Study on the uptake of learning from EuropeAid's strategic evaluations into development policy and practice

Annexes

June 2014

Study carried out on behalf of
the European Commission

Contracted Agency
IBF International Consulting
Contact Person: Giada Tu Thanh
tuthanh@ibf.be

Framework Contract No COM 2011/Lot 1

Specific Contract N° 2013/317962

Study on the uptake of learning
from EuropeAid's strategic evaluations
into development policy and practice

This Study was commissioned by the Evaluation Unit of the
Directorate General for Development and Cooperation – EuropeAid
(European Commission)

Jean Bossuyt, (ECDPM)
Louise Shaxson, Ajoy Datta (ODI)

TABLE OF CONTENTS

ABBREVIATIONS AND ACRONYMS

3ie	International Initiative for Impact Evaluation
ADB	Asian Development Bank
AF	Action Fiche
AfD	Agence Française de Développement
AfDB	African Development Bank
AIDCO	EuropeAid Cooperation Office
ASEAN	Association of Southeast Asian Nations
AusAID	Australian Agency for International Development
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (Germany)
BMZ-E	Evaluation and Audit Division of the Federal Ministry for Economic Cooperation and Development
BS	Budget Support
CAR	Central African Republic
CIDA	Canadian International Development Agency
CoA	Court of Auditors
CODE	Committee on Development Effectiveness (WBG)
CRIS	Common RELEX Information System
DANIDA	Danish International Development Agency
DEV	Directorate General for Development
DEVCO	Directorate General Development and Cooperation - EuropeAid
DFID	Department for International Development (UK)
DG	Directorate General
DGE	Director-General Evaluation (WBG)
DGMDP	Directorate General for Globalisation Development and Partnership in the Ministry of Foreign and European Affairs
DPME	Department of Performance Monitoring and Evaluation (South Africa)
EBP	Evidence Based Practice
EBRD	European Bank for Reconstruction and Development
EC	European Commission
ECDPM	European Centre for Development Policy Management
EEAS	European External Action Service
EP	European Parliament
EPPI	Social Research Unit and Evidence for Policy and Practice Information and Coordinating Centre
ESRC	Economic and Social Research Council
EU	European Union
EUD	EU Delegation
EVINFO	Evaluation summary following the standard DAC format
FC	Fiche Contradictoire
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
GP	General Practitioner
GTZ	Deutsche Gesellschaft für Technische Zusammenarbeit
IACDI	Independent Advisory Committee on Development Impact
IADB	Inter-American Development Bank
ICAI-ReD	Independent Commission for Aid Impact - Research and Evidence Division

IED	Independent Evaluation Department
IEG	Independent Evaluation Group
IF	Identification Fiche
IFAD-IOE	International Fund for Agricultural Development - Independent Office for Evaluation
IMF-IEO	International Monetary Fund: Independent Evaluation Office
IOB	Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (NL)
IPA	Innovations for Poverty Action
iQSG	Inter-service Quality Support Group
JEU	Joint Evaluation Unit
JICA	Japan International Cooperation Agency
JPAL	Abdul Latif Jameel Poverty Action Lab (MIT)
KfW	Kreditanstalt Für Wiederaufbau (German Development Bank)
KOICA	Korea International Cooperation Agency
KPP	Knowledge, Politics and Power
MTR	Mid-Term Review
NESTA	National Endowment for Science, Technology and the Arts (UK)
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NORAD	Norwegian Agency for Development Cooperation
ODA	Official development assistance
ODI	Overseas Development Institute (UK)
ODI-RAPID	Overseas Development Institute: Research and Policy in Development programme
OECD-DAC	Organisation for Economic Cooperation and Development - Development Assistance Committee
oQSG	Office Quality Support Group
RAND	Research and Development Corporation
RBM	Results Based Management
RCT	Randomised Control Trial
REF	Research Excellence framework
RELEX	Directorate General for External Affairs
ROM	Results Oriented Monitoring
RSS	Rich Site Summary
SADEV	Swedish Agency for Development Evaluation
SCIE	Social Care Institute for Excellence
SIDA	Swedish International Development Agency
ToRs	Terms of Reference
UK	United Kingdom
UNDP-EO	United Nations Development Programme Evaluation Office
US	United States
USAID	United States Agency for International Development
WB	World Bank
WBG	World Bank Group
WSIPP	Washington State Institute for Public Policy

1. LITERATURE REVIEW OF THE UPTAKE OF RESEARCH AND EVIDENCE IN DEVELOPMENT COOPERATION

1.1 Promoting research uptake

This review draws on research commissioned by the UK Alliance for Useful Evidence in 2013. It aims to provide inputs to analysis and recommendations of the study on the uptake of strategic evaluations within EC Development Policy. It is split into four sections: the first explores evidence production; the second assesses evidence communication and brokering; the third reviews policymaking and evidence use; whilst the final section looks at sustainability and capacity development.

A systematic review of 24 interview studies with health policymakers by Innvaer et al. (2002) indicates that the following factors positively or negative affect the use of research by policymakers.

Facilitators	Number of studies
Personal contact between scientists and policy makers	13
Timeliness and relevance of the research	13
Research that includes a summary with clear recommendations	11
Research that confirms current policy or endorses self-interest	6
Good quality research	6
Community pressure or client demand for research	4
Inclusion of effectiveness data	4
Total studies	**24**

Barriers	Number of studies
Absence of personal contact between scientists and policy makers	11
Lack of timeliness and relevance of research	9
Mutual mistrust between scientists and policy makers	8
Power and budget struggles	7
Poor quality of research	6
Political instability or high turnover of policy making staff	5
Total studies	**24**

Figure 1: Facilitators and barriers to use of research by policy makers, identified in a systematic review of 24 interview studies (Innvaer et al.. 2002).

The form of transfer and the willingness of policymakers to take action are important factors that influence the use of academic findings (Mermot, 2004).

1.2 Organisational learning/learning organisations

Rashman et al. (2008) define four broad areas of factors inhibiting and supporting organisational learning (figure 2):
- **Contextual factors** include both external environmental influences on the organisation and the organisation's policy and practice context;
- **Organisational characteristics** such as culture, leadership, resources to foster learning;
- **Intra and inter-organisational relationships** (trust, compatibility); and
- The **nature of knowledge**.

The concept of organisational learning remains vague in the literature. Most frameworks are normative and lack empirical foundation. It is not clear what role evaluations play within these frameworks. Within the aid discourse, evaluations are said to be useful, but certain characteristics and perceptions of evaluations may inhibit this – e.g. if evaluations are perceived as control or accountability instruments.

1.3 Evaluation utilisation

Cousins and Leithwood (1999) developed a framework to cluster different factors influencing evaluation use into two categories – evaluation implementation and decision and policy settings – with twelve characteristics. Characteristics include commitment/receptiveness to evaluations which covers the extent to which an organisation is resistant to evaluation. The characteristic political climate is defined as the political orientation of the people who commissioned the evaluation, the extent to which the decision maker is dependent on external sponsors, internal rivalries, budget fights and power struggles (Johnson et al.., 2009: 385).

Based on this framework, Johnson et al. (2009) summarised the findings of 41 empirical studies from 1986 through 2005 in a literature review.

Evaluation implementation			Decision and policy settings		
Characteristic	Number of studies	Relationship to evaluation use	Characteristic	Number of studies	Relationship to evaluation use
Communication quality	11	Most found strong positive relationship	Personal characteristics of evaluation user	9	Most found positive relationship
Timeliness	7	Most found positive relationship	Commitment / receptiveness to evaluation	8	Some found positive relationship
Evaluator competence	6	Most found positive relationship	Political climate	6	Most found positive relationship
Evaluation quality	6	Some found a (weak) relationship	Decision characteristics	5	All found positive relationship
Findings	6	Mixed conclusions	Competing information	3	Mixed conclusions
Relevance	6	Mixed conclusions	Information needs	2	Both found positive relationship
Credibility	4	Mixed conclusions			

Figure 2: Factors influencing evaluation use (Johnson et al... 2009)

1.4 Evidence production

'Good evidence'

Part of the challenge of making better use of evidence is the need to understand what constitutes reliable evidence of social change and how outcomes or results of specific interventions can be verified (Nutley et al., 2013). Overall, Nutley et al. (2013) argues that evidence's quality depends on a range of factors including what we want to know, why we want to know it and how we envisage that evidence being used. Moreover, what counts as good evidence will vary according to the context in which it is produced and used.

Many bodies (such as government agencies, service providers, professional associations, non-governmental organisations and academic organisations) provide practice recommendations often labelled as good practices, best practices, promising practices, research-based practices, evidence-based practices and guidelines, amongst others. Advice is often targeted to particular policy domains and/or specific target groups. However, there is often some uncertainty about the quality of supporting evidence for many recommendations, whilst there are no standards that could indicate the type of evidence underpinning the labels attached to particular practices.

Nevertheless, Perkins (2010) offers the following definitions in commenting on advice for children and family services, which help to standardise the way in which people talk about recommended practices:

- **Good practice**: we've done it, we like it, and it feels like we make impact;
- **Promising approaches**: some positive findings, but evaluations are not consistent or rigorous enough to be sure;
- **Research-based practices**: the programme or practice is based on sound theory informed by a growing body of empirical research;
- **Evidence-based practices**: the programme or practice has been rigorously evaluated and has consistently shown to work.

Evidence hierarchies

The US has gone as far as producing legal definitions of such categories, which has caused some unease. Even with the definitions above, there is uncertainty about what evidence would count as good enough to warrant such labelling. This raises the question of what criteria is used to make judgements about the rigour of the evidence base. Study design is often used as a key marker of the strength of evidence. Studies that conduct randomised experiments as well as systematic reviews and meta-analyses are usually seen as the most robust form of evidence whilst case study reports are often seen as the least – creating evidence hierarchies. However, there are a number of problems with rating the quality of research based on study design and subsequent evidence hierarchies.

- evidence hierarchies which place more emphasis on study design marginalise appraisals of how that design was implemented and how it fits with other studies on the same issue;
- they undervalue the strength of evidence produced by well-conducted and cost-effective observation studies which can provide stronger evidence and a more secure basis for practice recommendations than single Randomised Control Trials (RCTs). Even if it is an RCT, it can be dangerously misleading, whilst there are dangers in using meta-analyses when the interventions being studied are complex, variable and/or highly context dependent;
- systematic reviews of evidence about a particular practice or programme can overlook potentially important evidence, thus weakening the value of the synthesis;
- hierarchies based on study design tend to pay insufficient attention to programme theory – i.e. they tend not to unpack the relevant components of the black box in order to model multiple causal linkages and influences and thus gain a better understanding of how a programme works.
- Finally, hierarchies based on study design do not provide a sufficient basis for recommendations. The move from 'quality of evidence' to 'decision to apply' can never be a simple technocratic choice and will usually involve judgement and political considerations (as discussed above).

As such, Petticrew and Roberts (2003) argue that we need to think more in terms of a matrix rather than a hierarchy of evidence, even for seemingly straightforward questions about what works.

Mixed methods

What counts as good evidence will depend on the type of research, policy or practice question that is being answered. As such, Petticrew and Roberts (*ibid*) argue that policymakers and practitioners are interested in at least eight questions. 'What works?' will be only one of many concerns which policymakers will have – others being cost, acceptability and distributional affects, amongst others. RCT designs are inappropriate for answering half of these. The use to which the evidence will be put – whether it will be for generating options, making decisions, promoting on-going learning and development, deciding to continue, stop or scale something up, or developing more innovative ways of working – needs to be factored in as well.

One of the most effective parts of the government Office for Science in the UK is the Foresight Programme, which looks in-depth at emerging technologies and topics such as computer trading in financial markets, global food security and flood defences. Since the global financial crisis, interest in the use of foresight processes to help identify future trends in key policy areas has been renewed in the UK and other countries.

With this in mind, if policymakers and practitioners are interested in how or why a particular approach works, or exploring new innovations, then a broad range of evidence will be useful, including observational case studies, surveys and other qualitative research. If the question is about measuring effectiveness or impact, the key evidence is likely to come from quantitative studies and in particular experimental trials. There is a need to communicate risks and uncertainties with regard to evidence which states what might work and what might not (Sharples, 2013; Lenihan, 2013).

Notable Initiatives

- The Centre for Court Innovation uses multiple methods of analysis. In addition to randomised trials, the centre is committed to learning from other types of research including quasi-experimental designs, qualitative studies and process evaluations.
- The Social Care Institute for Excellence (SCIE), Social Research Unit and Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI) Centre are making good progress with blending different types of evidence – theory, qualitative, quantitative, practitioner knowledge in a way that could provide inclusive, but still reliable overviews of evidence.
- The Washington State Institute for Public Policy, in its efforts to identify what works, calculates the costs and benefits for Washington State and assesses the degree to which a set of policies is likely to affect big-picture state-wide outcomes such as crime or high school graduation rates. The final step involves testing the robustness of the findings, given that considerable uncertainty can exist in any estimate of benefits and costs.

1.5 Communication and brokering

Unidirectional communication

The main barriers to accessing and using research are said to be a shortage of time to engage with research, an overload of information to process and insufficient contextualised information for practice. Several authors have suggested that evidence being taken up does not just depend on the extent to which it is robust, but depends on targeting the right audience at the right time, who is presenting the evidence and how it is presented (not to mention whether it fits into existing assumptions and professional cultures). Communications efforts that are targeted, clear and concise can be powerful. As such, academic papers are more likely to be read if they are practical and sufficiently summarised.

Moreover, Mulgan (2013) suggests evidence needs to be made more quickly and easily available to decision makers. Google, he says, has played its part; global projects like the Cochrane Collaboration and the Campbell Collaboration have too. Initiatives like the National Health Service's (NHS) Evidence Services which provides health evidence in digestible forms to every doctor in the country could be replicated in other fields. Clearing Houses, used in the US and increasingly so in the UK provide information on specific programmes and interventions, rather than focussing on questions or general strategies. Whilst these have an advantage in providing support to applying evidence-based approaches, they often suffer from a shortage of rigorously evaluated interventions which meet the necessary standards of evidence – often resulting in sparse catalogues and accusations of being too exclusive (Sharples, 2013). All these strategies, however, require dedicated time, resources and skills which are not always available or valued.

Interactive communication

However, basic approaches to communication have a number of problems. Packaging and posting research passively is unlikely to have a significant impact on people's behaviour. Given that change is often a result of changes in behaviour amongst several actors and their relationships with one another, communicating research only to policymakers is unlikely to bring about desired changes. High quality evidence needs interpretation for guidance and conversion into meaningful materials. No wonder that research use has emerged as a largely social process, with interaction and relationships being key

factors in determining how evidence gets used and applied in practical settings. Recent research on General Practitioners (GPs) suggests that informal networks or local guidance may be more significant than guidelines from the National Institute for Health and Care Excellence (NICE) in adopting new drugs based on the latest evidence (Thomas et al., 2013). Programme funders in the UK often hire staff for their expertise and networks, from which a substantial amount of evidence is often acquired (Ravenscroft, 2013). Several funders take account of what other funders are doing, often through informal conversations.

Encouraging researchers, practitioners and policymakers to engage with one another can have several benefits:

- Research is often seen as something which is "done to" practice, rather than something which can improve it. The exception is medicine, where communities involved in delivering frontline services are more infused with a research-facing outlook. Collaboration between researchers and users has often been key.
- Practitioners gain a deeper understanding and sense of ownership of the findings. Hence, evidence can be integrated more relevantly and sensitively in professional settings.
- It can force practitioners to think more rigorously.
- It encourages those who plan and implement demonstration projects to be more disciplined about articulating measurable goals and objectives for their work.
- Researchers can benefit as they become grounded in the messy realities of day-to-day implementation, which makes their work more nuanced and their writing easier to read.
- Researchers who work with practitioners are less likely to hold new programmes to unrealistic standards of performance (Berman and Fox, 2013).
- Sustained engagement on the ground can give researchers and institutions credibility in engaging with higher-level policy-makers and processes. Hence, network-based approaches, which support direct engagement and dialogue between researchers and users, are proving to be particularly effective (Nutley et al., 2007).
- Having an opportunity for practitioners to participate in research is likely to create a more research-facing profession (Bell et al., 2010).
- Universities and research centres can help develop evidence-related skills in local authorities or other policymaking bodies.

Close interaction between researchers and users (such as practitioners and policymakers) can take place at various stages of the policy process, particularly the evaluation process, which is another key means of influencing uptake of findings. The structure of such a partnership can affect the way in which evidence can subsequently be used because it significantly helps to determine the design of the programme being examined and the scale at which it is evaluated. Typically, Abdul Latif Jameel Poverty Action Lab (JPAL)'s evaluations begin long before a programme is implemented. Researchers and implementing organisations have an extensive discussion of the underlying problem, and various solutions, along with the associated 'Theory of Change'. New forms of social media are changing the style and speed of interactions between experts, policymakers and the public, and suggest that the messiness of such online discourse is how we create capacity for more coherent exchanges, build trust, learn and digest (Bell, 2013).

Open policymaking provides opportunities to draw more systematically on engagement and dialogue with a range of stakeholders including the wider public. The UK government has been urged to, and have made clear their intention to promote a more open policymaking process. However, does this mean policymaking will be about open doors, welcoming in new perspectives, open minds, reflecting on the limits of central control and predictability, or transparent but closed windows, revealing policy but maintaining control of its contributors (Stilgoe and Burall, 2013).

Finding solutions to cross-cutting policy problems from obesity to ageing, food security and climate change, require insights from a range of disciplines from both social and natural sciences. Government analysts across the board are now working more closely with policymakers in multi-disciplinary teams. However, integrating an appropriate mix of expert advice from all the subject areas

is extremely challenging. Currently, the government's scientific advice tends to be structured by disciplines, reflecting the structure of a university, rather than blending disciplinary expertise and facilitating the pooling of analytical services between departments.

Facilitators and brokers

Nevertheless, researchers, policymakers and practitioners may feel it is not in their interest to engage with one another. Hence, third party brokers or intermediaries might be key in facilitating interactions. In fact, a large amount of evidence for social practice comes through the work of a range of intermediaries – third party brokers who bridge gaps between the creation of research knowledge and its use in practice settings.

Amanda Cooper highlights five characteristics of effective brokers or facilitators of research use – they have:
- an understanding of research methodology;
- a broad overview of the literature:
- a track record within academia and practice;
- sound interpersonal skills;
- an ability to translate and interpret complex information into meaningful materials for users.

Intermediaries are not necessarily unique bodies. They are internal and external (to government) and include the media, think tanks, lobby groups, professional bodies, private companies, local government as well as legislative committees, taskforces and government working groups. Mulgan and Puttick (2013), however, suggest an absence of such intermediary organisations and see this as impeding the effective use of evidence. Exceptions include SCIE and NICE, but there are no equivalents in sectors such as criminal justice or children's services, whilst universities have played only a small part in this intermediary brokerage role to date. In healthcare, medical schools have not only been the drivers of evidence production, but have also been at the heart of training, education and frontline provision. As a result, research and researchers are naturally in a position to be informed by the cues and prompts coming from day to day practice. In criminal justice on the other hand, social scientists are too often seen – unfairly or otherwise – as non-operational commentators who are out of touch with the realities of real world practice and lacking in credibility (Sharples, 2013).

Key players within policymaking bodies, such as advisors, are being urged to play a brokering role. Chief scientific advisors have traditionally been seen as scientists with considerable standing and authority. But Jasanoff (2013) argues no expert or institution can place himself or herself beyond critique. In fact, the authority of research evidence increasingly comes from its exposure to human interaction, review and scrutiny (Barder, 2012). Moreover, the role of scientific advisors needs to be understood within a rich eco-system of expert advice that includes analytical professions within the civil service, external advisory committees, policy tsars, national academies, learned societies, universities, NGOs, civil society organisations and many others. Moreover, as more of the execution of public policy is outsourced to agencies in the private and not-for-profit sectors, government advice is less able to engage with the complexities of delivery and implementation (Doubleday and Wilsdon, 2013).

Mulgan (2013) suggests that the key role of a scientific advisor is to act as an intermediary and broker rather than as an advisor. His/her skills need to be ones of translation, aggregation and synthesis as much as deep expertise. S/he needs to be a conduit of advice rather than a single expert opinion. Scientific advice is more often about helping policymakers to navigate options than it is about providing them with single answers. Pielke suggests that the role of the science advisor is a combination of championing the role of science in government, convening relevant expertise, and helping decision makers navigate wisely between the scope of available choices on a given topic and supporting the quality of evidence and expertise in government by "providing advice on advice".

Independence

Organisations which are most successful in linking the demand for evidence to its supply are those who are trusted beyond reproach (Lenihan, 2013). Independence is hence a key issue. Organisational or individual independence can take many forms, and can be achieved even for in-house or in-governmental bodies. The independence of the IMF's Independent Evaluation Office (IEO) is assisted by strict hiring practices that ensure its members are not influenced by career concerns. The mandate for Australia's PC is established by its own law, ensuring budget independence and placing limitations on situations in which officials can be terminated – protecting them from political manoeuvres. In Germany, Evidence-Based Practice (EBP) institutes which do not sit directly within a government department enable researchers to express their findings more freely.

Budget size is not as important as using resources wisely. Streamlined umbrella or co-ordinating organisations, which use deep and wide academic and private sector research networks, offer cost-effective ways of undertaking research and development work. The Bertelsmann Stiftung Indices are run by small compact teams which pull information and contributions from a very wide network of scholars. Organisations like the Washington State Institute for Public Policy (WSIPP) do the same, maintaining a large academic network and a small dedicated staff base. Multiple sources of funding (ranging from government to foundation grants and private sector contracts) can help intermediary institutions in maintaining their independence. In the case of the WSIPP, which was created specifically by and for government purposes, its budget is biennial and administered through a state university, and it has the ability to take on outside projects in proscribed circumstances.

Notable initiatives

- The Research Excellence Framework (REF) is placing greater emphasis on impact of research on the economy, society, public policy, culture and quality of life, with universities and funding bodies considering how their research could be put to better use (Sharples, 2013).
- JPAL communicates the results of its evaluations in a variety of ways. It makes an effort to make research accessible by extracting the most compelling results from longer papers and reports and presenting them in non-technical language. All evaluations are communicated through a 2-page summary consisting of the relevant policy questions, the programme being evaluated and the results of the evaluation, which are made available on the JPAL website.
- It also organizes outreach conferences to communicate the results of its affiliate's evaluations.
- JPAL also tries to promote active dialogue between researchers and policymakers to identify issues which are considered most pressing for both, by for instance hosting matchmaking conferences.
- It brings the policy perspective into research through the creation of special initiatives in focus areas like the adoption of agricultural technologies or ways to reduce corruption in public programmes. Initiatives are driven by funding organisations, policymakers and JPAL affiliates identifying key areas for future research, and inviting proposals from researchers working with field partners to propose evaluations that attempt to answer these questions.
- In the UK, NICE does a lot of work distilling unprocessed information into practical guidance and supporting materials, disseminating information across the diverse social care sector. Organisations like that can cross disciplinary and professional boundaries and enable synthesis, translation, and mobilisation of evidence.
- The recently disbanded National Police Improvement Agency has provided evidence-based guidance for practice, along with organisations such as the Universities Police Science Institute in Cardiff. Their Sixty Second Briefing on Signal Crimes provides an example of an evidence-based digest for police practitioners, based on a body of underlying research (Sharples, 2013).
- Non-governmental organisations such as the recently formed Education Endowment Foundation are filling some of the gaps in providing accessible guidance materials for educators.
- Project oracle seeks to establish a standard for evidence-informed decision-making on children and young peoples' policy in London. The project's delivery board represents cross-sector

leadership. It aims to match-make small charitable organisations with students and academics from local universities. In parallel, commissioners, philanthropic trusts and other funders and decision makers are engaged to help ensure they can interpret and use the evidence available.

- The project aims to turn commissioners into evidence champions through a fortnightly seminar programme. Content includes evaluation philosophy and techniques, RCTs and quasi-experimental methods and issues in implementation. The commissioners are invited to contribute, explaining their work and evidence needs.

- Berman and Fox (2013), in their review of the Centre for Court Innovation, suggest that working with government decision makers helps to ensure the relevance of the organisation's work. At the same time, however, it has the freedom to think beyond electoral cycles and pursue a long-term vision of justice reform. It is not subject to institutional constraints such as civil service regulations and has the room it needs to issue findings which are less than positive.

- Research and Development Corporation (RAND) has established the promising practices network on children, families and communities, to present evidence-based information on what works to improve outcomes for children and families in the US. The network's main product is the website www.promisingpractices.net – with a user-friendly interface that provides brief, easy to understand information for busy non-researchers. The primary audience for the website is policymakers, practitioners, the media and other decision makers who may not have research training. All content posted on the site must meet pre-established evidence criteria.

- Small teams of the Bertelsmann Foundation (a German think tank) are active in presenting their work and getting it in front of policymakers, as well as making it highly accessible through easy to use websites and downloadable reports.

- The Researcher in Residence programmes and Economic and Social Research Council (ESRC) knowledge exchange opportunities have helped to promote greater interaction between research producers, users and intermediaries.

- New media offers opportunities to gather and share evidence. Examples include:
  o Twitter and RSS feeds to stay in touch with policy and developments;
  o Short films about exemplar projects;
  o Writing regular blog posts rather than using traditional monitoring forms
  o The National Endowment for Science, Technology and the Arts (NESTA) has collaborated with writer Ben Goldacre on a new project called "Randomise Me" launched during the summer of 2013;
  o Crowdfunding and peer lending websites such as Buzzbnk, Kiva and Kickstarter are gaining popularity while crowdsourcing initiatives are taking off such as "Fix My Street" – offering models that might offer new insights about how evidence is generated, used and shared.

1.6 Policymaking and evidence use

Policy processes

Public policies have often been rolled out with little trialling or evaluation. Policymakers and frontline professionals in national and local government are not always aware of what is known globally in the fields they are responsible for. In many cases, important decisions are made by best guess and are overly influenced by politics, marketing and anecdotal evidence. Policymakers' and practitioners' own experience and that of colleagues are likely to have a significant bearing on their day-to-day work. What is done may simply be what has always been done. And new ideas may become fashionable regardless of whether they actually work. They may eventually be found wanting and abandoned only to be rediscovered later in amnesia. Meanwhile, interventions with a strong positive evidence of impact can still remain small scale (Mulgan and Puttick, 2013; Sharples, 2013). In sum, research evidence plays a relatively small part in informing decision-making. For instance, scientific advice to the UK Parliament is poorly understood and systematically overlooked (REF). Indeed, democratically elected politicians have the right to ignore evidence and often do (Mulgan and Puttick, 2013). In some cases, it may even be wise to (as the MRI scandal might testify to).

This is despite relatively large sums of money being invested in the generation of research. When evidence does appear to be used, with the mass of research it becomes tempting for policymakers and shapers to "cherry" pick evidence to back a particular perspective rather than consider the evidence base objectively as a whole. Further, research and other sources of evidence are rarely used in an instrumental way, but in more indirect and subtle ways. Some suggest there is a relative shortage of rigorous evidence about "what works" in what contexts and increasingly at "what cost". Others suggest there is insufficient evidence on the rationale underpinning an approach, evidence of impact, clear implications for practice as well as rich logistical information on implementation (training, materials, costs, management) which presents obstacles in taking evidence and applying it to real world contexts. There have been a number of individual, small-scale approaches which have emerged to find and use evidence. But without a coherent overall infrastructure, there is a danger of duplication and confusion amongst practitioners (Sharples, 2013).

Problems with using evidence

There are significant challenges, however, to using evidence systematically, including: commissioning models, budgeting, planning and organisational cultures that can block and prevent effective programmes from being adopted. There is often a dilemma about the weight and judgement attached to different forms of evidence. There is also uncertainty about whether an intervention that works well in one context will work equally well if applied in another. Evidence of what works elsewhere may trigger the "not invented here" syndrome, which can affect local willingness to act on findings from other areas. It is not surprising that many commentators suggest it is essential to understand why interventions are successful in order to build success factors into future delivery (see for instance, Johnstone, 2013). There may be contradictory advice from different sources adding to the confusion. Given this uncertainty, no wonder that many policymakers and practitioners rely on personal experience and advice from people they consider to be experts in the field. Given that experiential knowledge is often not captured by research, an absence of research evidence does not mean an absence of effectiveness.

Furthermore, evidence-based interventions are not a guarantee of improved outcomes. The environment in which a programme is being introduced must be considered. Are there procedures in place to recruit participants to a programme? Is there a suitable infrastructure in place to support delivery and training? Can the intervention be sustained over time? As ever, evidence-based programmes should not be seen as a panacea but as useful tools in helping professionals apply evidence-based practices.

Whilst evidence relating to particular strategies is useful in terms of guiding practice, it can often fall short in terms of how to apply that evidence in real world contexts. Hence, just as much effort and resources need to be allocated to the way evidence is applied as on what the evidence says. But the application of evidence will never be straightforward: experts are not always right and science can never replace values (Mulgan, 2013). Evidence is never definitive – what works now may not work in the future. What works in one region or nation may not work in another. We need continual challenge and an appetite for rigorous experimentation and evaluation to improve our understanding of policies, programmes and practice. The challenge is to get the right balance between intelligent supply of evidence and sensitivity to the conditions in which it is used.

Drivers of change

Austerity and decentralisation
Nevertheless, with increasing pressure on public resources, the government has committed to a series of wide ranging reforms with a renewed focus on achieving better social outcomes at a reduced cost, by directing money to programmes and services with the best chance of achieving impact. Moreover, decentralisation across many sectors has seen frontline organisations increasingly make strategic decisions about the services they provide – what is delivered, how they deliver it and by whom. With

this freedom has come increased responsibility to make informed choices. In sum, the desire to save money and improve performance, and to a lesser extent, train staff and advance professional standards is driving demand for better evidence and encouraging it to be acted on.

Leadership and culture

Johnstone (2013) suggests a need for organisational leaders to craft a culture that fosters effective use of evidence. This includes the need for leaders to insist that evidence is timely, relevant to the real world, credible and communicated effectively. This may involve protecting research and investment staff development budgets and requiring evidence statements in key policy reports. Leaders can create the conditions for effective use of evidence, setting expectations and promoting action by:

- using planning and decision-making methods which prioritise research and analysis inputs;
- making more of existing data, and ensuring systems and practices capture data which are most needed;
- including unit costs in service delivery;
- using rapid prototyping, with proportionate research and analysis to determine whether and on what scale to invest further;
- bringing data and analysis to life, with high standards of presentation;
- using cost-effective solutions to improve quality of evidence;
- drawing on wider evidence of what works from other areas, translated to fit local needs and circumstances;
- resourcing and embedding evaluation more generally;
- removing or reducing constraints on effective use of evidence by, for instance, ensuring systems and processes are in place to facilitate data sharing.

A strong leader can have a huge impact on organisational culture and effective communication to the public and policymakers. Continuity of leadership in an organisation beyond short political term limits can further institutional independence, allowing them to outlast political shifts and changes in national leadership. Leadership combined with humbleness, and a willingness to listen and learn from a variety of perspectives can improve the rigour of decision-making.

Over reliance on a strong leader can however hurt an institution in the long-term – especially if the good practices which have made them so effective are not institutionalised into the organisation for the long-term. As such Nesta have suggested there should be a 'Red Book for Evidence' which would be published alongside each budget or spending review, and would spell out what evidence was used to support significant spending decisions – in essence an audit trail. The broader aim should be that whatever major policy decisions are taken, there would be recognition of the relevant evidence set out. Some departments already do this – publishing evidence surveys as part of the process of producing green and white papers (Mulgan and Puttick, 2013).

Moreover, some have suggested that external bodies such as Ofsted could play a role in holding practitioners to account for their use of evidence in informing decision-making. However, this could see evidence informed practice being seen as part of a compliance culture, driven by government rather than a positive process led by professionals. Practitioners will often know more about what they are doing than researchers and policymakers. As such, the focus may not be so much on building an evidence base but matching practitioners and/or organisations who do similar work 'on the ground'. Moreover, organisations need some breathing room to innovate.

Notable initiatives

- The UK government has set up a series of 'What Works centres' inspired by the Cabinet Secretary's notion of a 'NICE for social policy'. Their primary role will be to make more of the existing evidence base and promote methodological developments that will strengthen future evaluation.

- Similar trends exist in the US, where Haskins and Baron (2013) suggest that the Obama administration has created the most expansive opportunity for rigorous evidence to influence social policy in the history of the US government by bringing evidence into the decision-making process.
- With local authorities experiencing huge budget cuts, they are increasingly using evidence to: improve targeting and increase value for money; apply lean thinking to drive out unproductive activities; prototype interventions; and engage people to do more to help themselves. Moreover, local authorities are placing greater emphasis on qualitative data and understanding the assets of local communities, not just needs defined by statistics and professional opinion (Johnstone, 2013).

1.7 Sustainability and capacity development

Sharples (2013) suggests that activities that could help in terms of capacity building include: training for professionals and leadership; organisational support for individual research users; embedding research in the daily work of organisations; establishing professional networks; and recognition by professional bodies of the status and importance of engaging with research.

UK Government departments are encouraging analysts to become more directly involved with local authorities as partners, either formally through joint initiatives or through employee volunteering. Skill gaps may however persist and limit the contribution that research and analysis can make. Activity to address skill gaps tends to be *ad hoc* and linked to specific organisational priorities whilst training budgets are typically constrained.

Users of research and other forms of evidence may well need to develop their evidence-related skills depending on their role. For instance, this could include training on interpreting and presenting data, challenging evidence, commissioning research, and judging what is likely to work in local circumstances. Adopting a more evidence-driven approach may mean senior managers such as director-generals or chief executive officers undertake critical self-appraisals, followed by training and mentoring on analysis and use of evidence.

Many organisations focussed on the evaluation of policy interventions do not employ external evaluations of their own work. A number of institutions do demonstrate the value of external evaluations (and self-monitoring) to promote learning and organisational adaptation. Where the budget to undertake independent external evaluations does not exist, other means of evaluation can be used to aid learning. Regular independent external evaluations (like those employed by the Dutch *planbureaus*) are useful not only to organisations themselves, but also to the societies that benefit from their institutional evolution.

Mulgan (2013) suggests that there is need to learn more systematically from existing research and practice, on why certain kinds of knowledge and advice are acted on, and others are not – evidence about evidence. For instance network analysis can inform our understanding of how scientific advice works in practice, and can assist in the construction of more intelligent systems of knowledge exchange (Cleevely, 2013).

2. EVALUATION SYSTEMS IN DEVELOPMENT AGENCIES

This section looks at the experiences of other development agencies with strategic evaluations. It examines the different stages of the evaluation process and how different agencies deal with specific challenges of uptake. Concrete examples of institutional practices which are relevant for the uptake system of the European Commission are put forward.

2.1 Institutional location and reporting lines

Issues

The institutional position within development agencies has largely been "resolved". Some are independent (as with most of the Evaluation Units of multilateral development banks); while some are positioned within HQ (mainly bi-laterals) often in an oversight, strategy or quality control division of the development organisation. Several Evaluation Units link directly to a management group which oversees programming and policy decisions. For instance, the DFID Evaluation Department and the Asian Development Bank (ADB)'s Independent Evaluation Department (IED) serve an oversight function for decentralised evaluations or cover more strategic and thematic topics beyond the scope of individual programme Evaluation Units or country offices.

Several bilateral agencies have split the evaluation function between different implementing agencies. Korea has one main Evaluation Unit in the Korea International Co-operation Agency (KOICA) and another in the Economic Development Co-operation Fund. USAID and Germany have similar set ups though Germany's evaluation department in BMZ has an umbrella function overseeing the main Evaluation Units in the implementing agencies (GTZ and KfW) – more info below. The Agence Française de Développement (AfD) is unique in that its evaluation function is housed with research, providing close links to broader learning agendas.

The process of clearing reports is often covered by evaluation policies. Central Evaluation Units tend to report to a high level, with two thirds of members reporting to the head of the development agency or the ministry (political level). All the multilateral banks report to the executive board, either directly – as in the case for the IMF, IADB, EBRD and the AfDB, or through a sub-committee on evaluation or development effectiveness (as for the World Bank and ADB). AfD reports to the executive board of the agency via an external evaluation committee made up of four persons from within the ministry and four external experts.

Two members – the special Evaluation Office of Belgium and the Netherlands IOB report directly to parliament, though their findings pass through the Minister – no changes can be made to the report before submission to policymakers. Several other Evaluation Units contribute to annual reports to Parliament on development results, including AusAID's Office of Development Effectiveness.

Examples

- **CIDA**: the evaluation directorate sits within the Strategic Policy and Performance Branch. Directorate reports to the President of the AID Agency (who then reports to the minister) who is also the chair of the Evaluation Committee. The directorate has no direct decision-making relations with programme branches.
- **Denmark**: the evaluation department is an independent specialised unit in the Ministry of Foreign Affairs. While it is part of the performance management framework, the evaluation department holds no responsibility for administration, implementation or monitoring of development co-operation. The head reports to the minister of development cooperation through the state secretary.

- **France/AfD**: the Evaluation and Knowledge Development Unit is part of the research department within the strategy directorate of the AfD, the Unit is separate from the operations departments and management. However, an external evaluation committee was being created at the time of writing. The evaluation committee was due to comprise four members from the supervisory ministries and four independent specialists and will be chaired by an independent expert. The Evaluation Unit was due to report to this committee, which will in turn report directly to the AfD board.
- **France/Evaluation Pole DGMDP** (Directorate General for Globalisation Development and Partnership) in the Ministry of Foreign and European Affairs: this is institutionally located next to two other Units – audit and management control and charged with improving effectiveness and performance of development assistance policies.
- **France/Evaluation Unit DG Treasury** – reports to the Assistant Secretary of multilateral affairs, trade and development policies within the Treasury DG. The Head of Unit participates in weekly senior management meetings.
- **Germany**: BMZ evaluation and audit division operates independently from BMZ policy divisions. It is a separate unit within the Directorate 'Central management and NGOs' and the head reports to the State secretary through the Director General.
- **Germany**: GTZ (as it was when the OECD undertook the study) Evaluation Unit was separated from the operational departments and reported directly to the office of the managing directors. The Head of Unit participates in meetings of senior management but not in the highest management team.
- **Sweden**: the Swedish Agency for Development Evaluation (SADEV) is an autonomous agency reporting to the Swedish government and is responsible for evaluating Swedish development assistance. It has decision-making powers and institutional independence.
- **Sweden**: UTV – department for evaluation of SIDA evaluates SIDA development activities. UTV reports to the director general of SIDA who reports to the Swedish government (Ministry for Foreign Affairs).
- **UK/DFID**: The evaluation department, like audit, is a separate unit within DFID. The budget of the department is determined by requesting funding, based on a work plan, from DFID's three year resource allocation, which is allocated from the central budget by the UK government treasury. The head of evaluation can select evaluation topics that are relevant to DFID's development effectiveness and protect the independence of the evaluation department evaluators and evaluation consultants contracted by the department. However, since the change of government in 2010, the evaluation department was split in two and then subordinated under the Research and Evidence division and IACI – the new watchdog institution respectively. A small team in ReD supervise evaluations carried out at the country level.
- **UNDP Evaluation Office**: the Director reports administratively to the Administrator but substantively to the inter-governmental governing body. The Director has full authority over the conduct and content of evaluations and presents them directly to the Executive board.
- **US**: Director of Foreign Assistance (DFA) - the evaluation head reports to the director of US foreign assistance.
- **US/USAID**: at the time of writing, USAID's central evaluation office was being re-established. Although it may have changed, evaluation reported to the Director of Management, Policy, Budget and Performance. The evaluation interest group and agency policy coordinating committee provide support to the evaluation function in USAID.
- **World Bank independent Evaluation Group** (IEG): this is headed up by the DG Evaluation (DGE), which reports to the board of executive directors through the committee on development effectiveness (CODE). Work programmes and budget are prepared independently of WBG management, under the oversight of the DGE for endorsement by CODE and approval by the Board.

2.2 Independence (and quality)

Issues

The OECD DAC 2002 suggests that an evaluation is independent when it is carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention. Independent evaluation implies freedom from political influence and organisational pressure, full access to information and full autonomy to carry out investigations and report findings. Independence must be balanced with the engagement of stakeholders and relationship building to ensure uptake and respect of evaluation findings.

Greater independence has been achieved through improved clarity regarding institutional positions and reporting lines and a better division of labour for decentralised evaluations – with Operational Units taking primary responsibility for internal or self-evaluations at the project level. The institutional position of most members' central Evaluation Units – either as an entirely separate entity or sitting in an audit-review or development effectiveness cluster – helps insulate evaluation from operational departments.

Evaluation manuals, policy and rules on staffing for the Evaluation Unit are used to avoid conflicts of interest. All but a few member agencies use external competitively recruited consultants as a primary means of creating independence. Or they rely on in-house staff with no connections to the operational activities in question. Reference or steering groups charged with quality assurance for the process are used frequently, sometimes incorporating external experts with agency staff to balance ownership or involved stakeholders with independent outsiders.

Most multilaterals have more elaborate checks in place for the head of evaluation – including term limits and a clause of no re-entry into the organisation from evaluation management. Most members have also created a substantial degree of independence in the reporting system to protect against the blocking of reports by mandating the Evaluation Unit to have full responsibility and publication rights to final reports without interference from management. Ethical codes of conduct further protect the evaluation process from interference.

Examples

- **WB IEG** adheres to independence in four areas: organisational independence, behavioural independence of evaluators and evaluation managers, protection from external influence and avoidance of conflicts of interest.
- **Sweden** created **SADEV** as a separate evaluation entity with a wide mandate to reinforce its independence.
- **UK DFID**'s independent advisory committee on development impact (succeeded by the ICAI) was an example of separate entity overseeing evaluation and development effectiveness.
- **UNDP evaluation office:** the director issues evaluation reports without clearance from UNDP management. The director is appointed by the administrator in consultation with the executive board, and is allowed to serve only two terms, without re-entry into the organisation.

2.3 Main functions

Issues

Most central Evaluation Units are mandated to evaluate at the strategic, sector or thematic level – assessing policies, aid instruments, and contributions to broader development effectiveness. They backstop programme or project level evaluations. Evaluation Units tend to play more of an accountability role, while also contributing to overall organisational learning and strategy. All

multilateral banks have independent evaluation departments with mandates to address issues of organisational effectiveness and impact. However, links between decentralised evaluations at the project or country level and the main central Evaluation Unit at HQ appear weak or ill-defined in many agencies. Most central Units do not have sufficient information on evaluations being carried out at the field level.

- In **Norway**, evaluation staff can comment and provide feedback on terms of reference drafted by country offices.
- **BMZ**'s Evaluation Unit is mandated to evaluate the effectiveness of the entire development assistance programme across several government departments.
- In the Netherlands, the evaluation department is charged with assessing not only official development assistance but also the impacts of other policies of the Ministry of Foreign Affairs.
- **SIDA** has a narrower mandate to assess the contributions of a single agency or ministry within the wider development cooperation framework.

2.4 Evaluation policies

70% of OECD-DAC member agencies have a single policy document to guide the work of the central Evaluation Unit and define the place of evaluation within the agency, through the form and legal status of these policies varies. In 2010, nine agencies were in the midst of approving, revising or updating evaluation policies. Aid effectiveness principles of harmonisation and alignment are increasingly present in evaluation policies. E.g. DFID's 2009 evaluation policy and SIDA's 2005 policy both emphasise the need to support partner country evaluation capacities and the importance of a collaborative approach to evaluation. Half of OECD DAC member agencies evaluation policies address the role of partner countries in evaluation, provide a mandate for joint evaluation and give Evaluation Units a role in supporting evaluation capacity development in partner countries.

2.5 Financial resources for evaluation

Information on budget allocations for the evaluation function is extremely difficult to collect and comparison is very challenging. However, data collected by OECD DAC suggests that development agencies and multilateral banks spend an average of US$ 5.1 million a year. Larger donors spend more in absolute terms. The World Bank Group spends the most – the IEG has a budget of US$31 million. The average evaluation department produces about 24 reports per year. But there is wide variation – IMF's IEO produces two evaluations per year whilst JICA produces 150 a year. Germany's three development Evaluation Units together produce 89 evaluations per year.

The average budget for a bilateral agency's central Evaluation Unit is US$ 2.4 million. Most Units have stagnated or increased only negligibly in recent years. The average ratio of the central Evaluation Unit budget is 0.16 per cent of ODA – evaluation spending correlates quite closely with overall size and resources of the agency or institution. Resources of the seven member multilateral development institutions are significantly higher on average than the bilateral agencies, and evaluation budgets are therefore larger on the whole. The average multilateral evaluation department has a budget of US$9.7 million and the AfDB has the smallest at US$ 4.5 million. The average evaluation cost as a proportion of the overall administration budget is 1.4%. The IMF spends the least with 0.5%, while the ADB spends the most at 2%. Regarding the overall spending on evaluation, ten respondents answered this question; the average member agency spends the equivalent of 0.47% of their aid budget on evaluation – with Austria spending 1.1% whilst AfD spends 0.52%.

The overarching trend is towards increasing the evaluation workload which has not been matched with a corresponding increase in resources for central Evaluation Units. In terms of determining evaluation budgets the trend is towards budgeting to meet the output objectives of the Units (based on a work plan with cost estimates, determined as part of the overall agency budget).

2.6 Planning and stakeholder involvement

- **CIDA**: the evaluation committee annually reviews a strategic risk based five year rolling plan. This plan is approved by the president of the agency each year. The plan is not publicly available but it is shared with other agencies via the DAC secretariat. Partner countries are involved when deciding to undertake a programme evaluation. Local consultants are hired in most programme evaluations as sector specialists or regarding cross-cutting issues such as gender equality and the environment.
- **Denmark:** the Evaluation Unit operates a two year evaluation programme developed through a complex consultation process involving a public meeting, presentation of a draft to the embassies and departments for comments and then presentation of the programme to the DANIDA board, which can make comments. Finally the minister sends the plan to the finance and foreign affairs committees of the parliament for additional comments. DANIDA frequently uses a participatory approach for evaluations of development cooperation. Approximately 50% of evaluations are conducted jointly with evaluation offices of other development agencies and the partner country is involved in about 80% of evaluations.
- **France/AfD:** the Evaluation Unit in AfD has a three-year strategic plan covering all their activities. Specific topics for strategic evaluations are selected annually and these plans are shared with the DAC secretariat. Depending on the type of evaluation and the partners, contact is usually made with local agencies and developing country partners to inform them that an evaluation is in progress.
- **France/Evaluation Pole, DGMDP:** a yearly evaluation programme is formulated by the head of the evaluation pole in cooperation with the field and central departments and is validated by an evaluation committee presided over by the Director-General of DGMDP, AfD and the Ministry of Economy participate in this committee. This is connected to a local steering committee which is comprised of French embassy staff and partner country stakeholders.
- **France/Evaluation Unit DG treasury:** 3 year evaluation plans are decided based on interviews with the operational Units at HQ and field Units. The Head of the Unit formulates initial plans which are then agreed with programme divisions. The proposal is then presented to the treasury and the economic policy director-general who approves the plan.
- **Germany: BMZ** evaluation division has a two-year work programme with some room for flexibility to accommodate joint evaluations or additional demand from the BMZ political level. The planning process follows a bottom up process in which every division is consulted. BMZ-E collects requests and formulates a proposal (including its own suggestions) which is presented to the state secretary who decides on the final programme. So far the evaluation programme is demand driven.
- **Germany:** the **KfW** evaluation department determines the evaluation agenda according to a set of rules for a multi-year plan, while the formal work programme is agreed on an annual basis. A stratified random sample of all finished programmes and projects is evaluated each year. There is a rolling planning for thematic and cross-cutting issues.
- **Germany: GTZ (now GIZ)** – projects to be evaluated are selected on the basis of a random sample of concluded projects within two sectors per year. Projects with a high potential for learning are added. The sectors are chosen on the basis of the evaluation interest of senior management within GTZ and agreed with BMZ.
- **Sweden: SADEV's** evaluation programme is determined by its DG on the basis of broad dialogue and suggestions from staff and interested parties both nationally and internationally. SADEV can focus on sector-wide, thematic, policy and impact evaluations, as well as on project activities.
- **Sweden: UTV** – when setting its programme UTV dialogues with all SIDA teams on possible evaluation themes. The strategic management group of SIDA may also suggest topics. UTV also discusses with international organisations and agencies to find interesting subjects for joint evaluations. UTV compiles these ideas and decides which should be put in the annual plan. The

plan is finally approved by the director-general of SIDA, along with the department's budget. Sweden//UTV are embracing utilisation focussed evaluation - "the evaluation process begins by having conversations with all Operational Units in SIDA, as part of the annual planning processes – asking what are their knowledge needs, what would they like to know and how can evaluation help to answer that. This generates a list of around 100 ideas for evaluations, from which the evaluation department choose 15 they will carry out. The ones which are not chosen to be carried out are subject to decentralised evaluations, for which the evaluation department give feedback and advice. The staff who proposed an evaluation which was chosen form the reference group for this evaluation. They must list intended use and users. If there are not enough users then it is dropped. Otherwise these intended users are then involved in drafting the ToRs, and work with the reference group throughout" (from Nordic Consulting, 2010). The evaluation is framed as a learning process, for example, bringing people together to discuss issues and comment on drafts. Preliminary findings are discussed with users in order to decide on recommendations (although evaluators keep the right to make their own).

- **UK/DFID:** the Independent Advisory Committee on Development Impact (IACDI) approves the evaluation work programme which is formulated annually for rolling two year periods. The programme is open to public consultation via the DFID website and the department also invites comments from a range of external partners – parliament, NGOs, independent experts. They expect to evaluate between 50 and 75% of total bilateral expenditure by the end of a five year period.
- **UNDP Evaluation Office (EO):** their two year work programme is developed by the EO based on consultation with UNDP regional policy and practice bureaus, UNDP management and UN sister agencies. The programme is presented to the executive board for discussion and approval in the annual report on evaluation prepared by the EO.
- **USAID**: the ideas for what kind of policy, sector, thematic evaluations to do come from senior staff members and administrative staff at HQ and at the Evaluation Unit.
- **WB IEG**: works from a rolling three year plan covering all evaluations and related activities including evaluation capacity development and communications. Work programme is prepared by IEG and endorsed by the Committee on Development Effectiveness (CODE) and the full board.
- **NORAD**: they broadly consult with key stakeholders in developing their evaluation programme to ensure there is interest in the topics to be covered (Grasso et al., 2013).

2.7 Types of evaluation

Single project level evaluations have largely been decentralised and incorporated into performance or results based management schemes in Programme Units, allowing central Units to take a broader look at development and agency effectiveness – and towards more strategic, thematic, sector-wide or country level evaluations. There has been some shift in moving towards using evaluation before and during implementation of projects (not just doing evaluation after projects have been completed). The World Bank is doing more 'real time' evaluations geared towards improving on-going development programmes and policies. Various evaluation types are now considered and the most suitable evaluation focus selected based on policy questions or accountability needs. E.g. Finland's policy provides a good example where evaluation approaches are adapted to the setting and programmes at hand in order to maximise the usefulness of evaluation.

The World Bank has initiated the Development Impact Evaluation (DIME) initiative, focussing on impact evaluation organised around thematic clusters which are coordinated across countries in different regions of the world. CIDA has a performance review system including all functions and instruments used by managers and staff of CIDA to assess agency development and operational results. This includes evaluation, internal audit, monitoring of policies, programmes, projects, institutions and operations of the Agency. Performance review provides work tools which enable CIDA's managers to better exercise their roles in achieving expected results. It involves several parts of the agency.

2.8 Communication of evaluation findings

OECD standards on communication issues

They are pretty comprehensive and include the following excerpts "conclusions, recommendations and lessons are clear, relevant, targeted and actionable so that the evaluation can be used to achieve its intended learning and accountability objective and that the evaluation is delivered in time to ensure optimal use of the results". Systematic dissemination, storage and management of the evaluation report is ensured to provide easy access to all development partners, to reach target audiences, and to maximise the learning benefits of the evaluations. Evaluations' findings should be presented "in an accessible format" with findings systematically distributed internally and externally for learning and follow up actions and to ensure transparency. In light of lessons emerging from the evaluation, additional interested parties in the wider development community are identified and targeted to maximise the use of relevant findings. Relevance and utility come up several times as factors which determine whether or not a given evaluation is worthwhile.

Issues

Donors are preoccupied with two challenges – using evaluation to measure attribution (causality between the action and an observed change in attitude or behaviour) and long-term effects, and how to develop the most coherent and appropriate evaluation policy for short-term awareness raising and learning activities outsourced to NGOs and grassroots organisations. There is a wide divergence among agencies on how communication is mandated, resourced, staffed and implemented. Communication departments which are larger, better resourced and more integrated within their respective organisations have stronger leverage – in influencing the institutional agenda-setting process, in accessing the kind of results information needed from programme colleagues and commissioning the kind of market research they need to develop targeted communication to donor publics.

Communicators in OECD aid agencies perform two main sets of overlapping functions – communicating about results and communicating for results. The former is often seen as external communication or corporate communication, whilst the latter is seen as programme communication and used as a tool for internal learning and engagement of stakeholders. There has been increased investment by donor agencies in building communication capacities (to publics i.e. communication about results). Anecdotal evidence indicates that aid agency communicators are beginning to collaborate more meaningfully with their evaluation counterparts.

Aid agency communication units usually have their own budgets for communication about results. Activities related to communicating for results tend to be managed by implementing agencies. In larger donors or implementation agencies, communication for results programmes are managed by project formulation and implementation departments and programme and advisory staff, sometimes on an *ad-hoc* basis. Field officers are also responsible for these tasks as well as regional or geographical offices of implementation agencies. Communicating for results tends to be financed from the core programme or project budget.

There are considerable disincentives within agencies to reporting bad results. Communication units seem to function primarily to put a positive spin on development work and are reluctant to tell it like it is as their job is to protect senior government officials and politicians. Open discussions about evaluation results can be tense, especially when there is weak mutual trust. Tensions and fears about negative results are commonplace when future funding depends on the success of the activity. But this might be alleviated if the evaluator, practitioner and the donor agree on the purpose of the evaluation, the criteria and the indicators from the very start.

If evaluators and communicators are to learn from each other, trust and build stronger collaboration, they will have to arrive at a shared understanding of the necessity as well as feasibility of communicating results deemed negative or sensitive. Efforts to deepen partner country involvement in evaluation, communicating evaluation results or indeed the wider results agenda have been slow in taking off. Dissemination reflects the overall purpose and differing types of evaluation. About two thirds of member evaluation policies defined agency learning as top or one of the top goals for evaluation departments. Staff and management at donor HQ and in the field are the best covered stakeholder group in the distribution of evaluation findings – through dissemination of reports, workshops, sending of summaries and other means. Units which did a lot of joint work with partner countries were more likely to systematically distribute evaluation findings to stakeholders in partner countries. Over half of member units send evaluation findings to stakeholders in partner countries and in donor countries. Only seven members said they include civil society groups in partner countries in their follow-up efforts. Some distributed evaluation findings to their own parliament, legislative bodies in partner countries and intended beneficiaries of their agency's development activities in partner countries. Even when evaluation findings are disseminated to or in partner countries, this is done in a top down way. Only a handful of development agencies have communication strategies, evaluation guidelines which specify communication, or guidelines for publishing results. Many agencies have no institutional policies or guidelines for communicating evaluation results.

Examples

- **Denmark:** the evaluation department contributes actively to the dissemination of DANIDA's own as well as other organisations' evaluation experience via workshops and seminars for staff in co-operation with the ministry's education section. Furthermore, the department assists DANIDA's centre for competence development in the dissemination of evaluation results and contributes to the incorporation of evaluation results in policies, strategies and guidelines.
- **France/AfD:** the budget for evaluation covers dissemination activities, including meetings or conference organisation and publication of knowledge products. Evaluation findings are shared with intended beneficiaries of French assistance and since 2007 with the general public, media, and civil society in France.
- **France/Evaluation Pole DGMDP**: evaluation conclusions and recommendations are made public, first through a meeting with involved stakeholders and then in electronic format on the French Diplomatic website. Since 2001 a review of evaluations is produced and widely distributed, presenting a summary of the most significant evaluations completed.
- **France/Evaluation Unit DG treasury:** when an evaluation is completed, English summaries and French full text reports are published on the unit's website, along with quality check list reviews. A workshop to internalise findings and a workshop in the partner country may be organised. A summary on key findings is sent to the director-general of AfD. Other relevant officials in the administration receive a copy of this note. An annual report on evaluation outcomes is sent to the parliament.
- **SADEV/Sweden**: they are required to distribute information on evaluation findings to parliament, government and other stakeholders in Sweden and internationally. Upon completion, all evaluations and reviews commissioned by SIDA are posted on the website. Seminars are also arranged for all evaluations. These are open for all SIDA staff and other development stakeholders who have an interest in the particular evaluation are also invited. On occasion, press releases are distributed to Swedish media. Reports are also shared with SIDA staff through newsletters, distributed to agency management staff in field offices, and included in SIDA annual reports.
- **UK/DFID**: their evaluation policy states that learning should be shared for wider benefit, encouraging citizens of developing countries to use evidence when appropriate for accountability purposes. It also states that DFID must professionalise the communication of its evaluation findings to increase the traction of recommendations and ensure accessibility to a wider audiences, and tracking systems in support of this aim.

- **UNDP EO**: All evaluation reports are made public and the EO also produces evaluation e-news for broad circulation and sends copies of evaluation reports to all programme units.
- **USAID**: reports are made available on the web, via electronic distribution, print media, and on a dedicated development experience 'clearing house'. USAID has a knowledge management programme – based in the Management Bureau and a Knowledge Management Reference Group is made up of representatives from USAID bureaus – it provides the knowledge management services – which aims to move USAID from a 'need to know' to a 'need to share' culture and has a library, for staff and partners, which is also open to the public.
- **WB IEG**: IEG reports are submitted directly by the DGE to the committee on development effectiveness and are disclosed to the public in line with its disclosure policy, which is being revised. Reports are posted on IEG websites, and many are published in book form. Results are disseminated through conferences and workshops, papers in journals, professional meetings and media releases and interviews. IEG designs and implements sophisticated communication strategies that involve interactive engagement with different stakeholders using a variety of tools, including social media. But few Evaluation Units and departments are as advanced in their communication efforts as IEG.
- **NORAD** has an established system for disseminating results of its formal reports through seminars and other mechanisms. The agency has had its evaluation reports posted to a range of websites which focus on issues of development and/or evaluation and research (Grasso et al., 2013).

2.9 Management response and follow up

Issues

Twenty members reported the agency had a management response mechanism in place. Such mechanisms most often consist of a written, formal response from the department or programme concerned and agreement on follow-up action on the recommendations made by evaluators. The degree of sophistication varies substantially and this seems to be an area of concern among a number of agencies. Some Evaluation Units are charged with tracking implementation of actions responding to evaluation recommendations. Some agencies reported they make management responses public. It seems that follow up actions are made public less frequently than management responses. Many suggested that the system doesn't work very well. Management response practices tend to be more developed than follow-up action monitoring or reporting.

Examples

- **Norway** is a good example of one of the more developed systems. Within six weeks of an evaluation report being completed, an official response from the Secretary General of the Ministry of Foreign Affairs is submitted to the relevant programme area, the NORAD Director-General and the Evaluation Department. One year later, the programme area concerned submits a report to the Secretary General describing actions taken in response to the evaluation. Evaluation staff strives to ensure programme staff buy-in, while protecting the independence of the consultants.
- **The ADB** and several other members systematically publish management responses along with evaluation reports on their websites, to ensure transparency and encourage use of evaluation findings.
- **France/Evaluation Pole DGMDP:** a monitoring table is used to verify annually the status of the implementation of evaluation recommendations or any difficulties encountered. The Evaluation Unit participates in the review of new projects to ensure that prior evaluation findings are taken into account at the planning stage.
- **CIDA:** all programme evaluations presented to the Evaluation Committee require a management response stating how recommendations will be implemented. The agency commonly uses evaluations as learning and management tools, particularly during considerations of new

submissions or formulation of new policies. A follow up process is in place to periodically monitor the extent to which the evaluated programmes are implementing recommendations.

- **Denmark:** at the conclusion of en evaluation, a follow up memo is prepared, taking note of DANIDA's management position on the conclusions and recommendations. The follow-up memo is discussed in the programme committee. Based on the discussion a short management response in Danish is prepared and approved by the Minister for Development Cooperation. The evaluation department undertakes to monitor the implementation of the follow up activities at regular intervals.
- **France AfD**: there was no formal management response or follow up system at time of writing.
- **Germany: KfW** Evaluation Unit has a number of instruments to support institutional learning, including discussion of the evaluation results with all stakeholders, dissemination of general conclusions, and learning by exchanging roles. Reports and presentations are used to inform agency staff and management, as well as the general public about evaluation results. There is no institutionalised process to enforce management responses to each and every evaluation.
- **Germany GTZ/GIZ:** a management response system was piloted and trialled in 2010 and 2011 and has been mandatory since 2012.
- **Sweden/SADEV**; a system for management response and follow-up is in place to respond to SADEV evaluation findings. Management responses to evaluations are included in SADEV's annual report.
- **Sweden/UTV**: a new system for management response and follow-up is operational. It responds to evaluations commissioned by UTV, reviews commissioned by SIDA departments and SADEV's evaluations. However the system has been found to come up short if it is expected to be the tool to ensure lesson learning. They are now looking at having management response workshops. Some management responses are available on SIDA's publication database online.
- **UK/DFID:** in DFID, evaluations have three functions: (i) accountability – implying symbolic focus on value for money, communication aimed at parliamentarians and taxpayers; (ii) programme improvement – implying focus on utilisation and communication with Programme Units/management and peer agencies and; (iii) aid effectiveness – implying focus on lessons learned, knowledge management.
- **DFID** assigns a lead director for each major evaluation to ensure that there is a management response to findings and recommendations and that those recommendations which are accepted are followed-up on. Follow-up of evaluations is tracked centrally by the evaluation department, verified by DFID's internal audit department, overseen by the investment committee and reported in its annual report. The evaluation department reports annually to the management board and the IACDI on evaluation at DFID. The annual report may include an assessment of the number and quality of central and decentralised evaluation processes and studies, follow-up to evaluation findings and recommendations as well as overall lessons learnt for DFID. The head of evaluation participates in key DFID decision-making committees and for a, such as reviewing new policies and country programmes, to help ensure evaluation results and recommendations are considered in DFID's major decision-making processes. All evaluations are published with the management response.
- **UNDP EO**: the UNDP management group reviews and discusses all management responses to EO evaluations – except the assessment of development results, which are country level programme evaluations (the management group is UNDP's most senior decision-making body, chaired by the administrator and includes the associate administrator and the directors of regional and central bureaus). The EO maintains a tracking system of management responses.
- **USAID**: a management response system is in place.
- **WB IEG**: the DGE is invited to some of the senior management meetings. The Management Action Record (MAR) in the bank and Management Action Tracking Record (MATR) in IFC and MIGA are the primary mechanisms for keeping track of the extent to which management has responded to IEG recommendations. Management responses are included in published reports, including print and online, but the MAR and MATRs are not public.

2.10 Link between RBM and evaluation

Results Based Management (RBM) systems which link with evaluation systems in many ways are much more soundly established at the project and programmes levels in most agencies. A clear definition of the complementary roles of central Evaluation Units and the wider RBM system has been established in most agencies where RBM is in place. This has led to a closer linking of evaluation with RBM and monitoring at the project level and a focus on wider strategic questions of development impact and effectiveness at the level of central evaluation departments.

2.11 Summary of lessons from other donor agencies

Key factors		Lessons and/or good practice from other donor agencies
The nature of the evaluation process	Usability of the results	All donors stress that the evidence from evaluations must be useable by the intended audience: DFID's evaluation policy (2013) states that evaluations should be designed to be of use to those affected directly and indirectly by the recommendations in terms of *timeliness* in the policy and programming cycle to inform decision-making, *relevance* in ensuring that evaluations ask questions to which users are keen to know the answers, and *quality* in that evidence must be credible and reliable and thus a legitimate basis for decision-making. This requires ownership by the relevant users in selecting the evaluation topic, designing the evaluation questions and responding to the findings & recommendations. • SIDA has an Evidence Group which takes the evaluation findings and formulates more specific targeted recommendations. These form a proposal which goes to the Director General for approval. • DFID has sought to ensure evaluation is built into programmes and policies at the outset.
	Quality of the evaluations	There are on-going and very robust debates among OECD-DAC network members; involving donor agencies and evaluators (e.g. 3ie, JPAL, IPA, individual consultants/contractors) around how to ensure quality of evaluations and the role of qualitative, quantitative and mixed methods approaches. There are no clear leaders in methodological development, though: • 3ie leads the development of mixed methods approaches (supported by DFID and two major philanthropists. • EuropeAid led the methodological development for Budget Support (BS) evaluations. DFID, SIDA, IFAD, WB IEG provide a high degree of methodological flexibility. The DFID Evaluation Unit focuses on tracking evaluation activity across the organisation and ensuring high standards.
	Independence of the evaluations	Greater independence has been achieved through improved clarity regarding institutional position and reporting lines and a division of labour for decentralised evaluations, with Operational Units taking responsibility for internal project-level evaluations. The institutional position of most central Evaluation Units helps insulate evaluation from operational departments: either positioning them in audit-review or development effectiveness clusters. • DFID: evaluation policy (2013) places a premium on independent evaluation rather than self-evaluation. ICAI is a separate entity overseeing evaluations and development effectiveness through independent assessments. • Sweden: SADEV created as separate evaluation entity with a wide mandate to reinforce its independence. • UNDP: director issues evaluation reports without clearance from UNDP management, to retain independence. • WB IEG stresses independence in four areas: organisational independence, behavioural independence of evaluators & evaluation managers, protection from external influence and avoidance of conflicts of interest. • Some donors such as GIZ, although seeing independence as important have prioritised utility ahead of independence Reference or steering groups are used frequently, incorporating external experts or involved stakeholders.
	Planning and stakeholder involvement	Agencies have different timeframes for their rolling evaluation plans: UNDP & France/DGDP (1 year); Denmark, BMZ & DFID (2 years), France/AfD & WB/IEG (3 years), CIDA (5 years). • SIDA has an extensive consultation process with all Operational Units as part of the annual planning process. This generates around 100 ideas for evaluation; Evaluation Unit chooses 15 to focus on, with intended users helping to draft the ToRs. Proposals which are not chosen are subject to decentralised evaluations where the Evaluation Unit provides feedback and advice.

Key factors		Lessons and/or good practice from other donor agencies
The nature of the evaluation process (cont'd)		• DFID (2013) looks for a stronger role for partner countries in fostering demand, building local capacity. Several donors engage potential users throughout the evaluation process through a Reference Group.
	Types of evaluation	Some shift towards using evaluations before and during project implementation (e.g. World Bank). Project-level evaluations are decentralised and incorporated into performance or results-based management schemes in Programme Units, allowing central Units to take a more strategic approach. • World Bank's DIME organises evaluations around thematic clusters which are co-ordinated across countries. • DPME (South Africa) offers a menu of 8 different types of evaluation to different Departments (see Annex). • IFAD's Independent Office for Evaluation conducts evaluations on loans, country programmes, projects and other non-project activities (e.g. policy dialogue) as well as corporate level evaluations focussing on a specific issue or theme.
	Communication of evaluation findings	Wide divergence between agencies on how communication is mandated, resourced, staffed and implemented. Communication serves two functions: communicating *about* results (external/corporate communication) and communicating *for* results (communication as a learning tool). • 2/3 of OECD-DAC members define agency learning as top or one of the top goals for Evaluation Units. • DFID evaluation policy does not specify how much to spend on communications, but states (paragraph 107, p17) that "the reporting requirements should be identified at the design stage and a reporting plan produced". However only a handful of agencies have communication strategies or requirements/guidelines for communicating evaluation results. In many agencies, including DFID, SIDA, GIZ, IFAD, the Evaluation Unit takes responsibility for ensuring that the findings and recommendations are brought up for consideration during discussions. GIZ brings in a facilitator to chair a learning café where internal managers discuss findings and formulate specific actions for which they are held to account The World Bank communicates its findings through a number of channels including high level meetings at HQ and/or in the field, posting evaluations online, using social media and Google ad campaigns, newsletters, an evaluation week and brown bag lunches. DFID channels evaluation findings through advisory networks, an evaluation digest, an internal community of practice for evaluations across DFID, and an evaluation café.
	Institutional location and reporting lines	Institutional positions vary from independent (multilateral development banks) to positioned within HQ (mainly bilateral). Denmark: evaluation department is an independent specialised unit within the Ministry of Foreign Affairs. • In the UK, the Independent Commission for Aid Impact (ICAI) has been set up scrutinise UK Aid and reports to Parliament through the House of Commons International Development Committee. • Germany: BMZ evaluation and audit division operates independently from BMZ policy divisions. • The Independent Office for Evaluation at IFAD reports to the executive Board of iFAD • Belgium & Netherlands: Units report directly to Ministers. • World Bank: IEG reports to the Board through the Committee on Development Effectiveness. All tend to report to a high level within the ministry, some to the executive board, some to the political level.
	Main functions of the Evaluation Unit	Most central Evaluation Units are mandated with evaluating at strategic, sector or thematic levels, assessing policies, aid instruments and contributions to broader development effectiveness. Tend to play more of an accountability role though do contribute to learning and strategy development. Multilateral banks have independent evaluations departments with mandates

Key factors		Lessons and/or good practice from other donor agencies
Organisational characteristics		to address organisational effectiveness & impact. • BMZ: the Unit evaluates development assistance programme across several government departments. • Norway: evaluation staff comment on ToRs drafted by country offices. • Netherlands: the Unit also assesses impacts of other policies e.g. Ministry of Foreign Affairs. • GIZ: there are two parts to the M&E group – an evaluation group which is responsible for undertaking strategic evaluations and another group to help Operational Units implement their own M&E systems. • The Evaluation Unit within DFID has ceased being a manager of strategic evaluations, which has been devolved to Operational Units and instead sets and ensures high standards. Units backstop project or programme evaluations, however links between decentralised evaluations and HQ are generally weak or ill-defined with many Units not having sufficient information on evaluations being carried out at field level.
	Link between RBM and evaluation	Where RBM systems are in place clear roles have generally been established with complementary evaluation processes. RBM is closely linked to monitoring at project & programme level, leaving the Evaluation Unit to focus on wider strategic questions of development impact & effectiveness.
Institutional incentives	Evaluation policies	Seventy per cent of OECD DAC member agencies have a single policy document in place to guide the work of the Evaluation Unit; half cover partner country involvement in evaluations. Aid effectiveness & harmonisation principles are increasingly present in these policies. • DFID (2013) sets evaluation in a 'spectrum of tools for enquiry', ranging from audit to research. • SIDA (2005) and DFID (2013) emphasises need to support country-level evaluation capabilities.
	Resource allocation for evaluations	Information on budgets for evaluation is very difficult to collect. Development agencies & multilateral banks spend an average of US$5.1m/year on evaluations and budget US$2.4m for the central Evaluation Unit. The number of evaluations produced per year varies from 2 (IMF IEO) to 89 (Germany) and 150 JICA).
	Management response	Most OECD DAC members have a formal management response system in place, but the degree of sophistication varies and many agencies consider the process does not work well. Management response practices tend to be better developed than follow-up action monitoring or reporting.
Contextual factors	External politics	Continuously changing or conflicting political priorities and agendas can both hinder and stimulate learning depending on how they are addressed. This affects all donor organisations to varying degrees. Most member states are under pressure to cut aid budgets to varying extents – these pressures seem to be particularly strong in the UK.
	Wider debates	All donor agencies are involved in the same global debates about evaluation, often through OECD-DAC. The debate around qualitative / quantitative / mixed methods approaches to evaluations and the role of impact evaluations (particularly RCTs) continues, with involvement from bilateral and multilateral agencies and global philanthropic organisations. There are no notable 'factions' around evaluation, except for methodologies for evaluating BS (see process tracking study), which appears to be drawing to a conclusion.
	External pressures	The Commission differs from other donor organisations in that it needs to co-ordinate with different Member States' views on evaluations – notably around BS where the politics of the aid instrument itself influenced the methodological development process.

Key factors		Lessons and/or good practice from other donor agencies
	Internal context	Organisational culture (hierarchical or flat, silo-ised or cross-boundary, single or multi-disciplinary) affects how evaluation evidence is shared. • France/AfD houses evaluation with research, within the Strategy Directorate, providing close links to broader learning agendas.

3. INSTITUTIONAL CASE STUDIES

Annex three looks into seven concrete case studies of how development agencies deal with specific dimensions of the evaluation process. The case studies were selected as follows: based on a first reading of a broad sample of development agencies with interesting evaluation practices a selection was made with the support of the reference group. Specific attention was paid to different dimensions of the evaluation system that are particularly relevant to the uptake challenges of the EC (e.g. participation, dissemination systems, management follow-up mechanisms, etc.). The review showed that these agencies face many of the same challenges the EU is facing today. Nevertheless, a number of interesting lessons can be drawn from these case studies.

3.1 Uptake of evaluation evidence: insights from GIZ

This is based on work currently being undertaken by Sabine Dinges, University of Bradford and GIZ. Used with her permission.

GIZ is one of the first organisations which institutionalised a systematic process to learn from its peace building evaluations at the institutional level. In 2010/2011, 8 independent evaluations took place in GIZ, the findings of which were summarised in a synthesis report. The recommendations in the synthesis report were discussed at the operational level with members from different departments in a 'learning café'. They formulated implementation agreements in a Management Response System (MRS) matrix which was then discussed with the respective management within each department. Finally the outcomes of these discussions were presented to GIZ's Executive Management Committee and Regular Operational Meeting where the formal decision on implementation was taken – see figure below.

Figure 3: Overview of discussion and decision-making process at the institutional level at GIZ

Characteristics of the recommendations

Recommendations that are easy to operationalize and implement were more likely to be adopted than the ones that were more difficult to put into practice – e.g. those that required inter-organisational collaboration. Within GIZ's MRS, implementation of agreements is followed up after a year – so it is rational to commit to agreements that are more easily achievable. Instead of discussing issues that sectoral departments could deal with on their own, they used the opportunity created by the evaluation to discuss issues that were perceived to be urgent and of high perceived institutional relevance, but often neglected. But recommendations that are not specific are less likely to be taken up especially if they require cooperation beyond a specific competence or thematic area.

Evaluations rarely produce new knowledge – the news value seems not be among the most decisive factors in the discussion process. But unambiguous evidence and clear recommendations are preferred to more cautious statements which stress context dependence. During the discussion stemming from the evaluation it is the stakeholders involved in the discussion who set the agenda, not the evaluation. As a result, people often discuss the strains they experience in their everyday work, not what has emerged from the evaluation.

Organisational factors

The organisation enables the strong influence of individuals and defines an internal division of labour and therefore roles and responsibilities in the implementation of agreements. Although personal interests, agenda and opinion can influence which recommendations are adopted and in which way, inter-departmental exchange during the learning café and through documents balanced this out to some extent.

Implementation agreements (i.e. what is agreed to implement) reflect the organisational structure: this means that recommendations that were not relevant to a specific organisational unit fell out in the final equation. Inclusion of the relevant senior manager in the process helped to institutionalise recommendations, but this might also have led to the exclusion of technical recommendations and a scaling-up from the topic of the evaluation to something that was relevant to a larger proportion of the agency (projects which promote peace and work on conflict to managing projects in conflict prone and fragile settings). The evaluation created an institutional pressure to react to its findings. However, organisational capacity to absorb, discuss and implement findings is limited – there is evidence of strong competition with other topics that might be perceived as more relevant by senior management.

Contextual factors outside the development agency

The merger of GIZ created a huge demand for specific internal resources. Moreover, influence from outside GIZ – BMZ was revising its concept in working in conflict prone and fragile settings – was seen as an opportunity to discuss and change some approaches.

Conclusion

- Decision-making depends on individuals who act as advocates for a topic.
- The evaluation does not generate new knowledge, but creates momentum and an opportunity for a discussion to take place within the institution.
- The synthesis comprising of eight independent evaluations created pressure to react to its findings – and given the discourse on aid effectiveness – GIZ could not afford not to react, at least formally.
- The evaluation did not set the agenda alone – various stakeholders were influential, so the evidence base of implementation agreements can be questioned.

3.2 Department for Performance Monitoring and Evaluation (DPME), South Africa

Institutional context

DPME was created in 2010 and is situated in the Office of the President. It has a wide remit, using performance monitoring and evaluations to report on government priorities and service delivery, with a budget of approximately US$ 20 million and 180 staff. The Evaluation and Research Unit (DPME/E&R) is situated within the DPME, with a staff of 11. It is responsible for co-ordinating evaluations, providing technical support where necessary (or advising on where to find it) and reporting on progress against any recommendations that are made. The Evaluation and Research Unit is analogous to EuropeAid's Evaluation Unit: DPME as a whole has a wider remit.

Planning evaluations

The Evaluation and Research Unit helps develop a rolling 3-year annual plan for evaluation, in line with the National Development Plan and the government's 12 priorities. In 2012/13 it planned eight evaluations, rising to 15 per year from 2013/14 onwards.

This evaluation plan covers six types of evaluation: diagnosis, design, implementation, impact, economic and evaluation synthesis. Departments are asked to submit programmes for evaluation on the basis that they pay for their own evaluations. If more than one department is involved, a lead custodian department is designated, which is responsible for the evaluation and for subsequent reporting on progress against any recommendations that emerge.

DPME will contribute up to a maximum of ZAR 500,000 (EUR 37,000) per evaluation, if necessary. Large or strategic programmes need to be evaluated every five years.
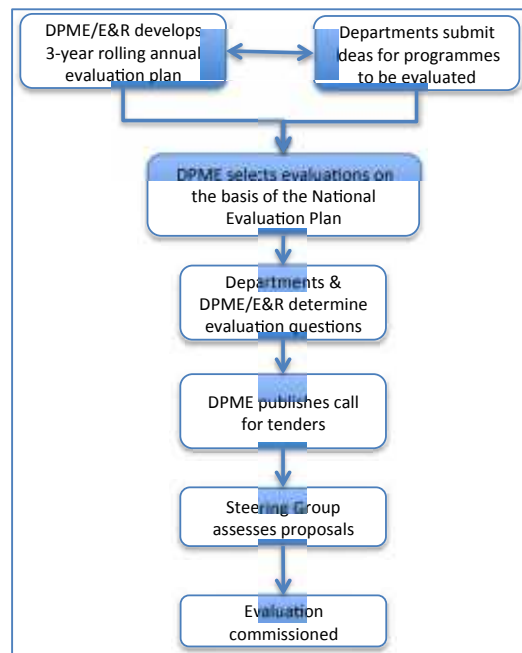
Figure 4: Evaluation planning in DPME

Commissioning evaluations

The ToRs are developed jointly by DPME/E&R, the managers of the intervention in question and key stakeholders. DPME/E&R provides technical support to ensure the questions are evaluable and gives advice on the broad methods which could be used. The tenders, published by DPME, ask the consultants to define in detail how they will address the questions. Once the consultants have been appointed and contracts exchanged, a week-long inception period is allowed, for consultants to conduct preliminary interviews and analysis and revise the proposal. The inception report is intended to be only about four pages long, solely for explaining the rationale for specific revisions to the proposal.

Once the Steering Group has accepted the inception report and revised proposal, the full evaluation is commissioned.

Managing evaluations

The Steering Group for each evaluation is constituted from the custodian department, DPME/E&R, partner departments, treasury, donors of evaluations or interventions, academics or sector experts, and wider stakeholders. Detailed ToRs for the steering group and its chair are set out: the Steering Group is expected to meet about seven times, though this changes with the needs of each evaluation.

The Steering Group provides advice to a bid adjudication committee on who should be appointed for the evaluation. Once this is done it then manages the evaluation, sometimes constituting a subgroup to work with the contractors on technical issues. However it does not just focus on the evaluation alone: its remit extends to approving the implementation plan, which is drawn up by the custodian department in the four months after the final report has been accepted.

Communicating the evaluation findings

Detailed guidance is given on communicating evaluation results including tailoring summaries of the findings to different audiences, work shopping the results with stakeholders, and developing a concise summary. There is no defined length for a full technical evaluation report: the normal format of report + annexes is used. The full technical report is accompanied by what is called a "1:3:25 report": a 25 page report with 3 page executive summary and a 1-page high-level summary. While the full technical

report has limited circulation, the 1:3:25 is widely circulated and also used to communicate the findings to non-specialist audiences.

Other communication is still a work in progress: DPME/E&R recognises that a major effort is needed in this field, but they have not yet set out a detailed plan.

Reporting on implementation

Figure 5: Reporting in DPME
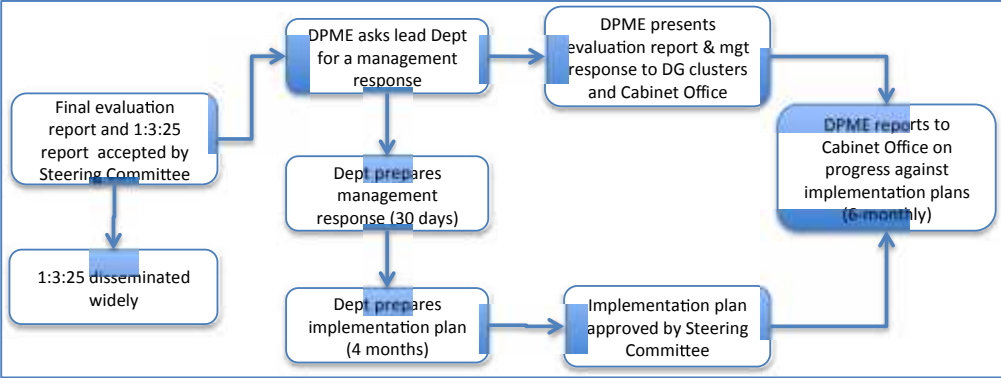
After the final report and the 1:3:25 have been accepted, DPME asks the custodian department for a management response. This is received within 30 days, after which DPME/E&R presents the results of the evaluation to the Cabinet Office and "DG Clusters", formal meetings of directors-general from across government. DPME also requests that the custodian department draws up an implementation plan within four months of the end of the evaluation. The implementation plan describes:

- The 'improvement objectives' which summarise the evaluation's recommendations into a workable plan and prioritise them (L/M/H).
- The activities that will be undertaken to achieve those objectives, defining responsibilities, deadlines, targets, what budget is available from where, and giving a current assessment of progress.

At 6-monthly intervals, DPME/E&R reports to the Cabinet Office on progress against these implementation plans.

3.3 GIZ

Institutional context

The M&E Unit in GIZ has two groups: the central evaluations group which manages corporate evaluations and is relatively independent; and the decentralised M&E group which provides advice to operational departments on how to implement M&E systems, and is not independent. There are eight people in each group which both report to the GIZ Board. They are independent from operations but not from GIZ. The evaluation group acts as a buffer zone between management, operations and evaluators, and they try to protect the independence of the evaluator.

Planning evaluations

Management are asked what questions they want answered. However, issues that might be of relevance at the time the evaluation is commissioned may not be so 1½ or 2 years later when the evaluation findings are ready. However, 90% of the questions they receive tend not to be evaluation questions. Hence, the M&E department together with evaluators decide which questions they will answer. Utility has been prioritised ahead of independence, although the evaluation group want valid and robust findings.

Managing evaluations

A reference group is formed made up of intended users, who are relatively high up in the GIZ hierarchy. They are identified by the M&E group. These include one of their board members, in order to help put a spotlight on the evaluation as well as, where relevant, sector specialists. They try to get senior management input during the process, but securing their time is challenging; when they do secure senior participation, discussions stay at a fairly high level.

There is no fixed methodology to follow to undertake evaluations, although robust findings must be ensured. There is a tendency to carry out smaller evaluations. The evaluation process is made up of a number of phases. After each phase the evaluators report back to the reference group. Corporate evaluations take one and a half years. Involvement during the process is very valuable for users, as opposed to simply being given findings at the end.

Communicating the evaluation findings

Responsibility for communication is taken by the evaluation group, which facilitates the process. A communications budget is built into the evaluation budget. External evaluators present the findings, after which their job is done. The main communication products include a five page executive summary and a four page briefing with graphics which are for conference type audiences. These are published on an internal website. They have a discussion as to whether GIZ should publish evaluations on the external website. There is a tension between being open to learning and transparent on one hand and protecting GIZ's reputation on the other.

Other channels include annual meetings of sector networks, where relevant evaluation findings might be presented. Networks have a mailing list. They are formal and often have a membership fee. In some cases they might meet biennially, with people meeting in Germany from around the world. How active the networks are often depends on the enthusiasm of the coordinator of the network.

Operational and regional departments are other channels for dissemination. People in sector departments are allies of the evaluation group as they are keen for project officers to incorporate learning from evaluation. Brown bag lunches aim to connect academic and practice oriented audiences. Although reports are accessible, there are multiple databases of information, which makes putting together a project proposal on a broad issue time-consuming. BMZ is working on creating one place where all reports can be accessed using a good search function.

Reporting on implementation

In a relatively new system, the recommendations in the report are discussed at the operational level with members from different departments in a 'learning café' (an interactive meeting to discuss recommendations). They formulate implementation agreements in a management response system (MRS) matrix which is then discussed with the respective management within each department. Finally the outcomes of these discussions are presented to GIZ's Executive Management Committee and Regular Operational Meetings where the formal decision on implementation is taken – see figure below. For the learning café, a professional facilitator is brought in – they might be from GIZ or an outsider. The evaluation group from time to time tries to trace the process – say after a year – to assess the extent to which implementation plan has been followed through.

Figure 6: Overview of discussion and decision-making process at the institutional level at GIZ

3.4 DFID

Institutional context

There is a strong commitment from DFID senior management to evaluate and ensure policies and programmes are informed by rigorous evidence. DFID's evaluation model has changed since 2011 when ICAI (Independent Commission for Aid Impact) was created. The evaluation department stopped being a manager for independent, corporate and strategic evaluations, and now operational departments (such as country offices and policy departments) commission their own evaluations. Policy areas often have their own specific evaluation strategy, e.g. there are strategies for governance and security, for private sector and growth, and for the humanitarian sector. Each country office develops its own evaluation strategy too. They are not able to evaluate everything they do, so they will have to make strategic choices based on major areas of spending, innovative areas of programming, the need to build up an evidence base in specific areas, capacity and partner interest amongst other criteria. As a result, programme leads are more aware of evaluation work throughout programme processes. This has encouraged more genuine dialogue on evaluation design, and more influence of evaluation on programme design.

The evaluation department now mainly tracks and monitors evaluation activity, sets standards, builds capacity across the organisation and assesses how decentralised evaluation activity aligns with DFID corporate priorities. However, there are a handful of policy and programming areas where the Evaluation Department have strategic evaluation commitments. These currently include a strategic evaluation on maternal and neonatal health, malaria and the international climate fund, which were selected on the basis of being political priorities.

Planning evaluations

There is a growing number of impact evaluations being undertaken. These involve the collection of a great deal of primary data, which DFID is committed to publishing under broader data transparency commitments, to allow for re-analysis and synthesis. DFID has also undertaken synthesis studies to look across a number of published evaluations.

Managing evaluations

Evaluation is built into programmes and policies at the outset, which helps to time and manage the evaluation process. Given the constraint on administration budgets, evaluation is charged to programmes, rather than administration, which has helped with developing an evaluation culture. All evaluations have their own governance arrangements. In some, the emphasis will be about ensuring independence of the evaluator, whilst in others it may be about bringing stakeholders together. DFID has flexibility in their approach to evaluation. First, the purpose of the evaluation is established, before an evaluation design is drawn up to respond to the evaluation questions.

At project and programme level, decision points are more predictable, but with higher-level policy issues, decision-making processes evolve and change shape. The evaluation manager must engage continually with the evaluation process, the consultants, as well as in internal decision-making processes in DFID to ensure that processes are aligned and that findings and recommendations are tailored to specific decision processes and audiences.

DFID has invested in building evaluation HR across the organisation. DFID see commissioning, managing, and using evaluation as a part of the core skill set of all their staff. There are evaluation specialists working in the evidence division as well as operational departments. Most country offices have evaluation specialists. There is an evaluation cadre which is open to all staff and accreditation at all levels. DFID currently has 150 accredited staff across the organisation. The frequency of accreditation has increased from an annual basis to quarterly.

Communicating the evaluation findings

There is a budget for communication and engagement built into evaluation budgets. Responsibility for promoting uptake is with the manager of the evaluation in DFID, who makes sure the evaluators have the right type of product and process in place. Potential evaluators are asked to set out in the tender document, a communication plan, include key actors, how they plan to engage with them, using what tools/channels.

Every DFID evaluation is published, which has required mechanisms to be put in place to secure clearance. All DFID evaluations have a management response, which is later published, and is discussed. There is a formal dissemination seminar, but the value of the uptake process is not necessarily during the seminar but in the detailed discussion which takes place during and after the development of the management response. The management response process can be lengthy, especially where evaluations are done jointly with other donors.

Evaluations are published internally on the 'evidence and policy exchange', a portal launched by the research and policy division, onto which a range of different evidence is uploaded. Some DFID staff consider the DFID website to lack a degree of 'user-friendliness'. The OECD database is considered by some DFID staff to have a better search function. Policy department staff play the role of knowledge brokers, and there is a good deal of experimentation taking place to drip feed knowledge to staff. For example:

- An evaluation newsletter which presents evaluation activity – both substance and methods
- An evaluation newsletter focussed on HIV and the theory of change, as well as climate and environment
- DFID works though its advisory networks, which act as knowledge brokers. There is a gender evaluation network
- The Africa directorate curates an 'evaluation digest', which picks up evaluation knowledge outside DFID
- Evaluation products are featured on the DFID intranet
- Every six weeks, there is a network meeting for evaluations across the organisation
- There is an 'evaluation café' convened to deal with thorny evaluation issues which people are confronting
- Professional development conferences on evaluation are held every year for members of the evaluation cadre. Professional development conferences held for other DFID cadres, include an evaluation component
- There are 'testing the consensus' seminars which bring people together to think about the implications of say a particular synthesis. Seminars in which academics and professionals discuss challenges, issues, etc. are common practice.

Handling negative findings has been challenging. The Evaluation Unit has tried to set clear standards which emphasise transparency, the need to put in adequate governance arrangements and discourage efforts to manipulate and bury negative findings.

Reporting on implementation

In shifting from an audit to learning culture, DFID no longer follows up on the management response after one year. This is something the audit department may consider.

3.5 IFAD

Institutional context

The Independent Office for Evaluation (IOE) which started in the late 1970s, became independent in 2003 whilst its functions have increased. It reports to the executive board of IFAD not IFAD management. The executive board is made up of representation from various countries. As such it has adopted a similar model to that of the World Bank's IEG. Although IOE interact with management, the analysis is owned by the evaluation office.

Planning evaluations

The IOE conducts evaluations on loans, country programmes, projects and other activities which are not project based, such as policy dialogue and corporate level evaluations (which are like institutional evaluations). They also conduct evaluations on issues at the global level within IFAD such as its work across the private sector, or the efficiency of IFAD processes. Moreover, they conduct syntheses of previous evaluations on common topics. They also study historical trends to see if issues have improved over time, remained the same or worsened.

To evaluate projects, they select a sample of projects which are then evaluated through a desk assessment of project completion reports and a field verification. They conduct about five country programme evaluations a year, selecting countries where IFAD intends to prepare a new strategy. The evaluation is then seen as a key input to strategy development processes.

Corporate evaluations are selected depending on issues raised by management. For instance recently, IFAD management felt that the policy on private sector intervention required a revision, so the IOE were asked if they could conduct an evaluation.

They have a multi-year rolling programme, but have recently introduced something akin to horizon scanning to assess key decision points for which evaluation evidence may be useful. Corporate evaluations are for instance ideally conducted two years in advance of key decision points.

Commissioning evaluations

Corporate evaluations start with a concept note which lists key evaluation questions. This is used to start a discussion within the IOE. The concept note is then developed into a full approach paper. The executive committee (a branch of the board) reviews this. Colleagues from other evaluation offices will occasionally be asked to comment on the approach paper.

Managing evaluations

There are strict guidelines for conducting project and country evaluations. However there is more flexibility regarding corporate level evaluations. All this is documented in an evaluation manual. Project evaluations take about six months, country evaluations take 12 months while corporate evaluations take 18 months. Evaluation processes usually comprise literature reviews and field visits. The evaluation team presents progress to the evaluation manager during the process to see if they are on track.

Communicating evaluation findings

When the evaluation office has completed an evaluation, it is responsible for organising discussion of the findings. Knowledge management and learning accounted for 21% of the budget in 2013. (This includes staff and on-staff costs and includes consultants). In 2014, this figure is projected to rise to 27% of the budget. The evaluation approach paper includes a modest communication plan. This includes, in broad terms, the key audiences with whom they need to engage, what kind of inputs are required from those audiences and ways to engage them (such as events). It includes a stakeholder mapping exercise. The report is often sent to key audiences before it is made public.

Once the findings have been drafted, there is a presentation to managers, and in some cases a mini workshop is held with colleagues from other international organisations to discuss results. The report is finalised and sent along with comments to management. The IFAD management then writes a management response, in which they explain how they will follow up on the evaluation recommendations.

Evaluation findings are set out in about 50-60 pages, with about 5-6 key messages. Most reports are discussed with the Board. They are given executive summaries with a 5500 word limit, with reports usually much shorter – 7 pages for a country level evaluation and 10 pages for a corporate level evaluation. There are also two page summaries, which highlight one issue which the evaluators consider management should focus on. IFAD has some budget for communication events. If held in Rome, they can be relatively cheap. In these cases, half a day is taken up with a presentation whilst the other half is taken up with a plenary meeting with the Board.

If a communication event is located overseas, it has to be part of an evaluation budget. For country level evaluation, there is a one day workshop, where the main findings are presented. There are working groups made up of different stakeholders (government, private sector, NGOs) to work on a few key themes and critique recommendations. There are on occasion, *ad hoc* events to present evaluation findings. The IOE has a Communication Unit, which helps. Occasionally, they hire editors from outside, especially when the audience is non-specialist. IOE puts their evaluations on a publicly accessible website, but they do not collect data to see which reports are most popular

The evaluation reports are also shared with two key evaluation networks: the UN evaluation network, and the evaluation group of the UN and multilateral banks. These networks have a few events a year where evaluations are presented. They also get invited to professional association events, especially the larger ones, where findings from select evaluations will, again, be presented. They publish an annual report that consolidates the findings of evaluations done on strategic issues. People from the IOE will be observers during country strategy development processes and feed in relevant findings.

Reporting on implementation

Every year the president of IFAD presents how management has followed up on the recommendations of evaluations conducted 12 months before. The IOE is able to comment on this report. IFAD is still fairly centralised. Until a few years ago, all decisions were taken in Rome. IFAD has only recently started opening country offices, but decision-making authority has not been delegated.

3.6 SIDA

Institutional context

In SIDA evaluations are mostly decentralised with programme managers managing evaluations and SIDA providing backstopping support. There is an independent Evaluation Unit which assesses Swedish development cooperation, with an internally located M&E Unit, evaluating programmes from 'within'.

Planning evaluations

Each year, heads of operational departments meet to identify the types of information they need. A small number of issues are ultimately identified which are relevant to a number of thematic areas and form the basis of a strategic evaluation.

Commissioning evaluations

A reference group is set up. The members draft ToRs. The challenge is that the evaluation issue can be fairly broad, so making it interesting and evaluable means the drafting process can take considerable time. External evaluators are then hired.

Communicating the evaluation findings

Once an evaluation is finalised, it goes to the 'Evidence Group'. This group has a conversation with representatives from different operational departments at SIDA about the conclusions and recommendations. Since the latter are not always specific, or well targeted, the group develops a set of actions which can be implemented with the resources available, stating which department is responsible along with a deadline. These actions form a proposal (using a pre-defined template) which goes to the Director-General for approval. Once approved, a set of decisions is sent to relevant operational heads of department requiring them to implement stated actions.

The evidence group, having drafted the management response, decide whether there should be further communication, internally or externally – the evidence group is formally responsible for this. SIDA's intranet is used to share the evaluation findings internally. A 2-page evaluation brief with findings, conclusions and recommendations is produced and disseminated across the organisation. On some issues, the evidence group might decide to organise a 'development talk', which is an external event open to the public and media to attend, or alternatively they might decide to organise an internal seminar.

Within the organisation there are informal thematic networks. The lead specialist usually determines how active they are. The evidence group has resources to set up meetings and workshops, and often meet virtually, through videoconference.

Reporting on implementation

Formal reporting for departments will include reporting on how they have implemented specific recommendations. The Financial Management Department at SIDA are developing a reporting format which will force departments to report on implementation of actions stemming from evaluations if relevant.

3.7 World Bank IEG

Planning evaluations

The World Bank IEG developed a criterion which identified a number of factors which are more likely to promote uptake of evaluations. These include for example whether there is a champion who could pick up recommendations, or whether the policy issue was of current significance. These criteria are used to identify strategic level evaluations which are then discussed with the Board, a relevant sub-committee, as well as the bank management in several rounds. The IEG may be approached by the bank management to do an evaluation, or they may do an evaluation which the management are opposed to. In either case, they have the independence to do an evaluation, although their budget is ultimately approved by the Board.

Commissioning evaluations

An approach paper is drafted and initially circulated between IEG and the management, before being approved by the Board and then published. There are few evaluation guidelines, and hence a high degree of flexibility. Evaluations are undertaken by bank staff or consultants who are familiar with the bank but do not necessarily have country expertise. The Board is seen as the main audience rather than the recipient of projects and programmes. Dissemination process is only vaguely spelt out at this stage.

Managing evaluations

Stakeholder engagement during the evaluation process is minimal. There is buy-in from the bank management to the evaluation – based in large part on the mutual respect between the task team and the evaluators, who are often well known amongst staff, and are often retired country directors.

Before the evaluation is finalised, a communications plan is prepared. This will include audiences, key messages and channels (including social media and whether to go back to countries they visited during the evaluation). Stakeholder analysis is fairly thin however. Every evaluation has a communications budget built in. Each department managing evaluations has to allocate resources for their communications. On a quarterly basis, the departments look at their communications plans and identify which evaluations they should focus on. The most critical decision is whether or not to go to the media – see below. About 10 per cent of the IEG's budget is spent on learning and communication. There is some demand internally for more of the budget to be spent on engagement rather than undertaking evaluation, but it is hard to justify further spending on the former as the Board has a narrow definition of what counts as value.

Evaluators may draft recommendations, but they are discussed in detail and finalised with the bank management. A trade-off is made between independence and utility. However, the quality of evaluations is seen as high and the IEG as credible. It uses a framework to assess if recommendations are worth making, and if implemented how they will know it has contributed to improvements. If recommendations from the study are not feasible, management come back with recommendations which are, but in the spirit of the original recommendation. The Board may well disagree with both the initial and revised recommendations. The detailed face-to-face discussion between the IEG, Board and management is a large part of the uptake.

Communicating the evaluation findings

When they IEG makes evaluations publicly available, they are published as part of a package, which includes the management response and a summary of the discussion with the board. On occasion there are high profile dissemination events, which are well thought through, and involve senior engagement from the bank – including managing directors and vice-presidents. However, this doesn't always facilitate discussion of the more detailed aspects of the evaluation findings.

The IEG does not return to the countries which were included in the study, but they do give feedback via video link and opportunistically as people travel. Typically it is the people who have led the evaluation in addition to the manager of the programme being evaluated who engage. The presence of the manager protects the evaluator's independence to some extent.

IEG staff are careful about going to the media as the latter can often interpret messages differently and damage reputations. If the IEG does go to the media, they send an overview document to relevant parties, whilst the full evaluation is put out on the website. The overview document is a self-standing summary of the findings, recommendations, management response and Board summary. They ensure that the report conveys the story, taking out features which could distort it or be taken out of context.

Internal communication:
- Evaluation products are posted online
- They have invested a lot in systems, improving their search engine, accessibility, structure and linking social media channels with the institutional database
- They pay for Google ad campaigns so evaluations come up when relevant search terms are entered
- There is an 'evaluation week' every year
- There are events held jointly with bank management which they stream to country offices on evaluation topics where there is something to be learned, or on issues which are likely to attract large audiences

- There is a newsletter which goes out to 40,000 people, which leads to big impact in terms of downloads, for which they have data
- There is a social media channel
- There are videos which are sent out featuring learning events
- There are brown bag lunches at Bank HQ where experts are invited to present, and which are streamed online.

Reporting on implementation

The Bank management take the lead on instigating changes which have been agreed. Senior leadership support is a key determinant of change. But policy change usually happens incrementally, and at country level there may be little space for change. Major evaluations are filed for four years. There is a management action record system. Each year, IEG makes an assessment of changes which have occurred in collaboration with the management, which is passed on to the Board.

4. INDIVIDUAL STUDIES OF 25 EVALUATIONS

This section sets out the desk reviews of 23 EuropeAid commissioned strategic evaluations. They are complemented by more in-depth process tracking reviews of 4 evaluations in a subsequent section. The selection of the evaluations was agreed with the Reference Group on the Study on Uptake, following the inception meeting. The selection of evaluations reflects the different types of strategic evaluations the EU produces, whilst maintaining a balanced geographic coverage.

All the case studies cover single evaluations with the exception of BS, which covers the four evaluations of BS conducted in Mali, Tunisia, and Zambia and draws on evidence from the most recent evaluation of BS in Tanzania (finalised in October 2013). This is to capture the methodological development which has been a key part of the BS evaluation process.

The review looks into four dynamics related to uptake, and are analysed through a document review:

1. **Process dynamics:** timing and planning of the evaluation, potential links with the programming cycle and/or policy processes, and the definition of evaluation needs.

2. **Response process and dissemination**: response of management and different institutional stakeholders to the evaluation, and consideration of the different ways in which the evaluation was disseminated.

3. **Programming dynamics:** analysis of the extent to which the evaluation was taken up in the programming and/or policy development processes of the EU.

4. **Qualitative dynamics:** the extent to which these evaluations were done in a way which is conducive to uptake (meeting evaluation needs, concrete and feasible recommendations, etc.).

Nr	Evaluation	Title	Year
Country level evaluations			
1	**China**	Evaluation of the EC Co-operation and Partnership with the People's Republic of China – ref. 1077	2007
2	**CAR**	Evaluation de la coopération de la CE avec la République centrafricaine – ref. 1263	2009
3	**Vietnam**	Evaluation of the European Commission's Co-operation with Vietnam – Ref. 1269	2009
4	**Nicaragua**	Evaluation of European Commission's Cooperation with Nicaragua – Ref. 1271	2009
5	**Nigeria**	Country Level Evaluation Nigeria – Ref. 1276	2010
6	**Egypt**	Evaluation of EC support to Egypt – Ref. 1281	2010
7	**Ukraine**	Evaluation of the European Commission's Cooperation with Ukraine – Ref. 1283	2010
8	**Malawi**	Country Level Evaluation – Republic of Malawi – Ref 1297	2011

	Regional evaluations		
9	**Central Africa**	Evaluation de la stratégie régionale de la CE en Afrique Centrale – ref. 1037	2006
10	**Central America**	Evaluación de la Coopéración Regional de la CE en America Central – ref. 1092	2007
11	**ASEAN**	Evaluation of EC Co-operation with ASEAN – ref. 1262	2009
	Sectoral evaluations		
12	**Water and sanitation**	Thematic Evaluation of the Water and Sanitation Sector – ref. 965	2006
13	**Energy**	Evaluation of EC Support to Partner Countries in the area of Energy – ref. 1192	2008
14	**Education**	Thematic global evaluation of European Commission support to the education sector in partner countries – Ref. 1282	2010
	Thematic evaluations		
15	**Good Governance**	Thematic Evaluation of the EC Support to Good Governance – ref. 884	2006
16	**Conflict prevention and peace building**	Thematic Evaluation of European Commission Support to Conflict Prevention and Peace Building – Ref. 1277	2011
17	**Human Rights**	Thematic evaluation of the European Commission support to respect of Human Rights and Fundamental Freedoms – Ref. 1298	2011
	Aid modalities		
18	**CSO channel**	Evaluation of EC aid 46 channelled through civil society organisations (CSOs) – ref. 1259	2008
19	**Investment Facility**	Mid-term Evaluation of the Investment Facility and EIB own resources operations in ACP countries and the OCTs – Ref. 1285	2010
20	**Budget support Tunisia**	Evaluation de opérations d'aide budgétaire de la CE à la Tunisie – Ref. 1286	2011
21	**Budget support Mali**	Joint Evaluation on BS to Mali – Ref. 1290	2011
22	**Budget support Zambia**	Between high expectations and reality: An evaluation of BS in Zambia	2011
23	**Budget support Tanzania**	Joint Evaluation of BS to Tanzania Lessons learned and recommendations for the future – Ref. 1321	2013
	Synthesis reports		
24	**Synthesis report BS Mali, Tunisia and Zambia**	Application of new approach to the Evaluation of BS operations: Findings from Mali, Zambia and Tunisia	2011
25	**Synthesis of the geographic evaluations managed by the Evaluation Unit**	Synthesis of the geographical evaluations managed by the Evaluation Unit during the period 1998-2006 – ref. 1256	2008

4.1 China

Name of evaluation	Evaluation of the European Commission's Co-operation and Partnership with the People's Republic of China
Language	English
EuropeAid reference nr.	1007
Type of evaluation	Country level evaluation
Year of evaluation	April 2007
Documents reviewed	• Main report and Annexes • Executive summary • Summary document • Presentations (by consultant and by EU personnel) • Fiche Contradictoire • Quality Judgement

Key points:

• *The evaluation was carefully specified to be forward-looking to take account of the EC's changing relationship with China: it was pitched at a very strategic level, and was not a 'traditional' country evaluation.*

• *Key lessons which are transferable to other countries are clearly set out in the evaluation report. However it is not clear how these lessons were supposed to be taken up, as there was no further record of dissemination after the country seminar.*

1. Process dynamics

Evaluation planning and timing

The evaluation covered the years 1998 – 2006, taking place between February 2006 and May 2007.

Links to policy and programming

There were no links to thematic or sectoral policy production, but the evaluation was designed to give insights into the design and implementation of the current CSP. The timing was scheduled to help contribute to the new CSP covering 2007-2013.

Definition of evaluation needs

The evaluation needs were carefully defined to take account of the EU's changing relationship with China, and there was not much change between the ToRs and the final evaluation questions. The ToRs (paragraph 22) clearly specified that the evaluation should be forward looking: providing lessons and recommendations for the continued support to the partnership with China.

Sector-specific questions:
• To what extent has the EC approach contributed to integrating China into the world economy?
• To what extent did the EC approach contribute to economic and social reform in China?
• To what extent has the EC approach reflected the need to ensure the environmental sustainability of economic development in China?
• To what extent has the EC contributed to promoting transition to an open society based on rule of law, democratic processes, and respect for human rights?
• How successfully did the EC mainstream the issues facing disadvantaged hinterland regions such as Western China, the Northeast, and the central Provinces into its approach?

Synthesis/generic questions:

- (Adapting to change) Has the EC's approach maintained its relevance by adapting to the evolving political, socio-economic and environmental challenges?
- (Mix of instruments & interventions) Were the mix of instruments and interventions (coordination, political dialogue, sector dialogues, TA, pilot projects, bilateral cooperation projects) appropriate to the national context and EU strategic policy aims?
- (Adding value) To what extent did the EC approach result in progress toward the goal of better integrating China into the world economic, political, and environmental community which would not have occurred absent the EC?

The evaluation was rated as "good" in the Quality Grid, which noted that it covered the ToRs and information needs to a "reasonable degree" but also that it went beyond its brief in discussing other topics of interest to the EUD such as developing an exit strategy.

2. Response process and dissemination

Fiche Contradictoire

The response process is detailed and follows the structure of the report's recommendations. The text discusses the issue rather than simply providing a technical response to each recommendation, providing a line of argument though not necessarily setting out a detailed programming-type response. The argument in the second column continues this, noting changes which have taken place in EU policy, such as the fact that several MS were phasing out development co-operation projects with China and the decision by the European Parliament to allow MS projects to be implemented through the EC.

In the main, the initial response agrees with the findings of the evaluation, though points out issues which the evaluation may not have considered. For instance, the evaluation noted that more attention should be paid to replication and roll-out during project formulation and implementation phases (R4, on p49 of the main report). The FC notes that the issue of replication is less sharp than it was due to the changing nature of projects which are being implemented; and also that there are timing issues as it is often not clear whether or not there will be anything to replicate until close to the end of the project which (means keeping a project running at a low level thus running counter to normal project management disciplines).

The follow-up does provide further information on what else has occurred since the evaluation was completed, noting for example the flexibility of particular funding instruments and how they can improve co-operation and dialogue. The follow-up does seem to take the initial response in the FC further.

Dissemination process

The evaluation itself and the various summaries (particularly the one on EVINFO) do a good job of identifying lessons which are transferable to other countries, including:

- **Improving the learning from EC-supported pilot projects**: in countries with distinct sub-national entities like China (i.e. regional governments), special mechanisms for planning and evaluating of pilot projects would help replication elsewhere.

- **Phasing out of development co-operation**: noting that planning for an exit does not appear automatically in a work plan - it needs to be explicitly put on the strategic agenda and then taken to the concrete levels of programming and implementation.

- **Co-operation among equals**: emphasizing that while countries may agree on equal partnership principles at a high level, this may not be reflected in the arrangements for implementation. In China, the high-level partnership of equals provided for expertise from the EC but not resources, which created frustration when it came to implementation.

No other uptake initiatives were seen – the slides from the presentation at the dissemination seminar were highly summarised, and do not really provide a useful basis for disseminating the recommendations. There are no minutes from the dissemination seminar. The EVINFO report is probably the best communication vehicle as it sets out the transferable lessons. The minutes of the reference group are not really about dissemination. The report identifies several potential audiences: policy maker and managers in the relevant external co-operation services, DG Trade and DG Environment, and EU MS active in China, but it is not clear whether there were any specific communication initiatives targeted towards them.

3. Programming dynamics

The evaluation was very much focused on informing the future of the EC-China co-operation programme, but there is nothing in the documentation to indicate clear links to programming, though there are several references to the need to better co-ordinate joint thematic programming between the EUD and MS and for strategic co-ordination of positions and initiatives.

4. Qualitative dynamics

Analysis of the recommendations

The evaluation questions were set in a fairly traditional way (responding to the DAC criteria, the 3Cs and the adding value question), but they were informed by a wider set of questions as set out in section 1 of the report: *What sort of cooperation programme is called for given the nature of China – an immense country and global economic power? Can EC spending on cooperation, small as it is relative to domestic resources, make any tangible difference in China? Is it coherent with the EC's Development Policy goal of reducing poverty? Does EC cooperation encourage China to act as a responsible global trade and economic partner, or does it merely amount to a subsidy to a formidable competitor? Does cooperation advance EC policy goals relating to human rights and open society?* The team pointed out that, since interaction between co-operation and other forms of engagement is crucial, it would not be possible to evaluate the co-operation programme in isolation. And because EC-China co-operation has been moving away from traditional project-oriented development assistance, the evaluation itself placed less focus on stand-alone projects and their performance than it did on the aggregated strategic level, assessing "*how the country strategies and agreements have served to steer, and give appropriate focus to, the country programme and its implementation*" (Main report, p2).

Scope of the evaluation

The evaluation question is very broad: what kind of partnership and co-operation programme with China makes sense, now and in the future? This translated into eight EQs which covered the main goals of EC programming, including integrating China into the world trade system, achieving social and economic reform, ensuring environmental sustainability and good governance, and the question of the needs of China's hinterland regions. Three synthesis questions were added (which are more similar to 'traditional' country evaluations); these looked at how the different instruments were combined, how flexibly the EC's approach adapted to change and whether the EC added value.

4.2 Central African Republic

Name of evaluation	Evaluation de la coopération de la Commission européenne avec la République Centrafricaine
Language	French
EuropeAid reference nr.	1263
Year of evaluation	2008-2009
Documents reviewed	• Final report • Annexes • Executive summary (English & French) • Summary document • Quality Judgement • Fiche Contradictoire

Key points:

- *The timing of the evaluation did not allow for smooth integration of the evaluation results in the programming for CAR under the EDF10.*
- *The evaluation captured the political and cooperation implications of the stabilisation of the country following the 2005 elections.*
- *Many of the recommendations which were issued by the evaluators feature in the new CSP. It is likely the result of both documents responded to the same political developments in CAR.*
- *It is possible that the evaluation was one of the ingredients of the programming discussions, however the document review cannot confirm this.*
- *Some of the EDF10 interventions were formulated (action fiches) when the evaluation was available. One of these action fiches makes mention of the evaluation's conclusions.*

1. Process dynamics

Evaluation planning and timing

The evaluation was part of the 2007 evaluation strategy. The evaluation was carried out in 2008 and covers the EC's cooperation with CAR in 1996-2007. The final report is dated June 2009.

Key dates:
- Inception note (according to ToRs): December 2007
- Dissemination seminar Bangui (according to ToRs): September 2008
- Date on final report: June 2009.

Link to policy and programming

The period covered by the evaluation was one of consecutive crises and insecurity. The evaluation notes that the EDF (7-9) programming procedures were "ill-suited" to respond to the cooperation needs (beyond development) which existed at the time. Improvement came after the 2005 elections with the EDF10 programming cycle (2008-2013). The programming of the EDF10 partly coincided with the evaluation process, but the CSP predates the evaluation, so results were not (visibly) taken up in the 2008-2013 cycle.

The evaluation signals that, for the period under review, EC cooperation followed programming documents too rigidly, preventing the EC from adapting its programming to address the stabilising needs of a fragile country. For example, "the first interventions under EDF 10 started in 2009, the population therefore would have had to wait at least until 2010 to feel the effects", which is a delay of 5 years since the elections. While the country held elections in 2005, a change in EU cooperation only came with the EDF10 programming cycle (2008-2013). The first recommendation of the evaluation is therefore *"to establish a procedure for derogation from the EDF's programming cycles following EU recognition that a partner country has entered a phase of instability"*.

The ToRs outline standardised evaluation needs for country strategy evaluations, which are based on OECD-DAC criteria and are generally clear and well defined. The evaluation questions in the report go into the focal and non-focal sectors of EDF 7-9, but also address the EC's actions in the field of conflict prevention, linking relief, rehabilitation and development and the EC interventions under the African Peace Facility:

"EQ7 – To what extent did the combination of EU instruments and other horizontal tools, such as the African Peace Facility or the EU external action contribute to the resolution of crisis in CAR?"

"EQ 10 – In view of the chronically instable context, has the EU approach, particularly with regard to good governance and the rule of law, integrated and anticipated dimensions of conflict prevention and LRRD (« Linking Relief, Rehabilitation and Development ») in the programming and implementation of its programmes?"

The quality judgment fiche acknowledges this explicitly and rates the "relevant scope" as "very good".

2. Response process and dissemination

Fiche Contradictoire

The response process follows the recommendations in a detailed though rather technical manner. The initial response agrees with most of the recommendations, except "specific recommendation 6" for which clear argumentation is given. The Commission services used the response process to (1) nuance the recommendations, and (2) **argue that those recommendations are already taken into account in current practice or are part of the EDF10 programming**. The Fiche Contradictoire does not set out new actions which specifically address the recommendations in the evaluation.

The response after one year is fairly detailed and outlines very concrete changes which were made since the evaluation was published, particularly with regard to the general (strategic) recommendations 1 (derogation procedures under the EDF) and 2 (Internal strategy document for conflict prevention and peace building). **It is difficult to identify these steps as an 'uptake' of evaluation results, mainly because these steps are part of two parallel processes**: the EDF10 programming and the creation of the EEAS.

Dissemination process

No dissemination documents are included in the annexes or the EuropeAid evaluation page. Apart from the standard EVINFO no details on the dissemination process are available for the document review.

3. Programming dynamics

Follow-up in programming and the mid-term review

The timing of the evaluation process coincided with the programming exercise of the EDF10, which did not allow for a formal uptake in the new CSP. The 2008-2013 CSP predates the final report of the Country Strategy Evaluation, but not enough for it to be taken up in the evaluation scope. On the other hand, the political stabilisation in 2005 meant that the **evaluation results and the new country strategy respond to the same political and co-operation developments.** In that sense, many of the evaluation results, especially with regard to conflict prevention and peace building are also present in the country strategy. A concrete example is the emphasis that the evaluation puts on the so-called *"pôles de développement (translation: poles of development)"*. The EDF10 fully adopts this approach across its interventions. However, **this cannot formally be attributed to the evaluation.**

No midterm review of the CSP is available.

While the country strategy pre-dates the evaluation report (and dissemination process), this is not necessarily the case with the formulation of EU interventions in the country **The Action Fiches of the EDF10** (though incomplete) are partly available online. All EDF10 Action Fiches include a section on lessons learned, however only one[1] of the publicly available project documents makes a direct mention of the evaluation:

"CAR - Rehabilitation of basic services and capacity building of local authorities in development initiatives (Phase 1): No. EDF/2009/021-431.
The evaluation of the cooperation of the European Commission with the Central African Republic for the period of 1996-2007 concluded that: the approach developed for the programming of the EDF10 - which was developed at the end of the period under review and gives priority to development of secondary towns (hubs of development) and the implementation of an integrated multi-sector approach at this level - appears to be a suitable strategic response to the context and needs of the country. It proposes to incorporate dimensions that have hitherto been developed separately: DDR2, LRRD3, local development, and decentralisation."

While this is not necessarily a sign of uptake, it is an example of how **evaluation results are used to support the approach** set out in the programme.

4. Qualitative dynamics

Analysis of the recommendations

The evaluation issues two sets of recommendations: (1) general recommendations for strategic action and (2) specific recommendations linked to the concentration sectors of EU cooperation with the country. The first set of recommendations is linked to the general conclusion that EU cooperation during the period under review largely failed to address the crisis situation in the country, and rather continued to focus on infrastructure as the country further slipped into crisis. The recommendations are both technical (increased flexibility of EU instruments) and strategic (stabilisation before infrastructure, which was the most important focal sector throughout the period under review). The Fiche Contradictoire confirms the feasibility as well as the relevance of the strategic recommendations.

Scope

The evaluation was praised in the quality assessment for moving beyond a mere performance assessment and fully integrating the conflict dimension, in its general as well as specific evaluation questions: *"The emphasis that was put on the conflict resolution and prevention has allowed original and relevant angle of analysis. (...) Other evaluation questions have taken the political situation into account of the country, either in the wording of the questions or in the judgment criteria."* This is also reflected in the strategic recommendations.

---

[1] Note that not all action fiches were publicly available for this document review. It is also beyond the remit of this review to develop a detailed calendar of when these programmes were identified and formulated.

4.3 Vietnam

Name of evaluation	Evaluation of the European Commission's Co-operation with Vietnam
Language	English
EuropeAid reference nr.	1269
Type of evaluation	Country level evaluation
Year of evaluation	2009
Documents reviewed	- Main report and Annexes - Summary document - Fiche Contradictoire - Quality Judgement

Key points:
- *The formal prescribed methodology was difficult to implement and had to be used flexibly.*
- *The evaluation recommendations do not seem to acknowledge the very complex political economy of the Vietnamese context and hence at times seem unreasonable.*
- *In some parts, the FC seems to suggest that the evaluation has played a legitimation role – validating actions which the Delegation had already been taking.*

1. Process Dynamics

Evaluation planning and timing

The field visit was between 13th and 27th Feb with discussion seminar on 16th June. The report is dated October 2009.

Links to policy and programming

The evaluation was done under time constraints, enabling it to feed into the mid-term review of the country programme. There were no wider links to policy production.

Definition of evaluation needs

The evaluation was based on nine evaluation questions, relating to the five DAC evaluation criteria – relevance, effectiveness, efficiency, sustainability and impact – plus coherence and EC value added. There was no information on who was involved and at what stages: the report in the methodology section mentions a launch note but this was not in the documentation available to the team.

2. Response process and dissemination

Fiche Contradictoire

Not all recommendations are referred to in the FC. Some responses suggest that the Delegation is already addressing issues raised in the evaluation, through specific projects, or had planned to do so. Where action has been taken, it is unclear whether that was going to happen anyway or whether it was because of the evaluation. Some responses are inadequate. They show how further actions have been taken to address the initial recommendations.

Dissemination process

Discussion seminar took place on 16th June 2009 – but there are no seminar minutes apart from the slide presentation.

3. Programming dynamics

Follow-up in programming and the mid-term review

4. Qualitative dynamics

Analysis of the recommendations

Report recommendations seem a bit unreasonable in places, and are based on getting the Vietnamese government to adopt liberal norms (e.g. the evaluation says EC should strengthen checks and balances - through NGOs and better PFM) in a Vietnamese political context which many say is largely elite driven. Also how does one promote 'full ownership', if it is a pre-requisite for effective programming? In some cases, recommendations are suggesting the Delegation continue to do things they have already started, making it easier for the Delegation to respond to.

On the EVINFO summary, the recommendations are not linked to the lessons – they address different things. The recommendations tend not to discuss major political issues even if many donor led reforms in Vietnam are highly contested. However, the report does mention that the policymaking process is opaque.

Recommendations addressed by the FC focus on strengthening links between the programme and poverty, global BS, sector support and broader planning and responsiveness issues. Most of these issues require political commitment, so the evaluation is unsure on how to classify them. The recommendations rarely address the 'how' question – how do you put into practice the recommendations? Who do you work with and what approach do you take etc.?

Other information
The methodology did not include review of academic literature – just project documentation. The report states problems with the quality of local statistical data. Many of the indicators turned out to be difficult to measure, or irrelevant, or the Judgement Criteria were found to have gaps. As a result, the evaluation team had to be flexible in the way they used the criteria.

Scope

Although the QJ report says the specific context of the country has been taken into account – this review doubts they have - i.e. the report expects independent NGOs to come up and hold government to account for its spending of general BS – which is unrealistic.

4.4 Nicaragua

Name of evaluation	Evaluation of the European Commission's Co-operation with Nicaragua 1998-2008
Language	English and Spanish
EuropeAid reference nr.	1271
Type of evaluation	Country level evaluation
Year of evaluation	2009
Documents reviewed	• Main report (English & Spanish) • Annexes A and B • Summary document • Fiche Contradictoire • Quality Judgement

1. Process dynamics

Evaluation planning and timing

The evaluation started mid-2008. A preparatory field visit was held in September 2008 and the main fieldwork took place during a three-week visit in January-February 2009. The final report dates from November 2009.

It is important to stress that the actual fieldwork coincided with a major crisis in donor relations. After the September 2008 municipal elections, the Commission, in coordination with MS, decided to suspend BS to Nicaragua's Sandinista government who had regained power in 2007 after two decades of conservative rule. While this must inevitably have complicated the fieldwork of the evaluation team, it surprisingly did not lead to any visible (even minor) change in the ToRs of the evaluation or in the nature of the EQs to be addressed by the team. Moreover, the report's conclusions and recommendations faithfully follow the initial ToRs and related focus on assessing the relevance and impact of EU cooperation between 1998-2008. There was apparently no attempt (or indeed request from the EU) to adapt or at least integrate the effects of the political crisis into the evaluation process (beyond general references to the crisis). The consultants also did not see the need to go deeper into the political crisis and above all into its effects on future cooperation. They could perfectly have exploited the available ToRs to take these wider issues on board (e.g. under the umbrella of "relevance" of the overall EU strategy).

Link to policy and programming

There are no links to policy, but a rather limited link to the programming cycle (see below).

Definition of evaluation needs

The evaluation was based on nine evaluation questions, relating to the five DAC evaluation criteria (relevance, effectiveness, efficiency, sustainability and impact) plus coherence and EC value added. While the nine EQs all cover relevant issues, it is surprising to note the rather low-profile status of the 'political dimensions' of the partnership. The focus is largely on traditional development issues, focal sectors, aid modalities and the 3Cs. There is a EQ dealing with governance, but one may have expected different stakeholders in the process (RELEX and the EUD) to insist on the need to look deeper into the wider political aspects of the relationship with Nicaragua - especially as tensions with the new government were already mounting over a year before the start of the evaluation. Yet this did not happen (despite two field visits and several RGs). The largely development-oriented focus was kept all along the process. Within that rather narrow remit, the evaluation was appreciated for "meeting needs" and for the "validity of the conclusions and the usefulness of the recommendations" (both were given a score of "very good" in the Quality grid).

It is also interesting to note that the ToRs explicitly requested the team to "evaluate whether the recommendations of the 2000 Country Level evaluation, the previous regional evaluation and the sector evaluations, in which Nicaragua was a pilot country, were useful and to what extent they have been taken into account in the current programming cycle, including the reasons in the context of the decision-making process". Reading through the main report, it does not appear that the evaluation team did this uptake analysis: references are spread over the report to various evaluations but not a systematic analysis of the evaluation as requested by the ToRs.

The RG comprised members of RELEX, EuropeAid, the EU Delegation and ECHO (as there was a question on humanitarian aid and LRRD) but there are no minutes available of RG meetings. An in-country seminar was envisaged but it is not clear if it was organised. The EVINFO faithfully reflects the report conclusions and recommendations. But, here as well, there was no attempt to link the evaluation to the significantly changed situation on the ground.

2. Response process and dissemination

Fiche Contradictoire

There was good buy-in, expressed in the form of fairly generic commitments to ensure follow-up if conditions allow. The FC contained constructive responses to the six main recommendations, some of which are "accepted", others "shared" or "fully shared" and still others "agreed under conditions". Also in this case, the absence of the whole political dimension of the partnership is remarkable, beyond the mention that BS is currently not a possibility in Nicaragua. The response is a classic example of a development-oriented, instrumental and technocratic approach to EU cooperation.

Follow-up in the FC was apparently two years later (reason unknown). It is difficult to see a direct link between the evaluation and changes in EUD policy and practice for several reasons:
1) the relative absence of a focus on the political crisis in the evaluation also means there are no recommendations that facilitate short-term management responses.
2) no attempt was made to go a bit deeper into the question how the EU could support governance (one of the 3 focal sectors) in such a difficult environment. If the team would have been able to do this, it might have resulted in much more interest for uptake of the related recommendations.
3) most other recommendations (e.g. improved coordination and task division or promotion of delegated cooperation) are linked to wider EU reform processes driven by HQ. If the EUD works more on these matters it is because there are higher drivers than simply this evaluation.

Dissemination process

It is not clear if a dissemination seminar took place. No other uptake initiatives were found.

3. Programming dynamics

Follow-up in programming and the mid-term review

A mid-term review of the cooperation with Nicaragua was done later in 2010. Just like the evaluation itself, the mid-term review exercise, is mainly focused on the development assistance provided and the need to use other aid modalities than BS. The political and risk analysis are limited. There is a rather general reference to one of the recommendations of the evaluation, i.e. a declaration on intent to explore the scope for project interventions in the domain of "rural development", especially for infrastructure and support to SMEs in rural areas.

In the follow-up response (two years later), the EU indicates that it will again consider the relevance of doing more in promoting "rural development" (e.g. by taking it up as a focal sector) during the next programming period "in view of the good results yielded during 2002-2006 and the positive feedback on the thematic projects in this area".

4. Qualitative dynamics

Analysis of the recommendations

Considering the fairly traditional "development" and "aid management" focus of the evaluation, the recommendations are concrete, logical and feasible. Yet the limited attention given in the evaluation to the "political dimensions" of the partnership and the effects of the crisis (including suspension of aid) also means a lot of issues have been "left out". This inevitably also reduces the relevance of the recommendations.

The recommendations do not really go beyond technical one-off adjustments to provide longer-term guidance about the fundamentals of the EU-Nicaragua partnership, especially the political foundations of the cooperation. There is no solid reflection as to what the EU can really do with governments which

do not respect fundamental values (as agreed upon in all kind of treaties and bilateral agreements) beyond the suspension of BS. There is no analysis of alternative channels to deliver aid (using CSOs or local authorities). The future choices of sectors is linked to technical aspects (EU added value and expertise, complementarity with other MS), not on the basis of a political economy analysis of potentially promising sectors. The evaluation also does not provide ideas on how to revitalise governance support in Nicaragua. In the follow-up response, the EUD limits itself to recognizing the challenges of operating in governance and seems satisfied that it could at least have an entry point, "a project with the National Police".

As mentioned before, the evaluation is only superficially grounded in a solid political analysis. The focus is really on delivery of cooperation. This means there are broad references to the political objectives underpinning the cooperation, quite some analysis of the EU responses strategies and then plenty of technical considerations (e.g. recommendations 4 and 5 on modalities of providing BS).

Within the logic (and limits) of a primarily development focused strategic evaluation there is a coherence between FC and the recommendations.

Scope

It could be argued that the non-adaptation of the ToRs to also consider the political crisis in Nicaragua (occurring at the time of the evaluation) was a missed opportunity for greater relevance and uptake. The same could be said of the apparent lack of follow-up of the "uptake clause" in the ToRs. It looked very promising to ask evaluators to globally assess the level of uptake of previous evaluations (Nicaragua was "over-evaluated" for a long time), yet this part of the ToRs has apparently been largely forgotten.

In terms of framing the evaluation in long-term country and co-operation dynamics, this occurs only to a limited extent, i.e. on where and how to spend aid in the future. The more fundamental aspects of the political partnership and the roles to be played by the EU (beyond coordination) are barely addressed.

4.5 Nigeria

Name of evaluation	Country evaluation Nigeria
Language	English
EuropeAid reference nr.	1276
Type of evaluation	Country level evaluation
Year of evaluation	2010
Documents reviewed	- Main report and Annexes - Executive summary - Summary document - Seminar presentation - Quality Judgement - Fiche Contradictoire

Key points:
- *In some cases conclusions are not backed up with sufficient data, whilst in others where it has been acknowledged that data has not been found.*
- *The report could have been better presented (both in written form and during the seminar).*
- *Some recommendations do not account sufficiently for the political context and that they may be difficult to implement.*

1. Process dynamics

Evaluation planning and timing

The main milestones were as follows:
- Inception phase, March-May 2009
- Desk Phase May-July 2009
- Field phase Sept-Oct 2009
- Synthesis phase Oct-Dec 2009
- Discussion seminar January 2010
- Final report dated May 2010.

Link to policy and programming

Part of the evaluation objectives was to look into whether recommendations from previous thematic and sector evaluations were taken into account in the current programming cycle.

Definition of evaluation needs

There were ten evaluation questions following the OECD/DAC criteria of relevance effectiveness, efficiency, impact and sustainability complemented by the EC criteria of coherence and 'EC added value'. EQs are in a text box on page 9 – EQ1 is about strategy, relevance and design, EQ2 is about efficiency, EQ3-EQ9 are about specific sectors and EQ10 is about coordination, coherence and complementarity. These came from the intervention logic. There is no mention of a reference group to help with the methodology and content.

2. Response process and dissemination

Fiche Contradictoire

Most recommendations were agreed with while only two were partially agreed with. In some cases recommendations affirm the Delegations work. In some cases, the FC recognises that the recommendation is not easily actionable or feasible - e.g. ownership is not merely increased through more active capacity building by external actors. It more appropriately says it needs to create the conditions for demand to emerge.

Some initial responses stated that new developments were already in line with the recommendations suggested. One response refined the recommendations in light of institutional changes in the EC (the merger of AIDCO and DEV and the creation of EEAS). In some cases, the response nuanced the recommendation and in some cases made detailed recommendations - providing a more holistic response. The report does sometimes recognise that the Delegation does not have the powers to recommend actions – e.g. in relation to staff recruitment.

In terms of the follow-up, when the recommendation is not rooted in the political context, and is unfeasible – the follow up becomes redundant. Follow up often shows that projects which follow recommendations have been instigated, but in one case it avoids talking about how the Delegation is coordinating EEAS and EuropeAid and talks briefly about how thematic projects are selected instead. In another case, the follow up highlights that HR continues to be inadequate.

Dissemination process

The report is well structured and presented (according to the quality grid). However, the QG suggested there are too many sub-headings, not enough visuals with poor links between the main report and the annexes. Before the end of the study a dissemination seminar was held in Abuja with all stakeholder groups and feedback from this seminar was reflected in the final report. No minutes were available in the report, but there is a set of slides from the presentation, which present findings by evaluation

question – which each include the logic (or log frame), judgement criteria, indicator, and an assessment against this, with a summary of limitations – there are a lot of slides as a result. It may perhaps have been better to start with the strategic recommendations (3 to 5) and retro fit the analysis from the evaluation. One page of the 2 and a half page summary is taken up with the subject of evaluation, its purpose and method. This does not seem to be particularly helpful to communicate key messages to busy staff, and it may be worth reformatting the summary document for country evaluations.

3. Programming dynamics

Follow-up in programming and the mid-term review

There is no information on this.

4. Qualitative dynamics

Analysis of the recommendations

There is little evidence of an evaluation of the quality of the analysis. The quality grid has been filled out in some detail, with concrete examples. Three aspects on the quality grid scored poorly (reliable data, sound data analysis and credible findings), whilst the others were good. In some cases conclusions are not backed up with sufficient data, whilst in others (where it has been acknowledged that data has not been found) no attempt has been made to find a substitute, nor have good enough explanations been provided.

There seems to be insufficient cause and effect analysis between interventions and their consequences. Comparisons (before and after, and amongst different countries) were lacking – there was also no mention of a theoretical counterfactual: i.e. what would have happened without the interventions? The context is not described and the conclusions and recommendations are not sufficiently rooted within the framework. The quality grid says the usefulness of the recommendations is good – well sequenced, and logically flowing from the conclusions.

Recommendations included: improving the support to specific partnerships, continuing assistance in focal sectors, promoting ownership (through more active approach to capacity building), adaptive designs and results based monitoring, better coordination between different parts of the EC which are providing funding, improving the profile of the EC in Nigeria (interestingly the perceived low contribution of the EC to specific outcomes is a profile issue not a substantive issue) and more support to regional integration. It is not clear however, how one promotes ownership through more active capacity building. Promoting ownership requires more political engagement, and perhaps less proactive capacity building. Additionally, it is doubtful whether shouting louder about an agency's aid is likely to get people to work together for the aid to be more effective.

Recommendations are clustered in four categories: principal recommendations of an overall strategic nature, general operational recommendations, sector specific recommendations and recommendations related to regional integration. The time frame for recommendations is the EDF10. The evaluation team found that all recommendations could be addressed without major difficulties. Most recommendations are addressed to the EC. But 8 out of 10 recommendations also involve national governments (so it is difficult to implement if they require government action). Recommendations are listed by priority from very high, to high and moderate importance. They seem to be fairly generic and seem to be based on a misunderstanding that building capacity can increase demand. There is very little talk of the context and political culture and the risk that these present to implementing recommendations. The analysis is overly technical.

There is little information on this.

With regards to the methods used – there was difficulty in accessing documents from the CRIS database, limited time for the field study, combined with belated Nigerian preparation and lack of knowledge of the results and impact of the budget line projects. The team developed compensatory measures for these constraints to the extent possible.

One of the conclusions states that impact could have been increased if there was better knowledge of the issues. This is a big assumption. A better conclusion might have been that interventions could be more practically implementable or politically supportable if knowledge of the issues was greater and better applied. It is impossible to predict impact, especially in a competitive and uncertain political system like Nigeria's.

4.6 Egypt

Name of evaluation	Evaluation of the European Commission's Co-operation with Egypt
Language	English
EuropeAid reference nr.	1281
Type of evaluation	Country level evaluation
Year of evaluation	2010
Documents reviewed	- Main report - Annexes - Methodology - Summary document - Quality Judgement - Fiche Contradictoire

See detailed process tracking study.

4.7 Ukraine

Name of evaluation	Evaluation of the European Commission's Co-operation with Ukraine
Language	English
EuropeAid reference nr.	1283
Type of evaluation	Country level evaluation
Year of evaluation	2010
Documents reviewed	- Main report - Annexes - Summary document - Fiche Contradictoire - Quality Judgement

Key points:
- *The evaluation could have been better timed – it was released just after the country strategic planning process and therefore opportunities were missed to consider the findings during important decision-making.*
- *Evaluation team could have been helped in their evaluation if they had been provided with monitoring data.*

- *The evaluation seems to have failed to capture the nature and impact of policy dialogue between the EC and the government.*
- *The recommendations frame issues technically and do not acknowledge that implementation requires the cooperation of a range of different actors – which is a political issue in itself.*
- *Findings do not seem to mention anything that the Delegation did not already know – rather they serve to affirm opinions in a relatively robust way.*

Process dynamics

Evaluation planning and timing

The report is dated as December 2010.

Link to policy and programming

It seems that the evaluation was undertaken just after CSP process, so at least one of the recommendations (to focus on social sectors) were redundant, as a decision had been made to focus on other sectors. Budget support had just started so it would have been difficult to make well-informed conclusions about that.

Definition of evaluation needs

The evaluation needs were not very clearly defined in the main report.

Response process and dissemination

Fiche Contradictoire

The response shows some buy in to the recommendations. It details what efforts are being made and what more needs to be done to address recommendations. The initial response for one recommendation showed that the issue was being addressed in the programming during the CSP process making the recommendation redundant, but also suggesting that the evaluation should have been done before the CSP process, rather than after. One response said the team did not understand the recommendation. Another response, about improving donor coordination is slightly defensive, saying the recommendation would be unlikely to work (it would have been useful to have the evaluator's response). Follow-up after one year indicates that real progress has been made in setting up structures for, e.g. high-level policy dialogue. However, putting too much emphasis on the aforementioned progress indicated would be inaccurate, as the Delegation seems to have been heading in the 'right' direction in any case. As such, evaluation findings inevitably endorse approach/direction of travel.

Dissemination process

There is no record of the dissemination process.

4. Programming dynamics

Follow-up in programming and the mid-term review

One of the recommendations suggests that ex-post monitoring exercises and evaluations of EC cooperation be carried out before formulation of CSP or NIP. But the response in the FC suggests that the Delegation does regular monitoring in any case.

5. Qualitative dynamics

There is mention of the participation of the reference group and the EC Delegation in Ukraine. All aspects of the QG score 'good'. The cause and effect links between interventions and their consequences could have been explained better. The nature and content of policy dialogue is not very well captured. A strategic evaluation is supposed to cover such issues – but the evaluation is largely based on documentation, and policy dialogue is not captured in documentation.

Analysis of the recommendations

There seems to be a disconnect between recommendations in the EVINFO summary and the recommendations in the executive summary of the main report which makes it a bit confusing as to which the FC responds to. The recommendations are clearly linked to conclusions. They are clustered and prioritised – recommendations give operational advice (QG). The recommendations suggest strengthening the outcomes through better policy dialogue, especially on PFM, and more attention on social development and improved efficiency of interventions (design, management as well as their implementation, coordination with other donors). Lots of recommendations are framed technically while ignoring the fact that advancements often need the cooperation of other government actors or donors, which is often very difficult. They also assume that the EC has the power to coordinate other donors.

Scope of the evaluation

There is no documentation which describes the scope of the evaluation.

4.8 Malawi

Name of evaluation	Country Level Evaluation: Malawi
Language	English
EuropeAid reference nr.	1297
Type of evaluation	Country level evaluation
Year of evaluation	2011
Documents reviewed	• Main report • Annexes • Summary document • Fiche Contradictoire • Quality Judgement

Key points:
- *It is not clear how/if the FC was quality assured: it responds selectively to the recommendations and the second column is not filled in.*
- *The evaluation reflects the intervention logic, but did not substantively question the wider political economy of aid to Malawi or the risks of non-delivery due to the termination of BS.*
- *However, at a technical level (i.e. if the political economy point above is ignored), the evaluation provided a useful validation of the choice of sectors.*

1. Process dynamics

Evaluation planning and timing

The evaluation took place between September 2010 and November 2011, with a field visit in May 2011.

Link to policy and programming

The dissemination seminar noted that the evaluation was well timed, as its results would feed into the strategy formulation for EDF11 in 2012.

Definition of evaluation needs

The evaluation fit the intervention logic outlined in the overview of EC-Malawi co-operation, in that it assessed the main focal sectors, which had been the focus of co-operation activities for over a decade. The report notes a striking consistency of support to Malawi throughout EDF 7, 8, 9 and 10 – the evaluation took place during the programming cycle for EDF10 – and all but one of the non-focal sectors identified in EDF9 were evaluated (the exception being micro-project programmes), with trade being evaluated though it only emerged in EDF10. Evaluation needs were well defined, setting out nine major issues for evaluation covering a wide range of issues:

- Relevance and responsiveness of EC co-operation, the breadth of co-ordination with government priorities and those of other donors, and whether it added value to other donor support;
- The effectiveness of its support to agriculture-led economic growth, including whether its policies and resources were used effectively to promote more sustainable practices, greater market competitiveness and mainstreamed cross-cutting issues such as gender and HIV/AIDS
- The impact the EC's programming made to food security at the national and household levels; ensuring that national food security policies were well managed, contributing to a diversified agricultural economy and providing an effective nutritional safety net for vulnerable groups;
- Whether road transport strategies were well targeted, considered national capacity, made direct contributions and particularly increased rural access;
- Whether the road transport policy was set in the context of a multi-modal transport policy and a broader need to improve trade via improving regional connectivity;
- Improving Malawi's ability to perform well in trade negotiations; improving its knowledge about trade flows, identifying the impacts of trade on the national economy, improving its ability to promote its own exports and supporting national involvement in regional programmed for trade & investment;
- Assessing whether BS has helped stabilise Malawi's internal and external financial balances; including the extent to which the predictability of aid flows has been increased by the timely release of tranches, and whether BS has contributed to better co-ordination of dialogue with other donors;
- Improving PFM and re-orienting Malawi's policy priorities towards poverty reduction; strengthening the PFM system so as to improve aid flows channelled through and managed by government financial systems and ensuring that national financial flows are channelled to health and education;
- Whether support to work on the rule of law and non-state actors has contributed to civil society's oversight of government and parliamentarian decision-making processes; raising NSA involvement in decision-making, contributing to an improved rule of law and regulatory framework, improving accountability and transparency;
- Whether the choice of aid modalities has improved the implementation of EC support; whether different aid modalities were considered, whether these modalities responded to emerging national needs and capacities of development partners, whether they were coherent with and

complementary to what other donors were doing and whether there were sufficient internal resources and capacities to use these modalities effectively.

The choice of issues to cover reflects the consistency of provision of aid to key sectors over the past decade, combined with the flexibility in aid modalities which has occurred over the same time.

2. Response process and follow-up

Fiche Contradictoire

The QJ scores the evaluation 'good' on all categories, which makes the issues raised below somewhat odd. Evaluations can be used to validate what is already being undertaken (which is perfectly legitimate) though given that the evaluation took place over a year it is not clear whether this is the case here, or whether the actions occurred as a result of the evaluation (or a mix of both). The FC notes several instances where the recommendations made by the evaluation are already happening (R1, R3, R6, R7, R11, R12).

The FC does not accord the same priority to the recommendations as the evaluation (the evaluation picks out four for special mention). Some of the detailed evaluation recommendations are not picked up in the FC (e.g. R3 makes detailed recommendations about improving the implementation design: a line by line response in the FC would not be necessary, but it could be useful to see how they have prioritised the detail within this recommendation. The FC has 'no comment' against R4, R14 and R15, which is odd considering that the evaluation uses the word 'unsatisfactory' against R4, and that R4 and R14 are linked (how can you improve your ability to address cross-cutting issues without a well-functioning M&E system?). The FC does not appear to be defensive, in that there are no major disagreements, but the observation can be made that it does not appear to give the same priorities to the issues as the evaluation does and that responding 'no comment' to three clearly stated recommendations is strange.
There is nothing in the second column of the FC: this does not appear to have happened. In contrast to other FCs, this one seems to be very thin.

Dissemination process

The dissemination seminar considered all the evaluation questions and recommendations. The minutes of the seminar are in note form, but it appears that comments/responses from participants were detailed, focused, and fully acknowledging where the report would be altered to take account new evidence or a different interpretation. The very initial response process during the seminar is difficult to follow.

There was some high-level disagreement with the evaluation's findings (notably around EQ9, about the rule of law and non-state actors), but primarily it seems that the comments from participants related to evidence or issues the evaluation had missed, rather than a major disagreement about the conclusions. This reinforces the observation that the QJ marked the evaluation as 'good' for all categories.

There do not appear to be any other uptake initiatives, though several recommendations are directed at EuropeAid as well as EUD. One in particular notes that limited staff capacity in EUD limits its ability to act effectively in so many sectors, and the FC notes that calls for projects are being centralised at HQ, the presumption being that this will free up considerable staff time to work on implementation rather than the project procurement process.

3. Programming dynamics

Follow-up in programming and the mid-term review

Two recommendations are reflected in the EDF11 programming guidelines:

- the EUD's internal capacity should be reflected in the choice and number of sectors selected. This does appear in the EDF11 programming document (Annex 3, section 2). Interestingly, the guiding principles for sector concentration and choice of sectors in EDF11 do not contain the words 'poor' or 'poverty'.
- there are risks in rapidly phasing out of a sector without an exit strategy: in fact the Malawi evaluation suggested reducing the number of sectors, the EUD response in the FC argued against this because of the risks.

4. Qualitative dynamics

Analysis of the recommendations

The recommendations appear to be feasible in the most part, though some (e.g. the recommendation to scale back the number of sectors the EUD worked in) fell in the realm of general guidance rather than specific recommendations. Otherwise the recommendations were concrete, with sub-recommendations in many instances. The QJ rates the evaluation as good against all its criteria, though notes three areas where the evaluation could have been strengthened:

- Identifying and analysing unintended or unexpected effects
- More emphasis could have been given to data from interviews
- The recommendations would have been improved if responsibilities to act had been given together with a timeline for implementation.

On the whole the EUD appears to have been reservedly satisfied with the evaluation. The recommendations do appear to go beyond the simply technical: there is nothing in the QJ or the FC to suggest that the recommendations are in any way lacking in this detailed knowledge. This may partly result from the consistency of the EUD's programming, which would give a consistent and coherent evidence base from which to draw their conclusions: it would have been relatively easy to get hold of the key documents.

Given that the evaluation questions were not directed at political issues, no political recommendations were identified. Some of the strategic issues addressed by the evaluation touch on political relationships between GoM and donors (e.g. the recommendation about better dialogue and co-ordination and the observation that with the IMF declaring Malawi to be off-track and withdrawing BS facilities, now might be the time to engage on key issues), but, primarily, the recommendations are technical in nature; there is little in them which examine the agents, structures and institutions which form the backbone of a political economy assessment.

Scope of the evaluation

The evaluation was closely linked to the list of EU interventions outlined in the various EDF cycles (7, 8 and 9) and to the intentions of EDF10. The primary structure of the evaluation is around the sectoral issues highlighted in EDF9, with additional questions around the coherence of EU aid, how well the different aid modalities work and whether sufficient attention had been paid to internal capacity to administer the range of aid modalities. The scope seems to be absolutely right in terms of how the provision of EU aid has been broadly consistent (in sectoral terms) and has been modified (in terms of delivery mechanisms) over the time period for the evaluation.

The evaluation did frame its evaluation within a long-term country and co-operation dynamic, in that it reflects what has been quite a stable set of co-operation activities over the past decade. The political economy of Malawi does not appear to be addressed, though the politics of aid disbursement (i.e. the fact that at the time of the evaluation the IMF had just declared Malawi to be off track) was identified as a useful window of opportunity to improve dialogue, keeping in mind that this meant that disbursement of BS would be suspended.

There is little in the evaluation which specifically looks at the political economy of Malawi, though there is a note that with the suspension of BS as a result of the IMF's declaration that Malawi was off track. This would be a good time to enter into dialogue on the broad themes of both economic and democratic governance. In its evaluation of whether EC strategy is responsive to national priorities (judgement criterion 1.3), the evaluation found that EDF9 & EDF10 CSPs and NIPs presented a 'sound and complete analysis' of the political, economic, social and environmental situation in Malawi. The evaluation appears to have understood that the issues it was asked to assess therefore reflected the local political economy, and did not make any further judgements on this.

However, the recommendations are on the whole fairly technical ones: though there is some analysis of the political economy of the agricultural or roads sectors (for example), this does not strongly make it through into the recommendations. This is in spite of EUD programming dealing with land reform, agricultural credit, irrigation, extension services, and a host of other issues with a potentially high degree of vested interests, clientelism, conceptual controversy and competition between state and non-state actors. The evaluation does state (Annex 2, p. 25) that "EC support attempted to address reform requirements of the agricultural sector with limited success as assumptions were frequently overoptimistic regarding political will, GoM capacity to translate policies into action and existing capacities within GoM." And key sector reforms such as land reforms or the establishment of rural credit institutions are not included into this new framework. EC sector support is therefore rather decreasing its support to key sector reforms, as political will by GoM currently does not seem to prioritise these long-term complex changes." (p. 26) and "The trucking industry is undoubtedly a powerful economic force with anecdotally political interest/influence." (p. 147) but there is no further analysis of how the local political economy could affect programming and implementation.

4.9 Central Africa

Type of evaluation	Regional
Name of evaluation	Evaluation de la stratégie régionale de la CE en Afrique Centrale
Language	French, English, Spanish, Portuguese
EuropeAid reference nr.	1037
Year of evaluation	2009
Documents reviewed	• Main report (French)
	• Annexes
	• Executive summary (English, Spanish, Portuguese)
	• Summary document
	• Fiche Contradictoire
	• Quality Judgement

Key points:
- *Timing of the evaluation was purposefully chosen to ensure inputs into the EDF10*
- *Political economy considerations appear to have been broadly overlooked, which may have limited the opportunities for uptake*
- *Evaluation needs appear to have been well-defined, but the evaluation was limited by a lack of local ownership which made data collection and interviews difficult*
- *The response in the FC suggests that recommendations were in line with what was already being planned for the EDF10, but in this reviewer's opinion did not go far enough in challenging the realism and feasibility of the EC's co-operation strategy (because of the weak political economy considerations).*

1. Process dynamics

Evaluation planning and timing

The timing of the evaluation (mainly 2006, with submission of final report in December 2006) was chosen on purpose to ensure inputs into the programming of the next cycle of regional cooperation strategies (EDF10). It could therefore be argued that this is an example of an "appropriate timing" of strategic evaluations –a key factor in terms of facilitating uptake.

Link to policy and programming

There is no link to policy production, but there is an explicit linkage to the programming cycle. The evaluation was supposed to draw lessons from the 8th EDF (1996-2001) and the then on-going EDF9 (2002-2007) to feed the programming process of the EDF10. In January 2006 the EC had organised a first dialogue with its regional partners in Brazzaville to give its first orientations on the new programming cycle. The further preparation of the EDF10 coincided with the evaluation process, providing in principle a conducive context for uptake of lessons learnt.

From the documentary analysis it would also appear that the EC had also a clear idea on the desirable focus of the evaluation. The ToRs clearly define the sectors on which the evaluation should concentrate (both longstanding focal sectors and "new" issues such as governance) as well as the type of issues to be considered (e.g. complementarity between regional and national indicative plans).

Definition of evaluation needs

They are well-defined considering the explicit plan to use the evaluation as a basis for the next programming cycle. It resulted amongst others in a rather limited set of EQs (5) and the development of a number of "hypotheses" (9) related to the EU interventions in the region, to be tested during the evaluation.

Regional evaluations of course face important logistic challenges as they have to cover several countries and reach out to a set of actors scattered around different places. In Central Africa these factors were even more constraining than usual (fragmentation of regional bodies, problems of insecurity, travel facilities, etc.).

It is interesting to note that among the main limitations, the report also mentions the difficulty of having interviews with key stakeholders. Apparently the team missed many opportunities to meet these actors either because they were not aware of the existence of regional meetings or because the EC services were not always helpful in communicating properly about these gatherings, facilitating the access of the evaluators and explaining the objectives of the evaluation.

2. Response process and dissemination

Fiche Contradictoire

Globally there is a fairly large buy-in, expressed in the form of fairly generic commitments to ensure follow-up. It would appear that the evaluation largely confirmed things that the EC was already very much aware of (e.g. the negative effects of the fragmented institutional landscape of regional bodies) working on (e.g. the EPA processes) or considering as a required change (e.g. a stronger focus on governance as a cross-cutting issue). So the evaluation does not appear to have raised "controversial" issues or brand new things, and this undoubtedly facilitated a broad convergence between evaluation recommendations and on-going actions at EU level.

The FC does not explicitly say "agree" or "disagree" with the various recommendations. It provides a rather long (often descriptive) narrative of general actions the EC is envisaging. Also the commitments for follow-up are there but in quite generic terms. But it is never defensive. In the follow-up after one year, the EC recognises the need to further think through its overall understanding of regional integration processes and related cooperation strategies AND to review more downstream implementation issues.

Yet when reading through the responses of the EC one notes again a relative lack of political analysis/realism. Both the evaluation report and the EC responses recognise this is probably the most fragile and poorly governed region. When one then looks at the strategies the EC wants to promote in that region there is a strong feeling that these policies will not be implementable. This holds particularly through for the EPA approach advocated by the EU (with hindsight, seven years later, the cooperation with that region is even more problematic and less successful than now).

It raises the question whether strategic evaluations "go deep enough" into issues such as the *relevance* of proposed EU strategies as well as the *feasibility* and *risk analysis*. These dimensions are present in the report, but yet they would perhaps need to be put more at the centre so as to allow for even more realistic and down-to-earth recommendations.

Dissemination process

No evidence that much dissemination took place. No indication of a seminar in Brussels or in the field (and probably this is quite difficult to organise, amongst others because of the cost). Yet the report mentions that the (2) focus groups organised in two countries to discuss the key hypotheses related to the EU intervention logic were quite formative moments.

Another real problem affecting strategic evaluations are regional level is the limited amount of actors involved. The report mentions the scarcity of HR involved in these processes as well as their relative isolation of staff working on the much more high profile national programmes. So the target group for "uptake" of this type of evaluations is also likely to be quite limited.

There is no evidence of other uptake initiatives.

3. Programming dynamics

Follow-up in programming and the mid-term review

The response (1 year later) of the EC suggests that most of the recommendations were in line with what the EC was already planning to do anyway in the EDF10.

4. Qualitative dynamics

Analysis of the recommendations

The recommendations can be considered to be globally sound. But, as mentioned above, they could have gone much further in challenging the realism/feasibility of the overall EC cooperation strategy in that region and in pushing the EU to lower its ambitions.

The recommendations are a mix of longer-term guidance based on a good knowledge of the region's realities (e.g. the recommendation to put governance at the centre of all interventions) and proposals of a more operational nature.

In terms of the ration of political, strategic and technical recommendations there is a good balance, though the political analysis could have gone deeper and led to even more down-to-earth recommendations regarding feasible intervention strategies. But, within the logic (and limits) of a primarily development focused strategic evaluation there is a coherence between FC and the recommendations.

Scope of the evaluation

The evaluation does go beyond its primary mandate as the two first set of recommendations challenge the EC to deepen its overall knowledge base on regional integration processes in Central Africa and rethink elements of its cooperation strategy.

The report partly manages to frame its evaluation in long-term country and co-operation dynamics.

The report seeks to go to the "essence" and question to some extent the overall relevance of the EU regional strategy in Central Africa. But in the authors' view it stops a "halfway" in terms of really drawing profound lessons of past cooperation thus possibly reducing the scope for a more fundamental rethinking of cooperation strategies.

4.10 Central America

Type of evaluation	Regional
Name of evaluation	Evaluación de la cooperación regional de la CE en América Central
Language	Spanish (evaluation); English (Summary), French (ToRs, FC and QJ)
EuropeAid reference nr.	1092
Type of evaluation	Regional evaluation
Year of evaluation	2006-2007
Period under review	1996-2006
Documents reviewed	• Main report • Annexes • Executive summary (English & French) • Summary document • Fiche Contradictoire • Quality Judgement • Regional Strategy Paper • Special supplement

Key points:
- *Timing was not ideal. The evaluation took place when the 2007-2013 regional programming was well advanced. However, since the RSP was not yet finalised, it was not included in the review*
- *The recommendations are feasible and primarily technical, no major strategic shifts are proposed. This is partly because of the regional integration dynamics in Central America are at the centre of this evaluation. The SICA, being a largely European financed institution (Spain and EU) has very clear objectives that are in line with the regional integration agenda of the EU. The more sensitive political and cooperation matters are left to the bilateral EU cooperation with Central American countries.*
- *While the evaluation touches on key regional and programming dynamics, the EC response to the evaluation indicates a lack of synergy with the two main relevant processes: EU-CA trade negotiations, and 2007-2013 programming.*
- *The evaluation is very focused on the regional cooperation strategies and interventions. Bilateral dynamics are largely ignored.*
- *There are some signs of uptake of evaluation results in the midterm review of the regional strategy paper (2011).*

1. Process dynamics

Evaluation planning and timing

The evaluation was part of the 2006 evaluation strategy. The evaluation was carried out in 2006-7 and covers the EC's cooperation with Central America in 1996-2006. The field phase took place in February 2007. The final report is dated July 2007.

Link to policy and programming

The 2007-2013 Regional Strategy Paper mentions that a comprehensive evaluation was in the making, though results were not available during the preparatory phase of this RSP, which took place in 2004-2005. The RSP does however mention lessons learned from the midterm review of the regional evaluation of EC support to Latin America (2005).

The ToR of the evaluation mentions that the "relevance, coherence and complementarity of the EU regional cooperation for the period of 2007-2013" is part of the scope of the evaluation, under the condition that the RSP would have been approved during the documentary phase of the evaluation. This was not the case.

In broad strategic terms, the objectives and priority areas of the 2007-2013 regional strategy are largely in line with the evaluation's recommendations. Especially the overall objective of "supporting the process of political, economic and social integration in the context of preparation of the future Association Agreement with the EU" is very similar to recommendation 2 of the evaluation "Adapt the Commission's strategy to the requirements of the EU-CA association agreement and CAFTA-RD".

Rather than an instance of uptake this is more the result of the fact that the evaluation and the preparatory work of the regional strategy were carried out within the same time frame, and therefore respond to similar political developments. The EU-CA trade agreement that was signed in 2011 had been in the works for quite some time and negotiations started in 2007. At the time of the evaluation it was clear that the partners aimed for a conclusion within the timeframe of the 2007-2013 MFF.

The regional dimension in Central America is also a very particular one. For a Latin American context, the formal integration process under the auspices of the Sistema de Integración Centro-Americana (SICA) is quite mature (customs union). Regional integration in Central America has been supported (and financed) by the EU since the very early days of the SICA. Even if political traction to further deepen regional integration remains highly uneven to this day, this gives the EU **one key institutional interlocutor** for regional cooperation and trade negotiations. Both the evaluation and the RSP recognise this.

Definition of Evaluation needs

The Terms of Reference set out standard evaluation needs with regard to the regional cooperation strategy (relevance, coherence, implementation, etc.). These are taken over in the evaluation itself. The Quality Judgment grid lauds the evaluators for carefully examining the role of Civil Society and the differentiation of support pre and post hurricane Mitch. Both of these aspects were mentioned in the ToRs, though not as separate components of the evaluation scope.

2. Response process and dissemination

Fiche Contradictoire

The Fiche Contradictoire, rather than taking over the literal translation in the Executive Summary, interprets the recommendations of the evaluation. The recommendations are generally seen as valid and feasible, but the responses clearly reflect the disconnection between the evaluation and the two major programming and policy dynamics: (1) the 2007-2013 Regional Strategy, and the negotiation of the EU-CA Association Agreement (concluded in 2011). Most of the responses indicate how the EU already addresses the recommendations in its regional cooperation and how it addressed many of the thematic components of the regional strategy recommendations in its bilateral cooperation.

Dissemination process

No dissemination documents are included in the annexes or the EuropeAid evaluation page. Apart from the standard EVINFO no details on the dissemination process are available for the document review.

3. Programming dynamics

Follow-up in programming and the mid-term review

The timing of the 2007-2013 regional strategy did not allow for a formal uptake of evaluation results. The preparatory work of the RSP partly predates the evaluation, and the evaluation results were not yet available when the RSP was drafted.

The Midterm review of the RSP and the Regional Indicative Programme for 2011-2013, however, directly refer to the evaluation: "The evaluation of regional cooperation during 1996-2006 demonstrated the importance and high relevance of EU regional cooperation for strengthening the

integration process in Central America. It stressed in particular the need to improve CA ownership of programmes, to respond to requirements arising from the free trade component of the Association Agreement, and to strengthen the environmental and risk dimensions of future cooperation, along with social cohesion and security."

On this basis the MTR suggests a number of quality improvements. The first of which is to work towards "sustainable finance mechanisms" for regional integration in Central America. This is **fully in line with recommendation 4 of the evaluation:**

"The cooperation should reflect the quality of the partner relationship. It is important that this cooperation be commensurate to the level of commitment and participation that national and regional institutions themselves show in the process of integration. It's equally important that the cooperation corresponds to the requirements and provisions of the Association Agreement EU-AC. Therefore, it is recommended to link the regional cooperation to the implementation of plans and programs by the regional institutions, to the effective participation of the Central American states in the self-financing of its agencies, and to the efficient and transparent management of resources "

4. Qualitative dynamics

The quality judgment awarded the evaluation an overall score of "good". Comments are mainly methodological and formal. The Commission accepted the evaluation results and conclusions as credible and valid, though criticised the evaluators for (1) weak methodology, (2) unsound analytical process, and (3) lack of clarity in the report.

Analysis of the recommendations

The recommendations are feasible and concrete. This is restated in the quality judgment, with reservations regarding the high number of recommendations and the lack of prioritisation. The recommendations are feasible and primarily technical, no major strategic shifts are proposed. This is partly because of the regional integration dynamics in Central America are at the centre of this evaluation. The SICA, being a largely European financed institution (Spain and EU) has very clear objectives that are in line with the regional integration agenda of the EU. The more sensitive political and cooperation matters are left to the bilateral EU cooperation with Central American countries.

Scope of the evaluation

The ToRs set out a broad evaluation scope for the evaluation, which includes analysing complementarity between national and regional interventions: "the relevance, coherence and complementarity of the Commission's cooperation with Central America (national, regional and thematic) for the period on 1996-2006."

The evaluation is very focused on the regional cooperation strategies and interventions. The complementarity question was partly integrated in the evaluation questions, but only with regard to "preparing the grounds for trade" (EQ4). EQ6 (transversal issues) and EQ8 (complementarity with other donors) miss this dimension, which is reflected in the response of the Commission services.

4.11 ASEAN

Type of evaluation	Regional
Name of evaluation	Evaluation of EC co-operation with ASEAN
Language	English
EuropeAid reference nr.	1262
Type of evaluation	Country level evaluation
Year of evaluation	2009
Documents reviewed	• Main report • Annexes • Summary document • Fiche Contradictoire • Quality Judgement

1. Process dynamics

Evaluation planning and timing

The evaluation was part of the 2007 evaluation programme as approved by ERDC. The report is dated June 2009. There appears to be close involvement of the reference group as well as the Joint Committee, which permitted high-level ASEAN MS representatives to provide feedback.

Link to policy and programming

The documentation provides no information on this.

Definition of evaluation needs

The evaluation is based on 9 evaluation questions, relating to the five DAC evaluation criteria plus coherence and value added. Evaluation questions, judgement criteria and potential related indicators were drafted based on the intervention logic, and discussed with members of the reference group and the evaluation managers.

2. Response process and dissemination

Fiche Contradictoire

The Fiche Contradictoire is slightly defensive in places, but with some justification. In some places the response is that "we're already planning to do something to address this". The response shows how projects, which have been set up at the time of the evaluation, are progressing and addressing issues raised in the recommendations.

Dissemination process

Preliminary conclusions and recommendations were presented to the 17th ASEAN-EC joint cooperation committee meeting in Feb 2009, permitting a range of stakeholders in particular ASEAN member state government representatives and ASEC officials. Discussions during this meeting were integrated into the report.

3. Programming dynamics

Follow-up in programming and the mid-term review

The documentation provides no information on this.

4. Qualitative dynamics

The evaluation followed the guidelines set by the joint Evaluation Unit. Some of the language is not completely clear and statistics on actual disbursements were difficult to obtain (even in case of projects which had been completed). ROM monitoring reports do not cover the entire period.

Analysis of the recommendations

The recommendations are split into 'improving cooperation' and 'thematic'. The recommendations in the EVINFO summary are not exactly the same as the ones pasted into the FC. In the conclusions and recommendations of the executive summary in the main report there are 5 global conclusions and 5 recommendations. The assessment of the EC programme is very ambiguous and vague (most probably intentionally so). It is not entirely clear on the difference between a lesson and a recommendation: this needs to be defined/clarified. Lessons are often a call to action, which is similar to a recommendation. To that end, the quality grid states that the recommendations are very general and hard to realise. A couple of examples are given (QG).

Recommendations do not acknowledge that making changes in say HR will not, on its own necessarily lead to better outcomes. Moreover, while some people say that risk needs to be better stated others will say that recommendations need to be better contextualised. Hence, recommendations do not acknowledge the often difficult institutional context in which the EC has to work.

This and other evaluations tend to talk a lot of improving visibility of EU work – however, it is not clear if this is something which the evaluators have looked at critically – e.g. how will visibility equate to better traction of reform efforts and dialogue? Also, in relation to suggestions of the EU reaching out better to the public - is there an analysis of the merits and drawbacks of doing so? Do these sorts of things interest the public (ASEAN integration) as a whole?

4.12 Water & Sanitation

Type of evaluation	Sectoral
Name of evaluation	Thematic evaluation of the water & sanitation sector
Language	English
EuropeAid reference nr.	965
Type of evaluation	Country level evaluation
Year of evaluation	2006
Documents reviewed	• Main report
	• Annexes
	• Summary document
	• Minutes of dissemination seminar
	• Fiche Contradictoire
	• Quality Judgement

Key points:
- *Absence of reliable data meant for inadequate methodology*
- *Analysis at a country level seem to remain at a general level*
- *Some recommendations seem to validate the way in which the EC is working in the sector whilst others are too vague and don't assess their feasibility and how they might be implemented.*

1. Process dynamics

Evaluation planning and timing

The launch note approved in December 2004 whilst the report is dated July 2006.

Link to policy and programming

Many of the recommendations were taken care of in the EDF10 it seems.

Definition of evaluation needs

The report addresses 9 evaluation questions, relating to five DAC criteria, the 3 Cs and crosscutting issues.

2. Response process and dissemination

Fiche Contradictoire

The initial response is fairly brief. Often, the recommendations suggest continued work in a similar vein. In some cases, the FC is slightly defensive, but agreeing with the recommendation – suggesting that issues are already being addressed. In some cases there is simple agreement. Follow up is more detailed (perhaps a different author completed this). In some cases, it states challenges they faced in implementation and how they have been overcome. Feedback from DG ECHO field experts contributed to the evaluation of proposals submitted to the last round of the Water Facility programme. Implementing regional water resource management principles and approaches are well known but very difficult, as it requires the involvement and coordination of experts from various sectors.

Dissemination process

The report is easy to read and logically structured. A seminar was held on 7th Dec 2006 in Brussels. Participants included the Evaluation Unit, the study team, representatives of DG AIDCO, DG RELEX, the water facility group, ECHO and the NGO sector. A member of the cabinet of the development commissioner, the Head of the Evaluation Unit and several members of the study team were present.

3. Programming dynamics

Follow-up in programming and the mid-term review

No information on this

4. Qualitative dynamics

In the QG, the evaluation scored poorly on reliability of data and sound analysis. The qualitative data was not backed up with quantitative data. Proxy indicators were not used when indicators were not sufficiently useful. This is to do with the composition of the set of indicators which was defined to confirm the proposed judgement criteria. Moreover, only selective use was made of the proposed tools: many statements were not based on factual evidence. Moreover, analysis in several country cases remained general. Problems might be due to the limited resources available to conduct the evaluation. Data related indicators were not available, making answering some of the EQ's very difficult and dependent on secondary data – hence the evaluation team had to adopt a cautionary attitude in presenting findings. Internal and external assessors were employed to ensure quality assurance. The internal assessor provided advice on the structure of evaluation, a set of tools for collecting and analysing information and ensured consistency of all outputs. The external assessor reviewed the quality of the final draft output.

Analysis of the recommendations

The recommendations follow logically from the conclusions and seem to be impartial. Although some are vague, e.g. what is to 'advance gender awareness' and how can it be done. It is difficult to implement thematic evaluation recommendations given that they have to be carried out by a multiple country Delegations. Many describe what not to do, but not how to do them. There is some talk of

'mainstreaming' – which is a very loaded term. There is a suggestion that recommendations are very general (seminar minutes) – one participant suggested that the thrust of several recommendations of this evaluation could be applied to other sectors like the transport sector. Finally, recommendations are not presented in any order of preference (seminar minutes)

Scope

The evaluation scores very good on 'scope' in the QG – it looks into relations with other EC policies as well as coordination and harmonisation with other donor interventions and partner government policies.

4.13 Energy

Type of evaluation	Sectoral
Name of evaluation	Evaluation of EC support to partner countries in the area of energy
Language	English
EuropeAid reference nr.	1192
Year of evaluation	2008
Documents reviewed	• Main report • Annexes • Executive summary • Summary document • Fiche Contradictoire • Quality Judgement

Key points:
- *The evaluation was extremely heterogeneous in scope; despite this the report is very well presented*
- *Most recommendations are process oriented, not surprising given that this is a thematic evaluation where content related issues are highly context specific*
- *Recommendations seem to have been followed up, but difficult to know to what extent this could be attributed to the evaluation.*

1. Process dynamics

Evaluation planning and timing

The evaluation was published in April 2008 and evaluates the 1996-2005 period

Link to policy and programming

No information on this

Definition of evaluation needs

The methodology section states how the evaluation is extremely heterogeneous in scope – having to evaluate activities contributing to quite diverse objectives – from ensuring the security of the supply of fossil fuels for Europe, enhancing nuclear security in former soviet union countries to facilitating access to energy for the poor in ACP countries. Heterogeneity was also apparent in the multitude of changes which took place in the sector during the evaluation period as well as the vast geography the evaluation was expected to cover.

The evaluation was based on 10 evaluation questions, relating to the 5 DAC evaluation criteria plus the 3 Cs - coordination, coherence and the value added of the Commission supported interventions.

2. Response process and dissemination

Fiche Contradictoire

This should be the responsibility of Thematic Unit on Energy – but the evaluation reports on the general doubt of the actual impact of DG Energy and their support efforts. All 22 recommendations are addressed. Most recommendations are agreed with (with nuances in some cases); some are partially agreed with, and there is one with which the commission services disagreed. Most recommendations are process oriented, there are a few which are about content – and in one case, there is strong disagreement from those filling in the FC, saying appropriate responses/policies will be context/case specific – global recommendations are thus not always appropriate.

Some recommendations present opportunities, whilst others acknowledge that they have already been taken up validating what is being done. There is a discussion of the pros and cons of the recommendations – e.g. on the need for more resources. In some cases, there is an indication that the service manager would like to see firmer (rather than tentative) recommendations, or that the scope of the recommendations should be expanded. Some disagree with the recommendation partially (e.g. on flexible instruments) saying that they have other problems (indicating perhaps a lack of understanding on the part of the evaluator?). A recommendation might be agreed with – but it will not say how they plan to do so – e.g. yes, lesson learning is good, but then they do not go on to elaborate on what they plan to do and how

On the follow up after one year – there is acknowledgement of recommendations being implemented – but as with all the other FCs without further analysis, we can not tell what factor the evaluation played in contributing to its implementation. Sometimes the response is fairly political, not providing details on how the recommendation was followed up. Not all recommendations were adapted after one year. In some cases, there is an explanation of the some of the politics on the ground which hinders recommendations from being implemented well

Dissemination process

No information on this

3. Programming dynamics

Follow-up in programming and the mid-term review

No information on this.

4. Qualitative dynamics

Analysis of the recommendations

The recommendations section is 22 pages long, containing 21 specific recommendations split into 5 subsections, each with a number of specific recommendations. There are recommendations aimed at high-level senior managers: such as those on formulation of energy policy as well as those at country level such as those on identifying and prioritising cooperation objectives. But, each stage of the programming cycle affects people at many levels in the hierarchy. There is no explicit mention of to whom the recommendations are targeted – although the QG says that recommendations are addressed to specific institutions. The recommendations address: 1) formulation of energy policy; 2) mutual understanding; 3) delineation of cooperation area; 4) prioritisation of cooperation objectives, resource allocation; 5) programming and implementation and 6) monitoring – mainly process type recommendations. However, in the EVIPNet summary there are four key recommendations – 1) on adopting more formal steps in its cooperation with partner countries 2) pay more attention to sound policies 3) develop up to date knowledge management systems and 4) draw lessons from experience through M&E.

The QG says that the evaluation could have assessed the impact of EC projects on poverty through indirect means/field observations. The evaluation draws on a strategic sectoral paper at country level, ROM reports, interviews and focus group discussions. The QG strongly praised the design of the methodology, analysis, findings, scope, data reliability, validity of the conclusions (very good), usefulness of recommendations, and clarity of reporting. The body of the report is short, concise and easy to read: structure is easy to follow; the summary is clear and presents the main conclusions and recommendations in a balanced and unbiased way. The table, which presents the link between the conclusion and recommendations, is commended – the executive summary is very well done (QG).

4.14 Education

Type of evaluation	Sectoral
Name of evaluation	Thematic global evaluation of European Commission support to education in partner countries (including basic and secondary education)
Language	English
EuropeAid reference nr.	1282
Type of evaluation	Country level evaluation
Year of evaluation	2010
Documents reviewed	- Main report vol. 1, 2a, 2b, 2c, 2d - Fiche Contradictoire - Quality judgement - Summary document - Minutes of dissemination seminar - Presentation slides (1: findings, 2: conclusions, 3: recommendations)

See detailed process tracking study.

4.15 Good Governance

Type of evaluation	Thematic
Name of evaluation	Thematic evaluation of the EC support to good governance
Language	English
EuropeAid reference nr.	884
Year of evaluation	2006
Documents reviewed	- Main report - Annexes - Summary document - Presentation seminar - Fiche Contradictoire - Quality Judgement

Key points:
- *Strong demand from different EC Units to carry out an evaluation in the early stages of supporting governance reforms meant that the ToRs were defined by a multi-actor group, and the EQs for the contractors to address*
- *The literature review helped frame the field of governance from the EC perspective (again, because the evaluation was commissioned in the early stages of supporting governance reforms).*
- *Within HQ the evaluation was clearly seen as a learning exercise from the outset with full buy-in from several different Units, which facilitated uptake of the results. However, some EUDs saw it as an audit of their performance and were somewhat defensive*
- *The political, strategic and technical recommendations form a single package and are related to each other.*

1. Process dynamics

Evaluation planning and timing

2005-mid 2006 (final report submitted June 2006).

This evaluation came at a time when the EC, like other donors, was expanding its activities in the field of governance (understood in the wider sense to include six clusters - democracy, HR, rule of law, civil society development, decentralisation and public administration/civil service reform). The issue of governance had gained prominence from the mid-1990s onwards – following rather apolitical/technocratic approaches to development cooperation. Ever since, some policy development on governance had taken place at the EC level (e.g. in the framework of negotiating a successor agreement to the Lomé Conventions or in specific areas such as the role of non-state actors) and in the field: the EC was experimenting with the first generation of governance support programmes in different country contexts.

Simultaneously, there was a strong demand within different EC Units to carry out a strategic evaluation on the early experiences in support of governance reform. A clear indicator is that extensive inter-service consultations took place on the ToRs of this evaluation. This multi-actor group even defined 8 EQs (reflecting the interests of different Units) which the consultancy team would have to address – a quite unique case of a "demand-driven" evaluation. (note that the evaluation team was free to add two EQs of its own choice).

Links to policy and programming

It was envisaged that the evaluation would take stock of the solidity of the EC policy frameworks, related to governance, and identify areas where future policy development would be required. To this end, the evaluation team was asked to carry out quite an extensive and "state of the art" desk study, on EC approaches to governance. This had not really been done before in a systematic way. In practice, it meant the evaluation also helped somehow to "frame" the whole governance field from an EC perspective (a by-product which was highly appreciated by the Evaluation Unit).

There is evidence of uptake of the evaluation to foster further policy development. The Quality Grid explicitly recognises that "this exercise has been useful for the Commission and the recommendations have been taken into account for the elaboration of the Communication on Governance (COM) 2006, 421".

This evaluation was clearly conceived as a learning exercise on a topic that had quickly moved to the forefront of the development agenda. Yet, it was also the evaluation's intention to ensure that the lessons learnt could be used to integrate governance more visibly and coherently in the various areas of cooperation programmes[2].

Definition of evaluation needs

Considering the extensive preparatory work by the inter-service group (facilitated by the Evaluation Unit) there was not only considerable "ownership" of the evaluation, but also a clear expression of needs.

Another factor that proved very helpful from an ownership and uptake perspective was the recent creation of a Thematic Unit on Governance in EuropeAid (E4) with the explicit mandate and resources to promote the operationalization of the new governance agenda (in all its components). From the outset, this Unit and its HoU championed this evaluation as it was clearly in their interest to have outcomes which could be used to develop the profile and mandate of the young Unit.

---

[2] For the purpose of the evaluation governance was seen as "a key value" of EU cooperation; a theme to be mainstreamed across sectors/instruments; and a direct field of assistance (i.e. dedicated governance programmes).

Among the limitations of the evaluation was the defensive attitude of several EUDs visited. They mainly saw this as an audit of their performance rather than as a collective learning exercise.

2. Response process and dissemination

Fiche Contradictoire

All in all there was a full buy in from several Units, as reflected in the FC.

The combination of high levels of ownership at the start of the process, the existence of a "champion" within EuropeAid (the E4 Unit), and the growing political demand for a stronger, and more effective EC stance on governance, all helped to ensure high buy-in for the evaluation process within the EC. The enthusiasm was less visible on the side of RELEX (poor attendance at RG meetings, partly as a result of a major restructuring process) and even lower on the side of partner countries (the team was even refused a visa to do a case study in Algeria).

There were also different uptake streams, which emerged from this evaluation (the consultancy , an organisation working in a structured manner on governance issues in the long-term provided the Team Leader for this evaluation). Two examples stand out:

1) One of the key findings outlined the limited integration of governance issues in EC sectors of operation (e.g. health, education, transport). The Thematic Unit E4 picked up this issue and started to develop a whole stream of work on sector governance which would culminate in the publication of a widely disseminated and used AIDCO reference document on "Assessing and addressing governance in sector operations" (2008).
2) The Cabinet of Development Commissioner Michel also used the evaluation to take bold new initiatives in relation to governance, including an experiment, which would turn out to be much less successful than expected (i.e. the so-called "Governance Facility" worth 3 billion euros for the ACP countries and based on an incentive mechanism)[3].

In this case the follow-up was a combination of these response strategies. Most of the recommendations found their way into the new Communication and into new programming instruments (such as the Governance facility), as one can see from the FC.

Dissemination process

All along the process there was quite some interaction with key stakeholders at the EC level and in the field – this was of course linked to the rather sudden high-priority status of the governance sector in EC discourse and the many operational questions EU Delegations had on how to work on this agenda in their country/region.

The dissemination seminar was attended by a wide range of stakeholders from different Units (RELEX, DEV, AIDCO) as well as external actors.

The evaluation was also presented at the 2006 European Development Days .

The fact that the consultancy had provided the Team Leader to some extent facilitated the uptake of the evaluation as the foundation was committed to very actively pursuing the implementation of several recommendations. This is of course not something consultancy firms can do. But, their mandate positioned them as "knowledge brokers" enabling them to increase awareness of for this evaluation and engage in a strategic partnership with AIDCO E4 Unit ensuring concrete follow-up on key recommendations (on decentralisation, sector governance, non-state actors, African mechanisms for promoting governance, etc.).

---

[3] The same consultancy did an assessment of the Facility in 2011 and it turned out to be a very poor instrument. With hindsight it could be argued that this whole initiative was an example of a "bad uptake" of evaluation findings.

3. Programming dynamics

Follow-up in programming and the mid-term review

It is difficult to assess the impact of this thematic evaluation on programming in the framework of a desk review. Some indicators may help to identify possible elements of follow-up "over time"

1) For two of the case studies done for this evaluation (Jordan, Angola) the EUD was very keen to engage in a reflection process, because they had just been through a mid-term review and were confronted with major problems in their governance support programmes.
2) The evaluation was one of the sources of inspiration for programming the above-mentioned Governance facility (though its implementation modalities were not necessarily coherent with the messages of the evaluation).
3) Over time the policy work on sector governance (as recommended by the evaluation) also led several EU Delegations to better integrate the governance dimension in future sector programming (e.g. in the transport sector).

In this context it is important to stress that the governance evaluation was only one of the inputs into rethinking EC practices related to governance, including at programming level. The impact was mainly in the form of (indirect) influencing of different dimensions of the huge governance agenda, with concrete changes only occurring after quite some time.

4. Qualitative dynamics

Analysis of the recommendations

Considering the feedback received during the dissemination seminar as well as the FC, the recommendations were seen to be highly relevant, concrete and feasible.

The recommendations address both the political preconditions for an effective governance action of the EC (i.e. an effective collaboration between the political arm, the development people and the administrative processes) as well as strategic and operational challenges. This combination of recommendations of a political nature and recommendations that focused on development cooperation was highly appreciated by several Units (including the Evaluation Unit). For the evaluators it proved that there is no need to shy away from "political issues" in strategic evaluations or to reduce them to a standard EQ on democracy and HR. It renders the task of the evaluation team more complicated but it adds, in our view, an important dose of realism into the evaluation report.

Just to illustrate this point: many staff from EU Delegations argued that it is very difficult for them to work on governance as there is no clear political mandate from HQ or the leadership of the EUD. This significantly reduces the operating space of EUD staff who are sensitive to governance issues (even if they do not work in the dedicated Governance Unit).

The political, strategic and technical recommendations form one package and all related to each other.

Scope

This evaluation was mainly a learning evaluation as it was accepted that the governance sector was a relatively new area'. So the traditional focus on performance and results was less dominant here.

The evaluation was clearly framed within long-term country co-operation dynamics. This was vital considering the highly political nature of governance support while also recognising all the practical challenges of designing and implementing governance programmes in often "hostile environments". These considerations were reflected in the recommendations issued.

4.16 Conflict Prevention & Peace Building

Type of evaluation	Thematic
Name of evaluation	Thematic evaluation of European Commission support to conflict prevention and peace-building
Language	English
EuropeAid reference nr.	1277
Year of evaluation	2011
Documents reviewed	• Main report • Annexes • Summary document • Presentation seminar • Fiche Contradictoire • Quality Judgement • Concept study main report • Concept study annexes • Concept study presentation seminar

See detailed process tracking study.

4.17 Human Rights

Type of evaluation	Thematic
Name of evaluation	Thematic evaluation of the European Commission support to respect of Human Rights and Fundamental Freedoms (including solidarity with victims of repression)
Language	English
EuropeAid reference nr.	1298
Year of evaluation	2011
Documents reviewed	• Main report • Annexes • Dissemination seminar presentation and minutes • Quality Judgement • Fiche Contradictoire

Key points:
- *The evaluation coincided with the Arab Spring, which led to a major upsurge in interest within the EU on how to better address human rights. As a result of this the intermediate results from the evaluation were fed into on-going policymaking processes.*
- *There was a tension between the need for an impact assessment and/or the need to consider the political implications of human rights programming. Initial interest across EuropeAid was limited apart from specialist Units, but the events of the Arab Spring brought about a sea change across the organisation, and also brought the need for political recommendations very much to the fore.*
- *The initial focus was on impact, yet as the evaluation became a source of inputs for redefining EU policy on human rights, the learning component became the priority with increased demands from EU Units to provide more specific details to support their own change processes. Uptake of evaluation evidence comes where evaluations are seen as learning opportunities and it is clear how they fit into wider change processes.*

1. Process dynamics

Evaluation planning and timing

The process was quite long: the ToRs were finalised in June 2009 and the final report delivered in December 2011. This is partly linked to the fact that two team leaders left during the inception phase, mainly because of disagreements on how to carry out this peculiar (highly political) evaluation and lack of familiarity with the EC Methodology (which led to tensions between the Evaluation Unit and the contracting firm). A new team was finally put in place, with a different composition and it could start working seriously from mid 2010 onwards.

A second source of 'delay' was linked to the events of the Arab Spring (that occurred in the midst of the evaluation). This led to a major upsurge of interest at EU level on how to better address human rights. As a result, the evaluation team was invited to participate in this fundamental 'rethink' of the overall EU approach to human rights. Exceptionally, the Head of the Evaluation Unit agreed that intermediary results could feed into these on-going policy-making processes. As a result, the consultancy company had to invest time in meeting these unexpected external demands for inputs and ideas. Furthermore, the drastic change in the EU perspective on human rights also meant the Team could now go deeper into some of the most sensitive issues and that they also had much more "liberty" to formulate political recommendations (in fact, there was a lot of pressure for the team to go as far as possible). This also explains why some extra time was taken at the end of the process to finalise the report (as the team was suddenly confronted with a much bigger 'demand').

Links to policy and programming

Initially no such link was planned. But the events in North Africa were a major 'wake-up call' exposing the limits of the EU policies pursued so far (with their preference for stability at expense of basic human rights/democracy). The Arab Spring led to several new EU Communications dealing with ENPI and broader policy frameworks which all drastically upgraded the status of human rights in EU external action. This changed the political climate, greatly facilitating the uptake of the evaluation outcomes (as a growing set of EU actors was now suddenly asking how human rights could be tackled differently in practice). There is also some link with the new BS guidelines as they now incorporate human rights issues (as advocated by the report).

In principle, the new EU policy framework, triggered by the Arab spring and partly inspired by the human rights evaluation, provides a conducive environment for a different way of programming future EU support, both in thematic instruments (EIDHR) and, more importantly, across all sectors of cooperation (i.e. the mainstreaming of human rights). In this desk analysis we cannot track possible effects on programming but there are many signs to suggest that things are evolving in a positive direction (e.g. the on-going search for ways and means to adopt "rights-based" approaches in sectors).

Definition of evaluation needs

The needs were quite well defined in the ToRs, covering both financial and non-financial instruments (e.g. human rights' dialogues), mainstreaming issues as well as the broader EU context for addressing human rights (e.g. link with other EU institutions, Member States, the UN system). As usual, the evaluation sought to both assess performance and identify learn lessons for the future. They also made the link between human rights and development (e.g. by emphasising the importance of social and economic rights or stressing the need to look at the mainstreaming of human rights in sectors).

The Evaluation Unit insisted heavily on focusing primarily on "impact issues" (compared to process issues or capacity questions). It was argued that the Unit wanted to rebalance things a bit and orient future strategic evaluations more towards providing evidence of impact. Four out of the ten EQs therefore sought to address impact.

The evaluation suffered from limited involvement from within the EU right from the outset. To some extent this may have been linked to the difficult start (with the departure of the two TLs). More fundamentally it reflected the rather 'marginalised' position of human rights in development cooperation. For many Units, human rights were seen as a specialist issue to be addressed (only/mainly) by dedicated Units (such as the human rights Unit in RELEX or the section dealing with human rights in the Thematic Unit on Governance within EuropeAid). All along the process, it proved difficult to mobilise a real RG (at the end of the day, 1-2 persons from dedicated Units acted as the de facto RG).

The case studies for this thematic evaluation could generally rely on stronger interest from the EUDs. Yet in the case of Vietnam, the evaluation was confronted with a "Chinese wall" between the political section and the development people when it comes to addressing human rights. The whole issue was deemed so sensitive by the EUD that the space to carry out a deep investigation was rather limited. This once again illustrated how difficult it can be to integrate political dimensions into evaluation work.

2. Response process and dissemination

As mentioned before, the initial interest was limited, right across the institution (with the exception of specialised Units). Yet after the Arab Spring and in the context of redefining the EU approach to human rights, this changed dramatically. All kind of doors were opened (including to the Cabinet of Ashton). There was also an informal brainstorming session on the intermediate findings, attended by more than 30 senior and middle management level people from RELEX/DEV/EuropeAid). During this gathering there was a lot of interest and agreement with the findings and conclusions.

Fiche Contradictoire

Both the quality grid and the FC suggest there was a large buy-in into the various recommendations. External events (Arab Spring) had created the conditions for a quantum leap in terms of EU approaches to human rights –as advocated by the report. It is not inaccurate to say that without these external pressures the very same recommendations would not have fallen on fertile soil and would have been described as "of limited operational use" because they were "too political" in nature.

So paradoxically, what started as an evaluation where one could fear the worst in terms of uptake, ended up having a very strong uptake in terms of changes in policies (see follow-up fiche with all the new EU communications/commitments) and also, in a slower pace, in practices on the ground.

In this case the follow-up was a combination of these response strategies.

Dissemination process

The sudden interest for improving the EU track record meant that this evaluation got into the limelight. Several unexpected opportunities for an interactive uptake emerged, including the above mentioned brainstorming on intermediate results. The dissemination seminar was highly attended and helped to broaden the audience of stakeholders (beyond the usual human rights specialists). It reflected again a fairly large uptake of the main ideas/recommendations (see minutes of the seminar).

Later on the evaluation team was also invited to present the main findings at the meeting of human rights focal points of EU Delegations in Brussels.

3. Programming dynamics

Follow-up in programming and the mid-term review

It is not possible to check here, but the drastically strengthened EU framework for dealing with human rights across instruments is inevitably going to impact on programming practices. It will take time of

course to translate the new policy into practice. Old habits will not go away quickly (including the reluctance of development people to work on human rights issues, the lack of overall EU coherence, the use of double standards) and in some EUDs changes may end up being more cosmetic than real. But human rights has now undoubtedly a much more prominent political profile in overall EU external action ensuring that consistent pressure is placed on the system to include the topic more forcefully and effectively in programming processes.

4. Qualitative dynamics

Analysis of the recommendations

Considering the feedback received during the informal brainstorming seminar on intermediate findings, the dissemination seminar as well as the FC, the recommendations were seen to be highly relevant, concrete and feasible (the quality grid considers them to be "excellent").

The recommendations were seen to address the systemic weaknesses of the EU system regarding human rights, including the need for a minimum coherence in EU external action as a precondition for effective human rights interventions. So it goes far beyond the 'downstream' implementation issues related to the management of the EIDHR (the thematic budget line for human rights). Having stressed the need for an overarching and coherent EU action, the recommendations then touch upon all other key factors for a successful human rights action, including a 'localisation' of the agenda (so as to tailor support to the specific realities of each country); a reinvigorated approach to political dialogue; overcoming the divide between HR and development; strengthened strategic alliances and creating a more enabling institutional environment within the EU to mainstream/promote human rights.

These are mostly political recommendations which require both short-term action and long-term transformation. The political, strategic and technical recommendations form one package and are related to each other.

Scope

Initially the focus was very much on impact. Yet as the human rights evaluation unexpectedly became a source of inputs for redefining the EU policy on human rights, the learning component gained prominence. In some cases, EU Units even pushed the team to provide more details on some recommendations badly needed for their own change process (e.g. on the type of incentives which could foster the mainstreaming human rights). At some stage, the team had to draw a line as these requests went far beyond the remit of an evaluation.

The evaluation was clearly framed in long-term country and co-operation dynamic: this was vital considering the highly political nature of human rights. This framing was reflected in the recommendations which were issued.

4.18 EC aid delivery through Civil Society Organisations

Type of evaluation	Aid modality
Name of evaluation	Evaluation of EC aid delivery through Civil Society Organisations
Language	English
EuropeAid reference nr.	1259
Year of evaluation	2008
Documents reviewed	- Main report - Annexes - Presentation seminar – operational support - Presentation seminar - Summary document - Quality Judgement - Fiche Contradictoire

Key points:

- *The Reference Group was split between those who wanted a pure "channel" evaluation, namely an evaluation of funds, which were channelled through CSOs, and those who wanted a wider remit to focus on political reforms. The evaluation ended up addressing both sides of the divide.*
- *While the comments in the Quality Grid suggested that few of the recommendations were immediately useful (because they required political action to move forward), in practice they appear to have been taken up widely and applied. An internal drive for reform coupled with a conducive external environment and pressure from CSOs themselves seems to have helped this.*
- *This raises the question about the purpose of strategic evaluations and whether they should seek to raise the more political issues to allow for real learning beyond direct operational needs. Experience with this evaluation suggests they should.*

1. Process dynamics

Evaluation planning and timing

The process started in mid 2007 and was concluded in December 2009 with a dissemination seminar in February 2010.

Links to policy and programming

There was no planned link with policy production, as the evaluation was mainly of an instrumental nature ("what is the added value of the civil society channel to deliver EU aid compared to other channels and how to use it effectively?"). Yet this instrumental focus was criticised by several RG members right from the outset, who felt that the evaluation should have gone broader than a mere "channel" evaluation (i.e. aid delivered _through_ civil society) and also encompasses the question of overall EU engagement strategies _with_ civil society. Also external stakeholders (in the first place civil society organisations) felt that the focus on using CSOs as aid delivery channel was too narrow/instrumental. The consultants tried to overcome this tension by looking primarily at "channel" issues while putting this analysis in a wider context (so that the broader question of engagement strategies could also be covered).

The resulting recommendations embraced both aspects (working "through" and "with" CSOs) by insisting on the need for a drastic revision of the overall EU policy framework for engaging with CSOs and, within that framework, using the CSO channel.

While some EC actors felt the evaluation had overstepped its mark or had not sufficiently included some critical aspects[4], others Units[5] as well as CSOs attending the dissemination seminar, felt this report gave them a wide range of strong arguments to push for a deeper (political) change in the relations with CSOs[6].

---

[4] The CSO Unit –in charge of managing the budget lines in support of CSOs- was primarily interested in aid delivery and the use of the CSO channel. They wanted concrete responses to their direct operational needs. Yet as mentioned above, it proved impossible (and also not desirable) for the evaluation team to artificially look at CSOs as a "channel" without also looking at the broader policy framework. The less operational recommendations were clearly less appreciated by the CSO Unit, which also considered the evaluation was too optimistic as to the capacity of local CSOs. The newly arrived Head of Unit considered that the "EC is criticized for not engaging more forcefully with CSOs but very often we simply cannot do it because the local organizations are too weak".

[5] This was the case of the thematic unit AIDCO E4 who very actively participated in the dissemination seminar and strategically used the evaluation to defend and extend its own reform agenda regarding civil society

[6] One of the recommendations was to organize a "structured dialogue" with CSOs (and other actors) to jointly agree on "new rules of the game" for EU engagement strategies with civil society. In 2009 this dialogue was initiated by the EC. IIs conclusions (see Budapest declaration) heavily influenced the subsequent elaboration of the above mentioned Communication.

While it is difficult to attribute things to the evaluation report[7], it is a fact that 3 years later the EC had drastically reviewed its overall policy framework, as reflected in the September 2012 EU Communication: *"The roots of democracy and sustainable development: Europe's engagement with civil society in external action".*

There is no direct link to policy and programming, but the new EU policy framework for engaging with civil society is now having a major impact on programming support to civil society.

Definition of evaluation needs

The needs were quite well-defined as this evaluation was part and parcel of a "package" of evaluations dealing with different "channels" for aid delivery (the other two evaluations looked at the UN channel and the multilateral banks). So there was a common thread running through these 3 evaluations.

Yet as observed above, several RG members felt this was too narrow and wanted to include other aspects. The Evaluation Unit rejected this but indirectly space was created to integrate these wider aspects into the evaluation so that the needs of other actors were also addressed (at least partly).

There was a parallel Court of Auditors (CoA) report assessing the performance of the EC in terms of promoting the participation of non-state actors in cooperation processes and providing capacity development support. The Unit refused to have formal exchanges between the two processes. The CoA report was finalised a few months later and arrived at similar conclusions than the evaluation (i.e. that the EU lacks clear engagement strategies with CSOs and therefore achieved less than optimal impact).

2. Response process and dissemination

The response process is quite interesting in this case as it illustrates many uptake dilemmas.

The Quality Grid considers the recommendations to be "good". However, in the text it is argued that "with one exception they are not immediately useful. They may be needed, but under the present circumstances, few of them can be applied". This statement reflects the above-mentioned split between those in the EC/RG who wanted a pure "channel" evaluation (of an operational nature) and those who wanted this evaluation also to focus on the needed political/policy reforms. The evaluation team decided to try to cover both aspects and that is why many recommendations indeed are of a more political nature. Those who filled in the Quality Grid apparently felt that the time was not ripe for these changes (even if they may be needed) so considered the report to be of limited operational use.

Yet this was not at all the perspective of other actors within the EC, who immediately after the publication of the report (and that of the CoA) started to push for an overall review of the EU policy towards CSOs (as advocated by the evaluation report). This internal drive, combined with a conducive international environment (Accra) and a real demand from CSOs themselves, triggered quite soon afterwards a real reform process within the EU. A "structured dialogue" was organised between March 2010 and May 2011 which helped to redefine the nature of the partnership relationship linking the EU and CSOs while also proposing all kind of changes in support strategies[8]. This, in turn, provided much of the substance for the September 2012 Communication.

---

[7] It is the Team Leader's view that the evaluation report was certainly one of the main sources of "push and inspiration" for the change because it came at a time that the international agenda was changing (i.e. Accra Action Plan, with its focus on country ownership and CSO participation); the limits of EU approaches to civil society became visible in various evaluations; CSO were becoming more vocal and demanding a more strategic engagement of the EU: the role of the Court of Auditors who addressed partly the same agenda and was also highly critical on EU performance, etc.

[8] For details see website information related to the "Structured Dialogue for an efficient partnership in development".

In that context, it is very interesting to see how elaborated the responses are of the Commission services. Despite major reservations in the Quality Grid on the "immediate" usefulness of the recommendations, the EC de facto agrees with all recommendations. It also explicitly states that several recommendations "*are very political in nature and of great importance. Any uptake can only be done within the new framework [of the EEAS] and in respect of political decisions to be taken by the appropriate authorities*". In practice, it means that EuropeAid agrees with the proposed re-orientations but that these matters require the political leadership (EEAS) to act on them before things can effectively change.

Fiche Contradictoire

The follow-up part of the fiche is very elaborated (9 pages). The EC basically responds that almost all recommendations have been effectively put into practice, making in the process a systematic reference to the new EU Communication of 2012 (which was in fact the "political response" EuropeAid operational people were waiting for).

This slightly contrasts with the rather lukewarm initial reception of the report by some Units. However, as the report was used by other stakeholders to trigger a major reform process, the follow up fiche looks completely different two years later. We see a full buy of recommendations deemed of limited feasibility in the Quality Grid.

There was a broad agreement on the outcome of the evaluations, but the EC response reflected a split between the value given to "operational recommendations", on the one hand, and "political-strategic recommendations". There was clearly some uneasiness to embrace the latter, as this was perceived to be out of the remit of EuropeAid. More fundamentally, this raises some questions about the very purpose of strategic evaluations: should evaluations not venture into "political" arenas and refrain from making fundamental remarks on changes needed (on the basis of the argument, used here initially, that such recommendations cannot be operationalised)? Or should strategic evaluations precisely seek to raise these more difficult political issues so as to allow for real learning (beyond direct operational needs) and over time enhance policies and practices?

In hindsight it could be argued that internal reformers and external pressures helped to 'uplift' the status of this evaluation. Without these forces, the uptake would have been limited (as can be inferred from the Quality Grid). Yet as this evaluation was used by these reformers and coincided with other push factors, the uptake process got a boost, both in strategic policy terms (see the new Communication) and in terms of short management responses). In fact, virtually all the changes advocated in the evaluation report (which were initially resisted in some corners) are now mainstreamed in the new policy.

Dissemination process

The dissemination seminar of February 2010 in Brussels was widely attended. It displayed the various 'camps' regarding this evaluation: the more Operational Units managing funds, the Thematic Units concerned with learning and above all the CSOs themselves, who were direct stakeholders. The role of the latter in the uptake process should not be underestimated. They formed a powerful ally to avoid a purely bureaucratic response to the evaluation.

The organisers of the Structured Dialogue extensively used the outcome of the evaluation to justify the organisation of such a dialogue.

It is interesting to note that several later country evaluations refer to this CSO evaluation when they address issues dealing with EC support to CSOs.

3. Programming dynamics

Follow-up in programming and the mid-term review

The CSO evaluation was not a thematic one, but related to "instruments and channels". A key expectation is that such evaluations help to reorient programming, in this case of the specific budget lines in support of civil society.

In an initial phase, the impact of the evaluation on programming budget lines was limited. Yet the Structured dialogue and the later Communication changed "the rules" of engagement with CSOs and are likely to have a profound impact on how the EC works with CSOs and funds them[9].

4. Qualitative dynamics

Analysis of the recommendations

See above. From a rather bureaucratic perspective, most recommendations were initially considered not to be "immediately useful" as they required political action to move forward. However, the choice of the evaluation to include clear political recommendations in the report somehow helped to trigger (together with other factors) a political drive within EC to tackle more fundamental challenges in relation to CSOs. This, in turn, helped to drastically increase the feasibility of the recommendations.

There were 4 types of interrelated recommendations

1) Overall recommendation, basically in the form of one core message: time to reform!
2) Political recommendations geared for the political leadership: create space for a different (non-instrumental) relationship with CSOs as actors in their own right
3) Strategic and operational recommendations (to translate this new political vision)
4) Process recommendations (on how to promote institutional change, mainly within EU Delegations).

The political, strategic and technical recommendations form one package and all related to each other.

Scope

The evaluation did go beyond its primary mandate. The evaluation team decided to move beyond a mere assessment of how the EC used the CSO channel, to also look deeper into explaining why the choice for a pure "aid delivery channel perspective" was no longer adequate and therefore doomed to have only limited impact, and be increasingly in contradiction with stated EU policies.

It could be argued that the evaluation team opted to carry out the evaluation with a "persuasive agenda".

The evaluation was framed in long-term country and co-operation dynamics, though it was a bit of a battle, as there were pressures from several Units to largely limit the evaluation to the more operational, aid delivery challenges related to CSOs. This was reflected in the recommendations.

---

[9] The evaluation had recommended such a (new round of) structured dialogue, amongst others to drastically review the modalities of supporting CSOs (beyond Call for Proposals). The Structured Dialogue led to the production of 12 operational fiches proposing plenty of new ways of supporting and funding CSOs.

4.19 EIB Investment Facility

Type of evaluation	Aid modality
Name of evaluation	Mid-term evaluation of the Investment Facility and EIB own resources operations in ACP countries and the OCTs
Language	French & English
EuropeAid reference nr.	1285
Year of evaluation	2010
Documents reviewed	• Main report (English and French) • Fiche Contradictoire

Key points:

- *The evaluation was able to make recommendations which linked two quite different programming cycles: the Commission's programming cycles over several years and the EIB's demand-led approach to programming.*

- *Because of its focus on the EIB's contribution to development outcomes, the evaluation was able to get to the heart of the relationship between the EC (especially country programmes), EIB and EEAS.*

- *The recommendations have resulted in very concrete changes which go beyond programming recommendations to sharing office space in order to maximise synergies between the three organisations.*

1. Process dynamics

Evaluation planning and timing

The evaluation was conducted in 2010 (final report dated September 2010) and covered the period 2003-2009. This covered the end of one financial protocol (2003-2008) and the beginning of the next one (2009-2013). It therefore was not able to inform the development of the second financial protocol, but it was a mid-term review which had been necessitated by the 2005 revision of the Cotonou agreement.

Link to policy and programming

One of the criticisms of the evaluation was that the Commission and the EIB seemed to act on parallel tracks with few synergies. This was partly due to the difficulty of matching a Commission programming cycle over several years with a demand-led approach to EIB programming. The first recommendation was to define guidelines with criteria for prioritisation of contributions to poverty reduction and to develop of a strategic approach at country level. From the FC this appears to have contributed to some significant changes; changing the ACP Business Plan, contributing to pillar 1 of the new REM Framework from Jan 2012 onwards, contributing to the relocation of some country offices, and sharing of the Country Programming Guidelines with the EIB. This has been a two-way process; the EIB has also put its sector policies (transport and energy) out for consultation for the first time. Other recommendations for better linking the EIB to policy dialogue have generated responses such as systematic consultation of EIB by EEAS and by country offices, office-sharing arrangements with EEAS, and EIB-EC-EEAS co-ordination meetings to be held annually.

So while there may not have been a direct link with programming cycles or policy production, there does appear to have been a real change in the way the EIB relates to country programmes.

Definition of evaluation needs

The evaluation needs are well defined: there are 10 EQs, 40 JCs and 127 Indicators. These were defined during the structuring stage of the evaluation, and agreed with the Reference Group. They cover the normal set of criteria for EC evaluations, the 5 DAC criteria plus co-ordination, complementarity and added value.

The evaluation team did not identify any issues which really sat outside the evaluation criteria so, in that sense, they could be taken to be comprehensive.

2. Response process and dissemination

Fiche Contradictoire

The evaluation could be seen as being quite critical in some aspects. There was an acknowledgement that, within a certain mandate, the IF/OR facilities had achieved what they set out to achieve (and in the case of the financial sector, had done it well) but that the overall mandate was really quite flawed. The evaluation showed that the EIB had not developed a strategy to maximise the contribution of IF/OR resources to sustainable development (Conclusion 1) and that the long-term sustainability of the fund was 'fragile'.

Given this, a defensive first response in the FC might be expected but in fact it was positive, mostly in agreement or partial agreement with the recommendations. The initial response is strategic rather than detailed, outlining the Commission's intentions (and assigning responsibility) but without setting out exactly what would be done. There is no QJ to indicate whether the Commission thought this was a good evaluation or not, but they did agree with most of the findings and disagreed with none of them.

The follow-up response is much more detailed than the initial response. It outlines several real changes which had taken place: some parts of the follow-up response indicate broader changes which had taken place to which this evaluation had contributed, but many of them appear to have been directly related to the evaluation recommendations.

Dissemination

There were three dissemination seminars in total: two in Brussels and one in Luxembourg where the EIB is based. None of the relevant documentation for these seminars was available, so it is not clear who attended and what was discussed.

The FC indicates that there was good uptake of the recommendations, and that concrete changes were made as a result. No other uptake initiatives were found in the documentation.

3. Programming dynamics

Follow-up in programming and the mid-term review

The EDF national programming guidelines have been shared with the EIB, but it is not clear the extent to which the recommendations for closer links have been taken on board. R7.2 suggests that the Commission could build on the EIB's expertise for financial sector and private sector analysis, and the EIB could call on Commission staff for country knowledge and local technical issues. The higher-level conclusions and recommendations call for greater collaboration during programming cycles.

4. Qualitative dynamics

Analysis of the recommendations

The recommendations all appear to be quite feasible, given that they were pretty much accepted and then acted on in a very detailed way.

The main thrust of the recommendations is that the EIB needed to be more strategic in its approach and to co-ordinate better with the Commission and the EEAS to ensure that they maintained a broad focus on developmental impacts rather than a narrow one that related only to their institutional mandate. The evaluators also commented on the long-term stability and sustainability of the operation.

There are few recommendations relating to national politics as these are arguably inappropriate; at the national level any political analysis would need to be led by the country programmes. However, the report did call for the EIB to be involved more in policy dialogue, and to devote more resources to

ensuring this happens at a country level. The response - that EIB staff could share office space with EEAS or national Delegations - seems to indicate that all parties are taking this responsibility seriously. The rest of the recommendations are a mix of strategic and technical.

The evaluation went beyond its primary mandate in that the report outlines the various strategic changes which need to be made for the EIB to function well as an instrument of EU development policy. The EQs show that there was a need for this: the first EQ relates to co-ordination and complementarity, for example, rather than being banished to the back of the report. Performance is assessed, but is done in the broader context of whether or not the EIB's strategic intent is the right one.

The recommendations were framed by the political economy of EIB/EC/EEAS relationships, not country relationships. As a review of the entirety of EIB operations, it would not be feasible to comment on the local political economy in each country. The direction to engage more closely with the EUDs should affect how the EIB takes into account local long-term country dynamics, but this will only play out in the long-term: it is not possible to assess that here.

The recommendations talk about stretching the EIB's mandate in asking it to take more account of how it can contribute to development outcomes rather than just align with them: i.e. that it takes a more active role in deciding what outcomes it is seeking (in collaboration with other IFIs, EUDs, EEAS and the Commission).

4.20-24 Budget support in Mali, Tunisia, Zambia, Tanzania (& synthesis report)

Type of evaluation	Aid modality
Name of evaluations	1. Joint evaluation of Budget support in Mali 2. Evaluation of European Commission BS Operations in Tunisia between 1998-2006 3. Between high expectations and reality: an evaluation of BS in Zambia (synthesis report) 4. A synthesis report from the OECD DAC network on development evaluation 5. Joint evaluation of BS to Tanzania: lessons learned and recommendations for the future
Language	Mali: French & English Tunisia: French & English Zambia: English Synthesis report: English Tanzania: English
EuropeAid reference nr.	Mali: 1290 Tunisia: 1286 Zambia: not numbered Synthesis report: not numbered Tanzania: not numbered
Year of evaluation	Mali: 2011 Tunisia: 2011 Zambia: no date, presumed 2011 Synthesis report: 2011 Tanzania: 2013
Documents reviewed	Main report: Tunisia, Mali, Zambia, Tanzania Annexes: Tunisia, Mali

See detailed process tracking study.

4.25 Geographic Evaluations Synthesis Report

Type of evaluation	Synthesis evaluation
Name of evaluation	Synthesis of the geographical evaluations managed by the Evaluation Unit during the period 1998-2006
Language	English
EuropeAid reference nr.	1256
Type of evaluation	Synthesis Report
Year of evaluation	2007-2008
Documents reviewed	• Main report

Key points:
- *The main purpose of this evaluation was not clear.*
- *The end result was that this became a methodological review and failed to produce any real analysis of the 40+ geographic evaluations which were covered.*
- *As a result there was little potential for uptake beyond the Evaluation Unit.*

1. Process dynamics

Evaluation planning and timing

The synthesis report was part of the 2006 evaluation strategy. The study was carried out in 2007 and covers the geographic evaluations managed by the JEU in 1998-2006. The final report is dated October 2008.

Links to policy and programming

It is unclear from the ToRs and the report what processes the study was meant to feed into. The ToRs outline dual objectives:

1. To analyse the evolution of the methodology used in evaluations since 1998 and propose paths for improvements;
2. To identify key lessons learned from the various geographical evaluations carried out so far.

The first objective is clearly aimed at **improving the internal methodological system in the Evaluation** Unit. However, the study was carried out one year after the revision of the methodological guidelines which are currently in use.

The second objective is to draw (general) lessons from a meta-analysis of the existing body of geographical evaluations; however this eventually took the form of a synthesis of conclusions and recommendations, which is of little use outside the EC evaluation system.

The ToRs initially required the consultant to "analyse the extent to which lessons learned from geographical evaluations have been taken into account in the last generation of programming documents (country strategy papers (CSPs) and regional strategy papers (RSPs))." From an uptake perspective this would have been useful, however, this was not part of the final report.

Definition of evaluation needs

The two objectives are translated in two main components: (1) a methodological and quality assessments of geographical evaluations, and (2) lessons learned from the main effects of the Commission's interventions relating to the evaluations under review. While the first component is very clear and well defined (both in the ToRs and in the report), this is not the case with the second component. In the ToRs this objective is set out in very general terms. The ToRs ask the consultant to "draw lessons from the main effects of the Commission's interventions relating to the 40 evaluations under review" and "identify good practices regarding the implementation of the Commission's interventions".

The **report interprets these objectives in a very liberal way**. Rather than a genuine quality judgment, the report makes a detailed *"conformity assessment"* to measure (numerically) the extent to which the evaluation reports conform to the methodology which the JEU provides. This seems a strange choice since the first task of the study was to assess the evaluation methodology over time. Since the following section equates quality with conformity, the study leaves little room to suggest improvements. The 'lessons learnt section' is little more than a summary of conclusions and recommendations of the "evaluations which do conform" to this methodology.

2. Response process and dissemination

The way the evaluation was carried out indicates little strategic thinking on the value and use of geographic evaluations; instead it shows a tendency for navel-gazing and a near-religious adherence to the evaluation methodology. Beyond internal use by the Evaluation Unit the purpose of this synthesis report is unclear. The report inscribes itself in a methodological self-perpetuating logic and largely fails to produce a real analysis of patterns and lessons emerging from the 40+ geographic evaluations it reviewed.

As a result the potential for uptake outside the evaluation system was quite limited. In addition, there was no reference group for the study. No signs of the study being used could be identified in the document review.

3. Programming dynamics

Follow-up in programming and the mid-term review

None seen

4. Qualitative dynamics

Analysis of the recommendations

See above.

Scope of the evaluation

See above: the scope was very limited indeed.

5. DETAILED PROCESS TRACKING STUDIES

The following section looks at four sets of evaluations: (1) BS evaluations, (2) education, (3) conflict prevention and peace building, and (4) Egypt. These more detailed process tracking studies are based on a documentary analysis, complemented by a series of in-depth interviews with the people involved in the production of the evaluations and the relevant EC and EEAS services. These process tracking studies look at the formal as well as informal processes of uptake which take place during and after the finalisation of the evaluations.

5.1 Budget support

This case study covers four evaluations of BS initiatives: in Tunisia, Mali, Zambia and Tanzania (Case Studies 21-25). It differs from the other process tracking studies in this report in that it covers more than one initiative, but this is because of the core component of methodological development throughout the first three studies and its consolidation in the Tanzania study.

Reasons for commissioning the evaluations

There were two distinct but linked reasons for commissioning the evaluations: to develop and refine the methodology and to evaluate specific BS programmes.

The EC and other donors had been funding BS initiatives for many years – since the structural adjustment programmes in the 1980's, but there were disagreements between donors on the methodology for evaluating their effectiveness and, linked to that, disagreements on whether BS (either GBS or SBS) was a useful mechanism for delivering aid. The Evaluation Unit at the time felt that the prevailing methodology put forward by DFID focused too much on process and not enough on impacts, and that a more rigorous methodology was needed to demonstrate whether or not BS had the impacts it was intended to. As the debates around the effectiveness of BS continued, the EC's Evaluation Unit led the process of developing a more robust methodology, working with OECD-DAC, other potential donors (the Dutch, German, Danish and Irish governments) and the consultants to develop a three-step framework which would be refined and debated after the results for the first three evaluations came in. Bearing in mind that BS comprises policy dialogue and capacity development as well as the disbursement of funds, the complexity and coverage of this framework meant that methodological discussions took a considerable amount of time.

Timing

Three evaluations were needed to fully test the revised methodology: Mali (conducted between 2009-11, covering the period 2003-09), Tunisia (2009-11, covering the period 1996-2008) and Zambia (2009-2011, covering the period 2005-2010). The choice of which countries to focus on depended, in part, on the need to evaluate a new forms of BS - the MDG contract. Previous types of BS had been for short periods, 2-3 years with annual review: under MDG contracts the BS period was extended to a 6-year programme with a mid-term review after 3 years, giving more predictability in terms of disbursement. Debates with donors about impact included debates about whether an MDG contract could be an answer to some of the shortcomings of shorter BS programmes.

The final choice of countries also depended on relationships with other donors and partner governments. The EC conducted the evaluation alone in Tunisia, collaborated with other donors in Mali (where there were many well co-ordinated donors supplying both GBS and SBS), and worked with the Dutch government in Zambia.

Users

There were three distinct sets of potential users: country offices which would make direct use of the evaluations, the Evaluation Unit which was keen to ensure a robust methodology was in place, and the BS team in EuropeAid which would work with country offices and the Evaluation Unit to apply the methodology to future evaluations. As BS was a high-profile instrument, others at the top of the office such as the Commissioner were also considered potential users of the evidence.

Other relevant events

The development of the methodology and implementation of the three evaluations was coloured by debates between donors on the effectiveness of BS, and by the relationships between individual donors and individual country governments. The view of the Evaluation Unit and other BS specialists was that there was an insufficient body of evidence on what BS was delivering, without a robust methodology it was impossible to judge, and that donors needed to commit to establishing this and testing it before making any judgements about effectiveness. Donors responding to domestic political pressures to reduce BS because of allegations of corruption or inefficiency were particularly difficult to keep on board during the long process of testing the methodology through the three initial evaluations. Within the EC there were also dissenting views, and the two Commissioners in post over the period of the three evaluations held different opinions about it.

Drafting of Terms of Reference

Terms of reference for each of the first three evaluations were developed in close collaboration with the consultants, as part of the process of methodological development. There was a clear sense that the methodology was being tested, particularly in terms of data availability and what this meant for how rigorously and comprehensively the methodology could be applied. For the Tanzanian evaluation the methodology was considered stable enough to be drafted by the Evaluation Unit, though it reflected years of close collaboration with the consultants who were thus very familiar with what they were being asked to do.

The evaluations

The three initial evaluations differed in what they were able to achieve. The Tunisia report was very well received and thought to be very comprehensive. The Mali evaluation ended up being more focused on process than the Evaluation Unit would have liked, but it encountered data collection problems which limited its ability to look at outcomes. The Zambia evaluation differed from the first two in that the consultants were asked to test some ideas from the Dutch government, which proved difficult to implement. But nonetheless the overall conclusion from all three was that the methodology was workable, and the results of the evaluations have been taken up and used elsewhere. A presentation to the Development Commissioner and the European Parliament with the results of all three evaluations was very well received and helped shape policy around BS.

The key issues for the evaluations were data availability and the ability of the consultants to conduct these very complex pieces of work: as with PFM, only a few consultants have the qualifications and experience to be able to conduct them. During the first three evaluations there was a good deal of learning by doing as data availability was often an issue, particularly when several donors were involved in BS activities.

The Tanzania evaluation was very well received in-country, with the donor group and government of Tanzania developing an action plan as a direct result of the evaluation process and report.

Feedback from the reference group

The Reference Groups comprised people from the Evaluation Unit, EUDs and – depending on the level of co-funding – representatives from other donors. The RGs for all evaluations were well aware of the methodological development during the three evaluations and the data collection issues the consultants faced, which made answering the attribution question difficult. The fact that the Reference Groups were comprised of people partly in HQ and partly in the field occasionally made communications difficult.

Analysis

The three-step approach developed during methodological discussions is (taken from the Tunisia report):

- Step 1: Evaluation of the specific intervention logic of BS, i.e. the bottom-up process covering the first three levels of the evaluation framework, from GBS/SBS inputs (Level 1), through outputs (Level 2), to induced outputs (Level 3).
- Step 2: Evaluation of the outcomes and impacts of the government strategy supported by GBS/SBS, i.e. the top-down process of the impact (Level 5) and outcomes (Level 4), from the strategy to its determining factors, including the achievement of government strategies in general, and particularly those affected by GBS/SBS.
- Step 3: Exploration of the links between the processes and results of GBS/SBS and the impact of the government strategy.

In line with this methodology each evaluation was structured around a series of EQs, consisting of the 'compulsory' DAC evaluations questions and further specific questions defined by the RG and the consultancy team, linked to the particular intervention logic of the BS operations covered by each evaluation. This translated into approximately 20 EQs in each country, with a variable number of JCs. The EQs covered both GBS and SBS.

Recommendations

Each report made country-specific recommendations relating to the three components of BS. OECD-DAC commissioned an overview report from the lead consultants from the Mali, Tunisia and Zambia evaluations, which also considered whether or not the methodology was robust and useful.

The messages emerging from the individual evaluations were complex and country-specific, but there were three headlines: first that there was evidence that BS was contributing to impacts; second, but that tranche disbursement on its own did not contribute to policy change but that it could reinforce change which had already been stimulated through policy dialogue. The third message was that the methodology was robust, that it was able to assess impacts as well as process, and that it could therefore be used as the basis for future evaluations.

Dissemination seminars

Individual dissemination seminars were held in-country for each evaluation, but the DAC overview report was presented to Commissioners and the European Parliament at a separate seminar in Brussels, hosted by the Belgian government. This was seen as particularly key to get messages out to the other donors that BS did have a role to play. For the Tanzania evaluation, an in-country seminar was complemented by another seminar held in Brussels, to which the Tanzanian government sent representation – seen as evidence of their high level of buy-in to the evaluation process.

Management response and implementation

No *Fiches Contradictoires* were produced for these evaluations: FCs are only applicable where the EC is the single donor, and are not used with multi-donor evaluations. This means that there is no formal management response process and therefore no way of formally assessing implementation of the recommendations.

Wider uptake

The outcomes of the three initial evaluations were taken into account in drafting the BS guidelines issued in 2012. Whether they have been translated into the identification and formulation of new programmes in the three countries is hard to say: in Mali and Tunisia the wider context has changed so dramatically that it is impossible to attribute change to a particular set of recommendations. Evidence from the Tunisia evaluation has been used to support arguments in favour of GBS as the conclusions were particularly strong; but in terms of the overall programming cycle the fact that we are reaching the end of the EDF10 means that there are few opportunities to bring the results of the evaluations into programming. There will be a better opportunity to do this in 2014/15.

The view within the Evaluation Unit is that the methodology for these full evaluations is stable and that there should be less emphasis now on developing it further and more on communicating it to the geographic desks, as there is a lack of expertise on BS in EUDs, particularly in Asia. In Africa the tradition of BS is much older, meaning that there is more expertise in Delegations and quality control in HQ is better. The movement of staff from dedicated Budget Support Units to geographic desks was seen as having hindered the uptake of results, as there are now fewer people dedicated to BS and ensuring the quality of BS programmes. The high level of staff mobility and the lack of rigorous quality control of BS programmes via the QSG process is another issue constraining the uptake of evidence from these evaluations. The danger is that without good communication of the evidence from these evaluations there will still be confusion between failure of BS as an instrument (which the evaluations do not support) and failure of BS because of how it is implemented and in particular the importance of the policy dialogue component of BS. There are still discussions within EuropeAid about whether BS is as efficient as other aid delivery mechanisms: the results from the evaluations suggest that it is, and those working on the detail of BS suggest that the evidence from the evaluations shows that the time has come to close down that particular debate.

The fact that an immediate effect of the Tanzania evaluation is the development of a multi-donor action plan provides evidence that there has been some uptake of the evaluation evidence, even if some of the recommendations in the evaluation (e.g. the relationship between tranche disbursement and PAFs) are not fully agreed with. Whereas the evidence is not widely seen to be useful it is doubtful that as much energy would have been put into the co-ordination between donors as was necessary to develop a shared action plan. Within the country it is hard to see what more could be expected. The Evaluation Manager for the on-going South African evaluation is attempting to ensure that the way uptake is expected to happen is considered by the consultants.

It is clear from the interviews that still more could be done to transmit the lessons to a wider audience. In addition, because of the historical interest of a wide number of donors in the process, interviewees have stressed that uptake by donors – particularly those MS who were sceptical of the methodology or BS as a whole – is as important as strengthening uptake within EuropeAid.

5.2 Education

Title ? Box of basic info as per other evaluations ?

Possible reasons for commissioning the evaluation

Informants have put forward two suggestions as to why an evaluation of the education sector was commissioned:

- Mainstream development discourse was dominated by the Millennium Development Goals (MDGs) in which education was key. The Development Commissioner was committed to seeing EC aid contribute to MDGs
- The EU had spent EUR 800 million on education, but had not commissioned an evaluation since the late 1990s – so wanted to know what effects their funding had on outcomes and how this had come about.

Timing issues

The evaluation could have helped inform 2007-2014 programming process if the results were released during 2005/06. However, it was commissioned in 2008 to evaluate the 2000-2007 cycle. But even then it was delayed due to the initial attempt being cancelled (see below). Despite the delay in commissioning, the period of evaluation (2000-2007) was maintained. When it was published in 2010, three years had elapsed since the programme period ended. So they were making recommendations for issues which had already been remedied.

Although the dissemination seminar minutes suggest that the results could feed into the next programming cycle, planning for the 2014-2020 Country Strategic plans was due to take place in 2012-2013, whilst the evaluation was published in 2010, coming, according to a couple of key informants, a couple of years too early. However, uptake could take time, with understanding of the recommendations taking possibly a year and another year required for directorates to negotiate and take action with regards to sector prioritisation.

The evaluation took 18 months in total, whilst the evaluators were contracted for about 70 days, making it difficult to maintain momentum for that length of time.

Nevertheless, evaluation results could be useful for the design of individual education projects and programmes which sit within the overall framework of CSPs, if the relevant staff is made aware of its availability.

Users

The Thematic Unit on Education in EuropeAid was seen as the key user – more so than geographic desks. The former was responsible for providing substantive guidance to Delegations through check lists and toolkits, whilst the geographic desks facilitated administrative linkages between the Delegations and Brussels. The Evaluation Unit contacted the education Thematic Unit in AIDCO when the former decided to undertake the evaluation.

Other relevant events

The evaluation took place at roughly the same time as a special audit by the Court of Auditors on EU support to the education sector. The CoA tried to avoid duplication, by for instance, not visiting the same countries the EuropeAid evaluation was going to. However, this stretched the Thematic Unit's workload (being the key users/stakeholders in the process) as they only had 5 people and were significantly involved in both efforts.

Formation of the Reference Group

The Reference Group (RG) was made up of relevant desks within the Thematic and Geographic Unit mainly from DG DEV with a few members from DG RELEX. The role of the Reference Group is laid out in the evaluation guidelines. However, these are not very clear. In practice, the evaluation manager guides the evaluation with support from the RG. The RG can influence the scope of the evaluation (i.e. what is dealt with and analysed) as well as the shape and content of the conclusions and recommendations.

Drafting of Terms of Reference

The Evaluation Unit took responsibility for drafting the ToRs. Although the QG notes that the evaluation went above and beyond the ToRs in setting out the international policy framework in the education sector, the evaluator suggested that the ToRs were relatively vague in the first place. Issues to be evaluated were not described in detail. The ToRs rather set the scene. In fact, section 5 in the ToRs set out very broad questions for the evaluation and explicitly noted that these would need to be refined by the project team.

The Thematic Unit on Education in EuropeAid wanted the evaluation to include tertiary education, as this was an emerging issue; EC education aid was increasingly being oriented towards higher education systems and helping recipient countries participate in ERASMUS. However, the Evaluation Unit for reasons unknown, was keen to limit the scope, with room for manoeuvre seemingly limited. The ToRs did not seem to refer to any long-term objectives for the education sector – this maybe because there weren't any.

Initial attempt at evaluation

The interest amongst the reference group and ownership amongst the Thematic Unit was medium to high at the very start of the process.

At first, the evaluation was cancelled, as the manager and reference group were unhappy with the inception report, whilst the inventory was incomplete. The questions were too many and too complex. It seemed the team were unfamiliar with the EC's methodological guidelines. The inventory guide was a list of all the EC funded education projects across the world with details of funding and country among other things..

Second attempt

A new team leader was brought in to help turn things around. Not all consultants were replaced: 2-3 remained from the initial attempt. There was a perception that some of the consultants were more competent than others.

The failure of the first attempt led to ownership reducing amongst the Thematic Unit and it being transferred back to the Evaluation Unit and possibly to the evaluators themselves. The evaluation team had to work very hard to establish credibility and repair reputational damage (presumably to the consortium which had the framework contract). Fortunately the lead evaluator was familiar with the evaluation guidelines (having helped draw them up) and knew what they had to deliver – ensuring that evaluation questions were well thought out, and which the RG were satisfied with.

They even tried to exceed expectations, by for instance, undertaking a survey amongst Delegations via the web, something which had not been done before. In fact the QG noted that the evaluation went beyond its remit (broad ToRs notwithstanding). The evaluation is framed in long-term EC cooperation dynamics, though more in terms of the global education sector. However, it does not discuss the specific politics of the global work on education, by for instance outlining differences in the approach between the EC and development banks or others donors. However, Annex IIa seems to have addressed differences in approaches between the EC and national governments.

Methodology

Words used by informants to describe the methodology included complicated, sophisticated and confusing.

Feedback from the reference group

There are a number of versions which the evaluators are asked to produce, each with a significant number of comments to address. The comments are often in their hundreds and often refer to choices of specific words and formulation of sentences, sequencing of text etc. As well as comments in the text, there are also separate sheets with comments. There were three versions of the inception report, two versions of the desk report, and three versions of the draft final report. The approval of the second version of the draft final report releases a considerable payment to the evaluator. There is usually a big difference between the first and second version.

However, some people came in at a later stage and were given permission to make comments by the evaluation manager – there do not seem to be clear rules about who can comment and by when. Latecomers, perhaps due to staff rotation, questioned issues which had already been resolved or signed off on or had little enthusiasm to begin with. The evaluation manager did not play a role in consolidating feedback from the commenters, with the evaluators left to deal with contradictory comments.

The intensity of the exchanges by email and the number of comments received may have signalled that ownership by the RG and Thematic Unit in particular strengthened over the course of the second attempt of the evaluation.

Analysis

Ultimately the evaluation questions seem to have been well defined and comprehensively answered: set out in nine EQs that cut across the DAC criteria, EC criteria (coherence, added value) and crosscutting criteria (3Cs, visibility and cross-cutting issues e.g. gender and HIV/AIDS). The nine criteria are (main report vol. 1, table 2):

EQ	Type
To what extent is EC support aligned to education development objectives in national development plans, such as PRSPs, and ensured coherence between EC development co-operation policies on education and other EC policies affecting education?	Relevance
To what extent has EC support to education contributed to improving access to and equity related to basic education?	Sector results
To what extent has EC support to education contributed to improving transition to secondary level (both lower and upper)?	
To what extent has EC support to education contributed to improving quality of education, at primary and secondary levels?	
To what extent has EC support to education contributed to enhancing basic education skills, especially literacy and numeracy?	
To what extent has EC support to education helped in improving education system service delivery and resourcing?	Governance and sector management
To what extent has EC support to education helped to strengthen transparency and accountability of the management of education service delivery?	

To what extent and how has the EC contributed to improving co-ordination, complementarity and synergies with MS and other donors in the education sector, in line with the Paris Declaration?	
To what extent have the various aid modalities, funding channels and instruments, and their combinations, in particular GBS/SBS/SSP/projects, been appropriate and contributed to improving access to, equity of and policy-based resource allocation in education?	Aid effectiveness

And a subsequent table shows how each of the EQs related to the DAC criteria, the 3Cs and the two value questions.

The QG says that it was a very good evaluation (some criteria were marked as good, but the overall quality rating was very good). This seems worth flagging so that others can learn; not only from the evaluation itself but possibly also from the way it was run.

The summary on EVINFO does not reference this: would it make people more likely to read it, and the conclusions, if it was flagged as being very good?

Recommendations

The draft analysis was in the region of 60 pages. Major points were identified for further consideration. From that the team on their own, considered what the EC could do to address these points, with information on who should do it and when. In fact the QG acknowledges (favourably) that the report assigns responsibility for implementing the different recommendations. The final recommendations were the result of a bargaining process between the evaluators and the evaluation manager/reference group.

The evaluators put forward a couple of high level political recommendations which addressed the EC's prioritisation of education issues and budget allocation, as well as a series of more technical recommendations aimed at programme managers in country Delegations regarding design and implementation. The way they are presented makes sense: the overarching recommendation is strategic, but it is broken down into up to four 'elements' which provide more concrete actions that can be undertaken. All recommendations appear to have been quite feasible, judging from the amount of concrete follow-up it is possible to see from the FC.

The evaluation set out the international policy framework in the education sector and how the EC supported it, and different country profiles which gave nuance to the recommendations. However, this did not make it through to the summary document, which therefore becomes more of a bureaucratic document than a strategic one.

Interestingly, one interviewee said they were able to remember the recommendations from the Court of Auditors special audit and not those of the evaluation, mainly due to the succinct and small number of key recommendations made by the former.

Dissemination seminar

The Evaluation Unit takes responsibility for organising the dissemination seminar. About 50 participants were invited by the Evaluation Unit. It was held in Brussels and lasted a day. This seminar was the last one to be undertaken in the old format where there were four sessions, each 1 ½ hours long – findings, conclusions, recommendations and Q&A – rather than focussing each session on a set of issues. This made the engagement between presenters and the audience challenging, as an audience member would often ask "so what do you suggest the EC do about access to education" with the response often being – "you'll have to wait till session three", by which time (say two hours later) most people had forgotten what the issues were.

A senior manager provided a well thought out, mainly positive response to the evaluation during the seminar indicating also that the Commission already taken steps suggested by the evaluation team.

The dissemination seminar minutes captured some of the possibilities for uptake, though at a strategic level rather than an operational level. The final section of the minutes of the dissemination seminar does note some specific instances of things which had been done since 2007 to improve education programming, but does not take a detailed forward look – as alluded to under timing above.

Management response and implementation

The evaluator's role ceased to continue at the end of the dissemination seminar.

The initial management response was rather formal and high-level; it does not disagree with the results of the evaluation but the response is strategic rather than substantive.

The FC sets out concrete actions which have been taken; all of which are in some way related to the evaluation. Given the number of other donors working on education the political recommendations (e.g. rebalancing sectoral priorities towards secondary education and VET) had to be agreed at a strategic level.

Many of the technical sub-recommendations appear to have been followed up: they are well laid out in the evaluation document. For instance: in R6, recommendations to assess the feasibility, validity and usefulness of indicators for education-variable tranche release in GBS and SBS have been addressed somewhat through a variety of guidelines, advice on choice of indicators and learning events with a focus on indicators. Further, R8 (integrate governance reform into education sector reform) has been partially addressed by papers on decentralisation in education, an education briefing on PFM and country work to strengthen the link between education as a vehicle for broader PFM reform and the reverse; using broader PFM reform to strengthen the education sector.

With other donors heavily involved in education, issues such as reconsidering sector priorities (R1), engaging with the FTI to promote reforms (R3), differential targeting of education sector support (R4) – pressure for change may have come from external actors as well as from the evaluation. Moreover, as alluded to above, many of these initiatives seem likely to have happened with or without an evaluation.

The follow-up after three years – dated 14 May 2013 (as opposed to the standard one year) gives a unique insight into what Commission services consider successful uptake: strategic policy responses, short-term management responses, amongst others things.

5.3 Conflict Prevention and Peace-building

Name of evaluation	Thematic Evaluation of European Commission Support to Conflict Prevention and Peace-building
Language	English
EuropeAid reference nr.	1277
Type of evaluation	Thematic evaluation
Year of evaluation	2010-2011
Period under review	2001-2010

Key Points

1. The conflict prevention and peace building evaluation is generally seen as a very useful document, not only by the Commission and EEAS services, but also by external actors. As one of the ingredients in the debate, it helped the EU define its role in the prevention of conflict in EU external action. At the time of the evaluation a number of important factors converged which greatly facilitated uptake of the evaluation:

- ✓ **Opportune timing.** The evaluation process started in 2007 and was therefore well underway when the EEAS was being set up. The evaluation was carried out at a time when the revision of the EU's conflict prevention and peace building architecture was very high on the agenda, and was therefore able to feed into this process extensively. After ten years of the Gothenburg process, the political traction to strengthen the EU's Conflict Prevention mandate was particularly strong. In the past few years, the EEAS and EuropeAid defined their mandate and developed a number of mechanisms which partially build on the recommendations of the evaluation. Two recent examples are a guidance note on conflict prevention, and a new conflict analysis framework.

- ✓ **Quality and content.** There is broad consensus on the quality and relevance of the evaluation. This is reflected in the Quality Judgment and Fiche Contradictoire, and is acknowledged by its users. This strengthened the credibility and usability of the evaluation. The strong focus of the evaluation on the integrated approach ensured the relevance of the evaluation for EU policy development.

- ✓ **Unique in its kind.** The evaluation was the first extensive analysis of the EC's interventions in the field of CPPB; as such it was able to uncover the full extent of the EC's previous interventions in this domain. This added to the sense of urgency to act on the evaluation and take into account its results.

2. **Evaluation process and policy development:** the production of the evaluation united key EU stakeholders in the reference group of the evaluation. As a result, many of the findings and results were taken up in the internal discussions which led to the 2011 council conclusions on conflict prevention and peace building. The evaluation is also one of the ingredients of an upcoming joint communication on the Comprehensive Approach to Conflict Prevention, which is currently being developed by the EC and the EEAS.

3. The evaluation is an example of the diverse ways in which strategic evaluations can feed policy and programming processes. In this particular case, the high profile which was given to the process itself maximised uptake when the evaluation was being carried out. After the evaluation was released it remained very relevant and is taken up in the discussion on new developments to this day.

Timing and Context of the Evaluation

The evaluation started during the Lisbon negotiations and took much longer than usual. It was carried out in three different steps which were deemed necessary to clarify the scope and conceptual basis of the evaluation. In 2007 a preliminary study was carried out with the objective of clarifying inventory and typology of the funding in the field of conflict prevention and peace building and to look at the existing policy framework. In 2010, the process entered a second stage with the recruitment of two experts to accompany the process. A concept study was first carried out to examine the concept of the "integrated approach" and work out an analytical framework. The actual evaluation was launched in July 2010; the final dissemination seminar took place in November 2011. As a result, the evaluation was considerably more costly than usual.

Political and Institutional Developments

As a result of the extended timeline, the process of carrying out the evaluation coincided with a number of key institutional and policy developments. The evaluation came at a time that conflict prevention was particularly high on the EU agenda, which brought it on the radar of both Commission and EEAS services. Three developments are particularly important in this respect:

1. **Gothenburg process review:** In 2010, the review of 10 years of the Gothenburg process was initiated. Even though the review process was discontinued[10], it put conflict prevention on the political agenda of the EU. Several member states pushed the EU to strengthen its conflict prevention framework.

2. **A new EU conflict prevention architecture:** Even though the process had started in 2007, in the midst of the Lisbon negotiations, the eventual evaluation was carried out in 2010. This meant that the evaluation was well underway when the EEAS and EuropeAid were established. By 2010, it had become clear that the two new institutions would share the conflict prevention mandate at EU level, and that a new conflict prevention and crisis management architecture would have to be developed. The reorganisation of the EU external action architecture coincided with the evaluation process and the relevance of the evaluation for these institutional developments was underlined by several interviewees.

3. **Council Conclusions on Conflict prevention:** In June 2011, the EU produced council conclusions on conflict prevention. At the time of drafting the conclusions, the evaluation was well advanced. Through the Reference Group there were many opportunities for exchange within the 'EU conflict prevention community'.

User base

The user base and target group of this evaluation includes:

1. European Commission Services: EuropeAid 07. Fragility and Crisis Management; EuropeAid B5. Instrument for Stability; and senior management.
2. EEAS services: EEAS K.2 Conflict prevention, peace building and mediation instruments, nuclear safety; EEAS VII.1 Crisis response planning and operations; and senior management.
3. Member states
4. Civil Society and specialised organisations such as the European Peace Building Liaison Office (EPLO).

Uptake of evaluation results

The evaluation was carried out during crucial times for the EU conflict prevention dossier, and is widely acknowledged as a very useful document. It is however not easy to pinpoint the exact contribution of the evaluation, nor can we simply trace back developments to the recommendations of the evaluation alone. The evaluation is part of a wider debate on the role of the EU in conflict prevention; it concerns institutional change and touches on the age-old discussion on the position of the EU in its foreign security policy. As a result it is only one of the elements which feature in policy discussions.

Nevertheless, a number of institutional and policy processes were cited as being *"well influenced by the evaluation"*. The general rule is that the evaluation acted as one of the ingredients leading up to a certain change. As an independent high quality evaluation, it has a particularly authoritative position in this discussion. The fact that the evaluation "followed the money", and unveiled how much the EU had already invested in it further added to its authority, and made it a document which "could not easily be ignored".

---

[10] EPLO (2012) <u>The EEAS and Peacebuilding One Year on.</u>

A new EU conflict prevention architecture

The high profile of the evaluation prompted a close follow-up of the process by the relevant Commission services. The Reference Group included decision-makers and the ensuing discussions were therefore able to feed into on-going debates and institutional changes.

EEAS and EuropeAid
The EEAS came in relatively late in the evaluation process because the relevant staff (RELEX and CMPD at that time) were not sufficiently available, and because consultation had been limited in the early stages. In the second stage of the process, following the recruitment of two experts, the rationale behind the evaluation became much clearer. The evaluation process was therefore able to feed into the reflection on the establishment of a dedicated Unit in the EEAS (currently K.2). The evaluation helped the EEAS team identify priorities and respond to the disentanglement of the institutional setup. Unlike EEAS K.2, EuropeAid's Fragility and Crisis Management unit (currently 07) recruited many new people from outside the institutions, in a way this facilitated uptake because the evaluation was not seen as criticising the Unit's activities. In addition, the evaluation provided an exhaustive overview of the Commission's track record, which could serve newly recruited staff.

Conflict Analysis guidance and framework
The EEAS and EuropeAid recently released a joint guidance note and framework for conflict analysis. The guidance note responds to Recommendation 2.2 *"develop and implement a systematic and structured approach to conflict analysis (…)",* and cites the evaluation on page two, outlining the EU's ambition to work more on conflict. The conflict analysis guidance is accompanied by joint trainings and workshops for staff in EU Delegations and HQ. The conflict analysis guidance is an important step towards a stronger EU conflict prevention function, and an early achievement of the new institutional architecture.

The programming guidelines for the next multiannual financial framework were published in 2013, staff working on conflict prevention were able to partly influence the guidelines, however, the conflict analysis guidance document was still under development when the programming guidelines were finalised, which limited the conflict analysis component to some extent.

EU policy development

June 2011 Council conclusions on conflict prevention
In June 2011, the EU produced council conclusions on conflict prevention. At the time of drafting the conclusions, the evaluation was well advanced. Through the Reference Group there were many opportunities for exchange within the 'EU conflict prevention community'. Several elements of the evaluation feature in the Council conclusions, including the emphasis on early warning and conflict risk assessment. This is not necessarily evidence of uptake; however, it does show that both documents paint a similar picture.

Upcoming communication on the comprehensive approach to external conflict and crises
A major contribution of the evaluation was that it clarified the integrated approach to conflict prevention, which was most useful for the on-going development of a joint communication on the comprehensive approach to external conflict and crisis. The lead services for this communication are EuropeAid 07 and EEAS K.2. The draft communication is currently in the consultation stage, and has been there for a while. EuropeAid sources mentioned that there are a lot of similarities between the evaluation and the upcoming communication, and a number of components which speak to the recommendations.

Recommendations

A number of other developments can partly be attributed to the evaluation. Table 1 gives a schematic overview of the evaluation's recommendations and all the relevant developments, based on a literature review and interviews.

Recommendations	Specific Component	Relevant Policy/institutional/programming developments
On the Commission's overall role and approach		
R1: Strengthen the Commission position as a key player in CPPB	Together, these elements constitute a good case for advocating a strengthening of the Commission's position as a key player with respect to CPPB. This strengthening should be done in coordination with the High Representative.	✓ New institutional architecture: EuropeAid 07, EEAS K.2 share the responsibility for strengthening the EU position as a key player
R 2: Strengthen the integrated approach	R2.1 Ensure clarification and common understanding of concepts among EEAS and Commission staff, including at operational level. (based on C2.1.1)	✓ Conflict prevention group ✓ Conflict Analysis Guidance ✓ Conflict Analysis Training and Workshops ✓ Upcoming communication on the comprehensive approach
	R2.2 Develop and implement a systematic and structured approach to conflict analysis, mainstreaming and "do no harm". (based on C2.1.2)	✓ Conflict Analysis Guidance ✓ Conflict Analysis Training and "light-touch workshops" in selected conflict countries ✓ Minor elements in the new programming guidelines ✓ Conflict dimension in oQSG
	R2.3: Create a comprehensive, easy and flexible early-warning system and make sure it is used. (based on C.2.1.3)	✓ Conflict prevention group and Crisis Management Board ✓ Early warning system in development ✓ Pilot in eight Sahel countries
	R2.4 Strengthen the synergies between the different geographical levels of intervention. (based on C.2.1.4)	✓ Limited progress
	R.2.5 Make sure that coordination mechanisms at all levels, but especially between the Commission, the EEAS, the EU Council and EU MS, go beyond a mere exchange of information and aim at enhancing complementarities at strategy and implementation levels. (based on C.2.1.5)	✓ Increased coordination between EuropeAid, FPI and EEAS ✓ Conflict prevention group ✓ Joint guidance ✓ Upcoming communication
On specific strategy issues		
R 3: Clarify role to play in conflict countries by focusing on crisis management efforts and on tackling directly the root causes		✓ Council conclusions on conflict prevention ✓ Upcoming communication on the comprehensive approach ✓ Emerging approach to fragility (new deal)

R 4: Leverage Commission's financial weight with non- financial support		✓ Scaling up of Mediation and dialogue work ✓ Upcoming joint communication on the comprehensive approach (right mix of instruments, combining cooperation, diplomacy and political dialogue)
R 5: Relevance of alignment		
On means and implementation		
R 6: Make appropriate means available	R6.1: Design and implement a specific human resources policy for intervening in a post-conflict or conflict (-prone) context (based on C.10.2)	✓ EEAS K.2 and EuropeAid 07 staff recruited ✓ Crisis response planners and project managers recruited under the IfS ✓ Conflict Analysis Training and "light-touch workshops"
	R6.2: Provide mechanisms to ensure effective knowledge management (based on C.10.2)	✓ Factsheets (EEAS) ✓ Training
R 7: Maintain "protective" character of procedures but make them swifter		✓ New programming guidelines
R 8: Better anticipate conflict-related challenges		✓ Early warning system in development ✓ New programming guidelines

External uptake

The evaluation is widely cited, both in EU documents and external documents. Member states and civil society took up the evaluation report in their positions on the EU's evolving conflict prevention architecture:

- A preparatory document for the UK balance of competencies report of 2013 cites the evaluation to argue that the EU could do more and better in the field of conflict prevention.[11]
- Over the years, the European Peace Building Liaison Office (EPLO) has used the findings and recommendations in a number of its actions[12] and has actively worked to *"ensure that the findings of this long-awaited evaluation are translated into concrete actions for increasing the effectiveness of the EU's support to conflict prevention and peace building"*[13].

These visible traces of uptake are a clear sign of the high profile of the evaluation, but are also an indication of its continued relevance.

---

[11] Conciliation Resources. (2013). Evidence for the Foreign Policy Report on the Balance of Competencies between the United Kingdom and the European Union: Focus on Peacebuilding and Conflict Prevention.

[12] See for example EPLO (2012) The EEAS and Peacebuilding One Year on.

[13] EPLO Blog. Evaluation of EC support to conflict prevention and peacebuilding.

5.4 Egypt

Name of evaluation	Country strategy evaluation Egypt
Language	English
Type of evaluation	Country evaluation
Year of evaluation	2010-2011
Period under review	2001-2010

Background

This evaluation was selected for a 'process-tracking' for two main reasons. First, Egypt had been subject to an earlier evaluation in 2004. It could be interesting to use the 2010 evaluation to assess the levels of uptake of previous recommendations. Second, soon after the completion of the 2010 evaluation, the Arab Spring erupted and the Mubarak regime fell. Subsequently, EU policies towards the MEDA region and the country changed quite fundamentally. This also provided an opportunity to check whether the 2010 evaluation had an influence on these policy developments.

There is, however, also a limitation to this case study. It proved difficult to involve key players in the analysis, as most of them had moved on and could not be traced back easily (e.g. the evaluation manager). Hence, the number of informants is limited (3). However, an extensive desk review helped to partly compensate for this.

Key Points

1. The 2004 country evaluation was appreciated as it came at an interesting moment in EU-Egypt cooperation. It helped the EU to start making the shift from a predominantly 'project mode' of operations to more structured forms of cooperation, focusing on policies, domestic reform agendas, BS and dialogue. This is an evolution which also took place in other countries of the MEDA region in the framework of the European Neighbourhood. Such a shift entailed both strategic and operational re-orientations of EU interventions. The Fiche Contradictoire shows that the EUD agreed with most recommendations and proposed a number of interesting reform paths to address the issues raised.

2. One of the reasons why Egypt was again selected for a country evaluation, five years later, was precisely the opportunity this would provide to check actual changes in EU cooperation approaches towards Egypt. The point was also made that the Evaluation Unit itself had insisted on the inclusion of two themes in the EQs, dealing respectively with human rights & democracy and with employment on the other side. The latter is an interesting choice as the topic was not a sector of concentration of the EUD. Yet it was perceived to be a fundamental development challenge for the country and a key component of the Association Agreement. It therefore presented a challenging entry point to question the overall pertinence of the EU intervention strategy.

3. Yet from the interviews it clearly appeared that the EU Delegation was not very keen to again be the subject of an evaluation. The demand clearly came from Brussels and ownership levels/interest at the level of the EU Delegation is reported to have been minimal.

4. This did not facilitate the implementation process. According to the evaluation team involved, there was a "blockage right from the start". The levels of collaboration received from the EUD would continue to be very limited throughout the process. The 'hostility' was not confined to the process

alone. It also expressed itself on matters of substance. At each step of the process, the EUD reacted rather defensively when policy choices were questioned (e.g. the rationale for the overall aid amount allocated to Egypt) and in some cases criticised (e.g. the programming process or the way BS was handled). This resulted among other things in a large number of comments on intermediate reports of the consultants. Inevitably, it also led to a rather protracted implementation process (the evaluation was part of the 2008 work programme, the field work was done in mid-June 2009 and the report completed in December 2010).

5.   In order to defuse the tensions and get the process rolling, the Head of the Evaluation Unit had to intervene on several occasions. He also decided to go personally to the dissemination seminar in order to ensure that things would run smoothly and a constructive dialogue could take place around the evaluation findings with all relevant stakeholders. The consultants are clear that this 'protection' by the Unit was key for them to go on and write an independent report. This complicated implementation process is not reflected in the quality grid, which gives overall positive scores to the evaluation.

6.   From this rather conflict-ridden situation a number of lessons can be drawn:

- The 2010 evaluation confirmed that EUD often see this type of exercise as an 'audit' rather than as a learning opportunity
- There can be a strong reluctance at the level of EUDs to really go deep into the underlying political/strategic choices which underpin a CSP. Yet this is, in principle, the very purpose of a strategic evaluation and also its potential added value – as it may allow 'anticipation' of changes in the country and the required adjustments of EU responses. As the consultant put it: "in order to raise the issue of the pertinence of the overall strategy you often have to go through the backdoor".
- The rather rigid methodology (with its intervention logic and EQs) is seen to further complicate such fundamental questions about strategy as it may lock the evaluation team into a straightjacket.

7.   In terms of uptake a distinction must be made between the two evaluations (2004 and 2010).

- Several of the recommendations of the 2004 evaluation (e.g. the need for an exit approach for the project mode, a stronger focus on dialogue) were picked up and led to gradual changes in EU approaches[14]. Yet this should not be solely attributed to the impact of evaluation. The move towards more structured forms of cooperation, based on a solid programming processes and dialogue with the partner country, reflected a general evolution of EU cooperation.
- The uptake of the 2010 evaluation was a different story. The tensions between evaluators and the EUD – which also concerned fundamental issues of substance[15] - were not conducive to ensuring a real follow-up to the evaluation. The long duration of the process (3 years) also meant that some of the findings/results were no longer relevant. In addition to this, barely a month later, the Arab Spring changed the whole political/institutional environment in the country and region. The events would spur a revision of the whole EU partnership with Egypt.

---

[14] The FC is interesting, because it 'tracks' what is picked up and what is not retained (without explanation). For instance, the suggestion to focus more on 'balanced social development' was not picked up (despite its critical importance as later events would demonstrate). The issue of human rights was picked up but the response strategy was mainly to use EIDHR. While this is undoubtedly an important instrument, the evaluation recommendation went much further than to support civil society through a set of projects. It could be argued that there was an 'uptake' here of the recommendations, yet one could certainly discuss the overall pertinence and coherence of the EU response provided (as later events would demonstrate).

[15] The dispute appears to have been the reluctance of the EUD to ensure effective policy dialogue around the BS provided. As mentioned in the report, the Delegation focused primarily on strengthening PFM through BS. They were not convinced about the pertinence/feasibility of engaging in policy dialogue.

In this process, space was created for the EU to reconsider wider political and strategic questions – which could not be optimally addressed in the 2010 evaluation for reasons mentioned above (see par. 6, second bullet) such as human rights, civil society, political/policy dialogue, employment and inclusive growth.

8. Discussions with staff from the EUD at the time of the evaluation helped to understand some of the broader dynamics which may condition uptake of evaluations. Three points can be highlighted:

- Strategic evaluations have the potential to help steer a debate on needed reforms in cooperation approaches. At the time of the second evaluation, some staff at the EUD felt that the portfolio still had important (self-standing) 'projects' (with limited perceived relevance) and BS in sectors where the EC lacked both expertise and leverage (considering the limited amount of funding involved). There was no explicit strategy to use EU resources in areas which were fundamental for implementing the Neighbourhood Policy – with its explicit political choice for promoting alignment of partner countries to the *'acquis communautaire'* and supporting systemic reforms. In such circumstances, a well-targeted strategic evaluation could be a useful tool to open-up the debate, adopt a helicopter view, put a number of 'existential questions' on the table ("what are we doing in this country?", "do we need so much money here?") and get a process of rethinking rolling. The 2010 Strategic Evaluation was not used for such purposes, for a variety of reasons, including defensive attitudes within the EUD and a lack of understanding of the potential value of strategic evaluations to trigger such change processes.

- The dynamics at country level tend to evolve all the time and often quite quickly. Both our programme cycle and the cycle of evaluations are not adapted to deal with these dynamics. This impinges negatively on the quality of our cooperation. It also reduces the chances of learning and use of evaluative findings. In order to have a real added value, strategic evaluations should be able to 'play shorter on the ball' and provide critical reflections on strategy at the 'right moment'.

- Another added value of strategic evaluations could be to better understand these local change processes. This will become increasingly important as future EU cooperation in many countries, including Egypt, will be based less on the 'amount of money spent' and more on 'how to influence structural reforms from the outside?' – including through the transfer of knowledge. In this context, it would be useful to have a new type of evaluation providing generic lessons on how to influence systemic changes. The uptake of such knowledge would be much easier than evaluations which are too comprehensive and arrive too late to influence real-time cooperation processes.

6. SURVEY OF EU DELEGATIONS

The study team, in cooperation with the European Commission launched a survey for EU Delegation staff in developing countries. The survey was sent to approximately 100 Delegations; a total of 99 responses from 54 different Delegations were recorded by 12 October 2013.

The survey built on a similar (internal) exercise that the Evaluation Unit undertook in 2007, when it sent a survey to EC and Delegation staff to map the use of evaluation results in their work. The current survey asks many of the same questions but goes deeper into how EU Delegation staff use evaluations and how they value the system which is currently in place.

The survey had multiple objectives:

1. Assess the evolution of the use of strategic evaluations in comparison with the 2007 survey results;

2. Determine the position of strategic evaluations in the wider knowledge setup in Delegations;

3. Examine the use of evaluations in EU Delegations and identify the extent of uptake in the different activities at Delegation level; and

4. Identify specific avenues for improving the uptake of evaluation results.

The survey consists of four sets of mainly multiple-choice questions. The first series of questions looks at the professional profile of respondents, their geographic distribution, professional interests and their familiarity with the evaluation process. This is necessary to connect the conclusions from the survey results with a specific target group of evaluation material. The second series looks at the knowledge and learning culture in the Delegations. The third series of questions goes deeper into how Delegation staff uses evaluation material in their daily work, and a final series of questions looks at the main bottlenecks and opportunities for improving the evaluation system.

Main findings of the survey include:

1. Since 2007, there is a high degree of continuity in how people use evaluations. The same is true for how the quality of evaluations is perceived.

2. Delegation staff are familiar with the system and use evaluations very selectively in their professional activities.

3. Evaluations are used as a complementary source of evidence.

4. Strategic evaluations in their current form are not adapted to the work environment in Delegations:

   • The timing of the evaluation cycle is not in tune with the programming cycle;

   • Limited involvement of Delegation staff in the design and implementation of evaluations has a negative effect on the relevance of evaluation results in Delegations;

   • The current format of evaluation reports does not facilitate smooth uptake of results.

5. The work pressure and management culture in Delegations is not always conducive to lesson learning.

Methodological notes:
1. Several respondents indicated that their survey response related to evaluations which were managed by the EU Delegation itself, such as project and programme evaluations. This indicates that EUD staff do not necessarily make a clear distinction between strategic and project/programme evaluations. It is impossible to distinguish between types of evaluation from the survey data.

2. In a number of questions, the respondents were asked to rate or attribute a numerical value to a statement. For comparative reasons, incomplete responses were excluded from the survey data. The graphs below are based on valid responses only; therefore the sample varies slightly from question to question. The highest number of invalid responses recorded was 12%. As a result this does not affect the overall validity of the survey sample.

6.1 Target group

The survey was sent to all Delegations in developing countries, the Head of Delegation was asked to send the survey to relevant colleagues working in the Delegation. The largest share of responses came from Sub Saharan African Delegations (40%), followed by Asia and the pacific, Latin America and the Caribbean and European Neighbourhood countries. The average respondent's experience in the EU Delegation system was **3.8 years. (Q2)**

Q1: Geographic distribution of responses

- Africa — 40%
- Asia and Pacific — 14%
- Latin America and Caribbean — 20%
- European Neighbourhood — 13%
- Middle East — 10%
- Europe — 1%
- Unknown — 2%

Q3: Previous experience of respondents in the EU system

- Yes, for DEVCO (DEV or AIDCO) — 60%
- Yes, for the EEAS (or RELEX) — 6%
- Yes, for a different agency — 6%
- No — 28%
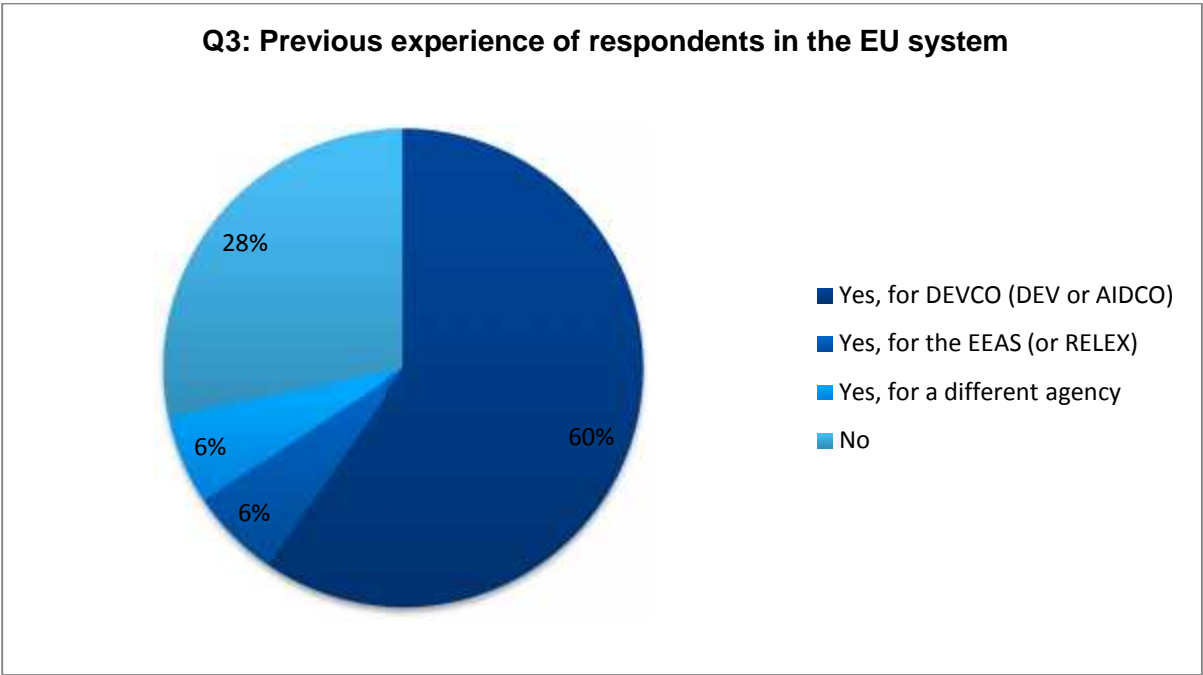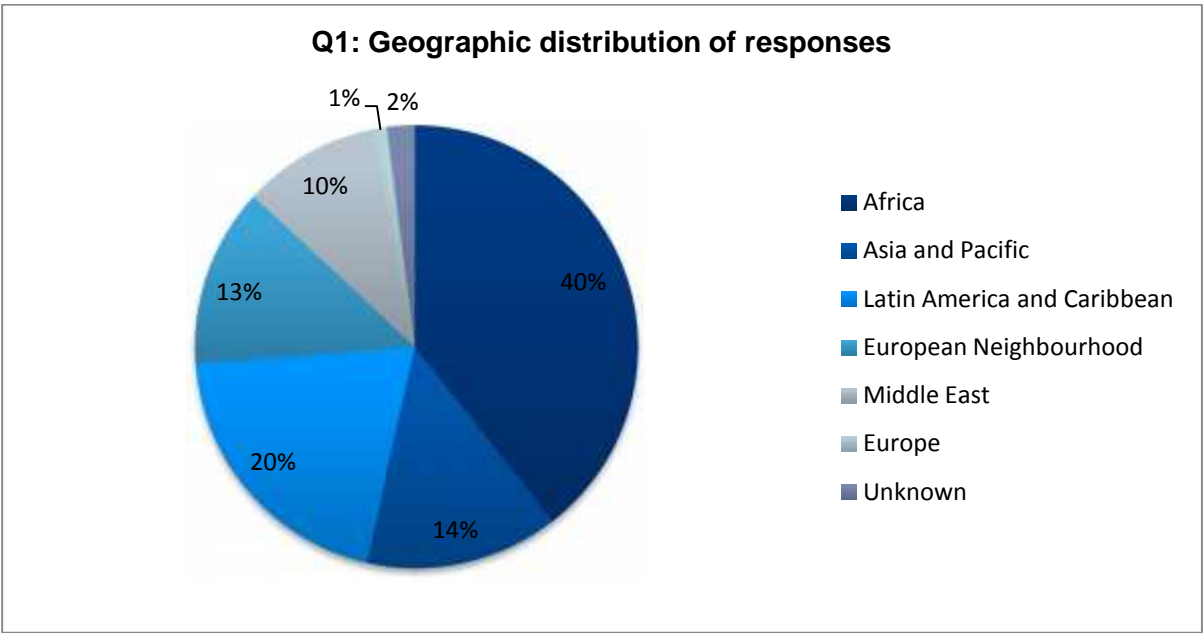
The main professional interests of the respondents are operational and development related. This is in line with the results of the 2007 survey[16]. There is a notable increase in respondents which are involved in policy and political dialogue, which reflects recent changes in EU external action.

Q4: Professional area of interest
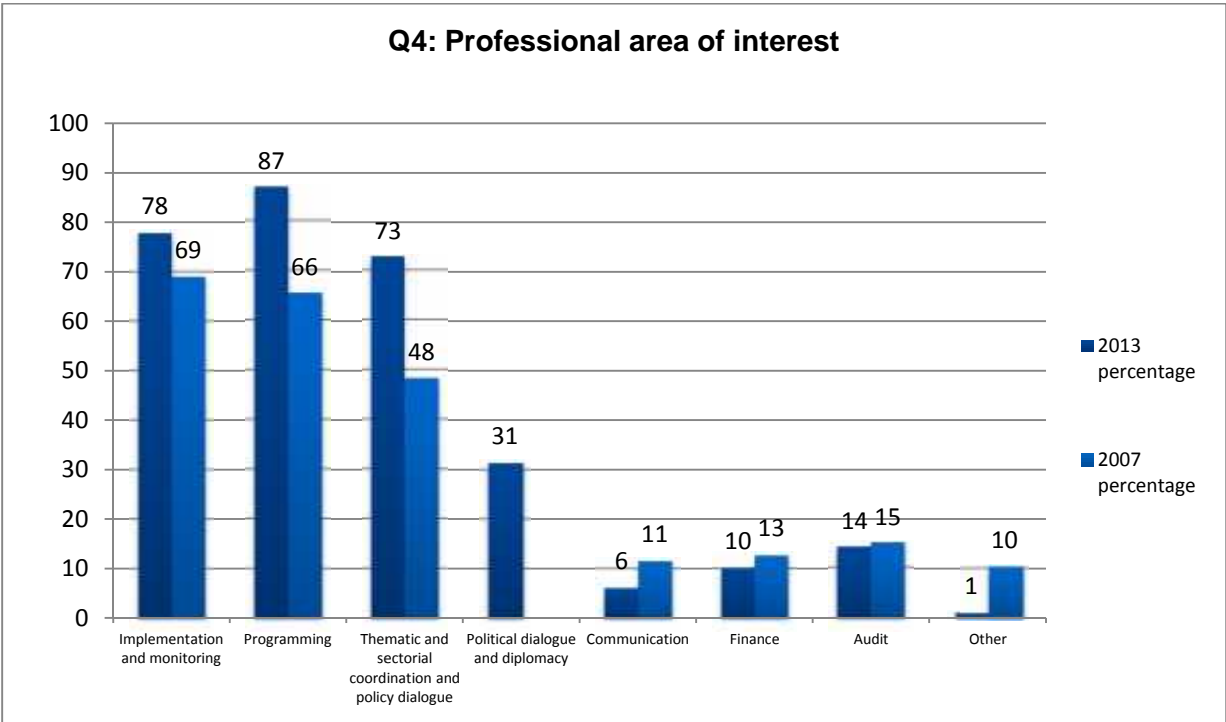
Most respondents (90%) have participated in at least one type of evaluation before. Country level evaluations take the lead, followed by sectorial and thematic evaluations, which reflects the composition of the evaluation programmes of the Unit. The two main evaluation-related activities are providing information to the evaluators (interviews) and commenting on the reports. This is also in line with the 2007 survey.

---

[16] 55% of the respondents in 2007 were Delegation staff, the rest worked in HQ.

Q5: Please indicate in which type(s) of strategic evaluations your delegation has participated in the past (percentage)

Category	Percentage
Country level evaluations	57,6
Sectorial evaluations (e.g. transport sector, health sector)	39,4
Thematic evaluations (e.g. human rights, good governance)	35,4
Regional evaluations (e.g. ECOWAS, ASEAN)	22,2
Evaluation of aid modalities (e.g. evaluation of budget support operations)	19,2
Joint Evaluation	1,0

Q6: How have you been involved in the evaluation process?

	2013 percentage	2007 percentage
I was interviewed by the evaluators	54	64
I commented on the report	51	61
I facilitated the evaluators' work in the EU delegation (2013 only)	40	
I participated in the final dissemination seminar	24	20
I was a member of the reference group	8	47
I have never been involved in any of the above activities (2013 only)	18	
Other or invalid	10	16

6.2 Knowledge and learning culture in EU Delegations

Delegation staff were asked to rate their agreement with five statements which relate to (a) the use of evidence and independent analysis in their daily operations (statements 1-2) and (b) the culture of learning in EU Delegations (statements 3-5). The first two statements received a fairly positive rating with mixed results for "management encourages the use of independent analysis in decision-making". This was not the case for statements 3-5 (learning culture). The survey results clearly indicate that staff feel that:

1. Knowledge needs are not being met in the Delegations
2. Knowledge transfer between rotating staff members is sub-optimal
3. The work environment is not always conducive to developing skills and expertise

This resonates in several of the individual statements which were recorded, often in relation to human resource constraints and information overload in the Delegations (see below).

Q7: Knowledge and learning culture in the EUD

Legend:
- 1 (completely disagree)
- 2
- 3
- 4
- 5 (fully agree)

Category	5 (fully agree)	4	3	2	1 (completely disagree)
Evidence informs policy decisions at the delegation level	15	34	27	18	2
Management encourages the use of independent analysis in decision-making	7	37	22	23	7
The work environment in the delegation is conducive to the development of staff knowledge and expertise	9	20	30	24	13
Knowledge transfer between staff (e.g. when new members arrive or others leave) takes place consistently	5	12	31	36	12
The delegation has all the mechanisms in place to meet my "knowledge needs"	2	9	31	34	20

Q8: What are the main sources of information for your professional activities?

	Interaction with colleagues (delegation and HQ)	Interaction with other domestic stakeholders (government, local authorities, civil society, etc.)	Interaction with other development practitioners (donor coordination WGs, etc.)	External research and documentation	EU produced or commissioned documentation (including evaluations)
1 (least important)	2	3	4	8	2
2	16	8	11	14	19
3	18	25	27	24	26
4	37	39	41	32	32
5 (most important)	22	20	12	17	16

A related question asked the respondents to identify their main sources of information:

1. Interaction with colleagues (Delegation and HQ);
2. Interaction with other domestic stakeholders (government, local authorities, civil society, etc.);
3. Interaction with other development practitioners (donor coordination WGs, etc.);
4. EU produced or commissioned documentation (including evaluations); and
5. External research and documentation

All five of these 'sources of information' are deemed important, which indicates diverse knowledge consumption. That said, **Delegation staff especially values interaction with colleagues and partners** as a source of information for their work.

6.3 Use of evaluation results

The majority of Delegation staff indicates using evaluations in their own work. The figure is only slightly lower than in 2007 (-6% find evaluations very or somewhat useful), which indicates that **staff use of evaluations remained largely the same in the past six years.**

Q9: How useful are evaluations for your work?

2007 Survey   **2013 Survey**

- Somewhat useful
- Very useful
- Rather not useful
- I don't know or evaluations are not relevant to my work
- Evaluations are useless (only 2013)
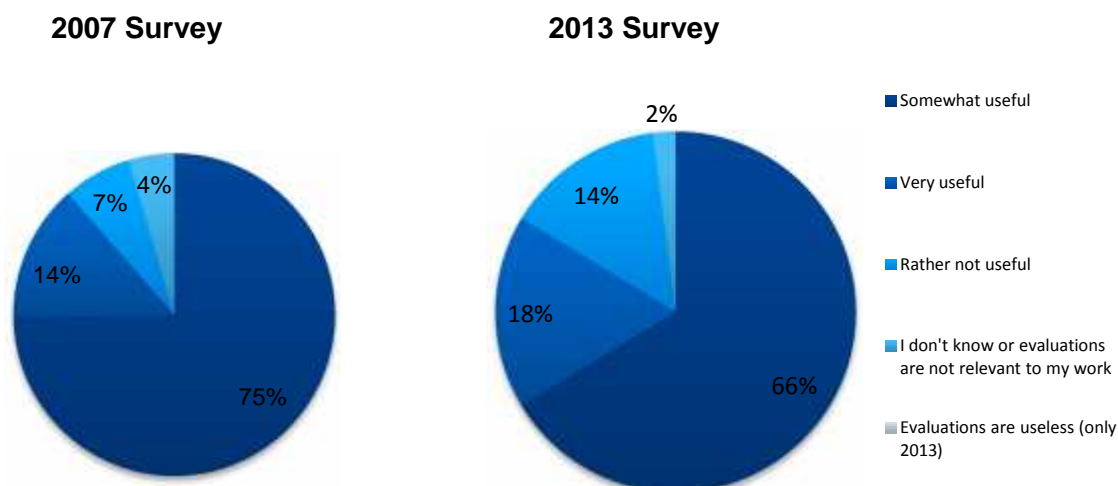
Q10: Where do you find evaluation reports?

- From the EuropeAid website
- During the final seminar
- Search engine
- From the evaluation unit as a member of the reference group
- from partners or colleagues (HQ and in country)
- CRIS
- Directly from the evaluators
- Delegation server/library

---
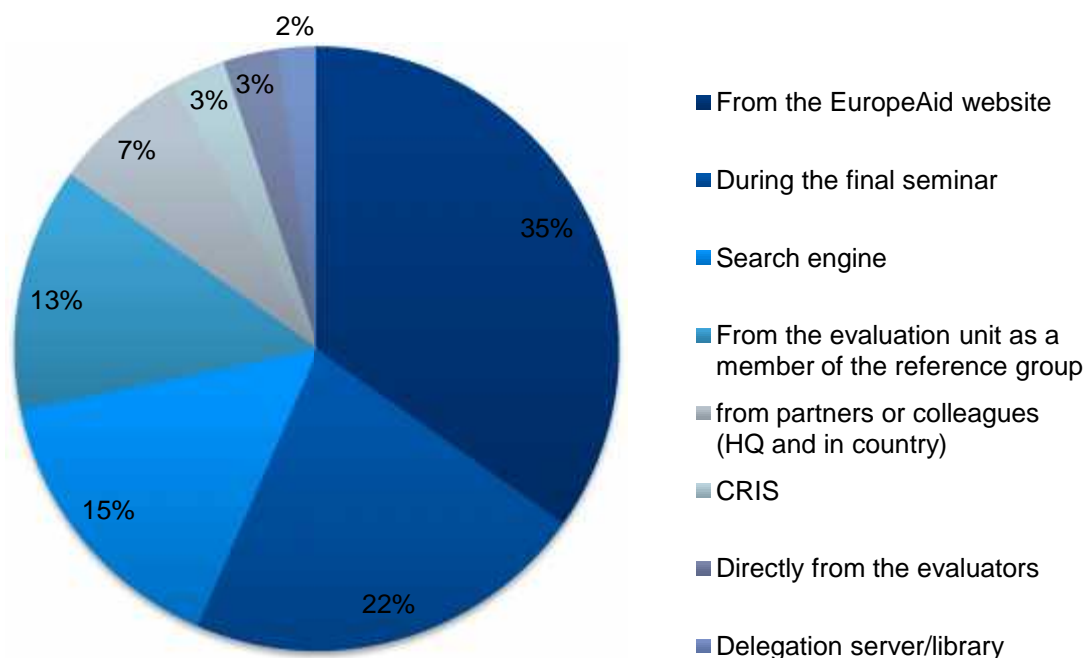
[17] Four respondents replied CRIS. This is most likely for project/programme evaluation reports, not strategic evaluations.

Delegation staff does not seem to have major difficulties accessing evaluations, though the majority does not rely on the EuropeAid website to retrieve evaluation reports. Hard copies handed out during the dissemination seminar prove useful (22%), and **only 51% look for the reports online.**

Q11: In your view, what are the main functions of strategic evaluations in EU development cooperation?
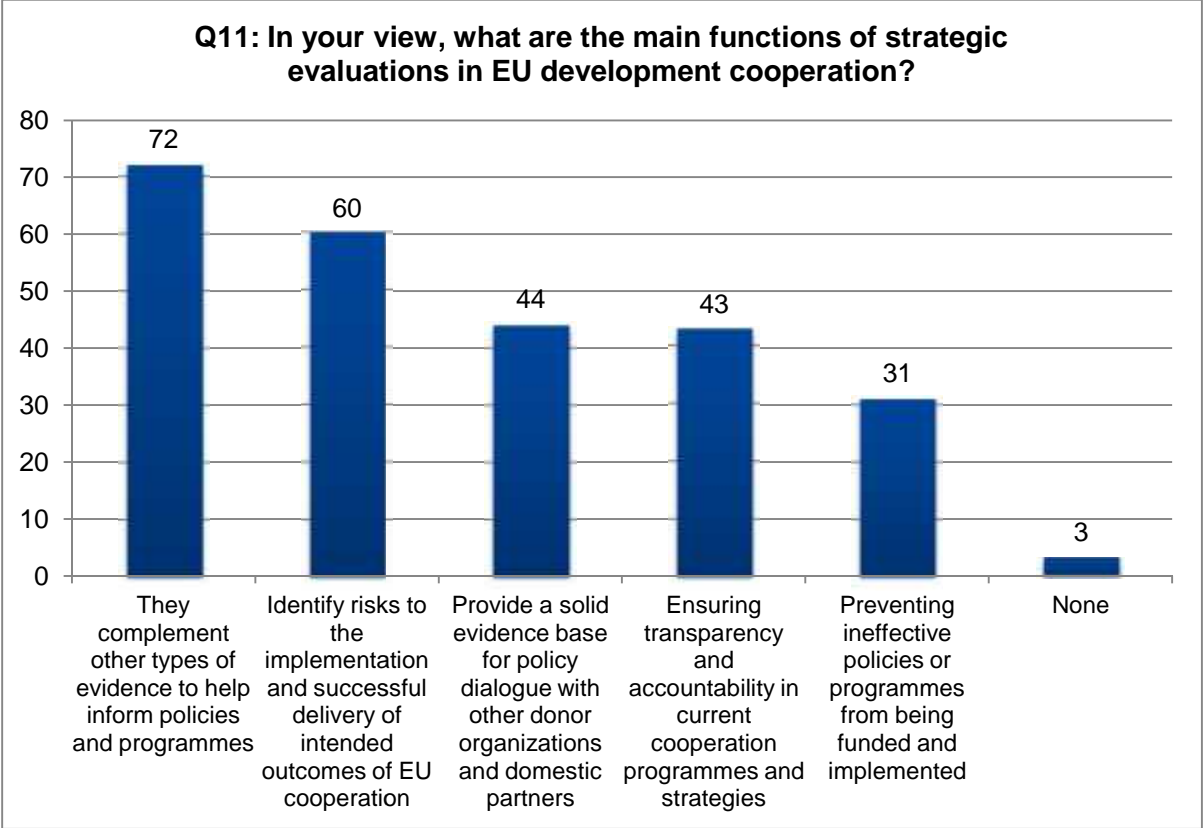
Staff in Delegations sees a broad range of uses for strategic evaluations, including risk identification (60%), ensuring transparency (44%) and preventing ineffective cooperation (31%). 44% of the respondents see evaluations as a useful evidence base to engage in policy dialogue with partners. **The main lesson we can draw from this question is that evaluations are one of many sources of evidence which are used in Delegations.** 72% indicated that evaluations are used in complementarity with other material.

Q12: Use of evaluations in respondents' own professional activities
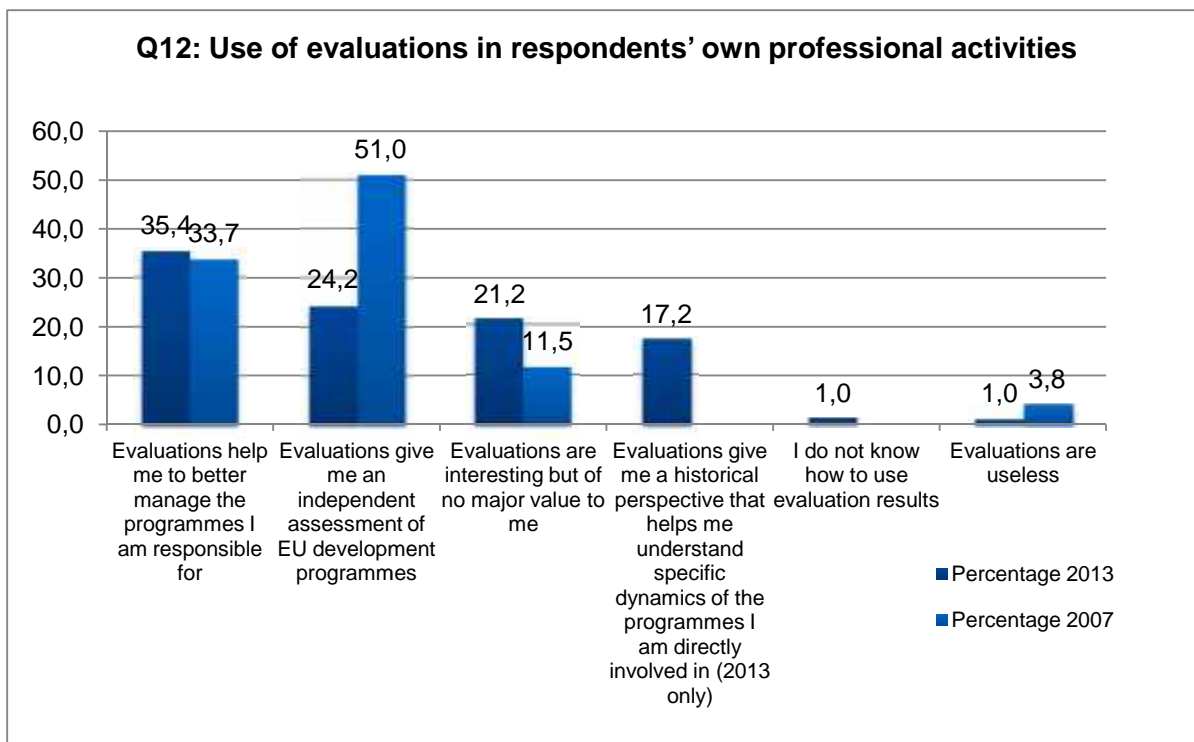
When asked how they use evaluations in their own work, 35% indicate using evaluations to help better manage the programmes in their area of responsibility. Given the high share of operational duties in Delegations this comes as no surprise. A remarkably smaller share turns to evaluations for their independent assessment of EU development programmes (24% as opposed to 51% in 2007). Only 21% indicate that evaluations are of no major value to their work and only 1% indicated that evaluations are useless. This shows that **the majority of Delegation staff see a clear use for evaluations in their professional activities.**

Quality of strategic evaluations

Q13: Has the quality of evaluations evolved over the years?
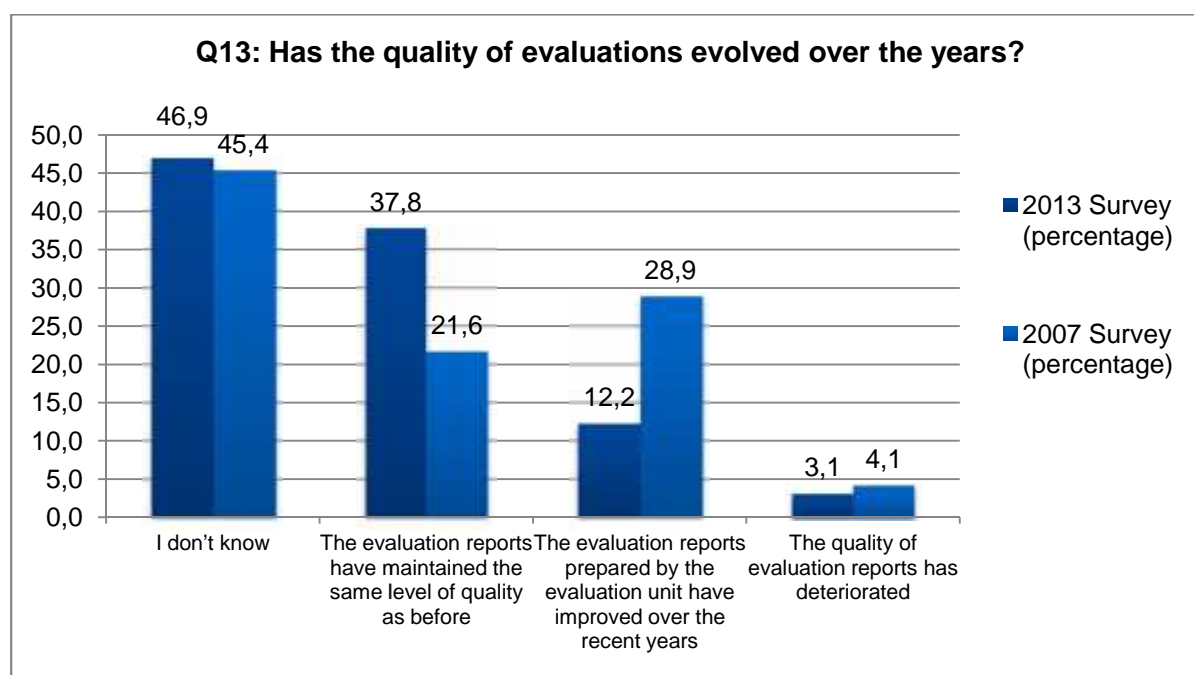
The quality of evaluations is often cited as a main factor in deciding their usefulness. The study team asked Delegation staff to judge whether the quality of evaluation reports has evolved over time. Almost half of the respondents do not have an answer to this question. This is most likely related to the high staff turnover in Delegations, and the level of seniority[18] of the respondents. There is a negative trend in whether people see an evolution in the quality of strategic evaluations. More people than in 2007 feel that the quality has been maintained and fewer people (12.5%) see improvement in the quality of evaluation reports. All this indicates that **the quality of evaluations stagnated since 2007.**

Q14: Which aspects of the evaluations are in need of improvement? (percentage)
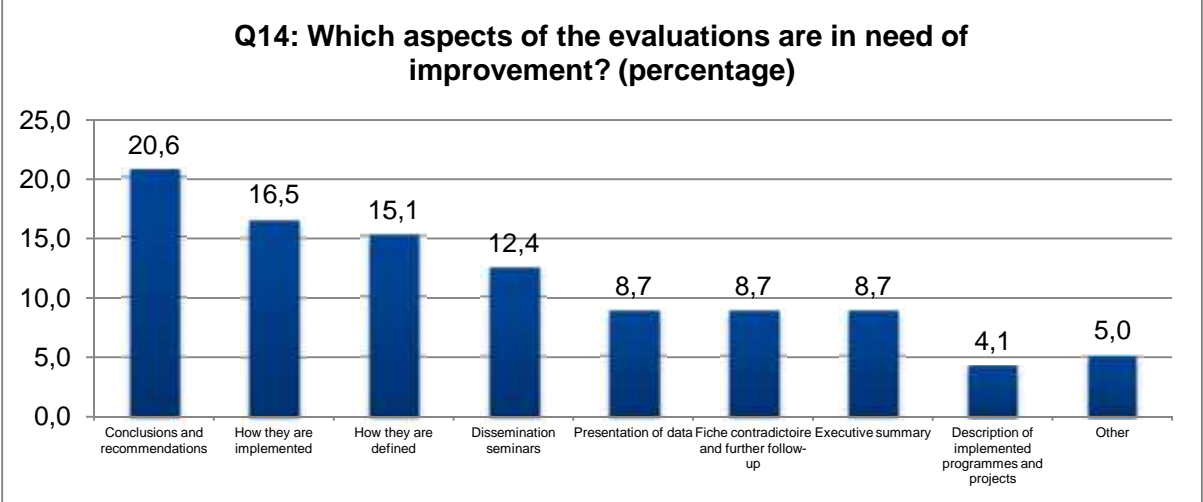
Staff in Delegations sees room for improvement in most components of the evaluation system. Most respondents indicate that the conclusions and recommendations leave room for improvement; this is in line with the results of the 2007 Survey. Interestingly, **many staff members also feel that the definition and implementation of evaluations is in need of improvement, much more so than the qualitative aspects of the reports**. Several individual statements support this finding (see below).

Q15: What are in your view the main characteristics of a good strategic evaluation?

---
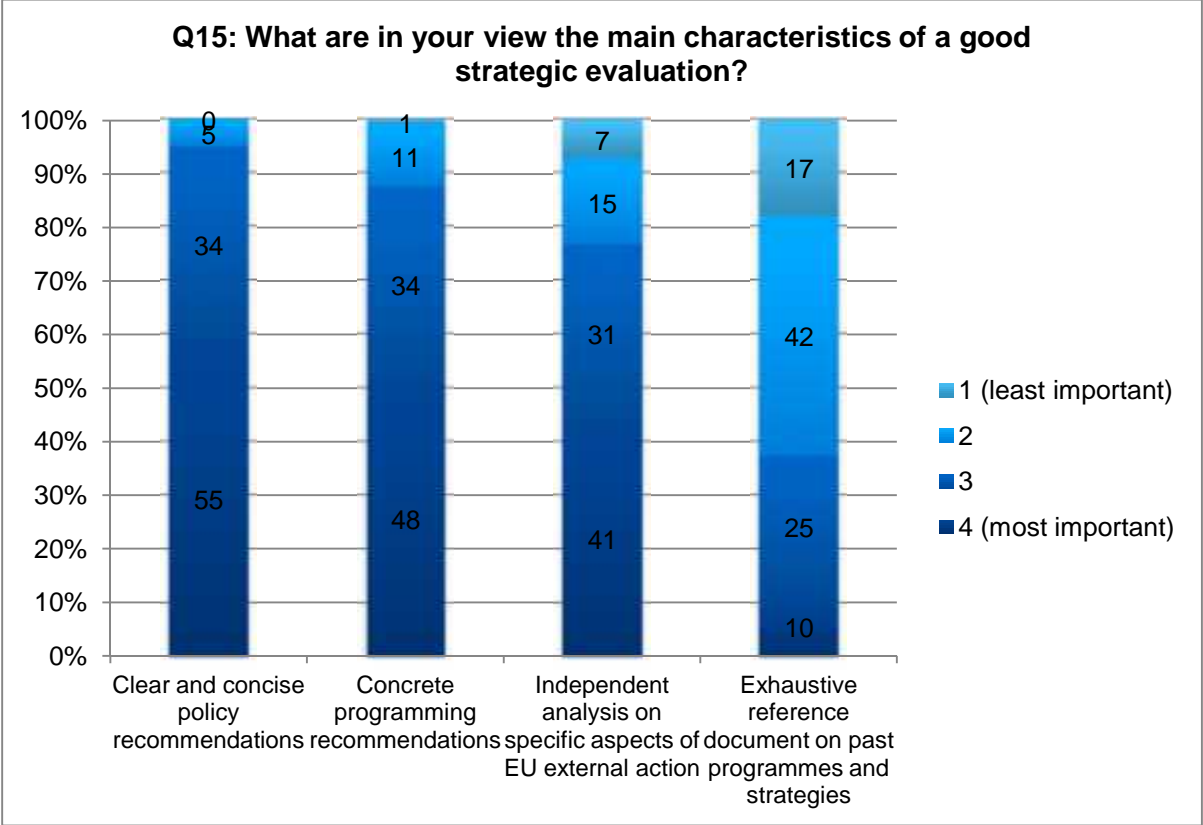
[18] Note that the average experience of respondents in the EU system was 3.8 years.

The survey tried to link Delegation staff's notion of quality to the different uses of evaluations in their work. Question 15 clearly identified that **quality is judged in terms of practicality.** Delegation staff particularly value clarity and concrete recommendations, while an exhaustive review of past programmes and strategies is seen as less important.

Q16: Are evaluation results discussed in your team when a new evaluation becomes available?

Legend:
- Systematically
- Often
- Occasionally
- Rarely
- Never

Values shown: 8%, 20%, 42%, 24%, 6%

Q17: Are evaluation results taken into account?

2007 Survey

Values shown: 17%, 64%, 15%, 4%, 0%

2013 Survey

Values shown: 44%, 39%, 9%, 8%, 0%

Legend:
- Systematically or regularly
- Occasionally
- Rarely
- I don't know
- Never

The survey also looked at the position of strategic evaluations within the Delegations. New evaluations are discussed in the Delegations on an ad hoc basis. Only 8% systematically put new evaluations on the agenda, and in **30% of the Delegations evaluations are rarely or never discussed**. On the other hand, **44% indicate that evaluations are regularly or even systematically taken into account**, which is a step up from the 2007 result (17%).

Q18: What determines productive use of evaluation results in your work?

Category	1 (not at all)	2	3	4	5 (very much)
Quality of the evaluation	0	11	24		55
Form and readability of the evaluation	0	2	19	43	26
Involvement of delegation staff and other potential users in the evaluation process	0	11	17	43	19
Management belief in the relevance and value of the evaluation	2	5	24	38	21
Involvement of stakeholders in the management of the evaluation	1	18	19	36	16
Timeliness of the evaluation	2	10	27	27	24
Delegation staff professionalism and experience with evaluation results	0	11	30	34	15
Institutional arrangements (dissemination seminar, feedback mechanisms and follow-up)	1	20	34	17	18

The productive use of evaluations depends on a number of important factors, including timing, formal dissemination, quality, Delegation staff competencies, etc. **Quality and readability are crucial for the Delegations. The involvement of Delegation staff in the evaluation process is seen as a key determinant for uptake, so is management belief in the relevance of evaluations**. Many respondents bemoan that evaluations are not carried out when needed (for example before programming). This is an indication of how EU Delegations see the use of these documents, but it also reflects the dominant method of working in Delegations. In many cases, the workload pressures Delegations into adopting a reactive course of action led by the programming cycle of EU cooperation.

Q19: What are the main bottlenecks that can limit the uptake of evaluation results, particularly in your line of work?
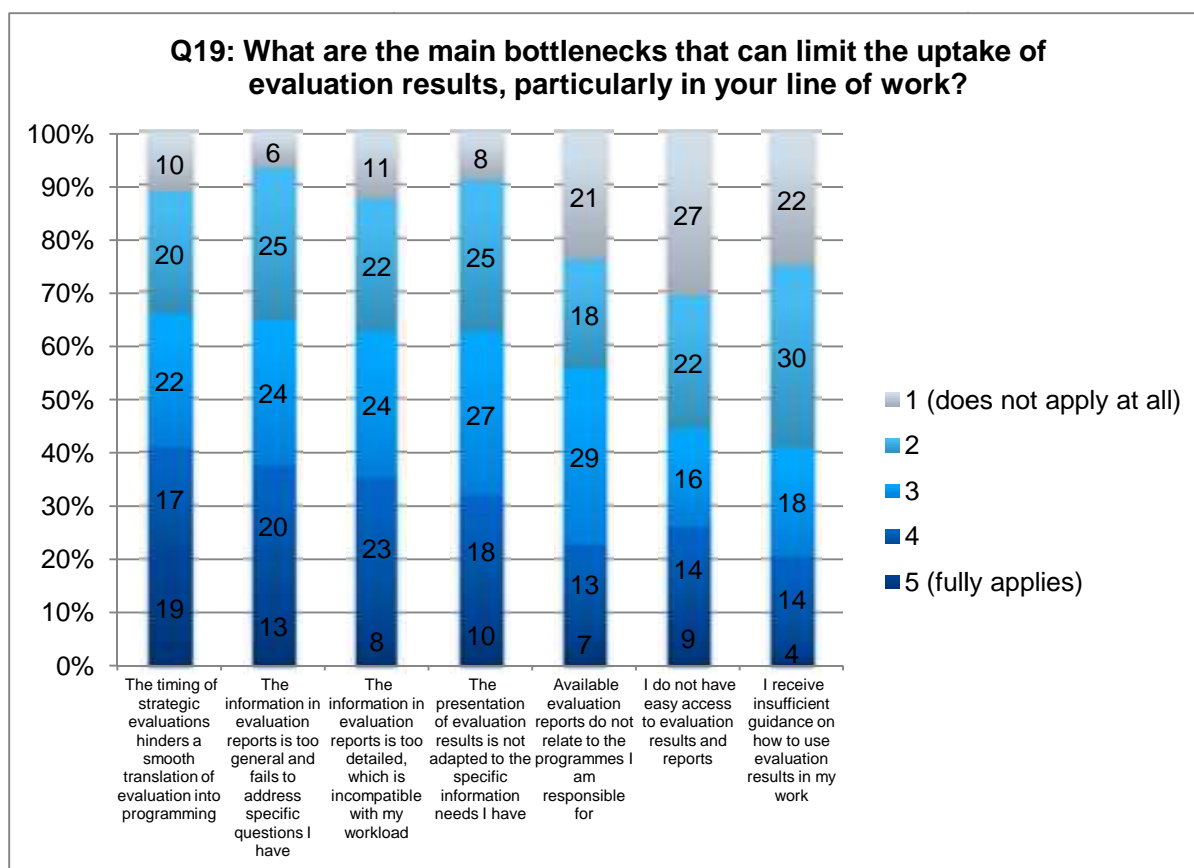
The study team asked Delegation staff to indicate in the extent to which a number of possible bottlenecks for uptake apply to their work. Interestingly, most of these issues seem to apply, and the first five statements in particular:

1. The **timing** of strategic evaluations hinders a smooth translation of evaluation into programming

2. The information in evaluation reports is **too general** and fails to address specific questions I have

3. The information in evaluation reports is **too detailed**, which is incompatible with my workload

4. The **presentation** of evaluation results is not adapted to the specific information needs I have

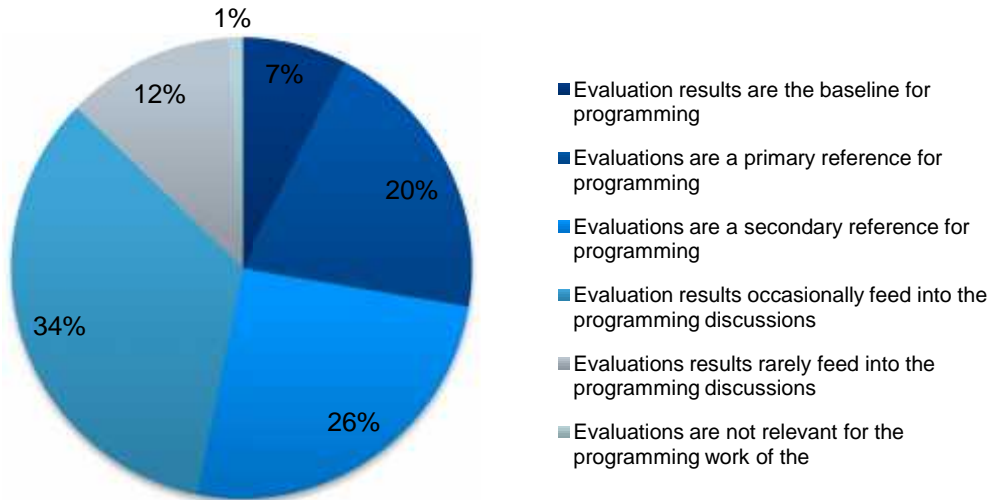5. Available evaluation reports **do not relate to the programmes I am responsible for.**

Evaluations and programming

Staff in both Delegations and HQ see the programming cycle as one of the primary opportunities for uptake of evaluation results. However, many factors curb a smooth translation of evaluation conclusions and recommendations into relevant programming decisions. The study team asked Delegation staff to what extent it uses evaluations in the programming work of the Delegation. **While evaluations are clearly seen as one of the ingredients in the programming discussions, they are only a secondary source of evidence to support decision-making. Evaluation results therefore only occasionally feed into programming decisions at the Delegation level.**

Country level evaluations are seen as the most relevant for programming, followed by sectorial and thematic evaluations. Regional evaluations are deemed less relevant. This may also be linked to the fact that not all Delegations have a specific regional mandate.

Delegation staff use evaluations in the programming phase and during the midterm review. The survey reveals that **evaluations are used selectively and in a very pragmatic way**: conclusions and recommendations and information on the impact of past interventions are the most relevant for programming. Other components are seen as considerably less relevant for the programming work of the Delegations.

Q20: How relevant are evaluations for the programming work of the EU delegation?

Legend:
- Evaluation results are the baseline for programming
- Evaluations are a primary reference for programming
- Evaluations are a secondary reference for programming
- Evaluation results occasionally feed into the programming discussions
- Evaluations results rarely feed into the programming discussions
- Evaluations are not relevant for the programming work of the

Values: 1%, 7%, 20%, 26%, 34%, 12%

Q21: In general, how relevant are the different types of strategic evaluations for the programming work of the EU delegation?

Legend:
- 1 (not relevant at all)
- 2
- 3
- 4
- 5 (very relevant)

	Country level evaluations	Sectorial evaluations (e.g. transport sector, health sector)	Thematic evaluations (e.g. human rights, good governance)	Evaluation of aid modalities (e.g. evaluation of budget support operations)	Regional evaluations (e.g. ECOWAS, ASEAN)
1 (not relevant at all)	5	2	1	7	16
2	6	5	8	12	26
3	14	23	28	22	15
4	28	28	31	24	22
5 (very relevant)	34	29	19	22	8
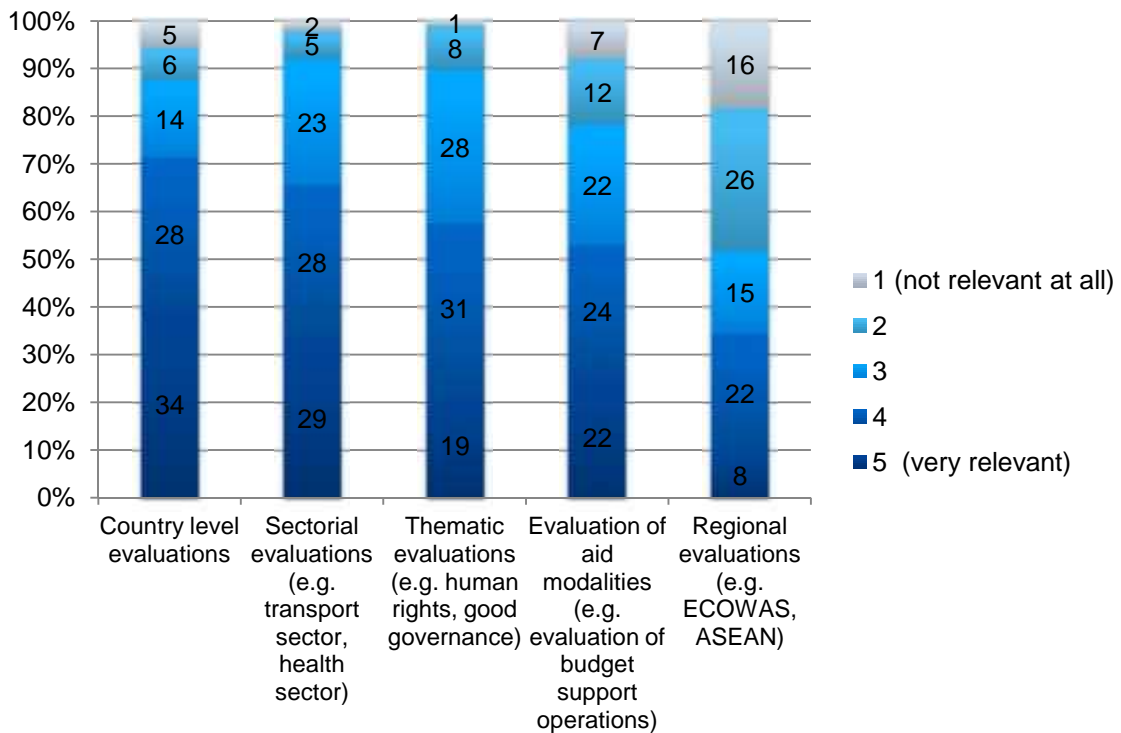
Q22: When are evaluations particularly useful?

- During the programming phase
- During the Mid-term reviews
- When looking for information about the effects and impact of EU cooperation
- Evaluations are not useful

1%
26%
44%
29%
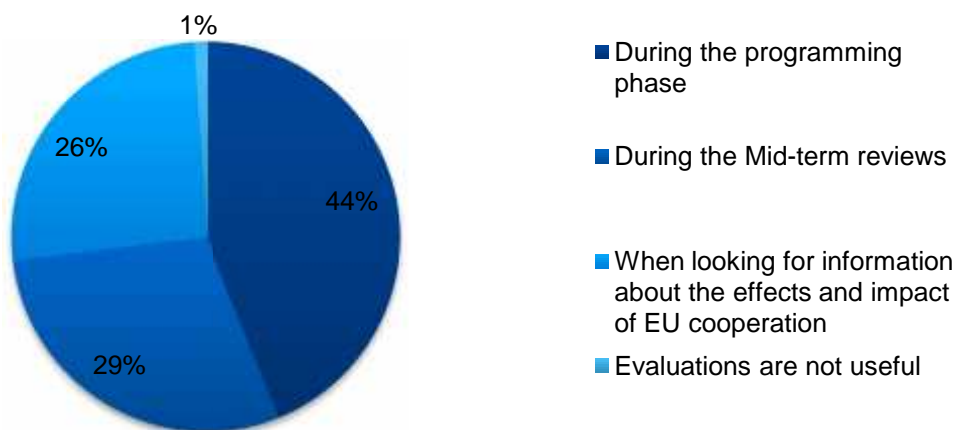
Q23: Which aspects of strategic evaluations are the most relevant for programming?

Legend:
- 1 (lowest priority)
- 2
- 3
- 4
- 5 (highest priority)
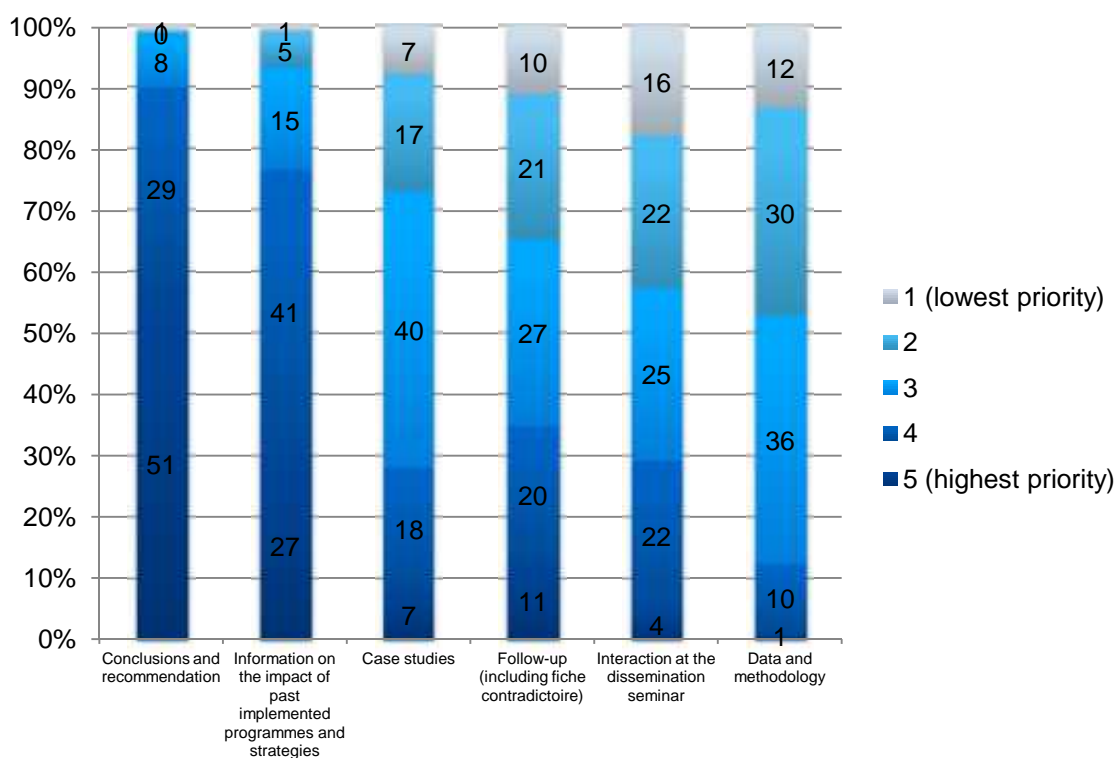
6.4 Improving the uptake of evaluation results

The survey collected 60 individual statements from Delegation staff on how uptake of evaluations can be improved. The question was "what do you think is the top priority for strengthening the uptake and productive use of evaluation material in EU development cooperation?" The responses yielded a wealth of different views on the use of evaluations, but also unveiled an important number of structural dynamics in the knowledge culture of the EU:

Q24: individual statements on how to improve uptake of evaluation results

24.1. Evaluations and programming
Timing and planning
The most important is to conduct evaluations at such a time that their information can be used for the programming phase
Thematic evaluations are essential during the programming phase of EU cooperation, to improve the impact of activities. Thematic evaluations should be done at the end of each EDF to improve programming accuracy of the next one.
Furthermore, at the time being, evaluations are implemented on an ad hoc basis and there is no formal framework which governs them: e.g. at the time of programming or programming review, each Delegation is deciding whether independent strategic evaluations are to be carried out or not, while it would be worth to integrate a compulsory evaluation in the programming cycle, thereby helping to foresee it well in advance. This also implies of course that the programming cycle - especially the determination of the MIP e.g. - is managed in full transparency and planned well in advance.
Decisions should be taken AFTER the evaluation is done; evaluators should be capable of telling us clearly if a project/approach is working, or not.
Timeliness of the strategic evaluations
Evaluation period for country evaluation: avoid long gaps between carrying out the evaluation and the country evaluation period to have up to date conclusion which can be used for new programmes.
Top priority is to have timely evaluations in the Programme Cycle to allow use of lessons learned and recommendations in the next cycle.
Intégrer les évaluations dans le cycle de programmation. Sauf erreur de ma part la programmation actuelle du 11ème FED n'inclut pas d'évaluation de la mise en œuvre du 10ème FED dans son calendrier. Il n'y a pas de point d'arrêt, de point de réflexion.
To ensure that major lessons learned are available before the programming of the next phase of aid, and not after.
Clarity and practicality
Evaluations should be concise, and recommendations should be very concrete and operational
Clear recommendations which can be easily used for programming
Best practices and lessons should be presented in a concise way"
Synthetic conclusions presenting clearly what worked or did not work, with practical recommendations on programming and implementation
The evaluation's conclusions and recommendations must give a clear list of lessons learned and provide a road map, with milestones, for future interventions funded by the EU.
Concise recommendations and advice on methodology
Other

Continuity and predictability of engagement in few key sectors.

In past projects all over and seldom prolonged engagement in same sector. Hence evaluation is interesting but will not be relevant as no continuation.

Mid term reviews could be linked to (major) Riders / reorientations of programmes

ROM missions I find very useful as they give a snapshot follow a standard methodology and are very structured. MTR/ETR often depend on quality of Expert who can be very creative. Hence comparison of evaluations often not given. It also takes much more time to digest information and understand what is being said.

At EUD level, mechanisms for better dissemination and use in programming.

Using evaluation results for new programming and assessment of EU cooperation based of evaluation findings

Ensuring that recommendations made from evaluations are implemented and factored in programming strategies and subsequently EU relevant staff is made aware.

24.2. Strengthening evidence-based decision-making

Political will and management support

The present decision mechanism for programming has not from my point of view integrated lessons learnt from Strategic Evaluations.
Our respectable institution should start a deep structural reform in order to become a Knowledge managed institutions and become hopefully a think tank.
We have the capacity but our programming / management system should be rethought.

Willingness to incorporate 'uncomfortable' findings from evaluations in forward programming (i.e. willingness to depart from standing "good" practice approaches and to abandon blueprints in favour of tailor-made programmes).

That the EU acquires sufficient political courage to act on the findings of professionally conducted independent evaluations

Management belief in the relevance and value of the evaluation, sufficient time and resources (Human Resources if internalised, Financial if outsourced) for programming, identification and formulation.

Political will of hierarchy.

The flexibility of the Management on giving programme managers ample space to have more say on the evaluation results

Support from the management. Management must learn to accept critics and learn from mistakes. It would be so motivating to have dedicated discussions at Delegation level on the results of strategic evaluation and how they should inform our programming and implementation.

Concise reports shared with geographical Units followed by formal instructions of the geodirectors

Other

Strengthening evaluation capacity in Delegations. Clearer expectations from the centre about how HoDs will be held responsible for the extent that they make evidence-based decisions.

- Developing knowledge management system, to make evaluation results available and used during programming and implementation
- Better sharing of evaluation reports internally and to stakeholders
- Develop strategy to follow up recommendations
- Develop repository of lessons learnt
- Introduce in the QSG checklist "if lessons on previous evaluation have been taken into consideration or not"

Consistent use of evaluations - their analysis and recommendations - on all levels.

24.3. Increase ownership of evaluations

It is clear that evaluations do not really make the Commission change its course. There are several reasons for this. First of all there is too little involvement of the people in Delegations and they have become too much of a theoretical exercise. Just the programme of the EDF11 shows that we do not learn any lessons and we have gone backwards on the 10th, which was a step backwards on the 9th. Good intentions are always violated for political or personal use.

Stronger involvement of DEL staff in the different phases of the evaluation process; improved quality of evaluation teams to result in higher quality reports (skills set of experts do not necessarily meet the requirements of the evaluation tasks); closer attention to follow up of evaluation recommendations during programming and mid-term review processes.

Involve Delegations in the evaluation design phase (results - outcomes definition)

Increased capacity of Delegation staff in the evaluation process and the uptake use of evaluation results. For evaluation exercises organised by the HQ where the Delegation is involved in the process (e.g. facilitating in country mission, involvement in the country mission briefing/debriefing meetings, etc.) it will be good to have the final report systematically shared to the Delegation for our own inputs and reactions (if any, and if these follow ups are clearly specified in the report).

The evaluation process should be considered not as a simple exercise which has to be done for instance annually - ROM. But as a tool to help us improve in the management of our projects (implementation and follow-up). Delegation staff should be present in the evaluation missions in other to have a better feedback and to be able to participate in the discussion with the different actors.

Close cooperation and coordination amongst HQ, Delegation involved and evaluators.

Close consultation with Programme Staff.

Undertake an open and inclusive evaluation of past programmes involving feedback from all stakeholders concerned (EU, government, NSAs and private sector). Make sure the results are shared in a public discussion workshop, where preliminary recommendations are presented before the evaluation report is finalised.

Effective management of these evaluations - with a active participation of key stakeholders, sensitisation/training of Del staff on value of evaluations, change of mind-set towards regular monitoring which will feed data/evidence into evaluations

I also find that in general we all use evaluations as a magic wand: we take them into account ONLY if they play in our team... for instance never really saw EuropeAid HQ happily taking up evaluations carried out at country level, but sure we have to study by heart what comes down from HQ ;)

24.4. Improve the evaluation process

Lighten the load

Secondly we have already an information overload. Evaluation results especially outside of direct connection to the programme are just additional volumes to read for which we simply do not have the luxury in staff of time resources.

Making it a more quick and less cumbersome process

Evaluations need to be less rigid in their methodology – standard ToRs for evaluations as prepared by the EU do not provide for differing focus in evaluation nor do they provide for the weight to be given to policy recommendations and programming recommendations – these are often general and lack analysis or depth.

Evaluation is a very time consuming activity so time is a key factor.   And of course enough human resources
Country evaluations: I have been through 2... Useless and draining.
Simplify the methodology and reduce the time needed for the evaluation (evaluators were using to "sophisticated" methodologies, too time consuming, and a long process which lasted over one year, examine the usefulness and added value of joint evaluation (added value of evaluation with of cooperation's with different periods)
Simplify the methodology and improve clarity
Review the methodology
The evaluation methodology should include different message formats for different uses.
Best practices and lessons should be presented in a concise way
(From the experience of a joint country evaluation) Improve the "Readability of the report": use concise style; include concrete recommendations and conclusion (not standard parts which could be applicable to any sector or country).
Clarity, conciseness, methodologies
Improve the quality of evaluations
Improve de knowledge of the evaluators
To improve the quality of experts under the Framework Contracts or to introduce greater flexibility for recruiting suitable experts by the Delegation.
Recruiting qualified experts producing high-level reports, in particular with regional or thematic evaluation as these implies very high-level capacity of analysis. Some evaluation reports do not go beyond the Delegation knowledge as they summarise the information the consultants have been briefed with.
Ensuring that the external evaluators are competent
Most evaluations are done by institutions which depend from further contracts from the EU. So, their evaluation is never independent or impartial. A kind of peer-review mechanism for evaluation methodology should be introduced
Quality of the evaluators
Increase the direct relevance of evaluations
Develop a methodology for evaluating BS operations. In the absence of this methodology, our Delegation has not had any evaluations in the recent past. Evaluations can also be used as a tool to strengthen or address difficult areas of policy dialogue. The absence of this tool deprives the Delegations from an outside expertise to assess and redirect our programmes.
Maintain the scope of evaluation under control
"Further to this can we please stop to have ""recommendations"" I find this totally out of the scope of evaluators (who should analyse and give the means for the organisation to decide and act).
Ensure a real follow-up
To establish a methodology for the follow up of the evaluation on the most relevant points to be considered after

the "Fiche Contradictoire"
Training of EU Delegation personnel to the follow-up and use of evaluations is crucial.
Promote the evaluation results discussion
Thoroughly ensuring follow up on recommendations (i.e. at QSGs on PIFs and AFs))
Training of EU Delegation personnel to the follow-up and use of evaluations is crucial.
Improve access to information
Share the evaluations data base
Access to research tools on EuropeAid evaluations per theme, countries, etc. to feed programming/formulation.
Awareness is important for users
Accessibility / information about the evaluations
Better communication on lessons learnt
Dissemination of the evaluation material at all levels.

24.5. Other

For example, start doing evaluation in Yemen. No evaluation has been carried out here since 2009.
The feedback provided by evaluations can be used to improve the quality of our development actions in terms of accountability and learning
Further to this can we please stop to have "recommendations" I find this totally out of the scope of evaluators (who should analyse and give the means for the organisation to decide and act).
The more the government is involved, the more the language gets diluted and the less useful the evaluation becomes.

6.5 Conclusions

Survey sample
The survey had a very high response rate in a short period of time, which suggests that Delegation staff are interested in the subject and see the value of using evaluations in their work. The survey sample has a representative geographical coverage and covers a wide range of staff profiles which work in EU Delegations and was therefore able to reliably capture the EU Delegations' experiences with the current system for strategic evaluations. Many of the questions were based on a similar survey for EU staff undertaken in 2007 (both HQ and Delegations). The responses show remarkable continuity in how staff assesses their use of evaluations.

Knowledge and learning culture in EU Delegations
There are clear indications that the use and transfer of knowledge and evidence in Delegation decision-making can be improved. There are several institutional barriers which limit the uptake of evidence from strategic evaluations. These include human resource and administrative constraints, but also a risk adverse management culture which does not systematically prioritise the use of evaluation evidence.

Use of evaluation results in EU Delegations

The majority of Delegation staff see a clear added value in evaluations as a complementary source of evidence in their work. Most of the respondents are familiar with evaluations and have been involved in at least one aspect of the evaluation process. Due to the high number of operational staff in Delegations, they judge quality in terms of clarity and practicality rather than methodological rigor, and particularly value clear and concise conclusions and recommendations. Delegation staff use evaluations in a selective manner for a variety of activities. The most important uses which were identified are programming, risk identification, policy dialogue and ensuring transparency and accountability.

The responses emphasised the importance of synchronising evaluations with the programming exercise[19]. The disconnect between the programming and evaluation cycles is seen as a serious problem and the primary bottleneck for uptake of evaluation results.

A comparison between the 2007 and 2013 survey results unveiled that EUD staff feel that the quality of evaluations has stagnated in the past six years. The past few years saw the consolidation of the 2006 evaluation methodology and the emergence of a rigorous evaluation system. The survey also indicated that heavy evaluation reports are not always relevant for the work of the Delegations, and that their knowledge needs are often much more specific and short-term. The survey responses identified a number of areas with room for improvement:

1. **Evaluations and programming:** Many Delegation staff members see the programming cycle as the main time to use evaluations in their work. Two aspects are particularly important in this regard: timeliness and practicality. Delegation staff lament the **lack of synchronisation** between the country programming cycle and the evaluation work programme, but also the timing gaps between evaluations themselves and the availability of results. In order for evaluations to feed into programming, they want **clear and concise conclusions**, with **concrete and practical recommendations** for planning future programming.

2. **Strengthening evidence-based decision-making:** in line with the survey results for question 7, Delegation staff feel that the use of evidence in EU decision-making is sub-optimal. Respondents see room for improvement in a more consistent use of lessons learned at different levels. This requires a strong signal from management in the Delegation and at HQ level and a higher degree of "political courage" to act on possibly inconvenient findings in evaluations.

3. **Increasing Delegation participation (and ownership):** Delegation staff see evaluations as a practical tool to assist them in their work, but criticise the **lack of involvement of local staff, both in the definition and in the evaluation process itself.** While there is some involvement (mainly of the cooperation section) in country evaluations, most of the Delegation staff is presented with the *fait accompli* of a country evaluation when it takes place. Given the practical use that Delegations foresee for these evaluations, increased participation at all stages of the process could benefit uptake a great deal.

4. **Streamlining the evaluation process:** respondents identified a strong need to improve the evaluation process itself:
   - Delegation staff experience evaluations as a time consuming and cumbersome process which drains resources.
   - The methodology does not contribute to the readability and practicality of evaluations, nor does it ensure the relevance of evaluations in a given country context.
   - The quality of the evaluators under the framework contract is not always ensured
   - Beyond the standard *Fiche Contradictoire*, the follow-up of evaluation results is often incomplete. An adequate system to ensure real follow-up is not in place.

---

[19] At the time of the survey, the EU was preparing the 2014-2020 MFF. It is possible that this influenced the results, mainly the individual statements that were recorded.

7. KNOWLEDGE TRANSLATION AND BROKERING IN PRACTICE

The literature on knowledge translation and brokering outlines the key functions which need to be performed for a more active approach towards uptake and its effectiveness. These are indicated in Figure 3 in the main report (and below) – how they are activated will vary between different organisations depending on their structure, resourcing and wider practices around evaluations. As noted in the main report, form follows function: the different functions need to be clarified before individual and team responsibilities can be elaborated.

Information intermediary	Knowledge translation	Knowledge brokering	Innovation brokering
Enabling access to evidence from evaluations	Helping people make sense of and apply evidence from evaluations	Improving the use of evaluation evidence by being involved in decision-making	Ensuring the organisational incentives are in place to support knowledge translation and brokering

Suggestions for how these roles can be elaborated are in the following table: these are supported by an analysis of the different communication channels which are currently available for disseminating the evidence from strategic evaluations and the extent to which they were found to be used in the 26 evaluations considered in detail in this study. As can be seen from the table, it does not appear that the full range of possible channels for dissemination and ensuring uptake is used extensively (NB: the information for this table was particularly difficult to find, if the relevant evaluation manager was not available).

Table 1: Suggestions for ensuring key knowledge functions are fulfilled within the Evaluation Unit

	Information intermediary	Knowledge translation	Knowledge brokering	Innovation brokering
Key functions	Enabling access to evidence from strategic evaluations	Producing short newsletters or digests, seminars or presentations (for dissemination)	Being actively involved in decisions around policy and programming to ensure evaluation evidence is considered by other Units. To be effective this needs to be an active rather than a passive process	Setting incentives to ensure that knowledge translation and brokering can happen effectively, looking for innovations and supporting their replication
What already exists in EuropeAid?	• All evaluations and supporting documents are posted online • Capacity4Dev hosts pages with information • EuropeAid intranet hosts '60 words' emerging from each evaluation	• A single dissemination seminar (occasionally more than one) • Info Point presentations • Pages can be written for Capacity4Dev	• Groups can be formed on Capacity4Dev • Ad-hoc brokering takes place by individual evaluation managers • Otherwise little evidence of active involvement in brokering has been seen • Thematic Units are key knowledge brokers, but their capacity has declined since the restructure	• Reorganisation of EuropeAid to enable the Evaluation Unit to have stronger linkages to senior management?
What do other agencies do?	• Publish evaluations online • All agencies have repositories of information similar to Capacity4Dev •	• Wide divergence between agencies on how this is resourced, mandated, staffed and implemented • DFID specifies that a reporting plan needs to be produced for each evaluation • DFID publishes an evaluation digest • Danida communicates its own evaluation evidence as well as that of other agencies • France/DGMDP: summaries of the most significant evaluations are produced and widely distributed • SIDA holds seminars, publishes newsletters • WB IEG designs sophisticated communication strategies	• Evaluation Units are widely responsible for ensuring evaluation findings are brought up in policy and programming discussions • GIZ management are asked what evaluation questions they want answered • GIZ brings in a facilitator to chair a learning café for internal managers • DFID runs an internal community of practice for evaluations, and evaluation cafes • GIZ holds annual meetings for sector/thematic networks where	• In GIZ utility is prioritised ahead of independence, although evaluations are robust • In agencies such as GIZ, DFID and the World Bank there is flexibility around the methodology to ensure evaluations are context specific and can be considered at key decision points • DFID senior management are committed to encouraging a learning culture across the organisation • Evaluation has been devolved to the programme level, enabling it to be built in to programming rather

	Information intermediary	Knowledge translation	Knowledge brokering	Innovation brokering
			relevant findings might be presented	than bolted on at the end • DFID mandates projects to produce short knowledge translation resources such as policy briefs • DFID resources knowledge brokers in its Research & Evidence Division • Evaluations which are decentralised may offer greater opportunities to broker knowledge to local stakeholders • DFID see commissioning, managing and using evaluation as part of the core skills set of all their staff. They have also brought in evaluation specialists and provide accreditation for those who want to be one. • In GIZ, resources are available for communication and brokering within the evaluation budget
Possible new functions within EuropeAid	• Capacity4Dev is a key tool: agree the role and purpose of different knowledge translation activities, how to store them and how to ensure that they continue to be actively 'sold'	• Evaluation managers & Reference Groups could use the table overleaf to produce a short reporting plan for each evaluation • Production of a regular evaluations' digest, with lessons from both EuropeAid and other agencies' evaluations (4-6 pages, six-monthly?) • Ad-hoc synthesis reports of key lessons from evaluations on a thematic basis (one per year?)	• Schedule evaluations such that the findings and specific evidence they generate can effectively feed into decision processes when needed • Set aside space for reflections on lesson learning within the Evaluation Unit • Construct formal spaces for lesson learning from evaluations around programming cycles (consider themes first; also consider hosting more than one space per theme) • Identify key findings from each	• Ensure that the non-use of robust evaluation evidence is seen as irresponsible (may need engagement between Heads of Unit) • Enable staff from Evaluation Unit to act as effective brokers by participating fully in key policy and programming discussions (this will need incentives from senior management).

	Information intermediary	Knowledge translation	Knowledge brokering	Innovation brokering
			evaluation (or any syntheses done) which need to be actively brokered, the key debates, audiences and timetables • Allocate responsibility to individuals within the Unit to do this brokering process	

Table 2: Assessment of the use of different communication channels

No.	Name of the evaluation	Fiche Contradictoire	In-country seminar	Brussels seminar	Country-level action plan	Video or audio interviews or presentations	Note to the network of EUD evaluation correspondents	Capacity4Dev: formed a group	Capacity4Dev: wrote a page	Capacity4Dev: asked journalists to write something	60 words (for EuropeAid intranet)	Info point presentation	Note to Parliament	Note to Court of Auditors	Note to MS Heads of Evaluation Units	Note to other stakeholders	Other dissemination
	Country evaluations																
1	China - 1077	X	X								X	X					
2	CAR - 1263	X	X								X	X					
3	Vietnam - 1269	X	X								X	X					
4	Nicaragua - 1271	X	X								X	X					
5	Nigeria - 1276	X	X								X	X					
6	Egypt - 1281	X	X								X	X					
7	Ukraine - 1283	X	X								X	X					
8	Malawi - 1297	X	X								X	X					
	Regional evaluations																
9	Central Africa - 1037	X															
10	Central America - 1092	X															
11	ASEAN - 1262	X															
	Sectoral evaluations																
12	Water & sanitation - 965	X		X							X						
13	Energy - 1192	X									X						
14	Education - 1282	X		X							X						
15	Transport - 1655	X		X							X						
	Thematic evaluations																
15	Good governance - 884	X		X							X						
16	Conflict prevention & peace building - 1277	X		X							X						

#	Evaluation																		Notes
17	Human rights - 1298	X		X							X								
	Aid modalities																		
18	CSO channel - 1259	X		X							X								
19	Investment Facility - 1285			X							X								(1) Dissemination seminar in Luxembourg
20	Budget support - Tunisia - 1286		?								X								
21	Budget support - Mali - 1290		X								X								
22	Budget support - Zambia		?								X								
23	Budget support - Tanzania		X	X	X	X				X									(communications still being developed)
	Synthesis reports																		
24	Budget support - Mali, Tunisia, Zambia																		(1) Presentation to Development Commissioner & Parliament (2) OECD-DAC Evaluation Insights note
25	Geographic evaluations - 1256																		

Note: 'country-level action plan' refers to the model being developed in Tanzania

8. SUGGESTED REVISED FICHE CONTRADICTOIRE

#	Activity	Responsibility	Timeframe	Follow-up one year later
1	*Reconsider sector priorities and overall funding profile for issue X…*			
	Detailed plan of actions to be taken to fulfil R1.1	Outline the responsibility for leading on each action or set of actions, and the associated budget (if necessary)	Month / Year	Record, with reasons, which actions have been achieved, which were changed and which have been dropped
	Comment from senior management Senior management should comment on the action plan and budget, noting any linkages with other Delegations, themes or sectors			**Comment from SM** Senior management comment on overall follow-up and any further activities
2	*With DPs, consider revising targeting and allocations to subsectors…*			
	Detailed plan of actions to be taken to fulfil R2	Outline the responsibility for leading on each action or set of actions, and the associated budget (if necessary)	Month / Year	Record, with reasons, which actions have been achieved, which were changed and which have been dropped
	Comment from senior management Senior management should comment on the action plan and budget, noting any linkages with other Delegations, themes or sectors			**Comment from SM** Senior management comment on overall follow-up and any further activities
3	*Link to ROM (country evaluations only)*			
	Explanation of what the results from the country evaluation means in terms of what needs to be monitored in future	Outline responsibility for taking any necessary changes to ROM forwards	Month / Year	Report on adaptations made with reasons
	Comment from senior management Comment on the strategic implications for evidence collection			**Comment from SM** As left

9. LIST OF INTERVIEWS

The study team interviewed over 50 individual stakeholders and held several focus group discussions. In order to maintain confidentiality the names and functions of the people in question are not listed, instead an indicative list of the type of actors is given.

DG EuropeAid
DG Senior Management
Senior and mid-level staff
Thematic, geographic and corporate units.

EEAS
Senior and mid-level staff; current and former reference group members
Thematic and geographic divisions.

Other EU and related
Staff working in EU Delegations
Former senior staff from the EC
Evaluation professionals; team leaders and consultants
Technical experts working on EU-funded programmes.

Non-EU agencies
Staff from the following agencies who work in either (1) independent units mandated to evaluate the work of a specific agency and/or (2) units within agencies which focus on evaluation and its uptake (including communication and learning): DFID, USAID, GIZ, IFAD, World Bank, SIDA, DPME (South African Presidency).