delivering the contracted services?

RCW

**The Review Team finds that the approach and sampling plan were robust and transparent for financial monitoring control purposes.** The WBG MA Contract Manager during the Inception Phase remarked that the MA was highly professional, collaborative, and flexible, working closely with the WBG to refine approaches. A robust Inception Report delivered by the MA at the beginning of the engagement illustrates a clear understanding of the objectives and evidences close consultation with the WBG to develop appropriate approaches. The Report clearly details: the scope of the assignment in line with

---

<sup>8</sup> Capacity building and knowledge transfer expectations found in the first contract from June 2016 – August 2018 were not detailed in the subsequent contracts from September 2018 – December 2019.

<sup>9</sup> The MA's investment window activities took place in 2018 and 2019 only.

<sup>10</sup> This KPI is found in each of the three MA contracts.

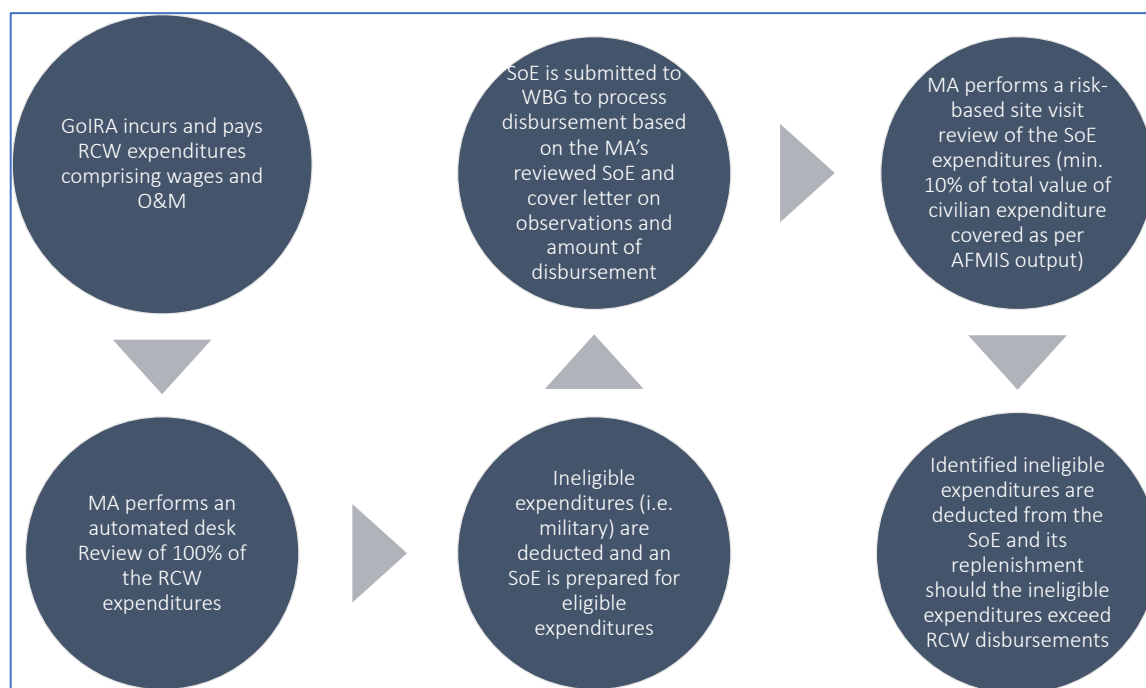
<sup>11</sup> This KPI is found in each of the three MA contracts. Yet, unclear why it was included in the second and third contracts given the CBKT plan was delivered in July 2016.



expectations set in the contract; the leadership team to be deployed and staffing plan; the methodology and sampling plan; workplan, including deliverables; CBKT approach; and staff duty of care protocols.

In essence, the MA was responsible for measuring the **amount of eligible GoIRA civilian recurrent expenditures that could be reimbursed by the RCW**. The Figure below summarizes the MA monitoring process employed:

Figure 1: MA Monitoring Process<sup>12</sup>, SoE (Statement of Expenditure)



The Operational Manual on Eligibility<sup>13</sup>, against which GoIRA expenditures were scrutinized, was developed by the previous MA. The eligibility criteria are summarized and simplified in the Focus Box below.

Focus Box 2: Summary Eligibility Criteria<sup>14</sup>

To be eligible for reimbursement from the RCW, civilian recurrent expenditures must be made in a manner consistent with Afghan legislation and related regulations. The applicable Afghan laws and regulations include (but are not limited to):

- Annual Budget Decree;
- Public Finance and Expenditure Management Law (PFEML);
- Civil servants law;
- Law on procurement and related regulations and circulars; and
- Circulars issued by the MoF.

<sup>12</sup> 2012 ARTF Fiduciary Framework, Annex 4, p.17; consultations with the MA (BDO).

<sup>13</sup> The Operational Manual on Eligibility was developed by the original MA (PwC Netherlands) and endorsed by the WB in 2010.

<sup>14</sup> ARTF Partnership Framework and Financing Program (2018 – 2020), p. 48 - 49.



A set of Fiduciary Standards were also agreed to between the Management Committee of the ARTF and GoIRA to provide a basis for GoIRA to demonstrate commitment to improving fiduciary performance with respect to government expenditures.

Expenditures made that are not consistent with the procedures agreed upon are considered 'ineligible' and are not reimbursed. Ineligibilities can be broken down into two main categories:

- Avoidable ineligibilities: the expenditure could have been eligible, but the process followed was not consistent with the agreement between the World Bank and GoIRA. For example, procurement procedures were not followed or documentation was not complete;
- Unavoidable ineligibilities: the expenditure is ineligible by nature, usually because of an ARTF limitation (e.g. the ARTF cannot finance military expenditures).

The expenditures were categorized in four broad groupings: 1) payroll-based salary expenditure (PBSE); 2) non-payroll-based salary expenditure (NPBSE); 3) O&M expenditure - excluding pensions; and 4) O&M pension payments. Each category of expenditure was then monitored based on:

- **Existence:** Government employees do exist and are paid remuneration and are being properly compensated for work performed by agreement back to attendance records. General administrative expenses are incurred for official purposes;
- **Proper authorization:** Transactions are justified, supported, authorized, and approved in accordance with established authorization limits (Procurement Rules and Regulations);
- **Completeness:** Transactions are completely recorded within three months of expenditure being incurred for central ministries in Kabul and within four months for provincial departments;
- **Presentation and disclosure:** Transactions are accurately coded within the relevant general ledger and budget activity code.

**The MA provided a clear justification for expenditure sample selection.** Transactions were scrutinized from the Afghanistan Financial Management Information System (AFMIS) and three sampling methods were applied: 1) targeted sampling, when O&M transactions met specific criteria (e.g. > USD 50,000); 2) statistical sampling drawn from the population after deducting the targeted samples; and 3) threshold sampling in order to ensure that a minimum of 10% of the value of auditable transactions was monitored. The sampling approach also included monitoring tests to assess the design and operational effectiveness of internal control systems and physical verification to ensure that employees existed.

The sampling above was modified from July 2018 to Dec 2019 to reduce the number of samples being tested in the RCW in order to accommodate IW-funded project testing, summarized as follows: statistical random sampling of the whole population; threshold sampling to ensure that a minimum of 10% of expenditure is tested overall; and conducting a countrywide sample of civilian GoIRA personnel for physical verification.

Finally, **the sample was of statistical significance.** The recommended confidence level was 95%. The confidence interval was calculated from the number of audits undertaken. The MA stated in its Inception Report:

*Interpreting the results from an audit of the sample is relatively straight forward. For example, if the confidence interval is 3% and that 23% of transactions were deemed to be ineligible from the audit, the interpretation would be that between 20% and 26% of all transactions are ineligible with a 95% confidence level. The confidence level indicates that if the selection of the randomized sample was undertaken 100 times, 95 of those times the average ineligibility rate would be between 20% and 26%. The key is not the point estimate, but rather the boundary that is identified.*

MA Inception Report, 2016



The MA and WBG former Contract Manager interviewed did not think a radical overhaul of the approach and sampling plan from previous contracting periods would be effective, as doing so would make it near impossible to accurately compare eligibility over time. Instead, the MA focused on making the process more efficient, such as increasing the threshold of targeted sampling of expenditures from USD 10,000 to USD 50,000. The MA observes that the approach in Afghanistan is somewhat unique in comparison to other countries given the very high number of eligible expenditure verifications. Yet, the MA remarks such an approach is commensurate with the context: high-risk, high corruption concerns, and low GoIRA staff capacity to perform such functions. As such, WBG and donor appetite for strict controls was and remains high.

Finally, the MA also consulted extensively with GoIRA national counterparts during the design of its bespoke CBKT plan (KPI2), in particular with the MoF, the Ministry of Education (MoE), and National Procurement Authority personnel. The MA sought feedback on the capacity gaps identified through an assessment of historical eligibility data as well as inputs into strategies to overcome gaps. The MA rightly observed that there could have been a conflict of interest in scrutinizing government expenditure while simultaneously working with government counterparts to improve government systems and staff performance. The MA also ensured its independence from capacity building activities with GoIRA staff with the establishment of a MoF led Ineligibilities Working Group. The MA supported the MoF in the design of capacity building aspects through the Working Group, while the MoF was responsible for conducting the capacity building activities with line ministries themselves.

*I/W*

**By-and-large, the Review Team finds that the project-level internal assessment controls were robust and filled a much-needed gap. Furthermore, the MA's flexibility to quickly take this task onboard in 2018 and pilot innovative approaches were welcomed by WBG Task Teams interviewed.** In total, the MA piloted the internal control assessments of 13 projects, eight in 2018, and five in 2019.

**The methodology employed by the MA was comprehensive.** It included: i) interviews with key personnel; ii) review of relevant documentation; iii) analytical review of data; and iv) sample testing<sup>15</sup>.

Each assessment included a review of: a) procurement and contract management; b) payroll and HR management; c) financial management; d) governance and oversight; e) complaints, fraud, and corruption; and f) monitoring and evaluation.

Exceptions<sup>16</sup> identified were defined as:

- **Deficiency:** internal controls, governance, and risk management practices are adequately designed, generally well implemented, but one or a limited number of issues were identified that warrant management attention;
- **Significant deficiency:** internal controls, governance, and risk management practices contain deficiencies, or a combination of deficiencies, that present a moderate risk to the achievement of the objectives or occurrence of material misstatement in financial reports; and
- **Material weakness:** internal controls, governance, and risk management practices have some weaknesses in design or operating effectiveness such that there is reasonable assurance that the project objectives may not be met or that the financial reports may contain material misstatement.

**However, some feedback on the methodology and sampling plan was provided by WBG Task Teams interviewed that could be useful ahead of full implementation under the current TPM agent contract.**

---

<sup>15</sup> The MA did not conduct audits or substantive test of transactions. Instead they selected a limited number of samples to test the operating effectiveness of key controls.

<sup>16</sup> MA ARTF Completion Report, p. 55.



First, there was some confusion among WBG Task Teams interviewed whether the summarized methodology above was fit-for-purpose for each project given unique project requirements and activities. The Review Team notes that financial management and disbursement approaches for the entire ARFT portfolio are the same, as noted in the financial management manual and disbursement guidelines agreed with the WBG. In addition, the methodology was adjusted based on project activities. For example, the Citizens' Charter Afghanistan Project (CCAP) internal control assessment largely focused on community development council (CDC) grants management (which was not part of other project assessments); whereas scholarships management was reviewed for the Higher Education Development Project (HEDP). Finally, the MA discusses and agrees to the specific scope with the WBG Task Teams during pre-planning meetings. Given the internal control assessments process was at a nascent stage of development in 2018 and 2019, it is likely that WBG Task Teams just require further clarity and understanding on how the assessments are made unique for their respective projects.

Second, some WBG Task Teams interviewed would have welcomed MA project site visits and face-to-face meetings with local officials. The MA relied mostly on a document review and interviews with national government representatives in Kabul. Much of the funding available for some IW projects was (and continues to be) decentralized at the provincial and even at the CDC level. According to some WBG Task Team members, speaking directly to those who disburse funds at a local level could have generated stronger reports. Yet the Review Team notes that in-person meetings with local officials were not part of the MA's scope. The SA was responsible for site visits and the MA was only required to review a sample of transactions for some sites visited by the SA to compare financial and physical project progress. Given that the role of the MA and SA have been merged into a single TPM contract as of early 2020, this issue has already been addressed as current TPM agent and/or its subcontractors will conduct the site visits themselves.

Third, the rating system noted above used to classify exceptions and more crucially, how it was applied was perceived by some WBG Task Teams to be excessively rigid. One WBG task team leader interviewed remarked: 'What might be a 'significant deficiency' in Europe cannot necessarily be deemed as such in Afghanistan'. Yet the MA remarks that the rating system reflects the fact that Afghanistan is a high-risk environment and it is not based on European benchmarks. Again, WBG Task Teams likely just require further clarity on how the rating system has been tailored to the context in Afghanistan.

Finally, and linked to point three above, the overall ratings were given to projects based on a limited sample. It was suggested by some WBG Task Teams interviewed that one cannot make, for example, a 'deficiency' judgement on this basis. The MA noted itself in its internal control assessments:

*It should be noted that this review is not an internal or external audit. It has not been conducted in line with International Standards on Auditing and had we done so, we may have found further issues to report. Consequently, we are not providing assurance, as this would be a matter for the Internal Audit department of the relevant ministry. However, this report may well assist Internal Audit in directing their testing program in the future. Our overall assessment, based on the limited scope of our ToR, of internal controls over project funds was that of a 'Deficiency'.*

MA Internal Control Reviews

#### 4.4. Deliver

The assessment of the MA's performance on service delivery is built on two primary questions: i) Did the MA deliver services in line with the objectives set out in the relevant contracts? And ii) Did the MA deliver on reporting requirements in line with the objectives and timelines set out in the relevant contracts? **The Review Team finds that the MA delivered well on its contractual obligations.** They produced monthly, quarterly, and annual reports as well as withdrawal applications and the CBKT plan. All of these products were delivered on time and in line with the approach detailed in their Inception



Report. Former and current WBG CMU staff observed that the MA was professional, working to the letter of the brief quite well. GoIRA counterparts found the MA to be more of a ‘supportive partner’ rather than an ‘auditor’. Furthermore, the MA proved to be flexible, a welcome attribute given the volatile nature of operating in Afghanistan and the changing priorities and demands of ARTF stakeholders.

The Table below provides a consolidated review of the MA’s performance against key objectives based on a thorough analysis of documentation and interviews with relevant key stakeholders.

*Table 2: Analysis of MA Contractual Obligations*

Objective	Performance
Provide independent monitoring and scrutiny of the government civilian expenditure	<p><b>Yes, the MA met this objective.</b> The MA delivered in line with the methodology and sampling plan detailed in the previous section. The MA provided clear details on the: headline eligibility amount and rate; breakdown by payroll and O&amp;M; and reasons for ineligible transactions by category (e.g. military expenditure, missing documentation, procurement noncompliance, etc.) on a monthly, quarterly, and annual basis. Furthermore, the MA conducted additional ‘special reports’ based on emerging issues from its analysis.</p> <p>Over the contract period (2016 – 2019), GoIRA submitted a total expenditure of USD 4.94 billion for reimbursement, of which USD 3.61 billion was assessed as potentially eligible for reimbursement, representing an average eligibility rate of 73% (Note that actual reimbursements under the RCW fell significantly below this total eligibility ‘ceiling’). By comparison, from 2002 to 2019, GoIRA submitted a total expenditure of USD 15.37 billion, of which USD 11.01 billion was approved, representing an average eligibility rate of nearly 72%.</p> <p>This improvement in eligibility over the contract period in comparison to historical averages is slight, but worth mentioning as the MA made significant efforts to work with GoIRA to submit missing documentation and then re-test rectifiable ineligibilities. This is particularly evident when analyzing the eligibility rate from 2018 to 2019, which improved from 72% to nearly 75%.</p>
Conduct site visits to government offices in Kabul and in provinces to substantiate validity and eligibility of expenditures submitted	<p><b>Yes, the MA largely met this objective.</b> The MA’s centrally and provincially based staff conducted in-person verification of the sample transactions in all but two provinces. By the end of the contract, the MA physically verified 30,056 employees. Yet, insecurity and a lack of personnel to meet verification demands hindered progress.</p> <p>The MA’s workforce comprised national and international staff based in Kabul, Nangarhar, Kunduz, Paktyia, Kandahar, Herat, and Balkh. The teams included auditors, statisticians, national monitoring, and headcount database teams and administration officers who were responsible for monitoring activities in the province in which they were based and adjacent provinces which reduced travel costs and had the aim of increasing the coverage of expenditure verification. National staff were able to visit 32 out of 34 provinces during the contract period. Due to the volatile security environment, the MA was unable</p>





	<p>to carry out any in-person testing in Nuristan or Paktika as well as some districts within other provinces at certain times.<sup>17</sup></p> <p>A total sample of 44,562 employees were selected for physical verification. Of this number, the physical location of 9,149 could not be reached due to security concerns, leaving an available sample for testing of 35,413. Of this number, the MA was able to verify that 85% (30,056) of employees were physically present without exception. The remaining 15% (5,357) could not be physically verified. In most instances, sufficient documentation was provided to justify reasons for non-verification. These included, 'on leave', 'school exams', 'employee abroad', and 'employee transferred to another location', among others.</p> <p>Despite provincial teams, physical verification of employees and expenditures in some insecure locations proved challenging for the MA given the sheer number of verifications to be completed in a timely fashion. As such, the SA supported this process near the end of the contracting period which improved verification in some hard-to-reach locations (the SA was able to conduct 92 site visits on behalf of the MA). This issue has already been addressed in the current TPM engagement as the enlarged TPM agent team is able to adequately carry out this function.</p> <p>Finally, as per the MA's Completion Report, follow-up with those that could not be physically verified was conducted 'whenever possible'. It was not, therefore, a contract requirement to physically verify the outstanding employees of the sample within the timeframe of the contract.</p>
Maintain civil service headcount database <sup>18</sup> with data entry and consistency checks	<p><b>Yes, the MA met this objective. Furthermore, the MA transferred responsibility of the Database to the MoF, a significant capacity building result.</b></p> <p>During the contracting period the MA supported database management functions, trained relevant MoF staff to enter, update, and maintain the Database, and conducted quality assurance of MoF work on the Database (including consistency checks).</p> <p>The MA then transferred ownership of the database and assets to the MoF in March 2019. This included a final handover manual detailing the suggested workflow, database design and maintenance, organizational chart, and MIS backup files, among others.</p>
Monitor withdrawal	<p><b>Yes, the MA met this objective. The MA kept a detailed record of RCW statements of expenditures (SoEs) and reimbursements by the WBG. This</b></p>

<sup>17</sup> Beyond Paktika and Nuristan, the MA was unable to physically verify employees in some districts within other provinces at certain times. In 2017 some employees in some districts in Farah, Helmund, Faryab, Jawzjan, Ghor provinces could not be verified. In 2018 some employees in some districts in Kapisa, Wardak, Logar, Baghlan, Ghazni, Paktya, Khost, Kunar, Badakhshan, Takhar, Kunduz, Sar-e-Pul, and Jawzjan provinces could not be verified. In 2019 some employees in some districts in Kabul, Wardak, Nangarhar, Laghman, Baghlan, Paktya, Khost, Kunar, Badakhshan, Kunduz, Balkh, Sar-e-Pul, Kandahar, and Jawzjan provinces could not be verified.

<sup>18</sup> The Headcount Database provides information on staffing levels and costs to various users and is used to generate monthly reports regarding compliance with Fiduciary Standards. As such it has proven to be a useful tool for controlling and monitoring expenditure on payroll costs for uniformed and non-uniformed employees across the government's line ministries and entities.



applications and disbursements between GoIRA and the WBG	information was consolidated at the end of contract in a user-friendly manner by: fiscal year; category (O&M or payroll); and percentages approved by the MA (e.g. 84% of payroll expenditures in 2018 approved).
Provide information to the MoF regarding high-value ineligible items	<p><b>The MA met this objective. The MA provided a Microsoft Excel spreadsheet of non-compliant transactions, including rectifiable issues, to the MoF on a monthly basis.</b> The MA then worked with MoF personnel to follow-up on rectifiable issues; in particular when documents were missing for high-value contracts. The MA had strong results in this area, as rectifiable ineligible expenditure reclassified as eligible expenditures improved from 64% in 2017 to 82% in 2018. Through three quarters of 2019, the figure stands at 59% and will likely be higher once final analysis is completed.</p> <p>MoF stakeholders interviewed remarked that they very much appreciated the information and support provided by the MA. Moving forward, they suggest it would also be useful to receive a <u>monthly</u> narrative report to accompany the detailed data that highlights key emerging issues by ministry, theme, and in order of importance. The Review Team notes that such briefing notes were prepared by the MA on a <u>quarterly</u> basis for the Treasury Department at the MoF for those entities with the highest ineligibilities to facilitate communication with the relevant line ministries. The briefing notes outline reasons for ineligibilities and included a listing of the top ineligible transactions.</p>
Keep the MoF and the WBG informed on emerging issues and progress	<p>Headline emerging risks and issues impacting the MA's activities were reported to the WBG on a monthly basis in progress updates and in more detail in quarterly summary reports. Moreover, the MA made suggestions for further analysis on the back of issues identified. For example, the MA produced a special report on potential unauthorized salary deductions in Baghlan province, and then further expanded this with additional samples country-wide. Headline reasons for ineligibility were packaged well in the executive summary of summary quarterly reports for ease of reading and other issues could be found in the details of the main body of the reports (e.g. physical verification issues, ongoing security challenges, etc.). The Review Team notes that moving forward the quarterly summary reports could benefit from having a stand-alone emerging issues and progress section similar to the monthly progress updates, albeit with more details.</p>
Work with the MoF to improve compliance with the eligibility criteria and strengthen its capacity	<p><b>The MA largely met this objective. The MA delivered a capacity building and knowledge transfer plan and conducted information sharing/capacity building activities in coordination with GoIRA counterparts. The eligibility rate continues to fluctuate, yet the average rate over the contract period (73%) is a slight improvement on the historical average (72%).</b></p> <p>The MA delivered a good CBKT plan. The plan outlined institutional, human resource, technological, and systems and process gaps, as well as strategies to overcome gaps and a suggested workplan of implementation towards improving the eligibility rate.</p> <p>*Please refer to Appendix B for details on the recommendations made in 2016 and the status of implementation.</p>





	<p>The MA worked with the MoF to establish the Ineligibility Working Group to share information, organize capacity building efforts (workshops), and work in a coordinated manner to improve the eligibility rate.</p> <p>The MA deserves credit for working consistently to support improvements in the average eligibility rate over the contract period in relation to the historical average, which can be correlated with capacity building efforts and its concerted effort to reclassify ineligible expenses. However, the historical rate continues to fluctuate, as it went from 70% (2016) to 76% (2017), then back down to 72% (2018) and up to 75% (2019). The MA remarks that the eligibility rate cannot improve solely on capacity building efforts alone. It may be worth considering whether the eligibility rate, which offers one summary measure for assessing the quality of some GoIRA control processes, could be incorporated as an indicator in relevant PFM capacity building initiatives or to inform reform targets. Second, a large proportion of ineligibility is due to noncompliance with WBG fiduciary standards which are not aligned to Afghan Law and therefore are deemed ‘unavoidable’. In 2019, for example, the eligibility rate would be 84% instead of 75% if unavoidable ineligibleities were removed.</p> <p>Yet, the Review Team finds that while the MA can clearly state that workshops and information sharing sessions took place (outputs), they cannot provide concrete evidence on whether or not MoF and line ministry capacity has ‘strengthened’ (outcomes). Yes, a correlation between capacity building and better eligibility rates can be made, yet the value of capacity building would be better understood if the MA found ways to measure increases in capacity (e.g. pre- and post- testing) and/or systematically collected qualitative insights from beneficiaries.</p>
Monitor IW project internal controls and transactions in coordination with the SA	<p><b>Yes, the MA met this objective. The MA conducted internal control assessments of 13 IW projects<sup>19</sup> in coordination with the SA in 2018 and 2019, and they submitted monthly and quarterly reports in line with contractual obligations.</b></p> <p>Furthermore, the MA’s internal control assessments of IW-financed projects are a good example of the MA’s ability to respond to the evolving needs of the ARTF. Broadly speaking, WBG stakeholders remark that when asked to adjust their scope or even do certain things outside of their scope, the MA proved to be flexible, willing to listen to constructive feedback and able to adapt and find solutions to meet demand.</p>

## 4.5. Results

The assessment in this section is built on one fundamental question of utility: **Did ARTF stakeholders derive value from the MA’s findings?**

In line with the ARTF administration agreement, the MA contract establishes the WBG and GoIRA as the MA’s direct audiences, but ARTF donors are a crucial additional audience given they fund the ARTF.

<sup>19</sup> AAIP; ARAP; ASDP II; CCAP; EQRA; FSP; HEDP; HEP; IRDP; NHLP; NHRP; OFWMP; and Sehatmandi.



Each derives different value from MA services and therefore how they perceived the usefulness of MA findings varies.

## WBG

**In regards to RCW, the WBG stakeholders interviewed broadly found the results useful.** The WBG principally derives recurrent expenditure monitoring value from answering this question: does the eligible expenditure ceiling exceed the total reimbursement commitment made by the ARTF RCW? This was the case each year throughout the contracting period under review as detailed in the Table below. For example, in 2018, GoIRA submitted USD 1.28 billion worth of expenditures for review. Based on an eligibility rate of 72% (28% ineligible) on the sample scrutinized, the MA deemed USD 361 million to be ineligible, for a net eligible expenditure of approximately USD 922 million. Given the WBG was to only disburse USD 159 million in 2018, this fell well below the total net eligible expenditure ceiling. With this in mind, the WBG and by extension the ARTF donors could rest-assured that their contributions in 2018 were disbursed against eligible expenditures.

*Table 3: Expenditure Claimed and Disbursed (USD millions). Source MA ARTF Completion Report*

	2016	2017	2018 <sup>20</sup>	2019
Expenditure claimed by the MoF	1,140.87	1,218.90	1,283.69	1,292.33
Less: Ineligible expenditure from MA review	-342.85	-292.88	-361.31	-324.17
Net eligible expenditure	798.02	926.02	922.38	968.16
Less: Total ARTF RCW disbursement by WBG	-553.69	-538.76	-159.38	-101.06
Net eligible expenditure over disbursement ceiling	244.33	387.26	763.00	867.1

Furthermore, the WBG points to the cost of the MA in relation to funds disbursed as a key metric of value. The MA cost USD 12.78 million over the contract period. Based on MA scrutiny of government civilian expenditure, the ARTF reimbursed USD 1.35 million. As such, the MA cost represents 0.94% of total reimbursement; or for each dollar spent on the MA, USD 106 was reimbursed. This cost is at the bottom range of other international development monitoring engagements analyzed as part of this Review, in which TPM cost 0.8% - 3.9% of funds disbursed (see Table 6 TPM Cost Benchmarking). As such, the WBG views this cost as money well spent. WBG staff interviews suggested it may be possible to better leverage the full scope of data provided in MA reporting to inform WBG-GoIRA policy discussions and/or to shape WBG programming. WBG Staff noted that achieving this outcome would likely require better integration of TPM reporting products and a dedicated dialogue between the MA and WBG-GoIRA project teams, which was expected to be one of the benefits of integrating management of the two contracts in the next phase of the TPM contract.

**In regards to the IW-financed internal control assessments which began much later in the contract, some WBG Task Teams felt that the process required further refinement to better reflect ‘ground realities’.** This was to be somewhat expected as the activity was in its nascent stage of development. However, **much value was still derived by both the WBG fiduciary team and Task Teams.** Each report was accompanied by practical and actionable recommendations for WBG Task Teams and GoIRA line ministry project teams to improve the status quo. For example, the MA found a ‘material weakness’ in regards to CCAP’s website security and outlined steps for improvement. This recommendation was accepted and the change implemented. It was one of 23 recommendations made, of which 20 were

<sup>20</sup> The 2018 and 2019 figures exclude IP DPG disbursements.



either accepted or partly accepted by MRRD and WBG Task Teams. Other examples include the ‘significant deficiency’ found in Naghlu Hydropower Rehabilitation Project’s (NHRP) M&E compliance or the ‘deficiency’ found in Afghanistan Rural Access Project’s (ARAP) spot checks on assets and inventories for financial management purposes. Furthermore, the MA consistently followed up on implementation of the recommendations with WBG Task Teams and GoIRA line ministries to ensure enhanced internal control effectiveness.

## GoIRA

**GoIRA considers the MA to be valuable because it is a condition for cost reimbursement, which is needed to meet civilian recurrent expenditure.<sup>21</sup> Interviews conducted by the Review Team suggest that GoIRA was able to leverage further value from the MA beyond this procedural necessity.** The GoIRA stakeholders interviewed remarked that the MA was viewed as a ‘partner’ rather than an ‘auditor’. Much of the value stemmed from the MA’s work through the Ineligibilities Working Group in which information sharing sessions were co-developed with the MoF Treasury Department for MoF and line ministry staff on: a) adherence to the Fiduciary Standards and understanding the risks relating to payroll and procurement systems; b) understanding of financial management, internal controls, and accountability; and c) understanding the rules of eligibility of expenditure. GoIRA found the forum and activities to be invaluable and would like the Ineligibilities Working Group to be an even more regular channel for communication, collaboration, and information sharing. Furthermore, the MA established an email helpdesk in 2016 to share training slides, briefing notes, and facilitate communication with GoIRA staff (which was monitored and queries passed to MA team members for response).

Generally speaking, GoIRA officials have more awareness, experience, and knowledge of their commitments to comply with rules and regulations, as evidenced by the achievements made on rectifying ineligibilities (see Deliver section). Finally, in regards to the IW-financed internal controls, the evidence is clear: GoIRA line ministry counterparts accepted most of the recommendations made by the MA and therefore derived value from the findings.

## Donors

**Donors interviewed remarked that they derived increased value from the reports shared with them over the course of the contracting period, but there is still further demand for heightened information sharing.** Overall, there appears to be a need for clarity/agreement among the WBG, donors, and the MA in order to bridge the gap between MA’s contractual responsibilities (i.e. what it does and does not do) and the donors’ information requests (i.e. what they would like in order to perceive ultimate value for money).

**Donors interviewed remarked that information sharing improved considerably in 2018 when they started to receive the full MA Reports.** Up until this point, donors received a two-page summary on a quarterly basis which included some topline information. In 2019, the ARTF Steering Committee members formally recognized the improvements in communications between the ARTF and donors. This in itself is a significant result over the contracting period.

**However, some donors interviewed stated that they would like the reports to contain progress (or lack thereof) on GoIRA public financial management (PFM).** The utility of the headline ‘eligibility rate’ metric was questioned several times. At first glance, reporting ‘eligibility rates’ seems like a sensible way of recording whether money is being spent correctly. However, there are weaknesses with this approach. First, meeting eligibility criteria is not identical to ensuring that funds are being spent on their intended purposes. To overcome this, the MA applied additional risk controls such as verifying procurement plans, budgets, and approval of requests, etc. for the sampled transactions. Second, as a metric, it has remained static. GoIRA’s eligibility ceiling has at all times significantly exceeded the amount of available

---

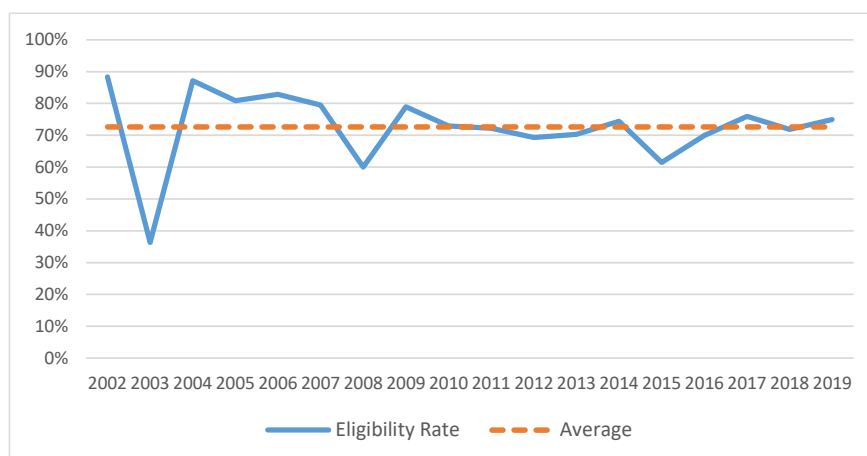
<sup>21</sup> The Review Team notes that ARTF contributions represent 18-20% of the government civilian recurrent expenditure.



ARTF recurrent financing. This could be interpreted as a significant and sustained success. Yet because the eligibility rate has not changed significantly over time, it fails to capture GoIRA progress in improving PFM and, to some extent, creates a sense of static reforms. As seen in the Figure below, the eligibility rate fluctuates over the years with an average of approximately 73%.

Figure 2: Evolution of Eligibility Rate 2002 – 2019

Source: MA ARTF Completion Report; MA ARTF FY1398 Annual Report.



WBG stakeholders noted a tendency among donors to focus on the eligibility rate rather than the eligibility ceiling and queried whether it was perceived as a proxy for corruption. While none of the donors who participated in this study mistook ineligibility for a measure of corruption<sup>22</sup>, they did note that their colleagues at headquarter-level in their respective home countries were often less discerning. The absence of alternative ‘headline’ indicators for the RCW meant that donor units in Kabul would report the eligibility rate back to their respective headquarters overseas. At this level, conceded one donor, the potential for misunderstanding was significant as government officials working at headquarter-level are less familiar with the context. Once the precedent for reporting eligibility was established, it became difficult to steer focus within the partnership towards other indicators. And as this indicator shows little signs of sustained improved government performance, donors continued to expect further assurances and greater levels of oversight. At the same time donors acknowledge that progress towards reducing government fiduciary corruption is a long-term objective and that improved GoIRA PFM sits outside the MA’s objectives.

The Review Team notes that the MA did make recommendations on how to make progress on eligibility through incentivizing GoIRA and finding mechanisms to digitize payments (reduce paper documentation). These recommendations and others are summarized in the Table found in Appendix B, delivered by the MA in 2016 to the WBG and in some cases reiterated in 2020 in their contract Completion Report.

Finally, as mentioned previously, the MA did produce data on rectifiable ineligible expenditures once reclassified and followed up as eligible (meaning, there is evidence to suggest that they took action with the MoF to correct mistakes on a monthly basis). It is unclear to the Review Team if these recommendations and progress on rectifiable expenditures were shared with the ARTF donors. What is clear is that there is demand for heightened information sharing on ‘progress’ and therefore the current TPM agent, the WBG, and donors should discuss exactly what information is needed and how often so that donors perceive more value for money.

<sup>22</sup> The Review Team acknowledges a reporting bias here. The donors who took part in this review were willing to do so because they were either a) currently involved in the ARTF, or b) had a strong interest in the program. It is unlikely that all donors were as well informed as this group, especially given the generally high levels of staff turnover in Afghanistan.



## 5. Supervisory Agent

### 5.1. Summary

Focus Box 3: Summary of SA Headline Findings

Review Question	Headline Findings
Did the SA <b>design</b> an appropriate and adequate methodology to provide the contracted services?	The SA designed a strong methodology suitable to deliver the services outlines in the contracts.
Did the SA <b>deliver</b> these services in line with contractual obligations and quality standards?	The SA delivered the services in line with its contractual commitments, and in many cases exceeded its original mandate.
Were the <b>results</b> of the SA's services useful to key stakeholder groups?	WBG and donors <i>generally</i> feel assured by the existence of the SA. The SA provided a good example of project monitoring and data collection to the GoIRA. It is unclear whether this contributed to an improvement in government service monitoring and delivery.

### 5.2. Introduction: What was the SA contracted to deliver?

The SA had a base contract and two extensions: November 2015 – October 2018; October 2018 – March 2019; and April 2019 – December 2019.

The base contract contains three overarching objectives:

- To provide critical data on small-scale infrastructure projects nationwide, including asset verification and quality assurance, as input to the WBG's implementation support and line ministries' own monitoring systems;
- To provide additional evidence to donors that projects were being implemented correctly, infrastructure was constructed properly, and social and environmental safeguards, gender issues, and select financial/fiduciary aspects were all considered;
- To provide line ministries with an example of project monitoring and data collection and showcase how such practices could improve project performance and results. The goal was to strengthen line ministries' own monitoring approaches and capacity.

Beneath these overarching objectives, the contracts outline a general scope of services for each of the individual projects funded through the IW. The scope of services varied from one project to the next, and from one contract to another as additional projects were added.<sup>23</sup> The Review Team does not analyze TPM performance against these individual project-level objectives as this is beyond the scope of services contracted for this Review.

<sup>23</sup> For a full list of projects covered by the SA, please refer to the 'Deliver' section of the report below.



The contracts provide a clear description of reporting requirements. Taken from the base contract, these are:

- a. An **inception report**, containing a detailed program of work, details of how the management information system will function, and a staffing plan;
- b. **Monthly reports**, prepared in consultation with the respective line ministries and WBG teams, providing a description of activities conducted and key findings;
- c. **Quarterly reports** produced every three months, summarizing key trends and lessons for each project, in a format that can be shared with the ARTF donor community, line ministries and WBG task teams;
- d. **Annual report** summarizing overarching trends and analysis, to be shared with implementing ministries, the WBG, and the ARTF donor community;
- e. A **web-based database and interface** was required to share data and reports with the WBG and counterparts in the line ministries and in particular the MoF;
- f. **Completion report** at the end of the contract period providing a frank assessment of TPM achievements, strengths, weaknesses, and lessons learned.

Interviews with the WBG CMU staff confirmed that the initial ToR in the overarching SA contracts were intentionally replicated from the pre-existing ToR for the previous SA firm<sup>24</sup> in order to ensure a comparable level of service delivery and outputs.

### 5.3.Design

Did the SA design and employ an appropriate and adequate methodology in delivering the contract services?

**In general, the SA employed appropriate and adequate methodologies in delivering contract services.**

At the portfolio-level, the feedback from WBG Task Teams regarding the quality of SA's methodologies is generally highly complimentary. Site selection was a participatory process, reflecting the needs and priorities of the WBG and line ministries, and the SA demonstrated flexibility in developing good data collection instruments, and piloting them pre-deployment.

#### *Site Selection and Sampling*

**The SA employed a robust methodology to select subproject sites for fieldwork.** As noted above, the SA worked closely alongside WBG Task Teams to design tailored methodologies for each project. The justification for site selection varied from one project to the next depending on a) the number of site visits pre-allocated by the CMU and b) the monitoring objectives for each project. The SA's Year IV Annual Report provides a detailed description of the variety of sampling methods employed. Projects with construction components were grouped by construction status using stratified random sampling. A combination of random or purposive sampling was then employed to identify specific sites for investigation. For those sites that did not contain a construction element, the SA either employed systematic random sampling, or targeted sampling, responding to direction provided by the WBG to investigate specific subproject sites.

It was not possible to provide representative sampling for each of the projects given the sheer number of different project types and differing stages of implementation and completion. Nor was it necessarily desirable to employ a random selection approach, since projects sought where possible to identify problems at subproject locations within contractual remedy periods. The SA therefore worked closely with WBG Task Teams to identify priority site visit locations to ensure that project results were captured at all stages of development from construction/inception to completion and O&M.

---

<sup>24</sup> IRD, now Blumont, held the SA contract from 2011 – 2015.



## Tools

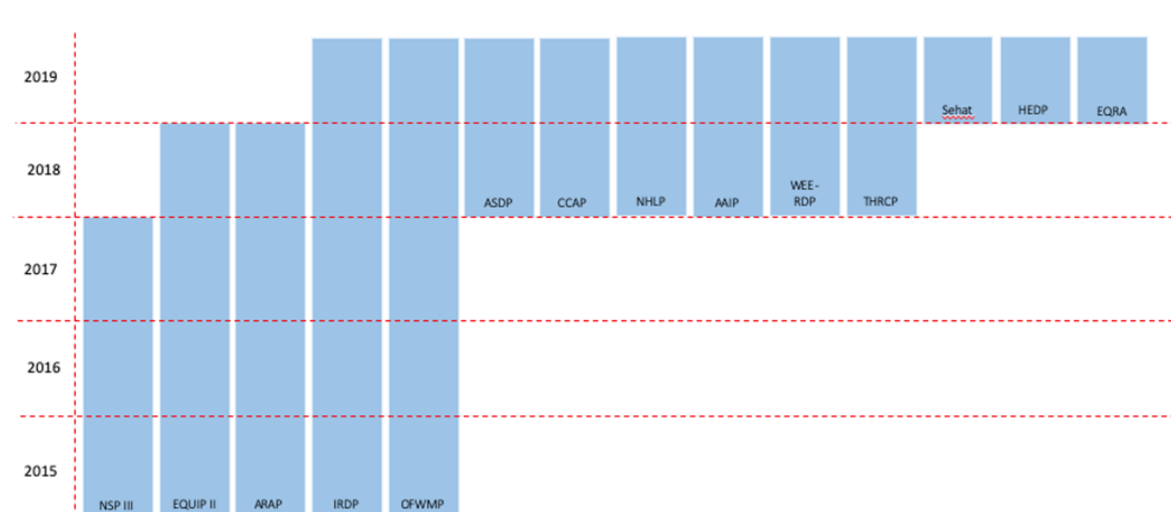
The SA employed a range of tools to meet the objectives of each individual project. The requirements of the SA contract expanded over time and varied between projects. Since 2016, the general package of data collected from each site visit usually contained infrastructure and non-infrastructure (environmental and social safeguards and gender) components, which required the SA to record observations and report deviations to the respective implementing line ministries, so that they could take remedial action.

- Infrastructure build quality** - subproject sites were graded using a 1-5 scoring system to denote whether construction was either 'very poor', 'poor', 'satisfactory', 'good' or 'very good'.
- Compliance with the World Bank's environmental and social safeguards** - indicators recorded included: use of land, the inclusion of women, the availability of a grievance redressal mechanism, and the presence of a first aid kit on-site.

The SA recorded the frequency and severity of deviations on infrastructure projects to help prioritize follow-up and rectification from GoIRA line ministries. Deviations were recorded as either *minor*, *major* or *life safety*, creating an easily navigable instrument to identify which issues needed to be prioritized.

At the design stage, the SA is to be commended for their flexibility in adapting to the growing number of projects that were added to their portfolio, and in adapting to the growing length, complexity, and diversity of tools required to conduct the monitoring visits. The chart below documents the growth of the portfolio of projects employing SA field monitoring over the course of the SA's contract period.

Figure 3: Growth of the TPM portfolio



In 2019, there was a step change in the volume of questions that needed to be addressed following the launch of the WBG's Environmental and Social Framework (ESF) in 2018. This was remarked upon by staff at the WBG, the SA, and GoIRA. According to the WBG, questionnaires to cover the new environmental and social safeguarding standards required between two to three hours to administer, in addition to pre-existing observation checklists and questionnaires. Interview fatigue in local communities is a significant risk with questionnaires of this length. One line ministry official engaged in the WEE-RDP reported that CDCs had expressed frustration at the volume of visits and, more importantly, the length of time taken to administer some of the questionnaires. This was especially true for women's groups, which usually take place in a member's house and which, in some cases, had been seen as a great imposition (see also the Focus Box on the Environmental, Social, and Gender Safeguards on page 30).





The Table below provides a summary of the main tools and approaches used for each project.<sup>25</sup>

Table 4: Examples of tools employed on each program

Project	TPM Activities
Education Quality Improvement Program II (EQUIP II)	Quality of construction; Presence and functionality of O&M plan; Availability and use of school materials and equipment; Teacher and student attendance; Contractor capability and performance; Payroll data verification; Environmental and Social Management Plan (ESMP) compliance.
Afghanistan Rural Access Program (ARAP)	Quality of construction; Presence and functionality of O&M plan; Contractor management efficiency, with focus on larger items in bills of quantity; Contractor capacity; ESMP compliance; Land tenure issues.
National Solidarity Program III (NSP III)	Infrastructure sites; CDC maturity assessment sites; PMU sites for systems monitoring.
Irrigation Restoration and Development Project (IRDP)	Sustainability of meteorological and hydrological stations; Canal construction; Presence and functionality of O&M plan; Environmental awareness.
On-Farm Water Management Project (OFWMP)	Quality of construction; Presence and functionality of O&M plan; ESMP compliance.
Afghanistan Power System Development Project (APSDP)	Electrical engineer report.
Citizens' Charter Afghanistan Project (CCAP)	Inspection of rural CDCs; Inspection of Urban CDCs; Inspection of Gozar Assemblies; CDC subproject construction.
National Horticulture and Livestock Project (NHLP)	Beneficiary perceptions; Subproject utilization.
Afghanistan Agricultural Inputs Project (AAIP)	Quality of construction; ESMP compliance.
Payroll Verification Assistance to BDO	Verify physical presence of GoIRA civil servants.

<sup>25</sup> ARTF Partnership Framework and Financing Program (2018 – 2020), p. 58





Women's Economic Empowerment Rural Development Program (WEE-RDP)	Interviews with CDC, social organizers, self-help groups, village saving and loan associations, and enterprise groups to understand successes and challenges with project implementation.
Trans Hindukush Connectivity Project (THRCP)	Road construction quality; Labour camp conditions; ESMP compliance.
Sehatmandi	Services and staff capacity; Assess medical supplies; Assess quality of recordkeeping.
Higher Education Development Project (HEDP)	Quality of dormitories, lecture and research facilities and dining rooms.
Education Quality Reform in Afghanistan (EQRA)	Construction quality.

### *Pilot Testing*

According to the SA's Completion Report (2019), **pilot tests were conducted with site visit tools once WBG Task Teams had approved initial draft questions.** Feedback was incorporated into questionnaires before full field deployment. The SA's first Monthly Report from December 2015 shows that pilot testing was carried out at 21 sites around Kabul, including each of the ARTF's five primary projects at the time. Both the WBG and GoIRA line ministry counterparts were involved at the time.

### *Data Collection and Analysis*

**The SA employed a range of methodologies to collect data from the field.** Broadly they fell into three categories:

- i) **Use of experienced national and expatriate male and female enumerators** (engineers and social scientists). These staff members were employed to conduct initial site visits, looking at performance and process monitoring.
- ii) **Participatory monitoring conducted by trained male and female Citizen Monitors (CMs)** from the subproject surrounding areas. CMs were trained by the SA's engineers to provide follow on verification and monitoring after the initial site visit had been conducted. The CMs provided weekly photographic evidence of site progress.
- iii) **Remote monitoring using technology solutions** (aerial/satellite imagery, etc.).<sup>26</sup>

Each site visit required a different configuration of team members. Some sites required only male engineers, while other may have required female social scientists to engage with male and female community groups. Working closely alongside government counterparts (see also 'Delivery' section below) the SA was able to identify site visit needs in advance and deployed teams accordingly.

In addition to the data collection teams, the SA employed a series of managerial layers and support units to ensure effective delivery of services, including:

- i) **Quality assurance team** – this team was responsible for designing the monthly site visit mission plans in coordination with the line ministries. This team also tracked the

<sup>26</sup> ARTF Partnership Framework and Financing Program (2018 – 2020), p. 57.



- rectification rates and reported to the WBG task teams. This team also acted as the primary contact with the line ministries.
- ii) **Program compliance team** – this team was responsible for ensuring accuracy and quality of all reporting, including subproject site visit report, monthly reports, quarterly reports and annual reports.
- iii) **Program management team** – this team provided oversight of the whole SA program, acting as the primary liaison with the WBG CMU and the TTLs. The program management team provided the final sign-off on all deliverables to the WBG.
- iv) **Program support team** – this team provided logistics and operational support to the other teams.

Taken together, the SA established a strong data collection, analysis, reporting, and quality assurance structure to ensure services were delivered on time and within budget.

### *Participatory Approaches*

**At an individual project level, the SA engaged closely with relevant stakeholders in the WBG and GoIRA to develop individual project-level scopes of work, and to ensure that implementation methods were suitably aligned with project objectives.**

**World Bank Group** - The WBG CMU was responsible for liaising with each TTL to agree on the allocation of site visits for each project. The SA was then engaged by the TTLs to agree on how best to distribute the allocation of site visits for their respective projects. Larger projects, like CCAP, were allocated a greater number of site visits than smaller ones, such as the WEE-RDP.

The SA's Completion Report (2019) outlines the typical process, beginning with an introductory planning session, and culminating in the preparation of all monitoring instruments, including engineering, environmental, and social science. According to the WBG TTLs who participated in this Review, this process was generally positive. The SA provided strong technical insights and suggestions, leading to detailed ToRs and work plans.

However, both WBG staff and the SA noted that collaboration varied depending on the interest level of the WBG TTL. Some WBG TTLs were highly engaged with the TPM process and played a highly active role in collaborating with the SA to develop ToRs and questionnaires. SA staff recall one TTL in particular who was extremely engaged with TPM, available for meeting in person or virtually whenever needed and pro-actively managing the relationship with the GoIRA line ministries to ensure follow-ups to deviations were carried out. Others were less involved, resulting in questionnaires that did not vary from one year to the next. Thus, the extent to which the SA engaged with the WBG during the design (and implementation) of the project was, in no small part, determined by the TTLs themselves. The Review Team finds that when given the opportunity to do so, the SA engaged WBG Task Team counterparts actively throughout the initial design phases of each TPM cycle at project level.

**GoIRA** - The SA engaged with GoIRA counterparts to refresh and recalibrate the methodologies employed for each project from the previous contract holder. The SA's first quarterly report, covering the period November 2015 to January 2016, states that a detailed review of the previous methodologies was carried out between the SA and GoIRA. One of the primary lessons from this exercise was the need for deepened collaboration between the SA and line ministry counterparts. On the basis of this review exercise, the SA embedded a full-time Knowledge Manager in each ministry to support the key pre- and post- site visit activities. Levels of coordination and collaboration with GoIRA generally remained positive over the course of the contract periods.



## 5.4. Deliver

The assessment of the SA's performance on service delivery focuses on the first of the SA's three overarching contractual objectives:

- *To provide critical data on small-scale infrastructure projects nationwide, including asset verification and quality assurance, as input to the WBG's implementation support and line ministries' own monitoring systems.*

The assessment is built on three questions:

- Did the SA deliver on services in line with the objectives and timelines set out in the relevant contracts?
- Did the SA deliver on reporting requirements in line with the objectives and timelines set out in the relevant contracts?
- Did the SA provide adequate data to enable WBG teams to make informed project supervision and implementation support decisions?

**Did the SA deliver on services in line with the objectives and timelines set out in the relevant contracts?**

The Review Team finds that the SA delivered services in line with the objectives and timelines set out in the relevant contracts. The SA exceeded the annual targets for site visits. The SA adopted inclusive approaches and provided flexible service delivery. The SA employed innovative approaches to data collection and made good use of open-source and bespoke technologies. Approaches were delivered in a manner sensitive to fragile and conflict affected states.

### Site Visits

The SA conducted site visits in all 34 Afghan provinces, meeting or exceeding its annual targets for site visits. In the final year of its contract, the SA was contracted to conduct 2,000 site visits; it delivered 2,224. The target for site visits in years I-III was 1,900.<sup>27</sup> In the Year III Annual Report the SA records that it conducted 1,949 site visits between November 2017 and October 2018. In Year II, the SA conducted 1,902 site visits and in Year I they conducted 1,967 site visits. The table below provides a breakdown of the number of site visits conducted for each of the ARTF projects that required field monitoring in descending order.

Table 5: Summary of site visits by project<sup>28</sup>

Project	Number of Site Visits
Citizens' Charter Afghanistan Project (CCAP)	2,078
Education Quality Improvement Program II (EQUIP II)	1,736
National Solidarity Program III (NSP III)	1,105
On-Farm Water Management Project (OFWMP)	525
Afghanistan Rural Access Program (ARAP)	431

<sup>27</sup> ARTF Partnership Framework and Financing Program (2018 – 2020), p. 57.

<sup>28</sup> SA ARTF Completion Report (2019).



National Horticulture and Livestock Project (NHLP)	400
Irrigation Restoration and Development Project (IRDP)	343
Education Quality Reform in Afghanistan (EQRA)	321
Women's Economic Empowerment Rural Development Program (WEE-RDP)	100
Payroll Verification Assistance to BDO	Equivalent of 92
Afghanistan Agricultural Inputs Project (AAIP)	63
Higher Education Development Project (HEDP)	19
Sehatmandi	15
Trans Hindukush Connectivity Project (THRCP)	11
Afghanistan Power System Development Project (APSDP)	8

An extensive review of project reporting material and interviews with WBG Task Team and CMU staff, the SA, and line ministries shows that the SA employed strong approaches to data collection requirements throughout the contract periods. Examples of good practice are highlighted below.

#### *Inclusive Approaches*

**The SA employed highly participatory and inclusive approaches throughout the contract period.** Notable examples include:

- **The SA employed CMs** to provide regular on-site monitoring and follow-up verification visits by people who live in the local area. The SA recognized early on that the single site visits they were mandated to do for each subproject site were insufficient to determine whether they were being delivered to agreed standards. CMs therefore filled an important monitoring gap by providing an additional layer of scrutiny and data. CMs were recruited through the local CDCs and Shuras, and were able to access locations that the SA and GoIRA's staff were unable to visit. A total of 213 CMs were recruited and trained over the course of the project. Reviewing the quality and rigor of the CM deviation reports from EQUIP, ARAP, IRDP, and OFWMP, it is clear that significant time and resources were invested in training and capacity building. CMs were tasked with collecting a broad range of data points, including photographs of specific architectural features, in such a manner as to show compliance with construction plans and build quality. Over 12,000 observations were recorded by CMs at 314 ARTF subproject sites.
- **The SA conducted fieldwork in all 34 of Afghanistan's provinces**, including in remote rural locations and hard-to-reach areas, ensuring that communities from diverse ethnic groups and socio-economic segments were included in the monitoring process.
- **Hiring female staff** - In total the SA trained and recruited 30 female enumerators in order to conduct site visits for the CCAP. Female staff were trained to a level sufficient to lead the monitoring of the WEE-RDP. Interviews with WBG Task Teams illustrate that the addition of female enumerators significantly improved the quality of site reports, as women were able to engage with female community members, which had until that point been impossible. WBG Task Teams interviewed note that the capacity of female enumerators was initially quite low, but acknowledged that their capacity improved significantly over the course of the contract periods. The SA disaggregated findings by sex where applicable to help ARTF stakeholders



understand the differential impact on men and women. Data was sex disaggregated in the annual reports where relevant to demonstrate gendered participation in ARTF projects.

*Focus Box 4: Environmental, Social, and Gender Safeguarding Standards*

**Headline: The Review Team finds that the SA generally collected good data on environmental and social safeguarding standards, but notes that this was a challenging undertaking as the scope of services gradually increased over the course of the contracting period.**

The SA was contracted to provide additional evidence to WBG and donors that projects were being implemented in compliance with environmental and social safeguarding standards.

In Year I, the SA graded environmental and social safeguards, and gender performance by project, providing an overview of general compliance.

In Years II and III, the SA provided a more detailed breakdown of environmental, social, and gender safeguarding trends. The reports looked at trends over four indicators: i) use of land; ii) the consultation of women on the subproject; iii) the availability of grievance redressal mechanisms; and iv) the presence of a first aid kit on-site. Deviations were recorded for each visit to denote whether and to what extent these standards were adhered to.

As noted in the main body of the report, the WBG Safeguards team recall a steep learning curve for the SA, whom they felt lacked the capacity to collect this type of detailed information. In response, the SA upskilled existing enumerators in social science techniques and recruited additional female staff to address the expanded scope of safeguards. After this, the WBG noted an improvement in data collection quality.

In Year IV the WBG updated all site visit instruments with project specific safeguards modules following the launch of the WBG's ESF the previous year, which contains 10 environmental and social standards (ESS). Each standard contains an extensive list of requirements for implementers (borrowers). Four of these standards apply to the ARTF projects:

- ESS 1: Assessment and management of environmental and social risks and impacts;
- ESS2: Labour and working conditions;
- ESS4: Community Health and Safety; and,
- ESS5: Land acquisition, restrictions of land use, and involuntary resettlement.

The SA incorporated these new elements into the existing site visit toolkit, significantly increasing the volume and complexity of data required from each visit. Depending on the type of site being inspected, the length of ESS questions ranged from just over four pages of condensed questions for the WEE-RDP project to over five pages of condensed questions for the Naghlu Hydropower Rehab. Project.

It is unclear from the data whether capacity to adhere to environmental and social safeguards standards improved for any stakeholder group, whether government or community. As noted above, this is partly due to the fact that the standards against which compliance was being measured changed over time. It is also a function of the fact that the number of variables within and between projects from one year to the next is too great to permit a clear analysis of trends.

### *Adaptability*

**The SA demonstrated flexibility throughout the contract period, responding to the growing scope of the TPM program and adapting to new requirements as they emerged.** WBG staff who engaged directly with the SA during the contract period made a point of highlighting the SA's flexibility. Some key examples born out of the interviews and program literature include:



- **Upskilling existing project staff** - In 2016, the SA was required to collect data to ensure compliance with environmental and social safeguarding policies. This required the recruitment and training of social scientists, qualified in administering such questionnaires. Engineers, originally recruited to conduct infrastructure assessments, received additional training in social science data collection methods, including interview administration. The WBG environmental and social safeguarding teams recall this required a lot of capacity building to ensure the reports delivered useful insight.
- **Inclusion of special projects** – the SA was required to deliver a range of special projects in response to a variety of needs that emerged during the course of the project implementation periods. Sometimes these special projects would be undertaken in response to issues highlighted through the grievance redressal mechanisms; at other times they were undertaken to provide further information or insight on points raised in the regular reporting schedules. The SA Completion Report (2019) provides a summary of the special studies undertaken over the contract period. These special projects varied significantly in size and scope. For example, in one special project, the SA conducted 60 site visits in Helmand and Kandahar to investigate emergency seed distribution. In another they conducted 30 site visits to examine the economic internal rate of return and cost effectiveness study for the CCAP. Again, WBG Task Team staff provided extremely positive feedback about the SA's flexibility and capacity to respond to new unanticipated project requirements.
- **Transition to online platform** – The SA initiated a transition to a digital deviation reporting platform, Ardea, in order to meet growing demand for real-time data and information sharing (discussed in further detail below).
- **Tool development** – The SA provided an array of different tools and field instruments over the course of their engagement. The tools grew in scale and complexity, particularly with the inclusion of environmental and social safeguarding questions, which were not part of the original package of site visit questions. As noted above, this required staff to be retrained in new techniques, as well as additional reporting and analysis. In Year III, for example, the inclusion of CCAP, NHLP and the Afghanistan Agricultural Inputs Project (AAIP) programs necessitated the addition of beneficiary perceptions, project impacts, and gendered participation.
- **Ad hoc presentations** - At the request of the WBG, the SA conducted 25 ad hoc presentations to ARTF donors on a range of subjects.

### *FCV Sensitivity*

The SA demonstrated strong capacity to deliver work in a manner appropriate to and sensitive of a country affected by FCV. The SA delivered fieldwork in all 34 provinces of Afghanistan, including some of the most volatile (Nuristan and Kunar) and some of the most kinetic (Helmand and Kandahar). The ability to operate in these locations - where many international organizations were not operational at that time - demonstrates effective approaches to working in challenging environments. Examples of sensitive FCV approaches include:

- **Agile implementation** - The SA liaised closely with GoIRA in order to avoid active conflict affected locations. According to the SA's records, 6% of site visits had to be altered due to active fighting between GoIRA forces and the Taliban, or when the Taliban simply closed road access. In these instances, the SA sought alternative site locations in the same district or provinces in order to maintain sampling integrity.
- **Female staff welfare** - The SA employed good approaches to ensuring the protection of staff undertaking site visits. Female staff, in particular, were vulnerable to threats and acts of violence, especially in Taliban controlled or highly conservative areas. To mitigate risks faced





by female staff members, each was accompanied by a *mahram*<sup>29</sup> and a male SA engineer. The SA engaged closely with local communities prior to conducting site visits to determine whether it was safe to deploy, and whether it would be appropriate for female enumerators to conduct work. Indeed, female enumerators were required to contact subproject focal points to check for any local issues (typically female CDC members, social organizers or female self-help group members).

- **Updating Personal Protective Equipment (PPE) Policy** - The Year III Annual Report describes challenges associated with the need for community members working on ARTF projects to wear PPE. According to community members, this enabled the Taliban to (mis-) identify them as government workers, thereby creating an unnecessary level of risk. The SA identified this issue and worked with the WBG to develop new PPE project policy better suited to the context.

### *Technology and Innovation*

**The SA actively sought opportunities to employ technology to deliver TPM services more effectively.** Specific examples of the innovative use of technology include:

- **The SA trained government staff in the use of Open Camera**, an open-source camera application for Android phones, which embeds geotags and time stamps in photographs to facilitate verification.
- **The SA staff employed Fulcrum**, a mobile and web-based application to capture and manage geotagged pictures. SA enumerators were required to use this software to take at least three GPS readings at each subproject site and at least one 'selfie' to verify that they themselves actually conducted the work.
- **The SA piloted remote sensing technology to improve data collection.** The SA developed bespoke equipment for canal sensor monitoring and environmental probe monitoring for the IRDP and OFWMP projects. Using locally sourced materials, the SA built two types of stationary and mobile sensors to detect motion, moisture content, relative humidity, air temperature, and solar radiation. Though the sensors themselves worked effectively, local community stakeholders were unable to prevent the sensors from being stolen or vandalized. The SA employed CMs to sensitive community leaders to the use of remote sensors, but theft and vandalism persisted, preventing a wider scaling up of operations.

### *Quality Assurance*

**The SA employed good quality assurance and data verification methods to ensure accurate reporting.** During site visits, enumerators were required to take a GPS reading of each project site, along with a selfie, to prove that they themselves conducted the work. Once the site visit was completed GPS coordinates were cross-referenced using satellite imagery to ensure the correct site location. The completed site visit reports were then reviewed by two Afghan staff members (the Program Compliance Manager and the Environmental and Social Safeguards Specialist) and by two international staff (the Project Manager and the Technical Manager). This is commendable practice, especially given the large quantities of data being submitted on a monthly basis.

### *Government Capacity Building*

The SA did not have an explicit capacity building mandate as this could (rightly) be perceived as a conflict of interest. **The objective of the TPM therefore was to provide an 'example' of good practice**, thereby building government capacity through a process of co-delivered activities (such as site visits, which were almost always carried out by the SA and government representatives at the same time), and the provision of tools to strengthen monitoring and reporting procedures.

---

<sup>29</sup> A close male relative - that is a husband, father, brother, or son.



**The Review Team finds that the SA performed this role well.** The SA developed a strong collegial working relationship with the line ministries. The SA embedded staff in implementing line ministries to ensure good communication and coordination of site visits. Interviews with government teams both within the MoF and the line ministries confirmed that the relationship with the SA was extremely positive. The Review Team detected only one area of potential friction between the SA and the line ministries: the late-sharing of monthly reports with the ministries. As the SA's monthly reports show, raw data and the reports themselves were transferred to the government for the previous month. This caused some frustration among government staff, who felt they were being 'sidelined'. This also resulted in data being out of date, as the government may already have redressed the issues raised in the reports. Nevertheless, the overall impression of the SA's performance remains highly positive across all government teams who took part in this Review.

The SA's contractual mandate to support government project monitoring capacity and to provide independent monitoring of those self-same activities required the SA to balance two roles that could pull in different directions: the former inviting close collaboration and the latter requiring distance and impartiality. The Review Team's research suggests that the SA maintained this balance effectively throughout the contract period. Site visits were planned in advance, so that government project teams would know when the SA was due to visit. According to SA staff, over time, this led to a reduction in the severity of deviations. This, they felt, was a more effective means of addressing deviations than simply trying to 'catch them out'. The SA was grateful of the leeway granted to them by the WBG to develop close working relations with the government. They felt this allowed them to pivot the relationship away from 'auditor-auditee' towards a partnership of equals focused on ensuring that projects delivered the desired results. Given the broader strategic goal of strengthening government monitoring capacity, the Review Team views this as a highly positive attribute of both the SA and the WBG's approach to TPM. The impact of the SA's support to line ministries is explored in the 'Results' section below.

### Did the SA deliver on reporting requirements in line with the objectives and timelines set out in the relevant contracts?

**Reports were produced on time, on a monthly, quarterly and annual basis.** A review of these reports contained on the WBG shared drive confirms this to be the case. The monthly reports contained detailed descriptions of the number of sites visited and the number of deviations uncovered. They contained analysis of the nature of deviations and a breakdown of deviations and rectification by ministry. They also contained sections on 'trend analysis', 'challenges', and 'recommendations' - all of which are well written and well presented. To this extent, the reports generally provide a good snapshot of each project at a given moment in time.

However, **some of the WBG Task Team staff interviewed remarked that the reports from the SA lacked analysis.** They felt the reports were too descriptive. At the same time, the SA felt that they were unable to provide further depth of analysis given the constraints of the reporting formats from the WBG, the volume of data they were required to collect, and perhaps most importantly of all, the sheer volume of reporting they were required to produce each month. Both parties acknowledged that reporting improved over time, yet both continue to feel that more analysis could be provided.

The Review Team's own analysis of the annual and quarterly reports acknowledges the challenge of reporting both breadth and depth across the IW portfolio. Projects funded by ARTF were highly diverse in nature, each with their own operational constraints and challenges, rendering comparisons between projects extremely difficult. Even analyzing trends, such as rectification rates, within projects is difficult as data collection instruments evolved from year to year. The SA provided strong justification for not tracking rectification rates on an annual basis in their 2019 Completion Report: *Comparison of annual rectification rates should likewise not be compared simply because each year the number of projects has increased under TPM [...] making such comparisons misleading.*





It is clear from reading the quarterly and annual reports that the SA had endeavored to provide an analysis of trends, but the analysis is not carried out consistently from one year to the next. Annual Reports from Years I and II contain a useful summary of infrastructure grade scores from each of the original five projects supported under ARTF, providing a helpful snapshot of annual progress. The scores are not repeated in Years III or IV. In Year I, the SA recorded deviations and rectifications for each project. In Year II this information is not recorded in the annual report, but in Year III it reappears as a cumulative tally of deviations and rectifications, making it impossible to detect yearly trends. The Year III Annual Report contains a useful categorization and summary of deviation causes (design, material and workmanship quality, operations and maintenance, environmental, gender, and social safeguards) yet this same analysis is absent from the other annual reports. It is unclear why these changes took place.

#### *Ardea: Digital Deviation Tracker*

**The SA delivered a digital deviation tracker in line with contractual requirements.** In Year III, the SA established a digital platform to track deviations - Ardea. This replaced excel tracker sheets, which, though comprehensive, were cumbersome to navigate and time-consuming to produce according to the SA. In Year IV, the SA reported that ministry project teams working on community or contractor construction activities transitioned to Ardea. According to the SA, the large majority of user accounts were assigned to GoIRA staff, who regularly used the platform to monitor their own projects. Uptake of Ardea by WBG teams was varied. According to the SA, there were very few WBG registered users, and interest in the platform from WBG staff was generally low, though by the SA's own admission they could have done a better job at educating WBG staff in using the data and the dashboards. None of the WBG staff the Review Team interviewed regularly used Ardea, relying instead on monthly, quarterly and annual reports. As noted previously, the WBG CMU itself acknowledges that engagement with the SA, and indeed the Ardea platform, varied considerably from one TTL to the next.

#### *Value for Money*

Assessing cost-effectiveness and value for money of the SA's work is extremely difficult due to the complexity and diversity of services provided under the contract and the unprecedented scale of operations within the WBG. Donors, in particular, would like to see a clearer calculation of value for money. They ask: do the efficiency and cost savings resulting from TPM outweigh the cost of TPM services?

Over the contract period, the SA cost USD 25.28 million. The ARTF disbursed approximately USD 965.3 million for closed and open IW-financed projects monitored by the SA over the same period.<sup>30</sup> **As such, the SA cost 2.6% of funds disbursed to projects monitored by the SA during this period.** This cost sits within the broad range of TPM costs described below in the literature review (0.8-3.9%).

**While all parties interviewed for this study recognize that TPM is expensive, perceptions among WBG staff were universally positive.** In the absence of other physical monitoring tools that could operate at the scale and with the degree of flexibility of the SA's program, they argue, the cost of TPM is entirely justified. Having conducted numerous large-scale quantitative household surveys in Afghanistan, the Review Team can confidently confirm that none of the information gathered by the SA could have been gathered through open-source data in the public domain. The government does not publish information of the kind collected by the SA online. It is unlikely the data could have been collected by INGOs / LNGOs, many of which struggle with their own internal monitoring requirements.

---

<sup>30</sup> Administrator's Report on Financial Status, July 2020. This is a rough calculation of the amount of money disbursed to IW closed and open projects that were supervised by SA during the contract period. The calculation only accounts for the years in which the SA was engaged on a particular project. For example, The Review Team calculated total funds disbursed to AAIP for 2018 and 2019 only and excluded 2015, 2016 or 2017 as the SA was not engaged during these years.



The WBG recognizes that there are concerns, both within the WBG itself and in the wider development community, that TPM is extremely expensive. However, as noted above, when viewed within the context of the total value of the IW, the SA contract value sits firmly within standard cost brackets.

**Did the SA provide adequate data to enable WBG teams to make informed project supervision and implementation support decisions?**

**Overall, WBG Task Teams felt that the SA provided adequate data to support supervision and implementation support decisions.** WBG Task Teams were not able to visit project locations to carry out site verification visits due to insecurity, therefore the ability to gather data through the SA provided Task Teams with an important additional layer of information to guide implementation. As WBG Task Team staff report, the SA became their 'eyes and ears' on the ground, providing critical field-level insight and data.

**WBG staff interviewed who no longer work in Afghanistan would like the same level of TPM support for their projects in their current duty stations.** In other countries, the number of site visits that WBG TTLs conduct varies from one project and operating context to another. In Nepal and Bangladesh, for example, one TTL reports that visits are conducted one to two times a year. The SA therefore provided a level of project oversight significantly beyond the scale and scope of anything delivered in other contexts. Arguably, the fragile operational context in places like Afghanistan justifies this additional level of assurance, as levels of state corruption are high and government implementation capacity are low. Nevertheless, the level of oversight the SA contract provided to the ARTF remains unprecedented within the WBG; even in other FCV contexts.

**Interviews with WBG staff show that expectations of the SA grew significantly over time.** By the time MSI took over the SA contract from IRD in 2015, there were already a suite of sophisticated data collection instruments in place, which generated large volumes of data. Over the course of its contracting periods, these tools became increasingly complex as additional layers of investigation were added – notably the environmental and social safeguards. As new projects were added to SA's portfolio (increasing from five in Year I to 12 in Year IV), both the volume and range of data collected on a monthly basis grew in proportion.

This resulted in reports that WBG TTLs viewed as descriptive rather than analytical, as noted above. The SA responded by bringing on additional internal resources to produce a deeper analysis of trends over time in both the quarterly and annual reports. By the end of the SA's contracting period in 2019, WBG Task Teams were generally satisfied with the depth of analysis in the reports they received.

Interviews with WBG Task Teams detected other minor criticisms, but nothing that undermined the fundamental assurance they felt at having the support of the SA for their program. Some of the WBG staff interviewed highlighted the potential for subjectivity when assigning severity grades to deviations. One enumerator may consider a deviation to be 'minor', while another may consider the same deviation at another site 'major'. The SA also recognized this limitation, and therefore stopped tracking grading trends in quarterly and annual reports after Year II. Grading was not dropped entirely as it was felt that within each project it still provided a useful and transparent framework for assessing deviations.

## 5.5. Results

The assessment of results focuses on two of the SA's overarching contractual objectives:

- To provide additional evidence to donors that WBG teams can effectively determine if projects are being implemented correctly, infrastructure is constructed properly, and social and environmental safeguards, gender issues, and select financial/fiduciary aspects are all considered;



- To provide line ministries with an example of project monitoring and data collection and showcase how such practices could improve project performance and results. The goal was to strengthen line ministries' own monitoring approaches and capacity.

The assessment is built on two questions:

- Did the SA's services provide donors with assurance that ARTF projects were being implemented correctly?
- Did the provision of SA's services improve the government's capacity to monitor and implement projects?

### Did the SA's services provide donors with assurance that ARTF projects were being implemented correctly?

Relying on qualitative feedback from interviews, **the Review Team finds that the use of a SA generally did provide donors with greater assurances that programs were being implemented correctly, though there were some wider concerns around the use of TPM in principle.** The findings below are presented with the caveat that these views are not representative of the full ARTF donor universe either currently or historically, but they nevertheless provide an informed perspective that should be carefully considered.

Donors had very little direct interaction with the SA, as would be expected as a service provider to the WBG. Donors therefore base much of their feedback on the quality of SA's reports, and the nature of the information they contained. On this basis, some donors felt the reports were perfectly adequate, while others felt that they needed to have more data, particularly around the rectification and follow-up rates.

It is well understood by all parties interviewed for this review that donor reporting is often a challenge in any context. Donors themselves recognize that they are a difficult audience to please because they have diverse agenda, reporting mandates, and political pressures. Some donors only have the capacity to absorb a topline summary report, while others have the capacity to review and analyze full-length reports. As one donor pointed out: *'the US have around 1,000 staff in country, compared to say the Norwegians, who had one'*. In this regard, minor criticisms about the SA's reports by donors are an inevitable function of reporting to multiple stakeholders.

Donors did, however, have broader concerns about the structure and format of the TPM program (both the MA and SA components). They felt that the MA component could have been strengthened by the inclusion of more site visits in their methodology (indeed, recognizing this shortcoming, the SA worked with BDO in 2018 and 2019 to conduct physical verification of payroll expenditures). At the same time, donors felt that the SA lacked the fiduciary expertise to detect financial discrepancies during its own field visits. The merger of the SA and MA contracts in 2020 onwards is therefore highly positive.

The prevailing understanding among WBG staff is that donors expect high levels of TPM in order to justify their continued funding support to the ARTF, and this understanding did appear to be borne out by interviews with donor representatives. The view, however, is not universally held. One prominent donor the Review Team interviewed felt that TPM should be reduced. They note: *'the key word in ARTF is trust'*. They expressed a desire to see the role of the TPM reduced and wanted to see evidence that TPM had increased government capacity. In particular, they asked for a cost-benefit analysis comparing the money spent on TPM to the money 'saved' through effective project implementation and monitoring, and through a reduction in corruption.

### Did the provision of SA services improve the government's capacity to implement and monitor projects?

The Review Team notes again that the SA's mandate was to provide an *example of good practice* to strengthen government's monitoring programs rather than to conduct training of government staff.



**The Review Team tentatively concludes that the TPM services did contribute to an improvement in the government's capacity to implement and monitor projects.**

The Review Team relies on qualitative interviews and anecdotal evidence, supported by proxy data pulled from the SA's reporting. On this basis, the Review Team cautiously notes that there is some evidence that government project and contract management capacity has been increased. Key evidence to support this conclusion includes:

- The SA's Completion Report suggests that Ardea significantly improved rectification speed (time taken to fix problems identified in the field visits). For example, when using the original tracker database in excel, OFWMP project deviations would take an average of 122 days to be rectified, compared to only 69 days after the transition to Ardea. This was largely due to the elimination of time lag between conducting the site visits, reporting to WBG first, and then delivering the reports to the line ministries. The SA Year III report records that deviation rectification rates had improved since Year I, though they note that rates varied considerably between projects, with OFWMP scoring the highest (48%) and CCAP the lowest (5%).<sup>31</sup>
- The SA used historic reporting data from IRD to identify early trends in government performance. According to their analysis, the SA did have a positive impact on the level of documentation at ministerial and community levels. Lack of documentation at project level constituted 50% of deviations under IRD's tenure of the SA contract. Under MSI, this figure was reduced to 40% after the first year (Year I Annual Report). This figure is not tracked further in subsequent years.
- WBG Task Teams who have been engaged with the ARTF over many years have observed improvements in government monitoring performance, principally in the quality of data and accuracy of reporting (varying across sectors and ministries).

---

<sup>31</sup> The SA's own analysis in the 2019 Completion Report acknowledges that assessing rectification rates was inconclusive - MSI (2019): Completion Report, p.40.



## 6. Literature Review of Third-Party Monitoring in FCV Contexts

---

This section of the report provides a summary of the wider literature published on TPM performance to determine how the contractors' approaches were similar/distinct from other TPM arrangements.

**Overall, the review team finds that the SA and MA services were broadly delivered in line with wider industry practices.** The tools and approaches employed by the SA and MA respectively sit squarely within well-recognized methodologies used by other TPM providers working in states affected by FCV.

The use of TPM has increased in recent years due to protracted insecurity across many development and humanitarian theatres of operation.<sup>32</sup> In the past, TPM was seen as a temporary measure to fill data gaps in the event of poor field access. However, as noted in the literature, TPM is increasingly viewed as 'the new normal' for operating in states affected by FCV.<sup>33</sup> The WBG alone has entered into 14 TPM contracts, active and closed, in seven FCV countries from 2012 to 2020<sup>34</sup>, and the WBG Strategy for FCV 2020-2025 recognizes the need for continued selective use of TPM services in FCV contexts for monitoring, compliance verification, and implementation support<sup>35</sup>.

As a result, there is a small, but growing body of literature on TPM approaches in fragile contexts. The review team conducted a literature review of available material to help situate the analysis of ARTF TPM performance in a wider context of other TPM projects. The diversity of TPM approaches (both in terms of scope and scale), combined with localized operational idiosyncrasies, precludes a straightforward comparison between the ARTF TPM program and TPM programs conducted elsewhere.

Nevertheless, the recent publication of the EU's *Study on Best Practices in Third Party Monitoring (August 2020)*<sup>36</sup> offers a strong synthesis of common practices based on the experiences and views of a wide range of donors, TPM implementers, academics, and experts. Reports published by other donors (DFID 2014; USAID 2020), implementers (WFP 2016a and 2016b: ICRC), and others (Integrity 2015; GSDRC 2013 and 2018; GPPI 2016), as well as the World Bank's own *Demystifying Third Party Monitoring in Fragile Situations*, provide further context for analysis.

Taken together, a number of common themes emerge from a reading of TPM literature, which provide an instructive basis for comparing SA and MA approaches during the design and delivery phases of the ARTF TPM to approaches employed in the wider development and humanitarian sectors:

- **Calibrating approaches to local context**

The EU's recent synthesis paper on TPM<sup>37</sup> explains the importance of an inception phase to calibrate TPM approaches. It notes that TPM bidders require an inception phase to adjust the approaches outlined in their original proposals. The report encourages TPM providers to conduct pilot studies to field test their tools and methods before the full implementation of the TPM activities. As noted previously in the report, the SA conducted a small-scale pilot of its tools before the commencement of regular TPM activities, ensuring a well-calibrated suite of tools. For its part, the MA refined its methodology for the RCW during the inception phase and piloted its internal control assessments

---

<sup>32</sup> GSDRC (2013): Remote management of projects in fragile states. GSDRC Helpdesk Research Report; WFP (2016): Monitoring humanitarian assistance in conflict-affected settings.

<sup>33</sup> GSDRC (2018): Approaches to remote monitoring in fragile states, p.3; DFID (2015): No Longer a Last Resort: A Review of the Remote Programming Landscape; Integrity (2015): Cross-Cutting Evaluation of DFID's Approach to Remote Management in Somalia and North-East Kenya, p.vi.

<sup>34</sup> World Bank (2019): *Demystifying Third-Party Monitoring in Fragile Situations: A Global Review*, p. 3. Countries include Afghanistan, South Sudan, Somalia, Cameroon, Iraq, Yemen, and Pakistan.

<sup>35</sup> World Bank Group Strategy for Fragility, Conflict and Violence 2020 – 2025, p. 47.

<sup>36</sup> EU (2020): Study on Best Practices in Third Party Monitoring.

<sup>37</sup> Ibid, p.28.



during the latter part of the previous contract period, which are to be further refined during the current TPM agent contract.

- ***Tripartite planning***

The EU report describes the importance of engaging with implementing partners (IP)<sup>38</sup> during the inception stage in order to ensure IPs are involved in the early design stages of the TPM process. As the report notes, this is a vital step in building a strong collaborative relationship between project partners. The ICRC's desk review of TPM approaches<sup>39</sup> draws a similar lesson, encouraging tripartite planning during the inception stages to ensure a common understanding of the scope of work, division of responsibilities, logistics and security between the donor, implementing partners, and TPM agent. The review team finds that both the SA and the MA actively engaged with government line ministries during the inception phases of their respective contracts, as described in further detail earlier in this report.

- ***The use of technology***

In a 2014 paper, DFID encourages the use of information and communication technologies in monitoring and evaluation (M&E) for peacebuilding. The paper records the structural challenges to systematically collecting and sharing information quickly and accurately.<sup>40</sup> The paper describes a number of tools that could be used to improve M&E, including interactive and online maps, the use of mobile phones, and survey digitization. A more recent paper published by the Global Public Policy Institute (2016)<sup>41</sup> provides a toolkit for using digital technologies for M&E in insecure environments. It includes descriptions of remote sensing technologies, digital data entry, location tracking, and the use of online platforms. Still more recently, in 2018, the Governance and Social Development Resource Centre (GSDRC) published a paper encouraging the use of ICT and big-data enabled approaches to facilitate near real-time data sharing. The ICRC's paper also encourages the use of relevant technologies where applicable.<sup>42</sup> Here again, the SA in particular, is to be commended for actively seeking, testing, and applying new technologies to improve efficiencies in data collection and reporting, such as the Ardea platform, which provided an aggregated platform for building near real-time data visualizations, and an interactive map.

- ***The importance of continued collaboration***

One of the key lessons from the EU's recent paper on TPM best practice is to engage a TPM provider with the 'attitude, chemistry and local connectivity to succeed'<sup>43</sup> in challenging environments. Recognizing this as a core tenet of good TPM, the review team commends both the SA and MA, who demonstrated strong operational experience and physical presence in Afghanistan throughout the contract period. Another key lesson from the EU paper is the importance of building a 'coaching' relationship with implementing partners. A paper published online by the World Bank in 2017 reiterates the importance of working in conjunction with implementing agencies' internal monitoring systems.<sup>44</sup> Again, both the MA and SA are to be commended for working in a collegiate and collaborative manner with the government line ministries, as noted by government officials themselves, as well as SA and MA staff.

---

<sup>38</sup> Ibid, p.29.

<sup>39</sup> ICRC: Third-Party Monitoring: Desk Review and Implementation Guidelines, p.7.

<sup>40</sup> DFID (2014): ICTs for Monitoring and Evaluation of Peacebuilding Programmes.

<sup>41</sup> GPPI (2016): *Technologies for monitoring in insecure environments*. Secure Access in Volatile Environments.

<sup>42</sup> ICRC: Third-Party Monitoring: Desk Review and Implementation Guidelines, p.9.

<sup>43</sup> EU (2020): Study on Best Practices in Third Party Monitoring, p.8.

<sup>44</sup> World Bank (2017): Third Party Monitoring in Volatile Environments – Do the Benefits Outweigh the Risks?



- **Need to develop standardized tools to compare and contrast TPM value**

A commonly cited concern in literature is the considerable cost of TPM in fragile contexts and in particular in Afghanistan. Yet the literature also reveals a lack of systematic and standardized approaches to monitoring TPM costs in relation to outputs/value (e.g. standardized value for money calculation). In absolute terms, when comparing the cost of the ARTF TPM to other programs in fragile environments, the ARTF TPM in Afghanistan is very expensive. The argument then made is that one should not look at the absolute costs alone but rather also compare TPM cost as a percentage of fund/program gross disbursement. As seen in the table below, the ARTF TPM is then of similar proportional cost to other TPM engagements.

Table 6: TPM Cost Benchmarking<sup>45</sup>

TPM Engagements	Total disbursements for TPM (USD millions)	Total Fund/Program Disbursement (USD millions)	Timeframe	TPM cost: % of total disbursement
Afghanistan Reconstruction Trust Fund <sup>46</sup>	38.1	2,312	2015-2019	1.6%
Somalia MPF <sup>47</sup>	10.2	255.3	2014-2019	3.9%
South Sudan WBG Portfolio <sup>48</sup>	2	165	2017-2019	1.2%
USAID Portfolio in Afghanistan <sup>49</sup>	38.85	3,200	2019 - 2024	1.2%
The Multi-Donor Trust Fund in Pakistan <sup>50</sup>	1.6	200	2017-2019	0.8%

Cost as a percentage of gross disbursement confirms the proportionality of scale, but the deliverables-based nature of most TPM contract arrangements means that costs are not directly comparable between contracts unless deliverables are also compared. A key lesson, therefore, for the WBG Afghanistan team is to consider working with the WBG's Corporate Procurement and FCV teams to consider developing tools to enable cost comparison between and among deliverables-based TPM contracts.

<sup>45</sup> The calculations made in this table are based on limited available data and should be considered indicative only.

<sup>46</sup> ARTF Administrator's Report on Financial Status, July 2020. Program disbursement calculation only includes open and closed IW-financed projects monitored by the SA.

<sup>47</sup> MPF Somalia Progress Report, December 2019.

<sup>48</sup> World Bank (2019): *Demystifying Third-Party Monitoring in Fragile Situations: A Global Review*, p. 30.

<sup>49</sup> USAID Afghanistan Monitoring, Evaluation, and Learning Activity (AMELA) service contract. Note that spend and portfolio value is an estimate as the contract is commencing now.

<sup>50</sup> World Bank (2019): *Demystifying Third-Party Monitoring in Fragile Situations: A Global Review*, p. 31.





## 7. Observations and Recommendations

---

Working in a highly kinetic conflict theatre on a high-profile, multi-stakeholder donor fund, the SA and MA generally performed very well. A thorough analysis of documentation and interviews with key informants reveal that both the MA and SA provided professional services, broadly met their objectives, were highly flexible and despite TPM being expensive, mostly represented good value for money.

The Review Team notes that the new TPM agent contract is already in place and that the WBG has already taken steps to address several of the challenges presented in this Review and lessons detailed in key documents such as Demystifying TPM, SIGAR, and the WBG FCV Strategy 2020 – 2025. First, the roles of the MA and SA have been merged into a single TPM agent contract to simplify management and improve efficiency of delivery. Second, donor access to information has improved, in particular in regards to fiduciary monitoring. Third, the WBG and the current TPM agent continue to innovate through technology to improve quality and reduce cost. Fourth, environmental, social, and gender considerations are front and center in contracting and future reporting. Fifth, demonstrating value for money through concrete analysis and programmatic recommendations is a key priority. Sixth, a clear verification strategy – including heightened physical verification – and sampling plan is in place. Seventh, coordination between the WBG, the TPM agent, and GoIRA has improved and remains a key priority (e.g. 2-day workshop in Dubai during current TPM agent Inception Period in 2020). Finally, an enhanced capacity building approach has already been included in the new TPM agent contract.

The Review Team applauds the WBG and current TPM agent for taking on board these lessons and therefore will not (re)recommend the above. Instead, the Review Team will use this section to detail additional elements that – if implemented – should support the WBG in Afghanistan and teams operating elsewhere in FCV contexts to design, deliver, and produce useful TPM results.

***OBSERVATION: There are opportunities to share lessons learned and best practices with other large-scale TPM engagements in Afghanistan.***

The WBG should encourage increased dialogue with GoIRA and other funders of TPM to ensure consistency where possible and complementarity where desirable. The ARTF and the Asian Development Bank's (ADB) Afghanistan Infrastructure Trust Fund (AITF), for example, could share lessons on TPM services now that the ADB is developing its own TPM approach. Although distinct Funds, they both provide on-budget support for infrastructure projects and have a common goal of supporting the development of Afghanistan. Moreover, some of the same donors support both Funds and the ADB sits on the ARTF management committee. Communication and coordination are already happening and therefore can be extended to TPM. Similar opportunities exist for sharing lessons with the USAID Afghanistan Monitoring, Evaluation, and Learning Activity (AMELA), which provides TPM across much of the USAID portfolio in Afghanistan.

***OBSERVATION: The ARTF TPM arrangement employs a wide range of methodologies and generates data that has applicability beyond the individual projects monitored.***

The WBG should ensure that learning from TPM monitoring informs program design and portfolio management. TPM agents in other contexts have been engaged to conduct a wide variety of evaluations, beyond implementation monitoring, including data collection on program performance, processes, and impact in support of strategic objectives. For example, in Somalia the World Bank explicitly leverages data collected by the TPM for the Multi-Partner Trust Fund (MPF) to build systemic change. The USAID AMELA contractor is engaged to carry out performance, impact, and special evaluations to contribute to the achievement of results defined in the Mission's development strategy. In Yemen, the WBG-administered TPM has been expected to use qualitative methods to verify if each



intervention is achieving its development impact and to confirm/quantify levels of beneficiary satisfaction.

TPM program management should aim to ensure that lessons from the monitoring efforts are extracted and disseminated across the portfolio, and contribute to shaping the strategic direction of portfolio planning. This is in line with findings from the 2019 Country Portfolio Performance Review<sup>51</sup> which highlighted the need to expand engagement with TPM data. It also resonates with feedback provided by both WBG and donors interviewed for this Review. The WBG may wish to leverage one or more of the current ARTF forums (e.g. donor briefings, standing meetings, the Country Portfolio Performance Review process) to encourage reflection on TPM findings.

***OBSERVATION: WBG Task Team Leaders operated in silo, managing TPM activities for their respective programs in isolation from one another.***

Strengthening coordination and collaboration between WBG Task Team Leaders should be a priority for the next phase of contracting. Given the diversity of projects funded by the ARTF IW, it is to be expected that TTLs operate with a degree of autonomy. TTLs have a wealth of experience, knowledge, and ideas about managing and executing TPM. WBG should build an internal community of best practice, encouraging lesson sharing between TTLs. This would help to ensure a consistent level of engagement by TTLs with the TPM process, strengthen project handover between TTLs, and may also result in cost-efficiencies as TTLs identify cost-sharing approaches to data collection.

***OBSERVATION: The ARTF TPM engagement has continued to grow in scale each year with no clear exit strategy.***

Heightened insecurity in Afghanistan has meant that more and more verification activities have been tasked to ARTF TPM agents. The World Health Organisation's literature review on humanitarian program monitoring in conflict settings<sup>52</sup> highlights the danger of falling into the 'remote operations trap' through an over-reliance on TPM, which makes it difficult to revert to traditional management and oversight structures. As such, it encourages the commissioning party to develop a clear exit strategy. With peace talks between the GoIRA and Taliban underway, there is continued hope that security conditions may improve. Should this be the case, the WBG will need to review the scope and scale of current TPM activities. Though the available TPM literature provides little guidance, the WBG may wish to consider building an exit strategy for TPM, highlighting the conditions (security, government capacity etc.) or targets that could trigger a gradual scaling down of TPM operations.

***OBSERVATION: At a global level, the WBG does not have a TPM agent performance measurement framework or consolidated list of operational and delivery standards with which to measure TPM agent performance.***

Such a framework would strengthen mutual accountability between the WBG and the TPM agent, and would provide greater clarity to government partners and donors as to the roles, responsibilities (and indeed limitations) of the TPM agent. The World Bank is well positioned to develop operational standards for TPM based on the available TPM literature, and the Bank's own recent publication, *Demystifying TPM*, which contains detailed lessons learned from TPM experiences in FCV contexts.

---

<sup>51</sup> A Country Portfolio Performance Review (CPPR) is an annual review of the active projects in the portfolio conducted by the WBG and the client (e.g. GoIRA) to ensure quality delivery and to identify any portfolio wide issues that need to be addressed.

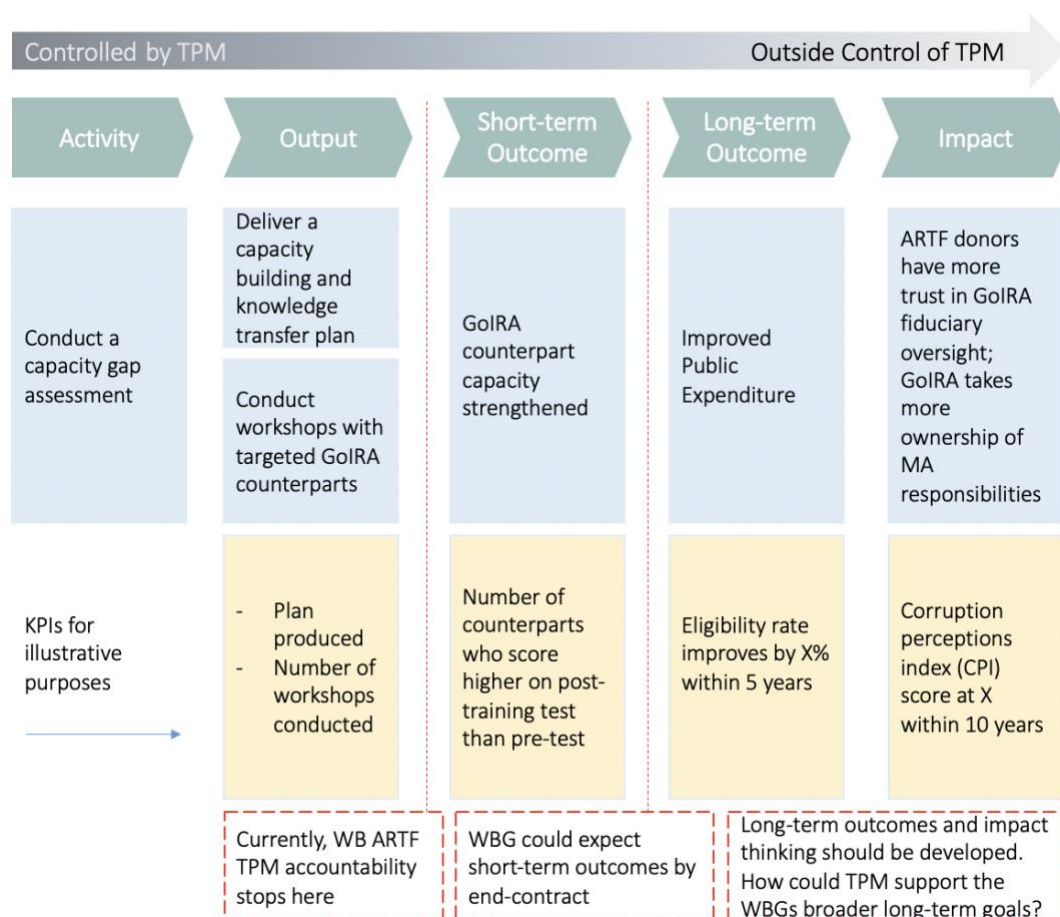
<sup>52</sup> WHO (2017): Humanitarian Programming and Monitoring in Inaccessible Conflict Settings: A Literature Review, p.9.



**OBSERVATION: Neither the SA nor the MA contracts contained outcome-level performance indicators.**

The contract for the MA contained only two key performance indicators (KPIs) at output level (see figure below). There were none in the contract for the SA. Outcome-level KPIs should be co-created with the WBG and the current TPM agent to: a) hold the TPM agent to outcome level results; b) set a path towards impact contributions; and c) have evidence to better demonstrate and communicate the value of TPM agents to donors.

Figure 4: Example of outcome-level performance indicators





## 8. Appendices

---

### 8.1. Appendix A: Detailed Methodology

We detail below the approaches and tools for this Review.

#### Research Framework

In order to review the performance of the ARTF TPM agents, the Review asked three overarching questions:

- i) Was the TPM approach well **designed**?
- ii) Was the TPM **delivered** to contractual obligations and quality standards?
- iii) Were the **results** useful?

These questions formed the basis of the research framework. The table below illustrates how the original research questions posed by the World Bank in the terms of reference for this body of work have been integrated into our research framework.

In answering these questions, specific lines of inquiry were pursued based extensively on a reading of available Bank literature on TPM (including *Demystifying Third-Party Monitoring in Fragile Situations: A Global Review*), a wider reading of TPM in FCV literature, and our own experience measuring the quality and impact of development programs.



Table 7: Research Framework

Supervisory Agent		
Research Pillar	Sub-Question	ToR Research Questions
Did the SA <b>design</b> an appropriate and adequate methodology to provide the contracted services?	<ul style="list-style-type: none"> <li>Did the SA design an appropriate and adequate methodology to provide the contracted services?</li> </ul>	<ul style="list-style-type: none"> <li>Did the contractors employ appropriate and adequate methodology in delivering contract services?</li> </ul>
Did the SA <b>deliver</b> these services in line with contractual obligations and quality standards?	<ul style="list-style-type: none"> <li>Did the SA deliver on services in line with the objectives and timelines set out in the relevant contracts?</li> <li>Did the SA deliver on reporting requirements in line with the objectives and timelines set out in the relevant contracts?</li> <li>Did the SA provide adequate data to enable WBG teams to make informed project supervision and implementation support decisions?</li> </ul>	<ul style="list-style-type: none"> <li>Did the contractors deliver on services and reporting in line with the objectives and timelines set out in the relevant contracts?</li> <li>Did the contractors employ appropriate, gender-sensitive strategies for reaching and engaging communities, including those in remote or hard-to-reach areas?</li> <li>Did the contractors respond to evolving WBG/GoIRA needs, including for adjusting TPM approaches to respond to earlier findings, in line with contract requirements?</li> <li>Did the contractors introduce innovations that may be relevant in the current phase of Afghanistan TPM, or for TPM programs elsewhere?</li> <li>Did the contractors effectively collaborate with relevant government stakeholders in monitoring processes, including follow-up, while maintaining appropriate independence?</li> <li>Did the contractors engage relevant community stakeholders in monitoring processes effectively?</li> <li>Were mechanisms established to follow-up on findings and recommendations effective?</li> <li>Were TPM data and findings of adequate quality, reliability, and relevance to the implementation support, fiduciary, and safeguards objectives of the contracts?</li> </ul>



		<ul style="list-style-type: none"> <li>• Were TPM findings and recommendations presented in a format that was usable to relevant WB and GoIRA teams?</li> <li>• Were TPM findings communicated in a manner that provided assurance to ARTF contributors regarding ‘value for money’ in ARTF programs?</li> <li>• Were TPM findings sex disaggregated, where applicable, to help ARTF stakeholders understand differential impact on men and women?</li> <li>• How were the contractors’ approaches similar to/distinct from other experiences of large-scale TPM described in research and literature on monitoring and implementation support in other fragile contexts?</li> </ul>
Were the <b>results</b> of the SA’s services useful to key stakeholder groups?	<ul style="list-style-type: none"> <li>• Did the SA’s services provide donors with assurance that ARTF projects were being implemented correctly?</li> <li>• Did the provision of SA’s services improve the government’s capacity to monitor and implement projects?</li> </ul>	<ul style="list-style-type: none"> <li>• Were TPM findings and recommendations accessible/understandable to ARTF stakeholders?</li> <li>• While contractual emphasis on capacity development evolved over time, did the contractors’ interaction with government, communities, and other stakeholders have impact on these stakeholders’</li> <li>• Capacity for monitoring the implementation of WBG-financed projects?</li> <li>• Capacity for adhering to fiduciary standards, including without limitation expenditure eligibility?</li> <li>• Capacity for adhering to environment and social safeguards standards, including without limitation gender requirements?</li> <li>• Were TPM findings and recommendations accessible/understandable to ARTF stakeholders?</li> </ul>
Monitoring Agent		
Research Pillar	Sub-Question	ToR Research Questions
Did the MA <b>design</b> an appropriate and adequate methodology to provide the contracted services?	<ul style="list-style-type: none"> <li>• Did the MA design an appropriate and adequate methodology to provide the contracted services?</li> </ul>	<ul style="list-style-type: none"> <li>• Did the contractors employ appropriate and adequate methodology in delivering contract services?</li> </ul>



<p>Did the MA <b>deliver</b> these services in line with contractual obligations and quality standards?</p>	<ul style="list-style-type: none"> <li>• Did the MA deliver services in line with the objectives set out in the relevant contracts?</li> <li>• Did the MA deliver on reporting requirements in line with the objectives and timelines set out in the relevant contracts?</li> </ul>	<ul style="list-style-type: none"> <li>• Did the contractors deliver on services and reporting in line with the objectives and timelines set out in the relevant contracts?</li> <li>• Did the contractors respond to evolving WBG/GolRA needs, including for adjusting TPM approaches to respond to earlier findings, in line with contract requirements?</li> <li>• Did the contractors effectively collaborate with relevant government stakeholders in monitoring processes, including follow-up, while maintaining appropriate independence?</li> <li>• Were mechanisms established to follow-up on findings and recommendations effective?</li> <li>• While contractual emphasis on capacity development evolved over time, did the contractors' interaction with government, communities, and other stakeholders have impact on these stakeholders' capacity for: <ul style="list-style-type: none"> <li>○ Monitoring the implementation of WB-financed projects?</li> <li>○ Adhering to fiduciary standards, including without limitation expenditure eligibility?</li> <li>○ Adhering to environment and social safeguards standards, including without limitation gender requirements?</li> </ul> </li> <li>• Were TPM data and findings of adequate quality, reliability, and relevance to the implementation support, fiduciary, and safeguards objectives of the contracts?</li> <li>• Were TPM findings and recommendations presented in a format that was usable to relevant WB and GolRA teams?</li> </ul>
<p>Were the <b>results</b> of the MA's services useful to key stakeholder groups?</p>	<ul style="list-style-type: none"> <li>• Did ARTF stakeholders derive value from the MA's findings?</li> </ul>	<ul style="list-style-type: none"> <li>• Were TPM findings and recommendations accessible/understandable to ARTF stakeholders?</li> <li>• Were TPM findings communicated in a manner that provided assurance to ARTF contributors regarding 'value for money' in ARTF programs?</li> <li>• Did TPM services, findings, and recommendations support enhanced 'value for money' in the ARTF-financed portfolio in line with contract objectives?</li> </ul>



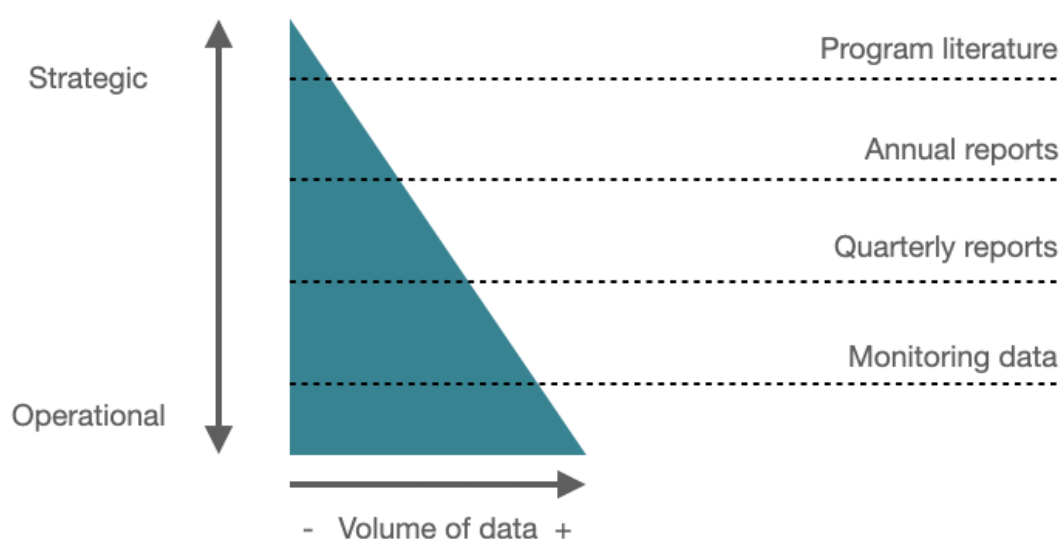


## Literature Review

Aleph Strategies conducted a literature review in two parts: i) a review of ARTF TPM agent documentation; and ii) a review of open-source TPM in FCV literature.

**ARTF TPM Document Review** - Aleph Strategies conducted a thorough review of program literature for the SA and MA components of the ARTF TPM. There was a large volume of information available for this Review ranging from the strategic to the operational. Given the high-level nature of this project, it was not possible to review this material comprehensively, nor would this be instructive. Instead, we prioritized the comparatively small number of mid- to strategic-level documents (e.g. contracts, completion reports, annual reviews) to provide a solid contextualized understanding of the World Bank's TPM approaches and standards, and contractual obligations of the SA and MA.

Figure 5: Volume of available information for the Review



Subsequent sampling was then a combination of purposive and randomized selection, directed principally by the key informant interviews. As is often the case with this type of research, stakeholders recommended specific reports or datasets for further investigation. For example, based on discussions, we conducted a deep-dive on particular project ToRs, quarterly reports, briefing notes, site visit reports, and monitoring data itself. If our spot-checking returned compliance issues, we investigated further, drawing on additional datasets (for example, ARDEA), follow-up interviews, and reports as necessary.

**TPM in FCV Literature Review** - Aleph Strategies also conducted a review of available open-source literature on third-party monitoring in states affected by fragility, conflict, and violence. A review of these documents helped to calibrate the research framework, and provided deeper contextual insight for determining whether and to what extent the ARTF TPM agents employed approaches that were similar to/distinct from other TPM arrangements in FCV contexts.

A select bibliography in Appendix C illustrates the range of documentation reviewed as part of this Report.

## Key informant interviews

Aleph undertook a broad range of stakeholder interviews during the data collection phase. All interviews were conducted via video-conference or telephone. Interviews were semi-structured, comprising a series of open-ended questions, permitting a free-flowing discussion. This approach allowed unexpected angles and unanticipated insight to emerge, and provided Aleph with the flexibility to pursue new lines of inquiry as they developed over the course of the discussion. Each interview



lasted approximately 1 hour and was conducted in English. There were multiple types of participants in these discussions as both ‘political’ and ‘technical’ views needed to be included.

Aleph consulted with 58 individuals. The WBG facilitated introductions with all key stakeholders to outline the purpose and objective of the Review and to secure the necessary permissions to interview certain stakeholder groups. The Table below details the key stakeholder groups, organizations, and number of individuals consulted. Personal information shall remain confidential.

*Table 8: Key Informant Interviews*

Stakeholder Group	Organization / Government / Ministry	Number of interviewees
Current and Former ARTF Donors	USAID, FCDO, EU, Norway, Canada	5
Current and Former Afghanistan CMU	World Bank Group	8
Current and Former Project Task Teams	World Bank Group: NSP, EQUIP, EQRA, OFWMP, WEE-RDP, CCAP	11
Environmental and Social Safeguards Teams	World Bank Group	5
Monitoring Agent	BDO LLP	6
Supervisory Agent	Management Systems International	6
GoIRA	MoF, MAIL, MRRD	10
External TPM Experts	Abyrint, ADB, consultants	7
<b>TOTAL</b>		<b>58</b>

Given time and budgetary constraints, Aleph could not consult with each of the 15 WBG project Task Teams and GoIRA counterparts. Instead, we selected 6 projects as seen in the Table above based on: a) number of site visits conducted; b) sectoral spread (education, agriculture, strengthening local institutions, etc.); c) gender focus (e.g. WEE-RDP to deep-dive on female enumerator approaches); d) contractual spread (projects that were monitored early in the contracting period, in the middle and in later years); and e) availability of key stakeholders<sup>53</sup>. GoIRA counterparts were then targeted based on the projects and WBG task teams interviewed. For example, MAIL was interviewed in relation to OFWMP; MRRD in relation to CCAP, NSP and WEE-RDP; and MoF in relation to the MA. Despite best efforts, representatives of the MoE were unavailable during the timeframe of this Review.

<sup>53</sup> As noted in the Limitations section: Some targeted interviewees did not respond to a request for interview or were unavailable during the timeframe of this Review. In a context such as Afghanistan, turnover can be high and therefore many targeted interviewees were no longer working in the positions they held at the time of project implementation. Interviews were conducted on a best-effort basis within the timeframe of the Review.



## 8.2. Appendix B: Status Quo of MA Recommendations

Table 9: CBKT Recommendations by MA

Capacity Development Indicators	MA Basis for Gap Selection	MA Strategy	Implemented by ARTF and/or GoIRA?
Supporting Documentation Missing or not provided (CD1)	Institutional: Ineligibilities are currently not taken into account in benchmarking the Incentive Program element of ARTF funding. Accordingly, in the absence of any other mechanisms, GoIRA is not significantly incentivized to reduce ineligibilities. Correspondingly, there is a lack of appropriate government policies/actions to encourage compliance, thus increasing the likelihood of fraud and other irregularities as a result of impunity (people feel they can get away with it). There does not appear to be any consequences for noncompliance. It is generally held that the non-provision of supporting documentation is, in some of the cases, a deliberate attempt to hide fraud/irregularities. However, this view is not supported by concrete substantive evidence.	Inclusion of ineligibilities in the IP benchmarking process; Comprehensive government policy to address /discourage ineligibilities, setting out measures in the event of noncompliance, and providing incentives for compliance; Setting up of a coordinating mechanism/Work Group to provide high level political impetus and coordinate efforts at addressing ineligibilities.	No.
	Technology: Technology does not appear to play a key role in explaining the ineligibilities. However, striving for a paperless environment will, in the long term, address the matter in a more efficient manner.	Long-term strategy to move towards Egovernment and paperless office environment.	No.
	Human resource management: While members of staff are generally aware of the requirement to retain documentary records, highlighting the severity and implications of noncompliance will provide impetus for change. In the longer term it would be possible for the provision of documentation for review to be built into a performance management approach for civil servants. This would be a long-term reform and require a wider HRM and cultural transformation in Afghanistan.	Highlighting of the importance and implication of retaining supporting documentation at training courses and in MOF Circulars.	Yes. Training courses conducted.



Noncompliance with procurement requirements (CD2)	Institutional: At a meeting with the Procurement Policy Director, Office of the President (National Procurement Authority), the issue was raised that while staff members were generally aware of procurement requirements, they were less willing to comply due to the benefits derived from noncompliance. The institutional gap highlighted above (i.e. absence of ARTF/GolRA policy incentive to encourage compliance), coupled with this self-interest, partly explains the ineligibility.	Inclusion of ineligibilities in the IP benchmarking process; Comprehensive government policy to address/discourage ineligibilities, setting out punitive measures in the event of noncompliance, and providing incentives for compliance; Setting up of a coordinating mechanism (Work Group) to provide high level political impetus and coordinate efforts at addressing ineligibilities.	Partially. Working Group established but ineligibilities were not included in the IP benchmarking.
	Systems & Processes: There are inadequate preventative controls in place, such as checklists to facilitate compliance with the procurement requirements.	Introduction of a Procurement Checklist incorporating compliance requirements should strengthen the control environment and bring it into alignment with procurement requirements.	Yes. Procurement Checklist workshop conducted and Checklist adopted by MoF.
	Technology: The procurement process is not automated and procurement, purchases, and payment processes are not integrated. The use of technology could reduce human errors, increase compliance rates by automating compliance, and increase efficiency.	National E-Procurement, and its integration with purchases and payment, should be considered as part of a long-term strategy to increase compliance rates and improve efficiency.	No.
	Human resource management: There is a knowledge gap in the requirements of the procurement law, and this has been compounded by recent changes to the procurement legislation. We understand that training in the new requirements of over 1,200 ministry staff has been conducted by the NPA, but there still remains a significant number of procurement personnel that have not received training. A harmonised approach to training is not in place.	Provision of additional resources to fast-track staff training in the new procurement requirements; Coordinating and/or harmonisation of procurement training activities by NPA & MA; Use of NPA's extensive training resources, at both central and provincial levels, for MA training.	Yes. Training courses were conducted.



Signature for receipt by employees of cash salaries or bank transfers reports not available (CD13)	<p>Institutional: The current ARTF institutional framework does not provide incentives to GoIRA to reduce ineligibilities as already discussed above. This, combined with the inherent fraud risks associated with cash handling, increases the likelihood that the ineligibilities indicate actual occurrence of fraud. Banking in rural areas is poorly developed, with the effect that a significant number of teachers are paid by cash. A solution being considered is M-Paisa, a mobile money transfer services offered by Roshan, the mobile phone operator. We understood that the process is slow and facing various regulatory challenges.</p>	<p>Inclusion of ineligibilities in the IP benchmarking process for ARTF funding; Regulatory fast-tracking of M-Paisa money transfer in rural areas; Introduction of policies demanding greater accountability from bonded salary trustees (e.g. demanding reimbursement for failure to produce the acknowledgment of cash receipt); Influencing of banks, through policy and incentives, to open more branches in rural areas.</p>	<p>Partially, as the WBG finances a project known as PAISA (Payments Automation and Integration of Salaries in Afghanistan).</p>
	<p>Systems &amp; Processes: The ineligibility primarily relates to missing (or non-provision of) acknowledgment of receipt for cash salary payments to teachers, and it is also understood that apart from the MoE, all other ministries pay overtime and allowances in cash even in situations where bank payment facilities exist. There is no integrated, national, IT payroll system at the ministries. Development of SQL/Access/Web-based systems to process payroll has been in-house and we note that formal system development processes and controls are not being applied.</p>	<p>Payment of overtime and allowances via bank as opposed to cash; Application of system development processes and controls to the systems under development; Development of a national, integrated, payroll system.</p>	<p>Partially. Payroll workshops and checklists developed. The payroll web-based system is implemented and available to be used by most of the entities in center and provinces but it is not mandatory. There are still some departments which are processing excel based payrolls for payments. So, we cannot consider this as national integrated system.</p>



	Technology: Banking and mobile technology is absent in rural areas as a solution to cash salary payments, though we understand mobile banking is being considered as noted above.	Development, harnessing and/or facilitation of technologies to reduce/eliminate cash salary payments in rural areas.	Partially, as the WBG finances a project known as PAISA (Payments Automation and Integration of Salaries in Afghanistan).
	Human resource management: There are gaps in the knowledge of inhouse Systems/IT personnel maintaining and developing the SQL system, and often we note there is one staff member carrying out system development functions, resulting in lack of accountability and/or appropriate segregation.	Recruitment of an International Payroll System Developer to spearhead the national development process, and development of a human resource strategy to address system development gaps and challenges.	Partially, as the WBG finances a project known as PAISA (Payments Automation and Integration of Salaries in Afghanistan).
Miscoding of expenditure (CDI 4)	This primarily relates to the miscoding of military expenses as eligible ARTF recurring expenditure. We understand that a budgetary code is not set up for military spending and inadequate guidance is provided on how these cost categories should be treated.	Setting up of budgetary codes for military expenditure that is not mapped to ARTF funding, and release of an MOF circular providing additional guidance to staff.	No. As per the Government Chart of Account, there are object codes for military expenditures and the spending entities can use should there is fund allotted under those codes. The issue of miscoding arises when a military expense is recorded under civilian code due to the lack of allotments and proper authorization has not been taken from MoF in advance. A proper guidance still needs to be in place in this regard.



Noncompliance with internal control policies and procedures (CDI5)	Systems and processes: This primarily relates to noncompliance with expense approval limits on delegated authority. The present system is not set up to flag delegated authority breaches as exceptions that should be addressed before payments are processed.	Development of a detect-and-stop system (e.g. via checklists) to address breaches in the internal control policies.	No.
	Institutional/Human resource management: There is a general unwillingness to comply with internal control policies where this conflicts with the self-interest of employees as discussed above, and staff members lack appropriate incentives to comply. In addition, there is some lack of awareness among some staff members, particularly in the provinces, of expense approval limits and delegated authority.	Inclusion of ineligibilities in IP benchmarking, in addition to increasing awareness of approval limits among staff via appropriate training and Circular distribution.	No.
Noncompliance with fiduciary requirements (CDI6)	This primarily relates to noncompliance with the requirement to book expenses within three months for ministries, and four months for provinces. GoIRA's systems are not set up to track and monitor compliance with this requirement, and to flag exceptions as breaches that should be rectified before payments can be processed. We also understand from talking to various ministries that the benchmarks are difficult to achieve in the Afghanistan context due to technological and logistical challenges.	Establishing a track-and-monitor system to ensure compliance with the fiduciary requirement. Revision of the benchmarks to take into consideration the challenges faced in Afghanistan.	Addressed through workshops and no longer a frequent breach.





### 8.3. Appendix C: Select Bibliography

#### ARTF TPM Document Review

Ardea platform - extensive review of platform data

ARTF (2018): Partnership Framework and Financing Program

ARTF (2019): Fiduciary Control Framework for World Bank-Financed Projects

ARTF (2019, 2020): ACREMAP Quarterly Reports

ARTF (July 2020): Administrator's Report on Financial Status

ARTF Scorecard (2018) Integrated Performance and Management Framework

ARTF Scorecard (2019) Integrated Performance and Management Framework

ARTF Website: artf.af for content, reports, and financial data

BDO (2016): ARTF MA Capacity Building and Knowledge Transfer Plan

BDO (2016): ARTF MA Inception Report

BDO (2016): Special Report on Additional Payroll Testing

BDO (2016, 2017, 2018, 2019): ARTF MA Monthly Reports

BDO (2016, 2017, 2018, 2019): ARTF MA Quarterly Reports

BDO (2017): ARTF MA Summary Annual Report (FY1395)

BDO (2018): ARTF MA Summary Annual Report (FY1396)

BDO (2018, 2019): ARTF MA Briefing Notes

BDO (2019): ARTF MA Summary Annual Report (FY1397)

BDO (2019): MA for ARTF Review of Internal Controls ARAP (MPW)

BDO (2019): MA for ARTF Review of Internal Controls ARAP (MRRD)

BDO (2019): MA for ARTF Review of Internal Controls CCAP (IDLG)

BDO (2019): MA for ARTF Review of Internal Controls CCAP (MRRD)

BDO (2019): MA for ARTF Review of Internal Controls HEDP (MOHE)

BDO (2019): MA for ARTF Review of Internal Controls NHRP (DABS)

BDO (2020): ARTF Monitoring Agent Final Financial Monitoring Report



BDO (2020): ARTF Recurrent Cost Window Detailed Annual Report (FY1398)

BDO (2020): TPMA Inception Report

Contract 8005414 SA TOR (Apr 19-Dec 19)

Contract 8005414 SA TOR (Nov 15-Oct 18 extended to Mar 19)

Contract 8005616 MA TOR (Jun16-Aug18)

Contract 8005616 MA TOR (May19-Dec19) - modification E

Contract 8005616 MA TOR (Sept 18-Apr 19)

Contract 8006437 Combined TPMA TOR (Jan 20-Dec 21)

ESF Questionnaires - randomized selection

MSI (2016): ARTF TPM Program Year I Annual Report

MSI (2017a): ARTF Year II Annual Report

MSI (2017b): CCAP Field Monitoring Form - Subprojects Transport Sector

MSI (2018): ARTF Year III Annual Report

MSI (2018b): Management Systems International Year III Work Plan ARTF Supervisory Agent - Third Party Monitoring

MSI (2019): Completion Report Afghanistan Reconstruction Trust Fund II

MSI (2019): Year IV ARTF SA Annual Report November 2018 - November 2019

MSI (2019a): ARTF Year IV Annual Report

MSI (2019b): ARTF II Final Site Visit Listing

MSI (2019c): ARTF II Completion Report

MSI Monthly Reports - randomized selection on quarterly basis

MSI Quarterly Reports - randomized selection of projects each quarter

Oyeyiola, A. (2016): Review of ToR and Scope of Work of the ARTF MA

SA Deviation Tracker Sheet (EQUIP-CM)

Scanteam (2008): Afghanistan Reconstruction Trust Fund External Evaluation

Scanteam (2012): ARTF at a Cross-Roads: History and the Future

Scanteam (2017): External Review ARTF, Taking Charge Government Ownership in a Complex Context



SIGAR 18 - 42 Summary of Recommendations, Responses and Actions

World Bank (2017): Third Party Monitoring ToR for Citizen's Charter Afghanistan Project (P160567)

World Bank (2017b): Third Party Monitoring ToR for Citizens' Charter Afghanistan Project (P160567)

World Bank (2019): Third Party Monitoring Terms of Reference (CCAP)

World Bank (2019b): Areas of Enhanced TPM Support (CCAP)

World Bank (2020): Implementation Support for the New Normal

World Bank Group (2020): Terms of reference for the Audit of the Bank's Supervision of Investment Project Financing (IPF) Projects in countries experiencing fragility, conflict and violence (FCV). Office Memorandum

World Bank South Asia Region, Operational Services and Quality Department (2014): Third Party Monitoring Programme for the Afghanistan Reconstruction Trust Fund: A Review

### TPM in FCV Literature Review

Aleph Strategies (2019): Mid Term Review of the Multi Partner Fund for Somalia

DFID (2014): ICTs for Monitoring and Evaluation of Peacebuilding Programmes

DFID (2015): No Longer a Last Resort: A Review of the Remote Programming Landscape

EU (2020): Study on Best Practices in Third Party Monitoring

GPPI (2016): Technologies for monitoring in insecure environments. Secure Access in Volatile Environments

GSDRC (2013): Remote management of projects in fragile states

GSDRC (2018): Approaches to remote monitoring in fragile stages

ICRC: Third-Party Monitoring: Desk Review and Implementation Guidelines

Integrity (2015): Cross-Cutting Evaluation of DFID's Approach to Remote Management in Somalia and North-East Kenya

SAVE (2016): The Use of Third-Party Monitoring in Insecure Contexts

USAID (2019): Afghanistan Monitoring, Evaluation, and Learning Activity (AMELA) service contract

USAID (2020): What are the ethical implications of using Third-Party Monitoring?: <https://usaidlearninglab.org/lab-notes/what-are-ethical-implications-using-third-party-monitoring>

WFP (2016): Internal Audit of TPM in WFP

WFP (2016): Monitoring humanitarian assistance in conflict-affected settings



World Bank (2017): Third Party Monitoring in Volatile Environments – Do the Benefits Outweigh the Risks?

World Bank (2019): Demystifying Third-Party Monitoring in Fragile Situations: A Global Review

World Bank Group Strategy for Fragility, Conflict and Violence (2020 - 2025)