



Ministry for Foreign Affairs
of Finland MFA



Department of Foreign Affairs of Ireland DFA



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Foreign Affairs FDFA
State Secretariat for Economic Affairs SECO

PEER REVIEW OF FOUR EVALUATION FUNCTIONS

Development Evaluation Unit, Ministry for Foreign Affairs, **Finland**

Evaluation and Audit Unit, Department of Foreign Affairs and Trade, **Ireland**

Evaluation and Corporate Controlling Division, Agency for Development Cooperation, Ministry of
Foreign Affairs, Switzerland

Evaluation Unit, Quality and Resources Section, Economic Cooperation and Development
Division, State Secretariat for Economic Affairs, **Switzerland**

November 2020

Consultant:

Patrick G. Grasso, Alexandria/USA

pgrasso45@comcast.net

SUMMARY

Background

Purpose of the Peer Review: Provide an impartial assessment of the evaluation functions of development agencies in Finland, Ireland, and Switzerland (in both SDC and SECO) in terms of the extent to which they meet internationally-accepted standards on: the enabling environment for evaluation, the independence of the evaluation system, the credibility of evaluations, and their utility and use.

Methods: Document reviews, key informant interviews, in-depth discussions with evaluation staff, and reviews of selected evaluation reports.

Peer reviewers: Directors of the evaluation units in Finland, Ireland, and Switzerland (SDC) reviewed each other in two-member teams; the director for SECO participated in study design and report reviews. A senior consultant assisted all three teams in common.

Major findings and recommendations

In general, the four units have favorable enabling environments, though with some areas for improvement, as discussed in the respective chapters

Across the four organizations, evaluation is accepted as an important contributor to ensuring projects and programs are meeting their objectives and proving beneficial in terms of development work. However, to the extent evaluation impinges on the related areas of foreign policy and trade there is less understanding of and support for the evaluation function. This is the case in both Finland and Ireland, though not in Switzerland.

Recommendation: To improve the enabling environment for evaluation in non-development areas in the countries where this issue arises, the evaluation units reviewed should consult with similar agencies in other countries, which have been successful in achieving this goal.

Across all four organizations independence is generally strong, but some areas could be strengthened, particularly policies on working with outside evaluators and preventing conflicts of interest

The evaluation units all are organizationally independent within their Ministries or Departments, and have demonstrated behavioral independence through candid reporting. However, policies and protocols are somewhat unclear on the extent to which unit staff can participate directly in the conduct of evaluations, and on specifically evaluation-related conflict of interest standards for external evaluators.

Recommendation: The evaluation units should develop policies or protocols that specify the purposes, extent, and limits to active participation by unit staff in the conduct of evaluations (e.g., joining field work, conducting interviews, contributing to products, etc.).

Recommendation: The evaluation units should ensure that their policies or protocols address specific issues of:

- conflicts of interest in evaluation, including those that may arise from external consultants' past or planned future work; and
- the recruitment and retention of staff with evaluation knowledge and skills, including possible exceptions to general policies of staff rotation in the organization.

The credibility of centralized evaluation is high in terms of quality, staffing, other resources, purposes, and use, but decentralized evaluation is less visible, so gets less attention

Stakeholders report that centralized evaluations from the four unit are regarded as highly credible. But much evaluation work is decentralized under the control of operating units. There is much less visibility around these evaluations, and even ambiguity about what counts as an evaluation at all.

Recommendation: The evaluation units should work with Management to ensure that evaluation policies are clear on what constitutes a decentralized evaluation, the quality standards for such evaluations, and the processes for assessing quality.

Finally, at all the organizations evaluation work by the evaluation function was perceived to be useful and was used by operational staff and Management, though this was mostly at the policy rather than operational level.

While Senior Management reported that they found the units' evaluation work useful, and actively used it, operational units reported it less relevant to their work. In addition, all four of the units raised concerns about reaching broader audiences for whom their evaluations could be useful.

Recommendations: evaluation units should consider strengthening the utility of their evaluations by, inter alia:

- using the planning process to:
 - ensure that the objectives and scope of the evaluation are clearly aligned with the information needs of identified internal or external stakeholders and audiences;
 - set priorities among those information needs to ensure maximum utility;
 - work with the evaluation team to clarify what data are needed to meet these information needs;
 - agree with the evaluators on how evaluation unit staff can participate in evaluation activities so they can ensure the work remains focused on those information needs; and
 - identify opportunities to employ evaluation findings and products for multiple purposes (including inputs to other evaluations, where appropriate) involving multiple stakeholders and audiences.
- improving communications and dissemination through:
 - developing a communication strategy that addresses such issues as how to reach internal and external stakeholder audiences;
 - including a communication/dissemination plan for individual evaluations from the start.
- strengthening the Management Response process by:
 - clarifying roles and responsibilities for Management and the evaluation unit;
 - setting timelines for both the formal response and agreed actions;
 - enhancing follow-up procedures, including tracking of actions taken; and
 - providing for regular reporting on progress and any shortcomings in implementation.

TABLE OF CONTENTS

SUMMARY	i
ABBREVIATIONS AND ACRONYMS	vi
OVERVIEW	1
Introduction: Why Conduct an Evaluation Function Peer Review?	1
Main Issues for the Review	1
Approach and Methods	4
Peer Review Team	5
Organization of this report	5
CHAPTER 1: CROSS-AGENCY FINDINGS AND CONCLUSIONS	6
Enabling environment for evaluations	6
Independence of evaluations and evaluation systems	7
Credibility of evaluations.....	12
Utility of evaluations.....	13
CHAPTER 2: EVALUATION IN FINLAND	16
Enabling Environment for Evaluation	18
Independence of Evaluations and the Evaluation System.....	19
Credibility of Evaluations	21
Utility of Evaluations	24
CHAPTER 3: EVALUATION IN IRELAND	28
Enabling Environment for Evaluation	29
Independence of Evaluations and the Evaluation System.....	32
Credibility of Evaluations	33
Utility of evaluations.....	35
CHAPTER 4: EVALUATION IN SDC	38
Enabling Environment for Evaluation	39
Independence of Evaluations and the Evaluation System.....	41
Credibility of evaluations.....	44
Utility of evaluations.....	46
CHAPTER 5: EVALUATION IN SECO	49
Enabling Environment for Evaluation	51
Independence of Evaluations and the Evaluation System.....	52
Credibility of evaluations.....	53
Utility of evaluations.....	55

ANNEX I: TERMS OF REFERENCE.....	57
ANNEX II: NORMATIVE FRAMEWORK	62

BOXES

1. Key Roles and Activities to Ensure a High-Quality Evaluation	8
2. Results-Based Management in MFA	14
3. Finland’s Evaluation Help Desk	18
4. Response to Forced Displacement	19
5. The Nexus of Development Assistance and Humanitarian Aid	48

FIGURES

1. Organization of the Ministry for Foreign Affairs, Finland	16
2. Ireland DFAT Organization Chart	28
3. EAU process overview	34
4. SDC Organization Chart...	38
5. Governance Chart for Evaluation at SECO WE	49

TABLES

1. Staffing and Costs for Evaluation Units	11
2. Roles for Audit and Evaluation	31
3. Criteria for Selecting Evaluations	36
4. SECO Management Response Follow-up Record	56

ABBREVIATIONS AND ACRONYMS

CLP	Core Learning Partnership
DAC	Development Assistance Committee
DFAT	Department of Foreign Affairs, Ireland
E+C	Evaluation and Controlling Division, SDC, Switzerland
EAU	Evaluation and Audit Unit, Ireland
ECG	Evaluation Cooperation Group
EF	Evaluation Function
EMS	Evaluation Management Services, Finland
EO	Evaluation Officer, SECO, Switzerland
EU	Evaluation Unit, SECO, Switzerland
EVA	Unit for Development Evaluation, Finland
EvalNet	OECD-DAC Evaluation Network
FCV	Fragility, Conflict, and Violence
IEG	Independent Evaluation Group, World Bank Group
KEPO	Development Policy Steering Group, Finland
MFA	Ministry of Foreign Affairs, Finland
OECD	Organization for Economic Development and Cooperation
SDC	Agency for Development Cooperation, Switzerland
SECO	State Secretariat for Economic Affairs, Switzerland
ToR	Terms of Reference
UNEG	United Nations Evaluation Group

OVERVIEW

The primary purpose of this Peer Review of the evaluation functions (EFs) of the development agencies of Finland, Ireland, and Switzerland is to provide an impartial assessment of each evaluation function, its strategic fit in the member Ministry, Agency, or Division, and both its good practices and areas for improvement to better support achieving the strategic objectives of development and development evaluation. The review is intended to address four overarching issues, as identified in the Terms of Reference (ToR): (1) the enabling environment for evaluations, (2) the independence of evaluation and the evaluation system, (3) the credibility of evaluations, and (4) the utility of evaluations.¹ In addition, each of the EFs also had specific issues that were addressed in its review.

Introduction: Why Conduct an Evaluation Function Peer Review?

The Organization for Economic Co-operation and Development's Development Assistance Committee (OECD-DAC) Network on Development Evaluation (EvalNet) encourages professional Peer Reviews of the evaluation function of its member countries and organizations as a way to "strengthen the evaluation function and promote transparency and accountability in development agencies."² These reviews are carried out to identify good practices and opportunities to further strengthen the evaluation units under review, with a view to contributing ultimately to improved performance in international development cooperation and humanitarian assistance.³

In light of this guidance, the evaluation functions of the development agencies of Finland, Ireland, and Switzerland joined together to conduct this Peer Review. The advantage of combining three Peer Reviews is that it allowed not only findings, conclusions, and recommendations related to each of them, but also cross-agency learning that can be beneficial for the broader development evaluation community. This review is the first Peer Review of the respective EFs outside the regular DAC Country Peer Review process.

Main Issues for the Review

Based on the Normative Framework outlined in Annex II of this report, the following main issues were used to guide the analysis and reporting for the Peer Review. Each chapter is structured around these issues.

Enabling environment for evaluations: The enabling environment includes such factors as requirements for explicit consideration of evaluation findings in developing programs and projects;

¹ These are defined in United Nations Evaluation Group, *Revised UNEG Guidelines for Professional Peer Reviews*, 2016.

² OECD-DAC Network on Development Evaluation, *Evaluating Development Co-operation: Summary of Key Norms and Standards*, 2nd edition, 2010. <http://www.oecd.org/dac/evaluation/dcdndep/41612905.pdf>

³ OECD-DAC Network on Development Evaluation, *Peer Review Reference Guide, 2019-20*, 2019. <https://www.oecd.org/dac/peer-reviews/DAC-peer-review-reference-guide.pdf>.

the degree to which such requirements are carried out in practice; the extent to which operational staff seek input from the evaluation office when considering new programs or projects; and the institutional support provided to the evaluation function by the agency in terms of the adequacy of financial and human resources, access to information, and willingness to cooperate in evaluation activities.

This is an important issue for evaluation because in the absence of a strong enabling environment the independence, credibility, and utility of evaluation work are likely to have little impact on the organization, its work, or its development outcomes, the *raison d'être* of evaluation.

Independence of evaluations and evaluation systems: The credibility of evaluation depends on independence to assure that the findings, conclusions and recommendations are as unbiased as possible. Independence has four major dimensions: organizational, behavioral, avoidance of conflicts of interest, and protection from external influence.⁴

- *Organizational independence* has to do with where in the parent agency the evaluation function is located, especially its reporting relationship to agency Management or some other body, such as an external board.
- *Behavioral independence* refers to the extent to which the unit has been willing and able to carry out its functions with candor, even when its findings may be unwelcome to the broader organization or its Management.
- *Avoiding conflicts of interest* focuses mainly on ensuring that those responsible for conducting and managing any given evaluation are not materially compromised because of current, previous, or projected relationships with the activities being evaluated or those in charge of them.
- *Protection from outside influence* involves assurance that the evaluation function is not unduly subjected to pressure from outside, including from Management, governing bodies, or other stakeholders, in the conduct of its evaluations. Of course, stakeholders will have an incentive to try to influence the conduct and findings of evaluations. But the issue here is whether the evaluation function has been given sufficient protection from these pressures to carry out its evaluation work without prejudice.

Credibility of evaluations: The issue of credibility is at heart about the quality of the evaluation products and the system in place to ensure high quality. Work of strong quality is likely to be seen as credible, and without credibility evaluations are not likely to lead to use by the development organization. Questions about credibility focus on whether systems are in place to ensure the professional competence of evaluators and the technical quality of the evaluations (design, methods, and conduct), the transparency and impartiality of the evaluation work, and the clarity of communication of evaluation results.

International development organizations such as the OECD-DAC Evaluation Network (EvalNet), UN Evaluation Group (UNEG), and Evaluation Cooperation Group (ECG) of the international

⁴ These criteria first were derived in World Bank Operations Evaluation Department, "Independence of OED" (February 24, 2003). <http://documents.worldbank.org/curated/en/912291468765325239/pdf/4200ED1Reach.pdf>

financial institutions (including the World Bank, Regional Development Banks, and International Monetary Fund) have developed quality standards for the conduct of evaluations.⁵ These standards are broadly similar and provide a sound basis for assessing the quality of evaluation products. They focus on assuring the competence of evaluators, use of appropriate methods, ethical and balanced treatment of stakeholders in the operations being evaluated, and systems of quality assurance by the evaluation unit.

Utility of evaluations: At the end of the day, evaluation is worthwhile only to the extent that it is useful and used.⁶ The import of this criterion is that to be useful (and used) evaluations must pay careful attention to the:

- multiplicity of audiences they necessarily address, such as parliamentary bodies, governing or oversight boards, senior management, operational staff, development partners, beneficiaries, the research community, and the general public;
- divergent information needs of those audiences, including which should be prioritized; and
- timeliness of their work to ensure it is available when needed, for example to inform emerging strategies or design new operations.

The key issue for the evaluation functions reviewed here is the extent to which their parent organizations have taken on board the findings, lessons, and recommendations from their evaluations and acted on them.

Evaluation needs to be seen as a continuous process, from issue identification through to institutional and individual learning. These criteria help to guide an assessment of how the EFs under review manage that process so as to produce relevant, independent, credible evaluations that are recognized by the relevant stakeholders as being of high quality and useful – and that are used.

⁵ OECD-DAC Network on Development Evaluation, *Evaluating Development Co-Operation: Summary of Key Norms and Standards*, Second Edition (2010), [link](#); United Nations Evaluation Group, *Norms and Standards for Evaluation* (2016), [link](#); Evaluation Cooperation Group, *Big Book on Evaluation Good Practice Standards* (2012), [link](#).

⁶ Patrick G. Grasso, "What Makes an Evaluation Useful? Reflections from Experience in Large Organizations," *American Journal of Evaluation* 24 (December, 2003). <https://journals.sagepub.com/doi/pdf/10.1177/109821400302400408>. There is an extensive critical literature on this approach, but it has been widely accepted within the development community.

Approach and Methods

In carrying out the review, the following tools and approaches were used:

Document reviews

A key source of information for the review is the documentary record on evaluation organization and management in each institution. Among the documents reviewed were:

- institutional policies on evaluation conduct and use;
- the Evaluation Function's mandate and related statements on organization, roles and responsibilities, reporting lines, staffing, strategy, and partnership agreements;
- plans, work programs, and supporting materials; and
- other documents deemed relevant by the Evaluation Function, Management, or other stakeholders.

Key informant interviews

Interviews provide some of the most important information for reviews of this kind because documentary evidence almost never allows for a complete understanding of what is happening on the ground. In particular, stakeholders are far more likely to be candid in interviews than in written comments or other documents. Panel members worked with their respective institutions to identify interviewees for different categories of stakeholders (Management officials, evaluation unit heads, operational staff, external stakeholders). A semi-structured interview guide was developed for each site and interviewee category.

In-depth discussions with evaluation unit staff

In addition to the key informant interviews the Panel members conducted group discussion sessions with staff of the evaluation unit for each organization, including both evaluators and support staff. This was an efficient and effective way to surface internal issues on the operation of the unit and its relationships with other stakeholders that are unlikely to be raised through other forms of data collection.

Review of selected evaluation reports

Finally, the panel reviewed a small sample of reports from each of the agencies for evaluations conducted during the study period; though the terms of reference set this period as 2010-19, the reports reviewed mostly were published after 2015. Key informant interviews helped to identify cases for this review, largely on the basis of their positive or negative receptions or other significant factors.

Peer Review Team

The review was carried out by the heads of four evaluation offices, with teams of two acting as peers for the third; a consultant served on all of the three teams. Members included:

- Patrick Empey, Deputy Director of Evaluation, Evaluation and Audit Unit (EAU), Department of Foreign Affairs (DFA), Ireland.
- Anu Saxén, Director, Development Evaluation Unit (EVA), Ministry for Foreign Affairs (MFA), Finland.
- Christoph Jakob, Deputy Head, Evaluation and Controlling Division (E+C), Swiss Agency for Development Cooperation (SDC), Ministry of Foreign Affairs, Switzerland.
- Johannes Schneider, Head of the Evaluation Unit (EU) and Deputy Head of Quality and Resources Section, Swiss State Secretariat for Economic Affairs (SECO), participated in planning the review and commenting on drafts.
- Patrick G. Grasso, former Advisor to the Director of the Independent Evaluation Group at the World Bank, acted as external consultant.

Peer Review Country Visits

- **Finland (MFA), December 2019**
Peers: Christoph Jakob and Patrick Empey, supported by Patrick G. Grasso
- **Switzerland (SDC and SECO), January 2020**
Peers: Anu Saxén and Patrick Empey, supported by Patrick G. Grasso
- **Ireland (EAU), March 2020**
Peers: Anu Saxén and Christoph Jakob, supported by Patrick G. Grasso

Organization of this report

This report consists of five chapters: a summary of the findings, conclusions, and recommendation from across the four evaluation agencies reviewed, and a separate chapter for each of those agencies. The Annexes provide the Terms of Reference and the Normative Framework for the review.

CHAPTER 1: CROSS-AGENCY FINDINGS AND CONCLUSIONS

One of the reasons for doing this Peer Review across four evaluation functions in three countries was the opportunity for the units to learn lessons and develop ideas for strengthening their own work from directly examining each others' experiences. This chapter presents the major conclusions (in bold) on each of the major issues cited in the Overview that were drawn from this exercise, along with the findings supporting each.

Enabling environment for evaluations

In general, the four units have favorable enabling environments, though with some areas for improvement, as discussed in the respective chapters

Operational managers and staff of the agencies working on development typically are conversant with the need for evaluation, and express support for the work of the evaluation units. Many have served on reference groups that provide inputs to and comments on evaluations at various stages of the work. This has familiarized them with how evaluations are carried out, and appears to have fostered understanding of both the process and the outcomes of evaluation work. Most reported that they have found evaluations useful, though how they have used the results was not always clear.

However, staff with backgrounds in diplomacy and policy, rather than development, often were less clear about, and supportive of, evaluation efforts. These are areas where the routines of evaluation, common to development agencies, are newer and less well-developed. Interviewees did not express hostility to evaluation, but rather a concern about how it might not fit as well with diplomacy and foreign policy activities as with development. A case in Finland highlights the issue; see Chapter 2. The issue also arises in Ireland, though not Switzerland.

However, in some other countries this works well. For example, a peer review of the Policy and Operations Evaluation Department (IOB) in The Netherlands found that its "evaluation system covers policy and operations, process and impact, and a range of interests and needs," and concluded that it "is an example of a world-class independent evaluation unit."⁷ Similarly, in Global Affairs Canada's "whole-of department approach, staff responsible for evaluating international assistance work side by side with colleagues evaluating foreign policy and international trade, and a strategy is in place to enhance learning across branches."⁸ These cases suggest that in the countries where this issue comes up, the evaluation units reviewed here may be able to learn from such successful cases.

Recommendation: To improve the enabling environment for evaluation in non-development areas in the countries where this issue arises, the evaluation units reviewed should consult with similar agencies in other countries, which have been successful in achieving this goal.

⁷ OECD Development Cooperation Peer Reviews: Netherlands 2017, p. 20.

⁸ OECD Development Cooperation Peer Reviews: Canada 2018, p. 2.

Independence of evaluations and evaluation systems

Across all four organizations independence is generally strong, but some areas could be strengthened, particularly policies on working with outside evaluators and preventing conflicts of interest

Organizational independence: The DAC Evaluation Network notes that independence “will best be achieved by separating the evaluation function from the line management responsible for planning and managing development assistance. This could be accomplished by having a central unit responsible for evaluation reporting directly to the minister or the agency head responsible for development assistance, or to a board of directors or governors of the institution.”⁹

The four EFs reviewed in this report provide variants on this guidance, but in each case they report to officials sufficiently removed from the activities being evaluated to meet the standard of organizational independence.

Behavioral independence: The review found that all four units were behaviorally independent. This was confirmed through key informant interviews, discussions with evaluation unit staff, and reviews of evaluation reports. Critical findings and hard-hitting reports were common across all four, and there were no indications that any of them were failing to meet behavioral independence standards.

During the review, the team was asked to consider whether active participation by EF staff in ongoing evaluations would compromise independence, especially behavioral independence. However, independence inheres in the organizations themselves, not in the external evaluators. Participation in the evaluation process does not compromise independence, but can bring to the work institutional knowledge that external evaluators often lack, making the final product more likely to be relevant, useful, and used. As the former World Bank Director General of Evaluation, Robert Picciotto, has argued, “having no connection or shared experience with intended users of evaluation findings, constrains access to information, evinces resistance and inhibits learning.”¹⁰

What is key is that the roles of the internal evaluation managers and external evaluators be clear. For example, Box 1 (below) was developed by the World Bank Group’s Independent Evaluation Group (IEG) as guidance for development evaluation units.¹¹ It offers practical steps that they can implement to manage evaluations without compromising independence. This requires a clear policy and practices that define the purposes, extent, and limits of such participation. However, current policies and protocols across the EFs are not clear on this issue, a gap that should be addressed.

⁹ *Evaluating Development Co-operation*, p. 8.

¹⁰ Robert Picciotto, “Evaluation Independence in Organizations,” *Journal of Multi-Disciplinary Education*, 9:20, pp. 18-31 (2013). https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/373/368

¹¹ World Bank, Independent Evaluation Group, *Managing Evaluations: A How-to Guide for Managers and Commissioners of Evaluation*, 2015, p. 27.

Recommendation: The evaluation units should develop policies or protocols that specify the purposes, extent, and limits to active participation by unit staff in the conduct of evaluations (e.g., joining field work, conducting interviews, contributing to products, etc.).

Box 1: Key Roles and Activities to Ensure a High-Quality Evaluation

The evaluation manager should:

- Ensure that the objectives of the evaluation are clear;
- Maintain ownership of the study by ensuring that decision-making responsibility is retained and that decisions are made in a timely manner;
- Negotiate expectations with stakeholders;
- Monitor the progress of the evaluation and provide relevant and timely feedback and guidance to the evaluator and evaluation team;
- Be open to suggestions from evaluators on possible solutions if problems arise;
- Discuss and ensure agreement on formal and informal communication protocols from the beginning;
- Ensure evaluators have full access to information as early as possible;
- Meet with evaluators and stakeholders to discuss draft reports and revisions;
- Approve the final report and organize a presentation of evaluation findings for stakeholders.

The evaluator (or evaluation team) should be encouraged to:

- Commit to conducting the evaluation within the allotted time frame and budget;
- Provide regular progress reports to the evaluation manager and communicating problems that require the attention of the evaluation manager immediately;
- Discuss the draft report and correct any errors or misinterpretations;
- Respond to comments and finalize the report.

Source: Independent Evaluation Group 2015, cf. footnote 11.

Conflicts of interest: Another area where existing evaluation policies across the four evaluation functions are not always clear is conflicts of interest. Such conflicts may arise from a number of sources:

For example, individuals who may have worked on a project or program years before may seem like good candidates to be involved in an evaluation. However, managers need to be careful because these types of situations can yield biased results. If the manager is not aware of the person's past involvement, the person being approached for the work should be forthcoming and offer to recuse him or herself from conducting or working on the evaluation.¹²

A related form of conflict arises if the individual or firm contracted to conduct an evaluation also has or is likely to be competing for other contracts with the unit responsible for the activity being evaluated. This would raise at least an appearance of conflict, creating risk to the perceived (or

¹² Independent Evaluation Group, *Managing Evaluations*, p. 30.

actual) independence of the evaluation, and raising a reputational risk for the EF. Given the finite supply of qualified development consultants and contractors, these kinds of conflicts are not uncommon. EFs typically rely on institutional procurement policies to fulfill this requirement, as in this reply to an inquiry from a potential contractor:

“If [an organization] is selected as a provider of administrative service for commissioned studies, it needs to follow the laws related to procurement. Considering that the selection of the performer of the commissioned study is finally decided by the [government agency], it is possible for the researchers from the same [organization] to be a performer of the commissioned study. However, the persons functioning as the providers of the administrative service and the persons authorizing their work are themselves disqualified. If the provider of the administrative service proposes experts from the same [organization], the provider of administrative service needs to sign a statement of no conflict of interest.”

However, such broad procurement policies generally do not address specifically the kinds of conflicts that affect evaluations. Therefore, at a minimum it would be useful if EFs supplemented procurement rules with specific policies and protocols for identifying and assessing potential evaluation conflicts, and to mitigate them where possible. These could be worked out with the institutional procurement offices to ensure they are in accordance with institutional rules and applicable statutes. A nice statement of the kinds of issues that should be covered is provided by the US Department of Energy:

Affiliations or activities that could potentially lead to conflicts of interest may include the following:

- Work or known future work for parties that could be affected by the individual's judgements on projects or program developments that the individual has been asked to review;
- Any personal benefit the individual might gain (or benefit of their employer, spouse or dependent child) in a direct or predictable way from the developments of the program/projects they have been asked to review;
- Any previous involvement the individual has had with the program/projects they have been asked to review, such as having participated in a solicitation to the program area that was subsequently not funded, or having a professor, student, or collaborator relationship with the program or its research staff;
- Any financial interest held by the individual (or their employer, spouse, or dependent child) that could be affected by their participation in this review; and
- Any financial relationship the individual has or had with [the organization], such as participation in research grants or cooperative agreements.¹³

Another potential source of conflict is implicit in the personnel rotation system used across the evaluation units. This can make it hard to recruit and retain staff with the right mix of evaluation,

¹³ US Department of Energy, “Program Evaluation: Independence, Conflict of Interest, Openness” (undated).

development, and institutional knowledge to meet the unit's needs, and could provide incentives for staff to be wary of the possible effects of hard-hitting reports on their career prospects.

Evaluation knowledge among staff is important primarily to allow them to contribute to the evaluation work, and to provide strong quality control to assure it is of high technical quality. Development knowledge, gained mostly through experience, helps the team to identify the key issues that require evaluation, and to ensure that evaluation reports are grounded in sector/thematic and country/regional realities. Institutional knowledge is a key to helping the evaluation team conduct its work as smoothly and efficiently as possible, and ensuring that recommendations and lessons from evaluation take account of the institution's capacity to put them into practice. Not everyone in the evaluation unit must have a strong background in all three areas, but the team as a whole needs to cover this ground. And the work environment within the group must foster sharing of relevant knowledge across the team to promote high performance in all three areas. Given the small size of the units, and the institutional rotation policies generally in place, one or two staff with strong evaluation training/experience who can act as a resource for the rest of the staff should be sufficient, but other members should be afforded adequate training to hone their own skills.

This has implications for the recruitment, training, and retention of evaluation staff. The unit may not always be able to recruit many staff with pre-existing evaluation expertise. As a result, it is important that evaluation units invest in training staff to at least a moderate proficiency in evaluation so that they can oversee the work of consultants. The extent to which such training can be provided depends on the resources that can be made available, given the multiple demands on the unit's budget. But this should be given as much attention as is feasible within that constraint. For example, Ireland's EAU staff all have a Masters degree in a relevant discipline, have attended the International Program in Development Evaluation Training (IPDET), and completed other evaluation training courses.

Retention is a more complex issue because without a career path evaluation staff move to other parts of the organization, bringing with them whatever evaluation skills they may have acquired and their institutional knowledge. From the broad view of the organizations, this diffusion of evaluation experience is beneficial, since evaluation work is going on throughout the organization. However, for the evaluation unit this is a serious challenge because it constantly must train new staff. It would be impractical to suggest that evaluation units be exempted from normal organizational assignment policies, or that an evaluation career be established; and in any case such ideas would be beyond the remit of this report. However, in order to promote good workforce planning it may be possible to work with Management to identify one or two positions that can be filled for an extended period by staff with evaluation expertise, including the possibility for promotions. This would serve not only a recruitment and retention need, but also help to re-inforce independence by having the option for some staff to build evaluation careers within the organization.

Recommendations: The evaluation units should ensure that their policies or protocols address specific issues of:

- conflicts of interest in evaluation, including those that may arise from external consultants' past or planned future work; and
- the recruitment and retention of staff with evaluation knowledge and skills, including possible exceptions to general policies of staff rotation in the organization.

Protection from external influence: All of the units reviewed report that, while stakeholders do attempt to influence evaluations through advice and comments, their organizational independence and enabling environments shield them sufficiently to protect against undue influence on the conduct, findings, conclusions, and recommendations from their evaluations.

Part of protection against such influence comes from the adequacy and security of resources available to the evaluation function. Overall, the EF staff report that they are adequately staffed in relation to current workloads, and that budgets are sufficient to carry out their institutional missions. (See Table 1, below.) However, that does not imply that they are working at an optimal level of effort; additional staff and budget resources would allow more thorough evaluation coverage of important issues, if they were made available.

Table 1: Staffing and Costs for Evaluation Units^a

Unit	Staff (FTE)	Budget	Cost per evaluation
Finland	5 full-time staff 8 consultant	€1.5 million	€350-500 thousand
Ireland	6.75	€230 thousand (consultant costs only)	€125-145 thousand (consultants + staff)
Switzerland – SDC ^b	5.4 (all E+C) 2.5 (evaluation)	€660 thousand (consultant costs only)	€70-200 thousand (consultants only)
Switzerland – SECO ^b	1.3	€220 thousand (consultant costs only)	€140-215 thousand (consultants only)

^a The figures in the table are not directly comparable across the units since the costs covered vary among them e.g. Finland's budget is the evaluation unit's entire annual budget; they are presented here only to provide basic descriptive data.

^b Calculated at 1.00 CHF = €0.94.

However, beneath the totals there are human resource issues that create risks to the ability of the units to function effectively. Staff rotation policies of the parent organizations often require experienced evaluation staff to leave for other parts of the organization, to be replaced by new staff with little or no evaluation training or experience. None of the organizations has a career path for professional evaluators, making it difficult to recruit individuals with the relevant skills. In addition, training is mostly spotty for incoming EF staff. All of this leaves the units vulnerable as they oversee evaluation processes and monitor the activities of external evaluators. (see recommendation in the previous sub-section.)

Credibility of evaluations

The credibility of centralized evaluation is high in terms of quality, staffing, other resources, purposes, and use, but decentralized evaluation is less visible, so gets less attention

A review of a small sample of centralized evaluation reports across the four agencies found that they generally met international quality standards. While the quality of individual reports varies, as is to be expected, the officials and others we interviewed expressed a high degree of satisfaction overall with the quality of the work. However, in some cases it was not clear exactly how the units were conducting quality assurance, though clearly there was some routine practice. Units should be sure their processes are well-documented and followed; a good example is Ireland's new quality guidance as discussed in Chapter 3.

However, while institutional evaluation policies typically cover decentralized evaluations, there is much less information on their quality and credibility. For one thing, evaluation policies generally do not define what constitutes a decentralized evaluation, so that a wide range of products, from mid-term reviews, to completion reports, to process reviews, to rigorous outcome and impact evaluations all were described as decentralized evaluations in interviews. The lack of clarity on what constitutes an evaluation is compounded by the lack of visibility of these studies, which makes it difficult to assess their credibility. This can be problematic when findings, conclusions, and recommendations from these reports are used as input into later evaluations, both centralized and decentralized.

The centralized evaluation units reviewed in this report differ in the extent to which they deal with decentralized evaluations. SECO, for example, keeps an inventory of all planned and ongoing evaluations, and publishes all completed evaluations in an annex to its Bi-Annual Performance Report. Finland conducts a meta-evaluation of decentralized evaluations every three years. But none of the four units has direct control over decentralized evaluations, which makes it difficult to maintain quality standards, especially since those overseeing these evaluations usually are operational staff who typically lack expertise in the design and conduct of evaluations.

Some central evaluation units do provide guidance and other technical assistance to operational units that commission evaluations, but this is mostly at the margin, since they do not have direct input to those evaluation processes. In Finland, for example, the evaluation unit helped to facilitate the creation of a help desk through an external management company (see Chapter 2). An assessment of decentralized evaluations was beyond the scope of this review, but this gap in the quality assurance system was readily identifiable.

Recommendation: The evaluation units should work with Management to ensure that evaluation policies are clear on what constitutes a decentralized evaluation, the quality standards for such evaluations, and the processes for assessing quality.

Utility of evaluations

Finally, at all the organizations evaluation work by the EF was perceived to be useful and was used by operational staff and management, though this was mostly at the policy rather than operational level.

Of course, this varies a great deal by individual reports. But the general consensus coming from the interviews is that the work of the EFs is worthwhile and contributes to the conduct of the organizations' work. Many of the operational staff interviewed reported that they perceived centralized evaluations as directed primarily at the concerns of Senior Management, often addressing policy-level decisions rather than operational needs at the project or program level. Country teams, in particular, noted that they often were asked to contribute information for evaluations, but that the final reports did not provide specific feedback or recommendations of use to them, leaving them feeling "taxed" without a commensurate benefit. By contrast, they saw decentralized evaluations as more directly relevant to their work. Implicitly confirming this, senior managers were generally more likely to cite specific uses they have made of the results of centralized evaluations.

This is not a novel finding. In general, it is true that centralized evaluations are geared toward the broader issues facing the organization than any one set of operations. One exception may be country program evaluations, which do address specific country issues, and typically are geared toward informing a new country program under development. Sometimes, this also is true for sector reviews. But in general, centralized evaluation does tend to be directed primarily toward issues confronting Senior Management at most development-related organizations.

This seems a reasonable focus, given that the organizations also carry out decentralized evaluations designed explicitly to address operational issues with projects and programs. In effect, there is a division of labor here that for the most part seems to be able to meet the evaluation information needs of the organizations at multiple levels. Indeed, it is nearly impossible for one report to be equally relevant at the operational and strategic levels, yet sufficiently deep in both areas.

Still, there are some things that the EFs can do to address this issue. For example, when an evaluation includes case studies of specific programs, projects, or other operations, in addition to the main report the EF could issue a set of case studies, with more in-depth information than in the main report. It also could have the evaluators provide briefings or other feedback to the operational teams specifically on their operations that were used in the evaluation. Some of this already is being done, but it could be more systematic.

One issue of common to all evaluation units is how to improve the communication about and dissemination of evaluation findings, conclusions, lessons, and recommendations to a range of audiences that might have an interest. In the subsequent chapters there are specific discussions of this issue for each unit. But the more general finding is that all evaluation units should recognize the need for a regularize set of communication activities to disseminate their work to audiences likely to be interested in and able to put to use their work. This requires both a general

communication policy and procedures to carry it out, but also a communication plan for each evaluation that is developed early in the planning process.

An important facet of use is the Management Response. This is the organization's formal process for taking on board the results of evaluation, accepting accountability, and demonstrating a capacity for learning. Management acceptance of recommendations and implementation of actions responsive to them is the most obvious example of utility. (See Box 2, below, for an example.) During the review, three issues about the Management Response were raised to one degree or another across the organizations: (1) lack of clarity about who is drafting and delivering the response; (2) frequent delays in delivering the response; and (3) difficulties in tracking implementation of agreed actions included in the Response.

These issues partly reflect process ambiguities that could be addressed through operational protocols. But they likely also are symptoms of a deeper problem with Management's ability or willingness to accept and implement evaluation recommendations.

The World Bank's IEG, confronting similar problems, conducted a series of case studies several years ago to determine what led to actual implementation of recommendations. They found that several factors are crucial:

- in-depth discussion with management during the drafting of recommendations;
- credibility of the evaluation results;
- sense of shared ownership over the evaluation and the findings;
- quality of the recommendations, specifically in terms of their timeliness, how actionable they are, and their cost-effectiveness, clarity, and coherence;
- advocates/champions supporting the adoption of the recommendations; and
- institutional incentives and accountability for adopting the recommendations.

Box 2: Results-Based Management in Finland's MFA

The Development Evaluation Unit produced three reports on Results-Based Management (RBM) in MFA's development work over the past decade. Two early studies, *Results-Based Approach in Finnish Development Cooperation* (2011) and *Evaluation of Finland's Development Policy Programs from a Results-Based Management Point of View 2003-2013* (2105) focused on helping to guide the development of a robust RBM system in MFA.

In interviews, a number of senior officials cited these as "high-quality" reports and good examples of how evaluation can influence MFA's work. One noted that the 2011 report, in particular, "provided a wake-up call to Management." Even though not all the recommendations were implemented, the report spurred constructive discussions across MFA that led to subsequent Management actions.

The most recent report, *How do we Learn, Manage and Make Decisions in Finland's Development Policy and Cooperation* (2019), looked at how much progress had been made, and found that RBM was having a "moderate level of influence on results information on both learning and decision-making", but also cited a number of continuing weaknesses, including a "lack of institutionalization of evidence-informed learning and decision-making at MFA."

Based on these findings, IEG developed a new process for developing recommendations, set a time limit of 90 days following discussion of the report at the Committee on Development Effectiveness for Management to submit detailed action plans and timelines for implementing recommendations accepted as part of the Response, and began reporting to the Committee on a quarterly basis the extent of implementation by Management.¹⁴ The World Bank's tracking system can be more resource-intensive. However, SECO has deployed a more streamlined approach that is discussed in Chapter 5 (see Table 4 in that chapter).

Recommendations: The evaluation units should consider strengthening the utility of their evaluations by, *inter alia*:

- using the planning process to:
 - ensure that the objectives and scope of the evaluation are clearly aligned with the information needs of identified internal or external stakeholders and audiences;
 - set priorities among those information needs to ensure maximum utility (see, for example, Table 3 in Chapter 3);
 - work with the evaluation team to clarify what data are needed to meet these information needs;
 - agree with the evaluators on how EVA staff can participate in evaluation activities so they can ensure the work remains focused on those information needs; and
 - identify opportunities to employ evaluation findings and products for multiple purposes (including inputs to other evaluations, where appropriate) involving multiple stakeholders and audiences.
- improving communications and dissemination through:
 - developing a communication strategy that addresses such issues as how to reach internal and external stakeholder audiences;
 - using the planning process to identify opportunities to employ evaluations for multiple purposes, and set priorities on whom the evaluations are expected to influence;
 - including a communication/dissemination plan for individual evaluations from the start;
- strengthening the Management Response process by:
 - clarifying roles and responsibilities for Management and the evaluation unit;
 - setting timelines for both the formal response and agreed actions;
 - enhancing follow-up procedures including tracking of actions taken; and
 - providing for regular reporting on progress and any shortcomings in implementation.

¹⁴ World Bank, Independent Evaluation Group, Managing Evaluations: A How-To Guide for Managers and Commissioners of Evaluation, 2015, pp. 41-47.

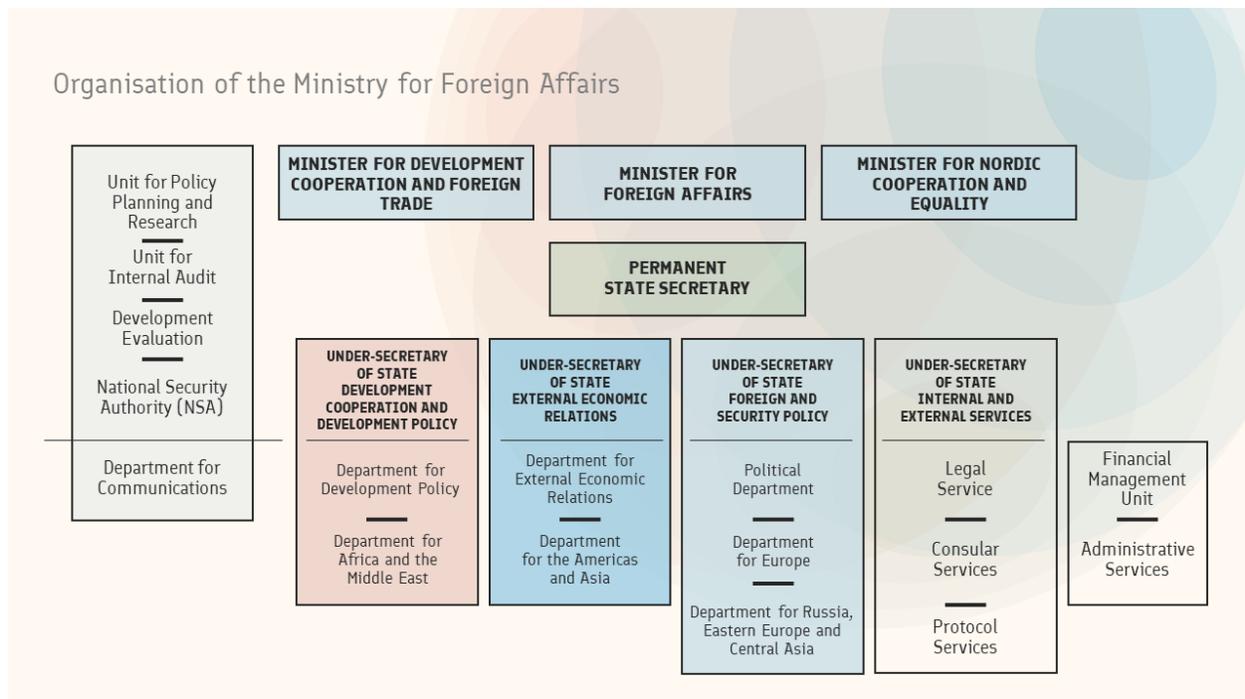
https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ecd_man_evals.pdf

CHAPTER 2: EVALUATION IN FINLAND

The Ministry for Foreign Affairs (MFA) manages and co-ordinates most of the Finnish development co-operation programs. Under the MFA, the Department for Development Policy is responsible for providing overall guidance on the implementation, planning, and monitoring of Finland's development co-operation policy, and holds direct responsibility for the operational activities directed to multilateral and civil society organizations, development co-operation, and humanitarian aid. Regional departments are responsible for the implementation of bilateral co-operation.

The institutional arrangements and responsibilities for evaluation are defined in the Decree on the Ministry for Foreign Affairs 550/2008(1280/2013) and the Evaluation Norm 1/2015. The Unit for Development Evaluation (EVA) is responsible for the development of the evaluation system, commissioning policy and strategic evaluations, and ensuring their effective use. It became an independent administrative unit in January 2014, reporting to the Under-Secretary of State for Development Cooperation and Development Policy. (See Figure 1.)

Figure 1: Organization of the Ministry for Foreign Affairs, Finland



Through its participation on the Intervention/Project Quality Assurance Board of the MFA, EVA is expected to ensure that evaluations are used for intervention planning, that the evaluability of new interventions is high, and that an initial evaluation plan is integrated into the funding proposal. Furthermore, EVA also is part of the Development Policy Steering Group (KEPO) of the MFA, and an expert member of the national Development Policy Committee to facilitate participatory evaluation planning and provide advice on evaluation-related issues.

The planning cycle covers three years, the first of which is elaborated in detail. The ministry-level evaluation plan is prepared by EVA in co-operation with policy makers, senior management, as well as implementing units. The evaluation plan is discussed with Development Steering Group (KEPO) and later presented to the Under Secretary of State for approval. Both centralized and decentralized evaluations are included in the plan. However, the actual planning and implementation of the decentralized evaluations is done by the implementing units themselves.

Uniquely among the EFs included in this review, EVA outsources conduct of its evaluations to a company that implements all stages of the evaluation process, from terms of reference (TORs) to final reports, through an Evaluation Management Services (EMS) contractor, based on a concept note drafted by EVA.

The purpose of this arrangement is to improve the quality of evaluations by securing the best possible evaluation professionals without bearing the time and resource burdens of the tendering process, thus maximizing the flexibility, efficiency and effectiveness of the MFA in planning and commissioning evaluations. The EMS contract is renewed each four years through competitive bidding.

EVA staff and other MFA interviewees report a high degree of satisfaction with both the quality and management of evaluations through this system. One especially appealing aspect is that the contractor is responsible for identifying team leaders and other members for each evaluation, using their networking capacity. Final decisions rest with EVA, but off-loading the search process is an especially useful element of this contract process.

One issue with this arrangement is that it may create some ambiguity about the role of the internal evaluation manager (EM) vis-à-vis the team and team leader. The EM's role is not clearly specified in the framework contract, so a wide range of possibilities could be compatible with it. A minimalist position would be that the EM plays essentially an administrative role, overseeing necessary paperwork, ensuring contract compliance, and the like. A more active role could include joining the team on field missions, primarily as an observer, to ensure the quality of the work being done.

Experience among consulting firms and other organizations involved in this kind of outsourcing indicates that Unit participation in the evaluation process as part of the managerial function—such as observing field missions, interviews, team meetings, and the like—can be beneficial in helping to ensure the quality of the finished product, while allowing the EM to identify any emerging problems early enough to rectify them. This intermediate role requires little background in technical evaluation skills, but at least some management experience likely would be helpful, particularly in helping to control costs, though staff often lack such skills.

Recommendation: EVA should ensure that its EMs either have experience in contract management consultants, or provide training in such skills for those who need it.

Enabling Environment for Evaluation

The enabling environment includes such factors as requirements for explicit consideration of evaluation findings in developing programs and projects; the degree to which such requirements are carried out in practice; the extent to which operational staff seek input from the evaluation office when considering new programs or projects; and the institutional support provided to the evaluation function by the agency in terms of the adequacy of financial and human resources, access to information, and willingness to cooperate in evaluation activities.

Box 3: Finland's Evaluation Help Desk

The Evaluation Unit in Finland's Ministry for Foreign Affairs has a contract with an external consulting firm for evaluation management services. In addition to managing the independent evaluation process for the Unit, it also supports other operational units in their decentralized program evaluation activities by providing capacity development tools and activities, evaluation guidelines, and a help desk services that provide particularized assistance. This function has proven popular among operational units. Which have provided positive feedback on its utility.

Overall, there is a strong enabling environment for development evaluation within MFA, though with areas that could be improved. This is conveyed, for example, by an Evaluation Policy requirement that every development project be evaluated at some point within its lifetime. Such evaluations may come at different stages, frequently as mid-term or completion reviews (Project Final Evaluations). In a few instances, impact evaluations have been commissioned. These evaluations are planned and implemented through the units managing the projects, not EVA. This has the benefit of making operational staff comfortable with the idea of regular evaluation, and contributing to their understanding of the need for and usefulness of evaluation. However, EVA does provide some support for these evaluations, including training and help desk services. (see Box 3.)

Helping to strengthen this environment are some tools that EVA has prepared to provide guidance applicable to both centralized and decentralized evaluations. Particularly noteworthy is a web-based *Evaluation Manual* that combines documents, graphics, and videos to walk users through a wide range of issues, from an explanation of what evaluation is and why it is important, through discussions of evaluation standards and ethics, evaluation questions, methods and designs, data collection and analysis, managing the evaluation process, to reporting and follow-up, and much else. It is both user-friendly and encyclopedic, and can serve not only as a training tool but as a ready reference for those planning and conducting evaluations.¹⁵

The Peer Review team found a notable culture of evaluation within the organization. Interviews with senior managers, staff responsible for managing projects and programs, and other officers found a broad understanding of and commitment to evaluation as a tool for conducting and improving MFA's development work, including the centralized evaluations conducted by EVA. Importantly, this extends to budget support for EVA work, all the way to the level of the Parliament, as discussed below. Finally, the size of the Ministry is small enough that staff throughout are

¹⁵ EVA, *Evaluation Manual*. <https://um.fi/development-cooperation-evaluation-manual>.

familiar with EVA's evaluation work through informal interactions, as well as more formal communications. This can be a double-edged sword, however, as discussed below.

One issue that has arisen is the scope of EVA's evaluation mandate. The Finnish International Development Agency was merged into the MFA several years ago, in part because of the implicit relationship between official development assistance (ODA) and broader foreign policy and trade-related activities. However, this necessarily means that in some cases evaluations of ODA work will touch on other these other areas of MFA responsibilities. In general, evaluation is not routine in those areas, and this can raise issues about the appropriate scope of evaluation work.

For example, a recent evaluation on forced displacement provided insights into Finland's work in this area, and the nexus between displacement and humanitarian aid. The findings and recommendations from this evaluation necessarily go beyond development assistance work and touch on issues of foreign policy and relationships with international organizations in non-development domains. (see Box 4.)

This is important because as international development programs increasingly intersect with those domains, evaluators will find themselves navigating what one interviewee called the "tricky" terrain that defines both the boundaries of their competence and their ability to formulate recommendations that can be implemented by all the relevant agencies. In practice, however, there is no need to change the formal mandate for EVA, since evaluations cover all departments within MFA already.

Box 4: Response to Forced Displacement

The *Evaluation on Forced Displacement and Finnish Development Policy* (2019), was designed to "assess how coherently Finland's development policy and its targets relating to forced displacement have been implemented and how the coherence could be enhanced." A number of interviewees cited this report as showing how the interaction of development and foreign policy activities can affect the work of evaluation units.

The report found that Finland, though "a highly respected development and humanitarian actor and advocate...has not developed clearly formulated and well-established approaches that effectively inform its policy making and programs for development cooperation and humanitarian assistance in a coherent and comprehensive fashion" in addressing issues of forced displacement and the humanitarian-development nexus.

Recommendation: EVA should consult with other evaluation units, such as those in The Netherlands and Canada discussed in Chapter 1, that have successful dealt with the non-

Independence of Evaluations and the Evaluation System

Organizational independence

As noted above, EVA reports to the Under-Secretary of State for Development Cooperation and Development Policy, who does not have responsibility for operations undertaken by the regional departments. This arrangement meets the criteria for organizational independence, though it might be stronger if the unit reported to the State Secretary or Minister for Development Cooperation and Foreign Trade.

Moreover, the budget allocated to the Unit appears to be adequate to meet its mandate. In interviews, EVA staff reported that they did not feel constrained by the budget to compromise the breadth or quality of the evaluations on which they were engaged. Interviews with interlocutors between MFA and the Parliament noted the strong support for providing adequate budgets for evaluation at all levels of government. This level of support also reinforces behavioral independence by removing a potential source of pressure on the unit

Behavioral independence

A review of a sample of reports and responses from those interviewed consistently emphasized that the Unit exercises its function with a high degree of independence. Indeed, the case of the evaluation of forced displacement discussed above is a good example of the assertion of independence even when questions were raised about whether EVA was going beyond its mandate. No document the Peer Review team examined, and no interviewee, cited any instance of independence being compromised behaviorally.

EVA annually develops a comprehensive three-year evaluation plan in consultation with potential users of evaluation; the plan includes decentralized evaluations to be undertaken by operational units. This plan, and any subsequent changes, is presented to the Development Steering Group (KEPO) for discussion, but approved by the Under-Secretary of State, who is responsible for ensuring the independence of the unit. The comprehensive plan for centralized evaluations is binding for the first year and indicative for the following two years. General evaluation principles guide the plans, including that the results will be available when the information is needed, though in practice this objective is often not met because of the lengthy time for evaluations to be completed. The plan also discusses related training.

In addition, a Development Policy Committee helps to bolster the Unit's independence by bringing together institutional members from academia, stakeholder groups, Ministries and Parliament to discuss evaluation reports and "translate" them into a more a more user-friendly form for policy-makers. The Committee also helps to gain support for EVA's budget, and can look into past recommendations to see what has been done in response. There may be scope for the Unit to make use of this Committee strategically to further promote its messages and encourage action on its recommendations, as discussed below.

Avoiding conflicts of interest and outside influence

Apparently, there is considerable interest in EVA's work among the extensive network of non-governmental organizations in Finland active on development-related issues. They do express their views, but there is no evidence that this has had any compromising effect on the independence of the evaluation process or products. Internally, interested units also may wish to influence EVA evaluation findings and recommendations, not only through the legitimate formal channels, such as commenting on draft reports, but by taking advantage of the relatively small size of the organization to exert informal influence. Despite this, the review did not surface any evidence of loss of independence from these possible sources.

However, there are personnel issues associated with rotation of staff from across MFA into EVA, as with other assignments. In addition to the potential for conflicts of interest from both past work and future opportunities this implies, these staff usually do not come with training in or experience with evaluation. This is discussed in Chapter 1, which includes recommendations on policies regarding conflicts of interest, as well as staff recruitment and retention.

Credibility of Evaluations

In interviews with operational managers and staff, the Peer Review Team was told that the framework agreement had helped improve the quality of EVA evaluations even as it has allowed the quantity to increase. Concomitantly, EVA staff reported that their administrative and procurement workload had gone down, leaving more time to attend to substantive tasks related to the evaluations. Both inside EVA and across MFA the framework agreement has been received positively.

However, this arrangement carries a number of risks for EVA. First, the EMS coordinator provides EVA with a list of three Team Leader candidates with a proposed approach from each. While EVA makes the final selection of Team Leader it is somewhat dependent on the external EMS coordinator for identifying a short list, raising the risk of real or perceived conflicts of interest, as discussed above.

Second, it also leaves EMS with the lead role in defining the scope of work, although EVA does produce a Concept Note at the start of the evaluation, and must approve the Terms of Reference and scope. This has important implications for the cost, timeliness, and utility of the evaluations undertaken. EVA's costs per evaluation are considerably higher than those for the other units. The reports also tend to be longer, denser, and more complex, sometimes appearing more as research products than evaluations *per se*. This likely reflects the fact that the work typically is led by university faculty members. (As discussed below, this also has implications for the utility and actual use of evaluations.)

A major way EVA could address this issue is by taking a more active role in the design of evaluations. The Concept Note is a vehicle that could be especially useful in directing the external consultants as they prepare the Terms of Reference and scope of work, proactively stemming the "mission creep" that has made some of the evaluations more expansive and expensive than necessary. It therefore should provide specific guidance on:

- the specific purpose and objectives of the evaluation, including a clear statement of why these issues are important to the work of MFA, not just to the field in general;
- the evaluation questions to be addressed, and any limitations on the scope of work to ensure that the focus remains on those questions;
- the expected use of the evaluation findings, conclusions, and recommendations, including, where possible, identification of specific users (such as MFA units or teams, external partners, policy-makers, and so on); and

- preliminary plans for how the evaluation results will be communicated and products disseminated.

The Concept Note thus can be a strong mechanism for guiding the work of the evaluators, including decisions on the Terms of Reference and scope of work. The TORs normally should conform to the Concept Note rather than adding to or subtracting to it, unless the evaluators make a compelling case that changes are needed to meet the stated purposes and objectives of the evaluation. Any agreement to such changes should be made only by the EVA Director, based on careful consideration of the trade-offs in terms of timeliness, costs, and staffing.

There may be some reluctance among EVA staff to do this because (1) staff without evaluation experience may be unwilling to challenge outside evaluators on technical issues, although EVA has made use of “critical friends” from outside the unit to advise the Evaluation Manager on these issues; and (2) there is some concern that a more active role would compromise the independence of the evaluation. But these concerns should not be determinative.

For one thing, while external consultant may have subject-matter and methodological expertise, EVA has a higher level of institutional knowledge. The means EVA readily can find out from internal MFA units with stakeholder interests in any evaluation what kind of information they will find most useful. A good place to start would be with the reference groups typically set up to advise on each evaluation. Reference groups are made up of MFA staff, and have input at five points:

- during preparation of the Concept Note;
- at the start-up meeting with the evaluation team;
- when commenting on the draft Inception Report;
- as part of a validation exercise when the team presents initial findings/data, and draft recommendations;
- in commenting on the draft final report.

There has been little research into the structure and use of reference groups, even though they are widely employed in evaluation work, including by the units covered in this review. So it is not possible to characterize any “norms” for such groups in terms of size, composition, skills, or roles and activities. However, several years ago the American Evaluation Association did publish a volume in its *New Directions in Evaluation* series that included some case studies and a proposed model for these kinds of groups.¹⁶ The authors of the latter note that, “who is invited to participate in [a reference group] depends on the needs of an evaluation study and the context within which the study is taking place.”¹⁷ They identify three main types of participants:

¹⁶ R. VeLure Roholt and M. L. Baizerman, “A Model for Evaluation Advisory Groups: Ethos, Professional Craft Knowledge, and skills,” in R. VeLure Roholt and M. L. Baizerman (eds.), *Evaluation Advisory Groups. New Directions in Evaluation* 136, pp. 119-27 (2012).

¹⁷ *Ibid.*, p. 120.

- experts with technical evaluation expertise (e.g., design, data collection and analysis);
- stakeholders with interest in the outcome of the evaluation;
- individuals who can help craft and disseminate the evaluation messages.

In the context of MFA and the other organizations reviewed here, technical evaluation expertise tends to be in short supply internally, so other resources may need to be brought to bear. EVA has introduced “critical friends” with evaluation and/or substantive expertise to assist the EMs in designing and managing evaluations. The stakeholders of interest are most likely to be MFA staff of the office(s) whose work is being evaluated, others with a direct role in the country(ies) or substantive area(s) covered by the evaluation, and perhaps those with responsibility for institutional policies that might be involved. Evaluation units typically do not have staff with singular responsibility for promoting their messages, but the parent organizations do, and these can prove helpful, even if not included on a reference group.

Properly constructed and used, therefore, reference groups can supply useful inputs to the design, implementation, and final product of an evaluation; give entrée to knowledgeable individuals and groups with a stake in the issues being evaluated who otherwise might not be as available to the evaluators; and engage stakeholders in the evaluation process to both foster the credibility of the results and create opportunities to disseminate them to the broader relevant community(ies) of practice. In addition, operational staff interviewed for this review who had participated in reference groups reported that they found the experience generally useful, both because they learned about issues relevant to their work, but importantly because they developed an appreciation for the evaluation process itself, giving them a clearer understanding of how it could contribute both to their own work and to the broader efforts of MFA.

Necessarily, the size and composition of such a group will vary depending on the scope and complexity of the evaluation. A review of work in one sector in a single country might require no more than three members, while a broad evaluation of a multi-sector effort across numerous countries would require many more. There is no simple rule of thumb here, except to recognize that there usually is a tradeoff in the size and efficiency of any group. Limiting stakeholder membership to the key issue areas, for example, might be helpful in managing this tradeoff, but in the end this is a matter of organizational politics as much as anything else, so knowledge of the organization and the key points of possible resistance/support is an important consideration.

Beyond that, a more active role for EVA staff in the conduct of evaluations, including participation in field visits or other evaluative work, does not undermine independence, and can serve both to better manage the evaluation and to help staff hone evaluation skills. Independence inheres in the evaluation unit itself, whether evaluations are conducted externally (as in the unit under review) or internally (as at the World Bank). The key is to define the roles of internal and external staff clearly so that expectations are harmonized and conflict minimized, and to have a strong commitment to the unit’s independence by the organization as a whole.

Recommendation: EVA should develop specific guidelines for:

- the selection of members of reference groups to ensure that a skill mix reflecting evaluation design and methods, stakeholders with substantive knowledge of the subject matter and countries covered by the evaluation and, if possible, with skill in shaping and disseminating messages; and
- conduct of the reference group, including which tasks are expected to involve formal meetings and which may not;
- clarifying the role EVA staff may play in working with evaluators in conducting field work and other evaluative activities.

Utility of Evaluations

Evaluation ultimately is expected to be used to achieve practical results, such as improving policies, programs, projects, and other activities in light of findings, conclusions, and lessons. In MFA, as in other organizations supporting international development, the key ways in which this is expected to happen are institutional learning and implementation of recommendations. Thus, it is important that the evaluation reports are not only independent and credible, but that they address issues important to MFA decision-makers in a timely way and with clear messages. This review found some issues in this regard, however. At EVA, some of the utility issues center on how managing the scope and complexity—and thus cost—of evaluations to improve utility may affect the independence of evaluations, how to improve utility through the management response process, and ways to communicate/disseminate learning from evaluation to audiences outside MFA.

Scope, complexity, cost, and independence

EVA's evaluations are more expensive than those at the other organizations. Interviews with MFA managers and staff, as well as a review of a sample of reports, suggest that in part these costs reflect how the evaluations are conducted. Compared with the other organizations, EVA evaluations tend to be more ambitious in scope, more complex in methods, and longer in duration. All of these drive costs, and also have implications for the utility of the resulting products.

In reviewing sample reports, one of the things that stands out is the amount of space in EVA reports taken up by discussions of the underlying issues and the methods employed. For example, the report on a 2018 meta-evaluation on the treatment of women's and girls' rights in Finnish development policy included a nine-page discussion of how the meta-evaluation was conducted, and ten pages on the context for the evaluation, in addition to several pages on the underlying theory of change.¹⁸ This kind of detailed exposition would be of interest to evaluators, researchers, and some readers interested in the policy background. But both sections clearly could have been much shorter, with the more detailed information moved to an appendix or made available online

¹⁸ Evaluation on Improvement of Women's And Girls' Rights in Finland's Development Policy and Cooperation, 2018. [https://www.shareweb.ch/group/Peer-Review-Evaluation-Function-between-IR-SF-CH/Shared%20Documents/Documents%20Finland/Central%20Evaluations/Evaluation Women and Girls rights NE TI 2018.pdf](https://www.shareweb.ch/group/Peer-Review-Evaluation-Function-between-IR-SF-CH/Shared%20Documents/Documents%20Finland/Central%20Evaluations/Evaluation%20Women%20and%20Girls%20rights%20NE%20TI%202018.pdf)

for interested readers. By front-loading these long discussions the report created a barrier for readers primarily interested in the analytical findings. Moreover, in interviews operational managers often characterized EVA reports as “theoretical”, almost certainly reflecting the heavy emphasis on the theory of change, broad context, and evaluation methods front-loaded in the reports.

What this suggests is some lack of clarity on the part of the evaluators about who was the primary audience for the report: MFA managers. It is not clear from the evidence, but some interviewee comments suggest that the focus on methods in the reports reflects the time and attention the evaluators spent on refining the methods. This possible misdirection of attention is reflected as well by the relatively long time-frames for some EVA evaluations, which often means that by the time they are completed interest that might have been high at the start might have waned or key decisions might already have been made.

At the same time, the high cost of many EVA evaluations limits the ability to conduct other work, and therefore the evaluation coverage for MFA. EVA’s dependence on the evaluation team leader for designing evaluations contributes to this problem, and the lack expertise among EVA staff, also appears to contribute to the high costs of the work.

A conclusion here is that EVA could exercise more control over the evaluations it commissions through the EMS. It likely can commission some evaluations that are methodologically simpler and less costly. But that would require that the ambitions of the work be more modest, as well. Evaluating big, complex programs is inherently more difficult and more expensive than doing reviews looking at management of projects, for example. From interviews it does not appear that there has been a discussion between EVA and MFA management on whether a greater diversity of evaluations along these lines would be useful, but that would be a logical way to begin addressing the issue.

EVA also has raised concerns about whether focusing too much on the utility of evaluations is a risk to evaluator independence. And, in fact there have been debates about this issue within the evaluation community for decades. But in general development evaluators across the international financial institutions, United Nations agencies, and OECD-DAC long have endorsed utilization as the key outcome of evaluation because their mandates are to use evaluation to promote accountability and learning, not to conduct basic research. The key is to conduct evaluations that address important issues for the parent organization, in a timely way, with high quality and credibility. Decades of experience demonstrates that this can be done without sacrificing independence. Ultimately, control over the scope, complexity, and cost by EVA are compatible with independence because, as noted above, independence inheres in the unit itself, not in the ability of external contractors and consultants to exercise unfettered control over the evaluation process.

Recommendation: EVA should explore ways to diversify its portfolio of evaluation products so as to address important issues more efficiently through:

- conducting some less complex and expensive evaluations;
- establishing through its EMS agreement requirements that evaluators employ the least-cost methodology that is adequate to answer the evaluation questions; and
- redesigning reports so that much of the technical material is moved to annexes or appendixes, allowing readers to focus on findings, conclusions and recommendations.

There are many examples among the development agencies, but in the end for the process to work it must be particularized to MFA, with buy-in by both EVA and Management.

Broader communication and dissemination

Finally, EVA is interested in extending the reach of its communication and dissemination activities to a broader audience in order to influence development policy through public discourse and to raise the visibility of MFA's development work, including achievements. Historically, this has been a somewhat neglected area among official evaluation units, partly because of reticence to participate directly in public debates, which could be seen as "taking sides" and impinge on the perception of objectivity and independence, and partly because most evaluation units lack the human resource skills to manage relations with the media and other actors of interest.

The main way to get media coverage of evaluation reports is to establish ongoing relationships with reporters and editors interested in the subjects of the evaluations. For example, when IEG decided to take advantage of a new, more outward-looking disclosure policy it identified a reporter at the *Financial Times* who had written a number of articles about World Bank work. They invited him to speak at a department retreat to explain to staff what makes a report of interest to reporters, and then kept in touch with him, alerting him to evaluations that might be of interest. This proved successful, and soon, working with a media relations professional, they had established a number of such relationships.

Identifying appropriate journalists could be done through a number of channels. Creating a contact information database of any who write about an EVA report or about Finnish development assistance, for example, which could be used to send notifications of new reports or short notes on important findings, using email or social media (reporters tend to be heavy users of Twitter, for example). Personal meetings with journalists are an important way both to establish relationships and to learn what specific issues are of interest to them. To do this kind of work, however, normally requires someone with experience in this area of work. That does not mean a full-time staff member would be required; given the volume of EVA output a part-time consultant likely would be sufficient.

But before pursuing such an effort, EVA would need assess the possible trade-offs. First, MFA has a Department of Communications as well as a communication policy and strategy. It would be necessary to determine whether those would require that such communications go through this

Department. If so, there is a risk that EVA's independence could be compromised, especially if that Department would control the content of any announcements, news releases, or other communications, or be responsible for screening contacts with journalists. Second, this work would require at least some financial resources, as well as the time and attention of the EVA Director. That's an opportunity cost in relation to the rest of the unit's work that would have to be balanced against the potential gains in terms of influencing policy and program decisions and informing stakeholders and the broader public. Third, EVA would have to weigh the reputational risk that it would draw negative attention to its work or that of MFA, however unintentionally.

Another outlet available to EVA is the Development Policy Committee, a group of about 80 individuals representing academe, stakeholders, and ministries. The Committee discusses EVA reports, "translates" them into non-technical terms for discussions with parliamentarians and other policy-makers, and does some checking on recommendation follow-up. It is possible to build on the Committee's membership and existing network of contacts to amplify EVA messages. To do this strategically, EVA would have to develop a clear statement of its communication/dissemination goals, identify the kinds of audiences it most needs to reach to meet those goals, and assess the extent to which the Committee can act as an appropriate conduit. The Committee could be a valuable resource if it is engaged with a strategic vision.

These suggestions should not be taken as formal recommendations, since a good deal of additional work would need to be done to assess the desirability and feasibility of carrying out such activities. Rather, these should be seen as suggestions on how to proceed, if EVA decides to explore these directions. However, EVA should already be considering ways to serve its multiple audiences.

All of the above can be summarized as advice to EVA to consciously consider all the ways its evaluation work can be used beyond the issuing of specific reports. For example, case studies conducted as part of an evaluation of a country-level program could also provide useful input to sector studies, and also could be published as stand-alone products that would serve the needs of both country and sector teams, as well as partner country officials, NGOs, academicians, and journalists with an interest in the specific subject. To do this effectively, however, EVA needs to assess the various audiences for any given evaluation and determine which are most likely to actually use the information, what format would be most useful for that audience, and even when the information would be most useful. This should be under consideration from the start of any evaluation and reconsidered over time routinely as conditions may change. A modest investment of time on such activities can yield a big return in evaluation use.

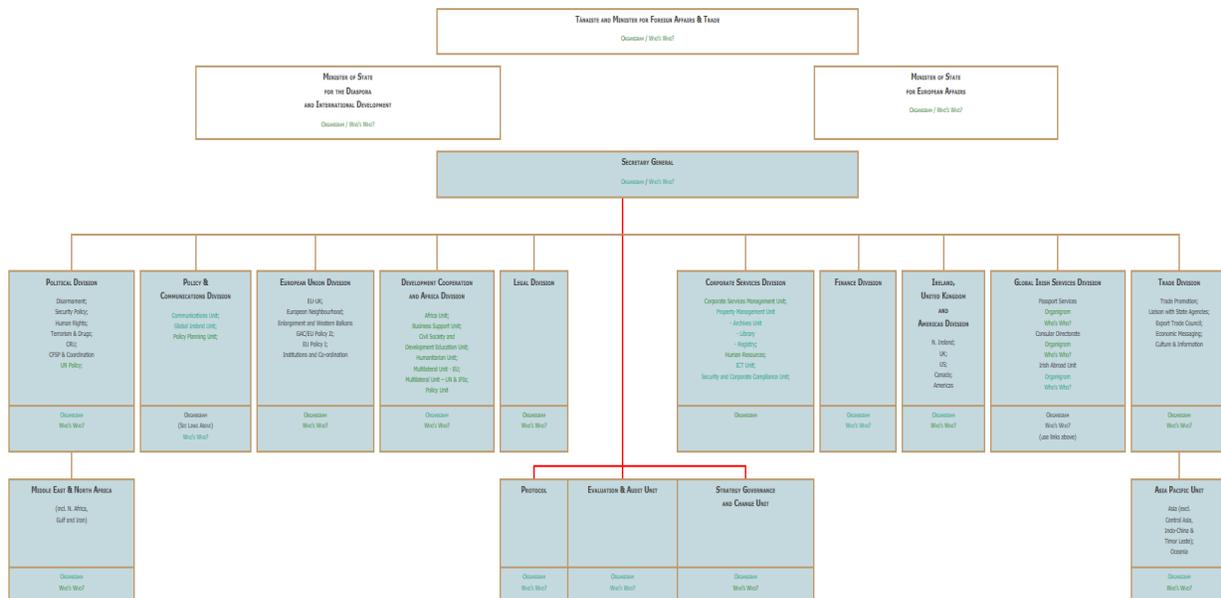
Recommendation: EVA should consider adopting the recommendations on utility in Chapter 1.

CHAPTER 3: EVALUATION IN IRELAND

The Development Cooperation and Africa Division (informally known as Irish Aid) is the Irish Government’s official development assistance program and is a division of the Department of Foreign Affairs (DFA). The Irish Aid Evaluation Policy (2007) defines evaluation in the Irish Aid context as being the systematic and objective assessment of the design, implementation and results of an ongoing or completed project, program or policy by assessing the effectiveness of the intervention against its stated objectives. Evaluations are overseen and planned for by the Evaluation and Audit Unit (EAU), a stand-alone unit that reports directly to the Secretary General of the Department. The Unit is mandated to evaluate not only the ODA managed by the Department, but also other activities across DFA. This is an expansion from its earlier role, limited to evaluating the work only of Irish Aid. In addition, EAU provides the internal audit function for the Department and the Unit is part of the overall Irish Government Economic and Evaluation Service, which is a network established by the Department of Public Expenditure and Reform to strengthen evaluation across the entire public sector (See Figure 2.)

Figure 2: Ireland DFA Organization Chart

HQ Organigrams



The Audit Committee reports to and advises the Secretary General, and provides an independent appraisal of audit and evaluation arrangements with a view to strengthening internal controls, fraud and risk management. The members of the Committee are fully external to the Department and have backgrounds in audit, governance, development, and public service management. The Committee meets at least six times a year.¹⁹

¹⁹ <http://www.oecd.org/dac/evaluation/network-member-ireland.htm>

In terms of planning, the Unit identifies evaluation topics that are responsive to the management needs of the Department, which are reflected in a three-year rolling work plan on strategic evaluations. Annual work plans are finalized in consultation with Senior Management and the Audit Committee and approved by the Management Board, chaired by the Secretary General. Plans typically include strategic evaluations of particular importance to the Department which are managed directly by the EAU. Criteria for selecting such evaluations are: policy and strategic relevance, utility, corporate accountability, financial significance, the level of risk and the cost-benefit (discussed further below). In addition, Divisions and Missions may also undertake and lead on decentralized “Operational Evaluations,” with technical support from the Unit. Finally, EAU provides technical support and training to Headquarters business units and missions abroad on evaluations.

Enabling Environment for Evaluation

The recent story of EAU begins with the loss of resources due to inattentive human resource management during the Irish financial crisis. By the beginning of 2017 evaluation staff numbered only 1.2 FTE, so capacity was sharply limited. Today it is up to 5.75, a remarkable recovery over a short period of time. This rebound in staff resources has allowed the Unit to develop an evaluation strategy to meet its broader responsibilities across the MFA beyond ODA.²⁰ For example, it has built a series of processes and tools to strengthen the evaluation function, including guides for assessing the quality of evaluations (discussed below), a report guide, and a report template.

Building on these successes, EAU currently is drafting a new Evaluation Policy, which will be complemented by the existing Operations Manual, published in 2012, that serves as the current policy document.²¹ A major reason it undertook this review is to provide recommendations that can inform that policy development and bring the evaluation function to a higher level of performance. Introduction of evaluation strategy, unit planning, work plans, quality assurance strategy, and tools shows great ambition, though it will be challenging to reach all the goals EAU is setting for itself.

More broadly, the Peer Review found that there is a good evaluation culture and considerable knowledge about the practice of evaluation within EAU. Members of the team report a high commitment to the Unit’s evaluation mission, and that team members work well together. Especially important, high-level Management interviewees expressed strong support for evaluation and its role in helping to improve the organization’s performance. In addition, the external Audit Committee provides support and advice to EAU. But unlike in Finland and Switzerland, EAU does not have a direct conduit to the country’s parliamentary body, the Dáil.

Currently, all contracts for funding from the Department include a requirement to conduct an evaluation when the contract is completed. However, as in Finland, the expansion of

²⁰ DFA, Evaluation and Audit Unit, *Strategy2020-23*, 2019

²¹ DFA, Evaluation and Audit Unit, *Evaluation Operations Manual*, 2012.

responsibilities into non-ODA areas is a work in progress. It is not clear from interviews the extent to which the staff focused on foreign policy and trade have absorbed evaluation into their processes and embraced an evaluation culture, even at the level of the Management Board. This kind of shift typically takes a long time, so it is not surprising that it remains a challenge at this point.

A second issue raised by EAU is whether the combination of evaluation and internal audit in the same unit is problematic. However, during the 1990s a small but lively literature developed around the apparent convergence of auditing and evaluation as performance auditing has become more widespread, particularly effectiveness audits. For example, number of national audit agencies, such as the US Government Accountability Office, UK National Audit Office, and Swedish National Audit Office, do carry out effectiveness audits and often program evaluations. In summarizing the findings from this literature, John Mayne concludes:

Finally, many organizations have placed audit and evaluation in the same organizational unit, often reporting to the same senior audit and evaluation committee. Again, it seems to me that the usefulness of having such a structural link depends on the roles seen for audit and evaluation. The more audit is seen as an aid to managers, the more reasonable it is to have the two closely linked. However, if audit is seen to primarily play an oversight function concerned with regularity matters, providing assurance to senior management on the robustness of the organization's systems and procedures, then the link with evaluation is more tenuous.²²

In reviewing documents and conducting interviews, it appears the combination of audit and evaluation in EAU is providing useful synergies that are appreciated by Management. For example, the Director General of Irish Aid specifically noted that he found the combination of audit and evaluation reports helpful to him in getting a full picture of what was happening within the agency. That said, it is incumbent on the Unit to make clear the respective roles audit and evaluation play; in this regard, Mayne provides a table that sets out some of the ideas coming from the literature. (See Table 2, below.) Of course, that table is a highly aggregated and simplified summary of the division of labor between evaluators and auditors; in practice, the work overlaps to different degrees in different agencies, as Mayne notes in the quotation above.

EAU also has established a presence for evaluation within the management structure of the Development Cooperation Division of the Department. The Director of EAU attends meetings of the Senior Management Group, which provides a comprehensive overview of its spending and management, helping to identify areas that need evaluation and to monitor follow-through on evaluation reports. The Director also sits on the Executive Management Board, a subgroup of the Department's Management Board.

²² John Mayne, "Audit and Evaluation in Public Management: Challenges, Reforms, and Different Roles, *The Canadian Journal of Program Evaluation*, 21:1, 2006, pp. 11-45. Quotation is from p. 38.

Table 2: Roles for Audit and Evaluation

Roles and Preferred Practices for Audit and Evaluation

	Optimal Roles	Optimal Performance Issues Addressed	Comments
External performance audit	<ul style="list-style-type: none"> external checking to provide assurance external assessing of performance 	<ul style="list-style-type: none"> management practices systems and procedures 	<ul style="list-style-type: none"> ought <i>not</i> to assess impact (other than as band-aid audit work); impact assessment should be a government responsibility; effectiveness audits risk audit's reputation external independence provides a significantly different perspective from evaluation
Internal performance audit	<ul style="list-style-type: none"> internal checking to provide assurance internal assisting and advising 	<ul style="list-style-type: none"> management practices systems and procedures 	<ul style="list-style-type: none"> impact/ effectiveness issues ought to be left for evaluation assurance to senior management ought to be the key role
Evaluation	<ul style="list-style-type: none"> assisting and advising assessing of performance 	<ul style="list-style-type: none"> impact of programs and policies why programs are not working 	<ul style="list-style-type: none"> organizational systems and procedures issues ought to be left to internal audit how to set up evaluation in an organization needs to carefully consider the intended purpose(s) of the evaluation activity

In addition, EAU benefits from reporting to the Audit Committee, which not only provides an independent review of DFAs risk management and internal control systems, but also of its audit and evaluation arrangements. Importantly, the Committee also follows up the recommendations from evaluations and audits and sees that they are implemented. The Committee thus can be a source of reinforcement not only for strengthening the enabling environment for evaluation, but in enhancing the use of evaluations by the organization.

An issue is lack of clarity on the role EAU can and should play in relation to non-EAU or decentralized evaluations and related studies. The lack of input into, or even awareness of those evaluations by the evaluation unit raises a potential institutional risk, particularly if the results of such studies are used in EAU independent evaluations. The ability of EAU to mitigate such risks is limited, but it can continue to provide guidance and possibly develop a help desk service, though the costs of the latter might be prohibitive.

Recommendation: EAU should:

- promote a policy for decentralized evaluations for the institution as a whole; and
- conduct a periodic review or meta-evaluation of decentralized evaluations and report to Management on any areas in need of improvement.

Independence of Evaluations and the Evaluation System

Organizational independence

EAU meets the strongest standard for organizational independence. It reports to the Secretary General, not to an operational manager or a lower-level senior manager. However, the presence of the Director of EAU on Management bodies, such as the Senior Management Group and Executive Management Board does raise the possibility that this independence could be compromised. That said, it is not without precedent; for example, at times the Director-General of the World Bank's Independent Evaluation Group (IEG) has participated in regular meetings of the Senior Management Team, including the President, with no apparent effects on IEG's independence. As reported above, the Evaluation Director in Finland also sits on Management boards without compromising independence. The Peer Review Team did not find any evidence of compromised independence in Ireland, either.

Behavioral independence

In interviews there was universal agreement that EAU is behaviorally independent and willing to "speak truth to power." A review of a small sample of reports confirms this view.

One issue raised by EAU, however, was whether this independence is compromised by the use of combined teams with both EAU staff and external consultants. As discussed in the cross-agency section above and in more detail in the case of Switzerland below, staff participation in the evaluation teams does not compromise independence, which inheres in the Unit, not in the external evaluation team.

That said, one issue that was brought to the Team's attention is a possible conflict between the internal manager and external team leader roles. This generally reflects some ambiguity about roles and responsibilities, which should be made clear at the outset of the evaluation, to the extent feasible (given that circumstances may change during the course of the work). For external consultants who do have concerns that their role is being compromised during the course of the evaluation there is a grievance procedure that can be accessed to resolve them.

Conflicts of interest and external pressure

The Peer Review Team found no evidence of evaluations being compromised by conflicts of interest or external pressures on the evaluators or EAU. However, the potential for conflict is implicit in the personnel rotation system, as discussed in Chapter 1. This issue relates to the recruitment and retention of staff with evaluation knowledge and skills in a system of Department-wide staff rotation. EAU has been able to recruit such staff, and to provide training where necessary.

However, retention remains an issue because without a career path evaluation staff move to other parts of the organization, bringing with them whatever evaluation skills they may have acquired and their institutional knowledge. As discussed in Chapter 1, while this diffusion of evaluation experience is beneficial for the Department as a whole, it presents a serious challenge to EAU,

which must recruit and train new staff. As noted, it would be impractical and beyond the scope of this review to suggest that evaluation units be exempted from normal organizational assignment policies, or that an evaluation career be established. But it may be possible to work with Management to identify a number of positions that can be filled for an extended period by staff with evaluation expertise, including the possibility for promotions. This would serve not only a recruitment and retention need, but also help to re-enforce independence by having the option for some staff to build evaluation careers within the Department.

Recommendations: EAU should work with DFA Management to ensure that HR policies or protocols enable the recruitment and retention of staff with evaluation knowledge and skills, including possible exceptions to general policies of staff rotation in the organization.

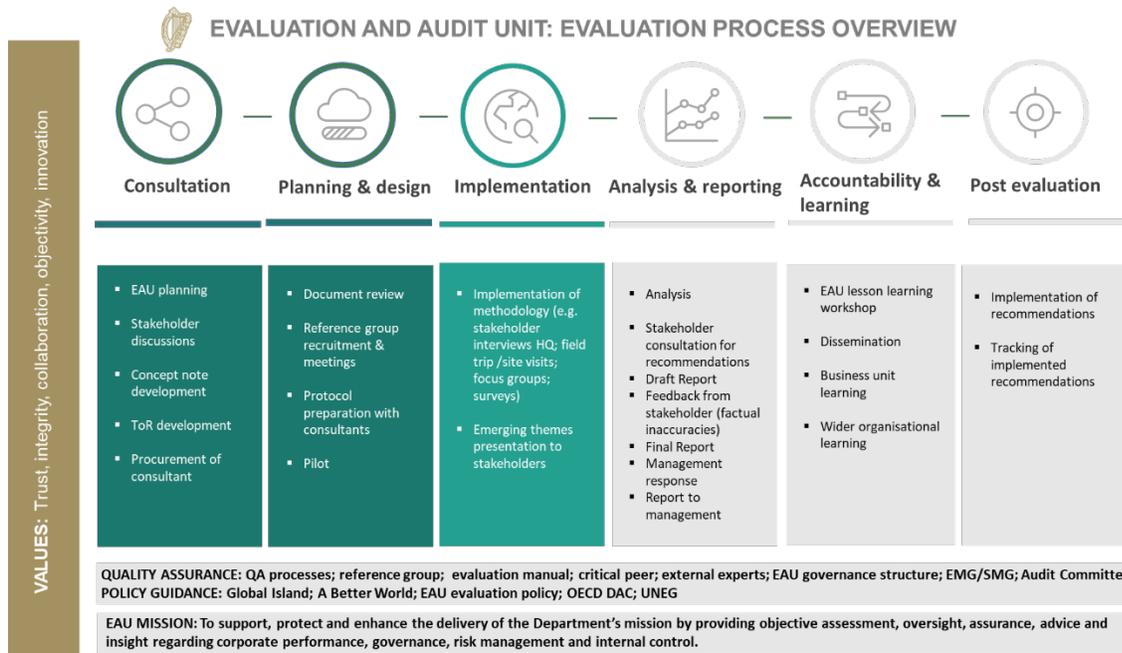
Credibility of Evaluations

As part of its multi-year rebuilding, EAU has produced a number of guidance documents, including one on quality assurance.²³ Though brief (six pages of substantive text) the paper discusses most of the major issues in the evaluation process. One particularly useful feature of the paper is a visual overview of the evaluation process. (See Figure 3.) This should be especially useful for operational staff whose work is being evaluated, allowing them to get an overview of the process from start to finish.

A shortcoming of the paper, however, is that discussion of the process of quality assurance is short, little more than half a page, and while it covers the main topics it does so at a fairly general level. Assuming this document is intended for non-EAU staff, that likely makes sense. Presumably the new policy, or related materials, will flesh out these steps for use by EAU staff to ensure that quality assurance is carried out to a high standard. Most important, the implementation of these quality assurance processes needs to be monitored to ensure they are working in practice, and to identify any needed changes based on experience.

²³*Evaluation and Audit Unit: Quality Assurance in Evaluation*, Brief guidance note, 2020.

Figure 3: EAU process overview



The credibility of EAU's work also is aided by participation of its staff on the evaluation teams. As discussed above, this brings to bear institutional knowledge that can ensure the evaluations are relevant to the operational staff and Management of DFA. For one thing, EAU should have a better understanding of what internal DFA units will find useful in an evaluation than external consultants, even those with substantive, as well as methodological, expertise. And if not, it readily can find out from those in the organization with stakeholder interests in any evaluation.

As with Finland's EVA, a good place to start would be with the reference groups set up to advise on each evaluation at various stages of the work:

- during preparation of the Concept Note outlining the purpose, objective, utility, main questions for the evaluation;
- at the start-up meeting with the evaluation team;
- when commenting on the draft Inception Report;
- as part of a validation exercise when the team presents initial findings/data;
- in commenting on the draft final report.

Research into the structure and use of reference groups is scarce, but what there is suggests such groups should include:

- experts with technical evaluation expertise (e.g., design, data collection and analysis);
- stakeholders with interest in the outcome of the evaluation;
- individuals who can help craft and disseminate the evaluation messages.

Similar to the other units reviewed, EAU is not the primary locus of country and substantive issue area knowledge, and nor in strategic communications. Reference groups can help fill any such gaps, providing greater credibility to its evaluations, and possibly helping to promote wider use of its work

In addition to reference groups, the Evaluation Unit in Finland has introduced “critical friends” with evaluation and/or substantive expertise to assist the evaluation managers in designing and managing evaluations. This also could be considered by EAU.

Recommendation: To promote the effectiveness of reference groups, EAU should develop specific guidelines for:

- the selection of members of reference groups to ensure that a skill mix reflecting evaluation design and methods, stakeholders with substantive knowledge of the subject matter and countries covered by the evaluation and, if possible, with skill in shaping and disseminating messages; and
- conduct of the reference group, including which tasks are expected to involve formal meetings and which may not.

It might be useful for EAU to get such groups engaged early in the process, including at the Terms of Reference stage. Currently, TORs are developed with input from relevant business units, which is good practice for both credibility and utilization. Advice from the Policy Development Group or the EMG also could help build ownership of the evaluation and buy-in to the results when the report is submitted. But Reference Groups also could provide additional insights from perhaps a broader perspective.

Utility of evaluations

EAU has taken a number of steps to ensure that its evaluations are useful and used. It consults closely with business units to identify evaluation gaps as part of its planning process, and on the TORs for individual evaluations. This is common practice. But it also has developed something less common: a set of explicit criteria for selecting evaluations to be carried out. (see Table 3, below.)

This is a good tool that might usefully be emulated by the other evaluation functions. Making explicit the criteria for selection not only provides guidance to EAU staff, but also makes the choices transparent and defensible to the organization as a whole, and even to external audiences. In particular, it highlights the possible uses for the evaluation. One thing that is not clear, however, is the extent to which these criteria are prioritized or weighted in the planning process; that could be made more explicit.

During the Peer Team’s interviews, assessments of the usefulness of evaluations varied. In general, more senior Managers often found them useful because they addressed strategic issues on which they had to make decisions. Mid-level managers, however, tended to find the evaluations

somewhat less immediately useful because they were less focused on operational issues. This finding is common across for evaluation units such as EAU, and reflects the role central evaluation units were designed to play in their respective organizations. Still, it is worthwhile to reflect on ways evaluation findings and recommendations could be used to better address the needs of operational staff better, such as by providing in-depth feedback to teams from case studies that looked at their own work. IEG has done some work along these lines, for example. (See the recommendation at the end of this section.)

Table 3: Criteria for Selecting Evaluations

1. Importance to Senior Management*	<ul style="list-style-type: none"> ▪ Is the evaluation likely to inform high level, strategic decision-making by Senior Management? ▪ Can the evaluation make a significant contribution to achieving the Department’s High Level Goals?
2. Importance to Business Unit	<ul style="list-style-type: none"> ▪ Is there clarity about how the evaluation will contribute to the work and effectiveness of the Business Unit?
3. Contribution to Evidence Base	<ul style="list-style-type: none"> ▪ Will the evaluation address an important evidence gap? ▪ Will the evaluation contribute to knowledge and learning?
4. Financial scale	<ul style="list-style-type: none"> ▪ Is the area a major focus of expenditure? ▪ Will the evaluation outcome influence future funding?
5. Inherent Risk	<ul style="list-style-type: none"> ▪ Is there a substantial level of risk? ▪ Could evaluation contribute to the mitigation of this risk?
6. Innovation	<ul style="list-style-type: none"> ▪ Is the evaluation innovative in its approach? ▪ Does it extend the reach of evaluation into new areas of DFA?
7. Feasibility	<ul style="list-style-type: none"> ▪ Is necessary data available and have baselines been established? ▪ Are relevant external stakeholders available to participate?
8. Timeliness	<ul style="list-style-type: none"> ▪ Will evaluation findings be available in time to inform a key decision?
9. Accountability	<ul style="list-style-type: none"> ▪ Is there a clear need to demonstrate upward and/or downward accountability?
10. Complementarity	<ul style="list-style-type: none"> ▪ Has the business unit established that there are no learnings available from other bodies of work carried out recently?

*Secretary General, Management Board & Senior Management Group

Dissemination and Communication

One area where EAU can make further progress is in disseminating its work and communicating its messages to key stakeholders. Currently, evaluation teams are expected to develop plans for dissemination and communication, but this normally is not an area of expertise for them. Typically, reports are posted on the Web, and some limited communications are carried out with operating units within the Department and some external audiences. However, there is no communication strategy to guide these efforts, so they are somewhat hit or miss.

The obvious solution is to bring on board staff with expertise in communications. But this can be costly, and in any case it is unlikely that a full-time staff person would be needed, given the volume of output produced. A possible solution is to work with an external communication consultant to develop a communication strategy that addressed both the objectives to be achieved and the resources needed to meet them. That would provide a strategic way to weigh the costs and benefits of alternative approaches and a logical basis for decisions on how to improve the communication function in the Unit. Of course, any decisions would have to take account of the communication policies of the Department, so would require consultation with the official communication unit. For broader knowledge sharing, EAU could consider the possibility of establishing a Help Desk or other knowledge management mechanism; a number of approaches have been tried at other development organizations (World Bank, United Nations, US Agency for International Development, as examples) that could serve as possible models.

Another issue is that, unlike the units in Finland and Switzerland, EAU does not have a direct line to parliamentarians, which further complicates its ability to get out the messages from its work to relevant audiences. This is an issue largely beyond the control of EAU because it has to do with the larger governance system. One possibility that was discussed during the Peer Review Team's visit to Dublin was enlisting the Audit Committee to help make this connection. That Committee provides its reports to Management, not to parliamentarians, so it is not clear that it could fulfill this role. Whether the Committee can or should do this, or whether alternative structures are available or could be created to do so, should be part of the development of any communication strategy for EAU, however.

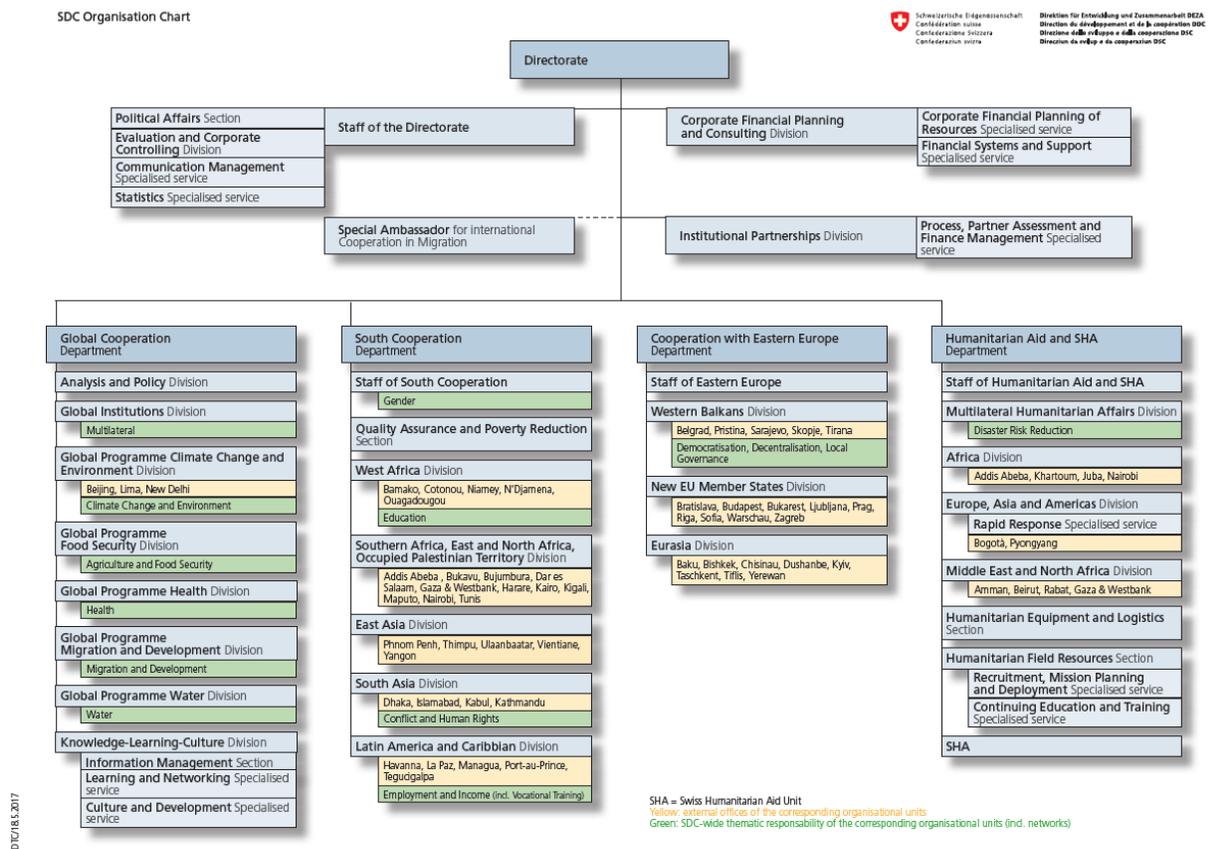
Recommendation: To strengthen its dissemination and communications of evaluation findings EAU should:

- work with DA's communications unit or an external expert to develop a communication strategy that addresses such issues as how to reach internal and external stakeholder audiences, including parliamentarians;
- use the planning process to identify opportunities to employ evaluations for multiple purposes, and set priorities on whom the evaluations are expected to influence;
- include a communication/dissemination plan for individual evaluations from the start; and
- consider adopting the recommendations on utility in Chapter 1.

CHAPTER 4: EVALUATION IN SDC

SDC differentiates two types of evaluations, depending on the commissioning entities: (1) centralized evaluations commissioned by the senior management and managed by the Evaluations and Corporate Controlling Division (E+C); and (2) decentralized evaluations commissioned by the operational units of the four SDC departments, or the Swiss cooperation offices abroad. The E+C is part of the staff of the Directorate. The Head of the Division reports directly to the head of the staff of the Directorate. (see Figure 4.)

Figure 4: SDC Organization Chart



In addition to overseeing the conduct of independent evaluations, E+C participates in international joint evaluations representing SDC;²⁴ disseminates evaluation results to the public and non – parliamentary commissions; and ensures transparent access to results. Furthermore, E+C is in charge of conducting strategic controlling, reporting to the board of directors on about 30 indicators covering SDC’s work. This is not a typical function for a central evaluation unit; usually it is handled by other units. We did not identify other examples of this arrangement, but this may be because the work is characterized differently in each organization.

²⁴ <http://www.oecd.org/dac/evaluation/network-member-sdc.htm>

In principle, there is no obvious reason why E+C's controlling work necessarily conflicts with its evaluation mandate. One issue that did come up was that the indicators are selected by Management rather than E+C, and therefore could compromise independence. However, since the indicators are used for management purposes, rather than to evaluate SDC's work *per se*, it is not obvious that providing this service impinges on the role of E+C to provide independent evaluations.

Based on the Federal Constitution and legislation, every four years the Federal Council and Parliament defines the strategic approach of Switzerland's international cooperation (Dispatch). E+C has the lead in writing the quadrennial final report that the Federal Council submits to Parliament on the implementation of the Dispatch.²⁵ The report explains how Switzerland helped save lives, provide access to education and healthcare for people living in poverty, create economic prospects, preserve natural resources, build countries' resilience to crises (climate, economic and financial), and promote democracy, peace and respect for human rights. This is not seen as an Evaluation function, as the report takes into account a lot of information in addition to independent evaluations, and is used for political communication.

Enabling Environment for Evaluation

There is a reasonably strong enabling environment for the evaluation function, though with room for improvement. E+C reports go not only to senior management but to Parliament, where there has been increasing demand for information on the use and effectiveness of development cooperation expenditures in order to foster accountability. At the same time, this creates a natural connection between evaluation and policy-making that is supportive of E+C's role within SDC in informing management decisions (or steering) and promoting organizational learning.

In this environment, E+C staff confirmed that they have adequate budget resources and access to the information they need to do their work. They also reported that operational staff generally cooperate with evaluation teams, even though there is a burden on them in doing so. Managers and operational staff expressed similar willingness to work with evaluators, and to see value in their work, when interviewed. But staff also cited some concerns arising from their general lack of formal evaluation training and the career prospects that are discussed below.

The Evaluation Policy, revised in 2018, includes an overview of the evaluation architecture in SDC and an evaluation framework based on OECD-DAC principles and criteria. Although it has a small section on decentralized evaluations, it mostly centers on centralized evaluation as conducted by E+C. In reviewing this policy, the team identified several issues. First, the policy is not clear about the distinction between centralized and decentralized evaluations. In fact, neither term is used in the document. The policy simply says that:

²⁵ *Switzerland's international cooperation is working. Final report on the implementation of the Dispatch 2017–20.* <https://www.eda.admin.ch/deza/en/home/publicationsservices/publications.html/content/publikationen/en/deza/rechenschaftsberichte/Schlussbericht-Umsetzung-Botschaft-2017-2020>

“SDC considers all evaluations to be strategic within its RBM system but differentiates between the following commissioning entities:

- **Evaluations commissioned by the senior management and managed** by the Evaluation and Corporate Controlling Division (E+C);
- **Evaluations commissioned by the operational units** of the four SDC departments, or the Swiss cooperation offices abroad.”

The lack of clarity on this issue has implications that are discussed below.

Second, the Policy also provides only a very general evaluative framework, reiterating OECD-DAC evaluation principles and criteria, but with no discussion of how these are to be operationalized within the context of SDC. For example, it provides for a Management response to evaluations, but is silent on who is expected to draft that response, what vetting process is to be used to ensure Senior Management approval, and what timeline is to be followed. This has led to instances in which E+C staff have found themselves assisting in drafting the response, which threatens independence; and other cases in which responses were long-delayed.

Third, aside from calling for a Management response, the Policy provides no overarching framework to ensure the input of findings, lessons, and recommendations from evaluation into operational work. What is needed is something like the Evaluation Operations Manual developed by the Evaluation and Audit Unit in Ireland. This can usefully fill most of these gaps, without the need to go through a process of developing a new Evaluation Policy, and presumably can be done more expeditiously.

Finally, the policy says nothing about the human resources required to carry out E+C’s responsibilities. In general, three kinds of knowledge and experience are relevant to SDC’s evaluation work:

- evaluation methods, techniques, and processes.
- development as a field, including sector/thematic and country/regional experience.
- institutional knowledge of SDC’s institutional culture, processes, and constraints.

Evaluation knowledge is important primarily to allow E+C to provide strong quality control; the actual evaluation work is done by external consultants, but the team needs to understand what they are doing and whether it is of high technical quality. Development knowledge, gained mostly through experience, helps the team to identify the key issues that require evaluation, and to ensure that evaluation reports are grounded in sector/thematic and country/regional realities. Institutional knowledge is a key to helping the evaluation team conduct its work as smoothly and efficiently as possible, and ensuring that recommendations and lessons from evaluation take account of the institution’s capacity to put them into practice. Not everyone in E+C must have a strong background in all three areas, but the team as a whole needs to cover this ground. And the work environment within the group must foster sharing of relevant knowledge across the team to promote high performance in all three areas.

This has implications for the recruitment, training, and retention of E+C staff. For the most part, current members of the team have had little training or experience in evaluation; rather, they tend to have backgrounds in development and related areas. It is unlikely that the unit will be able to recruit many staff with evaluation expertise; the number of such individuals available to SDC is limited, and there is no real career path within the organization for such specialists. As a result, it is important that E+C invest in training staff to at least a moderate proficiency in evaluation so that they can oversee the work of consultants. The extent to which such training can be provided depends on the resources that can be made available, given the multiple demands on the unit's budget. But this should be given as much attention as is feasible within that constraint.

Retention is a more complex issue because without a career path E+C staff move to other parts of the organization, bringing with them whatever evaluation skills they may have acquired and their institutional knowledge of E+C. From the broad view of SDC, this diffusion of evaluation experience is beneficial, since evaluation work is going on throughout the organization. However, for E+C this is a serious challenge because it constantly must train new staff; and, as discussed below, it is a potential threat to E+C's independence. It would be impractical to suggest that E+C be exempted from normal SDC assignment policies, or that an evaluation career be established; and in any case such ideas would be beyond the remit of this report. However, it may be possible to work with SDC Management to identify one or two positions that can be filled for an extended period by staff with evaluation expertise, including the possibility for promotions.

Recommendation: The evaluation unit should

- consider working with management on the possibility of having one or two slots filled by trained evaluators who could be assigned for a long term, outside normal rotation, and have promotion opportunities, and develop a job description, including qualification requirements, to recruit for the position(s);
- provide staff with appropriate training, within budget constraints, possibly before they rotate into the unit; and
- develop an Operations Manual to clarify the evaluation types and terminology to be used, provide guidance on the Management response process, and outline process to promote evaluation use.

Independence of Evaluations and the Evaluation System

Organizational independence

The organizational independence of E+C is strong. The unit reports directly to the Head of Staff of the SDC Directorate, not to a line manager. It has a secure budget for its evaluation work. During the visit to Bern, the Peer Review team identified no issues in this area.

Behavioral independence

Based on interviews and a review of a small sample of reports, E+C's behavioral independence also seems high, though with some potential issues. The reports are hard-hitting, and generally

address the issues raised for the evaluations straightforwardly. There is no evidence that behavioral independence has been compromised.

However, there are issues that were raised, mostly by E+C itself, that can affect behavioral independence. As discussed above, lack of formal training and staff rotation can be limiting factors within the enabling environment. And they also can threaten behavioral independence in two ways. First, lack of evaluation expertise makes it more difficult for E+C staff to provide effective oversight of the evaluations being conducted by external consultants, particularly in terms of quality control. Second, the need to apply for follow-on appointments has the potential to make staff wary of being engaged in negative reports over concerns for career mobility.

E+C's leadership of the quadrennial report of the outcomes of the Dispatch also raises questions of at least the appearance of a lack of behavioral independence. The report itself seems like a useful product, similar to those done at other institutions. However, in those cases the role of the evaluation unit is limited to supplying inputs rather than drafting what is in the end a Management report. For example, at the World Bank the Independent Evaluation Group (IEG) contributes evaluation findings and lessons to an annual report to the Board that is similar to the Dispatch report at SDC, but it does not draft the report; a Management unit has that responsibility. And in Finland, the evaluation unit also contributes to the report to Parliament, but the report itself is written the Department for Development Policy.

It is important to note that the Peer Review team did not observe, nor did any interviewee report, actual compromises of behavioral independence. However, the recommendations cited above are intended to address not only the enabling environment, but also the potential threats to behavioral independence that do exist.

Recommendation: E+C should seek to have Management give the responsibility for drafting the quadrennial report on the Dispatch results to another, more appropriate unit handling political communication, limiting E+C's role to providing inputs based on independent evaluations.

Avoiding conflicts of interest

The Evaluation Policy does not discuss conflicts of interest, but it is an important issue. From interviews it is clear that E+C vets external consultants to ensure that they are not evaluating activities with which they have a pre-existing relationship. This always is a danger for any organization using outside consultants as evaluators, particularly in fields where the choice of possible consultants is limited. It would be useful if any Operations Manual that might be developed were to provide explicit guidelines for assessing potential conflicts, however.

At the same time conflicts also can arise internally. E+C typically establishes a Core Learning Partnership (CLP), representing the primary intended stakeholders of the evaluation, to provide comments at a number of stages of the evaluation process: the evaluation design (Approach

Paper); preliminary findings; and draft report. However, the CLP members typically are closely associated with various aspects of the matter being evaluated. Their input is valuable in ensuring the relevance, quality, and usefulness of the evaluation, but they do come to it with a vested interest. This is a risk, but one that clearly is manageable, and from the Peer Review team's work there were no examples of compromised independence from this source. Conversely, participants in CLPs reported that they came to a fuller appreciation of evaluation as a result of taking part on a CLP.

In principle, based on the final report, the CLP assumes the responsibility for drafting a Senior Management Response (SMR). The SMR is subsequently approved by SDC's Board of Directors. However, the Peer Review team learned in interviews that often the process drags on for many weeks, or even months. The fact that the CLP is responsible for drafting the SMR is problematic. The CLP is seen by some as a "fan club" that will try to keep the status quo instead of implementing recommendations and proposed structural changes where needed.

Long delays in producing the SMR also delay publication of the evaluation reports. This is a vulnerability for E+C because it can limit the effectiveness of its independence; for example, such delays could be used strategically to blunt the impact of an independent evaluation on decisions that occur in the interim. Other evaluation organizations, such as Finland's, assign the responsibility for drafting the Management Response to a specific individual.

Recommendation: The Management response process should be reconsidered to provide for:

- a fixed time limit for Senior Management to produce the response, after which the report may be published without one if the SMR is not completed;
- appointment of a Management designee, rather than the CLP, to produce the SMR, with the ability to coordinate among the affected SDC units; and
- clear guidance on drafting the Management response in the Evaluation Policy, including that E+C staff are not to assist in producing the SMR, except through clarifying the report's findings and recommendations and discussing the feasibility of implementing recommendations.

In addition, the Peer Team learned that on occasion E+C staff are asked to help in drafting the SMR. This should be avoided because it raises questions about E+C's independence from Management. There is nothing wrong with E+C staff discussing possible responses with Management to clarify report contents or the feasibility of implementing recommendations. But staff should not actively take on the job of drafting SMRs.

Finally, as discussed in Chapter 1 the potential for conflict is implicit in the personnel rotation system, particularly as related to the recruitment and retention of staff with evaluation knowledge, given the SDC-wide system of staff rotation. E+C has been able to provide training where necessary, but retention remains an issue. Without a clear career path, evaluation staff move to other parts of the organization, bringing with them whatever evaluation skills they may have

acquired and their institutional knowledge. As discussed earlier in this report, this diffusion of evaluation experience is beneficial for SDC as a whole. But for E+C it means recruiting and training new staff to replace those moving out of the unit. The scope of this review does not extend to an exploration of SDC's general assignment policies, or career paths. But the findings do suggest that it may be possible to work with Management to identify a number of positions that can be filled for an extended period by staff with evaluation expertise, including the possibility for promotions. This would serve not only a recruitment and retention need, but also help to re-enforce independence by having the option for some staff to build evaluation careers within the Department.

Recommendation: E+C should work with SDC Management to ensure that HR policies or protocols enable the recruitment and retention of staff with evaluation knowledge and skills, including possible exceptions to general policies of staff rotation in the organization.

Protection from outside influence

During the review, the Peer Review team did not identify any evidence of undue external influence over E+C evaluations. Staff of the unit noted that, while SDC staff did often attempt to make a case for a particular interpretation of their work being evaluated, this was little more than “lobbying” and did not involve any form of pressure or compromise the work.

Credibility of evaluations

Interviews across SDC indicate that E+C's evaluations are seen as credible, even though quality inevitably varies among reports. The CLPs discussed above help to promote credibility since it is understood that they can signal potential issues throughout the process. The process of getting comments from Management and operational staff—and the way that E+C responds to those comments—also contributes to credibility. And the expertise of the external evaluators, particularly those who have institutional knowledge of SDC, also is a source of credibility. However, it is worth noting that all these factors are variables; they are not uniform across all evaluations, and it is necessary for E+C to maximize the credibility they confer through the choice of evaluators and CLP members, and in dealings with Management.

One issue that did arise was the difficulty E+C staff have in developing Terms of Reference (TOR) for the evaluations. Getting the TORs right is one of the most important steps in the evaluation process, but E+C staff generally lack the formal evaluation training and experience in conducting evaluations that make it possible to draft high-quality TORs. As noted above, one possible answer is training, budget permitting. It also could be possible to use an external consultant with evaluation expertise and experience to review TORs to ensure they are of high quality.

One potential threat to credibility comes from the fact that E+C uses decentralized evaluations to calculate a performance rate for SDC work, and often uses the results of decentralized evaluations in its own studies, but has no control over those evaluations. This means that it cannot vouch for the quality of that work, even while using the findings reported. The Quality Assurance Unit does

have a role in this regard, and has organized training and manuals to promote high-quality evaluations, but with 80-100 decentralized evaluations being conducted each year, there is no mechanism to assure their quality.

One possibility would be to have E+C take a direct role in validating the findings of decentralized evaluations, especially completion reports. This is the model that was developed by the Independent Evaluation Group (IEG) at the World Bank in the 1970s, and continues to this day. In that case, IEG independently reviews every project completion report, and its ratings are accepted as the official Bank ratings. However, this is a costly enterprise, taking up approximately 20 percent of IEG's annual budget. It is unlikely that E+C could adopt such a model, given its limited staff. The World Bank also has an internal quality function that regularly reviews the quality of such evaluations, but it is unlikely QA would have the level of resources necessary to take on that function, either.

This means that to the extent E+C relies on decentralized evaluations it takes on a risk that the work was not of sufficient quality to be relied upon, and thus a reputational risk to its own work. (This applies to SDC as a whole, as well.) There are several ways to mitigate this risk:

- SDC's evaluation policy could clearly define what constitutes an evaluation. Many decentralized "evaluations" can be described more accurately as reviews or other kinds of studies than evaluations.
- To the extent the budget allows, E+C could provide help in the form of a help desk, guidance, and/or training to operations staff. However, it is not clear that E+C has the capacity to conduct these activities, and any effort might not be widely accepted, given that the unit has no functional responsibility for such reports. A possible model is provided by Finland, where an external consulting firm provides a help desk function (See Box 4.)
- One thing it can do is to conduct a review of the quality of decentralized evaluations as part of its work program on a regular basis (e.g., every three to five years). The advantage of this approach is that it is much less resource-intensive than doing validations of each decentralized evaluation, and could be used to highlight for Management any quality issues with that work, which presumably would lead to efforts to improve on any deficiencies/

Steps such as these could help to support the credibility not only of E+C's own work but of all SDC's. Most immediately, they would make the performance rate reporting more credible, both inside and outside SDC.

That assumes E+C are committed to continuing to produce the Performance Rate data. As noted above, this has been a staple of World Bank evaluation work for nearly 50 years, but many national development agencies do not publish them. On the one hand, they do provide a quick overview of how well the development cooperation work is going, and particularly whether it is more or less successful over time. Many consumers of evaluation reports, not least parliamentarians, find these summary data compelling and useful. On the other hand, however, they are a relatively crude measure, since they do not take account of the non-quantifiable differences in importance among projects and programs. And they can be misused outside the organization to present an

incomplete or misleading picture. On balance, the organizations that traditionally have produced them see value in the exercise, and even have elaborated on it. IEG, for example, uses these data to produce a “dashboard” disaggregating the data by sectors, regions, and other factors.

Recommendations: In order to ensure the quality and utility of decentralized evaluations, E+C should:

- work with Management to revise the Evaluation Policy so that it more clearly defines what constitutes an evaluation, with special attention to decentralized evaluations;
- explore ways to provide assistance to those responsible for decentralized evaluations, possibly using the Finland example of an outsourced help desk for decentralized evaluations and a framework contract for quality assurance; and
- consider conducting periodic reviews of the quality of those evaluations and report the findings to Management, possibly as part of a meta-evaluation with major common findings and recommendations of decentralized evaluations (for example by themes).

Utility of evaluations

Overall, E+C’s evaluations have been found useful by various users. The topics for evaluation are demand-driven, so that they have a ready audience interested in their findings, lessons, and recommendations. The inclusion of CLPs helps to promote their relevance and utility within SDC.

But there are some issues with utilization of reports. For one thing, there are many audiences for any given evaluation, but their information needs are not uniform and no one evaluation is likely to satisfy all those users. For example, in interviews some operational staff complained that hosting an evaluation team in-country was time-consuming and sometimes onerous. Yet the reports tended to address issues at a much higher level of the organization than the country teams can affect, and often for thematic and institutional evaluations, the report simply treats their country work as a supporting case study for a broader evaluation, limiting its utility to them.

This is inevitable. But addressing the issue requires some clear thinking about priorities. In particular, whom does E+C intend to influence with its evaluations, both in general and for specific evaluations? In the case of Ireland there is a clear statement of priorities that guides these decisions transparently. This would seem to be a good practice. Normally, a central evaluation unit speaks primarily to the governing body and Senior Management of the organization. That is the strategic value of centralized evaluations. Decentralized evaluations typically are focused on more operational units within the organization. There is a rough division of labor here that, in most cases, makes sense. Any priority system would need to take that into account.

Then there is the question of which evaluations to undertake. As at most development agencies, E+C consults across the organization to identify pending issues that can be addressed by evaluation and determines which to conduct, given its budget and other constraints. One way to strengthen this decision-making process is to identify evaluations that may serve multiple purposes.

Utilization, however, requires a not only this kind of alignment with strategic institutional priorities, but also attention to communication and dissemination, and actual use. Understandably, this is not an easy issue. Formally, evaluation reports get to the appropriate individuals within the organization, are discussed by Senior Management, provoke a response, and ultimately affect institutional behavior. Beyond that, evaluation reports are disseminated to other interested officials, media, academicians and researchers, and to the general public. Most of these audiences are not the direct targets of the evaluations, nor are they expected to be the most important users (though sometimes they are).

Finally, E+C may wish to explore other methods for disseminating its reports and communicating findings to a broader internal and external audience. Mostly, it puts reports on the SDC website, which is necessary. But it may wish to explore using information on hits to that website to identify possible users who can be notified when new reports are available, or through a periodic electronic announcement of recent reports, for example. Admittedly, budgets are limited, but there are some inexpensive ways to help promote awareness and use of E+C evaluations. One possibility is to hold a public presentation of the evaluation, or a group of evaluations.

Even with all such efforts, it should be remembered that utilization ultimately depends on the ability and willingness of the line decision-makers to use evaluation findings, conclusions, lessons, and recommendations. The most obvious measure of utility is the extent to which Management uses the findings and recommendations from independent evaluation in subsequent work. This is why the Management response is an important part of the utilization process, and why it is necessary to ensure implementation of the response. E+C tracks such implementation and reports on actions taken until there is at least 80 percent compliance. Sometimes, even when there is broad agreement between evaluators and managers on these issues, adoption of evaluation results may be incomplete. (see Box 5, below.)

Recommendation: E+C should consider:

- identify cost-effective ways to better communicate the results of its evaluations both electronically and in other forms; and
- consider adopting the recommendations on utility in Chapter 1.

Box 5: The Nexus of Development Assistance and Humanitarian Aid

SDC Management welcomed a review of issues related to the “nexus” between development assistance and humanitarian aid as part of the Afghanistan Country Strategy Evaluation. E+C subsequently carried out a review “to assess whether and how SDC’s institutional and operational approaches to link humanitarian aid with development cooperation can be strengthened”.

The resulting report, *Independent Evaluation of the Linkage of Humanitarian Aid and Development Cooperation at the Swiss Development Cooperation* (2019), found that there had been “a continuous and positive reflection within SDC on how to strengthen the humanitarian-development linkages particularly in protracted crises,” but that the actual process changes at SDC “can be characterized as adjustments of existing instruments and processes” rather than addressing the “root” problem, barriers to working across instruments because of: “separate budgets, bureaucracies and supporting political constituencies...exacerbated by incompatible systems.”

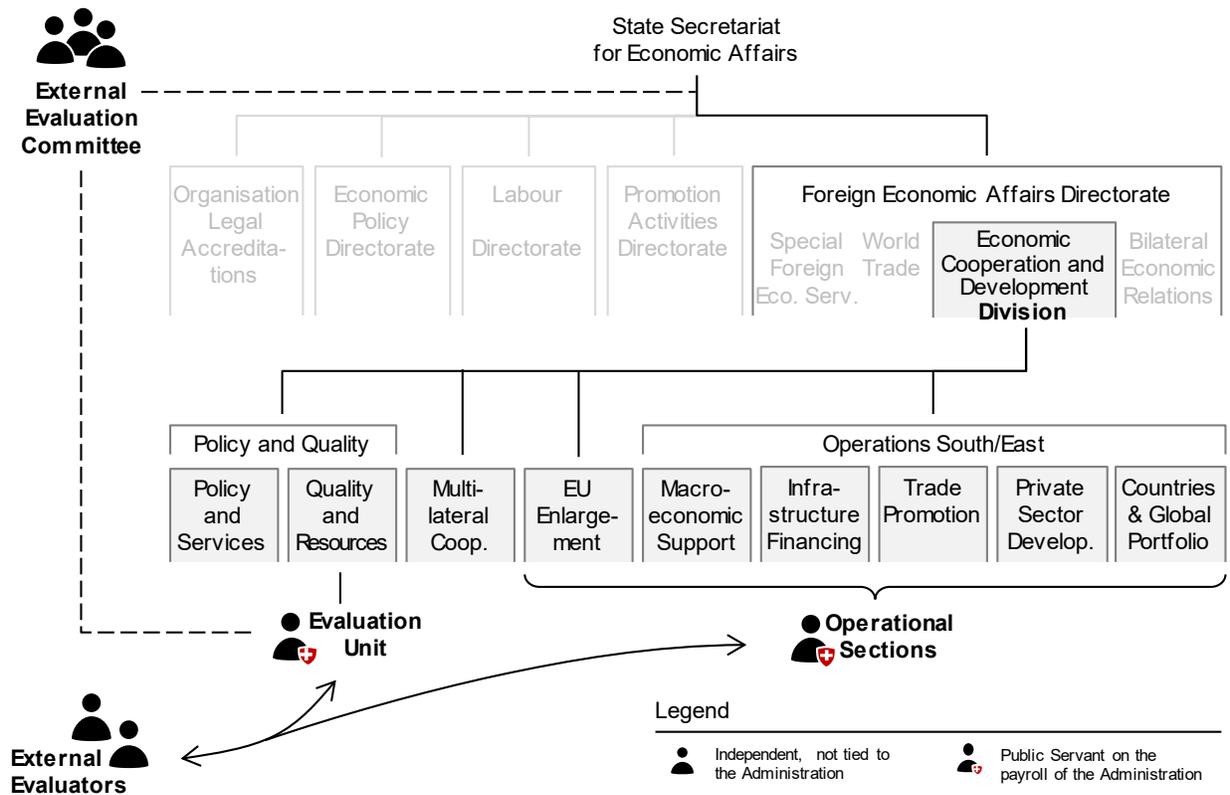
Management generally agreed with most of the findings from the study, and accepted all but one of the nine recommendations, noting that, “This shows a high level of convergence with the evaluation findings.” However, the recommendation it rejected arguably was the most important: “develop a strong business case to lobby the Swiss Parliament for the merger and creation of a single shared framework credit...to address the root cause of many of the current institutional and operational divides between the HA and development cooperation divisions identified by the evaluation.” Management argued that there were legal and practical difficulties, and that it could achieve better instrument coordination in other ways.

CHAPTER 5: EVALUATION IN SECO

SECO's Economic Development and Cooperation Division (SECO WE) manages economic development cooperation on behalf of the Swiss Government. It supports efforts of its partner countries to shape economic structural change and integrate into the global economy. It is, next to the Swiss Agency for Development and Cooperation (SDC), one of the three Swiss Government agencies providing ODA. The Evaluation Unit (EU) is headed by the Evaluation Officer (EO).

In 2009, SECO's evaluation system explicitly adopted a dual focus on both accountability and learning. The Evaluation Policy adopted in the same year declares that evaluation serves both purposes, indicating that evaluation goes beyond providing accountability reports, to contribute to SECO's decision-making process and to foster continuous improvements. The Policy also defines the responsibilities and organizational arrangements guiding the evaluation function in the Division of Economic Cooperation and Development.²⁶ (see Figure 5)

Figure 5: Governance Chart for Evaluation at SECO WE



²⁶ SECO, Directorate of Economic Cooperation and Development, *Evaluation Policy*, 2009.

The Evaluation Policy clearly distinguishes three types of evaluation:

- **Independent evaluations** are approved by the External Evaluation Committee, are commissioned and managed by the Evaluation Unit (EU), and are executed by external independent consultants.
- **External evaluations** are approved by the head of an operational section, are commissioned and managed by SECO Program Officers, and executed by external independent consultants.
- **Internal reviews** are approved by the head of an operational section and executed by SECO Program Officers.

The EU also is responsible for publishing the *Bi-Annual Performance Report*, providing a meta-analysis of the results of all external evaluations and internal reviews conducted during the elapsed year/s covered, and for arranging evaluation training and providing advice to the operational sections and program managers undertaking external evaluations and internal reviews.

Organizationally, the evaluation function is integrated into the Quality and Resources Section of the Division of Economic Cooperation and Development, which reports to Management. (See Figure 5.) In its capacity as part of this management structure, the EU consults with the Heads of Division and Management to plan the annual program of independent evaluations; consolidates a multi-year evaluation plan that includes independent evaluations, external evaluations, and internal reviews; reports on the results of evaluations annually, including assessments of the success rate of development interventions, the quality of evaluations, and progress on implementing recommendations. The EU also is responsible for disseminating the results of evaluations, for providing evaluation training and support to operational divisions, and for contributing to the SDC/SECO joint report on aid effectiveness. In turn, Management is responsible for promoting a culture of learning and accountability throughout the organization, responding to the findings and recommendations from evaluations, and providing the resources needed to carry out the evaluation function.

At the same time, the EU also acts as the Secretariat for an external and independent body, the Evaluation Committee. The EU, therefore, also reports to the Committee, which provides oversight of the evaluation function and in turn reports to the State Secretary (SECO Director). The Committee consists of experts and representatives from academia, business, civil society, international organizations, and Parliament, appointed by the State Secretary. The Committee is responsible for:

- Approving the program of independent evaluations and the related budget allocations.
- Commenting on the independent evaluations and related management responses.
- Reviewing and commenting the annual report on evaluation.
- Providing advice to ensure the quality of the evaluation function.

- Recommending to Management the allocations of additional budget resources to strengthen the evaluation function.

The Committee thus serves both an oversight role vis-à-vis the EU, and a supportive role in promoting the evaluation function. In principle, it is possible that having this dual reporting responsibility could be difficult for the EU to manage; however, interviewees reported that, in practice, this is not an issue, and has not caused problems for the evaluation function.

Enabling Environment for Evaluation

The Review found there is a strong enabling environment for evaluation within SECO. Both independent and project evaluation have clear and focused mandates, and the evaluators have access to needed information. SECO staff and Evaluation Committee members reported in interviews that evaluation is taken seriously by Management and staff, and that there is high interest in Parliament, where funding decisions take account of evaluation results. Evaluation staff are committed to, and reflective about evaluation, while operational staff actively engage with and support evaluation, and are familiar with evaluation concepts.

The EU has 1.3 FTE staffing and a budget amounting to around €220,000 per year. (See Table 1) This is considerably smaller than the other three units reviewed, in line with the relative size of SECO and the Quality and Resources Section within which it is housed, but the EO reports that these resources are adequate for the level of work currently undertaken. The budget depends on the evaluations the EU mandates and for which it requests funds during the annual budget cycle, and a team of external consultants is available for specific assignments as needed, typically for 20-50 days per year. The EU commissions a major evaluation each year focusing on one of the four pillars addressed by SECO in the current development cooperation Dispatch. This evaluation focus is welcome as it provides corporate coverage, yet could be considered as designed more for accountability purposes and less for learning. Also, it limits the capacity of the EU to address further issues in an evaluation. The limitation of one major evaluation per corporate pillar was widely acknowledged by interviewees, although there were no suggestions on how to address the issue. In addition, the small size of the unit limits the range of skills available internally, and tight human resources constraint affects the ability to provide additional training to address skill gaps.

As part of this review, the Peer Review Team was asked to consider how the EU could best put these resources to use. However, it appears that the EU already has a reasonable solution: rotating among the pillars of the institutional strategy, while using remaining resources for other evaluations, is a well-considered approach, with the limitation illustrated above. However, the Team understands that the new Dispatch will not be structured around those four pillars, which may complicate matters for the EU. This change in the Dispatch provides a good opportunity for the EU to work with the Evaluation Committee on a re-assessment of the current staffing for the Unit to ensure it is in line with anticipated needs. That would be in line with the Committee's responsibility to inform Management of resource requirements for the EU, and of Management's responsibility for providing adequate resources for evaluation.

Recommendation: The EU, working with the Evaluation Committee should undertake a needs analysis to ascertain whether additional resources are required to meet the independent evaluation needs of SECO under the new Dispatch. That analysis, which can be conducted by an external expert, should take account of the competing needs and availability of funds that could be directed to independent evaluation.

Independence of Evaluations and the Evaluation System

Organizational independence

As noted above, the EU reports on some issues to the external Evaluation Committee, which in turn reports to the State Secretary, while on others it reports to line Management. That line reporting relationship could compromise the independence of the evaluation function, but two factors mitigate this potential threat. First, the Quality and Resources Section where the EU is located does not conduct operations, and the EU reports through the Evaluation Committee to the State Secretary, who is above the division responsible for economic development cooperation. This is a strong form of internal independence in itself.

In addition, the Evaluation Committee provides additional support for independence, not only because of its mandate to promote independent evaluation, but because it is positioned to act as a check on any Management threats to independence. Of course, there always is the danger that reporting to such an external entity can leave the evaluation function isolated from the organization. However, that is not the case at SECO because of the other reporting relationship through line Management

In sum, the evaluation function in SECO has a strong level of organizational independence.

Behavioral independence

Interviewees uniformly confirmed that the evaluation function shows strong behavioral independence. One interviewee observed that SECO as a whole tends to be more driven by learning than by accountability *per se*, which may reduce conflict over evaluation findings, making behavioral independence easier to maintain.

One aspect of the evaluation function's work that could undermine behavioral independence is its participation in discussions on proposed projects. In principle, this is a good way to bring forward prior evaluation findings and recommendations relevant to new work, and to help promote a design that makes the project more evaluable. But it also could compromise independence because any subsequent evaluation could be wary of criticizing actions that came from evaluation function inputs. However, neither in documents nor in interviews did the Review Team uncover any case where this happened in practice.

Conflicts of interest

As with the other evaluation functions reviewed, SECO's does not make specific provision for avoidance of conflicts of interest in its policy or guidance documents.²⁷ That said, the Review did not find documented instances of such conflicts, nor were any reported in interviews.

However, one potential source for such conflicts comes from the process of identifying and recruiting evaluators to carry out the independent evaluations. Normally, the evaluation function invites possible evaluators to submit proposals to carry out planned evaluations, and selects the evaluator from among those who respond. Sometimes the selected evaluators have prior experience with SECO, including work on operations. However, in an independent evaluation the contract prohibits an evaluator looking at projects where he or she has had prior involvement. As noted below, there are advantages to such selections, especially because of the institutional knowledge the evaluators bring with them. As thematic experts often have prior exposure to development projects of SECO, the available pool of skilled evaluators for independent evaluations is often limited, due to conflict of interest clauses in the standard contracts of the EU.

External pressure

Neither document reviews nor interviews surfaced any undue external pressures on the evaluation function. However, there are potential sources of such pressure. The Evaluation Committee, for example, has been cited above for its role in helping to promote both organizational and behavioral independence. But its role in reviewing and commenting on the EU's work plans and products also gives its members an opportunity to press the unit to adopt a work program or operational practices in a way that compromises its independence.²⁸

While instances of such pressure were not identified in this review, it remains a potential issue that requires monitoring. Similar pressures could come from Management, of course. Fortunately, the double reporting lines would seem to create a system of checks and balances nicely suited to helping the evaluation function maintain its independence.

Credibility of evaluations

The evaluations conducted by the EU are regarded by those interviewed as credible and of high quality. A number of factors seem to contribute to this view within SECO.

First, through its activities within SECO the evaluation function is seen as organically connected to the organization, even while the evaluations themselves are commissioned by the Committee. One interviewee noted that the Unit is "not far from operations" because it gives advice on project evaluation terms of reference and log frames, for example, keeping it in "constant dialog" with

²⁷ A mention is however included in the draft template for project evaluations

²⁸ To be clear, this review did not find any instances of the Committee's members actually exerting such pressure, but experience with oversight boards at the MDBs, for example, does show that this can happen.

operations. This kind of continuous interaction is an important way to maintain credibility with operational staff, though it can cause challenges to independence, as discussed above.

Second, the evaluation plan is clearly aligned with the corporate strategy 2017-20. Not only does this enhance utility, as discussed below, but it also strengthens credibility by demonstrating that the evaluations' findings and recommendations are relevant to SECO's broader strategy.

Third, in designing its work program and evaluations, the EU consults with operational management and staff, and pays attention to their comments and suggestions. This not only helps to achieve strategic alignment, but gives the operational side of the organization a "buy in" to the evaluation work at an early stage. This could be a conduit for undue influence, but the review found no evidence that is the case in practice.

Fourth, some interviewees reported that the fact of having the Evaluation Committee overseeing the evaluation function is a source of credibility. The range of expertise among Committee members provided for assurance that there was sound oversight of the quality of evaluation work.

Finally, as indicated above, some of the external evaluators who conduct independent evaluations have SECO institutional knowledge. Even though this could lead to conflicts of interest, as mentioned, it also provides reassurance to operational managers and staff that the evaluators have a good grasp of the organization, its work, and the constraints under which it works, which reinforces the credibility of the evaluation and enhances the potential for institutionally-relevant recommendations.

Nonetheless, there are some areas where credibility could be strengthened. The *Evaluation Guidelines*²⁹ provide several useful criteria for assessing a draft evaluation report, for example. But they do not

include guidance on who should be engaged in that assessment. Identifying criteria for selecting reviewers to ensure that the appropriate units within SECO and others are able to provide feedback on draft reports would help to strengthen the credibility of the final product.

Another way to strengthen credibility is to establish an internal set of peer reviewers who can provide input into the independent evaluation throughout the process, from the development of TORs through the review of the final report. Other evaluation units reviewed do this as a matter of course for most of their evaluations. For example, in Finland "critical friends" help advise EVA staff throughout the evaluation process; these may be evaluation or subject matter experts, depending on the need in any given case. Having feedback from operational staff engaged in the kind of work the evaluation is expected to influence through a reference group also is a good way to improve the credibility of the final product and promote its utility.

There were some concerns expressed that the pool of external evaluators available to SECO appeared to be relatively small, so that the same ones seemed to be used several times, although

²⁹ SECO/WE, Evaluation Guidelines, undated.

<https://www.seco-cooperation.admin.ch/dam/secocoop/de/dokumente/resultate/evaluation/eval-guidelines.pdf.download.pdf/Evaluation%20Guidelines.pdf>

this may be a misperception by the interviewees. The EU reports that in recent years only one firm was involved in two centralized independent evaluations, however with different team members, and that the past four major independent evaluations were all contracted to a variety of international evaluation teams, with only one incorporating a Swiss organization as a sub-contractor. This illustrates the diversification of evaluation teams sought by SECO, and/or the limited availability of local independent evaluation expertise on SECO's key topics.

However, there is a great deal of work involved for the EU in soliciting and evaluating proposals to conduct evaluations by external evaluators and firms and completing the necessary arrangements to carry out this work. As noted in the overview report, Finland currently has a contract with an external firm to conduct these functions on its behalf. Early results appear to be positive from both the evaluation unit and operational side. It would be worthwhile for the SECO EU to consult with the Finnish MFA team to explore their experience with negotiating and implementing this contract, as well as the continuing response to the work carried out under it.

Recommendation: To further strengthen the credibility of the evaluation function products, the EU should:

- incorporate into the *Guidelines* criteria for selecting internal and external reviewers; and
- explore the possibility of adopting an approach like the Finnish model of using an external firm under a framework contract for selecting evaluators and conducting evaluations.

Utility of evaluations

Based on the interviews the Team conducted, the evaluations overseen by the EU were found to be useful by SECO Management, in particular. As is often the case, however, some staff reported that the reports were pitched at too high a level to be of much operational use to them. This is a recurrent theme with centralized evaluation, not only at the four organizations reviewed here but more generally. There is no obvious solution to this issue, except to recognize that centralized evaluations inevitably deal with broader policy and process issues, while decentralized evaluations tend to be more operational in scope. In any case, it is a common observation across the development evaluation field.

One factor for promoting utility is the alignment of evaluation work with corporate strategic orientation. As noted above, the new Dispatch will not be based on the four-pillar model, making it more complicated to ensure this alignment, but it currently is an important contributor to utilization and representativeness of evaluations.

A second factor is the high degree of buy-in to evaluation at all levels of the organization. As noted above, this is an important part of the enabling environment for SECO evaluations. Coupled with this is a broad awareness around the organization of not only the evaluation unit, but of the evaluations undertaken. In part, this is a consequence of the relatively small size of the organization, but it also reflects the engagement of the EU with the operational units as described above. These two factors are mutually reinforcing mechanisms for promoting the use and utility of evaluation, for example in the design of new projects and programs.

As noted in other sections, the most obvious form of utility is the way Management responds to evaluation recommendations and implements agreed-on actions to address them. In this regard, the SECO EU has developed an efficient tool for keeping track of these actions. (see Table 4.) This kind of record is an appropriate way to account for what actually was done to respond to the evaluations' findings and recommendations. The regular discussion of the implementation of recommendations in the Evaluation Committee helps to keep the momentum going. If such information were stored in a searchable database it would be especially useful.

Table 4: SECO Management Response follow-up record

Follow-up Management Response to the Independent Evaluation of SECO's Evaluation on [topic]						
1 Summary of the Status of Implementation						
[Text summary]						
2 Management response on recommendations made by the evaluation team						
Recommendation	Management response	Position of the External Committee	Responsibility	Date	Status	Comment
Recommendation 1						
Recommendation 2						

Finally, SECO's evaluation function shares the common problem of how to get broad dissemination of its reports and other products to the widest relevant audience. Currently, it works through the Committee to get its key messages to Members of Parliament, and to Senior Management in SECO. Its participation on various management committees, such as the Operations Committee, and other fora offer opportunities for further dissemination throughout the management ranks.

If the EU finds it beneficial to strengthen dissemination to others, particularly to operational staff or external audiences it would be useful to develop a communication and dissemination plan. Such a plan would identify what audiences the evaluation function seeks to reach, what their information needs are, and what products specifically geared to how they obtain and use information could be developed. That would imply working with an expert in communication to develop such a plan, and securing the funding needed to implement it.

Recommendations: The EU should

- consider adopting the recommendations on utility in Chapter 1; and
- explore further ways for disseminating evaluation results.

ANNEX I: TERMS OF REFERENCE

Introduction

The OECD-DAC Network on Development Evaluation (EvalNet) encourages professional Peer Reviews of the evaluation function of its member countries and organisations. Each Peer Review shall identify good practice and opportunities to further strengthen the evaluation function of its members under review, with a view to contributing ultimately to improved performance in international development cooperation and humanitarian assistance³⁰. Finland, Ireland and Switzerland recently joined together to conduct a Peer Review of their respective Evaluation Functions (EF). Supported by an independent consultant analysing relevant documents, the two other countries reflect the findings of a limited number of interviews with their own experience and international standards.

The Peer Review will be carried out according to the overall provisions contained in the UNEG Framework for Professional Peer Reviews of the Evaluation Function of UN organisations³¹. It will be the first Peer Review of the respective EFs outside the regular DAC country Peer Review.

This document provides the core Terms and Conditions (ToRs) of the Peer Review, which will be broadly common to all three members. This document outlines the generic purpose; the scope; the general approach; the methodology; and the composition of the Peer Review Team.

Purpose of the Peer Review

The objective of this Peer Review is to provide an impartial assessment on the Evaluation Function, its strategic fit in their member Ministry, Agency or Division, good practices and areas for improvement to better support achieving the strategic objectives of the Development Evaluations. It will be conducted according to the UNEG Norms and Standards, thus respecting the necessary degree of independence of the EFs.

The Peer Review will help the EFs to ensure that they are fully fit for purpose and positioned to make the best contribution to the work of their respective Ministries³. The Peer Review will pay particular attention on the need to build a culture of evaluation in their member organisations to promote accountability, strategic decision making and learning through integration of evaluations in project and programme cycle management.

³⁰ See UNEG, 2011, *UNEG Framework for Professional Peer Reviews of the Evaluation Function of UN organisations*, UNEG/REF(2011)

³¹ UNEG, 2011. UNEG Framework for Professional Peer Reviews of the Evaluation Function of UN organisations [Online]. Available: http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=945 ³ Ministry also comprises Agency or Division, depending on the country context.

Subject, Scope and Limitations

This Peer Review will centre its recommendations in the following areas:

- The strategic role of the EF in providing credible and useful information for senior management decision-making, in particular in the planning and preparation of its work programme and reporting;
- The EF's role and its interaction with strategic planning and results measurement functions;
- Adequacy of human and financial resources allocated to the EF;
- Deployment of the system for projects and programme self-evaluations and completion reports within the organisation including the validation system managed by the respective EFs if appropriate;
- Enforcement of the implementation of the evaluation recommendations to facilitate new strategies and interventions to improve performance and results;
- Intensification of the use of robust and credible evaluation methods, including in the field of impact evaluation.

The primary intended audience for the results of the Peer Review will be senior management as well as external stakeholders such as Parliament, Management Boards, and Advisory Committees etc.

The scope of the Peer Review is limited to the evaluation activities carried out from 2010 until 2019, considering the present set-up of the EF. It will include an assessment of the:

- Normative framework for evaluation: The impact of existing policies and procedures of the EF, including the extent to which they conform to international norms and standards;
- Management of the Evaluation Unit: The effectiveness of management arrangement, working procedures and the internal organisation of the EF in fulfilling the Evaluation Policy commitments and the achievement of strategic evaluation objectives;
- Evaluation planning: The methods and criteria used for strategic planning of evaluation activities and the extent they reflect the strategic priorities and directions of the respective organisations;
- Evaluation quality: This includes the quality and credibility of the evaluations undertaken under the auspices of the EF, taking into account the planning process, the conduct of the evaluations, the quality of the evaluation reports, the independence of evaluation teams, and ways in which the credibility and utility of reports is enhanced;
- Evaluation follow up and use: The Management Responses to evaluation reports and Action Plans for the implementation of the recommendations, including the follow up of the implementation of the recommendation. The use of evaluation evidence in the development of new policies, programmes and projects and in decision-making;
- External relations of the EF: with external stakeholders including national partners, donors, other partners, and the global development community.

The Peer Review will not be a fully-fledged evaluation that comprehensively evaluates practices, processes, and outcomes in depth, but rather will be descriptive in nature and be based on the Peer's own experience. The Peer Review Panel will report on the limitations of its work.

Approach

The Peer Review will use the approach developed in the UNEG Peer Review group, including the 'Stages of Peer Reviews' chart and according to the Normative Framework (Annex II). The panel will use a light version of the UNEG Peer Review, entailing a shorter period for the peer review, the use of existing Communication instruments such as a list of meta-evaluation criteria developed for prior peer reviews, and the use of the evaluation products of the organisations.

Core Assessment Criteria

The Peer Review will use the following criteria based on UNEG norms and standards:

- Enabling environment for evaluations: This could include assessing the prominence evaluations have in terms of learning, and how they influence decisions, policies and strategies.
- Independence of evaluations and evaluation systems: The evaluation process should be impartial and independent in its function from the process concerned with the policy-making and programme management. The planning and selection of evaluation subjects should be an impartial and independent process.
- Credibility of evaluations: Credibility requires evaluations to report successes and failures, as well as sufficient participation of stakeholders. This depends on the expertise and independence of the evaluators, as well as the degree of transparency of the evaluation process.
- Utility of evaluations: To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a timely, clear and concise way, fully reflecting the different interests and needs of parties involved.

Panel Composition and Responsibilities

The Peer Review Panel will consist of a senior external consultant and senior representatives from the EF of each organisation who will act as Co-Chairs. Panel members will not participate in the review of their respective EF.

The external senior consultant is required to work with and contribute to the analysis undertaken the Peer Review Panel. Using the Core Assessment Criteria, the consultant will conduct a systematic review of the arrangements for each EF in order to inform and advise the members of the Peer Review Panel on the key issues to be considered in the course of the Peer Review, and to assist the Panel in formulating and reporting its conclusions.

The panel is fully responsible for the quality and contents of the final report.

Terms of Reference for the senior consultant

The senior consultant will undertake the tasks set out below. The tasks will be executed in several phases:

Phase 1: Common Inception Phase (August and September 2019)

- Review ToR and receive briefing on assignment from the Co-Chairs, Peer Review Panel;
- Kick off meeting in Helsinki 27th – 28th August 2019;
- Collect relevant documentation concerning the evaluation functions, including relevant corporate policies, strategies and plans, and conduct preliminary analysis;
- Prepare short, common inception report outlining approach and work plan;
- Design, conduct, analyse and report on the survey on the demand and use of evaluations;
- Undertake systematic analysis of the documentation, drawing out the implications for the key topics to be considered by the Peer Review;
- Prior to country visit, prepare a preliminary assessment and an inception note (per country) for the use of members of the Peer Review Panel, and for consideration by staff of the Evaluation Functions.

Phase 2: Visit of the Peer Review Panel to each HQ (October 2019 to February 2020)

- Support the members of the Peer Review Panel during their visit to HQs;
- Participate in Panel interactions with staff.

Phase 3: Individual country reports

- Following each visit, draft the country specific issues report based on guidance and specific outline received from the panel.

Phase 4: Key issues report (April 2020)

- The consultant will draft a preliminary overall key issues report across all three countries based
- on a specific annotated outline received by the Panel;
- Building on the agreed overall key issues report, the consultant will prepare and submit the final Key issue report to the Panel (see deliverables).

Qualifications

- Advanced university degree in social sciences or related field;
- A minimum of 15 years professional experience in evaluation and results based management;
- In-depth knowledge of peer review process and assessments of evaluation function;
- Experience in conducting and managing evaluations in international settings and within bi-lateral organisations;
- Background in international relations and knowledge of current development issues is an asset.

Reporting

The senior consultant will report to the Co-Chairs of the panel on substantive peer review issues. The administrative issues, including contracting and payments, access to documents will be managed by the Swiss Agency for Development and Cooperation (SDC).

Key outputs include:

1. Common Inception Report (10pp);
2. Survey development and implementation/analysis;
3. Quality assessment of select evaluation reports;
4. Three inception notes – preliminary assessment prior to the country visits (Phase 2) (<10pp);
5. A preliminary key issues report after each country visit (Phase 3) (<15pp);
6. Draft Key Issue Report (Phase 4);
7. Report that will be published:
 - **Key Issue Report** (common issues of interest, lessons learnt and challenges / limitations of methodology (10pp);
 - **3 country specific annexes** (<15 pages per country).

Duration

The consultancy will start on 15.8.2019 and is expected to be completed no later than April 2020. The estimated level of efforts during this period is the **number** of working days as follows:

- Kick off meeting in Helsinki: 3 days;
- Common inception report (3 days);
- Desk reviews including quality assessment and inception report per country (5-8 days per country);
- Visit to HQ (4-5 days per country);
- Drafting of various report elements (5 – 10 days);

Time frame, target dates

Deadline	Activity
07.06.2019	Expression of interest by email.
11.06.2019	Deadline for submitting questions by email
14.06.2019	Sharing of the questions and answers with all the interested bidders
19.06.2019	Deadline for submitting offer by email
02.07. 2019	Awarding of mandate and notice to unsuccessful bidders
Mid July 2019	Signing of contract
15.08.2019	Start of the Mandate
27 th -28 th 08.2019	Kick off meeting in Helsinki in the week of 27 th -28 th August 2019

ANNEX II: NORMATIVE FRAMEWORK

The Peer Review Panel will use the Normative Framework below, which is based on an interpretation of UNEG norms and standards as relevant to the evaluation functions. X means relevant; (X) means that the answer is obvious but should be discussed.

	Questions	Norms	EVALUATION UNIT
Independence	What criteria determine the approval of evaluation plans?	1.2, 2.3	x
	Does the planning and selection of evaluation subjects result in a work plan that contributes to learning and accountability?	1.1, 4.1, 4.2	
	Are evaluation processes (planning and conduct) independent and impartial?	1.2, 7.1	x
	Who is finally responsible for ensuring the independence of Unit's evaluation function?	2.1	
	How is the Evaluation Unit organised and staffed?	2.3	x
	What is the basis for the Unit's evaluation budget (is it in function of respective evaluation programming or does the funding available drive what will be evaluated)?	2.3	x
	Who decides the Evaluation Unit's evaluation work programme (including the selection of subjects for evaluation)?	2.6	x
	Does the respective Evaluation Policy provide a clear explanation of the concept, institutional framework, roles/responsibilities and use of the evaluation function?	3.1	
	Do the Evaluation Policies conform to international standards?	3.1	(x)
	Where are the Evaluation Units located with respect to the Management and Governing Body?	6.1	
	To whom does the Head of the Evaluation Unit report?	7.1	
	Are evaluations publicly available?	10.2	(x)

Credibility	Does the normative framework provide good practice for evaluation processes, both centralized and decentralized?	3.1	x
	Is a system in place to ensure the professional competence of the evaluation team that is necessary for arriving at credible and accurate evaluation reports?	2.5, 9.19.3, 11.1- 11.5	
	Does the evaluation function provide an advisory role during the planning stage of undertakings to improve their evaluability?	7.1	
	Is a system in place to ensure the quality of evaluations in their design, methodology, conduct of evaluation and reporting?	1.2, 8.1	
	Are evaluations conducted transparently and impartially?	5.1, 10.1	x
	Are the criteria for selection of subjects of evaluation set in a way that ensures an impartial choice?	5.3	x
	Is a system in place to ensure conflict of interest is avoided in the identification and selection of consultants or evaluation services providers?	5.3	x
	How does the Unit ensure/enable the quality of evaluations in the design, methodology, implementation and reporting?	8.1	(x)
	Do evaluation terms of reference/approach papers spell out the evaluation methodology to be used, and if so does it ensure impartiality (e.g. approach to selecting case studies or site visits)?	8.1	
	Are evaluation findings communicated in an impartial way with adequate levels of technical and political credibility?	8.2	(x)
	Are there provisions that ensure evaluation staff has the right technical competencies?	9	
	Is a system in place to ensure that the evaluation process is clear and transparent to stakeholders?	10.1	
	Are evaluation terms of reference shared with stakeholders once finalised?	10.2	
	Is a system in place to ensure transparency in the reporting of evaluation findings and how comments are dealt with?	10.2	x

Utility	Is the purpose that evaluation fulfils for the Ministry clear at senior management and operational levels?	1.1, 1.3	x
	Is the respective evaluation function linked to the Ministry's results-based management system, if so in which way?	1.1, 1.3, 2.6	(x)
	How are evaluation recommendations used at the various management levels?	1.1, 1.3, 2.6	(x)
	Does evaluation feed into management and decision making processes?	1.2, 1.3, 4.1	x
	Is the selected sample of evaluation objects representative enough to enable comparative analysis and drawing lessons across the portfolio.	1.3, 1.5, 2.6, 4.2	
	Are evaluation reports easily accessible, e.g. through a searchable website?	2.7, 13.2	
	Does the Unit have an active policy/practice of disseminating evaluation findings and reports?	2.7, 13.2	
	Is the Evaluation Unit work programme made public?	4.1	x
	Does the evaluation process engage stakeholders in ways that make evaluations useful, while maintaining independence and credibility?	4.2, 5.1, 10.2	(x)
	Are evaluation reports easy to understand, to the point and present evidence, findings, conclusions and recommendations in a complete and balanced way?	8.2	
	Are recommendations concrete and action-oriented?	8.2, 10.2, 13.2	
	Are evaluation findings communicated in a useful, constructive and timely manner?	10.2	x
	Is there a system in place to ensure appropriate follow-up action?	12.1-12.3, 4.1	x
	Is there a management response system that ensures formal, corporate, substantive and timely management responses are given to evaluation recommendations?	12	(x)
	Is there a management response system that ensures that follow-up actions are taken, recorded/tracked, and reported on?	12	
How is evaluation knowledge shared? How does the Unit contribute to knowledge management system?	13.1	(x)	