

***DAC Evaluation Network:
Follow-up to the Paris Declaration on Aid
Effectiveness***

An options paper

Revised after discussion at the fourth meeting of the
DAC Evaluation Network, 30-31 March 2006

by

David Booth
*Overseas Development Institute,
London*

Alison Evans
*Institute of Development Studies,
University of Sussex*

**Revised Draft
May 2006**

Please also refer to accompanying Review of Literature

<p>Please send comments to: d.booth@odi.org.uk; a.evans@ids.ac.uk</p>

Contents

<i>List of boxes, figures and tables</i>	iv
<i>Executive summary</i>	v
1 Introduction	1
2 The challenge	3
2.1 The nature of the Paris commitments	3
2.2 From monitoring to evaluation	4
2.3 Outline of an evaluation framework	7
2.4 What needs to be evaluated?	11
2.5 Can it be done? Data and methods	15
3 Design of the evaluation work: Some options	18
3.1 Introduction	18
3.2 A matrix of options	19
3.3 Towards an optimal solution	22
3.4 Advantages of the suggested approach	26
4 Timetable, decisions and next steps	27
4.1 Summing up	27
4.2 Timetable	27
4.3 Managing the partnership	30
4.4 Options and next steps	30
<i>References</i>	33
Annex 1: A matrix of options for evaluation design	34
Annex 2: Persons consulted	36
Annex 3: Terms of reference for the Options Paper	38
Annex 4: DAC peer reviews contribution to monitoring the Paris Declaration <i>[attached file]</i>	

List of boxes, figures and tables

Box 1.1	The process so far	2
Figure 2.1	The Paris Declaration concepts	3
Figure 2.2	A possible evaluation framework for the Paris commitments	9
Figure 4.1	Timeline of proposed activities	28
Table 3.1	Scoring of proposed activities in terms of trade-off variables	26

The DAC Network on Development Evaluation

This Options Paper was prepared for discussion at the fourth meeting of the DAC Network on Development Evaluation, 30-31 March 2006, and to facilitate consultations with partners. It has been revised in the light of that discussion and further comments received. The Evaluation Network is a subsidiary body of the Development Assistance Committee (DAC). Its purpose is to increase the effectiveness of international development programmes by supporting robust, informed and independent evaluation. The Network is a unique body, bringing together 30 bilateral donors and multilateral development agencies: Australia, Austria, Belgium, Canada, Denmark, European Commission, Finland, France, Germany, Greece, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank, European Bank for Reconstruction and Development, UNDP, and the IMF. For further information on the work of the DAC Evaluation Network, please visit the website www.oecd.org/dac/evaluationnetwork or email dacevaluation.contact@oecd.org. The views in this draft options paper do not necessarily represent the views of the Evaluation Network.

Executive summary

This options paper was commissioned by the DAC Network on Development Evaluation to assist it in selecting an appropriate way of following up on the March 2005 Paris Declaration on Aid Effectiveness. The paper makes a qualified positive assessment of the “evaluability” of the Paris Declaration commitments, and considers various options for the design of a programme of evaluation work. The paper identifies a number of different options for action by the Network. It suggests that, given various constraints and trade-offs, an optimal approach would be a set of four connected evaluation activities, each capitalising upon a distinct type of strength.

The argument is developed in three main sections as follows.

The challenge

The Paris Declaration poses an outstandingly important challenge to the field of development evaluation. As well as clarifying concepts, the Declaration expresses a broad consensus how country policy ownership, aid alignment and harmonisation, and mutual accountability can work together, along with the strengthening of results-oriented management practices, to make aid more effective. The new consensus is not lacking in tensions. However, this in itself does not make the agreed commitments unsuitable as a focus for evaluation.

Plans for monitoring the Paris Declaration commitments are already quite advanced. However, evaluation is in principle different from monitoring. It is interested in “how” and “why” as well as “what” questions, and has to concern itself with underlying theories of change.

It is possible to visualise an evaluation framework (that is, a formalised results chain and corresponding evaluation questions) for the Paris Declaration commitments. An outline framework is provided to illustrate this point. It presents the 5 categories of Paris Declaration commitments as Outputs, with donor and country “political support, peer pressure and coordinated action” as the corresponding inputs.

The theory articulated by the rest of the framework says:

- ❑ that country ownership and the other outputs promised by the Paris Declaration would, if implemented together, strengthen country capacity to make and implement policies focused on development results and make good use of aid;
- ❑ that country capacity enhanced in this way would raise the quality of public investment and service provision, including regulation and institutional development for private investment; and
- ❑ that this would lead to better development results, such as growth and transformation, and the MDGs.

This “programme theory” may and may not be widely shared by the Paris Declaration signatories, and it is quite possibly at least partially wrong. However, a degree of uncertainty or disagreement about details of how policy or programme objectives have been intended to be achieved is not an unusual state of affairs at the beginning of an evaluation process.

Questions that could be answered by evaluation work based on the Paris Declaration framework include: 1) what is the range of country circumstances in which the specified linkages are expected to work? and 2) under what conditions do factors, either internal or external to the programme logic, result in particular benefits not being realised, or being realised in unexpected ways? For at least the next few years, the main focus of evaluation effort would need to be on the first two levels, Inputs and Outputs, including the interactions between the different Outputs (that is between the different sorts of behavioural changes that are the subject of commitments in the Paris Declaration).

The linkages further down the results chain would be unlikely to repay significant evaluation effort. This does not imply either that the approach is unconcerned with establishing a link to development results, or that the argument for more sophisticated impact evaluation is rejected. There is more than one knowledge gap about “what works” in development, and several complementary contributions are needed.

Evaluation questions about the Paris Declaration commitments should include both “are we doing things right?” and “are we doing the right things?”. This would result in valuable learning on how things should be done differently, and what else should be done. Questions about cost-effectiveness are relevant but may not be evaluable at this point. The function of evaluation work should be expected to be formative, not summative, at least until 2008 and probably for some time after that. This means that the main purpose is to feed practical learning back into the process of implementation.

The data needs of a Paris Declaration evaluation are potentially enormous. However, the types of questions that can be sensibly asked at this stage are limited, while the number of exercises already planned or underway that will generate relevant evidence is large. So this seems a manageable problem. Anyway, it would not be wise to let the whole evaluation effort rest upon a single information source, such as a survey or a set of commissioned country evaluations. In analysing data, the usual methodological snags will arise, but these are in some respects less difficult than they were for the recent GBS evaluation.

Options for evaluation design

There was a high degree of consensus among those consulted about this paper (see list at Annex 2) that preparation of a common evaluation framework is a desirable first step, likely to provide significant benefits on its own. Other evaluation design issues are less easily handled, including the scope of the work (units of evaluation, thematic scope); ways of ensuring quality and consistency; and suitable governance and management arrangements. On these choices,

there are trade-offs – between stakeholder buy-in and the “criticality” of evaluation content; between country-level learning and cross-country or global learning; and between keeping costs low and adding value in terms of knowledge and evaluation capacity development.

The various options can be set out as a matrix. This shows that it is impossible to think of a way of approaching the Paris Declaration evaluation that will satisfy all concerns to the maximum. The best that is possible is an *optimal* solution, one that scores reasonably highly on all important variables.

The following are the identified options for action by the DAC Evaluation Network:

1. Evaluation Network takes no action and plays no role at this stage;
2. Evaluation Network develops a common evaluation framework only;
3. Evaluation Network develops a common framework *and* takes a role in facilitating country-level evaluation work;
4. Evaluation Network develops a common framework and takes a role in facilitating country-level evaluation work *and* donor-agency case studies;
5. Evaluation Network only takes a role in synthesising available/ongoing evaluations (possibly in combination with option 2);
6. Evaluation Network adopts the suggested four-track approach as a whole (see below).

The paper suggests as the optimal approach working on a number of parallel tracks simultaneously. Four activities or sets of activities are outlined. These are visualised as feeding into each other but as not requiring a heavy apparatus of coordination, so as to maximise particular benefits and keep costs down.

The four activities are:

- a. **The development of a common framework** articulating the logic of the Paris commitments, including the implied linkages between aid effectiveness and development effectiveness.
- b. A series of **country-led country evaluation initiatives**, building directly on the joint monitoring activities being undertaken and other knowledge already available.
- c. A set of **donor case studies** which would look at the way in which the Paris Declaration is finding expression **across a sample of donor organisations** as opposed to countries.
- d. **A medium to long term programme of analytical work** designed to draw together and critically evaluate findings from a variety of sources with a bearing upon the common framework constructed by Activity (a). This component would need to be integrated with the Paris Declaration Medium Term Monitoring Plan.

Activities a) and d) would be designed to take advantage of the strengths of a top-down, but agile and consultative, way of working. Activities b) and c) would draw strength from being very stakeholder-led, bottom-up and responsive to local (country or agency) learning needs. Activity a) would contribute importantly to the other proposed activities and would need to be undertaken soon (in 2006) for this reason. Activities b) and c) are seen as necessarily based on a self-selection principle, with substantial room allowed for “localisation” of the common framework. The proposed Activity d) is not only intended to take advantage of the range of information sources likely to be available at key reporting points, but is seen as having some features of a meta-evaluation.

Timetable, decisions and next steps

A possible timeline up to 2010 is considered on the assumption that the full four-track approach is taken forward. This makes apparent the following:

- ❑ Work on the framework would need to be set in motion soon after the middle of 2006, with a view to discussion of an initial report at the November meeting of the DAC Evaluation Network.
- ❑ Adequate time would need to be allowed for the preparation of the country work, so that the concept of a country-led process is widely and well understood.
- ❑ The scope and nature of the proposed synthesis/meta-evaluation work and ways of integrating it with the Paris Declaration Medium Term Monitoring Plan (content of this plan is still to be decided) would need to be discussed with the Joint Venture on Monitoring the Paris Declaration (JV) at an early stage.
- ❑ The proposed country and donor evaluations should be raised in the framework of the JV and its parent body the Working Party on Aid Effectiveness (WP-EFF) at the earliest opportunity.
- ❑ Slippages in scheduling of activities and delivery of reports typically cause problems for tightly integrated global joint evaluations (e.g. GBS, CDF). The proposals made here depend on a less formal type of integration; however, problems could arise if the points above were not taken into account.

Depending on the decisions taken, the next steps might include:

- agreeing to one of the above options;
- establishing an Evaluation Network task team (with partner countries) with a remit to take this forward, in partnership with the JV and/or WP-EFF;
- organisation of consultations to agree the common framework for evaluation of the Paris Declaration commitments;
- developing a strategy and methodology for country and donor evaluations to be delivered in 2007.

1 Introduction

In March 2005, over 100 DAC members and partners agreed the *Paris Declaration on Aid Effectiveness*, thereby committing themselves to some 56 actions in 5 areas: ownership, alignment, harmonisation, management for results and mutual accountability. ***The Declaration was backed by a commitment to periodically assess mutual progress at country level in implementing the agreed commitments.*** Partner countries were to be in the lead and use would be made of appropriate country level mechanisms. The following elements were visualised:

- immediate steps to coordinate international monitoring of the 12 Indicators of Progress included in the Declaration;
- medium term monitoring arrangements to be agreed by the partnership of donors and partner countries hosted by the DAC, with widened country participation;
- use of DAC Peer Review and regional review mechanisms to support progress on the Paris agenda; and
- independent cross-country monitoring and evaluation processes to be explored “to provide a more comprehensive understanding of how increased aid effectiveness contributes to meeting development objectives” (Paris Declaration, p. 3).

This paper is concerned with the final point – the commitment to explore independent evaluation processes – taking into account the Declaration’s proviso that this should be undertaken without imposing additional burdens on partners. The paper was commissioned by the DAC Evaluation Network following a number of previous steps and the articulation of various additional concerns, as outlined in Box 1.1.

The paper is intended to assist the Evaluation Network and its partners to consider viable options for delivering upon the Paris Declaration commitment in respect of evaluation. It draws on an accompanying review of relevant methodological experiences and substantive lessons from the literature, and upon consultations with over 40 experts and stakeholders, mainly from the Evaluation Network, the JV and the WP-EFF, undertaken by means of face to face meetings or telephone conversations since January 2006 (see list at Annex 2). The paper

- ❑ considers how the intentions behind the Paris Declaration commitments might be conceptualised in results-chain terms;
- ❑ explores what a programme of evaluation work might add to the emerging monitoring framework of the Paris Declaration, including the indicator-based survey;
- ❑ assesses whether the Paris Declaration commitments are susceptible to evaluation, including the questions to be answered and the data and methods that could be drawn upon; and

- considers options and develops suggestions on the design and management of a programme of evaluation work meeting the various concerns that have been expressed by key stakeholders.

The paper does not provide a worked-out evaluation framework, but contains proposals on when and how such a framework might be agreed. It offers a qualified positive assessment of the “evaluability” of the commitments. It recommends an approach which seems to be optimal taking into account the various opportunities and constraints, and sets out options for the Evaluation Network in deciding whether and how to take this work forward. However, it would be hard to overstate the importance of undertaking further consultations on various aspects with all of the relevant parties. ***In all of these respects, the options paper is a point of departure and not a blueprint.***

Box 1.1: The process so far

In **June 2005**, the Evaluation Network held a preliminary discussion on how to maximise its comparative advantages in responding to the call for an evaluation process. A number of members indicated support for the development of a common evaluation framework. Members also noted the importance of partner country participation in such a process and of close coordination with the Working Party on Aid Effectiveness (WP-EFF) to ensure complementarity of monitoring and evaluation activities. Members asked the Secretariat to further explore and define possible ways forward.

Between June and October 2005, the Secretariat outlined this initial preliminary thinking to the Joint Venture on Monitoring the Paris Declaration (JV) in order to build collaboration from the start of the process. The JV welcomed follow-up in this area and encouraged the Evaluation Network to move forward. However, some concerns were expressed about the proposed framework approach. The Secretariat was urged to ensure a cohesive cross-country process, delivering findings as early as possible and going as far down the results chain as possible, but with minimum extra burdens for partner countries. The Chair of the Evaluation Network also outlined initial thinking at the meeting of the WP-EFF on 19-20 October.

Between October 2005 and January 2006, the Evaluation Network agreed terms of reference for a literature review and options paper. Following a competitive bidding process, the Overseas Development Institute (ODI, UK) were commissioned as consultants. The development of an Evaluation Network options paper was reflected in the draft WP-EFF Work Plan.

Between January and March 2006, Evaluation Network members were invited to join an interim task team on Follow-up to the Paris Declaration and to meet with ODI to steer the development of the options paper. The interim task team is an open membership and informal group and the first meeting was attended by Spain, Ireland, France, UK, UNDP, World Bank, and the Secretariat. The options paper is intended to chart a way forward through a consultative process and members of both the Evaluation Network and of the JV were invited to input via telephone interviews with the consultants. The draft options paper was discussed at the Evaluation Network meeting on 30-31 March and was shared with the JV.

Since March, further telephone interviews with civil society organisations have been undertaken by the consultants, and the paper has been revised in the light of comments received at the March meetings and in written form thereafter.

2 The challenge

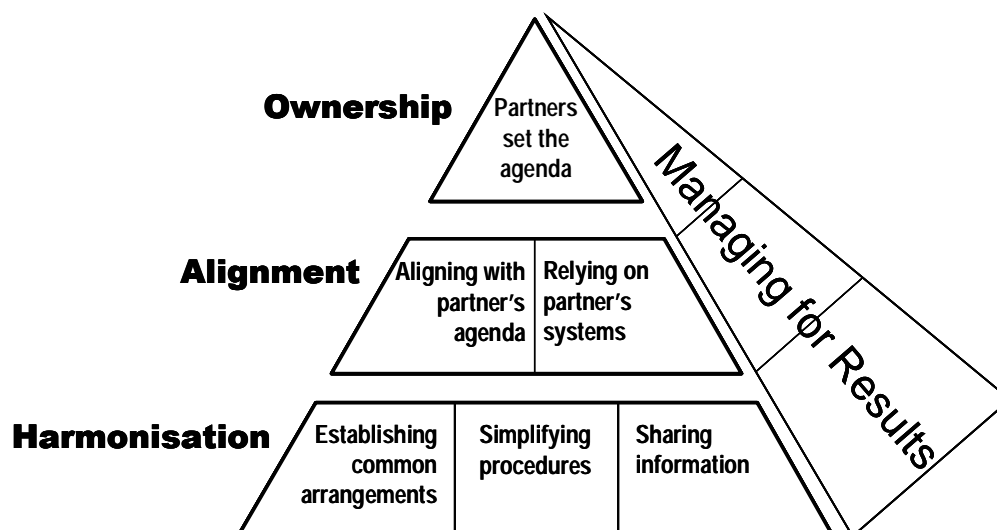
2.1 The nature of the Paris commitments

The Paris Declaration poses an outstandingly important challenge both to the world of development cooperation in general and to the field of development evaluation in particular. Compared with previous joint statements on aid harmonisation and alignment, it provides a practical, action-oriented roadmap with specific targets to be met by 2010 and definite review points in the years between. The number of countries and international organisations participating in the High Level Forum (HLF) and putting their signature to the joint commitments contained in the Declaration was unprecedented, reflecting a progressive widening of the range of voices included in major meetings convened by the OECD DAC.

The Paris Declaration may not address the full range of important concerns that experts entertain about the current state of development cooperation/aid (Rogerson, 2005). It is also important not to exaggerate the depth of the consensus at Paris, either between the different donor agencies or between the donors and partner countries. However, *the discussions before and during the HLF certainly served to clarify concepts and obtain wider agreement on the language being used*, as well as providing the action-oriented roadmap for both donors and partner countries. Conceptual confusion was noted only a few months earlier as a significant challenge (Balogun, 2005).

In various forms, the pyramid diagram reproduced as Figure 2.1 has been widely disseminated, providing a clear and accessible definition of the key terms country ownership, policy and systems alignment, and harmonisation, and the way these relate to each other and to the overarching theme of managing for development results.

Figure 2.1: The Paris Declaration concepts



Source: DAC (2004).

An important feature of the final Paris Declaration text is that ***it includes commitments not just on the established Harmonisation and Alignment agenda, but on five areas, including not just country ownership and results' management but also mutual accountability.*** The Alignment heading includes specific commitments on how to go about strengthening crucial factors such as country capacity and financial management systems.

The Declaration goes well beyond agreement on definitions. It expresses a shared view on at least the basics of how some central institutional variables fit together, and why they are important. In this way, it draws together international thinking on some of the core topics of concern to both sides of the official international aid relationship. In so doing, ***it provides an important reference point not only for the DAC members, but also for emerging aid donors and concessional funders throughout the world.***

The title of the Declaration conveys a simple but important message: aid will be more effective if the actions and behavioural changes listed as commitments under the five headings are undertaken, and less if they are not. Moreover, development results depend to a significant extent on the same variables. Underneath the consensus on these central propositions, there exist, no doubt, important differences of interpretation and emphasis.

This reflects several factors. First, there are some unexpressed but generally recognised disagreements about how the variables Ownership, Alignment, etc. relate to each other. There is not a single, universally accepted view on these matters. Second, these views are, in the main, practical axioms that form part of the current world view of particular agencies; they are based on experience, but not strongly rooted in a body of systematic evidence. Thirdly and most importantly, the “programme theory” or set of hypotheses that give the declaration its logic has not been fully articulated. ***The reasons why the Paris Declaration commitments seem to the signatories to be important ways of attaining greater aid effectiveness and development effectiveness have remained implicit.***

It might be argued that these features of the Paris consensus make the commitments an unsuitable focus for evaluation. We would disagree. We suggest below that, on the contrary, a degree of uncertainty or disagreement about details of how policy or programme objectives have been intended to be achieved is normal at the beginning of an evaluation process. Greater clarity and possibly consensus about such matters is one of the outcomes expected from evaluation work. The challenge represented by evaluating the Paris commitments is in this respect quite typical.

2.2 From monitoring to evaluation

Plans for monitoring the implementation of the Paris commitments are already quite advanced. A survey instrument is being piloted, with a view to setting a 2006 baseline in a significant group of countries (for full details see

www.oecd.org/dac/effectiveness/monitoring). Arrangements for a Medium Term Monitoring Plan are also under discussion within the Joint Venture on Monitoring with the participation of country representatives as well as DAC members and observers. In addition, the regular DAC Peer Reviews have been adjusted to take the Paris criteria into account (see Annex 4). The question we address in this sub-section is what evaluation should be expected to add to this existing effort.

We address this question at two levels. 1) In principle, evaluation and monitoring are distinct although related. 2) In more practical terms, the current Paris Declaration monitoring plans have a number of characteristics that need to be considered in mapping out an appropriate role for a programme of evaluation work on the Paris Declaration.

2.2.1 *Monitoring and evaluation in theory*

According to the DAC Glossary (2002: 27-28), monitoring is:

A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives [and] the use of allocated funds

Evaluation is:

The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results.

An evaluation “should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors” (2002: 21-22). Other standard definitions (e.g. Weiss, 1998: 330) also highlight that the purpose of evaluation is to contribute to the improvement of the programme or policy. ***In short, monitoring is expected to record how implementation is progressing, whereas evaluation questions are about whether the programme is working, how it is working or why it is not working.*** The principal basis of the distinction is the type of question asked, not the level of results considered.

While in principle the evaluation questions can be asked without a clear view of the original conception that inspired the design of the policy or programme, it is normal to begin an evaluation exercise with an effort to make explicit the causal model implicit in the design. This may be simple – in the sense that the programme documents include a Logical Framework or similar device for summarising the expected chain of results, and there is no dispute about this. Or it may be hard, in the sense that the “programme theory” is deeply embedded and needs to be painstakingly uncovered, or is the subject of different interpretations.

Much modern evaluation is considered to be “theory based”. This means that it involves as a key task bringing to the surface the programme or policy’s implicit theories of change, and then tracks the anticipated sequence of linkages

from inputs and activities to outcomes. Contrary to a common misunderstanding, theory in this context refers only to “the set of beliefs that underlie action”. Moreover, “The theory doesn’t have to be uniformly accepted. *It doesn’t have to be right*” (Weiss, 1998: 55, 338, *original italics*; Literature Review, Section 2.2).

In other words, ***it is not necessary a priori that everyone shares the same theory, or that it is robust.*** As indicated briefly above, one of the purposes of evaluation work is to focus stakeholders’ attention on what is supposed to be happening, based on what appears to be the logic of the programme, and to assist in the generation of stronger programme-theories and in due course better programmes. It is important to emphasise this to forestall the natural misunderstanding that the relationship between theory and empirical investigation is like that in economics and other quantitative social sciences, where theoretically-based propositions are expected to be quite refined before they are a suitable subject for testing.

In summary, evaluation is interested in the “how” and “why” questions, as well as the “what” questions. Therefore, it must concern itself with underlying theories of change. This suggests one element to guide thinking about what would be an appropriate contribution from evaluation to the Paris follow-up process.

2.2.1 *The Paris Declaration monitoring framework*

In general, a good supply of monitoring data is one of the preconditions of a successful evaluation. This obviously applies to the case we are considering. The Paris Declaration monitoring framework has identified a set of indicators for measuring the implementation of the commitments across countries and over time. While the emphasis in the monitoring framework is on measuring *what* is happening, an evaluation would be expected to go further to ask why some things are happening and not others, why they are taking place in particular directions and not others and, possibly, the effects on development results and impacts. This raises the question of how far an evaluation exercise could rely on the monitoring system to generate all of the necessary information on what is happening. This poses some difficulties.

When the initial monitoring is compared with the main body of the Paris Declaration, some issues are immediately apparent. The Declaration represents a compact between two principal parties, donors and recipient governments, with matching commitments on both sides. Of the 12 agreed indicators of progress, however, 8 refer to donor commitments, 2 refer to joint commitments and 2 refer to country commitments. Of the 12, 9 are being covered by a survey in which both governments and donors are participating, while progress on country development strategies, financial management systems and performance assessment frameworks are to be the subject of reporting by the World Bank.

There are good reasons why the initial monitoring arrangements have been designed in this way. Many donors and recipients think that the priority is to get behaviour changes on the donor side. Focusing the monitoring effort on the

donors and using existing systems is also a pragmatic way of handling the problem that country officials are already overloaded with information requests. These arrangements are consistent with the concern to reduce partner country transaction costs.

It seems clear that an evaluation effort would need to get a good balance between different sorts of reporting on different parts of the Paris agenda. The Declaration sets out mutual commitments that, it is strongly implied, will work if they are implemented *together*, but possibly not if they are done separately. If this is indeed the “programme theory”, the consideration of whether the commitments are working as intended needs to be centrally concerned with the relationship between the different donor- and recipient-side actions.

Clearly visible reciprocity would also be essential in an evaluation. It would be a mistake to set up a programme of work that was open to the charge that donors were being asked to assess country performance but countries were not being allowed to assess donor performance. Some of our interviewees considered that an evaluation initiative would be destined to fail if it did not involve partner countries in the definition of the agenda of questions, as well as including systematic efforts to collect data on donor behaviour.

In summary, the monitoring instrument is certain to prove a very useful source of systematic information, without which it would probably not be sensible to embark on an evaluation initiative at all. However, it is unlikely that the monitoring alone will provide sufficient data to underwrite an evaluation process. This does not necessarily mean that an evaluation would have to start from scratch and collect primary data for itself. ***It does mean that a wider range of source materials may need to be drawn upon.***

It follows from this that evaluation work on the Paris Declaration not only would have a somewhat different purpose than the monitoring exercise, but would also need to differ in respect of scope and information sources. This applies to the initial monitoring surveys in particular. Depending on the precise content of the Medium Term Monitoring Plan, which is still to be defined, it may also apply to that. According to some of our interviewees, the JV’s reporting on the medium term monitoring plan may be expected to go quite a lot beyond just saying which of the commitments are being realised and to what degree. ***Particularly if this is so, any Evaluation Network task team on the Paris Declaration should expect to have a very close relationship with the JV Monitoring in order to avoid any risk of duplication and ensure that evaluation is integrated with the Medium Term Monitoring Plan.***

2.3 Outline of an evaluation framework

It is not the purpose of this paper to define an evaluation framework (results chain and evaluation questions) for the Paris Declaration. For reasons that we explain further on, this is a task that needs to be undertaken in a highly consultative way, with involvement of a range of stakeholders including country officials and experts. However, it helps in taking forward the discussion

on whether evaluation can contribute in a useful way to Paris Declaration implementation to visualise what a framework might look like.

Figure 2.2 is an indicative *outline* framework. It would need to be developed, with work to spell out in more detail the various complexities of the expected causal links and to investigate the degree to which different Paris Declaration signatories share the same understanding of these issues.

A number of points need to be made about why the results chain is presented in the way it is in Figure 2.2.

2.3.1 *The results chain*

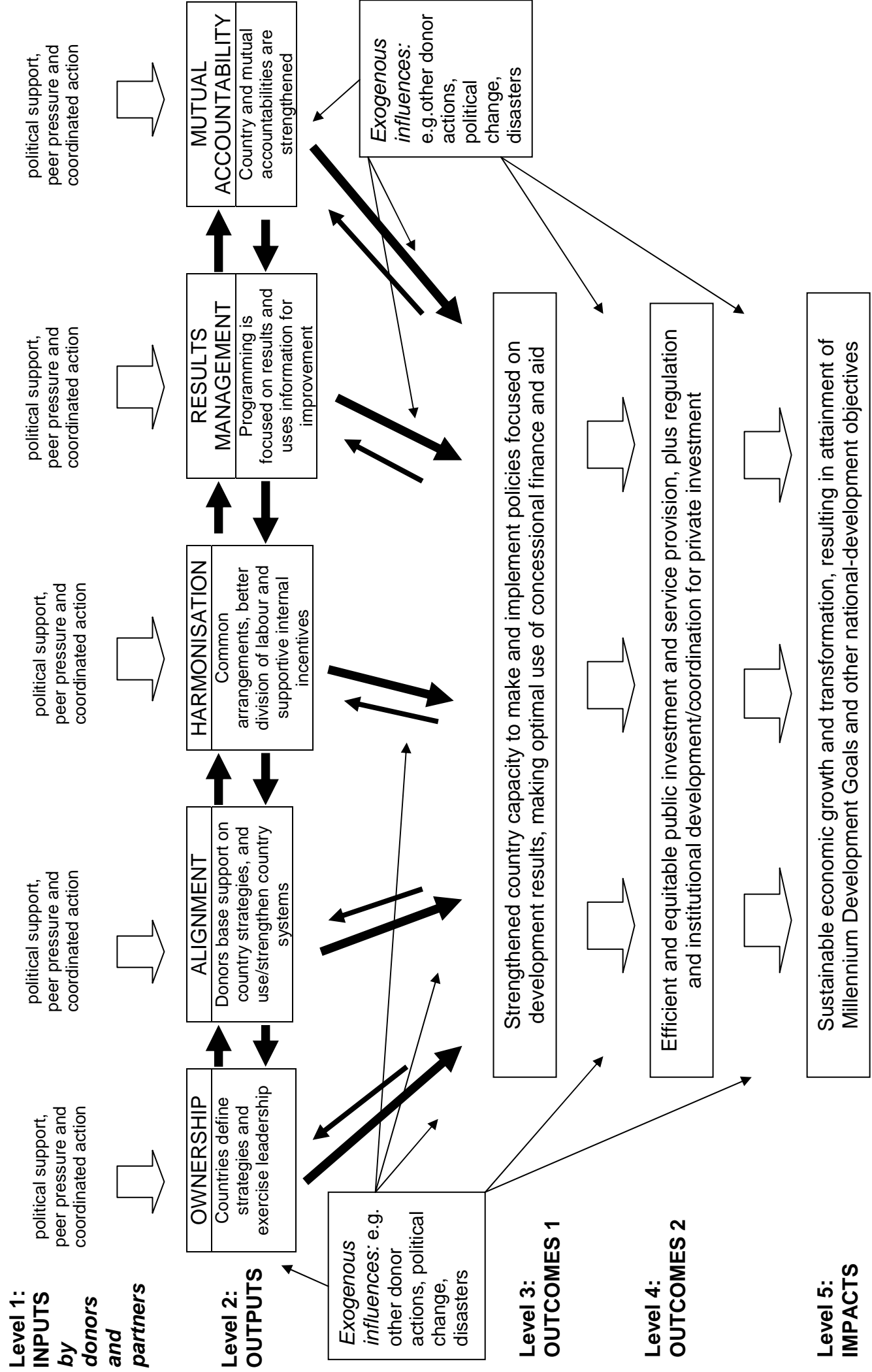
In the figure, the actions and changes in behaviour to which the Declaration commits the signatories are treated as “outputs”. The necessary “inputs” are identified, using the language of the Paris Declaration, as “political support, peer pressure and coordinated action”. It is assumed that this is a summary phrase that indicates the importance of a range of types of necessary input, on both the donor and the country sides.

As the arrows indicate, the results chain includes some very important relationships among the “outputs” (behaviour changes). These are probably reciprocal and cumulative – e.g. ownership is strengthened by alignment, and alignment needs a measure of ownership. It might seem tempting to consider some of the outputs as outcomes for this reason. However, this would depend on taking a view that goes beyond the current consensus. For example, we suspect some believe that ownership is lacking primarily because of weak alignment, while others would see ownership as largely independent of donor behaviour, and a function of local politics and history. Similarly, there are different views on how far harmonisation can contribute to improved outcomes if government leadership is missing. ***These are key evaluation questions, so the diagram does not suppress them but rather tries to bring them into view.***

It seems very possible that the indicated effects of the outputs on the outcomes are only expected where several of the outputs are realised to an important degree. This is hard to convey in such a diagram, but would need to be considered in any elaboration of the outline framework. It is also possible that there are negative feedback loops as well as positive ones. For example, policy ownership may be undermined by over-zealous pursuit of harmonised working which calls for the generation of additional “conditions”. This is something noted in successive Annual Reviews of Development Effectiveness by the World Bank. The ARDEs also document more generally the absence of evidence on the directionality of causation among institutional variables of this sort (see Literature Review, Section 2.1).

We suggest it is necessary to distinguish at least two levels of outcome to express how realisation of the Paris commitments is expected to make aid more effective, and how this results in greater development effectiveness. These are labelled Outcomes 1 and 2. This allows “impact” to be used in the conventional way to refer to the final level of development results.

Figure 2.2: An indicative outline for a possible evaluation framework for the Paris commitments



The theory articulated by the model says:

- ❑ that country ownership and the other outputs promised by the Paris Declaration would, if implemented together, strengthen country capacity to make and implement policies focused on development results and make good use of aid;
- ❑ that country capacity enhanced in this way would raise the quality of public investment and service provision, including regulation and institutional development for private investment; and
- ❑ that this would lead to better development results, such as growth and transformation, and the MDGs.

As suggested by the earlier discussion, ***this is a “theory” which may and may not be widely shared, and which is quite possibly at least partially wrong.*** Some academic specialists take the view that we still know relatively little about the drivers of institutional change in general (Evans, 2004). It seems entirely possible, for example, that what donors or recipients do outside of this particular programme logic (e.g. rewrite the constitution to include previously excluded groups, or radically change electoral rules to break political deadlock) is actually more influential on outcomes than what they do within it. That is, change that improves development outcomes at country level may occur without any change in partnership behaviour, at least not along the dimensions identified in the Paris Declaration.

In embarking upon evaluation work on the Paris Declaration, this needs to be fully appreciated. ***The benefits from a successful evaluation would not be limited to showing in a general way that the Paris Declaration is working or not working.*** They could and should include propositions of the type: “it is not working in the expected way because...” and “progress is being made in unexpected ways, or for other reasons altogether”. Such findings could have important implications both for partnership practices and for the monitoring framework.

2.3.2 Range of applicability

The framework is intended to be generally applicable as a basis for formulating relevant questions. However, some of the results linkages would not be expected to work at all under some moderately widespread country conditions – for example, where country leadership of a results agenda is completely improbable, or where the donor presence is not significant. One of the questions that evaluation work would be expected to address is under what range of circumstances does the model work, and where would a different set of commitments be more appropriate.

The current convention among DAC members is to treat the matter of the diversity of country conditions in terms of a simple two-way classification of normal and fragile states. The Paris Declaration has a specific sub-section under Harmonisation enumerating commitments appropriate in fragile states. But there is also an underlying recognition and consensus that the relevance of the general

prescriptions depends on country circumstances. Evaluation could be expected to give more sophisticated substance to this consensus.

2.3.3 “Risks and assumptions”

An important feature of the framework is the place given to “exogenous factors”, or necessary conditions other than those included in the main results chain (in logframe terms, Assumptions and Risks). In addition to the possibility mentioned above that significant changes happen because of things that actors do that are not included in the model, there is the possibility that expected effects do not occur or unexpected ones do occur because other necessary conditions are not met. These may be absolute barriers, conditions *sine qua non* of the expected outcomes. They may alternatively be inhibitors, which delay otherwise expected effects. And they may be contributory factors that make the main relationships non-linear. Both entrenched donor-side behaviours that are not covered by the Paris Declaration and purely contextual factors such as political events or economic shocks could be important in this way. The box on “exogenous influences” in Figure 2.2 is not exhaustive but is meant to draw attention to the range of such factors that may be important.

An elaborated framework would be expected to spell out in detail the exogenous factors expected to be important, drawing on comparative social science as well as stakeholder views. Evaluation work would be expected to assess the ways in which they have actually been determinant. The experience from the General Budget Support evaluation illustrates how important it is to be clear that these issues are as central to the evaluator’s task as whether or not expected positive changes can be shown to be occurring (Literature Review, Section 2.2.3).

2.3.4 Using the framework

Before being used to guide evaluation work, not only would the figure need to be made more elaborate. It would also be necessary to specify a set of corresponding evaluation questions, and indicators, proxy measures or checklists of empirical questions. We do not get into the discussion of indicators here but the next sub-section does undertake a preliminary discussion of evaluation questions.

A final comment about the use of an evaluation framework of this type is that a real-world evaluation would not be expected to focus on all of it. In fact, we recommend strongly (see below) that a common framework should be used selectively and for somewhat different purposes by different actors.

2.4 What needs to be evaluated?

The outline framework does not by itself suggest what should be the main focus of evaluation work. Nor does it tell us what would be the right range of evaluation questions to be addressed over a given period, such as the two years or so between now and the next High Level Forum in 2008, or between now and

the overall Paris Declaration target year of 2010. We discuss these issues briefly before going on to some of the more obvious methodological challenges.

2.4.1 *Where should the main focus be?*

It seems clear that for at least the next few years, the focus of any evaluation work should be on the top two Levels of the framework. That is, it should begin by establishing, with the help of the monitoring survey, how far political support, peer pressure and coordinated action (from countries and donors as appropriate) are working to get the behaviour changes to which signatories have committed. “How, why and why not” questions at these levels would then be addressed. We think this would pose an adequately challenging evaluation agenda, especially if delivery on the commitments is weak or uneven. Country representatives interviewed for the paper emphasised this quite strongly.

The nature of the interactions between the behavioural changes specified in the Paris Declaration under the headings of Ownership, H&A, RM and MA would also be suitable for early attention from evaluation teams. Once again, the questions about the results chain may be about why things are *not* happening as suggested by the model; but they would not be for that reason uninteresting. Firm findings along the lines of “the outcome is not occurring because ...” contribute significant learning. We need to be asking sensible questions about why things are not happening as well as questions about why they are (Balogun, 2005).

To be able to demonstrate that what has happened at Level 2 is indeed producing the beneficial effects in terms of country capacities to make and implement policies focused on results (Level 3) would be an important achievement. It would also be of great interest to be able to establish why any such effects were weak. Here it would be very important to consider exogenous factors as well as the endogenous inter-dependencies of the model. ***In contrast with Levels 1 to 2, Levels 2 to 3 involve quite an “advanced” set of relationships to evaluate.*** Unless the behaviour changes have happened in a mutually reinforcing way to a significant degree, the model does not predict any benefits at Level 3. ***It might be possible to get to this level for a few countries and/or sectors, but this cannot be taken for granted.***

The relationships from Levels 3 to 5 would not be expected to be a major focus of evaluation effort. Trends in these variables are not hard to establish. However, both time-lags and attribution are increasingly important at these levels. It does not seem likely that the payoff from attempts to track effects from the Paris-related outputs to these levels would justify the effort invested.

Emphatically, this does not mean that the framework is unconcerned with development results, or that there is nothing to be learned from impact evaluation. On the first issue, it may or may not prove possible to show that Level 3 Outcomes can be produced, without undue opportunity costs (see Evaluation Questions below) or negative side effects, by the Paris-related Inputs and Outputs (or that they cannot be, for specifiable reasons). However, if something of this kind were to be shown, that would be a substantial contribution

to understanding of aid effectiveness issues. This could only be said *not* to have a significant bearing on development effectiveness (Levels 4 and 5) if it were hypothesised that the strength of “country capacity to make and implement policies focused on results” is an insignificant factor in the performances and other outcomes and impacts specified further down the results chain. That hypothesis would be extremely hard to sustain.

It should be noted that this implies no necessary assumption that governments are actually motivated by the desire to reach the specific impacts indicated at Level 5. The assumption of the model is that if country capacity to make and implement results-oriented policies were improved in the indicated way, this would have *effects* of the kind specified further down the results chain.

On the second issue – the importance of impact evaluation – evaluation work on impacts does need greater and more sophisticated attention in development work (Levine, n.d.; Savedoff et al., 2005; Literature Review, Section 2.3.3). It is true that in a number of programme areas, we do not have very good evidence on what works and what doesn’t. However, this is an “evaluation gap” of a different order from the gap in knowledge that is being considered in this paper. Unless we learn how to get country policy systems that are oriented towards results and learning, there is going to be a limited audience for the findings of more sophisticated impact analysis. ***Efforts to improve understanding of “what works” at these two very different levels of the development enterprise are in this sense highly complementary.***

2.4.2 Evaluation questions

This sub-section is only illustrative. Formulating appropriate evaluation questions is a task that needs to be part of the collaborative and consultative process of elaborating the framework, along with investigation of the degree to which different stakeholders share the same “programme theory”.

In general, evaluation questions need to focus on particular observed trends or events. This may relate to a particular survey finding or report on one of the 12 indicators of the initial monitoring. It may involve, instead, an observed level of the change specified by one of the 56 Paris Declaration commitments not covered by the monitoring indicators (e.g. “Strengthen as appropriate the parliamentary role in national development strategies and/or budgets”, para 48). The general form of question would be about *how* this observed level or finding is happening, and, in particular, whether the answer is the one suggested by the model or not.

The emphasis would be different at different levels. Thus, Level 1 type questions would focus largely on what is happening or not happening and how/why. Level 2 would focus on the relationship between the “what/how/why” issues and key qualitative and quantitative change variables. Level 3 questions would be looking to ask whether things are changing in directions consistent with the programme logic and the degree to which that logic is complementary or

conflictual over time (because of differences of interpretation, inherent inconsistency or exogenous influences).

The World Bank's OED distinguishes two types of overarching evaluation question: "are we doing things right?" and "are we doing the right things?"

It seems likely that both types would need to be addressed in evaluating the Paris Declaration commitments. This would result in valuable learning on how things should be done differently, and what else should be done that is not included in current Paris agenda. There would also be a need to ask whether the monitoring system is measuring the right things in the right way from the point of view of ongoing learning and accountability.

A relatively advanced type of question that would arise if and when significant progress was apparent at Levels 1 and 2 of the framework is about cost-effectiveness. For example, if aid harmonisation were shown to be having noticeably beneficial effects in terms of a country's policy-making capacity, the question could arise whether the benefits are commensurate with the opportunity costs incurred. This might include whether the time and other resources invested in harmonisation could have been used to greater effect using more traditional project approaches, whether any gains from greater alignment needed to be set against increased fiduciary risk. Some of those consulted for this paper expressed strong interest in these types of issues. While they are undoubtedly important questions, we are not persuaded of evaluability in the short to medium term because of the measurement and counterfactual challenges that would be involved.

To complete this discussion, an important general observation needs to be made about evaluation questions and timescale. Experience with the GBS evaluation has confirmed the need to be cautious about the time it takes for changes in aid management to produce noticeable effects even at the intermediate levels of a results chain (Literature Review, Section 2.3.2). This suggests that it is important for all participants in any joint exercise to have realistic expectations on the likely form of evaluation findings.

Unless inputs have been in place for a minimum of five years, the function of an evaluation at country level is likely to be largely formative ("intended to improve performance") rather than summative ("to provide information about the worth of a programme"; DAC, 2002: 23, 35). What this means is that findings typically take the form of tentative suggestions about what seems to be working well and what doesn't. They do not typically involve firm claims about whether the approach is broadly right or not. In the case of the Paris Declaration, there should not be any expectation of summative findings ahead of the 2008 HLF.

The distinction between formative and summative evaluation is not primarily about Levels of the results chain. However, it is also the case that ***evaluation findings suggesting credible linkages from Levels 1-2 to Levels 3-4 in Figure 2.2 would begin to be possible after about 5 years of implementing the Paris commitments – that is, after 2010 – and not before***

2.5 Can it be done? Data and methods

This paper is concerned with the “evaluability” of the Paris commitments – that is, ***whether they are capable of being evaluated in a reliable and credible fashion***. So far we have suggested that the commitments do form part of a coherent set of intentions, in which inputs are being provided with a view to definite objectives. We have argued that the fact that the theory of change is largely implicit and not necessarily agreed in detail is not an obstacle, but a typical evaluation challenge.

Reliability and credibility depend on the availability of suitable data and the possibility of applying appropriate methods to settle questions about attribution. We make some brief comments on each aspect.

2.5.1 Data

The data needs of an evaluation of the Paris Declaration commitments are potentially vast. However, in view of the points above about the need for modest ambitions during the initial benchmarking period, this should not be over-stated. The practical considerations that are discussed in Section 3 also suggest that lack of data is not likely to prove an important constraint in practice.

For the reasons already given, a programme of evaluation work should not be expected to rely solely on the information generated by the initial monitoring programme of the Paris Declaration. Some additional inputs would be needed on in-country processes, in which the *interactions* of the inputs and outputs of the model can be observed and interrogated. We consider later what form country-level evaluation work might most suitably assume, because there are practical and policy constraints to be considered as well as data needs. We are also persuaded by those who suggest that any evaluation work would be ill-informed if it did not draw on at least some analysed data on cross-donor as well as cross-country performance at the input and output levels.

In other words, to be credible, evaluation work on the Paris Declaration could not rely on only one type of information source. We also think that it would not need to. The list of broadly relevant sources of data with which to address evaluation questions arising from the proposed framework is actually quite long. As far as we are aware, it includes:

- the monitoring survey;
- other country studies dealing with aid-relationship issues (e.g. Independent Monitoring Group [IMG] reports for Tanzania, Mozambique and Cameroon);
- the follow-up work on the fragile states principles;
- regional and country joint H&A monitoring exercises (e.g. Nicaragua, S.E. Asia);
- other country-level and cross-country evaluation work (e.g. in-country evaluation initiatives, the GBS evaluation, corporate performance evaluations by the IFIs); and

- studies of donor agency reforms in response to the H&A agenda.

It may be thought desirable for an evaluation of the Paris Declaration commitments to take the form of a series of country evaluations whose findings are capable of being aggregated upwards to provide a cross-country synthesis. This would perhaps be the model suggested by the GBS evaluation. However, leaving aside the lessons that are being drawn from that experience, there are questions about feasibility. To anticipate a little the argument of Section 3, a very systematic and directive approach to country evaluations may not be considered an appropriate way of following up the Paris Declaration. Generalising from country case studies is anyway known to involve quite daunting methodological challenges that it may be wise to avoid. White (2005) has argued that while there has been a real push for the use of “case studies” in evaluation, techniques for handling these are still relatively rudimentary (see Literature Review, Section 3.1.1).

In summary, an appropriate evaluation effort for the coming period does not need, and would be unwise, to rest on a single standardised set of country evaluations. It can and should draw on the variety of types of information from different sources. This might mean that the formative character of the evaluation effort would need to be stressed even more strongly.

2.5.2 Methods

If data constraints do not seem prohibitive, what about methodological snags?

A standard problem in this regard is the difficulty of handling the “counterfactual problem” – that is, whether there is a credible method of assessing what results would have been obtained in the absence of the programme inputs or other causative factors that are in focus.

In the GBS evaluation, this was a particular difficulty, given the lack of a suitable “control group” with features similar to countries receiving GBS except in this one particular. The counterfactual had to be handled more in the tradition of historiography and less in the manner of experimental science by means of intelligent conjecture (“what would have happened if ...”). This can work, so long as there is a sufficiently rich fund of testimony from well-placed observers, so that it is possible to try out alternative scenarios in a thorough way. It may be applicable to country work supporting the Paris Declaration follow-up.

Fortunately, however, systematic comparisons and the use of control groups are more feasible for parts of the Paris agenda (notably Harmonisation and Alignment) than for GBS. Different levels of H&A across sectors in a single country may well provide observed counterfactuals, particularly where the differences have been observed over a lengthy period (especially across the incumbency of different ministers or senior officials). If some controlled comparisons can be incorporated into some country evaluation work, this would significantly enhance its methodological strength and make possible a richer synthesis of relevant findings.

In both data and methodological terms, the Paris Declaration commitments do seem to be evaluable, therefore. However, there are important design options and tradeoffs to be considered before the implications of this conclusion become clear.

3 Design of the evaluation work: Some options

3.1 Introduction

On the basis of our interviews and consultations, there are some issues to do with the contribution of evaluation to the successful implementation of the Paris agenda on which there is almost complete consensus. This refers particularly to the usefulness of elaborating a framework that makes explicit the results chain implicit in the Paris commitments. This has wide support, on the understanding that there may be – and almost certainly are – different views about how exactly the components are supposed to work together to improve development results.

The proviso implies that *systematic consultations involving different donor and country stakeholders would be both a necessary feature of, and one of sources of benefit from, working on such a framework.* For both technical (evaluation science) and practical policy reasons, the idea of proceeding into an evaluation exercise without prior thinking about evaluation frameworks is unattractive. There is also little alternative to recognising that the evaluation work undertaken with the framework will be initially formative, and only much later summative, in character.

In other respects, however, there are genuine alternatives to consider. At three levels, there are basic choices to be made about the design of evaluation work on the Paris commitments, with fairly evenly balanced sets of advantages and disadvantages associated with each choice. The design choices to be made include:

- ❑ the **scope** of the evaluation, especially the selection of units of evaluation and the range of themes to be covered;
- ❑ ways of ensuring **quality and consistency** in the evaluation work; and
- ❑ the **governance and management** arrangements appropriate to this type of exercise.

The following subsection considers the arguments for and against alternative choices on the above design issues. It does so with reference to three sets of concerns that have been more or less strongly expressed in the discussions about the follow-up to Paris, within which it is recognised that there are some major trade-offs – that is, it is only possible to have more of one desirable quality at the cost of having less of another. The matrix formed by tabulating the three design choices against three “trade-off variables” is presented in full in Annex 1, and described in outline below.

In the final sub-sections, we make the case for a combination of options that appears optimal. Section 4 suggests a timetable and next steps.

3.2 A matrix of options

The matrix in Annex 1 sets out possible ways of handling the basic design questions that would maximise one or other of the following pairs of desirable but to some degree incompatible properties of the exercise:

- **stakeholder buy-in** to the evaluation process *versus* the “**criticality**” of the evaluation content;
- **country-level** learning *versus* **cross-country or global** learning; and
- low financial **costs** *versus* the **value added** to knowledge and capacities.

We now explain the reasons for believing that there may be trade-offs between these variables with respect to each of the basic design choices.

3.2.1 Stakeholder buy-in versus “criticality” of evaluation content

The importance of early stakeholder, especially partner country, engagement has been stressed by most of those consulted in the preparation of this paper. Good practice in undertaking development evaluation also underscores the importance of stakeholder buy-in, partly as a means of ensuring that evaluation work is *relevant* and that *uptake* of evaluation findings is good (see Literature Review, Section 3.2.1). Today, it is not only the recognition that aid is a partnership that has motivated change in development evaluation, but also the recognition that beyond partnership is the possibility of real country ownership.

In short, we are facing a crucial opportunity for recipient countries to take the lead in the design and implementation of evaluation in ways that support learning and more effective aid management at the country level.

At the same time, to be credible, evaluation must be seen as impartial and independent. Independence is not the same as isolation. It is essential that evaluation generates lessons that can be formative for those who design and manage aid programmes, and this means quite intensive interaction with programme managers and other stakeholders. However, we need ways to ensure that evaluators can remain detached while interacting intensively with stakeholders, so that evaluations achieve relevance while maintaining their *criticality* – the ability and willingness to judge in an objective and transparent fashion (Literature Review, Section 3.2.1).

This discussion suggests a good measure of stakeholder buy-in can be combined with the features associated with independent evaluation, and that indeed the quality of evaluation depends on achieving this. However, it is not suggested that this is easy. **In practice, if design choices are made with a view to maximising stakeholder engagement (referring to both country authorities and donors), this may make it hard to attain a high degree of criticality of evaluation content.** Column 1 of the matrix sets out how this could happen at the level of each design choice.

For example, maximising stakeholder engagement may well mean that the only unit of analysis for evaluation work is the country. As reflected in the Paris Declaration, both country representatives and many donor representatives see the country as the only appropriate unit of analysis. However, it is arguable that the ability and willingness to make objective and transparent judgements about the implementation of the Paris commitments would be guaranteed more effectively if evaluators were free to draw on assessments based on other units of evaluation – most obviously the donor agency.

It also seems clear that country-level buy-in from both country stakeholders and donors would be maximised by an evaluation design that permitted a radical “localisation” or “adaptation” of the evaluation framework and its management, so that the evaluation work can support a learning process attuned to the country change dynamic. However, if this leads to evaluation questions being negotiated in such a way that tough questions are avoided, some criticality in the evaluation will be lost.

Conversely, a more purposive design might enhance *criticality* but at the possible cost of *additionality* at the country level. In addition, maintaining a high degree of criticality would indicate a higher degree of centralised quality assurance and more standardised practices than a more localised approach geared to maximum buy-in.

We return below to how an optimal mix of these qualities might be achieved.

3.2.2 *Country learning versus cross country learning*

Although the Paris Declaration commitments are essentially global in nature and, therefore, do not offer a direct guide to specific country processes, the logic behind them makes primary sense at the country level. Interviewees were virtually unanimous in the view that the evaluation work should make a substantial contribution to country learning. Nevertheless, it was also widely accepted that to be useful in facilitating and guiding behaviour change on a global scale, evaluation should generate some cross-country lessons and yield some interim conclusions about whether the commitments are “working” in time for the High Level Forum of 2008. ***There is some necessary trade-off in meeting these concerns – between designing and managing an evaluation focused largely on supporting country-level learning and one in which a premium is placed on comparative and cross-country learning.***

The second column of the matrix therefore shows alternative ways of addressing the basic design choices which would favour one or the other of these qualities. For example, country learning is likely to be maximised if there is self-selection of participating countries. Similarly, flexible application of the common framework at country level would support country learning, recognising that some of the Paris commitments may be of particular local relevance and that there may not be a one-to-one correspondence between the commitments and the relevant in-country initiatives. However, the likely result would be that a lot would be learned about how to do things better in a few “model” countries. Less would be learned about the usefulness of the general approach – the correctness or otherwise of

the “programme theory” – underlying the Paris Declaration, the range of countries to which it can be applied and in what respects it needs to be changed.

In contrast, some form of purposive sampling would meet better the needs of comparative learning. Cross-country learning would imply a greater emphasis on centrally agreed procedures and standards, and a more centralised or international governance and management system for the evaluation. Opportunities for increasing understanding of the Paris issues at the global level would thereby be increased.

3.2.3 *Low costs versus value-added*

Development evaluations are generally costly, in both a direct and an indirect sense. Ensuring additionality requires careful design, management and implementation – all of which have high financial (and opportunity) costs. Typically, the indirect costs go unenumerated but find expression in the diversion of already scarce capacity within partner governments or in agency field offices. One effect is that the knowledge creation and dissemination that should go along with quality evaluation often gets lost. As noted at the beginning, the Paris Declaration text expresses in an extremely firm way the concern that any additional cross-country monitoring and evaluation should not impose additional burdens on partners.

As the third column of the matrix illustrates, **one way to comply with this last concern is to reduce severely the data needs of the exercise.** A commitment could be made to only make use of data already being generated by the survey of the 12 indicators, by existing country-level Harmonisation and Alignment monitoring initiatives, and by other completed or planned evaluations with a relevant thematic focus (e.g. those undertaken by the multilateral development banks). This would probably imply a more restricted effort to answer evaluation questions generated by the framework, with a focus only on piecing together the elements of a story about *Paris Declaration-like* behaviour changes. We have argued in the previous section that this is unlikely to be the preferred route.

At the opposite extreme, the evaluation could be deliberately ambitious, on the grounds that the scope for learning at this moment in the history of the aid relationship is large. The task would be to test as rigorously as possible the assumptions of a comprehensive model of behaviour change based on the Paris Declaration, with a range of purpose-built base evaluations covering countries and donors to make this possible. Costs would be high but justified.

Further down the matrix, corresponding choices are illustrated with respect to quality control and management. In the general field of evaluation, concerns about costs and benefits have led to a range of approaches to these issues which distribute responsibilities and costs differently among the participants. Several of these are candidates to be used in the Paris Declaration evaluation work.

There are examples of highly centralised evaluations where all the costs are borne by the development agency(ies) and independent evaluators from northern based institutions do much of the work. This reduces the costs for country

partners but also may reduce the benefits to them, including the possibility of building up capacity in evaluation management over time. Combined with a reduced evaluation agenda, this approach could enable the Paris Declaration work to comply fully with the injunction not to impose additional costs on country partners.

There are also hybrid models in which the evaluation is entirely delegated to a country-level donor and partner grouping, or where the evaluation is divided up into several parts each of which is managed differently and separately. These models spread the costs to a greater extent. They could be applied in cost-minimisation mode, or to support an ambitious programme promising substantial long-term benefits to all concerned, including enhanced capacities. ***The danger is that in looking to minimise additional costs, potential value-added would be lost, in terms of both what the evaluation can say globally and what is left behind at country level.***

The solutions proposed below reflect in part an effort to think creatively about how to minimise unnecessary additional costs, particularly for partners (but also for donors). This is moderated by concern that the processes and products that emerge are additional in terms of knowledge creation and material to development effectiveness.

3.3 Towards an optimal solution

The options set out in the matrix somewhat overstate the conflicts in order to be illustrative. What the matrix serves to do, however, is to illustrate the tensions that are inevitably present in as complex an evaluative exercise as the one being considered here. Few large joint evaluations have been able to resolve all the tensions involved (e.g. the CDF evaluation did not).

The important conclusion that we draw from the discussion of the matrix is that it is impossible to think of a way of approaching the Paris Declaration exercise that will make everyone happy and satisfy all concerns. ***The best that is going to be possible is an optimal solution, one that achieves a relatively high score, even though not the maximum, on all important variables.***

The idea that we now put forward is that the best way of optimising (and the most pragmatic approach) is to work on a number of parallel tracks simultaneously – that is, undertaking four different types of activity, recognising and giving full play to the distinctive advantages of each. Since what ultimately emerges will be the result of several rounds of learning, reflection and iteration, not only on the findings of the evaluation but also on the evaluation methodology and approach, there is much to be said at this point for spreading risks and not attempting to resolve all questions in a single “big bang”. While the suggested approach may appear complicated, we think it is in fact the most practical of the alternatives considered.

In general, the components are visualised as feeding into each other but as not requiring a heavy apparatus of coordination, so as to maximise particular benefits and keep costs down.

Four activities or sets of activities would be undertaken:

- a) ***The development of a common framework*** articulating the logic of the Paris commitments, including the implied linkages between aid effectiveness and development effectiveness.
- b) A series of ***country-led country evaluation initiatives***, building directly on the joint monitoring activities being undertaken and other knowledge already available.
- c) A set of ***donor case studies*** which would build on DAC Peer Reviews to explore in detail the way in which specific Paris Declaration commitments are finding expression ***across a sample of donor organisations*** as opposed to countries.
- d) ***A medium to long term programme of analytical work*** designed to draw together and critically evaluate findings from a variety of sources with a bearing upon the common framework constructed by Activity 1). This component would need to be integrated with the Paris Declaration Medium Term Monitoring Plan.

From the point of view of the follow-up to Paris as a whole, b) and c) may be considered “source” activities, whereas a) and d) are conceptual and synthetic. a) is seen as providing structure and a degree of guidance to b) and c), and these feed into d) along with other evidence and analysis.

3.3.1 Development of a common framework

This component would be most effectively undertaken by the Evaluation Network in intensive dialogue with a range of key stakeholders. ***The approach would be top-down but agile and consultative.*** The task would include facilitating a process, possibly including regional workshops, to agree a core set of evaluation questions and guidelines on methodological issues (e.g. use of sectoral comparisons) with donor and country stakeholders.

This is the activity that appears to command the most unqualified support on all sides. It is regarded as having value in its own right, assuming it is well done, as a means of taking forward dialogue, advocacy and critical reflection on the Paris Declaration commitments and broadening awareness of them, both among those already involved with the DAC and among “new donors”. To repeat the observations made earlier, the purpose would not be to impose a theory of institutional change and the aid relationship on everybody, overlooking its likely flaws. It is, on the contrary, to discover to what extent there is a common understanding of what the Paris Declaration means, where differences of emphasis or interpretation are significant, and the degree to which the different linkages are held to be robust on the basis of experience or research-based knowledge.

This activity would contribute importantly to the other proposed activities and would need to be undertaken very soon for this reason. It would provide a general guide for the country-led work proposed as b) above, and it would create the potential for significant global learning drawing on different types/levels of source analysis as proposed in d) above.

3.3.2 *Country-led country evaluation initiatives*

This activity would be very country-led and structured so as to contribute in important ways to country-level learning and not impose additional burdens on country officials. Countries would be self-selected, making good use of the networking already undertaken in the context of the JV and the WP-EFF. Although it would be better for global learning to have a purposive sample of countries undertaking evaluation work, it seems too early for the major efforts that this would involve.

Major efforts would be made to localise elements of the common framework in ways that suit country-specific circumstances, so that local learning needs are met as fully as possible. The framework would provide a core set of evaluation questions. To the extent these were incorporated by the country-led exercises, this would provide a basis for systematic cross-country analysis. However, country teams would be encouraged to go beyond the core to identify local evaluation questions of relevance to country specific processes.

Because of the demands and possible costs of carrying out a comprehensive country level evaluation of the Paris Declaration, ***country exercises could be limited to specific objectives within the Paris Declaration framework*** (e.g. alignment or harmonisation), to specific levels (e.g. Levels 1 and 2) or to specific sectors where there is a possibility of comparing better and worse performance in realising the Paris commitments. Stronger stakeholder buy-in would be expected than if the framework was applied rigidly across a sample of countries. There would be no expectation at this stage that the lessons from the individual country evaluations would be additive in anything but the most general sense. Some key principles of independence and objectivity would need to be seen to apply, but would have to be assured locally, with a limited amount of technical support.

Evaluation management would need to be delegated to the country level. Ideally this would be to an independent management group with some standing capacity to follow up on evaluation issues in the future. Financing could be managed by a donor fund to support the independent management group under the oversight of a delegated lead donor and/or a “lean and mean” steering committee at country level. Evaluation findings would need to be discussed at country level between donors and partners before being communicated to the DAC for information at the HLF 2008.

3.3.3 *Case studies of a sample of donors*

For reasons similar to those applying at the country level, ***agencies participating in this exercise would need to be self-selected, and implementation of an agreed set of enquiries would have to remain in the***

hands of the donors themselves. The reason for the Evaluation Network to concern itself with the promotion of this additional layer of follow-up to the Paris Declaration was expressed eloquently by several of the people consulted about this paper. The central point is that the reasons behind differential implementation of the Paris Declaration commitments, and what this might mean, are likely to remain obscure so long as the data are systematic only across countries and do not deal with issues such as HQ/field relations within agencies.

It is true that the main emphasis of the monitoring survey is on indicators that attempt to capture changes in donor behaviour, that there is an expectation that these results will be tabulated by donor as well as by country and that the DAC Peer Reviews have been adjusted to take the Paris criteria into account (see Annex 4). However, issues that we know to be important such as the internal incentive structures of the agencies will call for a thorough type of evaluative investigation, preferably covering both headquarters and country offices. **A collaboration between the Evaluation Network, the OECD Peer Review Unit and the WP-EFF might be the best way of taking this forward.**

Without this additional source of basic analysis to put together with findings from the country exercises, consensus on the Paris Declaration follow-up may be hard to maintain. Country representatives are likely to object that the whole evaluation exercise is unduly focused on them, and not enough on what many see as the principal causes of non-delivery on Paris Declaration-type commitments – donor policies. No less important, those undertaking activity 4) above need to be in a position to defend their independence and “criticality”. It is not clear that they could do so if they were wholly reliant on base evaluations about countries, supplemented by *ad hoc* sources on the donor side of the relationship.

3.3.4 Programme of synthesis and meta-evaluation work

This should have a largely top-down character in the interests of maximum cross-country and global learning about the relevance and effects of the Paris commitments. It could be managed by DAC-Evaluation Network in close partnership with the JV Monitoring. Inputs for the work would include findings from the country-led evaluations, completed and ongoing donor/joint evaluations that focus on aspects of the Paris Declaration agenda (e.g. ownership, partnership, GBS etc.) and IMG-type reports, as well as the donor-focused work-stream described in c) above. The use of multiple sources would avoid among other perils that of getting bogged down in insoluble discussions about the generalisability of country case study findings.

The activity might be properly described as meta-evaluation in two senses. It would involve aggregating and merging evaluation findings of different kinds on a single set of issues – probably selectively, not across the whole field of the framework. It would also involve assessing the quality, relevance and completeness of the inputs (including the monitoring survey) from the point of view of learning how to improve the global aid effectiveness agenda and its underlying “theory”.

Responsibility for this activity should probably rest with an independent evaluation team/group. Such a group would have the best chance of applying the necessary principles of criticality, additionality etc. and to ensure that findings are free of direct political interference and are sufficiently independent to be credible and legitimate. Steps would need to be taken to ensure that the evaluation work-plan is integrated with the JV Monitoring Medium Term Monitoring Plan (content of this Plan is still to be decided). ***There may be scope for combining the monitoring and evaluation plans into an integrated medium term monitoring and evaluation plan. Unified reporting to the 2008 High Level Forum would certainly be needed, so that delegates are not confronted with multiple reports.***

It makes most sense to see the evaluation process as a continuing activity, with results appearing at different points before and after 2008. However, a key point in the time-line is the end of 2007, when some substantive findings would be needed to feed into the 2008 HLF. This would be feasible within the suggested approach, where the “meta” work feeds off a variety of sources, several of which will be occurring anyway and is not locked into a complicated plan for aggregating country experience upwards on a fixed timetable. ***This is both an intellectually defensible strategy and a low-risk one from the point of view of having something significant to say before 2008.***

3.4 Advantages of the suggested approach

Table 3.1 summarises the ways in which the four suggested activities might be rated in terms of the main trade-off variables listed earlier. Our contention is that a single, unified activity would perform significantly worse in terms of the relevant criteria.

Table 3.1: Scoring of proposed activities in terms of trade-off variables

	Stakeholder buy-in	Criticality of evaluation	Country Learning	Cross-Country Learning	Cost minimisation	Value Added
Development of a common framework	√√	√√√	√	√√	√√	√√√
Country-led evaluations	√√√	√√	√√√	√	√√	√
Donor-focused thematic evaluations	√√√	√√	√	√√√	√	√√
Synthesis/meta evaluation work	√	√√√	√	√√√	√√	√√√

√√√ = High
 √√ = Medium
 √ = Low

4 Timetable, decisions and next steps

4.1 Summing up

This paper has considered appropriate ways of including evaluation work in the follow-up of the Paris Declaration on Aid Effectiveness. We have argued that the commitments outlined in the Declaration pose an outstandingly important evaluation challenge, and that in principle the commitments are susceptible to evaluation. Although the monitoring work already underway provides an indispensable input, an evaluation process would need to look beyond the monitoring information to be credible and command general support.

It is possible to visualise a framework for the Paris Declaration, on the understanding that this implicit “programme theory” is not the subject of a consensus in detail and is largely untested. The focus of evaluation work would need to be at the first two levels of the results chain, but findings about these levels would have definite implications for the probability of improved development outcomes and impacts. ***For a number of years, and certainly until the HLF of 2008, evaluation work would have a formative and not a summative purpose.*** The data needs and methodological challenges seem manageable, on the understanding that there is a component of cross-country and global learning that draws critically on a range of types of data and analysis.

There is general agreement on the desirability of preparatory work, carried out in a consultative way, on the evaluation framework (results chain, evaluation questions, indicators, etc.). On the other hand, there are trade-offs to be managed in settling upon a suitable evaluation design in respect of scope, assurance of quality and consistency, and governance/management arrangements. It is possible to visualise alternative design options that maximise particular qualities that stakeholders want to see. We have suggested that the optimal solution would be to work along four tracks simultaneously:

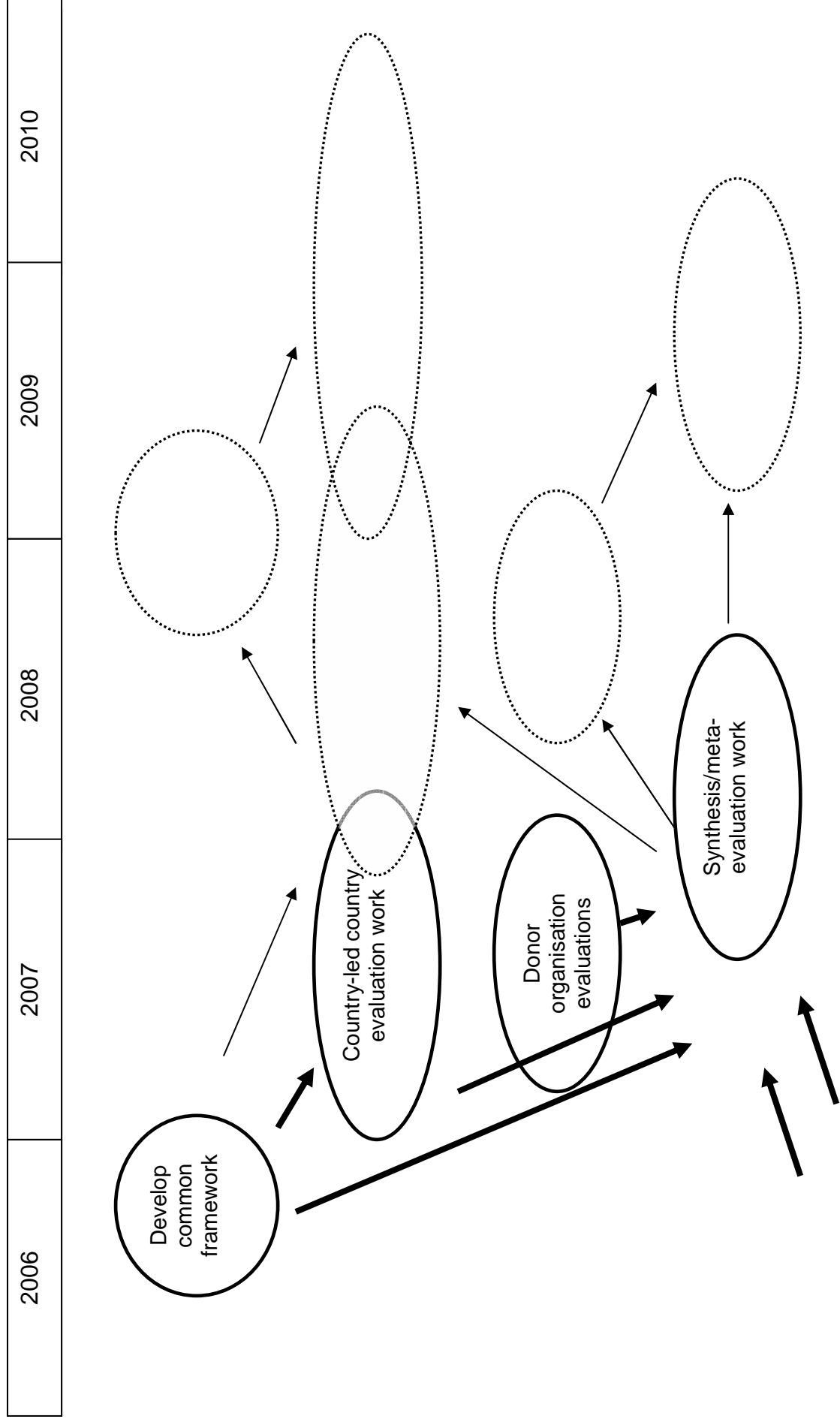
- development of a common framework;
- country-led country evaluation initiatives;
- thematic case studies across donors; and
- a medium to long term programme of synthesis and meta-evaluation work.

We have explained the rationale and key features of each of these four activities, including the way they are seen as relating to each other. However, it remains to be considered how this programme of evaluation work might take shape over time, and what the specific next steps can be agreed within the general approach.

4.2 Timetable

Figure 4.1 suggests a possible time-line for the proposed activities on the assumption that more than one of them is undertaken.

Figure 4.1: Timeline of proposed activities



As the dotted ovals and light-weight arrows imply, all of the activities are seen as ongoing or as likely to be repeated during the years to 2010. The arrows suggest some of the ways the proposed activities would inform or influence each other and feed off other processes happening independently. ***The diagram helps to make clear several important practical points about timing:***

- ❑ Work on the framework would need to be set in motion soon after the middle of 2006. That would make possible the discussion of an initial report at the November meeting of the DAC Evaluation Network.

This is essential if the framework is to have influence on the country-led country work, and if the latter is to produce reportable findings by the end of 2007, so that it can feed into the preparation of the 2008 HLF. Also, if it does not begin soon, there will not be enough time to carry the process out in the consultative way that is considered essential. It could helpfully occupy the few months until the delivery of the first round of analysed monitoring data at the end of 2006.

- ❑ Adequate time would need to be allowed for the preparation of the country work, so that the concept of a country-led process is widely and well understood.

It might be practical to combine country missions on the framework with joint activities at the country level to explain the approach being suggested for the country-led exercises, and in general build on the good DAC/country relationships that have been formed within the JV Monitoring and the WP-EFF.

- ❑ The scope and nature of the proposed synthesis/meta-evaluation work and ways of integrating it with the Medium Term Monitoring plan would need to be discussed with the JV at an early stage.

We see this activity as evaluation and not monitoring. Nevertheless, the proposed approach of using and critically assessing multiple information sources may facilitate its integration with the Medium Term Monitoring Plan being prepared by the JV, so that it forms part of a combined effort. It is extremely important to find ways of bringing together the different sources of analysis ahead of the 2008 HLF. A management and coordination mechanism needs to be established between an Evaluation Network task team and the JV.

- ❑ The proposed case studies of donor organisations should be raised in the framework of the JV and WP-EFF at the earliest opportunity.
- ❑ Slippages in scheduling of activities and delivery of reports typically cause problems for tightly integrated global joint evaluations (e.g. GBS, CDF). The proposals made here depend on a less formal type of integration; however, problems could arise if the points above were not taken into account.

4.3 Managing the partnership

The Evaluation Network does not have the same level of established and structured partnership with country representatives as the WP-EFF and the JV. In view of the practicalities, it would seem sensible that Evaluation Network relates to the *same* country representatives as the JV, and that meetings are scheduled in such a way that multiple journeys are not required for country representatives. If this seems impractical, an alternative would be for the Evaluation Network to “regionalise” itself and hold the necessary senior-level consultation meetings (e.g. on the framework and on the approach to country work) in different regional centres, so that country representatives do not need to travel so far.

The proposed activities would not seem to require the establishment of an international Steering Group for the evaluation work, beyond a small and informal Evaluation Network task team. In fact, this would be contrary to the recommended approach, which deliberately combines the virtues of a quite top-down and those of a largely bottom-up way of working through distinct activities, and tries to avoid imposing new burdens on country representatives. Ensuring that recipient voices are heard and have an influence on both the framework and the synthesis remains important. However, this should be done through intensive networking, both by the Evaluation Network task team charged with taking the proposals forward, and by the organisations commissioned to lead the implementation of these activities.

Intensive work would be required both to provide the “management of expectations” that has been suggested as a necessary element throughout this paper, and to convey to all concerned the spirit of the proposed approach to the overall task and the particular activities. ***This would be most appropriately delegated by the Evaluation Network to a small and highly motivated task team, with adequate representation of the major world language areas.***

4.4 Options and next steps

The first draft of this Options Paper was discussed at the DAC Evaluation Network meeting on 30-31 March. A large number of Evaluation Network members stated that they see this as a very important area of work and suggested that the Network starts substantive work in 2006 to ensure that useful formative findings are available to contribute to the 2008 HLF. Some members, however, expressed caution that the Network may be proposing to start work on evaluation too early in the overall Paris Declaration process. It was therefore agreed:

- ❑ to discuss the Options Paper with the WP-EFF at its meeting in July 2006 and to agree then which of the options should be pursued;
- ❑ to continue to coordinate with the JVM (who were invited to comment on the first draft) and to involve partner countries in any task team; and

- ❑ to carefully manage expectations about what kind of evaluative findings could be delivered by 2008 and how these will add value to the Medium Term Monitoring Plan to be developed by the JVM.

The following options for action are available to the Evaluation Network within the broad approach suggested by the paper:

- 1) *Evaluation Network takes no action and plays no role at this stage.* This would follow the view that it is too soon to expect any results from the Paris Declaration that are worthy of evaluation attention. On the other hand, it implies the risk that several years will go by without any systematic consideration of whether the path that has been embarked upon since Paris is the right one, and whether any corrective actions or additional efforts are needed.
- 2) *Evaluation Network develops a common evaluation framework only.* This is the activity for which there seems to be the most unqualified support, and it is one that the Network is well-placed to lead. On the other hand, a framework that is not applied may appear a rather limited exercise. The 2008 HLF is likely to expect some answers as well as questions.
- 3) *Evaluation Network develops a common framework and takes a role in facilitating country-level evaluation work.* Within the proposed approach, the Evaluation Network role is limited to adding an evaluation dimension to country-led joint monitoring processes that are already under way. This would be an appropriate, low-cost way of ensuring that there is a substantive evaluation input to the 2008 HLF. A limitation of this option is that it includes no evaluation focus on donor organisations, which could be considered one-sided in view of the Paris commitments to mutuality.
- 4) *Evaluation Network develops a common framework and takes a role in facilitating country-level evaluation work and donor-agency case studies.* This would provide a more balanced set of initial evaluation findings to report in 2008. Although in the proposed approach Evaluation Network would only facilitate the country studies and a set of donor case studies organised by the agencies themselves, it could be argued that this option would stretch the Network's coordination capacity.
- 5) *Evaluation Network only takes a role in synthesising available/ongoing evaluations (possibly in combination with option 2).* Given the fact that a good deal of relevant monitoring and evaluation work is going to be taking place independently, this is a feasible option, although probably only in combination with option 2, the elaboration of a framework. A significant limitation would be that none of the inputs would have been guided by the Network or assisted by its specific expertise.
- 6) *Evaluation Network adopts the suggested four-track approach as a whole.* As suggested by this paper, this would be a cost-effective and low-risk way of ensuring a significant evaluation input to the 2008 HLF, enabling some initial lessons to be learned and any necessary correctives to be

applied. The most important danger to be guarded against would be that of raising expectations that could not be met, given the limited and largely formative scope of the proposed work.

Depending on the decisions taken in respect of these options, the next steps to be taken might include:

- agreeing to one of the above options;
- establishing an Evaluation Network task team (with partner countries) with a remit to take this forward, in partnership with the JV and/or WP-EFF;
- organisation of consultations to agree the common framework for evaluation of the Paris Declaration commitments;
- developing a strategy and methodology for country and donor evaluations to be delivered in 2007.

References

- Balogun, Paul (2005) *Evaluating Progress Towards Harmonisation*, London: DFID, Working Paper 15.
- DAC, Working Party on Aid Evaluation (2002) *Glossary of Key Terms in Evaluation and Results Based Management*, Paris: OECD DAC.
- DAC, Working Party on Aid Effectiveness and Donor Practices (2004) "Draft Report on Aid Effectiveness for the Second High-Level Forum", Paper for Senior Level Meeting, Dec 8-9, Paris: OECD DAC.
- Evans, Peter (2004) "Development as Institutional Change: The Pitfalls of Monocropping and the Potentials of Deliberation", *Studies in Comparative International Development*, Vo. 38, No. 4, Winter.
- Levine, David I. (n.d.) "Learning What Works – and What Doesn't: Building Learning into the Global Aid Industry", Berkeley: Haas School of Business, University of California.
- Rogerson (2005) "Aid Harmonisation and Alignment: Bridging the Gaps between Reality and the Paris Reform Agenda", *Development Policy Review*, Vol. 23, No. 5, Sept.
- Savedoff, William D.; Ruth Levine; and Nancy Birdsall (2005) "When Will We Ever Learn? Recommendations to Improve Social Development through Enhanced Impact Evaluation", Washington, DC: Center for Global Development.
- Weiss, Carol H. (1998) *Evaluation*, 2nd edn., New Jersey: Prentice Hall.

Annex 1: A matrix of options for evaluation design

Trade-off variables Design choices	Stakeholder buy-in versus “criticality” of evaluation		Country learning versus cross-country learning		Low costs versus added value	
	<i>Maximum buy-in</i>	<i>Maximum criticality</i>	<i>Maximum country learning</i>	<i>Maximum cross-country learning</i>	<i>Least cost</i>	<i>Most added value</i>
Scope of evaluation (i) Unit of evaluation (ii) Thematic scope	<ul style="list-style-type: none"> One unit of evaluation: country ‘Localisation’ of the evaluation framework – few core evaluation questions and more scope for country specific questions to respond to specific knowledge gaps <p>Potentially high on additionality at country level but low on criticality</p>	<ul style="list-style-type: none"> Multiple units of evaluation: countries and donors Freedom of country and donor sponsors to negotiate the content of the evaluation is restricted, so that tough questions are not avoided <p>Potentially high on criticality but less additionality at country level.</p>	<ul style="list-style-type: none"> Countries select themselves Each country evaluation allowed to focus on different aspects of the common framework, as the purpose is to help local actors make progress faster <p>Potentially high on local learning but low on global learning</p>	<ul style="list-style-type: none"> Purposive sampling Evaluation uses several ‘building blocks’ to address a single set of core evaluation questions, and is directive about local adaptations <p>Potentially high on global learning but low on local learning</p>	<ul style="list-style-type: none"> Evaluation uses only existing data – monitoring of 12 indicators, country-level monitoring initiatives and ongoing thematic evaluations Task is to piece together a story about <i>Paris Declaration-like</i> behaviour changes <p>Potentially low money and opportunity costs</p>	<ul style="list-style-type: none"> Evaluation generates its own set of base evaluations for systematic analysis Task is to test and improve a comprehensive model linking aid effectiveness and development effectiveness <p>Potentially high money and opportunity costs</p>
Ensuring quality and consistency (i) Quality assurance mechanism (ii) Assuring consistency and comparability	<ul style="list-style-type: none"> Country-led steering/management group with oversight of quality assurance (internal to each evaluation) No common ToR <p>May affect criticality but increases sense of ‘buy-in’</p>	<ul style="list-style-type: none"> Centralised quality control through technical adviser or technical advisory group Common ToR, ground rules prepared centrally for country teams <p>Risks losing ‘local responsibility’ for a quality process</p>	<ul style="list-style-type: none"> Quality assured by local management group, possibly linked to ECD, building country evaluation capacity Each country devises own evaluation plan and timetable <p>May affect timeliness and thus possibility of synthesis</p>	<ul style="list-style-type: none"> Centralised quality assurance to ensure consistent treatment Common evaluation plan and timetable for country work <p>Contributes little to country capacity for evaluation management</p>	<ul style="list-style-type: none"> Centralised efforts limited to negotiating common ground rules for relevant evaluation work A limited set of core questions recommended for relevant evaluation work <p>Inexpensive but may not assure quality or consistency and comparability</p>	<ul style="list-style-type: none"> Take a major initiative on global and country evaluation capacity on the back of this evaluation Identify scope for local ECD and Independent Evaluation Groups in countries <p>High cost but would add value beyond quality of the evaluation</p>

Trade-off variables Design choices	Stakeholder buy-in versus "criticality" of evaluation		Country learning versus cross-country learning		Low costs versus added value	
	Maximum buy-in	Maximum criticality	Maximum country learning	Maximum cross-country learning	Least cost	Most added value
Governance and management arrangements <i>(i) Overall leadership</i> <i>(ii) Location of management</i>	<ul style="list-style-type: none"> Country-led steering group responsible for taking forward evaluation, building on work in progress. Management is localised, with technical assistance only <p>Likely to be good for local buy-in, but local management may find it hard to maintain criticality with respect to both government and donors</p>	<ul style="list-style-type: none"> Mixed international steering group, to neutralise particular interests Management of evaluation outsourced to independent technical group along the lines of the GBS evaluation <p>Pressures to achieve consensus on findings are reduced, which is good for criticality, but loss of buy in may affect relevance and uptake of findings in countries</p>	<ul style="list-style-type: none"> Active local steering guaranteed by self-selection of participating countries Localised management contributes to capacity development <p>Will result in a focus on what can be learned from 'model countries', not what can be learned across a range of countries</p>	<ul style="list-style-type: none"> Mixed international steering group, to support cross-country learning Management of evaluation outsourced to ensure purposive sampling and robust framework <p>Opportunities for cross-country/ global learning maximised, at possible cost to local learning</p>	<ul style="list-style-type: none"> Central costs kept low by light, largely top-down management structure, corresponding to a reduced evaluation agenda Country led work funded by local donor partners and managed locally to keep costs down <p>Costs of steering and management minimised</p>	<ul style="list-style-type: none"> Mixed international steering group and ambitious evaluation agenda implies substantial costs Outsourcing of whole of global exercise costs more <p>A relatively expensive option, corresponding to a much higher level of ambition, with both global and country capacity development</p>

Annex 2: Persons consulted

OECD DAC

Richard Manning, DAC Chair
Michel Reveyrand, France, Chair, WP-EFF
Josef Fuellenbach, Germany, DAC Delegate
Brian Hammond, Head of OECD Statistics
Simon Mizrahi, OECD Secretariat, Aid Effectiveness Division

DAC Evaluation Network (meeting of 8 Feb)

Finbar O'Brien, Ireland
Belén Sanz Luque, Spain
Daniel Kamelgarn, France
Alison Girdwood, United Kingdom
Sue Bassett, United Kingdom
Karen Kiernan, United Kingdom
Patrick Grasso, World Bank
Nurul Alam, UNDP
Hans Lundgren, DAC Evaluation Network Secretariat
Sebastian Ling, DAC Evaluation Network Secretariat

DAC Evaluation Network (telephone)

Eva Lithman, Sweden
Miwa Satoko, Japan
Niels Dabelstein, Denmark
Ted Kliest, Netherlands
Bruce Murray, Asian Development Bank
Saraswathi Menon, UNDP

Joint Venture on Monitoring Paris Declaration

Greg Briffa, United Kingdom
George Carner, USA
Jost Kadel, Germany
Elaine Venter, South Africa
Pham Thi Thanh, Vietnam
Mauricio Gómez, Nicaragua
Chris Hall, World Bank

Others

José Antonio González Mancebo, Spain
Paul Isenman, formerly DAC Secretariat
William Savedoff, Impact Evaluation Initiative
Andrew Rogerson, ODI, consultant to WP EFF
Andrew Lawson, ODI, consultant to GBS evaluation
Tom Beloe, United Kingdom (Bangkok office)
Sam Sharpe, United Kingdom (Policy)
Fiona Shera, United Kingdom (Policy)
Kate Tench, United Kingdom (Evaluation)
Helen Wedgwood, United Kingdom (Policy)
Régis Mabilais, Coordination Sud
Patrick Watt, Action Aid
Hetty Kovack, Eurodad

Annex 3: Terms of Reference (01/12/2005)

DAC NETWORK ON DEVELOPMENT EVALUATION FOLLOW-UP TO THE PARIS DECLARATION ON AID EFFECTIVENESS

Introduction

In March 2005, at the second DAC High Level Forum on Aid Effectiveness, over 100 DAC members and partners agreed the Paris Declaration on Aid Effectiveness (www.aidharmonisation.org). The Declaration contains over 50 commitments on ownership, harmonisation, alignment, results and mutual accountability.

Alongside its strong focus on monitoring the commitments, the Declaration states that the signatories will also explore independent cross-country evaluation processes – which should be applied without imposing additional burdens on partners – to provide a more comprehensive understanding of how increased aid effectiveness contributes to meeting development objectives.

Proposed Evaluation Framework

The DAC Evaluation Network (www.oecd.org/dac/evaluationnetwork) held a preliminary discussion in June 2005 on how it should respond to this call for an evaluation process. A number of Network members indicated support for jointly designing a common evaluation framework for evaluations of aid effectiveness at the country level. Network members stressed that the design of an evaluation framework should be undertaken with partner country participation and in collaboration with the DAC Working Party on Aid Effectiveness to ensure complementarity of monitoring and evaluation activities.

Questions and Challenges

There are various challenges to be addressed in developing and implementing an evaluation process:

➤ **Management Structure**

How should the evaluation process be managed – should the process be partner-country or donor led and/or managed by an Evaluation Network steering committee? A recent Evaluation Network report (www.oecd.org/dataoecd/28/9/35353699.pdf) on Joint Evaluations stresses the need to explore innovative ways of joint working that will minimise the costs and delays of intensive inter-agency coordination. The Paris Declaration also states that we should not impose extra burdens on partners. The proposed framework approach is one possible way of addressing this problem. Members would work together to develop an evaluation framework while its implementation might be delegated to individual or small groups of partners in different national contexts. The Network could then collectively oversee the production of a synthesis report.

➤ **Collaboration, Ownership and Independence**

How can partner countries participate in the design of the evaluation process and how can we build on partner country evaluation experiences and models? We envisage a process of ongoing collaboration with the DAC Working Party on Aid Effectiveness and also with partner country participation and ownership of findings in order to strengthen mutual accountability. The approach will, however, also need to ensure that the evaluations remain independent of undue influences from both donors and partner governments.

➤ **Cross-Country Process**

Should the approach be grounded on country-level evaluations building towards a synthesis report or on a more generic/thematic/global approach? The Paris Declaration calls for a cross-country monitoring and evaluation process. The monitoring process will collect data at the country-level and then aggregate-up across countries. The proposed evaluation framework approach could also serve this dual purpose - evaluations would need to be of practical value to country offices, partner governments and civil society while the framework would also be designed to make it possible to synthesize country findings as the basis for common conclusions and lessons learned.

➤ **Evaluating Outcomes and Impacts**

The purpose of the evaluation process is stated in the Paris Declaration as, “to provide a more comprehensive understanding of how increased aid effectiveness contributes to meeting development objectives”. There will likely be demand, from both donors and partners, for findings at the outcome and impact level. However, the higher up the results chain that the evaluation attempts analysis, the more difficult it will be to demonstrate causal links and to attribute effects. What is the optimal timing for delivering useful evaluation outputs, noting the DAC High Level Forum in 2008, and what is the balance to be struck between a formative process focussing more on lesson-learning and a summative process focussing more on mutual accountability?

➤ **Range of Evaluation Inputs and Outputs**

In carrying out this initiative, ways will need to be devised that will allow the incorporation of evidence from a wide range of inputs, including country and cross-country evaluations and reviews undertaken by bilaterals, multilaterals and partner countries. We also envisage a stream of outputs moving from a literature review to the framework to country and synthesis reports.

The Evaluation Network calls for consultant bids to undertake a literature review and prepare a concept paper exploring and defining possible ways forward, methodologies and next steps for developing, managing and implementing an evaluation process in view of these questions and challenges.

Activities

The consultancy contract will cover the following activities:

1. Undertake a focussed literature review on (a) emerging methodological lessons learnt in assessing and evaluating the principles of aid effectiveness, as contained in the Paris Declaration; and (b) substantive lessons learnt from past evaluations on aid effectiveness of particular relevance to this work.
2. Undertake telephone interviews with approximately 20 key informants, including representatives of partner countries, the DAC Evaluation Network and the DAC Working Party on Aid Effectiveness. The list of informants will be agreed with the Secretariat.
3. Undertake background conceptual thinking for either a logical framework, or causality tree or results chain approach for showing how the commitments contained in the Paris Declaration are intended to improve development outcomes/impacts;
4. Assess the capacity of the emerging Paris Declaration monitoring framework, including the Questionnaire and Peer Reviews, to capture information needed for the evaluation process.

5. Assess the evaluability, at the national and international levels, of the commitments made in the Paris Declaration.
6. Develop proposals for the design of the evaluation process/framework, including possible modalities for management and implementation.
7. Participate, and when necessary give presentations on work in progress, at up to three meetings at the OECD in Paris. These will likely include a one-day meeting with a small core group of 4-5 Evaluation Network members at an early stage of the contract and also a presentation at the DAC Evaluation Network meeting on either 30 or 31 March 2006. Feedback and input from these meetings should be reflected in the report.
8. Additional/emerging areas as agreed with the Secretariat.

Outputs

1. A focussed literature review (approximately 20 pages) outlining (a) emerging methodological lessons learnt in assessing and evaluating the principles of aid effectiveness, as contained in the Paris Declaration; and (b) substantive lessons learnt from past evaluations on aid effectiveness of particular relevance to this work.
2. A report (not more than 35 pages) on options for developing and implementing the evaluation process. As an outline/guiding structure, this should include the following sections:
 - Executive summary. The executive summary should include a clear presentation of the key options and recommendations for the design, management and implementation of the evaluation process that are being submitted for review, discussion and decision by the Evaluation Network.
 - An initial outline/overview (not more than 2 pages) of either a logical framework, or causality tree or results chain approach for showing how the commitments contained in the Paris Declaration are intended to improve development outcomes/impacts;
 - A review (not more than 8 pages) of the capacity of the emerging Paris Declaration monitoring framework, including the Questionnaire and Peer Review processes, to capture information needed for the evaluation process;
 - An assessment (not more than 5 pages) of the evaluability of the principles and commitments contained in the Paris Declaration. This should include an outline of the key evaluability challenges and options to address these challenges.

The core of the report will be a chapter with a number of alternative proposals for the overall form, focus, design and management of the evaluation process, including an analytic consideration of the proposed evaluation framework approach. This section should address the main questions and challenges - including those outlined in the above section on 'Questions and Challenges' - on evaluation design, management and implementation, and suggest options and recommendations for the way forward.