

IMF Fiscal Reforms in Southeast Europe Program: Phase III (SEE III)

Mid-term external evaluation
Period: January 2023 to April 2025

Evaluation Report
August 2025

ACKNOWLEDGEMENTS

[To be drafted]

CONTENTS

CONTENTS	III
ABBREVIATIONS AND ACRONYMS	IV
EXECUTIVE SUMMARY	1
1. INTRODUCTION	4
1.1 PROGRAM OVERVIEW.....	4
1.2 SUMMARY OF EVALUATION OBJECTIVES, APPROACH AND METHODOLOGY.....	5
2. AGGREGATED PROGRAM-LEVEL ASSESSMENT RATINGS	6
2.1 OVERVIEW.....	6
3. ASSESSMENT OF RA WORKSTREAM PROJECTS	8
3.1 KEY ACHIEVEMENTS.....	8
3.2 WORKSTREAM-LEVEL ASSESSMENT.....	8
3.3 COHERENCE.....	9
3.4 EFFECTIVENESS	9
3.5 EFFICIENCY	10
3.6 IMPACT	11
3.7 SUSTAINABILITY	12
3.8 COMMUNITIES OF PRACTICE	12
4. ASSESSMENT OF PFM WORKSTREAM PROJECTS	15
4.1 KEY ACHIEVEMENTS.....	15
4.2 WORKSTREAM-LEVEL ASSESSMENT.....	16
4.3 COHERENCE.....	17
4.4 EFFECTIVENESS	18
4.5 EFFICIENCY	18
4.6 IMPACT	19
4.7 SUSTAINABILITY	21
5. ASSESSMENT OF TP WORKSTREAM PROJECTS	23
5.1 KEY ACHIEVEMENTS.....	23
5.2 WORKSTREAM-LEVEL ASSESSMENT.....	23
5.3 COHERENCE.....	24
5.4 EFFECTIVENESS	24
5.5 EFFICIENCY	25
5.6 IMPACT	25
5.7 SUSTAINABILITY	26
6. REVIEW OF WORKSHOPS AND WEBINARS	28
7. REVIEW OF AFE GOVERNANCE AND OPERATIONS MANAGEMENT	29
8. FINDINGS AND RECOMMENDATIONS	34

ABBREVIATIONS AND ACRONYMS

BRFP	Briefing paper
CD	Capacity development
CEF	IMF Common Evaluation Framework, also Centre of Excellence in Finance
CIT	Corporate income tax
CoP	Community of Practice
C-PIMA	Climate PIMA
CRM	Compliance risk management
DP	Development partner, or development cooperation agencies
EC/EU	European Commission/European Union
ESC	Evaluation Sub-Committee
EUD	EU Delegation
EUR	IMF European Department
FAD	IMF Fiscal Affairs Department
FR	Fiscal risk
FRAT	Fiscal risk Assessment Tool
FRD	FR disclosure
FRM	FR management
FRS	FR statement
FRU	FR unit
FRWG	FR working group
FTE	Fiscal Transparency Evaluation
Fund	IMF
FY	IMF fiscal year
G&O	Governance and other
HCT	IMF Health Check Tool
HQ	IMF Headquarters
IFMIS	Integrated Financial Management and Information System
IMF	International Monetary Fund
JVI	Joint Vienna Institute
KEQ	Key Evaluation Questions for the SEE III evaluation
LTX	Long-term, regional or resident expert
MDA	Ministries, Departments, and Agencies
MoF	Ministry of Finance
MTEF	Medium-term expenditure framework
MTRS	Medium-term Revenue Strategy
OBL	Organic Budget Law
OECD	Organisation for Economic Cooperation and Development
OECD-DAC	OECD Development Assistance Committee
PD	SEE III program document
PEFA	Public Expenditure and Financial Accountability
PIM	Public investment management

PIMA	PIM Assessment
PIMIS	PIM Information System
PIMU	PIM unit
PIT	Personal income tax
PM	SEE III Project Manager
PM&A	Project management and administration
POA	TADAT Performance Outcome Area
PPP	Public-private partnership
Program	SEE III
Project	RBM country/objective
RBM	Results-Based Management
RMO	Reform management office
SC	SEE III Steering Committee
SCO	Swiss Cooperation Office
SECO	Swiss State Secretariat for Economic Affairs
SEE	ALB, BiH, MKD, MNE, UVK and SER countries
SEE III	Fiscal Reform in Southeast Europe Phase III program
SOE	State-owned enterprise
STX	Short-term expert
TA	Technical assistance
TADAT	Tax Administration Diagnostic Assessment Tool
TAR	TA report
TEA	<i>tax expenditure analysis</i>
TEAR	TEA and reporting
ToR	Terms of reference for the SEE III evaluation
TPU	Tax policy unit
VAT	Value Added Tax
W&C	Workshops and conferences
SEE III workstreams	
PFM	Public Financial Management
RA	Revenue Administration
TP	Tax Policy
SEE III Beneficiary Countries	
ALB	Albania
BiH	Bosnia and Herzegovina
MKD	North Macedonia
MNE	Montenegro
SER	Serbia
UVK	Kosovo

EXECUTIVE SUMMARY

The Fiscal Reform in Southeast Europe Program Phase III (SEE III, or program), implemented by IMF, and externally funded by the European Union (EU) and the Swiss State Secretariat for Economic Affairs (SECO), supports Revenue Administration (RA), Public Financial Management (PFM) and Tax Policy (TP)¹ reforms in six countries² in the Western Balkans (SEE countries). The program addresses, through capacity development (CD) support, the need in these countries for having efficient and effective fiscal institutions; mobilising revenue to finance expenditure policies; and implementing essential structural reforms. The current third program phase began in January 2023 and is expected to conclude in December 2026, and has a budget of US\$16.6m. SEE III is intended to contribute to the EU accession process and help align beneficiary country RA, PFM systems and tax policies and related capacity with international good practices. The program is closely integrated with the IMF surveillance and lending process.

This report sets out the findings and recommendations from a mid-term evaluation of SEE III, undertaken to assess the program's overall performance in achieving its mandate, and to formulate recommendations for improvement. The evaluation has been conducted through assessment, by applying five OECD-Development Assistance Committee (DAC) evaluation criteria³, of a sample of SEE III CD projects. The sample has been grouped into one country and one topical case study for each workstream, and a selection of project desktop reviews. Assessment has been undertaken through document reviews, interviews and surveys. The IMF proposed scoring system⁴ has been applied to determine aggregated scores for the three workstreams and the program. The review covers SEE III CD activities between January 2023 and February 2025.

The overall evaluation rating of all three workstreams and the program is between Largely Achieved and Partially Achieved⁵, and the RA and PFM workstreams were rated higher than in the previous phase on key criteria. The assessment reflects a strong performance of the program, especially in criteria such as efficiency, impact and sustainability, given the generally uncondusive reform environment in SEE countries. Compared to the previous phase, the RA and PFM workstreams were rated higher at the overall workstream level, and for these three OECD-DAC evaluation criteria, reflecting the benefits of effective CD delivery, and increased maturity of counterpart institutions and establishment of legislative frameworks and processes initiated earlier in the program.

A number of strategic initiatives in the current phase have contributed to the strong results, building on the proven CD design and delivery approaches from previous phases. These include the fielding regionally-stationed experts (LTXs) for in-person missions; focusing resources on important topical areas in each of the workstreams to implement a regional approach to providing CD; increased emphasis on peer-to-peer learning, including the introduction of structured communities of practice in the RA workstream, and fiscal risk peer exchange and European Infrastructure Governance network for PFM; and increased focus on capacity building to promote sustainability. These approaches have reinforced the program's strengths from the previous phase, which included the designing of interventions based on detailed diagnostics; providing very high-quality technical advice; harnessing multiple CD delivery channels that complement each other, including HQ missions, LTXs, STXs, workshops and access to courses. Partnerships with the Joint Vienna Institute and the Center for Excellence in Finance supported the latter, and also helped locate LTXs in the region.

¹ The Tax Policy workstream was introduced to the program in the current phase.

² Albania, Bosnia and Herzegovina, Kosovo, North Macedonia, Montenegro, and Serbia

³ Organization for Economic Cooperation and Development, Development Assistance Committee evaluation criteria evaluated being coherence, effectiveness, efficiency, impact and sustainability. Relevance was excluded from assessment as it was rated excellent during SEE II evaluation and IMF (in consultation with EU and SECO) did not expect any change for Phase III.

⁴ Based on responses to key evaluation questions for each evaluation criterion in the evaluation terms of reference, scores and qualitative gradings are awarded as: 4 = Fully Achieved; 3 = Largely Achieved; 2 = Partially Achieved; and 1 = Not Achieved, against objective and outcome targets for the current phase.

⁵ Ratings based on rounding decimal places to nearest whole number.

The overall performance of the RA workstream has been driven by a combination of strong reform design, close alignment with country priorities, and effective long-term engagement. Key positive drivers include the use of embedded LTXs, which ensured continuous support tailored to institutional context; the establishment of reform management structures that enhanced coordination and accountability; and the strong technical quality of CD inputs, including on CRM and IT modernization. The workstream's responsiveness to country-specific needs, through tools like TADAT and MTRS, also enabled well-sequenced reforms. On the downside, progress was sometimes constrained by political will, turnover of staff, capacity gaps in IT implementation, and uneven reform ownership at the mid-management level in some administrations. In a few countries, the pace of institutional change was also slowed by delays in decision-making and reform fatigue following externally driven projects.

The SEE III PFM workstream demonstrates significant achievements across the main program CD areas. With respect to the two focus CD areas: in PIM, SEE III has built on the baseline information provided by the PIMAs undertaken under SEE II and the more recent C-PIMAs. In fiscal risk (FR), SEE III has achieved a broadening and deepening of fiscal risk management in three important dimensions: i) the setting up of institutional frameworks has progressed significantly, with many beneficiaries having put in place a fiscal risk unit/working groups; ii) beneficiaries apply in a risk-focused manner and routinely update analytical tools, especially for SOEs; and iii) disclosure is progressing substantively, with most beneficiaries producing fiscal risk statements either as self-standing output or as chapters in a Fiscal Strategy, with a number also publishing them. Contextual factors adversely affecting CD performance have been the lingering effect of the budgetary disturbance caused by COVID-19, and the difficult global economic context including accelerating inflation and rising real global interest rates. Internal factors have included serious capacity constraints in several countries in relation to deepening of required analysis for PIM and FR.

TP workstream projects have produced various important achievements within the first two years of the workstream's integration into the SEE III program, successfully capitalising on earlier IMF HQ CD. Institutional capacity has been built or enhanced in almost all SEE III countries, through the establishment or structuring of tax policy units, and the training and use of TP tools, such as simulation modelling. Tax Expenditure Analysis (TEA) activities have generated TEA Reports in several countries, at a first iteration or supporting future iterations. In some cases, TEAs have been institutionalized with annual updates, assigned responsibilities within the MoF, and integrated into reporting of budget documentation. CD on implementing TP tools across all tax types has been provided in a range of areas, including VAT, corporate income tax, personal income tax and immovable property tax. Some of the targeted outcomes for the workstream are more susceptible to changes in political sentiment, whilst others to capacity limitations. The dual focus on capacity building and policy advice provides the workstream with a stable foundation.

Whilst SEE III program management and reporting is generally good, some areas merit consideration for further strengthening. The recommendations from the evaluation, presented below, are intended to address these:

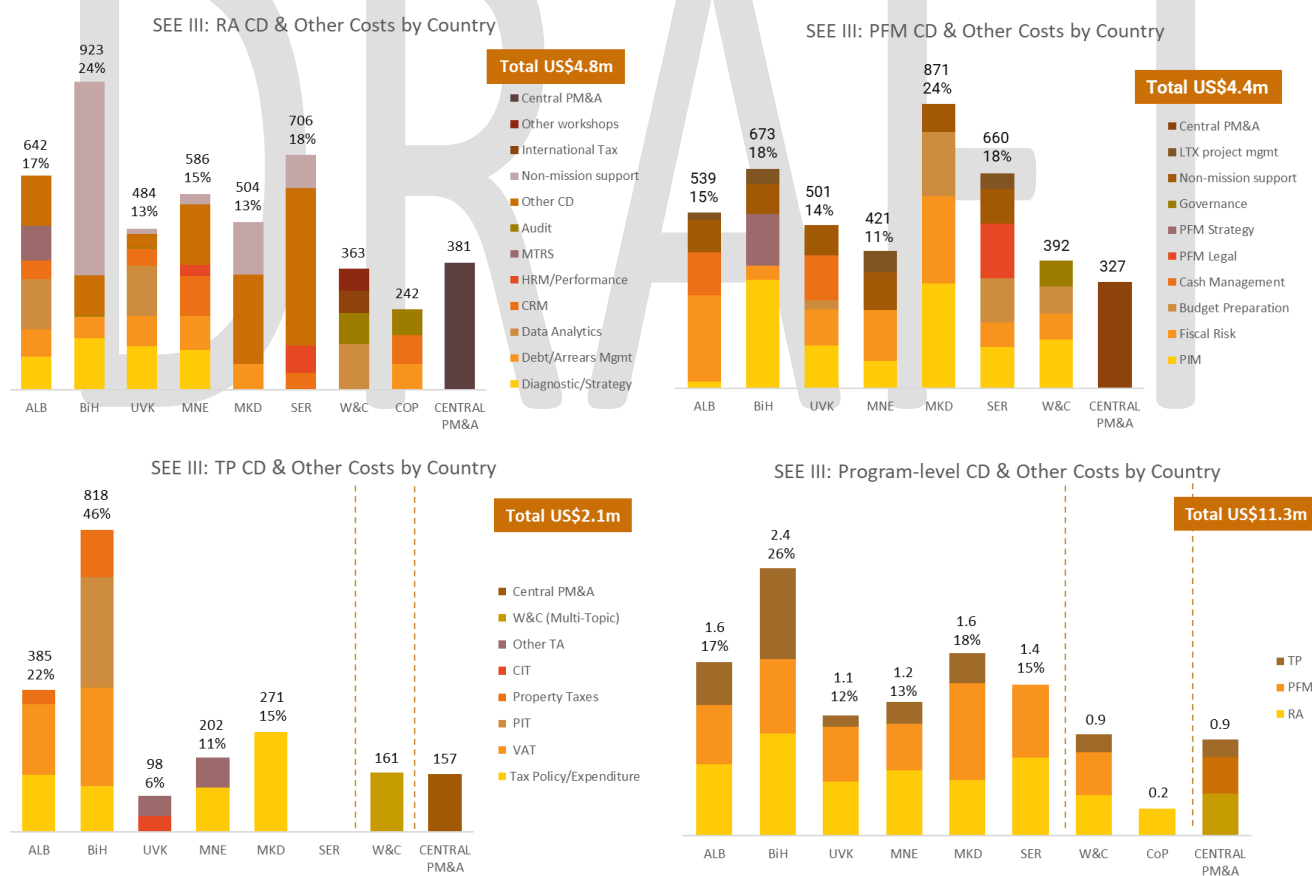
1. Improve the use of the RBM logframe system within its current constraints to provide more accurate, timely, consistent and actionable information. This would principally involve more diligent and consistent use of milestones; systematically rating all indicators; and ensuring consistency between workstreams in recording information.
2. IMF should consider updating its annual progress reports to incorporate more information linking current period activity with achievements targeted for the phase, and bring out more regional-level and cross-workstream issues to the extent not already provided. The logframe, which should be better linked to the contents of the main report, would preferably include information about inputs, activities, outputs, outcomes and targeted impacts/objectives. Current focus on country-level information should be extended to the program as a whole. The reporting process should be automated.
3. IMF should consider engaging a project coordinator to support the SEE III technical team, to strengthen project administration and management, and release time of specialists to focus on providing CD advice.
4. IMF should minimise the levels of costs in CDMAP unallocated to specific CD projects, or project management and administration. This would be to address historically large balances of such costs in the activity reports.

5. IMF should consider in its dialogue with relevant counterparts further highlight the negative role of inadequate human resource allocation and poor HR practices as a major barrier to reforms and CD effectiveness. Where possible formal commitment to address this problem should be secured. Allocation of CD should generally more closely reflect adequacy of resourcing and potential for achieving results.
6. In developing a strategy for the rest of the current phase within the available funding envelope, IMF and donors should consider both the impact on current CD projects, and the implications of leaving any significant gap, or less than a critical level of activity, on any future program.

1. INTRODUCTION

1.1 PROGRAM OVERVIEW

The Fiscal Reform in Southeast Europe Phase III (SEE III, or program) is a capacity development program and multi-partner trust fund established by the International Monetary Fund (IMF, or the Fund). It supports Revenue Administration (RA), Public Financial Management (PFM) and Tax Policy (TP)⁶ reforms in Albania (ALB), Bosnia and Herzegovina (BiH), Kosovo (UVK), North Macedonia (MKD), Montenegro (MNE), and Serbia (SER), (hereafter referred to as SEE countries). The program addresses, through capacity development (CD) support, the need in these countries for having efficient and effective fiscal institutions, which are essential to deliver macro-economic stability and sustainable growth; mobilise revenue to finance expenditure policies; and implement essential structural reforms. The European Union (EU) and Swiss State Secretariat for Economic Affairs (SECO) are external donors. The current third program phase began in January 2023 and is expected to conclude in December 2026. The SEE III Program Document (PD) provides for a four-year budget of US\$16.6m, of which US\$15.5m is allocated to CD activities⁷, the balance accounting for IMF Trust Fund Management Fee.



Source: IMF data, evaluator analysis. Figures in US\$'000. Jan 2023 - Feb 2025

Central Project Management and Administration (PM&A) includes project management, cost of Annual Coordination and Steering Committee Meetings, and Mid-Term Evaluation. We believe it does not include PM&A costs incurred by LTXs

Non-mission support reflects LTX costs in delivering CD and (for RA) project management, unallocated to specific projects or PM&A
For RA, Other CD includes PIT, HQ CD delivery, multi-project missions, VAT, IT, investigation, LTO, BPR, and Organisation

⁶ The Tax Policy workstream was introduced to the program in the current phase.

⁷ Between RA, PFM and TP at a ratio of 40:40:20 respectively

1.2 SUMMARY OF EVALUATION OBJECTIVES, APPROACH AND METHODOLOGY

This mid-term evaluation has been conducted primarily for learning purposes. It aims to formulate recommendations that inform and further strengthen future CD, including SEE III's possible top-up. The evaluation terms of reference (ToR) set out its objectives and methodology, and this was updated in the inception report. A bottom-up project-based approach for assessing the CD performance has been taken, with a focus on outcome and impact levels. Other key evaluation objectives are: follow-up on the implementation of recommendations from the mid-term evaluation of the previous phase; providing insights on SEE III CD contribution to structural reform in the Western Balkans, including alignment with the EU New Growth Plan; efficacy of the program's regional approach; and assessment of the opportunity to scale up the program. CD project assessments and ratings has been undertaken through a detailed review of a sample; with reference to five⁸ of the six OECD-DAC evaluation criteria; and eight⁹ Key Evaluation Questions (KEQ). The IMF Common Evaluation Framework (CEF) and its accompanying Guidance Note have provided the framework for Project-level assessment, with the Fund's Results Based Management (RBM) logframe providing information on indicators for assessing effectiveness. The evaluation has also reviewed SEE III governance and operations management with reference to four additional ToR questions, although these have not been to be rated. During inception the Evaluation Sub-Committee (ESC) also suggested focus on particular areas, including program management, fund management and 'burn rate', logframe reporting, and impact.

The assessments presented in this report are based on the review of a sample of SEE III CD Projects¹⁰ as of February 2025. These include one country and one topical case study for each of the three workstreams, and five desktop reviews each for RA and PFM, and two for TP. These case studies and desktop reviews covered 37 Projects: 13 for RA, 15 for PFM, and nine for TP. The combination of case studies and desk assessment Projects was selected based on a range of criteria, with material weighting to: significant resource deployment; geographic and topical distribution of the workstream portfolios; and, importantly, projects where success and CD-delivery challenge-related lessons could be learnt.

A range of evaluation tools has been deployed to undertake CD project assessments. Extensive documents have been researched, including those related to SEE III governance, management and reporting; CD projects and associated RBM logframe data; strategies, policies and related reports produced by CD counterparts; relevant IMF policies and surveillance and other reports; and background information about the case study countries. Interviews have been undertaken with FAD staff and experts working on SEE III; relevant IMF European Department (EUR) mission chiefs; IMF Resident Representatives in case study countries; Steering Committee (SC) members and local EU Delegation (EUD) and Swiss Cooperation Office (SCO) staff in case study countries; and other Development Partners (DP). CD counterparts and external stakeholders were interviewed during country case studies visits. A total of 90 individuals were interviewed separately or as part of groups. The evaluation also relied on surveys of SEE III CD counterparts; IMF staff and experts; DPs, including SC members and EUD and SCO staff; CoP participants; and workshop participants. For the SEE III CD counterparts survey, 71 invitations were sent out, eliciting 40 responses (response rate of 56%); for the IMF survey, the corresponding numbers were 66, 40 and 61%; for the DP survey, 33, 19 and 58%; CoP survey numbers were 65, 35 and 54%; and workshop participants, 169, 33 and 20%.

Projects have been rated and scored for each of the five assessed OECD-DAC criteria based on an approach guided by the ToR. Each criterion has been awarded a qualitative grading ('rating') with a numeric 'score' of between 4 and 1 based on an assessment of the responses to the relevant KEQ. These are, 4: Fully achieved; 3: Largely achieved (more than 75% to 99% of target); 2: Partially achieved (26-75%); and 1: Not achieved (less than 25%). Achievement is in relation to targeted RBM logframe objectives (proxied through RBM outcomes and their selected indicators) for effectiveness, and a considered realistic optimum in relation to other DAC criteria. Scores have been averaged on an unweighted basis to arrive at aggregates for each Project, workstream and the Program. The aggregated scores are calculated to one decimal place. Achievements from CD delivered prior to the TP workstream joining the program have been reflected in the ratings.

⁸ Coherence, effectiveness, efficiency, impact and sustainability.

⁹ The original ToR provided for 10 KEQs, and these were consolidated into eight at inception.

¹⁰ Defined as RBM logframe country/objectives

2. AGGREGATED PROGRAM-LEVEL ASSESSMENT RATINGS

2.1 OVERVIEW

The aggregate rating for the assessed OECD-DAC criteria at the SEE III program level is between **Partially Achieved and Largely Achieved**¹². The RA and PFM workstreams have similar average score, although there are some variations attributed to individual OECD-DAC evaluation criteria between them. The TP workstream, new to the program, scores marginally lower, for reasons explained later in this section. The overall score reflects a strong performance of the program, especially in criteria such as efficiency, impact and sustainability, given the generally uncondusive reform environment in SEE countries. The aggregated OECD-DAC ratings reflect an improvement from the previous phase in key evaluation criteria. The average score for RA and PFM workstreams was 2.5 in the previous phase, compared with 2.7 in Phase III.

OECD-DAC AGGREGATED SCORES ¹¹ FOR SEE III WORKSTREAMS AND PROGRAM				
	RA	PFM	TP	SEE III
Coherence (COH)	2.8	2.8	2.8	2.8
Effectiveness (EFT)	2.5	2.5	2.3	2.5
Efficiency (EFF)	3.0	2.5	2.6	2.7
Impact (IMP)	2.6	2.5	1.9	2.4
Sustainability (SUS)	2.7	2.4	2.3	2.5
Overall	2.7	2.6	2.4	2.6

As with the previous phase, a number of common external factors negatively affected the evaluation ratings for the SEE III workstreams, albeit to different extents. Although the adverse effects of COVID-19, which had an important impact on CD delivery in Phase II, were significantly reduced, persistent factors affecting SEE countries in varying degrees included: i) the political economy of the countries, some of which saw a number of changes in government, delaying policy decisions and disrupting implementation of CD in ministries and agencies; ii) resource and capacity constraints at CD counterpart institutions; iii) high staff turnover, at both junior implementation and senior decision-making levels, often linked with political changes; iv) in some countries, reform fatigue, with many years of changes with the aim of EU Accession, combined with in some countries a loss of motivation (exacerbated by the potential fast tracks being given to Ukraine and Moldova), albeit EU Accession continues to be a key objective of economic reforms; and v) weak governance and high levels of corruption, also affecting political decision-making; limited or no involvement by the authorities in promoting donor coordination. In the DP survey undertaken by the evaluation, 16/19 respondents felt achieving results from CD in the region was ‘Quite difficult’.¹³

Nevertheless, recent economic developments have concentrated authorities’ minds on implementing reforms. These include the continuing economic pressures ensuing from the COVID-19 crisis; negative global economic developments; and donor funding reductions to different degrees. As a result, SEE governments are more focused on implementing reforms that generate revenues and better managing their spending. More effective, sustainable and impactful CD projects would be expected to ensue as a result.

¹¹ For each Project (defined as RBM logframe objective/in a country) assessed, an average of the OECD-DAC criteria scores has been calculated to provide an overall project-level score (based on the methodology in the footnote below). The scores for all Projects within each workstream (for each OECD-DAC criterion and overall project scores) were averaged to provide a set of workstream-level scores. This exercise was repeated for all Projects to arrive at the program-level scores. Finally, the criterion score for all projects were averaged to arrive at the assessed portfolio-level scorings and ratings for each workstream. Simple averaging has been applied for aggregating all scores. Average scores have been rounded up or down to the nearest integer to attribute a qualitative rating.

¹² Evaluation ratings are, 4 = Fully Achieved; 3 = Largely Achieved; 2 = Partially Achieved; and 1 = Not Achieved.

¹³ The other three responded ‘Quite easy’. No responses on ‘Very easy’ or ‘Very difficult’

The drawbacks were at least partially mitigated by IMF CD delivery approach through SEE III. These were principally: i) RA and PFM workstreams built their work programs on strong foundations laid in the previous phase, while TP had benefited from continuity of earlier HQ CD delivery in selected countries; ii) the work programs were designed robustly, guided by detailed diagnostics such as the Tax Administration Diagnostic Assessment Tool (TADAT) (for RA) and PIMA (for PFM); iii) very high technical quality of advice was delivered by IMF, engendering confidence in counterparts and willingness to translate CD support to action, constraints permitting; iv) a combination of IMF HQ missions, LTX, STX and regional workshops complementing each other. Additionally, major developments in the current phase were: i) regionally based LTXs, bringing benefits such as greater understanding of counterparts, more frequent CD missions, and ability to provide support between missions (the previous phase also had LTXs, although fewer in number, and they could for the majority of the duration only provide remote support due to COVID-19); more strategic concentration on a few cross-cutting topics, especially in PFM; introduction of communities of practice (CoP) in the RA workstream, which has progressed significantly in some topics on promoting and facilitating peer-to-peer learning, and generally adoption of more peer-to-peer CD delivery approaches in the other two workstreams.

The presence of regionally based LTXs has had a major positive impact on CD delivery across the workstreams, while HQ staff continued to be valued for their technical expertise. LTXs have a stronger grasp of local context, enabling them to tailor advice to country-specific circumstances. This was evident both from the results of evaluation project assessments, and feedback from a range of counterparts. The CD counterpart survey respondents¹⁴ highlighted the top three benefits of LTXs as being: i) more effective implementation of CD advice (27%); ii) that they understand issues and capabilities and therefore can give more pertinent advice (20%); and iii) they help counterparts better understand issues (18%). Interestingly, the benefit of being able to implement advice quicker came only fourth, with 17.5%, suggesting that there may be a natural pace of implementation in many SEE III counterpart institutions that cannot be exceeded with more CD resources. This latter point was further supported by a different question on demand for CD, where half the respondents indicated they would like more CD, with the other half saying same or less. Nevertheless, the evaluation has not been able to establish the additional benefits compared with costs of deploying three vs say two LTXs per workstream. Although LTXs have played a major role in the program's successes, the complementing advice of HQ continues to be highly valued by CD counterparts, especially for the quality of their technical expertise. Their survey rated the effectiveness of HQ missions the highest, with almost 80% considering them to be 'highly effective' (74% thought LTXs were highly effective).

¹⁴ Based on 40 responses

3. ASSESSMENT OF RA WORKSTREAM PROJECTS

The findings for the RA workstream presented in this section are based on assessment of a sample of i) one country case study covering four projects on: data analytics, debt management, diagnostic and strategy, and CRM; ii) a CRM topical case study covering CD projects four countries and the CRM CoP; and iii) five desktop reviews in five countries on: personal income tax (PIT) and Medium-term revenue strategy (MTRS), debt and arrears management, diagnostic and strategy, reform, and organisation and performance management. 13 Projects were assessed, covering 38% of RA country-specific CD spend (a relatively low coverage compared with the other two workstreams, reflecting the more distributed nature of the RA CD portfolio).

3.1 KEY ACHIEVEMENTS FROM SAMPLED PROJECTS

The RA workstream has delivered measurable progress in strengthening core capacities, institutional resilience, and reform momentum across the SEE countries. A defining achievement has been the establishment and operationalization of reform management structures, such as Reform Management Offices (RMO), which have improved internal coordination, reform tracking, and alignment between tax policy and administration. These structures, now institutionalized in most countries, have provided a much-needed anchor for reform continuity in politically fluid environments.

In CRM, SEE III CD support has been both strategic and catalytic. All countries now have functioning CRM frameworks, with several establishing dedicated CRM units and applying structured risk segmentation to high-risk sectors. Peer learning through the CRM CoP, combined with benchmarking tools such as TADAT, has driven a shared regional understanding of good practices and helped translate diagnostic findings into actionable reform priorities. Progress in digitalization and analytics has also been a standout theme. SEE III CD supported the rollout of new or upgraded IT systems, often integrated with improved data analytics, which has strengthened both enforcement and taxpayer services. In one country, for instance, digital tools have started enabling more risk-based audit selection, while in another, data integration initiatives were shaped around CRM needs.

The workstream has also contributed to improved donor coordination and reform coherence, with LTXs actively engaging in multi-partner forums to reduce fragmentation and align efforts. The ability to tailor CD support to national reform priorities, through for instance collaboration between TP and RA utilising MTRS as one example, has helped sustain momentum even amid political turnover or institutional resistance. While workforce constraints and uneven political buy-in remain ongoing challenges, the workstream's achievements reflect a strategic shift: from *ad hoc* interventions to longer-term, system-building reforms. The cumulative effect has been a more reform-minded, data-driven, and outcome-focused revenue administration landscape across the region.

3.2 WORKSTREAM-LEVEL ASSESSMENT

The overall performance of the RA workstream has been driven by a combination of strong reform design, close alignment with country priorities, and effective long-term engagement. Key positive drivers include the use of embedded LTXs, which ensured continuous support tailored to institutional context; the establishment of reform management structures that enhanced coordination and accountability; and the strong technical quality of CD inputs, including on CRM and IT modernization. The workstream's responsiveness to country-specific needs, through tools like TADAT and MTRS, also enabled well-sequenced reforms.

RA WORKSTREAM		
AVERAGE		
2.7		
COH	EFT	EFF
2.8	2.5	3.0
IMP		SUS
2.6		2.7

On the downside, progress was sometimes constrained by political will, turnover of staff, gaps in IT implementation, and uneven reform ownership at the mid-management level in some administrations. In a few countries, the pace of institutional change was also slowed by delays in decision-making and reform fatigue following externally driven projects. At the country and topical levels, CRM emerged as a strong area of progress across most countries, largely due to the clarity of reform objectives, technical soundness of advice, and regional peer learning through the CRM CoP. The development of CRM frameworks, establishment of specialized units, and improved data use were consistently rated as high-performing components. Reform management structures contributed positively in settings where they were fully embedded but had less traction in countries where leadership commitment was less stable. IT modernization showed mixed results: progress was notable where reforms were integrated into broader strategies and supported by strong project management; but in some countries, ambitious IT agendas faced delays due to procurement hurdles or lack of technical capacity. Overall, variation in project ratings across countries reflected the interplay between institutional readiness, leadership stability, and the presence of well-sequenced reform plans underpinned by robust diagnostics.

3.3 COHERENCE

The RA workstream demonstrated strong internal coherence, with CD activities well-aligned to IMF policy advice and diagnostic tools such as TADAT. In some countries reform plans were refined based on these assessments, with priorities feeding directly into capacity development efforts. In one country, the MTRS process served as a key integrator, linking policy objectives with administrative reforms. Coordination between IMF surveillance, CD, and lending ensured that SEE III support was embedded within broader macro-fiscal and PFM frameworks. LTXs played a crucial role in connecting CD with HQ guidance and in addressing emerging issues like staffing, IT risks, and compliance gaps. 11 of the 13 Projects were assessed as Largely Achieved for coherence in RA, with two achieving Partially Achieved. The latter reflected areas where institutional instability, e.g. leadership changes or coordination challenges, affected the consistency and synergy between IMF CD activities and national reform priorities. In such cases, CD interventions, though still relevant, were less uniformly delivered and achieved limited impact.

RA COHERENCE
AVERAGE
2.8

External coherence was generally strong, supported by close coordination with DPs including the World Bank, SECO, and the EU. IMF LTXs facilitated regular collaboration with donor-led initiatives and joint participation in tax policy and PFM platforms, reducing duplication and enhancing complementarity. In some countries for example, IMF-supported CRM reforms were aligned with World Bank IT modernization efforts, while SECO-funded projects built on IMF diagnostics. Some challenges arose when donors independently pursued IT projects without full integration into tax reform priorities. However, IMF engagement in planning often improved coherence. Overall, the workstream showed a pragmatic, proactive approach to donor coordination, fostering more harmonized region-wide reform implementation.

3.4 EFFECTIVENESS

The effectiveness of RA CD interventions is evident in the advancement of key institutional reforms, strengthened organizational capacities, and improved operational practices across multiple areas. Notable achievements include the establishment and enhancement of RMOs, the development of CRM frameworks, and improvements in audit and collection methodologies. These efforts have fostered more systematic, results-oriented approaches to revenue administration, with significant progress in strategic planning, reform coordination, and the adoption of risk-based methodologies. CD initiatives such as training, technical assistance, and workshops have been well-received, contributing to positive shifts in mindset among senior management and technical staff, who demonstrate increased ownership of modernization initiatives. Seven of the 13 RA Projects assessed for effectiveness were rated Partially Achieved, while six were assessed as Largely Achieved. Common themes among the Partially Achieved ratings, most of which were attributed to desktop reviewed Projects, included diagnostic frameworks that had been introduced but not yet fully resourced: for instance, compliance programs that lacked detailed treatment plans, or compliance committees that met infrequently, limiting the realization of full outcomes despite evident

RA EFFECTIVENESS
AVERAGE
2.5

structural progress. Conversely, the Projects achieving a "Largely Achieved" rating for effectiveness were characterized by strong institutional ownership, consistent leadership and governance, integrated approaches (combining TA, training, and institutional support), and adaptation of CD to local contexts. CD counterpart survey responses highlight internal institutional constraints, such as staff turnover and limited resources, as key bottlenecks to fully realizing the effectiveness of capacity development. 15 out of 24 RA respondents identified these factors as significant obstacles to implementing CD advice.

However, the pace and extent of achieving intended outcomes vary across countries and institutions due to persistent challenges. Political interference, limited human resources, and organizational resistance to change have, at times, constrained the full realization of reform objectives and slowed improvements in revenue performance and administrative efficiency. External factors, including economic shocks and shifting government priorities, have further complicated the operating environment, making it difficult to attribute broader macro-level outcomes solely to CD efforts. Despite these constraints, the interventions have laid important groundwork and generated momentum for ongoing reform. Beneficiary feedback from interviews indicates that training and advisory support have positively influenced daily operations and reform implementation, while emphasizing the continued need for support, particularly in change management and addressing institutional bottlenecks, to ensure sustained results.

3.5 EFFICIENCY

CD inputs were generally well-aligned with expected outputs and outcomes, achieving results in a cost-effective manner relative to feasible alternatives.

The deployment of LTXs based in the region proved especially efficient, despite being relatively costly, leveraging their local presence and contextual knowledge to provide timely, tailored support. This modality enabled quicker feedback loops and adaptive management, enhancing the relevance and impact of interventions compared to more centralized fly-in and fly-out models. Operational delivery maintained good-quality outputs and consistent backstopping, ensuring SEE III CD remained responsive to evolving country needs. The mix of in-country and remote support helped optimize resource use, balancing intensive on-site engagement with cost and time efficiencies. While timeliness was generally satisfactory, occasional delays occurred due to external factors such as political dynamics or coordination challenges with other partners. All 13 Projects assessed were rated Largely Achieved for efficiency, and this was the highest rated OECD DAC criterion for the workstream. This rating can be attributed to several factors: the strategic deployment of LTXs, whose regional presence allowed for continuous engagement and stronger relationships with local counterparts, ensuring interventions were closely aligned with national priorities. Integrated delivery modalities, combining in-country and remote support, provided flexibility and enabled timely responses to emerging needs while optimizing resource allocation. Adaptive management practices ensured interventions remained relevant and effective, even in the face of unforeseen challenges. These elements collectively contributed to the efficient delivery of CD interventions, ensuring that resources were utilized effectively to achieve the desired outcomes.

RA EFFICIENCY
AVERAGE
3.0

Key lessons emphasize the importance of robust coordination among all stakeholders to avoid duplication and enhance alignment. Realistic planning that anticipates potential delays, coupled with continuous monitoring and feedback mechanisms, proved critical in identifying challenges early and adapting delivery approaches promptly. These practices have been instrumental in maintaining operational efficiency and sustaining momentum throughout the CD program.

While comprehensive cost-effectiveness analyses were not systematically conducted across all interventions, evidence suggests that CD support positively impacts tax revenues over time. Additionally, long-term CD interventions, such as those involving resident experts, have been shown to significantly and positively increase tax revenues, underscoring the importance of sustained engagement. These findings highlight the value of investing in CD initiatives, as they contribute to long-term revenue mobilization and institutional strengthening.

3.6 IMPACT

SEE III RA IMPACT

IMPROVED TADAT ASSESSMENT SCORES IN SEE III COUNTRIES 2015-2023

The cumulative impact of IMF-supported capacity development activities in the SEE III countries since 2015 is reflected in measurable improvements across most dimensions of tax administration performance as evidenced by TADAT assessments conducted between 2015 and 2023. Analysis of countries with multiple assessment cycles shows a consistent upward trend in scores across the majority of the nine Performance Outcome Areas (POAs). Specifically, seven out of nine POAs demonstrated net score increases, indicating sustained progress over time, while only one POA showed no change and one experienced a slight decline. Notably, areas such as timely filing (POA-4), efficient revenue management (POA-8), and supporting voluntary compliance (POA-3) showed the strongest gains, with several indicators improving from moderate (C) to good (B) or high (A) ratings. However, some indicators experienced fluctuations: for example, a few high ratings in areas like taxpayer registration (POA-1) or audit effectiveness (POA-9) were lowered in later assessments, reflecting challenges in sustaining certain gains. Similarly, while some countries made significant progress in ensuring accurate and timely filing, others faced setbacks in selected indicators related to arrears management or dispute resolution (POA-5). Newer indicators introduced in recent assessments, such as risk-based audit selection under POA-6, digitalization of taxpayer services under POA-5, and enhanced tax arrears management, have generally shown limited progress, underscoring the evolving challenges in these emerging focus areas. Overall, the trajectory of improvements across multiple TADAT dimensions underscores the positive and sustained impact of the IMF’s phased and adaptive CD interventions in the region.

Progress in revenue administration often hinges on enhancing ICT systems and processes, an area where many, including those in the SEE countries, have faced challenges. Among the highest-performing entities, as indicated by TADAT assessments, one achieved this through a partial upgrade of its CRM systems, while another developed its own software in-house. Maintaining a previously achieved grade is as challenging as earning it initially; therefore, sustaining top-tier ratings (A and B) represents a significant accomplishment. Notably, several entities have experienced declines across the POA ranges, likely due to reasons given above and due to updates or redefinitions of POAs, with the last update occurring in 2019. While other sections may highlight improvements in timely filing and payment, the critical factor appears to be the timing of the last review, which dates to 2022. This suggests that observed TADAT assessment changes may reflect the outcomes of SEE Phases I and II rather than Phase III.

SEE III CD initiatives have contributed positively to strengthening macroeconomic and financial stability by supporting key reforms in tax administration, debt management, and CRM.

These reforms have enhanced revenue collection effectiveness and fiscal sustainability, both vital for maintaining macroeconomic balance. While direct measurement of impacts on indicators like GDP growth or tax-to-GDP ratios is complicated by external shocks such as economic downturns and the COVID-19 pandemic, the foundational reforms, particularly in audit, CRM, and debt management have set the stage for long-term inclusive growth by fostering fairer tax systems and improved taxpayer engagement. Tax revenue has remained broadly stable, with the region recording a tax-to-GDP ratio of around 31.2 % in 2022, which recovered to approximately 31.6 % in 2023, reversing a slight dip from the pandemic period. Countries with stronger CD support saw higher ratios, with top performers reaching between 36 and 38%, suggesting that CD in tax administration, audit, and risk management likely contributed to more effective revenue mobilization.¹⁵ In the CD counterpart survey, a large majority agreed that the RA workstream contributed meaningfully to institutional strengthening over the long term [18/24], and many viewed it as contributing to broader economic development [16/24]. The majority of RA Projects, eight out of 13, were rated Largely Achieved for impact, the remaining Partially Achieved.

RA IMPACT
AVERAGE
2.6

¹⁵ Western Balkans Competitiveness Outlook 2024: Regional Profile, OECD; and International Monetary Fund, Government Finance Statistics Yearbook and data files, and World Bank and OECD GDP estimates. Tax revenue (% of GDP). World Bank

Nevertheless, the pace of impact contribution has been uneven, with some reforms experiencing delays or partial implementation due to political interference, capacity constraints, and legacy system challenges. Lessons learned highlight the critical importance of sustained high-level ownership and political commitment to maintain reform momentum. Additionally, adaptive approaches that respond to emerging challenges, alongside alignment of CD efforts with broader macroeconomic policies and DP programs, are key to achieving meaningful and lasting impacts. Conversely, political interference often led to shifts in reform priorities or delays in decision-making processes, hindering timely implementation. Capacity constraints within institutions further resulted in challenges in managing and sustaining reforms effectively. Additionally, legacy system challenges, such as outdated technologies and processes, impeded the adoption of new practices and tools. These obstacles were particularly evident in areas requiring significant institutional change, such as tax administration and CRM. Despite these challenges, the initiatives achieved positive outcomes in areas where reforms were implemented effectively, demonstrating the potential for sustained impact when these barriers are addressed.

3.7 SUSTAINABILITY

The sustainability of CD benefits across the RA workstream presents a mixed but cautiously optimistic picture. Institutional reforms and improved processes, such as the establishment of RMOs, clearer strategic roadmaps, and adoption of modern IT systems, have laid a foundation for lasting improvements. Strong high-level buy-in and alignment with national priorities, including commitments to regional integration and donor frameworks, further support the medium- to long-term preservation of benefits. This has led the majority of the objectives rated, nine out of 13, being assessed as Largely Achieved.

RA
SUSTAINABILITY
AVERAGE
2.7

However, sustainability risks remain due to challenges such as political interference, frequent turnover in senior management, old and ineffective ICT systems, aging workforces without robust succession planning, and limited internal reform mindsets. These factors threaten to erode reform momentum if not carefully managed. Feedback from interviews with LTXs and STXs highlights that while agencies recognize the need for reform, ownership and acceptance tend to develop gradually, underscoring the importance of sustained leadership engagement. Coordination with DPs and maintaining a supportive external environment are also critical to reinforcing gains. Without ongoing CD delivery, particularly in change management, ICT modernization, and human resource development, the net benefits risk diminishing over time. Tailored follow-up interventions focused on institutionalizing reforms and strengthening internal capacities will be essential to secure lasting impact.

3.8 COMMUNITIES OF PRACTICE

The CoPs have successfully fostered valuable networks for practitioners to connect beyond formal training events, encouraging cross-country learning and the sharing of good practices adapted to SEE countries’ context. Survey feedback highlights CoP participants’ appreciation for the practical focus of discussions and the opportunity to benchmark progress and challenges with peers. Notably, the CRM CoP received consistently high praise for its practical orientation, peer learning, and effectiveness in helping participants apply risk management methodologies, contributing to tangible improvements across the region. Despite these strengths, challenges persist in maintaining sustained engagement and consistent participation. Variations in institutional capacities, reform priorities, and resource constraints, alongside competing demands on officials’ time, limit the regularity and applicability of CoP activities. Practical barriers such as delays in critical system implementations, differing national regulations, human resource shortages, and evolving legal frameworks further complicate the translation of CoP lessons into practice. Strengthening facilitation and follow-up mechanisms remains important to better convert discussions into concrete, context-sensitive actions. Lack of participants’ authority implement reforms has also been highlighted as a challenge.

Many improvement efforts are already in progress and these initiatives could be developed further. Building on these, enhancing facilitation with consistent leadership and expanding interactive formats like case study clinics can deepen practical learning. Improving follow-up with clearer accountability and progress tracking will better ensure the translation of insights into action. Tailoring content more precisely to diverse institutional contexts and leveraging asynchronous collaboration tools can accommodate participants' varying needs and schedules. Strengthening connections between CoPs and national reform agendas would also support ownership and sustainability. Ongoing feedback and evaluation will help the CoPs adapt effectively to emerging challenges. Consideration might be given to partnering with CEF, which has a long track record and experience of establishing similar CoPs, and could contribute some of the 'softer' aspects of building such networks sustainably.

The CoPs participants survey elicited additional improvement suggestions. These included allowing countries to propose topics for IMF prioritization and focusing seminars on practical case studies addressing regional challenges; hosting seminars in a rotational format across Western Balkan countries was recommended to facilitate direct observation of tax administration practices; and the exchange and adaptation of operational guidelines between administrations, along with more study visits to countries with advanced risk management practices. Overall, participants expressed a strong desire for CoP activities to be even more practical, actionable, and closely aligned with the real-world challenges faced by tax officials.

Assessment of the CoPs against OECD DAC criteria indicates that they are highly relevant to regional priorities and effective in fostering knowledge exchange and peer support, although measuring direct reform outcomes is more challenging. CoP activities have generally been cost-efficient through virtual platforms and regional workshops, but intermittent participation and coordination gaps sometimes reduce overall efficiency. Moreover, survey participants indicate preference for in-person contact. While the CoPs contribute to capacity building and innovation diffusion, greater emphasis on linking outputs to institutional changes would enhance impact. Ensuring long-term sustainability will require embedding CoPs within formal frameworks and securing dedicated resources.

The CoPs logframe requires strengthening to reflect more closely the objectives of these networks. At present, there is a single logframe for all three CoPs, which is more appropriate for a stand-alone workshop. The single targeted outcome is 'The participants value this training' with feedback scores being the indicator, and the milestones relate to number of events held. Given the wider ambitions of CoPs, a logframe with more robust outcome indicators and milestones is needed, to reflect defined SMART objectives of the initiative.

KEY LESSONS AND RECOMMENDATIONS FROM THE EVALUATION OF SEE III RA PROJECTS

- **Evidence-based diagnostics:** Across SEE countries, IMF CD interventions were consistently informed by rigorous diagnostic tools such as TADAT and IMF surveillance reports, ensuring tailored and context-specific support. Embedding tools like TADAT and fostering regional communities of practice enhance knowledge sharing and continuity of reform efforts.
- **Institutional reform support:** The successful creation and operationalization of Reform Management Offices in multiple countries provided institutional anchors for reform coordination and monitoring.
- **Coordinated modernization:** IT modernization and digitalization efforts advanced in coordination with partners like the World Bank, SECO, and the EU, avoiding duplication and improving data analytics and taxpayer services.
- **Strengthening political buy-in:** Persistent political interference and frequent changes in senior management remain a challenge across several countries, undermining reform continuity and momentum. Enhanced efforts to secure sustained high-level ownership are needed.
- **Better coordination on IT projects:** While modernization efforts are progressing, occasional misalignment and fragmentation have occurred when donors implemented IT-heavy projects independently. Earlier and closer coordination among all stakeholders is essential to maximize synergies.
- **Reform resistance is real and requires robust change management:** CD support alone is insufficient. Embedding comprehensive change management strategies, including stakeholder engagement, communication, and capacity building, is essential to overcome entrenched inertia and resistance and to ensure capacity gains translate into sustained reforms.
- **Human resource constraints are critical bottlenecks:** Aging workforces, skill gaps, and poor succession planning threaten reform sustainability. Dedicated investments in HR policies, targeted recruitment, retention incentives, and systematic succession planning are non-negotiable for long-term impact.
- **Invest in non-traditional CD areas:** Allocate resources to areas beyond conventional technical assistance, such as organizational culture, communication strategies, and human resource development, to address deeper reform challenges. This will require more proactive coordination with DPs.
- **Clear, realistic, and collaboratively developed milestones are crucial:** Milestones must be jointly agreed with all key stakeholders to ensure buy-in and ownership. Overambitious or poorly defined milestones lead to delays and frustration, while well-structured milestones enable effective tracking, motivation, and timely course correction.
- **Sustained engagement in complex topical areas is needed:** Areas like compliance risk management and IT/digital systems require ongoing support due to their technical complexity and the evolving nature of reforms, emphasizing the importance of medium- to long-term CD strategies.
- **Increase program flexibility:** Design adaptive interventions with continuous monitoring and feedback loops to respond to evolving political and economic contexts.
- **Digital transformation must be holistic:** Investment in new technologies should be matched with legal, procedural, and human capacity reforms; otherwise, digital initiatives risk becoming underutilized or unsustainable.

4. ASSESSMENT OF PFM WORKSTREAM PROJECTS

The findings for the PFM workstream presented in this section are based on assessment of a sample of i) one country case study covering five projects: FR, PIM information systems (PIMIS), Climate PIM assessment (C-PIMA), Organic budget law (OBL), and budget preparation; ii) a FR disclosure (FRD) topical case study¹⁶ covering CD projects in five SEE countries¹⁷; and iii) five desktop reviews in five countries (on: i) PFM strategy, ii) PIMA, iii) cash management, iv) PIM, and v) SOE FRD). 15 RBM logframe Projects were assessed. The sampled Projects cover 55% of PFM country-specific CD spend.

4.1 KEY ACHIEVEMENTS FROM SAMPLED PROJECTS

The SEE III PFM workstream demonstrates significant achievements across the main CD areas¹⁸. Notable achievements include:

- **Under PIM**, SEE III has built on the baseline information provided by the PIMAs undertaken under SEE II and the more recent C-PIMAs. Building on the recommendations provided by these assessments, good progress has been made in advancing the work of the PIM units (PIMU), or equivalent organisational entities, in strengthening the database available and the analysis applied to support more effective public investment.
- **Fiscal Risks:** SEE III has achieved a broadening and deepening of FR management in three important dimensions: i) setting up of institutional frameworks has progressed significantly, with five beneficiaries having put in place a FR unit (FRU)/working group in the last two years; ii) beneficiaries apply in a risk-focused manner and routinely update analytical tools, especially for SOEs. Currently, FAD's SOE Health Check Tool (HCT) is applied by four beneficiaries targeting the largest SOEs; iii) disclosure is progressing substantively, with eight beneficiaries producing FR statements (FRS) either as self-standing output or as chapters in a Fiscal Strategy, with six also publishing them; this progress is achieved in an area where, by nature, disclosure keeps being sensitive due to potential market repercussions. The advances in disclosure have been accompanied with an extension of the scope of FR analysis to include SOEs, public-private partnerships (PPP), macroeconomic risk, budget implementation related risks, as well as legal and climate risks.
- **Budget structure and management systems:** Progress in respect of reform through the introduction or strengthening of the multi-annual budget through the medium-term expenditure frameworks (MTEF) as a necessary condition for enhancing the treatment in the budget of public investment programs.
- **The country case study highlights progress in a difficult CD implementation environment.** The country experiences: i) very low baseline in terms of the capacity in government (specifically MoF) to manage the more technical aspects of fiscal policy and budgetary management; ii) lingering impact of COVID-19 and ensuing adverse global events including rising global real interest rates and inflation; and iii) impact of elections held in May 2024, which effectively put a stop to reform work for nine months. In spite of these challenges, SEE III CD has made important progress in several areas: i) a plan has been prepared for aligning the antiquated and ineffective budget management system to meet the requirements of the new OBL; ii) a start has been made in developing PIM capability, which has been hitherto fragmented with little or no central control on the quantity or composition of public investment; and iii) progress has been made in initiating understanding of the main issues surrounding FR management (FRM), in terms of data collection, analysis and eventually reporting, with a FRS being prepared for 2024 and 2025 fiscal years.

In this section the rationale behind ratings have been illustrated with reference to the country and topical case studies.

¹⁶ In executing the topic review, it was found to be necessary to address not only the narrow topic of fiscal risk disclosure but also the progress made across the set of components of FR management, including the scope and depth of FR analysis. This was necessary as FR disclosure is only possible on the basis of pre-established qualitative and quantitative fiscal risk analysis.

¹⁷ Two of the projects were assessed as part of country case study and desktop reviews, and the findings and results were incorporated into the topical case study.

¹⁸ PIM and FR are strategic objectives across all countries, complemented with country-specific objectives.

4.2 WORKSTREAM-LEVEL ASSESSMENT

Within an overall situation where SEE III has delivered inputs largely as planned, the evaluation ratings achieved under the PFM workstream have been primarily determined by contextual and in-country factors. The former includes the lingering effect of the budgetary disturbance caused by COVID-19 and the difficult global economic context including accelerating inflation and rising real global interest rates. Internal factors have included serious capacity constraints in several countries in relation to deepening of required analysis for PIM and FR. In the country case study, the main impediments to rapid progress lay in delays in the completion of higher-level former reforms (OBL, and the delivery of the Integrated Financial Management and Information System (IFMIS)) as well as in issues arising from the 2024 elections and the difficulty experienced in engaging at political level in a complex political-economy environment. In the FRD topical case study the main challenge lay in the lack of skilled personnel to staff the FRUs or functions, partly caused by frequent staff transfers within the public service. In spite of these constraints good progress was made in several countries. The FRD topical case study was rated Largely Achieved, and the country case study and average desktop reviews rating both at between Largely and Partially Achieved.

PFM WORKSTREAM		
AVERAGE		
2.6		
COH	EFT	EFF
2.8	2.6	2.5
IMP		SUS
2.5		2.4

The PFM workstream has progressed in each of the countries included in the FRD topic review. However, the content and rate of progress has differed across the countries reflecting differences in baseline and country-specific challenges. In the more advanced SEE countries, the institutional base for FRM has been established now for several years (contributed to by CD from earlier SEE program phases), though it has taken time for appropriate staff complements to be placed in the FRUs and there is a perennial problem of capacity being undermined by frequent staff transfers. In the less advanced countries, in which FR management is a relatively new field, it has taken SEE III time to gain critical understanding and buy-in of beneficiary governments for them to recognize the importance of FRM. Initial concentration on the most evident sources of FR (such as SOE financial liabilities) has been followed by an expansion of the agenda to cover other forms of fiscal risk. Achieving progress in the quantification of identified FR has typically been slow, as a result of the inherent technical difficulty of analysis and complexity of the quantitative assessment tools made available under SEE III¹⁹. It is only once the basic analysis of FR is in place that attention can be usefully applied to FRD (both of which are components of FRM). So, unsurprisingly, progress in these areas has been slow to develop, albeit IMF has highlighted that all SEE countries are being assisted to make progress in tool management. It must also be recognised that there are sound reasons for the adoption of a cautious approach to FRD, especially as this applies to commercial operations of SOEs and PPPs, where typically it is not considered appropriate to disclose the risk posture of individual enterprises.²⁰

Overall progress in CD implementation has been relatively slow in the case study country, although there is an increase in momentum. The key problems explaining the slow progress have arisen from delays in higher level reforms which have undermined the ability of the specific reforms in the SEE III topical areas to progress as expected. In particular, two areas of delay have impacted on the program: i) delays in the full roll-out of the OBL, which was scheduled to have been completed by January 2025, a target which has now been extended by one year; ii) initial delays and final breakdown of the procurement of an IFMIS system to replace the antiquated and fragmented government legacy accounting system. However, the evaluation witnessed an increase in momentum of PFM reforms since the last quarter of 2024, by which time the new administration was somewhat settled in.

The state of politics and influences deriving from political economy play a major role in all SEE countries taking a variety of forms. The electoral process has an important impact on the feasibility of progressing the PFM agenda. Typically, important activities required to advance PFM agendas such as the passing of legislation and regulations, and the signing of approvals on key policy issues at the political level, fall into abeyance during the electoral process. This can be for as long as nine months, allowing for the campaigning period the presence of an interim technical government

¹⁹ When the SOE-HCT is used, it is embedded as part of MoF-standard reporting and monitoring, lending sustainability.

²⁰ Beyond disclosure, risks need to be mitigated and accommodated, ideally embedded in the budget cycle. IMF has pointed out that this last part is still particularly weak in SEE and going forward, there will be more focus on risk mitigation.

to hold elections, and the period of the incoming administration becoming conversant with PFM issues and their importance. Where elections result in a change of government the standstill can be quite prolonged. Incoming administrations typically want to take the time necessary to assess initiatives started by their predecessor. In one sampled Project the entire contract for the development of a new IFMIS was cancelled by the incoming minister, greatly impeding delivery of SEE III CD. The change in the governing party from a pro-reform government to one less oriented to reforms of the type supported by SEE III can also have a broader effect, or, of course, *vice versa*.

Weakness in the interest in or low prioritisation can also lead to stagnation or backsliding in institutional developments at the heart of PFM reform. This can take the form of failure to provide staffing to newly established units, for example for PIM and FR, leading to a common scenario of units being established but with only skeleton staff, which is not addressed until some years have passed. It also takes the form of tolerating rates of staff transfer from critical institutions which are not consistent with the effective functioning and development of those institutions, leading to sluggish or delayed reform.

4.3 COHERENCE

The main drivers of the evaluation assessment of coherence have been based on the relations between the Fund and the government, and between the Fund and the collaborating DPs. Internal coherence within the IMF has proven to be strong, itself driven by the procedures set in motion within the IMF to ensure close collaboration departments in the design monitoring and oversight of programs including the SEE III. Relations with other DPs have largely been based on the collaboration with the EU and SECO, reflected in program financing, which has provided a basis for collaboration that would have been difficult to achieve if these players had operated distinct programs in the areas currently covered by SEE III. There have been issues relating to coordination with the World Bank, especially in PIM CD implementation. The FRD case study and desktops were rated Largely Achieved on coherence, and the country case study between Largely and Partially Achieved.

PFM COHERENCE
AVERAGE
2.8

Internal coherence has been very effective with both the overall scope of SEE III and in terms of technical topics included in the program, which are strongly aligned with policy priorities identified, and more detailed attention paid to issues under each topic arising from the Article IV and policy lending processes. The FRD case study highlighted that SEE III CD support on FR has been closely tailored to the circumstances of the individual countries and based on a sequencing of forms of support appropriate to the situation on the ground and the capacity to advance their systems. SEE III CD support has also been closely linked to the analytical findings and policy recommendations arising from successive IMF Article IV reports, which in recent years have placed great emphasis on the need for effective FRM systems. There has been good coordination of CD and policy activities of different branches of the IMF, especially in the linking of training courses, in country and regionally.

External coherence has also been assessed relatively positively, albeit it is less robust than internal coherence. The positive assessment in part reflects the fact that several of the areas addressed under SEE III are recognized as falling squarely under IMF areas of comparative technical competence. In the FRD case study, given the dominant role of IMF as a setter of policy agendas and provision of FRM CD, there has been less interaction with other DPs than in some other elements of the PFM agenda. SEE III FR CD is important in assisting countries to comply with the requirements of the EU for access to EU development funds and eventual compliance with the *acquis communautaire* as part of the EU entry. For the country case study, the program has made strong efforts to engage with the administration in what is sometimes a difficult political economy context. In one country the SEE III modified its planned program to avoid overlap with an ongoing program in FR. There are concerns about potential overlaps in PIM CD delivery, where the World Bank has a PIM program in the region, also funded by EU. Although IMF and Bank responsibilities have been informally delineated, there have been occasions that SEE III CD counterparts have indicated that they are already been engaged with the World Bank on a topic where a SEE III intervention was planned. In some cases, coordination also takes place with SECO-funded bilateral projects, covering *inter alia* also PIM and FR, the two core areas of the PFM workstream. Sharing of TARs and timely notification of meetings has been highlighted by a number of EUDs and SCOs as areas that require attention.

4.4 EFFECTIVENESS

Effectiveness has largely been driven by in-country factors, including the availability and competence of technical staff in the CD counterpart institutions, and political will to implement reforms.

In some instances, the level of technical input provided by the SEE III missions was initially at too advanced a level in relation to the technical standard and training of the beneficiaries (IMF has indicated that this was because planned technical improvements did not materialise). There are examples where this weakness was recognised and corrected in follow-up missions. In several cases the evaluation assessed effectiveness score exceeds the IMF RBM self-ratings, as a result of two considerations: the lapse of time since the latest RBM self-scoring, and taking into consideration on-going progress which indicates the likelihood of achieving the stated targets in due course. In any event, the shortcomings in CD absorption environment have been at least partially compensated the high quality of the technical inputs from LTXs and IMF staff. The effectiveness of the FRD case study was scored Largely Achieved, the country case study Partially Achieved, and the desktops averaged between the two.

PFM EFFECTIVENESS
AVERAGE
2.5

The topical and country case studies highlight the main challenges in realising PFM outcomes. In the FRD case study, relatively slow progress is explained by several factors: i) the very concept of FR is complex and initially difficult to grasp even for experienced financial management staff; ii) the assessed SEE III countries faced an acute lack of experienced staff able to undertake analysis of FR; and iii) the building of institutional capacity has moved slowly across the region. A common situation was the early initial forming of an FRU lodged in the MoF but starting with only a single staff member. In most countries this situation has improved a little over the past five years. Progress is nevertheless being made in the majority of the SEE countries towards meeting RBM targets. In many cases the scores achieved are relatively low (in the region of 2) reflecting the fact that many activities are on-going and in general there is a tendency for delays in the implementation of FR milestones and outcomes. Three of the SEE III countries are well advanced in meeting RBM targets on the latest assessment. For the case study country, effectiveness has been adversely affected by the structural and contextual challenges outlined above and notably delays relating to the implementation of the OBL and the development of a modern IFMIS. Effectiveness has varied significantly across the topic areas, with good progress achieved under PIM/PIMIS as the PIMA and C-PIMA were completed; a Decree on Public Investment was prepared and approved and new guidelines on appraisal of public investment were also drafted and approved; a PIMU in the MoF has slowly built capacity; and under FR, where SEE III CD has supported the legal establishment and functioning of a new Fiscal Risk Working Group (FRWG), a fiscal risks chapter has been included in the country's fiscal strategy. IMF has also pointed out that SEE III work on PIMIS in the country has become even more important in the face of IFMIS delays, and a functioning PIMIS will be critical in bridging the gap.

4.5 EFFICIENCY

The main driver for the efficiency rating has been the occurrence of delays in the ability of countries to receive missions and implement agreed work programs.

The evaluation noted an acceleration in the adoption and implementation of reforms relative to the situation recorded in the mid-term evaluation of SEE II. This acceleration is believed to have arisen as the technical units established under SEE II gained staff, competence and experience. It was also driven by the formalization, through laws or decrees, of requirements for documentation and reports to be included in the budget documentation. This was apparent in the areas of FRD, where SEE countries have progressively introduced legal or regulatory requirements for submission of specific documents as part of the budget documentation. In the countries assessed by the evaluation SEE III LTXs have played a critical role as the dominant agents for CD delivery in-country and are the authors of the majority of TARs. Their inputs have spanned in-country visits (the majority) and online engagements. The technical quality of the TA reports is uniformly high. In some cases, as highlighted above, it was found that presentations and delivery of training were at too advanced a level for the participants and government counterparts and a simpler approach needed to be adopted. The evaluation found it was not in a position to make an explicit comparison between the inputs of experts from different sources. On efficiency, the FRD case study scored an average of 3.0, and the country case study and desktops both averaged 2.4.

PFM EFFICIENCY
AVERAGE
2.5

Institutional and staff factors in the topical case study, and slow reform progress in the country study, have influenced efficiency of assessed SEE III PFM interventions. The FRD case study evidenced that there has been overall advancement in the take up of SEE III recommendations by authorities, though FR progress is typically slower than planned, owing to institutional and staff weaknesses in the relevant bodies. In several countries in the region SEE III has fielded multiple FR missions in the past year, which in our opinion, is potentially a sign of efficiency, in terms of ensuring that the benefits targeted to be achieved from SEE III CD outputs are actually achieved. Efficiency has also been enhanced by the adoption of combination of relatively high cost in-field missions and lower cost online trainings and engagements. The ability of online engagements to be effective is mainly confined to the relatively advanced countries in terms of the institutionalisation of the FR system as they have a greater capacity to follow and benefit from online instruction and discussions. The main source of pressure on efficiency in the case study country has arisen from the delays experienced in reform progress which arise from contextual factors and capacity constraints in the MoF. The country has received the highest level of support of the countries participating in SEE III in the PFM workstream, but has consistently displayed delayed and partial success in achieving topic outcomes. However, this limited results picture should not necessarily be interpreted as CD delivery inefficiency in the country. If due recognition is given to the exceptionally low starting point in most key aspects of fiscal and budgetary management, it is assessed that the SEE III resource endowment to the country should be higher than others, albeit tempered with management of risk, especially given the resource constraint facing the program.

4.6 IMPACT

SEE III PFM IMPACT
STRENGTHENING OF FISCAL RISK CAPACITY IN SEE COUNTRIES
<p>The real achievement of SEE III in the topic area of fiscal risk is best seen by taking a step back and reviewing what has changed over the past 5–10-year period. From this perspective it is evident that important progress has been made in developing fiscal risk management in the region. A decade ago, there was almost no attention or capacity established to cope with fiscal risk, which was emerging as a critical issue as SEE countries faced large unexpected contingent liabilities falling on the budget. As of 2025, this situation has changed in all the countries included in the evaluation.</p> <p>The development of fiscal risk management capacity has typically progressed through stages, starting from creating an appreciation of the importance of fiscal risk for fiscal management, identification of the nature and scope of different forms of fiscal risk (notably including SOE liabilities), macro and budget related risks, public expenditure arrears and legal liabilities and more recently, climate-related fiscal risk. The set up of institutional arrangements (functions, units, working groups) follows. The next stage witnessed support for building capacity for quantitative assessment of fiscal risks, including the application of several of the six fiscal risk tools developed by the IMF, and notably the Fiscal risk Assessment Tool (FRAT) and State-owned enterprise SOE-HCT. Only once some capacity has been developed in fiscal risk analysis and quantification can the requirement for fiscal risk disclosure be introduced to the reform agenda. In most countries improved fiscal risk disclosure has only started to be implemented in the past couple of years, including the issuance of an annual Fiscal Risk Statement and inclusion of fiscal risk in the documentation of the MTEF.</p> <p>The progress achieved over the medium-term in development of fiscal risks management in the selected countries over the past 5-10 years should be seen as a major success. The IMF through its policy advisory role has been the lead international organization in terms of putting fiscal risk management on the international agenda, and in the development of appropriate tools for undertaking qualitative and quantitative analysis of a country’s fiscal risk profile, and through the role played by SEE II and III in providing practical support to the selected countries to build the required institutional capacities to manage their fiscal risks. There is still a strong need to strengthen fiscal risk disclosure within the wider scope of fiscal risk management. Disclosure has necessarily been the later field of development since it depends on the achievement of acceptable levels fiscal risk identification and quantification. The end result of fiscal risk disclosure playing a significant role in advising the fiscal management authorities on the need for a precautionary approach to fiscal strategy and budget development is within sight and is likely to be manifested in at least some countries in the region within the coming 2-3 years. In parallel, it would be important to act upon identified fiscal risks, adopting risk mitigation measures where required (SOE, PPP, etc).</p>

The expected impact of the PFM projects assessed under the evaluation is largely driven by the state of development, including skilled staffing, of the key technical institutions responsible for implementing the recommendations of SEE III missions. Where the institutions are reasonably positioned, which is true of an increased number of technical units compared to what was observed under SEE II, the prospects for impact are good. The anticipated impact of SEE III projects varies significantly across the countries reviewed with generally greater than expected impact in the more developed countries as recommendations are rapidly adopted. SEE III CD impact is likely to be particularly strong at the institutional level. Impact at the macroeconomic level is difficult to measure but is expected to increase over time. This will be driven by better consideration of contingency liabilities in fiscal strategy and budget preparation processes as FRSs become more quantified and are built into the annual budget process and the MTEF. SEE III projects have limited impact on inclusive growth. An attempt to support Spending Reviews in one country has failed to make progress, perhaps indicating that this is not a high government priority. Among the PFM projects assessed, the FRD case study was rated Largely Achieved for impact, and the country case study and desktop reviews averaged between Largely Achieved and Partially Achieved.

PFM IMPACT
AVERAGE
2.5

Typically, this takes the form of the preparation of an annual FRS, which in some countries is attached to the budget documentation while in other countries it continues to be a stand-alone document. Achieving impact from the activities undertaken for FRD requires a certain level of stability in the institutions established for FRM. Such stability is prevalent to a varying degree in the countries included in this topic review. Typically, institutional stability is centered on the existence and staffing achieved in the dedicated FRU or FRWG. For more substantial impact of SEE III FR projects an emphasis should be placed on building the capabilities of the staff in the units, and indeed the wider pool of technically competent staff within the MoF, given the frequent transfers of individuals between units and departments. One country in the topic selection has opted not to establish an FRU, but has instead opted to create a legally mandated FRWG. While this approach has managed to quickly assemble the available skills across multiple departments related to FR, it has also raised questions of capacity and impact. The FRWG in this case is not legally empowered to hire staff. Technical analysis must be undertaken by the active members of the group as an additional delivery. This situation is likely to have a limiting effect on impact. IMF has pointed out that such working groups are an accepted institutional platform to reach out to multiple agencies and units, particularly at the start of the complex FR work.

With respect to the topical case study, the initial signs of SEE III CD support for FR making an impact on the budgetary systems are beginning to show. Typically, this takes the form of the preparation of an annual FRS, which in some countries is attached to the budget documentation while in other countries it continues to be a stand-alone document. Achieving impact from the activities undertaken for FRD requires a certain level of stability in the institutions established for FRM. Such stability is prevalent to a varying degree in the countries included in this topic review. Typically, institutional stability is centered on the existence and staffing achieved in the dedicated FRU or FRWG. For more substantial impact of SEE III FR projects an emphasis should be placed on building the capabilities of the staff in the units, and indeed the wider pool of technically competent staff within the MoF, given the frequent transfers of individuals between units and departments. One country in the topic selection has opted not to establish an FRU, but has instead opted to create a legally mandated FRWG. While this approach has managed to quickly assemble the available skills across multiple departments related to FR, it has also raised questions of capacity and impact. The FRWG in this case is not legally empowered to hire staff. Technical analysis must be undertaken by the active members of the group as an additional delivery. This situation is likely to have a limiting effect on impact. IMF has pointed out that such working groups are an accepted institutional platform to reach out to multiple agencies and units, particularly at the start of the complex FR work.

The achievement of impact in the country case study has varied significantly across the topic areas with important positive if delayed results under PIM and FR. A less successful topic with limited impact is the support provided to the government to create a framework for spending reviews. This led to the creation of a Spending Reviews Unit in MoF but the unit has only a single staff member and no practical progress has been made to date, though the World Bank may provide more substantive support in this area. The prospect for impact is greater under PIM as the PIMU is now increasing its staffing capacity. However, it has made little progress towards executing its mandate, citing the delays in the implementation of the IFMIS as a key constraint inhibiting the creation of a public investment database. The IT required for development of the database is being provided through a tailored application of the IMF PIMIS, an Excel-based application designed specifically for the purpose in hand. Given the reality that there will be no IFMIS for several years the PIMU needs to modify its approach to data collection and start exercising the legal mandate provided by the Organic Budget Law, to demand information from Ministries, Departments, and Agencies (MDA) on their existing and planned investment portfolios. The other bright spot for impact is the newly created but now well-functioning FRWG, albeit with the reservations cited above.

4.7 SUSTAINABILITY

The main drivers for sustainability are similar to those for impact, namely the stage of development and technical capacity of the key institutions involved in the various reforms, and consistency of political support to the institutions.

In spite of the slow progress observed in some of the weaker countries, the evaluation believes that the outcomes and potential impacts have good prospects for sustainability. This is attributable to several

factors: i) the increased understanding of the technical aspects of the reform issues addressed by top management in the MoFs; ii) the steady progress achieved in progressive strengthening of the key technical units, which in many cases have progressed markedly since the evaluation of SEE II; iii) the beneficial effect of continuing IMF surveillance, especially through the Art IV process; and iv) the important commitments in several of the reform areas arising in countries under IMF programs, and requirements related to access to EU development funds which provide a strong incentive for the countries. Given the importance of stability in the staffing of the technical units, which is a necessary condition for the build-up of institutional capacity, there is probably scope for the IMF to engage more directly at a high level, typically in the MoF, to obtain assurances of staff stability at least for the period of implementation of the respective SEE III programs. Sustainability can also be advanced by encouraging the technical units (such as the PIMUs and FRUs) to establish networks of competencies within and across the branches of the MoF and with other relevant ministries and agencies. The evaluation encountered examples of this approach which can be very useful in situations of extreme skilled personnel shortage. The FRD case study scored 2.6 for sustainability, and the country case study and desktop reviews both averaged 2.2.

PFM SUSTAINABILITY
AVERAGE
2.4

The FRD case study has highlighted that in the majority of SEE countries the efforts for reform of FR management have reached the point of achieving high likelihood of sustainability even in the absence of continuing IMF support. This assessment is based in part on the degree of institutionalisation of FRM, with dedicated FRUs now present in almost all countries with encouraging signs of increased staffing, contributed to by external pressure and incentives outlined above.. However, first, in all countries the rate of future implementation progress will be lower if IMF support is discontinued, and second, there is an exception in one country assessed, where there is a significant possibility that the progress made to date will be lost in the absence of continuing support from SEE III.

Sustainability of SEE III CD projects in the case study country depends on achieving results in two dimensions. First, the establishment of permanent institutional features to support fiscal and budgetary management. These can take the form of the legal framework, and the embedding of dedicated entities, such as the technical units and legally mandated functions represented in the country by the PIMU and the FRWG respectively. The progress made in these areas lends support to the prospect of a degree of sustainability for the results of the program, Second, sustainability requires the building of capacities at the individual and institutional levels. This presents a major challenge in the selected country as the national skill pool with relevant training and experience is very small, and civil service personnel management procedures tend to weaken institutional capacities through frequent staff transfers.

KEY LESSONS AND RECOMMENDATIONS FROM EVALUATION OF SEE III PFM PROJECTS

- **Good use of diagnostics:** The program for PIM has successfully built on the information base generated through the PIMA and C-PIMA assessments undertaken earlier. Specifically, this has informed the PIM reforms through the assessment systematic recommendations which have formed the basis for the prioritisation of reforms supported by PIM reform units. Introduction of PIMIS, for example in the case study country, has also been important.
- **Linking PIM with budgeting:** The program has been effective in addressing the need for improvements in PIM reform to be accompanied by reforms to the budget process to permit the effective and more orderly inclusion of public investment into the budget. This has largely depended on the development or strengthening of the medium-term elements of the budget process.
- **Staged and sequenced approach to development of fiscal risk** has been used starting from initial basic training and progressing to broadening the scope of concept of fiscal risk to include areas such as SOE, PPP, macro, legal, government arrears and finally climate risk. This stage has used the FRAT tool in several countries to good effect.
- **Tailoring CD to low capacity environment:** In the country case study undertaken by the evaluation it was found that SEE III has been effective in assisting the authorities in a relatively weak country to focus on the most important PFM challenges faced by the country, which centered on development of the instruments to increase overall fiscal control and sustainability.
- **Flexibility in CD implementation:** SEE III has shown flexibility in the acceptance of an institutional approach to management of fiscal risk which differed from the standard approach of early development of a fiscal risk unit in the MoF, instead supporting the functioning of a legally mandated Fiscal Risk Working Group.
- **Collaboration with DPs:** SEE III needs to ensure that there is a high standard of collaboration and communication with the other leading development partners, and specifically the co-financing agencies of the EU and SECO. The program needs to ensure that the wider restrictions on the distribution of technical documents, and weaknesses in communications with DPs, are not leading to failure to share critical documents and findings with the SEE III partner DPs. SEE III donors, at the same time, need to ensure all local offices are appropriately informed about agreed procedures.
- **More CD to weaker countries merited on occasions:** SEE III should continue, to the extent possible in constrained financial situation, to support the weaker countries in the region. In the stronger countries a degree of internally driven support for PFM reform is present which is absent in the weakest countries, where the role of the IMF in PFM reform remains critical. In the country case study, it was argued that there would be backsliding in the event of a reduced SEE III presence.
- **Regional course to complement TA:** The key challenge to be the focus for SEE III is the need to continue the development of capacity in the new institutions created under SEE, notably the PIM and Fiscal Risk Units. The avenues for delivering this focus include in-country and regional training courses, which are already proving effective, continued support for the agenda and deliverables of these units, and engagement with the authorities to obtain changes in civil service management that can reduce the current destructive level of transfers into and out of the units.
- **Networks of competence:** The program should support the development of structured networks of competencies within and between the government institutions as a means of increasing access to skilled personnel who are not directly involved in the technical units. In one country this is effectively being done through a fiscal management committee with membership from all the main relevant departments within government. This principle can be extended to other institutions such as the central bank and academic centres of excellence.

5. ASSESSMENT OF TP WORKSTREAM PROJECTS

The findings for the TP workstream presented in this section are based on assessment of a sample of i) one country case study covering four Projects on tax policy unit (TPU) capacity building, tax expenditure analysis (TEA), VAT tax expenditure modelling, and immovable property tax); ii) a tax expenditure analysis and reporting (TEAR) topical case study covering CD to four countries; and iii) two desktop reviews in two countries, on personal income tax (PIT) and corporate income tax (CIT) respectively. Nine Projects were assessed. The selected sample represents 70% of TP workstream CD spend on SEE countries.

5.1 KEY ACHIEVEMENTS

TP CD projects have progressed and produced various important achievements within the first two years of the workstream’s integration into the SEE III program. Institutional capacity has been built or enhanced in almost all SEE III countries, through the establishment or structuring of TPUs, and the training and use of TP tools, such as simulation modelling. TEA activities have generated TEARs in several countries, at a first iteration or supporting future iterations. In some cases, the TEA has been institutionalized with annual updates, assigned responsibilities within the MoF, and integrated reporting into budget documentation. In three countries, TP provided extensive TA advice for the modernization of the immovable property tax, supporting the operation of a more buoyant revenue source for subnational governments.

CD on implementing TP tools have across all tax types been provided reflecting external drivers of reform (EU accession) and identification of significant expenditure costs to the budget. Placing TP in a wider context, such as data generation and exchange (statistics agency); implementation consequence (revenue administration); and budgeting cycle (MoF budget); is work-in-progress.

Generating a “trusted advisor” role for SEE III in TP has been achieved, primarily through the deployment of an experienced regional LTX who has been able to respond rapidly to CD requests, and generate both a better implementation environment and guide appropriate and relevant interventions, through a wider understanding of the region.

5.2 WORKSTREAM-LEVEL ASSESSMENT

The average rating of all OECD-DAC criteria assessed for the TP workstream is between Largely and Partially Achieved. A number of factors inhibit a higher overall score. TP, as a distinct workstream within the SEE program, is new but IMF CD tax policy assistance in SEE countries has a longer tenure and capitalizes on earlier works in the region. Though building on previous HQ support, TP is not able to harness the reform momentum generated under previous phases as can RA and PFM. TP is delivered in five of the six countries, and has not been fully active in two out of five jurisdictions. The TP is broadly at an early stage of delivery, with the results horizon somewhat distant. This is especially so for the TPD objective related to improving tax policy design, where much of the impact and sustainability will be evidenced by subsequent political approvals and legislative changes that are neither guaranteed nor immediate. Several countries have TP initiatives stretching back up to eight years. With a dual emphasis on both institutional strengthening (TPM) and improved tax policy (TPD) objectives, for most countries the former is a precursor, as the building of institutional capacity is undertaken first and is expected to contribute to improved tax policy design, but over time. The progress to date would reasonably suggest that more impact and better sustainability will be seen in the future.

TP WORKSTREAM		
AVERAGE		
2.4		
COH	EFT	EFF
2.8	2.3	2.6
IMP		SUS
1.9		2.3

5.3 COHERENCE

The coherence of TP workstream CD interventions assessed was generally strong internally and anchored on a number of emerging external drivers. Internally, the Projects demonstrated consistent strategic alignment between their core objectives of strengthening institutional tax policy functions and advancing analytical tools (notably for TEA and PIT reform). The two topical areas were designed and delivered in a way that supported one another. TP CD for PIT in one of the desktop review projects, for example, was complemented by follow-up TEA support, structured over three phased missions with consistent oversight by the same LTX, helping sustain the institutional narrative and analytical capacity over time. Eight of the ten TP Projects assessed were rated Largely Achieved for coherence, with two Partially Achieved. Good coherence was attributed to appropriate relevance to each country’s overall TP environment and challenges, with a mixture of direct linkages to external drivers (such as MTRS/RGF, Fiscal Strategy, budget laws, and EU accession requirements) and general assessment of needs. The lower scores reflect limitations on achieving coherence caused (in specific cases) by poor political ownership and lack of commitment to a clear TP reform agenda.

TP COHERENCE
AVERAGE
2.8

The workstream’s ability to achieve coherence in CD delivery varied between jurisdictions. The need for improved TP was well recognised: One country respondent stated that “we lack a comprehensive and institutionalized framework for evaluating tax expenditures. The SEE III intervention has directly responded to this gap.” Half the countries received well-sequenced CD support that responded to and built upon prior IMF surveillance findings, MTRS (in one country), and previous CD engagements. In these countries, the technical products such as TEA templates, revenue modelling guidance, and legal diagnostics, were embedded within broader reform roadmaps and gained traction due to consistent government engagement. The regional LTX’s involvement further enhanced coherence by ensuring that country engagements reflected evolving institutional conditions and avoided duplication or overlap in mission content. With respect to external coherence, SEE III TP CD delivery was generally well aligned with EU priorities and the World Bank’s PFM reform initiatives, although not always explicitly so. One country evidenced the TEA activities in both national budget documentation and donor coordination platforms. In contrast, some CD delivery was less directly connected to national strategic documents, and coordination with donors was more *ad hoc*. Where there was a lack of a functioning TPU and/or limited MoF engagement, the overall coherence was weaker or less sustained over time, notwithstanding well-prepared CD outputs.

5.4 EFFECTIVENESS

The effectiveness of the TP workstream is assessed to be between Largely and Partially Achieved. Overall, the workstream has contributed to building national institutional capacity for TPA and design, although final reform outcomes remain uneven across jurisdictions. Of the ten Projects rated, one was rated Not Achieved; five Partially Achieved; and four Largely Achieved. Intermediate outcomes were achieved when CD was associated with capacity-building, particularly in technical understanding, use of models, and diagnostic capabilities. It was less effective in achieving legislative or structural reform outcomes during the project period, though in several cases, this remains a reasonable expectation over the medium-term. Effectiveness was highest where CD was matched by institutional readiness and where local officials were empowered to act on the recommendations. The Not Achieved rating resulted from lack of institutional readiness to undertake TP functions, which prevented any substantive CD engagement.

TP EFFECTIVENESS
AVERAGE
2.3

The ten sampled Projects reflected two RBM logframe objectives, which performed differently across the assessed countries. Under the TPM objective (strengthening institutional and analytical capacity), most jurisdictions showed progress in developing foundational tax policy functions. Many engaged actively in CD on TEA and microsimulation modelling. There was evidence of sustained interaction with SEE III advisers, technical understanding by mid-level officials, and in some cases, the production of draft reports or tools that have since been integrated into policy workflows. However, the realization of the TPD objective (improved tax policy design and implementation) has been more limited to date. While several jurisdictions received sequenced CD support on PIT, VAT, and other reform

areas, relatively few translated the advice into enacted legal reforms during the evaluation period. This reflects, in part, structural and political constraints, including limited high-level commitment, fragmentation within finance ministries, and institutional instability. A related challenge is that demonstrating effectiveness is dependent in many cases on highly sensitive decision-making by government: TPU officials of themselves do not decide and implement tax policy.

5.5 EFFICIENCY

The CD delivery model for TP was well-structured, with appropriate use of expert and IMF staff time, analytical tools, and sequencing to manage resource constraints while maintaining engagement across multiple countries. Four Projects were rated Partially Achieved, and six Largely Achieved. These ratings reflect an approach that involved initial fact finding (including data needs), training on use of models (with data) and, through application of the models, increased awareness of current impacts or production of relevant information (such as TEARs) to inform policy choices.

TP EFFICIENCY
AVERAGE
2.6

A key driver of efficiency was the use of an LTX, implementing well structured CD delivery, supported by specialists from HQ and STXs, and using standardised tools. One jurisdiction analysed three taxes but at different levels using tools appropriate to the available usable data. This allowed for continuity across missions, institutional memory, and the reinforcement of core messages and frameworks. The LTX’s ongoing engagement also reduced the need for repeated onboarding and reorientation, especially in jurisdictions where reforms were complex or politically sensitive. The presence of a single advisor across multiple country engagements enabled the program to streamline diagnostics and ensure consistency in methodologies: an approach that contributed to both time and cost savings. Mission structuring followed a sequential format, typically organized around diagnostic, reform design, and implementation phases. This modularity allowed for customized CD delivery, where missions could be adjusted in length and focus depending on the reform stage and institutional readiness. Teams were often kept small but technically robust, with targeted HQ or STX expertise deployed as needed. Analytical tools, including TEA templates, microsimulation models, and diagnostic matrices, were developed in standard formats and then tailored to national contexts. This approach ensured that CD was not only rigorous but also efficient in terms of staff input and delivery timelines. Efficiency was further enhanced through informal but regular liaison between CD teams and government counterparts between missions.

Efficiency was nevertheless reduced by several factors: lack of data exchange mechanisms with external bodies delayed the availability of data to be used in modelling exercises; absorptive capacity of MoFs was affected by changes in staff and the resulting need to bring new staff up to speed; in limited cases, MoF staff were overburdened with competing tasks that reduced their engagement during missions.

5.6 IMPACT

The long-term impact of the TP workstream under SEE III is emerging, with early indications of positive institutional and technical effects. The workstream’s upstream nature means that its primary outputs, diagnostics, tools, and capacity building, would not be expected to generate immediate macroeconomic or distributional outcomes, but instead lay the foundation for longer-term reforms. Two of the Projects assessed were rated Not Achieved, both in the same jurisdiction; seven Partially Achieved; and one Largely Achieved. The ratings of Partially Achieved reflect: CD activities being conducted relatively recently; the indirect links to external drivers which could embed impact; varying institutional absorptive capacity to robustly continue the TP activities, and the distant horizon (beyond the evaluation period) before evidence will occur of policy change or sustained capacity enhancement. The Not Achieved rating was attributed to the lack of operationalised TPU that could meaningfully absorb institutional capacity strengthening.

TP IMPACT
AVERAGE
1.9

In several jurisdictions, SEE III CD has contributed to the establishment or strengthening of tax policy functions, particularly in areas such as TEA and microsimulation. These capacities are now embedded in some MoFs, supporting more evidence-based fiscal policymaking. Policy dialogues supported by SEE III CD have also helped shape reform narratives and raised awareness of distributional and efficiency concerns in the tax system. The undertaking of TEA, and (in some cases the production and internal use of tax expenditure reports), coupled with the use of scenario-based policy tools, has supported (or will support) governments in assessing reform options. While legislative adoption of tax reforms may not yet have followed, the ability to model, evaluate, and communicate policy trade-offs has improved. This improved analytical capacity may support more equitable and sustainable tax policies over the medium-term, through creating a more informed decision making in the policy making process. There is also some evidence of influence on public finance debates, including the use of tax policy diagnostics in budget preparation and fiscal planning, and peer learning. In some jurisdictions, TP CD under SEE III was aligned with broader fiscal strategies, including medium-term revenue frameworks and structural reform agendas. Where such alignment occurred, the likelihood of long-term policy impact is stronger. Tools and approaches introduced in one jurisdiction were, in some cases, replicated in others. These cross-country spillovers, facilitated by the TP LTX and harmonized tools, have encouraged a degree of convergence in reform approaches.

While macroeconomic impact remains difficult to attribute directly to SEE III TP CD (and will remain so for a period), the workstream has started contributing to building the institutional and analytical underpinnings of tax reform. In this way, the program supports longer-term goals of improved revenue mobilization (at both at national and subnational levels of government), greater fiscal equity, and more resilient public finance systems. The extent to which these impacts are realized in practice will depend on the continuation of reform momentum, sustained institutional ownership, and integration of tools and diagnostics into the policy cycle.

5.7 SUSTAINABILITY

The sustainability of benefits generated by TP CD is assessed as between Largely and Partially Achieved, with positive variation where assistance was embedded in institutional frameworks and reform agendas. The TP workstream has laid a credible foundation for ongoing tax policy reform, with some jurisdictions likely to preserve and expand upon the benefits without further support. TP CD sustainability is enhanced where institutional structures are maturing and domestic incentives for reform remain strong. Of the ten TP Projects rated, two were assessed to be Not Achieved (in same country); three Partially Achieved; and five Largely Achieved. The extent of establishment of the TPU, its capacity to adopt and implement TP functions through the use of analytical tools, and the preparation of tax policy options linked to achievement of externalised requirements, informed the ratings.

TP SUSTAINABILITY
AVERAGE
2.3

Sustainability was most evident where CD advice was institutionalized within formal structures, such as dedicated TPUs, and aligned with ongoing fiscal frameworks or external reform anchors (e.g. MTRS). In these cases, CD counterparts demonstrated the ability to independently update analytical models, produce policy reports, and apply tools within routine fiscal processes. Factors such as staff continuity, structured inter-agency collaboration, and embedded technical practices contributed to greater resilience of CD outcomes. The long-term use of tools such as microsimulation models and TEA templates also supports sustainability. Instances where models were adopted into budget documentation, or where ministries undertook follow-up analyses without further external prompting, reflect a credible trajectory of institutionalization. Continued use of these tools was associated with earlier intensive engagement and repeat missions, often facilitated by the LTX who helped ensure technical transfer and context-specific adaptation.

However, in some cases, the absence of a clearly designated TPU, limited interdepartmental coordination, or insufficient follow-up has placed sustainability at risk. CD outputs in these contexts are more vulnerable to staff turnover (especially in small units, where the impact of staff turnover can be particularly severe), political shifts, and discontinuities in reform momentum. Even where technical gains were made, lack of formal mechanisms or mandates for continued use of tools has impeded their integration into decision-making systems.

KEY LESSONS AND RECOMMENDATIONS FROM THE ASSESSMENT OF SEE III TP PROJECTS

- **Coherence with policy and legislative drivers:** Effectiveness and impact has been highest where TEA has been integrated with fiscal planning and budget documentation, backed by strong policy interest. Coherence was most notable where CD support aligned with broader tax policy reforms and EU accession benchmarks. Efficiency was bolstered by continuity of expert support and integration of TEA.
- **Multi-modal delivery:** the use of multi-modal delivery of CD, anchored around an LTX resident in the region, has been the foundation of the successful delivery of timely, appropriate and relevant CD. It has enabled rapid response and reaction to MoF requests, and provides for the LTX to generate a thorough knowledge, not just of the TP technical requirements but of the political economy surrounding both the CD content and method of delivery. A position of “trusted and reliable adviser” features in how the MoF perceives the SEE III CD.
- **Skills development:** Delivery of CD features both knowledge transfer and skills development, through ownership and management of the activity resting with MoF, and SEE III focussing on enabling MoF to undertake the technical activities. Hands-on modelling and benchmarking (compared to basic explanatory training) significantly improved technical competency within the TPUs. Ad hoc follow-up CD has been provided (at MoF request) in the case of VAT modelling, and this demonstrates (and deepens) the partnership between SEE III and MoF and supports the eventual achievement of outcomes. There is evidence of MoF embarking on subsequent TEAs.
- **Peer Learning:** Complementary regional CD through the 2024 TP conference has enabled peer-to-peer learning and awareness of potential approaches to TP reform (including lessons learned), particularly where the SEE III countries face similar challenges and share many tax, legal and administrative commonalities.
- **National ownership:** Ministerial direction and ownership of TP can be fragile, due to repeated changes of Minister, often with a short tenure. Where TP involves multiple agencies within government, such as property tax, or sectoral VAT treatments, clearer identification of risks that are likely to impact high-level decision making could assist in planning and designing CD.
- **Mitigating risks to institutional capacity:** Where CD delivery features the creation or enhancement of institutional capacity, and given the historic challenges of staff retention, the approach (within MoF) to documenting how that capacity is used could be more clearly articulated (and thus contributing to future measurement of the capacity uplift).
- **Donor coherence:** Coordination with EU and World Bank programs could be enhanced to embed TEA in wider PFM reform and EU acquis requirements.
- **Documentation and dissemination:** TEAR publication remains limited; open sharing (voluntary or as a regulatory requirement) would support transparency and peer learning.
- **RBM logframe and Strategic Objectives:** Use of the RBM framework should be reviewed, to ensure consistent approach to TPD and TPM outcomes, that clearly differentiates between the foundational institutional capacity building (TPM) and the subsequent development of tax of policy, based on institutional capacity and use of tools. Also, consider more concrete metrics, with shorter horizons, for capturing the effectiveness and impact of TPD CD – measuring through the passage of laws severely limits how the CD can be measures and likely omits much valuable CD.

6. REVIEW OF WORKSHOPS AND WEBINARS

Regional workshops and webinars have been used in different degrees by the SEE III workstreams to complement CD. The partnership with JVI and CEF has facilitated the hosting and the recruitment of presenters in workshops. Some of the workshops have involved participants from a wider range of regional countries than SEE III institutions, but with the core and target audience being the latter. In addition, LTXs have made presentations at workshops and seminars organised by JVI and CEF, targeted at an audience beyond SEE. RA workshops have focused principally on the workstream's three CoPs, and these are covered by the evaluation in Section 3.8. The workstream also organised under SEE III a series on international taxation, and a workshop on GovTech during the evaluation period. The PFM workstream organised four workshops on FR, including a regional peer exchange; one on PIM; one on anti-corruption; and one on climate and gender-based budgeting.

A review of a sample of workshops from the workstreams indicated high quality presentations and good participant feedback. The PFM climate and gender budgeting workshop, held in partnership with EU and SECO and hosted at CEF, covered overview of climate budgeting, and in the region, assessing climate risks, tagging, implementation challenges, working groups to develop action plans, and strategic budgeting approaches. The quality of presentations reviewed was very high. The workshop attained an overall satisfaction rating of 4.9 out of 5. Participant feedback suggested that future sessions should maintain high-quality expert facilitation and continue the strong emphasis on practical application, while expanding regional networking opportunities to strengthen cross-institutional collaboration. Additional suggestions included developing follow-up activities to provide ongoing implementation support and incorporating emerging best practices and innovations to ensure future workshops continue meeting evolving needs and challenges in climate budgeting. The TP workstream held a multi-topic workshop at JVI, structured as a peer event to share experiences. Topics covered included common issues related to EU integration and tax harmonization; reform of taxation of the profits of multinational enterprises and investment tax incentives; design and implementation of property tax reforms; options for rethinking the tax mix; and improving tax transparency through the publication of tax expenditures assessments reports. A review of the presentations highlighted relevance, including good analysis of regional data to stimulate discussions. Feedback highlighted significant gaps in the region on tax policy capabilities and demand for increased tax take; importance and value placed on peer learning; and the significant resources required to craft strategies and options for tax harmonisation for EU accession. A review of the RA Transforming Public Finance through GovTech workshop held at JVI also highlighted strong performance. The objective of the course was to provide participants with comprehensive knowledge about the benefits of digitalization and policies to support faster digital adoption by households, firms, and governments. The sampled presentations were relevant and of high quality. Participants scored content, presentation methods and value at between 4.7 and 4.9 out of 5, and a pre- and post-workshop test saw an increase of average scores from 51% to 74%.

The evaluation survey of participants also provided generally positive feedback about SEE III workshops. Participants rated inputs and outputs, such as quality of presenters, training material and relevance of topics extremely highly, with outcomes such as implement-ability of actions discussed, and linkage with SEE III CD, somewhat lower. Main benefits were seen as increased awareness of international practices and gaining of technical knowledge. Around 2/3 thought they could only implement less than 50% of lessons in their workplace (decreasing to 42% in two years), the main barriers being lack of authority to make the necessary changes, and lack of budget and resources. Main suggestions for improvement included making them more practical, narrower and more in-depth, more relevant to the institutions, and follow-up with CD.

7. REVIEW OF SEE III GOVERNANCE AND OPERATIONS MANAGEMENT

This section addresses the Governance and Operations Management related questions in the ToR, and additional points raised during the inception phase. The contents are intended to supplement the SEE Phase II recommendations follow-up table.

SEE III governance and operations management is generally strong. The combination of technical expertise and project management tasks adds effectiveness to the PM role, enabling more informed steering of program activities and guidance of experts. Intervention design and redesign can be more efficiently undertaken, and more appropriately targeted. There is good coordination of activities between workstreams, with PMs meeting weekly to discuss and resolve strategic and operational issues, thus ensuring coherence in the workstream activities. Significant efforts are made in internal communications, with EUR, and other parts of FAD. The feedback from EUR staff was that the SEE III program adds considerable value to its surveillance and Program activities. The presence of LTXs on the ground provides more current and deeper insights. The program has also made concerted efforts in communicating with local EUDs and SCOs, especially through the LTXs. Although there are areas where improvements can be made, worth noting that in the DP survey, 2/3rd of the nine EUD/SCO respondents indicated that they were either fully satisfied, or mostly satisfied with SEE III cooperation. This foundation can be built upon to further strengthen program management in a number of areas, as highlighted later in this section.

SEE II evaluation recommendations have been implemented to a great extent, many explicitly, some more so in principle:

REVIEW OF IMPLEMENTATION OF SEE II EVALUATION RECOMMENDATIONS

- No formal Political Economy Analysis has been undertaken, as envisaged in the recommendation, nor a Governance Diagnostic, as the IMF response to the recommendation indicated might be considered. Nevertheless, as highlighted in the SEE III implementation status update, surveillance and other alternatives are relied on. The presence of LTXs on the ground has also helped develop improved understanding of political risks. The PFM workstream also held an anti-corruption workshop twice. We additionally understand that practical steps are also taken to manage such risks, for example, implementing shorter-term projects in higher risk countries, or slowing down CD delivery around elections. Nevertheless, greater consideration might need to be given to potential relative risks regionwide, when allocating CD resources. For example, allocating 27% of country-specific CD resources in a workstream to a single country may be too big an exposure to potential risks, irrespective of the prevailing reform appetite (our case study on that country concludes that that it continues to merit a high level of CD, perhaps not that high).
- We have seen some evidence of *ad hoc*, formal capacity gap diagnostics being undertaken, although not in a systematic and comprehensive manner. In any event, again, our reviews suggest that especially with the presence of LTXs, advice is in the current phase tailored to counterpart capacity, and this is evidenced by CD counterpart feedback in the evaluation survey (although in one country case study advice was pitched at too high a level, but corrected in the next mission). Limited progress has been made on softer skills including change management, and collaboration with CEF in this area is not in-depth, or systematic. Perhaps it is for SEE donors to consider supporting a separate vehicle to deliver CD on softer skills as this is not an area of IMF core competence, strongly encouraging close collaboration with IMF's expertise of providing more technical CD.
- There has definitely been greater focus on consolidation and capacity building in the current phase, and more LTX engagement as recommended has supported this objective (see below for more on LTXs). Peer-to-peer exchanges, more formally and broadly introduced in the RA workstream, and also undertaken in PFM, for example in FR, are in high demand from CD counterparts.
- The recommended increased focus on CRM in the RA workstream has been implemented, with CRM CoP the most mature, and CRM being one of the regionally implemented CD topics.
- With respect to integrating better data analytics and CRM CD, the evaluation assessments saw two countries where STXs were implementing
- We haven't seen any evidence of the nature or degree of strategic collaboration between SEE III and EU envisaged in the related recommendation.
- The recommendation related to milestones and outcomes had two elements: i) that CD implementation plans be developed in a more realistic way, taking into consideration counterpart capacity, and this appears to have taken place; and ii) that RBM logframe and outcomes and milestones also set realistically, and discussed with counterparts. The latter does take place, but RBM logframes continue to experience a number of shortcomings, which are discussed later in this section.
- On the embedding of cross-cutting experts such as anti-corruption, climate and gender, whilst not progressing to the degree of implementation envisaged in the recommendation, we believe progress has been satisfactory.
- Regional aspects of the CD program have been strengthened significantly, through a combination of: i) implementing selected CD topics regionally within each workstream; ii) engagement of regionally based LTXs; and iii) promoting CoPs and peer-to-peer exchanges, especially in RA, and also PFM (with one TP workshop designed to bring out regional commonalities and promote peer discussions).

The TP workstream has added an important dimension to the SEE III CD offering, although its benefits are yet to be fully realised. The workstream fills an important gap in meeting the program goal of building efficient and effective fiscal institutions in SEE III countries. The dual-track complementary objectives, of building tax policy analysis in the ministries so that they can make more evidence-based and informed policy decisions; and in parallel providing tax policy advice in specialised areas, in principle reinforces the objectives of the other two workstreams. There are also important areas where joint TP and RA, as well as TP and PFM workstreams can work together²¹ to enhance mutual effectiveness, as identified in the SEE III PD. It should be considered, nevertheless, that building TPU capacity requires a relatively long-term horizon, especially in a number of SEE countries where shortage of staff, skills and high turnover is particularly acute. Fuller contributions to the program goal will be realised once the TPUs mature into relatively autonomous functions. To-date there have been limited opportunities for the TP workstream to proactively coordinate its CD activities with the other two workstreams (the passing of an OBL with PFM workstream CD, to be followed by TP CD work with the ministry to analyse implementation options, is one of a few relatively rare cases). There are opportunities for the workstreams to craft CD that reinforce each other, especially between TP and RA, but with a relatively limited budget, the workstream may not be able to pursue both its own capacity building and providing targeted policy advice objectives, as well as work on joint interventions, effectively.

IMF should strengthening its internal SEE III program accounting to ensure allocation of costs to projects. A breakdown of SEE III costs by activity highlighted that not all PM&A costs are captured (especially those related to LTXs) accurately. Also, a significant proportion of costs, although notionally allocated to countries, are not attributed to specific CD projects. Allocation of costs more accurately to countries and CD projects would enable closer monitoring of fund utilisation, and strengthen program management. Moreover, more accurate recording of PM&A costs would provide a more realistic estimate of overheads (at present the target at 4.5% of CD costs is, we believe, unrealistically low). We consider time not allocated specifically to specific CD delivery, or project management activities as defined, should be minimised. This will require more accurate and diligent timekeeping by LTXs and HQ staff on multi-topic missions through the intended use of the TRACES time recording system, and clearly timekeeping obligations should not become overly burdensome. Outputs could be included in the annual SC reporting, and used for internal resource allocation decisions.

The use of RBM logframes within SEE III requires improvement, within the bounds of the current system, for it to produce more consistent data, act as a more effective indicator of progress, and function as a project management tool. SEE III project managers have limited scope for changing parameters of the RBM system. Many of the functionalities are hardcoded into the system by ICD, and catalogues for objectives, outcomes and indicators are set by the CD departments, including FAD, which can be updated periodically (with input from CD programs such as SEE III). The program has flexibility in defining and updating milestones, and the IMF RBM governance guidelines see milestones as the main tool for tailoring logframes to program activities. Some of the wider system's shortcomings include: i) inability to set project start and end dates, with the system assuming the start and end dates of the program being that of the project, and setting these as objective and outcome target dates; ii) the system not maintaining an audit trail, so, for example, it is not possible to track progress in achieving project target thresholds, or delays in target milestone dates; iii) the outcome rating date is automatically updated to the latest date a logframe is downloaded, so it is not possible to establish the correct date of the last update; iv) milestones are linked to outcomes, nor outcome indicators, and for outcomes that have more than one indicator, it is not possible to track progress toward them separately; v) it is not possible to create more than one project in a country with same objectives and outcomes, so, for example, in BiH, it is not possible to separately track progress in CD delivered on the same topic to the different jurisdictions; and vi) the outcomes and indicators rating system is not granular enough to adequately demonstrate progress during a program (with especially the rating of 2 marking achievement of between 26% and 75% of the targeted outcomes). SEE III outcomes and related indicators are generally fine, but some, for example in TP, would benefit from

²¹ (RA and TP) Administrative feasibility: RA support for implementing revised tax policies; Joint diagnostics and gap analysis: RA provides tax gap, compliance, and data analytics support to TPUs; Shared digital infrastructure: TP relies on RA systems (e-filing, data warehouses) for analysis and forecasting; Institutional collaboration: Co-design and data sharing between RA and TPUs. (TP and PFM) Revenue forecasting: TP reforms directly influence PFM budgeting tools like MTEF; Fiscal transparency: Tax expenditure (TE) analysis supports transparent budgeting and spending prioritization; Green finance: TP on environmental taxation complements PFM's Climate-PIMA and greening investment planning; Integrated institutional reform: TP, RA, and PFM work coherently under the ministries of finance and with EU alignment goals.

more granularity given the extended period that may be needed to translate policy advice to laws to implementation, to targeted outcomes (e.g., improved distribution) to be realised. Also, many outcomes relate to formal assessments, e.g., PIMA, PEFA, FTE, or TADAT indicators, that require follow-up implementation to realise higher-level outcomes that can meaningfully contribute to impacts.

There is flexibility within these limits for SEE III to improve its use of the RBM systems, in particular with the use of milestones. Key areas that might merit attention include: i) establishing periodic milestones, at least annual ones, preferably more, to align for example with TAR implementation plans; ii) updating milestone target dates to realistic ones in the future if they have been missed, to keep them relevant; iii) defining milestones in a manner for which achievement could be easily assessed, either qualitatively or quantitatively; iv) more diligent use/rating of outcomes and especially outcome indicators; v) that they are indeed properly defined to assess implementation of CD recommendations (e.g., one of the projects simply listed the annual publication of a report initially supported by SEE III CD as a series of milestones, which is not appropriate); and vi) taking steps to ensure consistency of rating approach between workstreams. Consideration should also be given to developing some impact-level indicators for the workstream and the program, although IMF has indicated that this is not possible as there are only indicators in the RBM system at the outcome level. Based on CD counterpart survey responses, and experience of an impact study undertaken as part of another IMF CD program, institution-level, rather than a macroeconomic-level indicators, may be more appropriate for impacts. The RBM system does provide objectives, at a higher level than outcomes, and although some programs measure them²², in practice objectives are difficult to target, especially as they do not provide indicators.

Donor reporting, especially in relation to logframes, could be further improved. The main body of annual progress reports is informative, with a prioritisation on highlighting results. There is potentially some scope for making it more strategic by providing information on the target for the project at the end of the phase and progress along it vs plan. Additionally, donors might benefit from a more strategic section, that highlights cross-linkages, lessons, and issues encountered at a regional level, as well as on interlinkages of activities between the workstreams. IMF has highlighted that the SEE III regional notes could serve this purpose. The logframe could be strengthened by including more information about missions and other CD activities, outputs, and outcomes at different levels, with an appropriate degree of summarisation: The historic RTAC annual reports, which provided summarised information on milestones and outcomes, could be a guide. Currently much of the required information is provided in the main report in various sections, but these are difficult to link together. IMF has highlighted that significant work has been undertaken to streamline the logframe contents from that presented in the previous phase, and much effort invested to ensure the logframe contents conform with donor (especially EU) needs. However, feedback, especially from EU, suggests further tailoring is needed. Moreover, many of the logframe definitions do not translate easily to CD areas, so linking them (as in the workplans section in the main report) could improve understandability. Additionally, for reasons stated in the previous paragraph, the rating system does not enable easy assessment of progress in achieving outcomes between years, with the significant majority of ratings remaining unchanged between years. More systematic use of milestones, presented in a summarised manner, would be one option for addressing this information need more granularly. IMF has indicated that considerable resources would be needed to meet these requirements. We do not fully agree. Provided time and milestones data is input as intended, almost all the reporting could be automated (our recent evaluation of an RCDC and current evaluation of another IMF CD program highlights that this can be achieved through the use of, for example Power BI integration with CDMAP).

The program's efficiency could be strengthened further with dedicated administration support. We agree with the IMF approach of FAD staff with expertise in the relevant technical subject(s) taking leadership of the program and workstreams. A dedicated project management and administration officer, preferably but not necessarily with some logframe experience, could provide support to the PMs and LTXs on a range of areas. A review of relevant tasks currently being undertaken by HQ staff and LTXs would be desirable, but activities undertaken in that role could include: i) managing the RBM entries and following-up with staff for information as necessary; ii) maintaining a document filing and archiving system (harmonisation of file-naming system between the workstreams would facilitate this process); iii) acting as a help-desk to, for example, EUD and SCO staff with document related queries and informing them in advance of missions; iv) supporting LTXs on administration matters, such as contracting of STXs; v) arranging travel of

²² Mainly through calculating a weighted average of related outcomes.

HQ staff and LTXs; vi) arranging meetings etc. for missions for LTXs (and HQ staff to the extent not being handled by others currently); vii) collating information and helping prepare internal and external reports; viii) supporting the logistics and preparation for annual coordination and SC meetings; etc.

Greater emphasis may be needed to be placed on addressing human capacity-related weaknesses in counterpart institutions, and allocation of resources to countries and institutions may require tailoring to reflect related risks. A common barrier to realising sustainable results efficiently across the workstreams is the inadequacy of human resources devoted to carrying out necessary functions and implement CD recommendations. This problem is often exacerbated by high staff turnover, frequently caused by staff changes directed by management. IMF has taken steps to address this matter, e.g., by suspending CD activities until a counterpart unit is adequately staffed, or escalating the issue within the institution (albeit the latter often with limited success). Given its considerable impact on CD effectiveness, efficiency and sustainability of the program as a whole, this problem may require a further increase in prominence in the dialogue between IMF and the authorities, both at the SEE level, as well as wider IMF institution-level consultations with the senior-most level of representatives from the governments. In addition, resource allocation between countries and counterpart institutions may need to take into consideration authorities' commitment to adequately staffing relevant units, and implementing HR policies that are conducive to institutional stability, effective CD absorption and implementation, and demonstrating results from SEE III CD. Having said that, there may be cases (e.g., the PFM case study country), where institutional weaknesses may warrant the provision of more intensive CD, but such exceptions should be clearly justified.

SEE III is projected to see a significant reduction in available budget for all workstreams in the current fiscal year, with the RA and TP workstreams having no funding available for the eight months of FY2027. The reduction in budget and available funds has been caused by the following:

- The annual financial projections in the PD assumed that the program would taper down in the last two years, with RA reducing LTXs from three to two in the third year, and none in the fourth year²³; PFM budgeting for a total of eight LTX years, with the number of LTXs reducing from three to one over the four years; and TP allowing for one LTX over the four-year period. It is unusual to see a program's activity reducing so significantly over time, risking loss of momentum and needing to leave CD projects uncompleted. Budget constraints would normally result in a lower but relatively even level of activity throughout the program, rather than starting with a higher spend in early years, and then tapering off. The annual PD projections were based on the premise, also stated in the PD, that IMF would seek other sources of funding, and use own funds for emergency needs. There was also an implicit assumption on the part of IMF that top-up funds might be available from the existing SEE III donors, although any related discussions do not appear to have been documented. In the event although the total funds contributed to SEE III CD activities from IMF and other donor sources has been material, it has not been enough to offset the reduction in annual budget and spends in FY2026 and FY2027.
- The potential impact on CD activity of the PD-projected reduction in available resources in the last two fiscal years of SEE III has been further exacerbated by the actual spend being greater than such projections. Although there was underspend compared with PD projections in the last four months of FY2023 as the phase started, both FY2024 and 2024 saw actual spends at 21% and 18% higher than PD projections. The remaining funds for the rest of the program are consequently even lower than that provided in the PD. The cumulative spend over PD projections until FY2025 is a relatively modest 7% of the US\$15.5m budget for the current phase. However, the impact of the difference on CD activities is compounded by the fact that the annual projections for the last two years show relatively low, and decreasing, availability of funding.
- There is no documentation of the reasons for the actual spending being more than that projected in the PD, although we understand that part of the difference for the TP workstream is accounted for LTX cost being higher than that provided for standard costing. On whether SC approval was sought or needed for the actual level of annual spending being higher than that in the PD, IMF has highlighted that financial information presented in the PD is not intended to represent annual budgets, but only indicative projections representing budget estimates.

²³ This is from Appendix 5 of the PD. Para 68, however says that the number of RA LTXs would be reduced from three to two over the period)

Moreover, IMF has pointed out that unlike RCDCs, SC endorsement of annual budgets is not anticipated or required (following year budgets are presented as part of annual progress reports for SC information only). This is in principle inconsistent with the provision in the SEE III operational guidelines, which envisages donor approval for reallocating budgeted costs between lines (although not between years), which would not normally be expected if the SC had no part in approving the budget in the first place.

- The following observations are worth making relating to financial management: i) it is relatively risky to make projections on the basis of assumptions or understandings without firm commitments and should preferably be avoided; ii) there appears to be uncertainty between IMF and SEE III donors about the nature of annual budgets presented in the progress reports and whether the latter's approval is required; iii) to the extent that PD annual financial projections are intended to be indicative only, consideration might be given to presenting firmer annual budgets to donors relatively soon after the program starts and adequate plans have been made (in this case the program started in January 2023, and the annual projections and overall financial position was discussed at an SC meeting in November 2024).

Other points:

- SEE III CD continues to be well-aligned with supporting member countries meet their EU accession objectives in all three workstreams. Work in the PFM workstream, such as on PIM, is supportive of the New Growth Plan (NGP). The evaluation has not discerned the NGP being a major factor in CD demand or delivery during the current phase of the program.
- We believe there is limited scope for significantly scaling up RA and PFM workstream CD delivery, given CD counterpart absorptive capacity, and other CD providers in the space. There may be opportunities for increasing TP CD, provided: i) a careful audit is undertaken to ensure ministries dedicate adequate staff to TPUs to enable the provision of effective and sustainable CD; and ii) additionality considerations are properly addressed as previously tax policy advice was being provided directly by IMF HQ.
- Interviews with EUDs and SCOs, SC members, and DP survey responses suggest persistent difficulties in accessing TARs from the Partner Connect/Teams system. Also, a number feel inadequate notice is given ahead of missions. Also, more engagement could be had between IMF HQ staff and SECO/EU HQ staff, both in respect of strategic issues, but also with specialists in relation to technical matters.

8. FINDINGS AND RECOMMENDATIONS

Given that the current phase of the program is expected to end in December 2026, and further IMF fiscal CD to SEE countries might be rolled into a planned RCDC, no major strategic changes are proposed. The previous phase evaluation recommendations focused on program strategy, and they have been broadly implemented. The evaluation has not identified any immediate need for strategy shifts. The principal recommendations from this evaluation therefore focus on more management and operational issues, aimed at strengthening the program’s effectiveness and efficiency in the short-term. In addition to the formal recommendations, some suggestions are provided later in the section, which IMF and the SEE donors may wish to consider. Finally, each of the workstream review sections (Sections 3 to 5) provide recommendations specific to those workstreams.

	Findings	Recommendations	Priority
1.	The CDMAP RBM logframe system, whilst providing a good basis for collating CD project related information, suffers from some shortcomings. Many of these are outside the program’s control. However, the current system could be used more effectively to provide more useful information, both for FAD internal use, and donor reporting, on CD project performance, and potentially also as a project management tool. Most importantly, milestones, which the program has the ability to tailor, can be used more extensively and consistently to generate information on project progress and as a project management tool. The current practices do not fully comply with the IMF RBM Operational Guidelines.	<p>Improve the use of the RBM logframe system to provide more accurate, timely, consistent and actionable information by ensuring:</p> <ul style="list-style-type: none"> i. Milestones are used more extensively, and set at least annually (preferably more frequently), e.g., using the TAR action plans as a guide. ii. They are current, and review and update target dates periodically at set dates (e.g., every six months to a year). iii. Milestone definitions are appropriate to assess CD implementation, and their achievement can be easily assessed. iv. Other indicators, such as those related to outcomes are also rated consistently. v. That a consistent approach to establishing milestones and outcomes is used in all three workstreams, so that outputs are comparable. <p>Additionally:</p> <ul style="list-style-type: none"> vi. Change the logframe for CoPs to one that is more appropriate for the project. Request a catalogue change from FAD. vii. Keep a periodic record of the logframe to preserve an audit trail. 	High
2.	Donor reporting is currently relatively static, providing a snapshot of the program’s position at the end of the year. It could capture the regional dimension more fully, and provide more strategic information to readers. The reporting format of the logframe is difficult to link with the main report; it does not reflect progress in project implementation as the ratings do not move much between years; the information presented is at a country level and does not capture the regional dimension; and importantly, it does not function effectively in serving its purpose as a logical framework, i.e., logically linking activities, outputs, outcomes and impacts. Much of the information needed is already presented in the main report, or is available in CDMAP.	<p>IMF to consider updating the contents of its annual progress reports to incorporate more information linking current period activity with achievements targeted for the phase, and bring out more regional-level and cross-workstream issues to the extent relevant and not provided elsewhere. The logframe should include information about activities, outputs, outcomes and targeted impacts, including, on a consolidated basis, at the program level (as envisaged in the RBM Guidelines).</p> <p>One potential option to consider would be to create, for each project, a template that, drawing on information from CDMAP, automatically populates data related to workstream, country, CD area; inputs (e.g., Field Person Weeks, or costs); activities (missions); outputs (e.g., TA reports); milestones (current period, next period, total number); outcome indicators and outcome(s); objective; ratings (previous, current); etc. The data from each project sheet could then be combined by different</p>	High

	Findings	Recommendations	Priority
		groupings, and varying degree of abstractions to arrive at workstream, country, and program-level indicators. Selected areas such as missions, plans vs actual.	
3.	Managing a multi-country, dispersed program, with a range of internal and external stakeholders; engaging in a range of activities and events; generating various types of information; and ensuring activities are coordinated at the administrative level requires a level of project management support that SEE III currently lacks. Consequently, resources such as LTXs and PMs need to spend time on matters that can be delegated. Moreover, time constraints can at time impair performance, e.g., in dealing with EUD and SCO needs in a timely manner. An illustration of the range of tasks the project manager or coordinator can provide support on is provided in this report.	IMF should consider engaging a project administrator or coordinator to support the technical team at HQ and LTXs. An option IMF may wish to consider is to engage a person in one of the SEE countries, perhaps on a contract basis. This would be more cost-effective than hiring as staff someone at IMF HQ.	High
4.	CDMAP activity reports show a high amount of unallocated or misallocated time costs, or costs in general categories, especially relating to LTXs. This practice leads to understating the true costs of CD projects; understating or overstating of project management costs; and potentially reduction in accountability for LTX time. It is understood that PMs monitor LTX activities and performance closely, and the FAD Resource Management function has recently taken steps to ensure staff and LTX time is accounted for more precisely. There would need to be a balance between attaining a degree of accuracy in this area, and the administrative burden on staff and experts.	IMF should minimise the levels of program costs unallocated to specific CD projects, or project management and administration.	High
5.	Limited human resources and high staff turnover at counterpart institution is a consistent barrier to efficiently achieving sustainable results in most SEE countries across all three workstreams. Whilst wider economic and social factors play a role, many of the challenges could potentially be addressed through increased senior-level commitment to devote adequate resources and improving institutional HR practices such as avoiding frequent staff changes and appropriate incentives for good performers. Exceptions may be needed in instances where lack of capacity may be caused by external factors outside authorities' control.	IMF should consider in its dialogue with senior representatives of counterpart institutions and governments, at the SEE III program level and in wider HQ consultations, further highlighting the role of inadequate human resource allocation and poor HR practices as a major barrier to reforms and CD effectiveness. Where possible formal commitment to address this problem should be secured. Allocation of CD resources to countries, institutions and projects should generally more closely reflect adequacy of resourcing and potential for achieving results.	High
6.	Current financial projections suggest that the level of CD activity is likely to reduce significantly over the remaining course of the program. This raises risks relating to: i) existing projects being left partially completed, and thereby impairing their prospects of delivering results and achieving sustainability; and ii) there being a significant gap between the current program ending, and any new one recommencing with those or related CD	In developing a strategy for the rest of the current phase within the available funding envelope, IMF and donors should consider both the impact on current CD projects, and the implications of leaving any significant gap, or activity falling below a critical level, on effectiveness and efficiency of the remaining period of the current phase, and on any future program.	High

	Findings	Recommendations	Priority
	<p>activities. In the latter case, continuity and momentum risk being lost, with greater investment needed to restart the project. It would also take any new LTX or expert significant time to restart the process of gaining trust and building confidence before most counterparts engage seriously in taking and implementing their advice.</p>		

The following are some of the additional points that have arisen from our reviews of SEE III CD projects and the program that IMF might consider addressing. These are offered by way of suggestions.

Suggestions
<ul style="list-style-type: none"> IMF should consider closer collaboration with CEF to draw on its expertise in designing, creating and supporting the operation of communities of practice to further strengthen the operations and sustainability of the RA CoPs.
<ul style="list-style-type: none"> Given the relatively frequent technical problems reported by EUD and SCO staff in accessing documents from Partner Connect/Teams, consideration should be given to hosting the documents to an alternative, appropriately secure, platform. Explore feasibility of sending out auto-email notifications on upload to those that have access. Also consider a help desk for queries on this and other topics (especially if a project administrator is engaged).
<ul style="list-style-type: none"> ICD should consider making available to program managers materials related to logframe theory and project and program cycle management. The objective would be to raise or refresh awareness about logframe structures, logic, role of assumptions and risks, and how logframes can be used as a M&E and reporting tools. To the extent such information is not available already, this would enable the more stripped-down IMF RBM logframe to be viewed in context. In addition, consideration might be given to providing overviews of the logframe and results frameworks used by major IMF donors, so that a more informed dialogue can be undertaken with them on reporting. On project cycle management, the major components, including the components going into intervention design; the different elements and stages of monitoring and evaluation; approaches to (at least theoretically) managing results chains; and approaches to managing program funds; would be useful background material. This should preferably also include concepts related to results-based management as applied for development interventions, and how the CDMAP system is intended to support such practices. The RTAC Handbook covers some of these areas, but not in our opinion sufficiently, and providing briefly the theoretical contexts and some examples of other practices would be helpful.
<ul style="list-style-type: none"> With respect to financial management matters: i) Tapering down of budgets towards the end of a program carries various risks, such as leaving implementation of CD advice remaining incomplete, and loss of momentum. Where such projections are based on anticipation of resources being available in the future, these should be relied upon only on the basis of firm commitments; ii) IMF and SEE III donors should clarify matters such as whether annual budgets require SC approval or endorsement, any requirement for reallocation of funds between years, and update the SEE III operational guidelines accordingly; and iii) to the extent that PD financial projections are considered indicative only, IMF should consider providing SC with a firm annual budget for the prevailing year within a set period of the program starting.
<ul style="list-style-type: none"> IMF should consider entering into discussions with key DPs, especially World Bank, on how collaboration on CD activities could be further enhanced to leverage each other's engagements. In addition to talking and sharing of information, this might involve identifying and addressing barriers to closer, more systematic engagement; how to best involve WB HQ and local offices; and developing joint plans of action. Responsibilities would be assigned in the respective institutions for monitoring action and corrective action agreed where appropriate. We understand that the RA workstream has already embarked on such a process with the Bank, and the other workstreams could follow.