Mid-Term Evaluation of FIRST Phase III Operations (2013-2017)

Volume I: Final Synthesis Report

REPORT | MAY 2017

Contents

1 INTRODUCTION ... 1

1.1 Background ... 1

1.2 Evaluation Objectives and Methodology ... 1

2 PROFILE OF FIRST ... 4

2.1 FIRST Initiative Objectives and Model ... 4

2.2 FIRST Governance and Management ... 5

2.3 FIRST Windows and Portfolio ... 6

2.4 FIRST Expenditures ... 6

3 CONCLUSIONS ... 7

3.1 Context ... 7

3.2 FIRST Initiative ... 7

3.3 FIRST Projects ... 9

3.4 FIRST Programs ... 9

4 RECOMMENDATIONS ... 11

5 LESSONS LEARNED ... 17

Exhibits

Exhibit 2.1 FIRST Phase III Status as of 31 December 2016 ... 6

Exhibit 2.2 FIRST Commitments and Actual Expenditures (December 2016) ... 6

1 Introduction

1.1 Background

The Financial Sector Reform and Strengthening Initiative (FIRST) is a multi-donor program aimed at supporting low- and middle-income countries (LICs and MICs) in building stable, more efficient, and inclusive financial systems. Since its inception in January 2003 it has funded more than 700 projects and programs in over 120 countries, and disbursed over USD 150 million between 2002 and June 2016.

FIRST is currently in its third five-year funding cycle (Phase III, January 2013 – December 2017) with support from five donors: the Department for International Development of the United Kingdom (DFID), the Ministry of Foreign Affairs of the Netherlands (MFA), the State Secretariat for Economic Affairs of Switzerland (SECO), Germany's Federal Ministry of Economic Cooperation and Development (BMZ) and the Ministry of Finance of Luxembourg. The World Bank (WB) and the International Monetary Fund (IMF) are the implementing partners. As of December 2016, FIRST had mobilized USD 98.7 million (including USD 10.5 million rolled over from Phase II) for Phase III operations.

This mid-term evaluation of FIRST was commissioned by the FIRST Governing Council (GC) to assess the implementation of the Phase III strategy to date and to draw lessons to inform the strategy going forward. It focused the assessment at two levels: (i) FIRST's Business Model, governance, and management; and (ii) Projects/Programs' performance. The evaluation was expected to provide a sound basis for assessing the introduction of the Programmatic Window and the corresponding results framework (eligibility criteria). It was also expected to gather and incorporate lessons learned, including feedback from the 2015 Consultative Group (CG) meeting. The mid-term evaluation was pushed back to year 4 of the Phase III cycle in order to capture preliminary outcomes from programs, which have implications for Phase IV design. The evaluation was carried out by Universalia Management Group Ltd between September 2016 and February 2017. The Terms of Reference (TOR) are presented in Volume II, Appendix I.

The report is presented in two volumes: a Synthesis Report in this Volume I, and a more detailed Technical Report in Volume II.

This final report reflects feedback from the Program Management Unit (PMU) and FIRST stakeholders on draft reports submitted 3 March, 20 March and 18 April 2017.

1.2 Evaluation Objectives and Methodology

Evaluation Matrix and Objectives

The evaluation was guided by an evaluation matrix (see Volume II, Appendix II) that outlined the evaluation foci, key evaluation questions, sub-questions and illustrative indicators. The matrix guided data collection and analysis and the development of standardized interview protocols and a survey instrument.

This Mid-Term Evaluation assessed:

(i) **FIRST project and program performance** was assessed based on the standard evaluation criteria of the Organization for Economic Co-operation and Development (OECD)/Development Assistance Committee (DAC) – Relevance, Effectiveness, Efficiency, Impact, and Sustainability. The assessment focused on FIRST projects that had been completed as of June 2016, and a sample of active programs with a cut-off date of December 2016.

(ii) **FIRST's Business Model and management effectiveness** were assessed against the performance targets adopted in the Phase III Strategy, including the extent to which Consultative Group (CG) recommendations have been implemented.

(iii) **FIRST Initiative performance** was assessed based on OECD evaluation criteria – Relevance, Effectiveness, Efficiency and Sustainability.

(iv) **FIRST gender mainstreaming approach**, adopted in December 2015, was reviewed to assess the appropriateness of its early implementation and provide recommendations for adjustments or improvement.

Sample and Methods

The evaluation sample was 47 catalytic projects that had been completed at June 2016. All projects in the sample were implemented by the WB, as no IMF implemented projects had been completed at that time. Ten of the 47 projects were excluded from the review: three because the projects were cancelled; one because there was no team member with the relevant technical expertise to rate the project; one because it was embedded in an ongoing program and five because the deliverables were in languages other than English and French.

The evaluation used a mixed methods approach to strengthen the reliability of data and increase the validity of findings and recommendations. This helped to broaden and deepen understanding of the processes through which results were achieved, and how these were affected by the context within which the program was implemented. The approach also allowed for triangulation of data from a variety of sources. Methods included interviews, country site visits, and document review.

- **Interviews with key FIRST stakeholders:** The evaluation team conducted individual interviews throughout the evaluation with more than 95 individuals representing a cross-section of FIRST stakeholders, including: FIRST GC members, current FIRST managers and staff, WB staff, representatives of national governments and Central Banks that received FIRST support, and key stakeholders in the financial sector (see Volume II, Appendix III for a list of stakeholders consulted). Interview guides were drafted, tested and revised for various stakeholder groups, using the evaluation matrix as the guiding framework (see Inception Report, December 2016). The information and opinions collected during the interviews were aggregated and synthesized and informed the findings of the evaluation.
- **Document reviews:** The evaluation team reviewed more than 200 key documents, including strategy documents, reports, governance documents (e.g. meeting minutes), lessons learned pieces and Financial Sector Development Implementation Plan (FSDIP) documents. The evaluation team also reviewed the project approval documents, Client Surveys, GRMs and deliverables for 38 FIRST

projects and five FIRST programs, including more than 400 additional documents. Information gleaned from these reviews was summarized according to the key questions and sub-questions in the evaluation matrix. A list of documents reviewed is presented in Volume II, Appendix IV.

- **Country visits:** The evaluation team conducted country visits between November 2016 and January 2017 to Rwanda, Lesotho and Morocco to obtain the perspectives of 46 key stakeholders on activities financed by FIRST in those countries (see Volume II, Appendix V for list of projects and programs visited).
- **Project and Program reviews:** The evaluation team's financial sector experts reviewed and rated 37 completed FIRST projects and five FIRST programs (see Volume II, Appendix V). The reviews were based on documents including proposals, aide memoires, Annual Monitoring Reports (for programs), client surveys, Grant Reporting and Monitoring (GRM) reports, and the outputs delivered by WB consultants. The evaluation team rated catalytic projects using a weighted average on the following criteria: relevance (30 percent), effectiveness (30 percent), efficiency (10 percent) and sustainability (30 percent).
- **Survey of FIRST Clients:** The evaluation team conducted a survey of FIRST clients using Fluid Surveys software. The survey was sent to 44 FIRST clients, with a response rate of 43 percent.
- **Validation of preliminary findings and emerging conclusions:** The evaluation team presented and discussed preliminary findings and conclusions with selected members of the FIRST GC and the PMU in February 2017. The information gleaned informed the Evaluation Report.

Limitations

Limitations to the evaluation are described below.

- Number of country visits: Due to budget constraints, in-country data collection was limited to three countries in Africa and the Middle East and North Africa (MENA) region. As a consequence, evaluation findings and conclusions may not do justice to stakeholder views and experiences in other parts of the world, namely Asia, Latin America and Central and Eastern Europe.
- Catalytic project performance and ratings: The completeness of the catalytic project reviews was limited by the quality and completeness of information found in project files, particularly the GRM reports, and by the incomplete and modest feedback from many FIRST clients through the client survey (which was disseminated in English only).
- The WB's systems, the Portal and the TF system (including GRM) do not require or provide opportunities to document changes to a project during the course of implementation.
- Some FIRST Phase III activities were at relatively early stages of development/implementation at the time of the evaluation, which limited the depth and/or scope of review. This included FIRST's gender mainstreaming approach, as well as the performance of FIRST's Programmatic Window and the FSDIP initiative.

While the processes used by the WB and IMF to implement FIRST projects and programs are similar, there are some distinctions in the terminology used by each implementing partner which may not be fully reflected in this report. The evaluation team requested PMU and implementing partners' feedback to correct any such errors and omissions.

2 Profile of FIRST

2.1 FIRST Initiative Objectives and Model

FIRST has maintained the same objective since its 2007 Charter:

> To help strengthen financial systems in low- and middle-income countries so that they make a strong and positive contribution to economic growth and help reduce poverty and income inequality (FIRST 2013 Charter, p.3),].

The guiding document for FIRST's priorities and structure is the FIRST Charter, most recently updated in July 2013 for Phase III. The 2013 Charter is complemented by the Phase III Strategy, which describes how FIRST operates and delivers technical assistance (TA), outlines the successes and challenges coming into Phase III and describes the internal activities FIRST engages in, in parallel to delivering TA, such as knowledge management and monitoring and evaluation.

FIRST is guided by the three basic principles: a demand-driven approach with open access; additionality; and harmonization, transparency, and accountability.

FIRST Sectors and Activities

Since its creation, FIRST has supported financial sector reform by providing financing for TA that is used by clients to obtain the advice they need from consultants or WB/IMF staff. FIRST relies on the WB and the IMF for all service delivery and monitoring of implementation.

FIRST finances TA for strategy and policy advice, in particular on financial sector development plans following Financial Sector Assessment Programs (FSAPs). FIRST also provides financing for advisory services to strengthen legal, regulatory and supervisory frameworks, on financial sector market and product development, and on institution and capacity building.

FIRST Business Model

FIRST is demand-driven in that it does not source projects but responds to funding requests, and uses an open, flexible model. The following definition of the Business Model was drafted based on the FIRST Phase III Strategy and consultations with the PMU during the Inception Phase and subsequently refines throughout the evaluation. FIRST's Business Model is characterized by two windows, the Catalytic Window and the Programmatic Window:

- The **Catalytic Window** was created to support rapid response projects to fill gaps. Projects respond to targeted requests in follow-up to FSAPs and other needs based on country strategies. They are single projects that focus on a particular topic and follow up, and are approved based on the expected outcomes and catalytic potential.
- The **Programmatic Window** was created in 2013 to support programs that provide sectoral support and assistance in comprehensive financial sector country reform. The programs respond to

programmatic requests for either FSAP follow-up, or for a set of connected, multi-year interventions in the context of country strategies.

FIRST Funding Model

FIRST's funding is separated in four 'buckets': LIC Catalytic, MIC Catalytic, LIC Programmatic, and MIC Programmatic. There are separate funding agreements for each bucket, and funds cannot be moved between buckets. The funding model is governed by Bank-executed trust funds; donors may withdraw from the trust funds at any time and have a proportional share of the remainder returned upon exit. The other key characteristic of the Funding Model is that since inception 85 percent of donor contributions are allocated to the WB, and the remaining 15 percent to the IMF.

FIRST Monitoring and Evaluation

In Phase III, FIRST introduced a new approach to monitoring and evaluation of FIRST interventions. Part of this exercise included the design of a results chain for the FIRST Initiative, as well as monitoring and reporting guidelines for FIRST projects and programs. While the results chain identifies outcomes and impacts for the Initiative, FIRST did not have a common, approved results framework for Phase III to use as the basis for the evaluation of the Initiative.

2.2 FIRST Governance and Management

The Governing Council (GC)

FIRST is guided by a GC comprising one representative of each of FIRST donors, a representative from the WB, and a representative from the IMF. The GC is chaired by one of the donors on a rotating basis, elected by the GC donor members. The GC meets annually and as necessary to review and approve the FIRST strategy, priorities and operating procedures, the annual strategic plan and budget, and project/program proposals to be financed. It reviews the Initiative's performance on the basis of an annual report and supporting information on the activities financed through FIRST during the previous period.

The Project Advisory Committee (PAC)

The PAC is composed of the Finance and Markets (F&M) Global Practice (GP) Senior Representative, a Senior IMF staff representative, and the Program Manager. The PAC provides information to the GC reflecting recommendations by the WB and the IMF for their respective proposals and coordinates the selection and implementation of activities. The PAC provides technical inputs into project/program design, reviews the appropriateness of activities, and makes recommendations for GC approval.

The Secretariat

The WB administers FIRST's day-to-day operations through a Secretariat housed in the WB's headquarters in Washington, DC. The Secretariat operates under applicable WB policies and procedures and is also accountable to the GC to report results of FIRST activities against targets, and abide by the 2013 Charter. The Secretariat is headed by a Program Manager and is referred to as the PMU. The Program Manager reports to the F&M GP as part of the WB hierarchy, and is accountable to the GC for the execution of FIRST's strategy, work program and budget.

2.3 FIRST Windows and Portfolio

Exhibit 2.1 FIRST Phase III Status as of 31 December 2016

WINDOW	PHASE III # OF ACTIVITIES	PHASE III APPROVALS (USD MILLIONS)			TOTAL PHASE III FUNDING	
		TOTAL	WB	IMF	USD	% COMMITTED
LIC Catalytic	66	**21.4**	18.8	2.6	31.8	67%
LIC Programmatic	10	**16.3**	13.0	3.3[1]	19.9	82%
MIC Catalytic	55	**15.2**	12.5	2.7	18.2	83%
MIC Programmatic	4	**7.0**	7.0	0	9.4	75%
Total	**135**	**59.9**	**51.4**	**8.6**	**79.3**	**76%**

2.4 FIRST Expenditures

Exhibit 2.2 FIRST Commitments and Actual Expenditures (December 2016)

COST CATEGORY	COMMITMENTS (USD MILLIONS)	ACTUAL EXPENDITURES AT 31 DECEMBER 2016 (USD MILLIONS)	ACTUAL EXPENDITURES/ COMMITMENTS
WB Project and Program Approvals	51.4	24.9	48.0%
IMF Project and Program Approvals	8.6	2.168	25.0%
PMU	18.7	9.6	51.0%
M&E	0.8	0.1	12.5%
WB Admin Fees	2.1	1.9	90.0%
Totals	**81.6**	**38.7**	**47.0%**

[1] The LIC programmatic total committed of USD 3.3 million also includes a knowledge product.

3 Conclusions

This section presents the conclusions of the evaluation, which are based on the main evaluation findings (Appendix I).

3.1 Context

The global financial crisis (GFC) in 2008 drew attention to the need for sound financial sector infrastructure and a robust regulatory framework and supervisory apparatus to support financial sector development. While most LICs and MICs were not directly affected by the proximate causes of the GFC, many were caught by the second round effects. Lessons from the GFC, coupled with financial distress in many MICs and LICs arising from the collapse of commodity prices, political instability, or poor fiscal policies, increased the demand for financial sector technical assistance spanning the foundational or prerequisite topics of financial sector infrastructure, legal and regulatory frameworks and supervisory capacity, in addition to a focus on expanding the outreach of the formal financial sector. As international standards for financial stability and consumer protection have evolved rapidly, the gaps between current and best practices have widened in many LICs and MICs and new threats to financial stability emerged, including cyber-security. These and other emerging needs (such as strategies to mitigate the effects of de-risking, personal privacy and security and consumer protection) fueled the demand by LICs and MICs for technical assistance and donor support to address such challenges.

LICs and MICs are also demanding support for financial sector reform in other areas including financial inclusion and long term finance. Given global emphasis and commitments related to increasing the access of individuals and small and micro-businesses to useful and affordable financial services, LICs and MICs require assistance to develop comprehensive national financial inclusion strategies. Given the inadequacy of long-term financing relative to demand, increasing numbers of countries are taking steps to expand their financial sector beyond deposit-taking-based intermediation which requires advisory support to create legal frameworks for issuance of capital market instruments, market conduct regulation, payments and settlement infrastructure among others.

3.2 FIRST Initiative

In this context, the objective of the FIRST Initiative, to support LICs and MICs in building stable, more efficient and inclusive financial systems is relevant to the evolving needs of LICs and MICs alike. In addition, it helps to fill the support gap for developing financial sector strategies exacerbated by the focus of FSAPS on systemically important financial sectors. Consulted stakeholders confirmed the continued relevance of FIRST principles, while noting some concerns about how the demand-driven approach principle is operationalized. Stakeholders appreciate the relative flexibility of FIRST support to respond to their evolving needs (broad scope, duration of support, funding needs) as well as its ability to accommodate the needs of MICs and LICs. The Initiative is congruent with FIRST donor priorities vis-à-vis financial sector reforms and with the priorities and expertise of FIRST's implementing partners, the WB and the IMF. Unlike other existing Trust Funds in the F&M GP, FIRST has a broad, flexible approach that accommodates multiple geographic regions, contexts and thematic foci. In our view, a broad scope is justified by the unknown challenges and opportunities that can emerge from evolving global and national contexts. While other sources of support may exist to address similar financial sector reform needs, the needs exceed the available support. While FIRST has several distinguishing characteristics, there is mixed understanding and value placed by FIRST stakeholders on the need for FIRST to have a clear brand.

The FIRST Initiative has been effective in realizing its operational targets. There is ample evidence that individual FIRST projects and programs are having positive outcomes and some catalytic effects, but it is not possible to assess the overall developmental performance of the Initiative given the absence of an approved results framework for Phase III and the limited observation period for Phase III projects that are recently completed or still ongoing. FIRST responded to Phase II evaluation recommendations deemed relevant by the GC, although the evaluation team considers a couple of recommendations still relevant, which are identified in the Evaluation Report. FIRST has responded to or is working on addressing the action items of the 2015 Rabat CG meeting.

FIRST management costs are reasonable, given the heavy emphasis on project and program monitoring and evaluation. However, the Initiative has several inefficient processes of concern to GC members and FIRST implementing partners alike, mainly related to the FIRST funding model and FIRST project and program review and approval processes. FIRST stakeholders are taking actions to address some of the most significant noted shortcomings.

While the FIRST Charter clearly articulates the roles of the GC and the PMU, interviewed WB Managers and TTLs report some variations in fulfillment of the roles and responsibilities of the PMU and implementing partners during and after project implementation, including for quality assurance, reporting and follow up. The GC engages actively in FIRST operational activities, at some cost to FIRST's strategic management. The role played by the PMU related to FIRST project and program review, approval and monitoring is generally appreciated by interviewed TTLs/IMF staff. Donors have ongoing concerns about the challenges faced by the PMU staff in managing their dual reporting relationships with the WB and the GC.

There are several concerns about the sustainability of the Initiative beyond Phase III. While FIRST donors were successful in meeting the financial target for Phase III, the GC has not been successful in attracting any new donors during Phase III. Furthermore, while FIRST's current donors have been very supportive of FIRST's work, there are relatively few of them; changes in their domestic priorities and resource constraints may have a significant effect on FIRST's sustainability. A final concern relates to any prolonged pause between the end of Phase III and a future phase; this occurred following Phase II, and contributed to significant staff turnover in the PMU and slowed or stopped project funding.

The review of management processes used to design, plan, manage, monitor and report on the FIRST Initiative identified improvements from Phase II in monitoring and reporting operational targets, but also noted several remaining shortcomings with the FIRST design. These pertain mainly to the disconnect between FIRST's demand-driven principle and its business model; the limited FIRST support for diagnostic work that can be done in projects and programs which adversely affects information needed to support financial sector reforms; the absence of an approved results framework with defined indicators to guide, monitor and report on FIRST's developmental performance as an overall; the heavy emphasis on the results of individual FIRST projects and programs and FIRST Operational Targets in tracking and reporting on FIRST performance; and the development of Knowledge Management objectives and action plans that are premised on the mistaken view that FIRST itself creates and is a repository of intellectual property.

While the FIRST Charter indicates that the PMU will prepare annual strategic plans for GC approval, to date, FIRST has relied on its multi-year strategy rather than annual workplans to guide its operations. As a consequence, there is no formal, clear basis to guide FIRST priorities and monitor implementation for the year. The PMU reports that it is working on developing an annual workplanning process.

3.3 FIRST Projects

All FIRST catalytic projects reviewed were implemented by the WB as no IMF-executed projects were complete at the time of the evaluation. A review of project documents indicates that they were highly relevant in FIRST client countries. While the majority of reviewed projects were recently completed, there was evidence that the majority generated outputs and are likely to produce preliminary outcomes. Some interviewed WB TTLs and PMU staff feel that FIRST has been instrumental in leveraging support and has played a catalytic role for follow-on initiatives by governments and other development partners. In the ratings of project effectiveness, 20 percent of projects received low ratings due to unrealistic objectives, and/or missing information in project documents, particularly GRMs.

Surveyed FIRST clients were positive about the TA received. The majority of respondents indicated that the quality of the legal-regulatory advice, strategies, methodologies and tools, and capacity building activities was Good or Very Good. Both the survey data and interviews confirmed that clients appreciate the performance and quality of WB consultants and the WB/IMF project teams, who received even higher survey ratings.

Overall, FIRST catalytic projects appear to be managed efficiently. However, based on Phase III experience to date, the prescribed timeframe for completion of catalytic projects is not realistic. While still at relatively early stages of maturity, the results of the majority of reviewed Phase III projects are likely to be sustained.

The FSDIP is an effective tool to meet client demand for assistance in developing a reform strategy in the absence of an FSAP. Potential constraints to successful FSDIPs are the Catalytic Window limit of 30 percent of the budget for diagnostic work, and the budget ceiling of USD 500,000. The limit on funding for WB staff time resulted in most FSDIP work being undertaken by teams of consultants. Removal of the limit on WB staff time will lessen the constraints seen in several projects where the authorities' timing had to be adjusted in light of the limited availability of suitably experienced consulting teams.

3.4 FIRST Programs

Reviewed FIRST programs are highly relevant to client needs for financial sector development, and build upon previous diagnostic and analytical work. The oldest FIRST programs in Rwanda and Morocco are making significant progress towards realizing planned results including several draft laws and regulations related to the new stock exchange law and real estate investment trusts law (Morocco) and amendments to laws in the insurance sector and a new law on banking (Rwanda). The efficiency of FIRST programs was affected by delays in implementation in all five FIRST programs due to internal factors (change in government, time required to pass legislation, country capacity) and external factors (an earthquake in Nepal). This raises questions about the set timeframe of three years for the Programmatic Window. The sustainability of results for FIRST programs depends on the passing of laws and reforms, which may take years, and continued capacity building of national stakeholders is required to ensure all components are implemented after completion.

FIRST pays significant attention to quality at the project and program inquiry and proposal stages. Most of those interviewed feel the approval process could be streamlined without compromising quality and the PMU is currently developing and testing a streamlined approach. While FIRST design, monitoring and evaluation systems emphasize results, established systems work more effectively for FIRST programs than for FIRST projects.

4 Recommendations

The recommendations of the evaluation, which complement suggestions made throughout the report, are grouped into those that pertain to Phase III, and those that relate to the design of any subsequent phases of FIRST.

Phase III

Recommendation 1: The GC, with PMU support, should take steps to streamline the FIRST funding model into a maximum of two buckets designated for LIC and MIC needs.

During the course of the evaluation, donors and other key stakeholders identified multiple limitations with the current FIRST funding model which has six funding buckets. This contributes to inefficiencies associated with managing so many trust funds. It also reduces the ease with which funds can be "repurposed" for FIRST programs and projects if predetermined allocations for LICs, MICs, projects and/or programs are no longer deemed appropriate. Given that FIRST's current donors have varying priorities for LICs and MICS, we suggest that the FIRST business model should comprise no more than two buckets designated for LICs and MICs.

Recommendation 2: The PMU should evaluate and report on the results of efforts made to expedite the FIRST project approval process and consider additional options to increase the efficiency of the process.

During the meeting of GC members in December 2016, the PMU proposed to pilot test a streamlined approval process for FIRST projects and programs. In addition to decisions made by the GC on that process, the GC should consider some additional refinements such as increasing the Program Manager's approval threshold so that it is in line with practices in other trust funds, and consider the possibility of establishing a "fast track" application mechanism for standard/ routine projects.

Given PMU estimates that the budget available for approval of new FIRST programs and projects will be fully allocated by March 2017 and December 2017 respectively, there is very little time to test the streamlined procedures on FIRST programs, and also little time to test them on FIRST projects. The results of such tests will likely have greater benefits for future phases of FIRST than for the current phase.

Given the limited time in Phase III to test such streamlined procedures, it is important that the PMU define the objectives of the streamlined process, and communicate the process to potential project applicants over the next 10 months. It should also evaluate the results of the process against envisaged objectives, and recommend any changes that can inform FIRST project approval processes in the future.

Recommendation 3: The GC, with support from the PMU, should review and revise FIRST roles and responsibilities so that they support a clear, more efficient and effective implementation of FIRST.

Interviewed WB stakeholders report that there is considerable variation in the fulfillment of roles and responsibilities of WB staff in FIRST project monitoring. This is partly due to established caps on WBG staff time. It may also be due to the absence of clearly defined responsibilities for the PMU and implementing partners respectively in the FIRST Charter, reviewed Administrative and Transfer Agreements. Our review of the two documents provided by the PMU indicates that roles and responsibilities are defined in very general terms (focusing on supervision and reporting); they do not clarify the respective responsibilities of the PMU and the two implementing partners.

Given plans to expedite the FIRST approval process for projects and programs, the GC and PMU should review the continued relevance of GC responsibilities related to approval, quality standards and monitoring and evaluation frameworks for FIRST projects and programs (items 4.1.5 (vii) and (x) in the 2013 Charter). The GC should consider a more hands-off role that would allow them to provide views on a no objection basis. The GC and PMU should clarify article 4.1.5 (iv) which describes GC responsibilities related to the annual strategic plan and budget, given that no such plan currently exists. If recommendations made elsewhere in this document related to adoption of an annual workplan process are accepted, an annual workplan could replace the reference to an annual strategic plan in item (iv).

Finally, the GC should assume a more strategic role in FIRST governance. During Phase III, document reviews and interviews with the GC and PMU indicate that the GC has paid considerable attention to FIRST operational matters. While the GC has the authority to determine its priorities, there is a need for the GC to pay greater attention to strategic matters such as the overall performance of FIRST, including its continued relevance and viability given the evolving contexts within which it operates, as well as the overall effectiveness and efficiency of the FIRST Initiative. While interviewed donors raised such questions during the course of the evaluation, it is not evident that these matters are sufficiently discussed during GC meetings. We encourage GC members to discuss how they may play a more strategic role vis-à-vis FIRST in the future. In this vein, one immediate priority relates to the diversification of FIRST sources of funding as discussed in Recommendation 7.

Recommendation 4: Building on the noted accomplishments and lessons in tracking the performance of FIRST projects and programs to date, the PMU should refine its current approaches to monitoring and evaluation of FIRST projects and programs so that they are more effective, efficient and support learning.

One of the notable changes between Phase II and III has been the attention paid to results-based management at the level of FIRST projects and programs throughout the project and program cycles. The PMU is to be given credit for the rigour with which projects and programs are designed.

The updated M&E framework (2016) includes a variety of strategies to track and report on results over time. The PMU has learned with experience that the effectiveness and value-added of such strategies varies. For example, the cost-benefits of FIRST client surveys are not evident. The GRM reports are not particularly helpful in documenting the planned to actual progress of FIRST projects, partly due to the timing of such reports but primarily because of the limitations arising from the way such reports are prepared. The GRM's reviewed by the evaluation team were often incomplete, missing important changes in the project over time, and did not provide the information required by the template. Despite the strong emphasis on results in FIRST design documents, the completed GRM reports paid inconsistent attention to planned/actual results and indicators.

The PMU should consider abandoning its client survey and replacing it with more formal post-reviews of projects 18 months or longer after completion, as outcomes and impacts in financial sector development are generally not observable over shorter time horizons. Given that resources are limited, the PMU should consider the value of periodic reviews of clusters of projects within selected countries (such as occurred in this 2017 external evaluation). Such reviews could be carried out by the PMU with support from thematic experts as required. Rather than waiting for an external evaluation to take place, these reviews could be staggered over time and used to assess FIRST performance in different contexts.

Recommendation 5: The GC should consider adopting a set of much more focused knowledge management objectives that are better aligned with the FIRST delivery model.

The knowledge management outputs could be better focused on the strengths of the FIRST model, provide enhanced reporting to donors and other stakeholders on success stories, and better inform GC meetings.

To the extent that additional topics can be identified where there may be unique lessons learned from FIRST projects or programs – possibly the FSDIP as a partial substitute for the in-depth review of an FSAP – it would be appropriate to plan and complete lessons learned notes.

It would also be appropriate to support preparation of a modest number of best practice notes, leveraging WB and IMF expertise and drawing on multiple FIRST-funded projects for examples. Input from the relevant technical experts in the WB and the IMF should be sought to identify suitable topics.

Additionally, FIRST could prepare and disseminate a number of project or program-specific case studies, particularly to illustrate some of the distinct features of FIRST. These could include focusing on the demand-driven aspect, and potentially the gap-filling and flexible nature of the funding by selecting projects that had sought but failed to obtain funding elsewhere.

Recommendation 6: In the interest of designing projects more effectively, FIRST caps on diagnostic work and WB staff time should be eliminated.

To date, the FIRST design imposes a limit on the diagnostic work that can be done in projects and programs, which can be a constraint in providing support for development strategies. While in some countries without recent FSAPs there may be an appropriate body of previous TA reports to provide the necessary diagnostic foundation, in other instances there will be a need for more extensive diagnostics.

In catalytic projects, projects have been subject to a cap of 30 percent of total project cost; furthermore WB TTL cost is capped at 34 percent of total project cost. These two caps have limited WB experts' ability to conduct diagnostics that are critical to the effective design of a project. In our view, these caps should be removed to accommodate the specific needs of a project.

In FSDS/FSDP and FSDIP projects this constraint has been commonly dealt with by identifying the need for a baseline study as an initial action plan item in a particular area such as financial literacy. While a reasonable work-around, this means that the action plan items are often not as detailed as would be desirable, since developing a more detailed set of actions is deferred until completion of the diagnostic work. This in turn may affect implementation, as in most cases completion of the diagnostic work will be dependent on finding development partners prepared to fund the necessary TA. To address this need, it will be important for the WB and the IMF to work together to a stronger, more systematic diagnostic approach prior to a potential Phase IV.

Future Directions

Recommendation 7: Given the generally positive results of the Phase III mid-term evaluation, the GC, with support from the PMU, should play an active role in securing funds from existing and new donors in support of a future phase of FIRST.

While the majority of FIRST projects and programs are still being implemented, this evaluation reports good to excellent progress with completed projects and programs in terms of their relevance and effectiveness, and they also have a reasonable likelihood of sustained results. Noted limitations relate to the efficiency of FIRST management processes and the modest attention paid to the FIRST Initiative as opposed to FIRST projects and programs. In our view these shortcomings can be resolved with GC and PMU actions and support.

Looking ahead, the key challenge facing FIRST relates to the sources and amounts of funding it can secure for the future. To date, FIRST's current donors have not been in a position to make financial commitments for a possible Phase IV because their own domestic priorities may be changing, and because some are waiting for the results of this Phase III evaluation. FIRST's continued reliance on a relatively small number of donors makes it vulnerable. Given that interviewed GC representatives strongly believe in the relevance and value of FIRST support to LICs and MICs, the GC could and should play a strong, proactive role in identifying funds for a future phase of FIRST from existing and new sources.

Recommendation 8: The GC, with support from the PMU, should ensure that the design of a future phase of FIRST encompasses a clear, approved results framework for the Initiative, as well as systems to monitor, report on, and evaluate its overall developmental results.

While some efforts were made to develop a results framework for Phase III, it was never approved by the GC. As a consequence, Phase III reporting has focused mainly on the performance of individual FIRST projects and programs as well as the status of FIRST operations. While such information has its own valued purposes, it does not fully support GC responsibilities to review and evaluate FIRST's overall performance, as stated in the 2013 FIRST Charter, item 4.1.5 (ix), or their stated interest in knowing the extent to which the Initiative is effective.

In line with our view that the distinction between the Catalytic and Programmatic Windows is artificial (see Recommendation 9), the evaluation team suggests that the design of a results framework for a future phase of FIRST be grounded in a set of outcomes that are common to FIRST projects and programs as there is considerable overlap, with both projects and programs likely to support long term reforms. Similarly, projects are equally as likely to support legal, regulatory and supervisory capacity business, standard setting and product development.

In developing a results framework, the GC and PMU should consider the following suggestions:

- The PMU should work with stakeholders to unbundle the implicit complex theory of change that lies behind this capacity building Initiative. For example, there are multiple steps between providing technical assistance, increasing capacities in targeted institutions, improving institutional performance, and ultimately affecting economic growth and poverty in a country. Defining these steps would help identify the length of time typically required for such changes to take place, and could simultaneously help set more realistic timeframes for FIRST funded projects and programs. In our view, the existing results chain does not adequately reflect these complexities (see sidebar for other comments).

Observations about FIRST's existing results chain

Instead of expecting that national authorities are "informed" about international good practices as in the current results chain, it would be better to target "implementing" good practices.

As is currently done in GRMs (at least the ones that are well-completed), in assessing results for the entire FIRST Initiative it would be necessary and appropriate to note reasons why there were limitations on achievement of expected outcomes.

"Improved soundness of prudential regulation and supervision" might be disaggregated into "improved legal and regulatory framework" and "enhanced capacity of regulatory and supervisory agencies." "Improved financial infrastructure" remains appropriate (although for both catalytic and programmatic).

"More long-term financing and investment vehicles" is not likely to be observable over any reasonable period during and after a project or program; progress towards this outcome should actually be observable in the other three outcomes.

- FIRST's results chain and any future results framework should not distinguish between catalytic and programmatic outputs and outcomes as there is considerable overlap, with both projects and programs likely to support long term reforms. Similarly, projects are equally as likely to support legal, regulatory and supervisory capacity business, standard setting and product development.
- Any future results framework for FIRST should include indicators that would be used to measure qualitative and quantitative progress of the Initiative over time. At present, no such indicators exist.

In our view, the use of targets in any future FIRST results framework should be selective and grounded in reality, rather than in theoretical assumptions. For example, stakeholders could consider targets that focus on the percentage of supported countries or projects/programs that enhance their capacities or performance in certain areas, rather than attempting to forecast the specific changes and varied changes in the financial sectors. Given the demand-driven nature of the FIRST Initiative, it will be difficult if not impossible for stakeholders to establish such baselines and targets for the Initiative.

The current set of outcomes reflects the challenges in identifying observable targets that can be attributed to FIRST projects and programs while recognizing that in most cases achieving outcomes is not within FIRST's direct influence. Nevertheless, it would be appropriate to include more specific outcomes, as is done in the results chain in the approval of individual programs and projects. While in some sense these may be aspirational for FIRST overall, there would be benefits to including outcomes that would help to document the performance of FIRST overall. (In March 2017, the PMU requested additional feedback from the evaluation team on how to develop an appropriate results framework for the Initiative. (Throughout April and May 2017, the evaluation Team Leader worked actively with the PMU providing feedback on the development of the results framework.)

In addition, in designing any future phase of FIRST, we suggest that the PMU and the GC members review the need for and role of operational targets in management and reporting practices. In our view, the current Operational Targets list is very activity-based and process-oriented and while it may provide a checklist for the PMU in managing FIRST, it does not reflect FIRST's commitment to embracing results-based management.

Once FIRST has a results framework, it should also pay explicit attention to monitoring, evaluating and reporting on the overall FIRST Initiative, including its relevance, effectiveness, efficiency and sustainability, and take measures to make its reports more analytical, commenting on, for example, notable trends and lessons learned.

Finally, given that it takes time for projects and programs to achieve outcomes and demonstrate impacts, the terms of reference for future external evaluations should permit reviews of FIRST's cumulative support (i.e., across FIRST phases) to selected countries, rather than limiting the evaluation scope to those activities funded in the current phase.

Recommendation 9: In designing any future phase of FIRST, the GC, with support from the PMU, should takes steps to ensure that the FIRST design and practices are based more on client needs and country context.

The evaluation found that there is a significant contradiction between the FIRST's principle that emphasizes the demand driven nature of FIRST, and FIRST's current design and operational practices. This is evident in how FIRST Operational Targets are defined, as well as in how several of the criteria and guidelines for project/program design are articulated. In our view, the FIRST business model should reflect FIRST principles more closely: for example, project/program size, duration and targets should be realistic and reflect client needs and objectives, rather than being artificially designed to reflect predetermined criteria. FIRST should also accommodate a stronger, more systematic diagnostic approach rather than relying on pre-determined, inflexible support formulae. Finally, in our view, it is not evident that the current distinction between the Catalytic and Programmatic Windows is necessary or helpful. We encourage the PMU to identify a more fluid model that accommodates varying needs of clients.

In addition, in keeping with the principle of demand-driven programs, there continues to be a rationale for FIRST to have a broad scope given the unknown challenges and opportunities that can emerge from evolving global and national contexts. While other sources of support may exist to address similar financial sector reform needs, the needs exceed available support.

Recommendation 10: FIRST should continue to offer the FSDIP product in a prospective Phase IV.

In our view, the FSDIP is particularly relevant for LICs and smaller MICs that are unlikely to be targeted for FSAPs. Lessons from the earlier experience with FSDS/FSDP projects are already reflected in the FSDIP selection criteria. Possible enhancements to the program would include expanded funding for diagnostic work, possibly in the form of a pre-approved budget contingency for diagnostic studies to be identified as review of prior TA work and stakeholder consultation takes place during the initial stages of the FSDIP. To address this need, it will be important for the WB and the IMF to work together to a stronger, more systematic diagnostic approach prior to a potential Phase IV. Removal of the hard cap on catalytic projects and the sharp distinction between Catalytic and Programmatic Windows recommended in Recommendation 9 would also help to ensure that FSDIPs are designed wholly to meet client needs rather than being tailored to fit within the selection criteria.

5 Lessons Learned

Initiative, program, project design

- To be effective, it is important for projects and programs to have a design that is realistic given available resources, expected timeframes, and partner capacities.
- The relevance of projects and programs is enhanced when they are grounded in an adequate diagnosis of client needs and context.
- An initiative's potential performance is enhanced when an agreed and commonly known basis for assessing its performance (including clearly articulated results, indicators and/or targets) is used for monitoring and reporting over time.

Managing for sustainable results

- National stakeholders are more likely to own and support program and project plans if they are appropriately consulted in the design and review of such projects and programs over time.

Evaluation design

- Evaluation results are most meaningful /useful when the evaluation's objectives are congruent with the stage of evolution of the project, program or initiative and the evaluation scope. For example, a mid-term evaluation is not likely to identify impacts in new projects, program and/or initiatives.
- To be most useful to stakeholders, evaluation TORs should prioritize what the client is most interested in, and therefore how the evaluation budget should be allocated.

Appendix I Evaluation Findings

The key evaluation findings are presented below.

Finding 1: The majority of reviewed Phase III projects were highly relevant in client countries.

Finding 2: The majority of FIRST projects were effective in that they have generated outputs and are likely to produce preliminary outcomes. The ratings of overall effectiveness of some projects were affected by unrealistic objectives and limited information on results in the GRMs.

Finding 3: Overall, FIRST catalytic projects appear to be managed efficiently. However, based on Phase III experience to date, the prescribed timeframe for completion of catalytic projects is not realistic.

Finding 4: While still at relatively early stages of maturity, the results of the majority of reviewed Phase III projects are likely to be sustained.

Finding 5: The FSDIP is an effective tool to meet client demands for assistance in developing a reform strategy in the absence of an FSAP.

Finding 6: Reviewed FIRST programs are highly relevant to client needs for financial sector development, and build upon previous diagnostic and analytical work.

Finding 7: The oldest FIRST programs are making significant progress towards realizing planned outputs and some outcomes, demonstrating early signs of effectiveness.

Finding 8: The efficiency of FIRST programs was affected by delays in all five of the programs reviewed.

Finding 9: The sustainability of results of many FIRST programs depends on the passing of laws and regulations, which may take years, and on the capacity of national stakeholders to ensure all components are implemented after completion.

Finding 10: FIRST pays significant attention to quality at the project and program inquiry and proposal stages. Most of those interviewed feel the approval process could be streamlined without compromising quality.

Finding 11: While FIRST design, monitoring and evaluation systems emphasize results, established systems work more effectively for FIRST programs than for FIRST projects.

Finding 12: The objective and principles of FIRST remain very relevant given the financial sector reform needs and resources available to support the needs of LICs and MICs.

Finding 13: FIRST’s broad scope, geographically and thematically, as well as its flexibility to respond to country financial sector reform needs, distinguishes it from existing WB Trust Funds in the F&M Global Practice. Moreover, the demand for financial sector stability strengthening exceeds additional support that may become available if and when the FSSF Trust Fund is established by IMF.

Finding 14: The limited number of FSAPs, combined with FIRST budget caps on diagnostic work that can be done in projects and programs, adversely affects the availability of diagnostic information to inform country financial sector reforms.

Finding 15: There is mixed understanding and value placed by stakeholders on the need for a FIRST brand.

Finding 16: While there is ample evidence that FIRST projects and programs are producing a broad range of outputs for FIRST clients, having positive outcomes, and resulting in some catalytic effects, it is not possible to assess the overall performance of the Initiative given the absence of an approved results framework.

Finding 17: FIRST has been generally effective in meeting its operational targets.

Finding 18: FIRST responded to Phase II evaluation recommendations deemed relevant by the GC, although the evaluation team considers the other recommendations are still relevant. FIRST has responded to or is working on addressing the action items of the 2015 Rabat Consultative Group meeting.

Finding 19: FIRST management costs are reasonable, given the heavy emphasis on project and program monitoring and evaluation.

Finding 20: While the costs of FIRST management are reasonable, the Initiative has several inefficient processes of concern to GC members and FIRST implementing partners alike.

Finding 21: GC members engage actively in FIRST operational activities at some cost to FIRST's strategic management.

Finding 22: The role played by the PMU Secretariat in supporting FIRST project/program design and approval is appreciated. There are some concerns about the roles of the PMU and TTLs/IMF staff during and after project/program implementation.

Finding 23: While funding for the duration of Phase III is secured, funding for future phases of FIRST is less certain.

Finding 24: As currently designed, the FIRST business model lacks sufficient congruence with FIRST's demand-driven principle and it is not evident that the distinction between the Catalytic and Programmatic Windows is necessary.

Finding 25: While FIRST's operational targets are used in tracking performance, stakeholders have some concerns about the appropriateness and number of targets, and the focus on operational targets contributes to a culture that emphasizes activities rather than results.

Finding 26: The design of Phase III has a strong emphasis on results at the project and program levels but pays limited attention to the overall results and performance of the FIRST Initiative.

Finding 27: The Knowledge Management Action Plan is premised on the mistaken view that FIRST itself creates and is a repository of intellectual property.

Finding 28: FIRST is in the process of developing an approach to mainstreaming gender in financial inclusion projects. While the approach is reasonable, it is premature to assess the effectiveness and results of this approach.

Finding 29: To date, FIRST has relied on its multi-year strategy rather than annual workplans to guide its operations. It has not introduced any systems to guide its activities on either a semi-annual or annual basis.