exPost
ExPost

A Comparative Study of

# Evaluation Policies and Practices in Development Agencies

**Marta Foresti**

with Christina Archer, Tammie O'Neil and Richard Longhurst

odi
Overseas Development
Institute

GROUPE AGENCE
FRANÇAISE DE
DÉVELOPPEMENT

**Disclaimer**

# Foreword

The question of whether 'aid works' is at the heart of the political debate. On the one hand, we have seen commitments to increase aid, designed to achieve the MDGs, and some evidence that the actual flows are beginning to rise. New instruments to deliver aid and support developing countries are also being introduced, for example the airline tax proposal pioneered by France. On the other hand, public expenditure is facing tight constraints in all western countries, not least as the outlook for the global economy becomes somewhat more pessimistic. It is not surprising that development ministers and the public who elected them are facing hard questions about impact and value for money.

Evaluation is clearly a strategic feature of this debate. In particular, expectations are higher than ever before with regard to the knowledge, evidence and expertise that evaluation should deliver in support of policy decision-making and resource allocation.

The fact of the matter is that despite agreed overall frameworks and common standards - such as the DAC criteria for evaluation - evaluation practice in development agencies varies very much, depending on a number of factors: resources invested, independence of the evaluation function, methodologies adopted, to name a few.

This study was commissioned to map and compare evaluation practices across development agencies, primarily with a view to stimulate an internal debate within AFD on its evaluation systems during a time of reform. However, we believe that the report can stimulate a wider debate within the development evaluation community as it complements the existing literature on donor policies and the individual case studies of good practice.

Nine agencies were reviewed as part of this study. The results show that there is indeed variation among the strategies and practices adopted by different evaluation units, in terms of both internal arrangements and different roles and responsibilities fulfilled. This is partly explained by the fact that units are increasingly expected to fulfil a variety of different roles and to engage in a wide range of activities. However, evaluation units share a number of common features and challenges, and are on a similar journey (although at different stages), from a relatively straightforward model of project evaluation aimed at internal management and accountability, towards a more complex model of policy, country and partnership-led evaluation, which requires new skills, roles and organisational arrangements.

Summing up, the report describes an apparent disconnect between the rhetoric on the strategic and growing importance of development evaluation, and evaluation practice in many development agencies. This "institutional evaluation gap" calls for greater attention to institutional approaches to evaluation, arrangements and capacity. And perhaps for a more collective effort among the key players in development evaluation.

The study has been a useful and interesting collaboration between the Agence Française de Développement (AFD) and the Overseas Development Institute (ODI): we were able to draw on our respective experience and expertise and we hope that other agencies will find these findings interesting.

Pierre Jacquet
Chief Economist, Agence Française de Développement

Simon Maxwell
Director, Overseas Development Institute

## Acknowledgements

# TABLE OF CONTENTS

**List of Tables**

**List of Boxes**

# Executive Summary

This report aims to understand experiences and challenges in implementing development evaluation reform and to assess a variety of solutions, looking at different development institutions. It hopes to lead to improved knowledge of evaluation systems in different development agencies and to provide some pointers to support AFD and other development agencies in their evaluation reform processes.

An overview of the evaluation units in the institutions under study looks at: accountability; information and knowledge; partnerships; capitalisation of experience; and feedback. The report compares policies and practices across institutions, looking at different choices and their consequences, as well as strengths and weaknesses. The analysis also identifies the main internal and/or external factors involved for each institution. The report reviews management arrangements and structural position of evaluation units; their main features, processes and tools; and practices involved in commissioning, managing and supporting evaluation processes. It does not aim to identify good practice or to set standards for the AFD to follow.

The report is based on full case studies (including primary data collection) of five evaluation units at DFID, Sida, the World Bank, KfW, and the AfDB. Staff for interviews mainly came from the evaluation unit. It also covers four desk-based studies (relying mainly on documentary analysis) at the EU, the IMF, Danida, and Oxfam. In addition, a small number of key informant interviews were carried out with leading development evaluation experts.

## Evaluation policy, strategy and mandate

Evaluation policies, mandates or similar organisational documents refer to the overall purposes and objectives of the evaluation function for the particular agency. Most agencies are either reviewing these or waiting for them to be approved. If there is no evaluation policy or mandate, evaluation units refer to other organisational or procedural documents. Most such documents refer to both accountability and learning as the main objectives of evaluations, although there are some differences regarding to what degree each should be fulfilled. Although accountability and learning should be mutually reinforcing, some degree of confusion persists among staff, related to the (perceived) tension and trade-offs between the two.

Agencies without an agreed policy sometimes experience confusion regarding the evaluation unit's mandate and/or roles and responsibilities.

## Independence, integration and isolation of evaluation

With the exception of the multilateral banks, units have all undergone significant organisational changes in recent years. In most agencies, evaluation units are not subordinate to the management, operational or policy departments of the agency; the head does not report to management. Some report to a governing body; some are headed by staff recruited from outside and who cannot be employed by the agency at the end of the mandate; and some sit within the line management structure of the agency. Sida has set up an external and separate organisation (SADEV) to look at the totality of Swedish aid. Sitting outside management and policy structures allows for greater independence, one of the cornerstones of the evaluation function in aid and development agencies. Yet in some agencies' experience, this presents risks of isolation from operational units.

Behavioural independence is more prominent in some agencies than in others. In a number of agencies some degree of tension exists between the independence of evaluation and its integration with other functions. Meanwhile, evaluation units normally rely on consultants to provide an external and objective assessment in their evaluations; those who rely on regular commissions can be under considerable pressure not to be too critical. Concerns were expressed that independence could jeopardise integration and "voice" within the agency.

In the lending agencies, the unit budget is approved by the Board. However, in most bilateral agencies, the unit "bids" for its annual budget, as do other departments. Increasingly, units are submitting workplans.

### Types of evaluation and other activities

Overall, most agencies are focusing increasingly on policy, strategic and country evaluations. Some are still involved in project evaluation, though to a lesser degree than in the past. This can imply that evaluation units now have fewer opportunities to engage directly with field offices and operational departments.

Most evaluation units carry out on average between five and ten country programme evaluations every year; these aim to review and guide overall approach and strategy. Timing of these is not always consistent with country strategy and planning processes, however. The shift towards strategic, policy and thematic evaluations can present some challenges, particularly when findings are perceived as 'political' or it is not clear who should lead. However, it has generally been accepted that this shift is necessary in order to influence policy development and organisational choices. For this to happen, decisions on which policies or themes to evaluate should be taken in consultation with relevant staff in operational and policy departments.

Joint evaluations have also become more frequent. In line with the principle of harmonisation enshrined in the Paris Declaration on Aid Effectiveness, the increasing emphasis on joint evaluation is seen as inevitable and, in principle, a much needed effort on the part of the donors. Actual experiences have been mixed.

The main focus for all agencies is on "independent evaluations"; internal reviews, self-evaluations, monitoring systems and, increasingly, project evaluation and completion reports are increasingly the responsibility of management, operational departments and country programmes. Quality of the latter is often a matter for evaluation units to deal with, sometimes through quality control systems and ratings. Most agencies rely on the OECD/DAC Evaluation Principles as a foundation for quality rating systems. Overall, there is more focus now on rating systems, management follow-up, quality control and knowledge management, to ensure that evaluations contribute to a "learning culture" within the organisation and to some concrete proposals to influence practices and behaviours. There is, however, more work to be done here.

In some agencies, the unit is responsible for producing some form of global report or synthesis. In addition, many agencies have mechanisms and processes to monitor management follow-up and response; these are not always formal or structured.

### Evaluation methodologies

Agencies tend not to adopt a uniform methodology, although almost all use the DAC criteria as a common basis. Efforts are being made to ensure more unified and consistent approaches, to make evaluations more comparable and transparent. Usually, individual consultants propose a methodology which is then discussed and agreed in consultation with the units. Quality can be an issue here too. Impact evaluations are an area of increasing interest. Some agencies are already carrying these out, or at least highlight impact as a key feature.

### Roles, responsibilities and capacity of staff

Specific roles and responsibilities of staff vary considerably. They can include: staff involvement in evaluations; consultations with other departments on specific evaluations; tendering processes; and evaluation management, training and capacity building. Seldom independent evaluations are managed and carried out by unit staff. Increasingly, some form of direct staff involvement in evaluations is considered an opportunity to fos-

ter individual and institutional learning, although most agencies rely on external consultants to carry out all evaluations.

Evaluation units deal with the tendering process, including development of ToRs and recruitment and management of consultants. In most cases, ToRs are adapted to specific requirements, although there can be a lack of flexibility on this written into tendering procedures. Level of consultation with operational and policy departments varies across agencies although consultations are more intense during implementation, reporting and dissemination. Difficulties may arise in this process because units do not have a high level of access to country programmes.

Several units look after the evaluation capacity building, both within and outside the agency, of in-country researchers/consultants, partner governments or internal staff.

Internal staff capacity is considered critical: the number of staff in evaluation units is often inadequate, leaving staff members with little time to engage in the content of studies while overburdened with administration. In addition, staff do not always have previous evaluation experience. Attracting high-calibre staff and limiting staff turnover is a challenge, although many agencies have tried to tackle this by raising staff profile and credibility. There is also recognition of the need to bring in new skills, such as in learning and knowledge management.

### Communication, dissemination and use

This is an area of growing importance for all agencies, in part motivated by a general awareness of current limitations. Only a few agencies have an explicit disclosure or communication policy/procedures, although almost all agencies publish evaluation reports. In some cases, management response is also published. Products range from reports, to synthesis papers, policy briefs, newsletters and e-bulletins, as well as annual reports. Other initiatives include dissemination seminars, both internal and external, special events, conferences, or press releases/conferences.

It is important to recognise that the pressure to communicate effectively with very different audiences can result in dilution of the more challenging findings or over-simplification of the more technical or rigorous aspects of evaluation. This could undermine the impartiality or objectivity of the evaluation process.

Internal feedback and learning often remain very weak. Some efforts are being made to improve management response and follow-up, as well as information management systems and databases. However, this does not reach the wider internal audience (made up of those who could still benefit from sharing lessons learnt). There are some systematic efforts to improve this situation, but these remain few.

### Concluding remarks

The picture that emerges here is a rather composite one. Evaluation units share a number of common features and challenges, and are on a similar journey (although at different stages), from a relatively straightforward model of project evaluation aimed at internal management and accountability, towards a more complex model of policy, country and partnership-led evaluation, which requires new skills, roles and organisational arrangements.

However, there is considerable diversification, in terms of both internal arrangements and different roles and responsibilities fulfilled. This is partly explained by the fact that units are increasingly expected to fulfil a variety of different roles and to engage in a wide range of activities. Differences lie in: roles and responsibilities beyond traditional evaluation; reference audiences; types of evaluation; and tools and products.

Against this background, the evaluation function within aid and development agencies emerges as a function "in search of identity". In some cases, the search is quite advanced; in others, this is a work in progress. Such searches are exemplified in the frequent changes agencies have undergone in recent years.

Overall, there is an apparent disconnect between the rhetoric on the strategic and growing importance of development evaluation, and evaluation practice in development agencies. More investment is needed, both financial and (more importantly) in terms of human resources and capacity. This disconnect could be further investigated and assessed through an

evaluation initiative, aimed at assessing the "goodness of fit" between the current arrangements and the objectives of evaluation units.

This "institutional evaluation gap" calls for greater attention to institutional approaches to evaluation, arrangements and capacity: Investing in the quality of evaluation should be a priority for most agencies. In addition, greater attention needs to be paid to the nature, origin and potential of the demand side of the evaluation equation: in recent years, there has been an increased demand for transparency and accountability of aid and development interventions. This has the potential to inject some new energy into the practice and use of development evaluation which, in time, should be reflected in the performance of evaluation units in all development agencies. ■

## 1. Introduction and Background to Study

The AFD evaluation system[1] is currently in a transition period, halfway through reform. At this particular stage, the AFD needs to consider its challenges and options by looking at the experience of other donors or development institutions. The purpose here is not to benchmark against a golden standard, but to better understand shared challenges and difficulties and to assess a variety of solutions. The AFD thus commissioned ODI to carry out a descriptive study on evaluation policies and practices in different development institutions. This study aims to provide a better understanding of the different evaluation systems, to deepen the dialogue between AFD and other donors, to enrich AFD evaluation reform and to offer food for thought to the evaluators' community.

**Box 1**

**Evaluation in AFD: a reform process**

At the beginning of 2006, the AFD decided to revitalise its evaluation function and, at the same time, to reorganise its Evaluation Unit. Up to then, the evaluation function had been based mainly on in-house project evaluation, driven and implemented by the Evaluation Unit, directly placed under the authority of the Head of Strategy. About 15% of projects financed by the AFD were subjected to final evaluation. Project evaluations were presented and discussed in the Evaluation Committee, chaired by the General Management.

The general assessment was that the AFD did not make the most of its evaluation system. The Evaluation Unit was isolated from the rest of the institution. Feedback was weak. The Evaluation Committee was difficult to mobilise.

In 2006, the Evaluation Unit was integrated as a division into the Research Department, part of Strategy. Since then the revitalisation of evaluation has been based on four principles:

• sharing evaluation;

• synergy between evaluation and research;

• focus on capitalisation and institutional learning;

• mixing external and internal analysis.

▶

---

[1] According to the DAC, "the term 'evaluation systems' refers to the procedural, institutional and policy arrangements shaping the evaluation function and its relationship to its internal and external environment—the aid agency, Parliament, other members' evaluation departments, etc. This includes the evaluation function's independence, the resources it relies upon and, not least, cultural attitudes to evaluation. Evaluation systems influence the demand for evaluation as well as its use, including particularly the dissemination, feedback and integration of evaluation findings by operations and policy departments. They not only affect the quality of evaluation but are also of key importance to the impact of evaluation products and results." ("Evaluation Systems in DAC Members Agencies", a study based on DAC Peer Reviews, presented at the Second Meeting of the DAC Network on Development Evaluation, Paris, 9 October 2004, p.4).

The AFD has also launched a number of new evaluation tools, including:

• **decentralised evaluations:** evaluations will be commissioned by geographical departments and local agencies, shared with local partners and entrusted to external experts, giving a priority to local experts;

• **rigorous impact evaluation:** the AFD engages in experimental exercises in scientific impact evaluation, in partnership with specialised academic teams and interested local partners;

• **strategic evaluation:** the Evaluation Unit continues to commission and pilot strategic evaluation, under initiatives of management and of supervising ministries;

• **thematic capitalisation:** the Evaluation Unit is developing thematic capitalisation, based on comparative analysis of groups of development operations, completed or ongoing, financed by the AFD or other partners.

This study looks at issues raised during the reform of the AFD's evaluation system in the light of the evaluation policies and practices of other donors.

The expected outcome for the AFD is a better factual understanding of the evaluation systems of a certain number of comparable institutions. From a more analytical standpoint, it is also expected that the AFD will understand how evaluations are organised in the institutions studied, and how they contribute to a certain number of crosscutting functions, in particular the following:

• accountability
• information and knowledge
• partnerships
• capitalisation of experience
• feedback.

The main objective of the study is to compare evaluation policies and practices across a range of agencies by highlighting the different choices made, their consequences in terms of "service rendered", and how the institutions feel about the strengths and weaknesses of their respective systems. This comparative analysis will also attempt to identify the main internal factors (pertaining to the missions of the institution) or external factors (national policies) accounting for the choices made and any identified differences. The specific focus of the analysis includes the following factors:

• key features of the evaluation function within the organisation, including management arrangements and placement within the organisational structure;
• main aspects of evaluation systems, processes and tools;
• practices involved in commissioning, managing and supporting evaluation processes.

The study is based on a combination of "full case studies" (including primary data collection) and desk-based case studies (relying mainly on documentary analysis) of the evaluation units (EUs) of the following development agencies:

• **Full case studies:** the Evaluation Department (EvD) at DFID; the Department for Evaluation and Internal Audit (UTV) at Sida; the Independent Evaluation Group (IEG) at the World Bank; the Independent Evaluation Department (FZE) at KfW; and the Operations Evaluation Department (OPEV) at the AfDB.

• **Desk-based case studies:** the Joint Evaluation Unit (JEU) at the EU; the Independent Evaluation Office (IEO) at the IMF; the Evaluation Department (ED) at Danida; and the Programme, Learning and Accountability Team (PLA) at Oxfam.

In addition, a small number of key informant interviews were carried out with some leading development evaluation experts. A full list of respondents is attached in Annex 2.

The full case studies involved in-depth semi-structured interviews[2] with staff from relevant EUs and, to a limited extent, with staff from outside the EUs. It is important to recognise that, in accordance with the Terms of Reference, the analysis presented in this report is based mainly on the opinions, perceptions and feedback of EU internal staff, while providing an overview of the views and opinions held by staff from other departments or by external actors. In this respect, the findings proposed by this study are limited in scope and could be reinforced by further research on wider perceptions of the role, performance and results of EUs in different agencies.

In relation to the comparative nature of this study, it is also important to clarify that the objective here was not to produce an overall assessment of the different evaluation systems put in place by the different agencies, but rather to provide a comparative analysis of some of their features, including management arrangements, main processes, tools and methodology adopted, and practices involved in commissioning, managing and supporting evaluation processes. This study is primarily descriptive in nature and does not aim to identify good practice or to set standards for the AFD to follow.

Section 2 of this report provides an overall profile of the EUs considered by the study. Section 3 provides an overall analysis of the key findings. In Section 4, we draw some conclusions based on the study findings and identify some trends and challenges for the future. ∎

---

[2] An overview of the case study methodology is in Annex 3.

## 2. Summary Profile of Evaluation Units

In this section, we present some of the key features, policies and activities of the EUs considered by the study. Table 1 below summarises the basic features of each EU, including position in the organisation, staff, annual budget, whether it has an agreed evaluation policy or mandate and a common methodology for conducting or commissioning evaluations.

Most evaluation units sit outside the management, policy or operational structures of their agencies, with the exception of DFID/EvD and Oxfam/PLA. When outside management structures, the heads of EUs report directly to the Board of Directors or the equivalent governing body (WB/IEG, IMF/IEO, AfDB/OPEV, KfW/FZE, Sida/UTV), or to the responsible Minister, Secretary of State or Commissioner (EU/JEU, Danida/ED).

The **team size** of EUs varies according to several factors, including: the number of evaluations commissioned or carried out each year (see Section 3.3); whether evaluations are carried out in-house or commissioned to external consultants; and the kind of activities managed by the EUs beyond external or independent evaluation (e.g. information management systems, self-evaluation and monitoring, annual reporting, project evaluations, etc.)

The **role, objectives and mandate** of the EUs is not always determined by an agreed official policy statement or mandate. This has implications for overall strategy and direction: it affects not only the position and role of the EU within the agency, but also the perceptions and attitudes of its staff (see Section 3.1).

Finally, very few agencies have **a standard common methodology** for evaluations. This is explained partly by the diversity of evaluations being commissioned, and by differences in terms of specific activities carried out by EUs. When an EU has a standard methodology, this tends to be related to specific activities, such as project evaluations (KfW) or annual reporting (Oxfam). Although IEG at the World Bank and the IEO at the IMF do not have a standard methodology for all evaluations, a detailed methodology is specified in the issue paper developed prior to the evaluation, agreed by different stakeholders (in the case of IEG by the Board via CODE). The JEU at the EU is the only agency providing common methodological guidance for geographical and policy evaluation. In most other cases, the methodology for evaluations is outlined in the ToR and further specified by the external consultants hired.

## Table 1. Basic features of EUs considered by the study

| | Agency and EU | Position of EU | Staff | Annual budget[3] (US$m) | Evaluation policy or mandate | Common Methodology and Guidelines |
|---|---|---|---|---|---|---|
| **Bilateral agencies** | Sida – UTV | Secretariat reporting to Board of Directors (previously in Policy) | 13 ft | 2.76 | Yes, 1999 | No, common methodology but DAC criteria; Sida manual for evaluation |
| | DFID – EvD | In Finance and Corporate Performance Division, reporting directly to its Head who reports to DG for Corporate Performance (on Management Board) | 20 ft | 7.2 | Under development, not finalised, 2002 | No common methodology |
| | Danida – ED | Reports to Secretary of State – Head of South Group. Outside operational/management structure | 8 ft | 3.5 | Yes, revised in 2006 | No common methodology but DAC criteria; guidance for evaluation and project completion reports |
| | IMF – IEO | Reports to Executive Board – but works at 'arms length' | 13 ft | 4 | No, but has mission/value statement | Methodology defined in issue paper |
| **Banks and lending agencies** | WB – IEG | Reports to Board of Directors, through its Committee on Aid Effectiveness (CODE) | 90 (WB) 120 (incl. MIGA and IFC) | 22 | Yes, DGE mandate/ToR 2005; updated operational M&E policy 2006 (pending adoption by board) | Methodology defined in issue paper; methodological guidance on impact evaluation |
| | AfDB – OPEV | Reports to Board of Directors, through its Committee on Aid Effectiveness (CODE) | 21 ft | 4.7 | Under development, Evaluation Guidelines 1996, Directive of OPEV 2002 | No common methodology; guidance on project completion reports, RBM, private sector projects |
| | KfW – FZE | Reports to Board of Managing Directors, outside management structure | 7 ft, 1 pt | ? | No, refers to 1970s guidelines for BMZ, GTZ, KfW | Yes, common methodology for all end of project evaluations |
| **Multilateral** | EU – JEU | Reports to three Commissioners, administratively attached to EuropeAid | 15 ft | 5.18 | Yes, multi-annual strategy 2002-2006 | Yes, common methodology and guidance for policy and geographical evaluations |
| **NGOs** | Oxfam – PLA | Reports to Director of International Division | 3-5 ft | ? | No, but guidance in the Mandatory Procedures | Yes, but only for programme annual impact reports |

[3] Including staff costs.

# 3. Main Findings

In this section, we analyse the main findings of the comparative study, drawing on both the desk-based and the full case studies.

## 3.1 Evaluation policy, strategy and mandate

As highlighted in Section 2, not all EUs have adopted an official evaluation policy or a specific mandate. Notably, these are currently being reviewed in some agencies, or are pending official approval by the Board or management (e.g. DFID/EvD, AfDB/OPEV, WB/IEG). In the case of Sida/UTV, there is an evaluation policy in place, but this was felt to be out of date by many respondents. In the absence of an evaluation policy or mandate, EUs refer to other organisational or procedural documents offering guidance on evaluation activities and priorities.

Evaluation policies, mandates or similar organisational documents refer to the overall purposes and objectives of the evaluation function for the particular agency. Most of these refer to both **accountability and learning** (sometimes referred to as knowledge, e.g. Danida/ED) as the main purposes or objectives of evaluations. However, there are some differences among the agencies under study. Although Danida's Evaluation Policy states that "generation of knowledge" and "generation of accountability information" are the two main objectives of evaluations, it also recognises that "individual evaluations will not necessarily fulfil both". The KfW/FZE policy has a greater emphasis on accountability to the German Federal Government and the general public, focusing on impact and cost effectiveness information. However, the focus

of KfW/FZE evaluations partly shifts from ex-post project evaluations to thematic evaluations; lessons learnt are seen as a growing priority. Finally, according to the AfDB 1996 Evaluation Guidelines, the main objective of evaluations is to ensure that resources are efficiently used, projects effectively implemented and projects objectives achieved.

Despite the generally agreed consensus that accountability and learning are both key objectives of evaluation, reinforcing each other rather than conflicting with each other, some degree of **confusion** still persists among staff inside and outside EUs. This confusion is sometimes related to the shift (or perceived shift) occurring over time, away from an emphasis on accountability towards an increasing role for lesson learning. In the case of AfDB/OPEV, it was found that this was still a relatively recent trend. In most other agencies, the "two-tier" approach of accountability and learning has been consolidated over time, and is now enshrined in evaluation policy as well as in practice. However, this does not always prevent perceptions of tension between the two and the consequences that each entails, among staff both inside and outside the EUs. Different views can be held in relation to what should be the main purpose of evaluation in a particular agency, with people taking sides in favour of either accountability or learning, often depending on their location with the agency.

Perhaps not surprisingly, it appears that, in agencies with no **agreed or established evaluation policy**, there is more confusion or there are tensions around the role and mandate of the EU. In contrast, one of the main benefits of an agreed and established policy or mandate is that it can clarify the respective roles and responsibilities of EUs and of the main related staff. Examples of EUs with an agreed policy and a clear

mandate are Danida/ED, Sida/UTV and the WB/IEG. The latter's roles and functions are defined by the mandate of its

Director General, Evaluation (DGE), as well as by its Operational Policy endorsed by the Board (see Box 2).

---

**Box 2**

**WB/IEG: evaluation mandate and policy**

The World Bank's Executive Directors have approved the Director General, Evaluation Mandate (revised following the transition from OED to IEG), together with the revised ToR for the IEG.

**Mandate of the Director General, Evaluation:** The DGE is directly responsible to the Executive Boards of Directors of IBRD/IDA, IFC and MIGA for:

• assessing whether the World Bank Group's programmes and activities are producing the expected results;

• incorporating evaluation assessments and findings into recommendations designed to help improve the development effectiveness of the World Bank Group's programmes and activities, and their responsiveness to member countries' needs and concerns;

• appraising the World Bank Group's operations self-evaluation and development risk management systems and attesting to their adequacy to the Boards;

• reporting periodically to the Boards on actions taken by the World Bank Group in response to evaluation findings, and on the measures being taken to improve the overall operations evaluation system, including dissemination and outreach activities;

• co-operating with the evaluation heads of other international financial institutions and development assistance agencies;

• encouraging and assisting developing member countries to build effective monitoring and evaluation associations, capacities and systems.

**ToR of the IEG (WB):** The IEG performs the following principal functions:

• making periodic assessments of the adequacy and effectiveness of the Bank's operations evaluation;

• developing risk management systems in light of the institutional objectives of the Bank;

• carrying out reviews of the Bank's completion reports and other self-evaluations, and undertaking performance assessments on selected completed projects, impact evaluations, and evaluation studies focusing on ongoing or completed operational programmes, policies, strategies, processes and partnerships;

• assisting member countries and development partners to develop their own operations evaluation capacities;

• disseminating evaluation findings within the Bank and the wider development community.

Finally, the Operational Policy on Monitoring and Evaluation endorsed by the Board distinguishes between monitoring and self-evaluation activities, carried out by Bank operational staff, and independent evaluations, which are carried out by IEG.

*Source: adapted from www.worldbank.org/ieg/dge_mandate.html.*

## 3.2 Independence, integration and isolation of evaluation

**Independence** is one of the cornerstones of the evaluation function in aid and development agencies. However, organisational structures, cultures and practices in the agencies considered by this study vary considerably. A detailed analysis of the types and degrees of independence of EUs was not a specific focus of this study. However, we present below a synthetic analysis of the different aspects relating to independence which emerged during the study. In particular, we analyse the experiences of the nine agencies in relation to three main criteria for independence[4], namely: i) organisational independence; ii) behavioural independence; iii) avoidance of conflicts of interest; and iv) protection from external influence[5].

● **Organisational independence:** in most agencies, the location of the EU follows the so called "separate unit model"[6]. This means that the EUs are not subordinate to the management, operational or policy departments of the agency, and that the head does not report to management. Agencies currently adopting this model include: Danida/ED, Sida/UTV, DFID/EvD, KfW/FZE, AfDB/OPEV, WB/IEG, IMF/IEO and EU/JEU. However, there are significant differences between them. In the case of the multilateral banks and lending agencies, the head of the EU reports to a governing body, namely the Board of Directors, usually through a specific committee. The head of the EU is appointed on a fixed mandate by the Board. For the IMF/IEO and KfW/FZE, the head of the EU is recruited from outside the agency and cannot be employed by the agency at the end of the mandate. In other cases, the head of the unit reports to the relevant Minister/Secretary of State or Commissioner, sometimes through a Board of Directors. Finally, it is worth noting the cases of DFID/EvD and Oxfam,

where the evaluation unit sits within the line management structure of the agency.

Sida/UTV is the only agency to have recently set up an **external and separate** organisation (SADEV) as a mechanism to look at the totality of Swedish aid, including impact of foreign policy and funds going through the Ministry of Foreign Affairs, although funds channelled through Sida/UTV can have independent evaluations commissioned through the UTV. The experience of SADEV is still too recent to provide further insights on its potential benefits or constraints.

It is worth noting that the EUs of bilateral agencies and NGOs have experienced many changes in recent years, particularly in relation to their position and role within the organisational structure. However, as noted above, the general trend is towards the "separate unit model" and an increased independence from management structures.

● **Behavioural independence:** this refers to the capacity and space for the evaluation function to critique the agency's work and to make recommendations for improvements. This appears to be more prominent in certain agencies than in others. A self-evaluation on WB/IEG independence[7] noted that their "reports routinely critique the Bank's work at the project, country, sector and global levels. While recognising the Bank's achievements, these reports also address its shortcomings, draw lessons from its experience and make recommendations to advance its development work beyond current practice." In other cases, it was less clear whether the EUs are effectively able to influence the agency's work; it was reported that sometimes it can be difficult to convince management of the value of evaluation and, as a result, the use of findings can be limited. More generally, in a number of agencies it was reported that **some degree of tension exists between the independence of evaluation and its integration with other**

---

[4] We do not consider avoidance of conflicts of interest in our analysis owing to to insufficient data available from some of the agencies.

[5] "Independence of OED", OED REACH, February 2004: at http://lnweb18.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/

[6] "The separate unit model implies that responsibility for planning evaluations rests with a specialised unit which is managerially separated from the ordinary line positions of the organisation responsible for the intervention or programme

being evaluated. Following this model the responsibility for programme activities and the responsibility for planning evaluation of these activities are separated within the organisation" (Annex 1 in Lundgren, H. et al. (2003), "A Review of Evaluation in Danida", Ministry of Foreign Affairs, report to Danida).

[7] "Independence of OED", OED REACH, February 2004.50B9B24456B788BE85256CE0006EE9F8/$file/Independence_Reach.pdf)

**functions**, sometimes resulting in isolation of the EU from the rest of the agency. Examples of these kinds of tensions emerged in the DFID/EvD and AfDB/OPEV case studies. As one respondent put it, "being involved is as important as being detached"[8].

A somewhat different dimension of behavioural independence is one that relates to the "true" **independence of evaluation consultants** who carry out independent evaluation on behalf of the EUs. With the exception of WB/IEG, EUs normally rely on consultants to provide an external and objective assessment in their evaluations. However, it was pointed out that consultants who rely on regular commissions from a particular agency can be under considerable pressure not to be too critical in their analysis. Some worries were expressed in relation to "the usual consultants" who, in order to get the contracts, are not always incentivised to provide objective and critical views of agency actions.

● **Protection from external influence:** this relates to independence in setting the work programme, in defining the annual budget of the EU, and in hiring arrangements. In the case of the lending agencies, the budget is approved by the Board, independently from management decisions. However, in most bilateral agencies, the EU "bids" for its annual budget, following the same procedures as other departments/units. Increasingly, and as a result of the shift towards policy and strategic evaluation, EUs are submitting a workplan outlining their main evaluation initiatives as planned for the next year or two. The EUs of the banks tend to set this workplan largely independently, whereas most bilaterals engage in extensive consultations with other departments and field offices[9]. In relation to staff hiring, the multilateral lending agencies have the most specific procedures. The IMF/IEO puts a particular emphasis on externally recruited staff (see Box 3 below).

---

**Box 3**

**Independence of the IMF/IEO**

The IEO is a permanent unit which operates independently of IMF management and operational staff, reporting directly to the Executive Board but working at "arms length" from it. This independence is guaranteed through both its structure and operations:

i) The Director is appointed directly by the Executive Board for a period of four years, renewable for a further three years. The Director is an official of the IMF but not a staff member, and cannot be appointed to the IMF staff at the end of his term.

ii) The Director of the IEO is solely responsible for the selection of IEO personnel (including external consultants) on terms and conditions set by the Executive Board, with a view to ensuring that the IEO is staffed by independent and highly qualified personnel. It is intended that, when fully staffed, the majority of IEO personnel will come from outside the IMF. In addition, IEO staff report exclusively to the Director of IEO, not to IMF management.

iii) The budget of the IEO is subject to the approval of the Executive Board, but its preparation is independent of the budgetary process over which IMF management has authority.

iv) The IEO's work programme is established by the Director in light of consultations with interested stakeholders, from both inside and outside the IMF. The work programme is presented to the Executive Board for review but is not subject to the Board's approval.

v) The IEO has sole responsibility for drafting evaluation reports, annual reports, press releases and other public statements. Board, staff and management can comment but have no right to insist on changes.

*Source: http://www.imf.org/external/np/ieo/index.htm.*

---

8  In the case of WB/IEG, the issue of isolation was mentioned in relation to the external environment, given the emphasis on internally produced reports and initiatives, which do not always involve external experts or audiences.

9  In Danida, about 70% of the programme is based on suggestions or requests from other departments.

## 3.3    Types of evaluation and other activities

An overview of main activities carried out by the EUs of the agencies considered by the study is summarised in Table 2 in Annex 1.

Overall, EUs of most agencies are focusing increasingly on policy, strategic and country evaluations. Some agencies, such as the WB/IEG, KfW/FZE and AfDB/OPEV, are still involved in project evaluation, though to a lesser degree than in the past. Project evaluation is now being carried out more and more by operational units and country offices. In the past few years, joint evaluations have also become more frequent.

Some agencies are undergoing **significant changes and transitions** in relation to their activities and priorities. In particular, the EUs of KfW and AfDB, which used to focus mostly on ex-post project evaluations, are now moving towards the more general trend of programme and policy evaluations. At the WB, the IEG is responsible for carrying out desk-based reviews of all completed projects. However, only 25% undergo a field-based assessment – a project performance assessment report (PPAR) – compared with approximately 40% in 1997. The EUs of Sida/UTV and other bilateral agencies no longer conduct individual project evaluation.

In most agencies, the main focus is on **"independent evaluations"** (usually country, policy/thematic or strategic), whereas internal reviews, self-evaluations, monitoring systems and, increasingly, project evaluation and completion reports are normally the responsibility of management, operational departments and country programmes.

The **quality** of these self-evaluation activities is often a concern for EUs, which frequently have a role of quality "control" or "assessing" M&E activities carried out by other departments. In some cases, quality control systems and ratings are well established, forming an integral part of the EU's core activities, for example at the WB/IEG and Sida/UTV.

In others, quality control and ratings are an area of growing importance. For example, DFID/EvD is currently hosting a secondment from the Quality Assurance Group (QAG) of the World Bank; the latter is supporting EvD in developing a quality assurance system. In the case of AfDB/OPEV, concerns were expressed about the weakness of the quality control function.

EUs can have different roles in relation to **quality assurance and rating systems.** WB/IEG applies ratings[10] to its own country assistance evaluations (CAEs), as well as a means of verification of project self-evaluations carried out by other departments. The Sida/UTV Rating System (SiRS) is a key management tool. It consists of electronic rating software, generating reports completed by head office desks each year for all projects worth over SEK3M lasting longer than two years[11]. EU/JEU regular internal monitoring is supplemented by an external monitoring system of contractors who analyse project progress. All monitoring reports are entered into the Common Relex Information System. Finally, at Oxfam, changes have been recently introduced to enable better indicators/milestones for better quality analysis. Within each Regional Management Team (situated in the field) there is a Regional Programme Quality Manager who can assist countries in carrying out reviews and evaluations. In general, agencies rely on the **OECD/DAC Evaluation Principles** as a foundation for quality rating systems.

Independent **country programme evaluations** are a major focus for most EUs, which carry out on average between five and ten country evaluations every year. These tend to be pitched at the strategic level, aiming to review and guide the agency's overall approach and strategy in a particular country. However, difficulties are often reported in relation to

---

[10] The ratings for bank and borrower performance are: highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory and highly unsatisfactory.

[11] SiRS involves an assessment of two major dimensions: the performance against targets and workplans; and the risks that the project/programme will not achieve its objectives.

the timing of country evaluations, which is not always consistent with country strategy and planning processes. In the case of the World Bank, it was acknowledged that CAE findings can become available too late to effectively inform or influence countries' strategy development processes. Experience of country-led evaluations is still very limited.

There is a clear trend towards an increasing focus on **strategic, policy and thematic evaluations**, mostly as a result of the shift towards "evidence-based policymaking" and in response to the recent introduction of more target-oriented policy implementation plans. This can present some challenges, particularly when findings are controversial or perceived as "political". In such cases, findings are not taken on board or followed up by management or operational staff. Despite these challenges, there is some consensus among staff in EUs that this is the right direction to follow, as policy and thematic evaluations have the potential to influence policy development and organisational choices. For this to happen, it is of paramount importance that decisions as to which policies or themes are to be evaluated are taken in consultation with relevant staff in the operational and policy departments, and that mechanisms are in place to ensure involvement of and regular feedback from relevant staff and partners.

The number of **joint donor evaluations** carried out each year is also on the increase in all agencies. Danida/ED in particular is leading the way, with up to 50% of its evaluations being conducted jointly in the past year. Most respondents agreed that, in line with the principle of harmonisation enshrined in the Paris Declaration on Aid Effectiveness, joint evaluations can ease pressure on country programmes and partner governments, give more credibility to evaluation, and act as a lever on other agencies, mostly through the DAC. In this respect, the increasing emphasis on joint evaluation is seen as inevitable and, in principle, as a much needed effort on the part of the donors. However, actual experiences have been mixed. A number of problems with joint evaluations were identified by almost all donors (see Box 4).

---

**Box 4**

**Experience with joint evaluations**

In the past few years, most donor agencies have been involved in joint thematic or country evaluations. Mostly, these are carried out under the auspices of the DAC, although sometimes there can be a bilateral arrangement between two donors or a specific initiative (e.g. between the development banks).

The main **expected benefits** of joint evaluations include:

• donor harmonisation and alignment;

• greater learning and sharing between donors;

• greater legitimacy and impact;

• broader scope;

• reduced transaction costs;

• eased pressure on partner countries and programmes.

However, the actual experience of most donors in conducting joint evaluations has been mixed. Challenges include:

• programmes not always/yet being harmonised, so it is difficult to carry out a joint evaluation;

• difficulties in harmonising approaches and tools, namely: ToR, timetabling, methodology, reporting, content and formats, publication and dissemination arrangements. This can result in delays in the process;

• reducing messages to lowest common denominator to avoid disagreement;

• weak accountability and co-ordination (particularly when more than two donors are involved;

• difficulties in agreeing quality control;

• actual increase in transaction costs owing to extra meetings and other communication exchanges to achieve consensus;

• evaluations being very resource- and labour-intensive, not always realistic.

Because of these difficulties, some donors have experimented with different approaches, including organising joint missions and joint teams but producing separate reports to maintain independence and avoid lengthy processes in agreeing on content (World Bank and IMF).

In some agencies, the EUs are responsible for producing some form of **global report or synthesis**. However, there is considerable variation here. The WB/IEG produces its Annual Review of Development Effectiveness (ARDE), which synthesises recent evaluation findings, as well as its Annual Report on Operations Evaluations (AROE), which reviews the progress, status and prospects for M&E of the development effectiveness of World Bank activities and examines how effectively evaluation findings are used to improve results and enhance accountability. Sida/UTV produces the Annual Report of Sida Evaluation Activities, targeted at Sida departments and peer organisations. Although it mainly provides information on UTV activities, it also gives an update on all departmental evaluation and review activities and progress against the annual Evaluation Plan. In the past year, UTV also produced an internal annual report at the request of the Ministry of Foreign Affairs. This synthesises the year's evaluation study findings and management responses.

Many agencies have mechanisms and processes to monitor **management follow-up and response** to evaluations recommendations. These are not always formal or structured mechanisms. In a number of cases, concerns have been expressed in relation to the EUs' role and capacity (or lack of it) to follow up and monitor management response. In a number of agencies, including DFID/EvD, Sida/UTV and AfDB/OPEV, efforts are being made to improve these systems. In Sida/UTV, for example, the Evaluation Policy formally requires that an action plan be produced six months after the

completion of an evaluation, summarising: i) an overall assessment from Sida's point of view of the relevance, accuracy and usefulness of the evaluation and its findings; ii) a point-by-point response to the recommendations, and whether they are accepted or rejected; and iii) a summary action plan, with a completion date for each action. Sida's Director General will then issue "instructions to the organisation", in the form of an action plan detailing who will follow up what and whether any deadlines are attached to any of the points. Responsibility for follow-up then lies with the targeted department. UTV has no responsibility or involvement in this process; a recent evaluation found that the system at present is not achieving its objective of feeding into learning or decision-making, is not used effectively by managers, does not enhance dialogue and partnership, and is not routinely followed up on. At the World Bank, IEG recommendations from strategic evaluations are endorsed by the Board through CODE. Management response is included in the report. The IEG then monitors management response and progress is reported to CODE each year.

## 3.4    Evaluation methodologies

As explained in Section 2, agencies tend not to adopt a **uniform methodology** to their evaluations, although almost all of them use the DAC criteria as a common basis. Efforts are being made by a number of agencies to ensure more unified and consistent approaches, with the objective of making eva-

luations more comparable and transparent. As a result, a number of agencies are currently involved in developing methodological guidelines/manuals, mostly available on EU websites. When evaluations are carried out by individual consultants, it is expected that they will propose a methodology which is then discussed and agreed in consultation with EUs. For the IEG, where all evaluations are carried out by in-house staff, the methodology is included in the issue paper and approved by the Board through CODE. As KfW/FZE has so far mostly been carrying out ex-post project evaluations, a uniform methodology has been developed under the criteria of effectiveness, relevance/significance and efficiency. Oxfam is considering a proposal to develop generic indicators across key programme areas, to provide upwards accountability.

The overall **quality** of evaluation processes and products remains a challenge for some agencies, with different kinds of measures being taken to address this. In the case of the AfDB/OPEV, the main concerns are quality and the backlog of project completion reports. DFID/EvD has recently introduced a quality assurance process (with support from the World Bank) which entails a standard review process for each major evaluation report, involving both EvD and non-EvD staff. A recent review of Oxfam M&E and learning found that Oxfam did not always have enough evidence to show that it was making a difference, and that a greater focus on quantitative data was needed. As a result, a number of changes have recently been introduced, including greater support to country programmes to set better quality indicators and milestones.

Quality and methodological concerns are not limited to evaluation reports and related activities, but increasingly include other areas of growing importance in most EUs, such as management response systems (see Section 3.3), rating systems, knowledge and information management, and lesson learning. Validity, reliability and consistency of systems for **rating** projects and programmes are areas of specific interest for the multilateral banks. At WB/IEG, work is underway to ensure better consistency between the different rating systems

applied by management, the QAG and IEG. OPEV is currently refining its project rating system, which is perceived as generating excessively high ratings.

**Impact evaluation** is an area of increasing interest and attention in development agency EUs. A number of factors contribute to this, including initiatives at the World Bank on impact studies[12] and impact evaluation[13] and, most recently, at the Centre for Global Development[14]. Some of the agencies considered by this study are already carrying out impact evaluations, or consider assessing impact to be one the key features of evaluation (e.g. ex-post evaluations carried out by KfW/FZE all focus on assessing impact, albeit using a lighter and less scientific approach than the World Bank model). Other agencies are planning to invest more in impact evaluation in the future: Oxfam will undertake impact assessment for large donor-driven programmes and it is currently carrying out a feasibility study for a randomised control evaluation of a large livelihoods programme in West Africa.

However, a number of challenges for impact evaluation were identified, including:
• Impact evaluation should be planned at the design phase of a programme, so that the necessary arrangements for baseline data, control groups and cohorts can be made, and should involve partner countries.
• Sometimes, agencies are not directly responsible for implementing programmes and they support the work of others (e.g. NGOs, national governments, etc.) In these circumstances, assessing and attributing impact can be difficult.
• Impact assessment can be expensive and time consuming. Furthermore, results only become available after programmes are completed; this is not the most useful approach for improving practice.

---

[12] See the DIME initiative: http://go.worldbank.org/HIYKB2QV00.

[13] For IEG work on impact evaluation, see http://www.worldbank.org/ieg/ie/.

[14] "When Will We Ever Learn? Closing the Evaluation Gap": http://www.cgdev.org/section/initiatives/_active/evalgap.

• Not all agencies and partners in developing countries have the required technical capacity and resources to carry out impact evaluations.

• As a result, impact evaluations are not always considered a realistic option for some of the agencies considered by the study.

## 3.5 Roles, responsibilities and capacity of staff of EUs

The specific roles and responsibilities of staff of EUs vary considerably from agency to agency. Roles and responsibilities include: staff involvement in evaluations, consultations with other departments on specific evaluations, tendering processes and evaluation management, training and capacity building.

Approaches to **direct staff involvement in evaluations** are particularly diverse among EUs of development agencies. In the case of WB/IEG, independent evaluations are always managed and carried out by IEG staff, with external consultants involved only on very specific technical tasks or as advisors/technical experts. The evaluation report is always written by the "task manager", who is internal to WB/IEG. Increasingly, direct staff involvement in evaluations is considered an opportunity to foster individual and institutional learning, as in the case of KfW/FZE. Approximately half of the evaluations are done by external evaluators, the other half is done by experienced internal project managers who are seconded for the purpose of one evaluation to FZE (following very tough independence criteria to internal staff). This kind of secondment proved to be a very efficient way to ensure "Institutional learning" in KfW. Similarly, Oxfam is increasingly seeking to involve its staff and partners in evaluations, although so far this has been applied mostly in humanitarian evaluations, because donors often require external consultants to carry out country-managed development evaluations. The IMF/IEO seeks to

involve a mix of external consultants and its own staff in all evaluations and Sida is increasingly relying on SADEV staff. Finally, the EU/JEU and DFID/EvD rely on external consultants to carry out all evaluations.

The **tendering process** for external evaluations is a primary responsibility of the EUs, including the development of ToRs and the recruitment and management of consultants. In most cases, ToRs are adapted to the specific requirements of an evaluation. Tensions can arise between complying with tendering procedures, which specify that once a contract has been awarded you can not alter the ToR, and leaving enough flexibility to be able to design the study according to emerging needs and priorities.

The agencies considered by the study adopt different approaches and procedures for **consultations and decision making** around individual evaluations. In most cases, EUs are responsible for proposing country programmes[15] and policy evaluations as part of the annual planning process, which is approved by the Board or by the head of the department. Level of consultation with operational and policy departments varies across agencies. In some cases, ToRs are developed in consultation with the relevant programme desk or policy unit (e.g. Sida/UTV). However, in most agencies, the development of the ToR is the responsibility mainly of the EU. In the case of the multilateral banks, the ToR is developed in close consultation with CODE.

Consultations and negotiations between the EUs and operational and policy departments are more intense during the evaluation **implementation, reporting and dissemination phases.** Such consultations are seen as crucial to ensuring ownership and buy-in by all relevant stakeholders within the agencies and to maximising the potential of follow-up on findings once the evaluation is completed. However, difficulties

---

[15] In some cases, these are part of a rolling plan (e.g. WB/IEG CAE evaluations take place every four to five years in accordance with the country planning process); in other cases, selection criteria are agreed between management and EUs (DFID/EvD is currently reviewing its country evaluation selection criteria).

may arise in this process. One critical issue for DFID/EvD is in persuading country programmes to be involved as case studies in policy evaluations. EvD does not have the same level of access or mandate as the Internal Audit Department when it comes to engaging country programmes, which can be reluctant to be involved in thematic evaluations, because they are too busy with other initiatives, or because the topic of the evaluation is too sensitive or because the timing is incompatible with other programme priorities.

Several EUs have direct responsibility for **capacity building** initiatives, both within and outside their agencies. Danida/ED adopts an innovative model in that it aims to contribute to developing the evaluation capacity of researchers and consultants based in partner countries by ensuring that they are involved in all Danida-commissioned evaluations.[16] For WB/IEG, the focus is on building Southern governments' M&E capacity by providing training and other professional development initiatives. In addition, IEG has been supporting university-based initiatives to develop M&E capacity of evaluators from around the world. Both Sida/UTV and DFID/EvD have promoted internal evaluation training initiatives for staff outside the EUs, although these tend to be farmed out to external consultants. Finally, the EU/JEU has planned an evaluation training module, currently under development.

The **capacity of staff of EUs** is critical in relation to their roles and responsibility. There are two dimensions to the issue of staff capacity. With the exception of the WB/IEG, the number of staff either employed or collaborating with EUs is often inadequate in relation to the tasks and responsibilities expected of an EU. In DFID/EvD, for example, the increasing number of evaluations carried out each year and the limitations in staff capacity imply that, for the most part, staff are involved in the tendering process and in day-to-day management of evaluation contracts, with too little opportunity to engage in the content of the studies and sometimes in the follow-up and dissemination. Over time, such limited engagement combined with relatively

administrative-heavy tasks can cause frustration as well as some concern among EU staff in relation to professional development and career opportunities within their agency.

The **evaluation skills** of the staff of EUs are another constraint. Staff skills sets are very wide-ranging and, overall, cover three main domains: research and analytical skills, sectoral skills, and management/programming skills. Increasingly, EUs attract highly specialised professionals with advanced university degrees. However, staff members do not always have previous evaluation experience or background. In some cases, this is because staff are not expected to be involved directly in evaluation studies; in others, this owes mainly to the strong technical profile of the agency, which may require sectoral expertise over generic skills (e.g. economists at AfDB/OPEV and KfW/FZE, though they also must have extensive evaluation skills). In Sida/UTV, the emphasis is on ensuring that staff have enough skills to be able to engage not only in the management but also in the content side of evaluation studies, even though they are not expected to be leading evaluations.

Attracting high-calibre staff in different areas of expertise and limiting **staff turnover** has been a challenge for many EUs. In recent years, efforts have been made to address these challenges and to raise staff profile and credibility across each agency. In the case of WB/IEG, for example, the seniority and level of expertise of IEG evaluators is considered to be a key asset as well as an important factor in determining that evaluations findings and recommendations are taken on board by management. Finally, it is increasingly being recognised that new professional profiles and skills are required, such as learning and knowledge management: the most recent appointment to AfDB/OPEV is a knowledge management officer, as part of a wider effort to raise the profile of lesson learning within AfDB, a recognised deficiency noted by successive reviews of operations.

---

[16] Approximately half of the consultants hired by Danida each year are from partner countries.

## 3.6    Communication, dissemination and use

An overview of the main features of communications activities carried out by the EUs of the agencies considered by the study is summarised in Table 3 in Annex 1.

**Communicating, disseminating and follow-up** of evaluations are an area of growing importance in all EUs. Evolutions in this domain reflect perhaps the most distinctive trend across the agencies considered. Interestingly, this goes beyond the traditional focus on disseminating evaluation findings to include a more comprehensive approach aimed at effectively communicating a variety of evaluation initiatives (reports, events, tools, etc.) with a focus on reaching out to wider audiences both within and outside the agencies.

The emphasis on communication and dissemination is in part motivated by a general awareness of the current **limitations**: in most agencies, it was reported that more needs to be done to ensure more effective learning from evaluation, particularly within agencies, where opportunities to feed back on and discuss evaluation findings are often limited. With regards to external audiences, it was lamented that the findings of evaluations are too rarely disseminated, acknowledged and used, either by specialist audiences or by the general public.

Only a few agencies have explicit **disclosure or communication policy/procedures**, notably the WB/IEG and the IMF/IEO. However, almost all agencies publish all their evaluation reports and post them on their websites, with the exception of Oxfam, where plans are underway to post all global strategic evaluations on the website in the near future, and of KfW/FZE, which only publishes summaries of reports, owing to confidential information on partners often included in the full reports. In some cases, the management response to evaluation findings is also published.

Efforts are being made by most agencies to ensure that published information is **accessible** to a wide audience,

including the public and non-specialists. Evaluation reports usually follow a common format, including an executive summary of the main findings. They are increasingly available in English as well as in a few other key languages. In the case of Danida, for example, all documentation is published in English, with the Danish version made available in the case of particularly strategic or important evaluations. Evaluation reports, annexes and other evaluation products are disseminated more and more via CD-ROM, although a few copies are printed out and distributed to targeted audiences (e.g. to country programmes in DFID).

Beyond the publication and dissemination of reports, EUs are engaged in a variety of different initiatives aimed at **reaching out to wider audiences**, both internally and externally. Although not systematically, many agencies organise dissemination seminars to report on key findings from evaluation activities. These are both internal initiatives, to contribute to feedback and learning within the agency, and external initiatives, typically involving the consultants directly responsible for the evaluation, academics and policymakers. Sida/UTV, DFID/EvD and other EUs organise regular seminars and are seeking to develop a more systematic work programme for communication and dissemination.

Evaluation **products** are also increasingly diversified, with a view to maximising dissemination and reach, ranging from traditional evaluation reports (with increasing focus on accessible formats and layout), to short synthesis papers and policy briefs (e.g. the Quick Turn-Around Notes of the WB/IEG), newsletter and e-bulletins, as well as annual reports summarising and assessing main evaluation activities and lessons learnt (e.g. the Annual Report of Sida Evaluation Activities and the WB AROE). The WB/IEG is increasingly involved in organising special events and conferences based on its evaluation initiatives, either at the World Bank or in-country, often in partnership with other donors and national government.[17] Finally, some EUs are seeking to engage more with the media to disseminate their findings, mostly through press releases and press conferences.

[17] See, for example, the recent IEG/Norad conference on fragile states held in Ethiopia in July 2007: http://www.worldbank.org/ieg/licus/conference.

## Box 5

**Communicating evaluation at the World Bank**

Over the past few years, the IEG has invested a lot of effort in improving the way in which it communicates within and outside the World Bank. Crucially, the focus has shifted towards a clearer understanding of the use that can be made of evaluation knowledge, with client feedback as the building block to achieving such understanding. The current approach of IEG, therefore, goes beyond disseminating and communicating evaluation findings, with an aim of engaging the client throughout the evaluation process.

IEG strategy for use and engagement has a number of key components:

• increasing awareness of IEG studies among World Bank staff;

• addressing timeliness of reports and findings;

• targeting both internal and external audiences;

• sharper and clearer key messages targeted at specific audiences;

• a clear disclosure policy ensuring maximum transparency;

• the use of a wide set of tools and media.

On of the key feature of the IEG strategy is an innovative and effective use of different media. The various activities planned before and after the launch of a new report are planned as a **communication campaign**, increasingly involving web-based, press and live events. In addition to the launch on the IEG website, efforts are made to ensure that the new report is linked to other internal and external websites. Steps are taken to ensure that Google and other search engines pick up the report and an e-mail campaign ensures that targeted individuals are made aware of the new product. Press releases are shared with key media contacts and audio interviews are organised to ensure maximum reach. Finally, live events such as conferences and seminars are considered central to ensure use and client engagement, beyond their function to disseminate and share experience among specialised audiences. The various phases of this campaign are closely monitored through web-based and other techniques to keep the momentum going beyond the first few days following the release of a report.

As a result of this focus on evaluation use and client engagement, the IEG has been revising the format of some of its products, leading to (among others) a new four-page standard executive summary for all evaluation reports, as well as the sequencing of foreword, summary, management response and recommendations.

The issues of **internal** communication, feedback and learning deserve particular attention. In many agencies it was noted that, despite recent efforts to address these (see Section 3.3), internal mechanisms and opportunities to provide feedback, to exchange views, to learn from lessons and to capitalise on them for the future remain very weak (for example in Sida and AfDB). It is important to distinguish here between targeted dissemination and feedback, aimed at engaging relevant stakeholders directly affected by an evaluation or responsible for their follow-up, from the more general objective of lesson learning across the organisation. In relation to the former, some efforts are being made to improve management response systems and their follow-up, as well as information management systems and databases. However, these are targeted to specific audiences within the agencies and are not suitable for engagement with a wider set of actors with no direct know-

ledge of, interest in or responsibility for a particular evalua-tion (but who could still benefit from sharing lessons learnt). As discussed above, efforts are being made to provide opportunities for internal learning and exchange, although not (yet) systematically.

KfW/FZE is particularly committed to internal dissemination and learning: as discussed in Section 3.5, independent KfW staff being directly involved in carrying out evaluations. Discussion of findings with a wide range of people within KfW is seen as key to institutional learning, so presentations are not limited only to the sector or to regional divisions; FZE staff and evaluators also present findings at various working groups and other sectoral meetings. This is regarded as the most intensi-ve discussion and dissemination channel, as all the operatio-nal departments of the sector will be addressed.

A final point to note in relation to communication and disse-mination regards the potential tensions that this can generate in relation to other aims and objectives of EUs. Although not specifically critical in the agencies considered, it is important to recognise that sometimes the pressure to communicate effec-tively and in a simple and accessible manner to very different audiences (parliaments, the general public, civil society etc.) could result in dilution of the more challenging findings or over-simplification of the more technical or rigorous aspects of eva-luation. This could undermine the impartiality or objectivity of the evaluation process. ∎

## 4. Emerging Trends, Differences and Main Challenges

### 4.1    Overview of main findings

The main aim of this study was to provide some factual analysis of the policy, strategy and organisational arrangements of the nine EUs considered by the desk-based and full case studies. The comparative analysis conducted, although not comprehensive, allows us to identify some emerging trends and some areas of differences, as well as some of the main challenges faced by the EUs of the selected development agencies.

With the exception of the multilateral banks, the EUs considered have all undergone significant **organisational changes** in the past few years, particularly with reference to their position within the agency's structure. There does not appear to be an "ideal" location for the EUs: on the one hand, sitting outside the management and policy structures allows for a greater degree of independence and, potentially, more direct access to strategic and political decision makers (e.g. ministries, Board of Directors, etc.); on the other hand, there is a risk of becoming too isolated and removed from the operational units, country programmes in particular. This raises two potential problems: first, the EUs may struggle to access and engage with relevant policy or operational departments. Second, and perhaps most importantly, the potential isolation of the EUs may undermine the follow-up and use of evaluation findings. It should be noted that, although changes and reviews are underway in many agencies, most bilateral agencies' EUs and the EU/JEU are currently located outside the management structure and report directly to the Board of Directors or the Minister.

Further to these structural changes, it appears that many EUs are engaged in **wider change and reform pro-**

**cesses**, with many policies currently being revised, strategies being developed and new internal functions, such as knowledge management, quality control and rating systems, being introduced. One of the challenges posed by such intense "change processes" is a certain lack of clarity in relation to the EUs mandate vis-à-vis other relevant departments or services within the relevant agency. This appears to be most problematic in agencies with no clear or recent policy statement on evaluation and its functions and mandate.

In line with most recent trends in development strategy and recent debates in the literature on development evaluations, most EUs are either phasing out or significantly reducing their direct involvement in **project evaluations** (see Section 3.3). None of the EUs of the bilateral grant-making agencies are currently involved in project evaluation. In some cases, like Sida, project evaluations are no longer under the responsibility of the EU. In Sida, they are now fully integrated into PCM and, hence, are directly managed by operational departments. Development banks are still carrying out some project evaluation. KfW/FZE is planning to reduce the number of project evaluations and to put a stronger focus on programme and policy evaluations. The two multilateral banks continue to carry out project evaluations, although the number has significantly decreased over the past few years.

Although the shift away from project evaluation is generally recognised as a "good move", and a necessary step in line with recent development trends (including more sector-wide initiatives, joint donor evaluations and country-led approaches), some respondents highlighted that this was not without its consequences. In particular, EUs have fewer opportunities to engage directly with field offices and operational

departments, which risks further isolating the evaluation function from where the "real action" is taking place.

In most cases considered by the study, it is acknowledged that **accountability and learning** are both key purposes of evaluation and that they should be regarded as mutually reinforcing rather than alternatives to each other. This is increasingly being integrated into official evaluation polices adopted by these agencies (see Section 3.1), which are increasingly coherent in relation to key purposes, aims and objectives of evaluation. However, in some of the interviews with staff from within and outside the EUs, the "trade-off" between accountability and learning was often mentioned as a challenge, often revealing quite diverging views among staff as to what they believe the purpose of evaluation should be. Perhaps most importantly, in some cases this lack of clarity in relation to evaluation purposes is seen as a factor impeding concrete choices in relation to what kind of evaluations should be carried out, by whom, with what methods, etc. Over time, it will be important that the same level of coherence achieved at policy level be reflected at the more operational level and shared with all staff and partners.

Similarly, confusion can arise in relation to the merits and limitations of **independent versus external evaluations**. In some agencies, the fact that evaluations are always carried out by external consultants is considered the cornerstone of independence. In others, namely at the WB/IEG, independence is conceived as based on a selection of key criteria, mostly related to the internal governance structure and mandate of the EU (see Section 3.2). In general, concerns were expressed across the board that independence could be at the expense of integration and "voice" within the agencies, leaving the EUs increasingly isolated from the rest of the agency and too far from the field to be considered relevant.

Increasingly, EUs are engaged in "non-evaluation" activities such as rating systems, management follow-up, quality control and knowledge management. These new functions are likely to become more centre stage in the years to come and will require new skills, methods, procedures and products. In relation to knowledge management in particular, various agencies are currently involved in a variety of initiatives aimed at ensuring that evaluations contribute to a **"learning culture"** within the organisation, seeking to build on the abstract notions underpinning the conception of a learning organisation, to create some concrete proposals to influence practices and behaviours. However, more could be done by EUs in this respect, as they could play a vital role to ensure that development practice is increasingly evidence-based and learns from its own experiences, including both successes and failures.

The **professionalisation, skills development and career prospects** of staff of EUs remain a concern, despite recent efforts in many EUs to raise staff profile and to reduce turnover. For many development professionals, the prospect of writing ToRs and managing an ever-growing number of tendering processes and contracts with consultants is not seen as particularly strategic in their professional development. Furthermore, the incentive structures of most agencies do not necessarily reward those in the evaluation departments, which, as a result, are not able to offer clear career opportunities for staff members. Finally, with the notable exception of the WB/IEG and IMF/IEO, tasks assigned to staff in EUs only rarely offer an intellectual challenge, as opportunities to engage with the content of evaluation activities are very limited, if at all available.

Two issues were notable for their absence in the analysis of the EUs considered by the study: this clearly does not necessarily mean that they are not considered relevant by those consulted during the case studies. Very few references were made to the role of local partners in evaluation, and the challenges associated with this. This is particularly surprising given the increasing attention paid by EUs to local and national evaluation processes. Second, it is hard to detect from the analysis the consequences of major evaluation initiatives involving several agencies and (normally) addressing topical issues in development policy. It was unclear, for example, what the likely consequences were of the recent GBS evaluation carried

out on behalf of several donors. What are the implications of its findings for budgetary support? What are the main lessons learnt for the EUs and the agencies involved more generally?

## 4.2    Concluding remarks

The picture that emerges from the analysis presented in this report is a rather composite one. On the one hand, the EUs of the agencies considered by the study share a number of common features. Their mandate is grounded in the emerging consensus developed over the past twenty years, that evaluation has an important and strategic role to play in international development. Furthermore, all agencies share similar challenges in relation to the size of their international portfolio, the tensions among operational, policy and communications functions, and their role vis-à-vis other departments and organisational priorities. Likewise, they are all on a similar journey (although at different stages), from a relatively straightforward model of project evaluation aimed at internal management and accountability, towards more complex models of policy, country and partnership-led evaluation that require new skills, roles and organisational arrangements.

Despite these significant commonalities, a closer examination of some of the key features, mandates, choices and internal arrangements of the EUs considered by the study revealed a significant degree of **diversification**, both in terms of internal arrangements (e.g. management arrangements within the EUs) and in relation to the different roles and responsibilities fulfilled by the EUs. This is partly explained by the fact that the EUs are increasingly expected to fulfil a variety of different roles and to engage in a wide range of activities. This study found that some of the more significant examples of diversification within and between EUs include:

• roles and responsibilities of the EUs, beyond traditional evaluation, to include quality assurance, project ratings, meta-evaluation, capacity building, etc.;

• reference audiences, beyond management and parliaments, to include politicians, civil society organisations, academics, the general public, governments of partner countries, citizens of partner countries, etc.;

• types of evaluation, increasingly moving away from project evaluations, to include policy, strategic, country-led, joint and impact evaluations;

• tools and products, beyond evaluation reports, to include management response systems, knowledge management systems, quality assurance frameworks, etc.

Against this background of diversification and complexity, the evaluation function within aid and development agencies emerges as a **function "in search of identity"**; in some cases, the search is quite advanced and the role, mandate and capacity of the EU is relatively well established. In other cases, the search is still very much a work in progress, and this can result in frustration and confusion, primarily among the EU staff but also in other parts of the organisation.[18]  A good example of the consequences of the "search of identity" underway in many EUs (although mostly related to bilateral agencies) lies in the frequent changes that they have undergone in recent years both in relation to their position in the organisational hierarchy/structure (e.g. DFID/EvD), trying to address the trade-offs between being independent from the management structure and being influential within the organisation, as well as in relation to changes to the internal management arrangements (AfDB/OPEV).

Overall, the analysis conducted suggests an **apparent disconnect** between the rhetoric and discourse on the strategic and growing importance of development evaluation, and the realities of evaluation practice as experienced by the EUs. In order to meet the ambitious expectations set out by the international community and to fulfil the growing number of roles and responsibilities expected of the EUs, more investment is

---

[18]  It is worth reiterating that in this study we did not systematically consider perceptions of the EUs held by staff from other units/departments. However, when we did have the chance to interview non-EU staff, issues surrounding clarity of mandate and confusion over roles and responsibilities emerged.
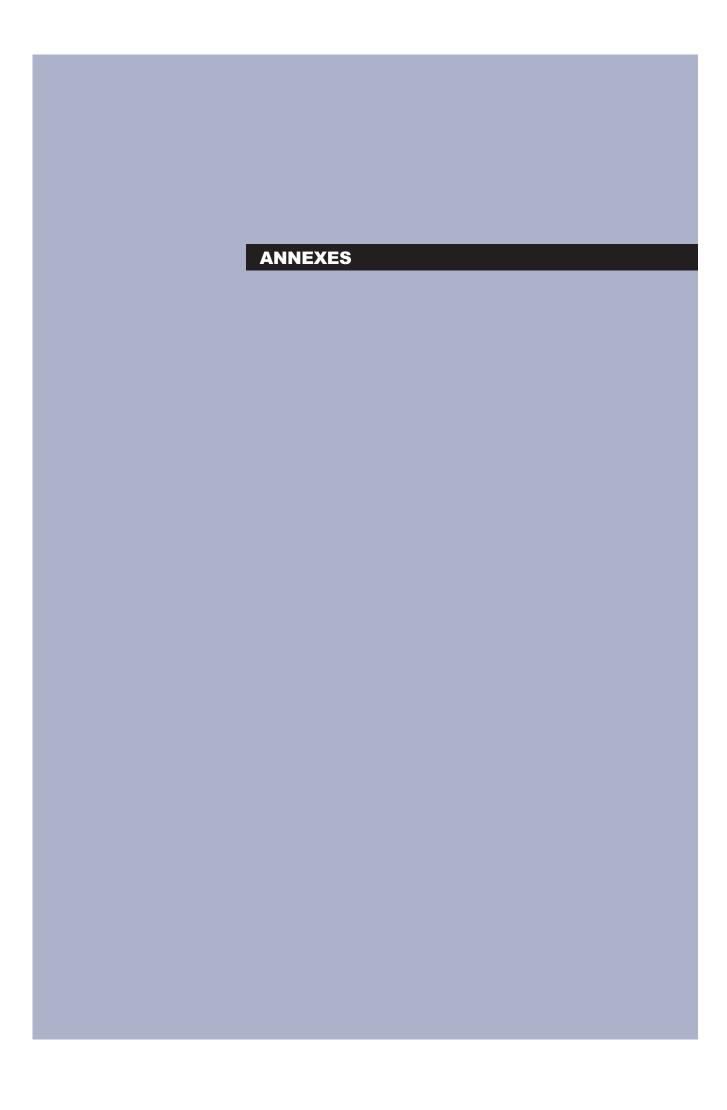
needed, not only in terms of financial resources to fund what are increasingly large-scale and complex evaluation exercises, but (and crucially) also in terms of human resources and capacity. Overall, feedback from the EUs showed that they lacked both financial and human capacity to fulfil their objectives. This disconnect between the rhetoric and the reality of development evaluation could perhaps be further investigated and assessed through an evaluation initiative, aimed at assessing the "goodness of fit" between the current arrangements of the EUs and the objective that they should pursue.

In conclusion, we would propose that a key dimension of the **"evaluation gap"**[19] in international development and aid evaluation is an institutional gap; there is a need for greater attention to the institutional approaches, arrangements and capacity to successfully address the challenges set out by the international community. Investing in the quality of evaluation should certainly be a priority for most agencies.

In addition, greater attention needs to be paid to the nature, origin and potential of **the demand** side of the evaluation equation. Contrary to most domestic policies, development assistance does not lend itself to routine scrutiny of the treasuries and, sometimes, of the parliaments in donor countries. However, in recent years, civil society in the North and South, citizens and partners have been increasingly demanding greater transparency and accountability of aid and development interventions. This has the potential to inject some new energy into the practice and use of development evaluation which, in time, should be reflected in the performance of EUs in all development agencies.

■

---

[19] See the work of the Centre of Global Development initiative on Closing the Evaluation Gap, http://www.cgdev.org/section/initiatives/_active/evalgap.

# ANNEXES

# Annex 1. Activities of EUs

## Table 2. Main activities of EUs

| | Agency/ unit | Types of evaluation | Quality control/ratings | Management response/system | Tools and methods |
|---|---|---|---|---|---|
| **Bilateral agencies** | Sida – UTV | – 40-50 evaluations p.a., 50% led by UTV<br>– No project evaluation<br>– Focus on hematic/sectoral/regional<br>– Increasing number of joint evaluations<br>– Decreasing number of internal evaluations | – Considers DAC standard criteria<br>– UTV and stakeholders engage with consultants in ongoing manner<br>– UTV can be involved in department-led evaluations (may train in future) | – UTV signs off but not involved in response<br>– Response requirement: no set format<br>– Action plan for input within 6 weeks<br>– DG action plan and instructions<br>– Concerns on learning and follow-up | – DAC criteria<br>– No set methodology<br>– UTV to explore impact evaluation<br>– UTV Evaluation Manual (2004) and SiRS used for department-led evaluation |
| | DFID – EvD | – EvD does independent evaluations (country: 5-6 p.a.; policy/thematic: 6); joint evaluations increasing in number<br>– No country-led evaluations yet<br>– EvD may be involved in country programme internal evaluations | – Steering groups maintain quality with EvD<br>– QA by non-participating EvD member (DAC standards)<br>– EvD being used more for internal evaluations and training; ratings for some internal evaluations (to be improved) | – EvD signs off final report<br>– EvD issues recommendations matrix for target audiences for response within 4 weeks<br>– EvD not responsible for follow-up but sometimes involved | – DAC criteria<br>– No set methodology: depends on type<br>– Guidelines available for internal evaluation |
| | Danida – ED | – ED does external evaluations (ex-post and ongoing; geographical, thematic); about 8 p.a.<br>– Recent shift from project to thematic, can be multidisciplinary<br>– Prioritises joint evaluations (about 50% of total) | – Policy (2006) and Guidelines (1999) to clarify<br>– Training outsourced<br>– Stakeholder engagement depends on type<br>– ED does quality control<br>– Aim to build capacity in developing countries (50% of consultants) | – ED signs off final report<br>– Follow-up memo issued to relevant departments and ED controls implementation | – Evaluation Policy (according to DAC) and Guidelines<br>– No set methodology<br>– ED guidelines in 2006 for internal programme completion reports |
| **Banks and lending agencies** | IMF – IEO | – IEO covers independent evaluations, small number per year on IMF mandate issues<br>– ToR prevents IEO from "interfering with operational activities, including programmes | – Staff and member govts plus external observers are involved, IEO signs off<br>– Procedures in place for comment | – Management prepares comments<br>– Executive Board views comments and IEO responses; Board conclusions not binding | – No information regarding methods and tools readily available<br>– Specific guidelines not referred to on website |
| | WB – IEG | – IEG responsible for independent evaluations: project assessment reports, country assistance evaluations, sector and thematic evaluations, corporate evaluations, corporate reviews<br>– All project completion reports reviewed and validated by IEG but decreased % of project assessment reports over the years (1 in 4) | – Six-point rating scale for outcomes.<br>– IEG reviews self-assessments at end of project reviews.<br>– Increased consistency in ratings with quality assurance function | – Final report presented to Board through CODE<br>– CODE signs off report and management response<br>– Management response and plans for follow-up included in report<br>– IEG and CODE follow up on recommendations | – Toolbox for each type of independent evaluation<br>– Common ratings for outcomes<br>– Focus on impact evaluation |

▶ **Table 2**

| | Agency/ unit | Types of evaluation | Quality control/ratings | Management response/system | Tools and methods |
|---|---|---|---|---|---|
| **Banks and lending agencies** | AfDB – OPEV | − Move towards country/thematic/sector studies<br>− Joint evaluations<br>− Backlog in internal/self-evaluation<br>− OPEV covers independent evaluation at project/meso level | − Follow ECG standards<br>− QA and ratings weak for projects; peer/internal reviews envisaged<br>− Training not significant<br>− Can have limited input<br>− Timing and independence a problem | − Only higher-level evaluations subject to management response<br>− Feedback/follow-up not institutionalised or systematic<br>− Lack of clarity in OPEV relations with Board | − Guidelines: evaluations (through ECG); results-based management; etc.<br>− Impact assessment seen as unrealistic at this stage<br>− Country assessment guidelines in draft |
| | KfW – FZE | − Only ex-post evaluations, 3-5 years after completion<br>− 120 carried out in 2006<br>− FZE to look at thematic studies, etc.<br>− Some joint evaluations | − FZE responsible for quality control<br>− Internal staff carry out part of the evaluations as a kind of training<br>− Internal and external stakeholders involved | − FZE signs off report<br>− Informal management response | − Uniform methodological approach including impacts<br>− DAC criteria and KfW/DAC core crosscutting issues<br>− "State-of-the-art" sector standards |
| **Multilateral** | EU – JEU | − JEU looks at country (5-6)/regional (2-3); sector (17); instrument/channel (7); and regulations (14)<br>− Programme/project rare, done by delegations<br>− ROM looks at results based on sub-sector information (1,500 p.a.)<br>− Some joint evaluations | − QA covered by JEU<br>− Many stakeholders involved | − JEU presents evaluation findings to RELEX family of Commissions who take action on conclusions and recommendations<br>− Feedback is followed up by JEU one year later | − Methodological framework and tools led to new evaluation manual (2005)<br>− Guidelines and tools on website<br>− DAC criteria |
| **NGOs** | Oxfam – PLA | − Works on formal systems and informal learning and sharing<br>− Programme, thematic and strategic evaluation<br>− Impact evaluations to be carried out for large donor-driven programmes<br>− To strengthen local-level support<br>− Some joint evaluations | − New direction (former approach not robust or systematic) recently introduced (June 2007)<br>− New quality indicators and milestones<br>− Prefer using internal staff because of learning opportunities, though global strategic evaluations are carried out by external consultants | | − No formal policy but some "mandatory procedures" (not systematically implemented)<br>− Operational guidelines for M&E and learning<br>− New guidance to come on M&E requirements<br>− Recently accepted need |

## Table 3. Communication activities carried out by EUs

| | Agency/ unit | Disclosure and publication policy | Dissemination | Communications | Follow-up/use |
|---|---|---|---|---|---|
| **Bilateral agencies** | Sida – UTV | – No policy in place<br>– All evaluations in same format and must comply with Access to Information legal requirements | – Dissemination plan in process<br>– Ongoing review of dissemination methods<br>– UTV librarian in charge of publication and dissemination | – Communication plan in process<br>– Website; a few in hardcopy; CD ROMs<br>– Outputs depend on study<br>– Seminars, press conferences | – More to be done on lesson learning and using findings<br>– SADEV to take responsibility for accountability; UTV to look at learning<br>– Integration into decision making through MRE but this has been slow and follow-up limited (under review)<br>– UTV links with Swedish NGOs through training, but not systematic<br>– Evaluation report published |
| | DFID – EvD | – No policy in place | – EvD involved in dissemination | – Full reports on website; printed copies to country offices and others<br>– Targeted seminars, lunchtime talks, newsletters, intranet, press packs, PRISM<br>– Communication strategy being drafted | – Lessons from evaluation do not always feed into decision making<br>– EvD feeds lessons into Corporate Strategy Group and Programme Guidance Group; IA aims to help improve performance through lessons learnt<br>– EvD has role regarding ongoing monitoring for implementation under review<br>– Evaluation report published 2004 |
| | Danida – ED | – Every evaluation must have résumé in English and Danish and conform to form | – No fixed policy<br>– ED contributes actively to dissemination | – No fixed policy<br>– In Danish if thought to be of interest to broader public<br>– Workshops and seminars | – Best practices compiled and formulated |
| **Banks and lending agencies** | IMF – IEO | – Standard rules for publication | | – External stakeholders can comment<br>– Readable summaries; outreach conference | – Annual report sets out main conclusions and recommendations to assess implementation |
| | WB – IEG | – Yes, disclosure policy on reports and management response agreed by Board in 2006 | – Recent efforts to improve dissemination and communication<br>– Client engagement is key<br>– Diversification of products, importance of targeting different audiences | – Different media (including web and press)<br>– Both internal and external audiences are key<br>– Press releases and "web campaigns" on report release | – Data about usefulness and strategic value of IEG reports still patchy but increasing effort to monitor it (e.g. website statistics)<br>– Quality of IEG product rated highly by clients<br>– Targeting awareness of different key actors<br>– Some worries that despite efforts IEG reports not seen as strategic by Bank staff at different levels (particularly CAS)<br>– Board uses IEG reports |

▶ **Table 3**

| | Agency/ unit | Disclosure and publication policy | Dissemination | Communications | Follow-up/use |
|---|---|---|---|---|---|
| **Banks and lending agencies** | AfDB – OPEV | – No policy in place | – Multiyear knowledge management strategy now underway | – Website<br>– Multiyear knowledge management strategy now underway<br>– Publications and workshops, but not systematic | – No review of effectiveness owing to complexity and lack of experience<br>– OPEV carried out some review of evaluation results<br>– Weak and incomplete feedback system<br>– Multiyear knowledge management strategy now underway<br>– Quarterly evaluation report |
| | KfW – FZE | – Full reports are internal only as confidential information is included (short version for public) | – Various departments also used in dissemination (policy/development economics/strategy) | – Website in German and English<br>– Discussion with wide range of stakeholders at meetings and workshops | – FZE works on QAG to see if experience is taken on board<br>– Not obligatory to comply with QAG recommendations<br>– FZE takes charge of co-ordination of BMZ evaluations<br>– Evaluation report biennially |
| **Multilateral** | EU – JEU | – To follow format | | – Available on website since 1997 (previous reports in hard copy) | |
| **NGOs** | Oxfam – PLA | Strict sign-off procedures make it difficult to post on website. All global strategic evaluation will be posted in the future | | – Very few evaluation reports on website<br>– Some country programmes have published their own reports or posted elsewhere | |

# Annex 2. List of People Consulted

## AfDB

### OPEV

| | |
|---|---|
| Mr Douglas Barnett | Acting Director |
| Mr Mohamed Manaïi | Chief Post- Evaluation Officer |
| Mr Anthony Curran | Senior Evaluation Officer (Knowledge Management) |
| Mr Detlev Puetz | Principal Post- Evaluation Officer |

### Non-OPEV Operations

| | |
|---|---|
| M Philibert Afrika | Director Operations Policies and Compliance Department ORPC |
| Mr Frank Black | Director, Regional Department South 2 (ORSB) |
| Mr Mahdi Ahmed Ismaëil | Principal Financial Analyst Governance and Financial Reforms Department |
| M. Gilbert Mbesherubusam | Director, Infrastructure Department OINF |
| Mr Mohammed M. Youssouf | Manager, ISCD.1 Human Resources |

### Executive Directors (Board)

| | |
|---|---|
| M Emmanuel Carrèere | Alternate Executive Director, France, Belgium |
| Mr Richard Dewdney | Executive Director for Germany, Netherlands, Portugal and the UK |
| Aud Marit Wiig | Executive Director for Denmark, Finland, India, Norway, Sweden and Switzerland |

## Danida

| | |
|---|---|
| Niels Dabelstein | (Former) Head of the Evaluation Department |

## DFID

### EVD

| | |
|---|---|
| Nick York | Head EvD |
| Robin Russell | Deputy Head |
| Julia Compton | Senior Evaluation Manager |
| John Heath | Senior Evaluator and Quality Assuror (WB secondee) |
| Jeremy Clarke | Senior Evaluator |
| Kerstin Hinds | Team Leader |
| Alison Girdwood | Team Leader |
| Iain Murray | Programme Manager |

**Non-EVD**

| | |
|---|---|
| Richard Calvert | Director Finance and Corporate Performance Division |
| Mike Noronha | Head of Internal Audit |
| Sue Owen | Director General, Corporate Performance |
| Mark Lowcock | Director General, Policy and International |
| Sam Sharpe | Acting Director, Policy and Research Division |
| Paul Godfrey | Head of country team, Asia Division |

**EU**

| | |
|---|---|
| Jean- Louis Chomel | Head of Evaluation Unit – Relex |

**KfW**

**FZE**

| | |
|---|---|
| Joachim Heidebrecht | Vice President (Managing Director) |
| Ulrich Jahn | Senior Project Manager |

**Other Departments**

| | |
|---|---|
| Dr Philipp Lepenies | Senior Economist, Policy Department |
| Bettina Zoch | Intern, Policy Department |
| Dr Jüuergen Welschof | Division Chief, Competence Centre Water Resources Middle East/Policy Division |
| Dr Martin Raschen | Principal Country Manager, Bangladesh Nepal |

**Oxfam**

| | |
|---|---|
| Annabel Wilson | Senior Advisor, - Programme Learning and Accountability Team |

**Sida**

**UTV**

| | |
|---|---|
| Eva Lithman | Director Department for Evaluation and Internal Audit (UTV) |
| Stefan Molund | Deputy Head UTV |
| Mattias Lindgren | Evaluation Officer |
| True Schedvin | Evaluation Officer |
| Fredrik Uggla | Evaluation Officer |
| Maria Elvira | Librarian, UTV |

**Non-UTV**

| | |
|---|---|
| Anders Berlin | Deputy DG, Director General's Office |
| Anders Molin | Head of Health Unit,  Department for Democracy and Social Development |
| Cecilia Chroona | Programme Officer, Eastern Europe/Central Asia, Regional Department for Europe |

## World Bank

**IEG**

| | |
|---|---|
| Patrick Grasso | Advisor |
| Alain Barbu | Manager, Sector and Thematic Evaluation Group |
| Ajay Chhibber | Director |
| Ray Rist | Senior Evaluation Officer |
| Keith Mackc Kay | Senior Evaluation Officer |
| Niels Fostvedt | Consultant |
| Gita Gopal | Lead Evaluation Officer |
| Klaus Tilmes | Manager, Knowledge Programs and Evaluation Capacity Development Group |

**Non IEG**

| | |
|---|---|
| Susan Stout | Manager of Results Secretariat |

**Others**

| | |
|---|---|
| Robert Picciotto | Evaluation expert and former director of OED |

# Annex 3. Methodological Guidance for "Full" Case Studies

## Purpose and objectives

The "full case studies" (FCS) are the main sources of information for the benchmarking study. In particular, the full case studies should provide analytical as well as detailed descriptive information on the agencies' evaluation function and units in order to inform a comparative analysis to be developed in the final report.

In accordance with the ToRs, the main objectives of the FCS are to analyse:
– key features of the evaluation function within the organisation, including management arrangements and placement within organisational structure;
– main aspects of evaluation systems, processes and tools;
– practices involved in commissioning, managing and supporting evaluation processes.

## Main questions to be addressed

The main questions to be addressed by the case studies are set out in the ToRs (Section 6). A generic interview schedule based on the questions set out in the ToRs is in Annexe 1. The schedule should be used as a basis for all individual and group interviews and it should be adapted depending on the role, area of expertise and position of the respondents. The schedule should be sent in advance to all organisations involved in FCS.

## Methodological guidelines

Two main methods for data gathering and analysis will be used in the FCS.

1. Review of secondary sources, mostly documents produced by the evaluation units of the different organisations considered. This will be initially based on documentation available on the website, complemented by reports/documents made available directly from the organisations. These documents will include: strategy/policy documents, methodological guidelines, reporting procedures and formats, review and evaluation reports, etc. All electronic copies collected will be included in a CD ROM.

2. Collection of primary data, mainly through individual or group interviews. These should include: staff from the evaluation units, ranging from management to technical/operational staff; relevant staff form other units and departments working with the evaluation unit (in particular the managers of the department where the evaluation unit sits); any other staff member from policy or operational departments who could be useful sources of information. A combination of individual and group interviews is recommended, although this will depend on the availability of staff in the organisation.

A contact person in the evaluation unit should be identified and, in the first instance, he/she will be asked to provide assistance to set up the interviews. However it is important that during the visit to the evaluation unit opportunities be sought to engage with any staff member or associate who is willing to collaborate and share their views and experiences.

In addition to the "core" set of interviews with evaluation unit staff members, ad hoc interviews will be carried out with other key informants (e.g. previous heads of evaluation unit, members of DAC Evaluation Network, experts and evaluators, etc.).

## Outputs

The main output of each FCS is a detailed analytical report based on the information collected and the documents reviewed. The report should be 10-15 pages maximum and structured around the main sections and sub-sections outlined in the generic interview schedule (see Annex 1). The report should also include the following annexes:

– list of all people consulted/interviewed, including role/title;

– detailed programme/agenda of meetings and interviews;

– list of all documents consulted (including URLs where they can be located). All collected documents will be included in a CD ROM that will be annexed to the final report of the benchmarking study).

## Issues to be covered in the case studies and key questions for interviews

### Section 1. Background information

*NB: Basic information on this section should be gathered from the website. In case the information is not available, please do send relevant material to the researcher prior to the interviews.*

1.1 Key features of the evaluation function

• Is there a policy/strategy for evaluation? If so, what are its key objectives and features?

• What different types of evaluation are conducted (e.g. independent, internal reviews, country, sector, cross-cutting issues, etc.)?

• What are the key features of the evaluation unit/team/function? (i.e. number of staff and composition of the team/organisational structure; annual budget; staff profile?

• Where does the evaluation unit/team/function sit in your organisation? (e.g. In which department? Is it only a headquarter function or also in field offices? etc. ). In particular, what is the relationship between the evaluation function and the overall management structure?

### Section 2. How is evaluation implemented?

2.1 Roles and responsibilities

• What is the specific role of the evaluation unit? (i.e. commissioning, technical support, managing, training and capacity building, dissemination, etc.)? What is the role of other units/individuals in these processes?

• Who commissions evaluations and who carries them out? Who determines when evaluations take place? Who determines the quality of the evaluations and approves reports?

• How do you address the tension between the independence of evaluation and its integration with other functions/services/operations?

2.2. Methodologies, tools, instruments

• Is there a general/standard methodological approach for all evaluations?

• Is there a dedicated methodology for impact assessment?

• What methodologies are used for specific evaluations such as sector, country, instrument evaluations?

• What has the experience with joint evaluations been?

• Is there a dedicated process or system for project evaluation (e.g. annual reports, project completion report, ratings, etc.)? Is it comprehensive (i.e. all projects) or selection-based (i.e. a sample every year)?

• What are the main feedback mechanisms to ensure that lesson learning results from evaluations?

Section 3. Use of evaluation findings

3.1. Dissemination, publication and communication

 • Are all evaluations published? If not, what are the criteria? What are the main barriers involved in publishing evaluation findings?
 • Is there an overall annual report? What is it based on?
 • How are evaluation findings disseminated and communicated within and outside the organisation (e.g. seminars, etc.)?

3.2 Relationship between evaluation and decision making, other operational functions and external actors

 • How are evaluation findings integrated into decision making processes? Who is responsible for this? What are the main challenges (e.g. tension between independence and involvement in decision making, timing of evaluation findings versus timing for decision making, etc.)?
 • How is evaluation integrated with other key operational functions such as planning, monitoring and information systems, results-based management, auditing and "value-for-money" systems, etc.)
 • What is the relationship between the evaluation function and other actors, such as parliaments, other government departments (the Treasury, in particular), civil society, partners, etc.?

*ODI, October 2006*

# Annex 4. Terms of Reference

## 1. Background

AFD's evaluation system is currently undergoing changes. The following options, in particular, are being looked at:

– the launch of decentralised external evaluations (commissioned by the geographic departments);

– reorientation of the structure of evaluations towards a capitalisation of experience and lesson learning, on the one hand, and towards upstream functions of evaluation (methodology) and downstream functions (dissemination), on the other hand;

– the reform of evaluation governance and reporting bodies (Evaluation Committee);

– the development of project completion reporting and the project rating system.

The evaluation system is also confronted with important methodological challenges coming up with the diversification of AFD financing and, particularly, the increasing part of budget assistance, programme assistance and financial intermediation.

AFD would also like to invest significantly in a better understanding of the impacts of its actions.

Finally, the development of partnership relations in the area of evaluation is very desirable, first with project owners, where evaluations would be considered as part and parcel of the development action, but also with other development financing partners.

In the context of its evaluation system reform, AFD wishes to have a better understanding of the evaluation systems of a certain number of donors and, in particular, of how such insti-

tutions respond to the questions that frame AFD's ongoing reflection on its own evaluation system.

## 2. Objectives of the study

The study involves benchmarking the issues raised by the reform of AFD's evaluation system in light of the evaluation policies and practices of other donors.

The expected outcome for AFD is a better factual understanding of the evaluation systems of a certain number of comparable institutions. From a more analytical standpoint, it is also expected to provide a deeper understanding of how evaluations are organised in the comparable institutions studied, and of how they contribute to a certain number of cross-cutting functions, notably:

– accountability

– information and knowledge

– partnerships

– capitalisation of experience

– feedback

Finally, the objective of the study is also to compare the different evaluation systems studied by highlighting the different choices made, their consequences in terms of "service rendered" and how the institutions feel about the strengths and weaknesses of their respective systems. This comparative analysis will also attempt to identify the main internal factors (pertaining to the missions of the institution) and main external factors (national policies) accounting for the choices made and any identified differences.

## 3. Expected outcomes of the study

The report should contain the following, in an order to be determined by the consultant:

– a summary sheet describing the main features of evaluation systems, including references and sources, for those institutions that are the subject of desk-based case studies;

– a detailed report for each institution subject of full case studies; this report should be based on the questions listed in Section 6 below. The reports should describe the overarching logic of the evaluation function for each institution, highlighting the choices made, the results obtained, as well as any tensions between objectives, areas not sufficiently covered, difficulties and ambiguities encountered;

– a comparative analysis of the different systems studied, highlighting available options, factors explaining the choices made among these options, strengths and weaknesses of different systems, consequences in terms of functioning, publications and role of evaluation.

A copy of the collected documentation shall be submitted along with the report to AFD.

## 4. Use to be made of the study

The study is designed, first, to inform AFD's internal reflection and to shed light on its choices in matters of evaluation. Nevertheless, the main report (cf. Section 3 above) will be made available to the institutions visited, as well as to the DAC network of evaluators. Should AFD judge it useful, the report could also be published in its collection comprising publications on project experience capitalisation.

## 5. Organisation of the study

The study will be carried out by a consultant and based on case studies.

The five following institutions will be the subject of full case studies, based in particular on interviews with their evaluation services.

The donors AFD has in mind are:
– DFID (United Kingdom)
– Sida (Sweden)
– KfW (Germany)
– IEG (World Bank)
– African Development Bank

The four following institutions will be the subject of desk-based case studies, based on documentation and key informants:
– Danida (Denmark)
– Save the Children UK (or Oxfam)
– IMF
– EU

A total of five institutions will be visited. The consultant will supplement the information thus collected with documentation, notably peer review data from the DAC and expert interviews and focus groups. Interviews with evaluation experts (such as the head of the DAC evaluation group) could also be envisaged for the mission.

If the consultant judges it necessary, AFD will inform the evaluation units of the different institutions of the exercise by sending them the Terms of Reference of the study, and let them know that the consultant will be contacting them to arrange interviews. It will be up to the consultant to organise his/her schedule to meet with the different evaluation units concerned for work sessions of not less than half a day. It will be entirely

up to the consultant to decide whether or not he/she should conduct interviews outside the evaluation units of the institutions.

The consultant shall inform AFD of any meetings arranged. If necessary and possible, an Evaluation Unit executive-level officer shall accompany the consultant during these visits.

The consultant shall collect from the donors visited any relevant documentation (evaluation policy, methodological framework, examples of evaluations, etc.), which he/she shall submit to AFD upon delivery of the study.

The consultant shall write up his/her report from the information collected from the donors interviewed.

The consultant shall naturally feel free, in his/her comparative analysis, to seek and use information sources other than sources listed above.

## 6. Points to be covered in the full case studies

The interviews and focus groups carried out as part of the case studies will cover the following main questions:

### Organisation of evaluation systems

**Evaluation policy:** What are the goals of evaluations? Does the institution have an evaluation policy? Formalised or not? A strategic document or a set of instructions? Nature and contents of the policy. Does the current policy cover the goals?
**Different types of evaluation:** What are the different types of evaluation conducted by the institution? Centralised, decentralised? Project, sector, instrument or thematic evaluations?

**Responsibility for evaluations:** Who commissions and conducts evaluations? Who does the actual evaluating? External or internal evaluators? Consultants or researchers? If necessary, according to the different types of evaluation.
**Role of the structure dedicated to the conduct of evaluations:** What are the missions of the structure charged with conducting evaluations? How does it conduct them? What problems does it encounter?
**Position in organisational structure, budget, staff:** The place of the evaluation structure in the organisation chart of the institution, approximate annual budget for consultants, staff, number of evaluators, staff profile.

### Evaluation governance and cycle

**Types of evaluation:** Do the different types of evaluation (centralised, decentralised) have different governance and cycle modalities?
**Committees and bodies involved in the evaluation system:** What are the different internal and external bodies involved in the evaluation cycle? What are their roles?
**Programming of evaluations:** How are evaluations programmed? Systematic, at random, well thought-out, etc.? Who decides?
**Schedule of implementation:** At what point in the cycle of actions (projects, programmes, etc.) does the evaluation take place? Mid-course evaluation? Operational evaluation (during the preparation of a subsequent phase)? Who decides? Who conducts the different types of evaluation?
**Submission of evaluations:** How are evaluation wrap-ups organised? How are evaluations submitted?
**Dissemination:** What is the policy on the dissemination of evaluations? Are there any obstacles to dissemination? What are these? Is there a tension between communication policy and the policy on evaluation dissemination?
**Feedback and monitoring:** How is feedback organised? A vertical approach (written recommendations) or a horizontal

approach (capitalisation of experience)? Evaluation monitoring and/or recommendations? Does the feedback system appear satisfactory? Why?

### Methodology and scope of evaluations

**Typical methodology:** Are there typical evaluation methodologies based on evaluation type? Position relative to the DAC criteria?

**Evaluation of programme assistance:** Is there a dedicated methodology for programme assistance evaluation? Experience in the area? Any special problems?

**Impact evaluation:** Is there a dedicated methodology for impact evaluation? How is the information on the impact of interventions assessed? Experience in the area? Any special problems? Projects?

**Expanded evaluations (sector, country, instrument, etc.):** Are there any dedicated methodologies for expanded and strategic evaluations? What is the experience in this area? Any special problems?

**Joint evaluations:** Practices and experiences with the evaluation of co-financing and parallel financing operations. Other experiences with joint evaluations.

### Integration of evaluations in decision-making and operating processes

**Co-ordination with operational departments:** How is co-ordination with operational departments done? Independence vs. integration? Is this perceived as a tension? Problems encountered in the internal integration of the evaluation unit.

**Co-ordination with monitoring and results measurement system:** Co-ordination of evaluations with monitoring information system? With results measurement system (Results-based management)? Problems encountered?

**Relationship with the outside world** (other institutions, parent ministries, parliament, civil society, other donors, etc.):

Description of relations of the evaluation unit with external partners. Problems encountered.

**Partnership with beneficiary countries in the evaluation process:** Link between evaluation and recipient institutions? Does evaluation have a partnership objective? Problems encountered.

### The work of the evaluation unit

**Project completion reports:** Existence of project completion reports (PCRs), coverage rate, use of PCRs by the evaluation unit, problems encountered?

**Rating:** Description of the rating system used at project completion and after completion? Role of the evaluation unit in the rating system? Use of rating. Problems encountered?

**Publications of the service:** Notes, evaluation reports, capitalisation of experience reports, communication publications, others?

**Annual reporting:** Is there an annual evaluation report? Type, target public, objective?

## 7. Methodology, schedule and budget for the consultancy

The consultant shall prepare a proposal including description of activities, timeline, methodology, composition of the team and budget based on these terms of reference.

# Acronyms

| | |
|---|---|
| AFD | French Development Agency |
| AfDB | African Development Bank |
| ARDE | Annual Review of Development Effectiveness (World Bank) |
| AROE | Annual Report on Operations Evaluations (World Bank) |
| BMZ | German Federal Ministry for Economic Co-operation and Development |
| CAE | Country Assistance Evaluation (World Bank) |
| CODE | Committee on Aid Effectiveness |
| DAC | Development Assistance Committee (OECD) |
| Danida | Danish International Development Agency |
| DFID | UK Department for International Development |
| DG | Director General |
| DGE | Director General, Evaluation (World Bank) |
| ED | Evaluation Department (Danida) |
| EU | European Union (in context) |
| EU | Evaluation Unit (in context) |
| EvD | Evaluation Department (DFID) |
| FZE | Independent Evaluation Department (KfW) |
| GBS | General Budget Support |
| GTZ | German Technical Co-operation |
| IBRD | International Bank for Reconstruction and Development |
| IDA | International Development Association |
| IEG | Independent Evaluation Group (World Bank) |
| IEO | Independent Evaluation Office (IMF) |
| IFC | International Finance Corporation |
| IMF | International Monetary Fund |

| | |
|---|---|
| JEU | Joint Evaluation Unit (EU) |
| KfW | German Development Bank |
| M&E | Monitoring and Evaluation |
| MIGA | Multilateral Investment Guarantee Agency (World Bank) |
| NGO | Non-governmental Organisation |
| ODI | Overseas Development Institute |
| OECD | Organisation for Economic Co-operation and Development |
| OED | Operations Evaluation Department (World Bank) |
| OPEV | Operations Evaluation Department (AfDB) |
| PCM | Programme Cycle Management |
| PLA | Programme, Learning and Accountability Team (Oxfam) |
| PPAR | Project Performance Assessment Report (PPAR) |
| RBM | Results-based Management |
| QAG | Quality Assurance Group (World Bank) |
| SADEV | Swedish Agency for Development Evaluation |
| Sida | Swedish International Development Cooperation Agency |
| SiRS | Sida Rating System |
| ToR | Terms of Reference |
| UTV | Department for Evaluation and International Audit (Sida) |
| WB | World Bank |