



# **Mid-term Evaluation of the 2000-06 Structural Fund Programmes**

*with executive summary*

***IQ-Net Thematic Paper 11(2)***

Philip Raines and Sandra Taylor



***IQ-Net  
Improving the Quality of Structural Fund  
Programming through Exchange of  
Experience***

**European Policies Research Centre**  
University of Strathclyde  
Graham Hills Building  
40 George Street  
Glasgow G1 1QE  
United Kingdom

Tel: +44-141-548 3339/3955

Fax: +44-141-548 4898

E-mail: [john.bachtler@strath.ac.uk](mailto:john.bachtler@strath.ac.uk)  
[sandra.taylor@strath.ac.uk](mailto:sandra.taylor@strath.ac.uk)

June 2002



## ***PREFACE***

The research for the following paper was undertaken in preparation for the sixth meeting of Phase II of the Objective 2 exchange of experience network IQ-Net, which took place in Luleå, Norrbotten, Sweden on 5-7 June 2002.

This paper is a product of desk research and fieldwork visits among national and regional authorities in Member States (notably in member regions of the IQ-Net consortium) as well as commission services in Spring 2002. The field research team comprised:

- |                                                 |                                        |
|-------------------------------------------------|----------------------------------------|
| ▪ <i>Professor John Bachtler (EC)</i>           | ▪ <i>Ruth Downes (Austria, Sweden)</i> |
| ▪ <i>Professor Henrik Halkier (Denmark)</i>     | ▪ <i>François Josserand (France)</i>   |
| ▪ <i>Laura Polverari (Italy)</i>                | ▪ <i>Philip Raines (UK)</i>            |
| ▪ <i>Mary Louise Rooney (Flanders, Germany)</i> | ▪ <i>Sandra Taylor (Spain)</i>         |

Thanks are due to everyone who participated in the research. The European Policies Research Centre also gratefully acknowledges the financial support provided by participating regions and by the European Commission (through European Regional Development Fund co-financing for regional subscriptions) in funding the production of this report and the organisation of the Luleå meeting.

For further information about IQ-Net, and access to the full series of Thematic Papers, visit the IQ-Net website: <http://www.eprc.strath.ac.uk/iqnet/>.



**Structural Fund  
Programming Update:  
December 2001 – April 2002**

**TABLE OF CONTENTS**

|                                                                        |               |
|------------------------------------------------------------------------|---------------|
| <b>Executive Summary .....</b>                                         | <b>1</b>      |
| Introduction.....                                                      | 1             |
| Context of the 2000-06 mid-term evaluations.....                       | 2             |
| Organising the mid-term evaluations.....                               | 2             |
| Designing and undertaking the mid-term evaluations.....                | 4             |
| Using the findings .....                                               | 7             |
| <br><b>1. Introduction.....</b>                                        | <br><b>9</b>  |
| <br><b>2. Context of the 2000-06 Mid-Term Evaluations .....</b>        | <br><b>11</b> |
| <br><b>3. Organising the Mid-Term Evaluations .....</b>                | <br><b>14</b> |
| 3.1 Configuration of the proposed evaluations .....                    | 14            |
| 3.2 Dividing the different roles in the mid-term evaluations.....      | 21            |
| 3.3 The timetables for evaluation .....                                | 29            |
| 3.4 Evaluation budgets.....                                            | 30            |
| 3.5 Selecting evaluators .....                                         | 30            |
| <br><b>4. Designing and Undertaking the Mid-Term Evaluations .....</b> | <br><b>32</b> |
| 4.1 Overall approach: defining evaluation methods.....                 | 32            |
| 4.2 Context: how the mid-term evaluations are conceptualised.....      | 35            |
| 4.3 Style: choices about how to undertake evaluation.....              | 41            |
| 4.4 Scope of the evaluations: the aims to be addressed.....            | 43            |
| 4.5 Tools for evaluation .....                                         | 54            |
| <br><b>5. Use of Findings .....</b>                                    | <br><b>60</b> |
| 5.1 What are the outputs of the evaluation process?.....               | 60            |
| 5.2 How evaluation findings can be disseminated .....                  | 60            |
| 5.3 How the evaluation findings can be used.....                       | 62            |
| <br><b>6. Conclusions .....</b>                                        | <br><b>70</b> |



# Mid-Term Evaluation of the 2000-06 Structural Fund Programmes

## EXECUTIVE SUMMARY

### Introduction

As the IQ-Net partners are currently planning for the mid-term evaluations of the 2000-06 programmes, it is timely to discuss the new challenges arising from the current round of evaluations. Preparations are already well underway as partners work towards the deadline of the end of 2003 to complete and submit their mid-term studies to the EC. The evaluations have much at stake and entail a complex series of required tasks. They also present an opportunity for programmes to take stock, and make decisions on how best to get the most benefits from the remaining years of Structural Funding in the current round.

It is important that the evaluations are effectively managed. This paper is intended to contribute to partner preparations by reviewing what has been done to date, identifying the main remaining challenges and discussing the options for maximising the value of the evaluations. To do this, the paper considers the evaluation process under four key headings, each with their own set of important questions for partners and stakeholders:

➤ **The context of evaluation:**

What are the main challenges in the current round of mid-term evaluations? What are the major differences with previous rounds?

➤ **The organisation of evaluation:**

How will the evaluations be configured?

What organisational structures will be put in place to manage the evaluations?

What are the respective roles and responsibilities of the Commission, the Member States and actors within individual programmes?

➤ **The design and undertaking of evaluation:**

What choices will affect the design of individual evaluations?

What are the different options for undertaking the evaluations and what are the costs and benefits associated with each?

What tools and techniques are available to evaluators?

➤ **The use of the findings**

What is the value of the different types of evaluation output?

What plans are being made for dissemination?

How will the evaluations be used in allocating the performance reserve and improving individual programmes?

### Context of the 2000-06 mid-term evaluations

Perhaps the most important question with which to begin is: **why do the mid-term evaluations matter?** The answer differs for the EC, Member States and individual programmes, as each set of actors has different objectives needing to be fulfilled, potentially more so than in previous mid-term evaluations. At **the level of the EC**, the Commission requires the evaluations for monitoring the efficiency and effectiveness of programming, the analysis required for the Third Cohesion Report, and making any operational adjustments to the programmes. For their part, **the Member States** will be looking to the evaluations to inform the process of allocating the performance reserve, and to feed into national policy debates on the future of the Funds. Lastly, **the individual programmes** will want the evaluations to advise on how to adjust the programmes, ensure effective and efficient programme management, meet the regulatory requirements, and help the programmes to have durable legacies.

Despite the complex range of interests, **partners are in a better position to undertake successful evaluations**. Whereas in the past, evaluations have sometimes been greeted with reluctance (and seen as Commission-driven exercises), this time, the overall attitude is predominantly positive and engaged. Actors at all levels are not only more familiar with the operational requirements for conducting the evaluation, but have access to larger pools of evaluator skills and experience. Indeed, as a whole, programmes are more easily evaluated this time, given the improvements that have been made to monitoring and data storage systems. Lastly, there are powerful incentives to encourage programmes to pursue robust, professional evaluations, not least their role in informing the allocation of the performance reserve.

### Organising the mid-term evaluations

Member States have considerable flexibility in organising the mid-term evaluations, but **two main options** have emerged. Most have chosen to undertake a separate evaluation for each programme. This approach has the advantages of tailoring evaluations to programme needs and encouraging greater engagement of individual programmes. However, a few Member States have decided on a single evaluation covering multiple programmes (Flanders, Sweden, Denmark), which can have benefits in financial as well as organisational terms. To increase the value of the studies overall, some Member States are also planning meta-evaluations to bring together lessons drawn from all the mid-term evaluations. Links to other evaluations running in parallel (of both Structural Fund programmes and domestic policies) are also being considered by some partner regions.

Another aspect of organisation is whether to give **special attention to specific issues** within the evaluations. Options for individual thematic or cross-cutting evaluation exist, giving programmes the potential to assess any new strategic or operational issues. Consequently, some regions are considering separate evaluations for particular themes while others are identifying certain thematic issues to be highlighted in the programme evaluations. The horizontal themes are particularly important in this regard, and have been emphasised by the



Commission, although most programmes appear to be favouring their integration into mainstream evaluation.

In terms of **roles and responsibilities**, a clear designation of tasks between the EC, the Member States and the individual programmes is emerging.

- Hence, at EC level, the **Commission** has multiple roles. It has a regulatory requirement to be consulted by Member States on a range of areas including evaluation terms of reference, budgets, tendering procedures, methods and the quality of reports. Commission services have also been an important source of advice, notably through key guidance papers (such as Working Paper 8) which have generally been regarded as highly constructive.
- The **national authorities** have had a notable role in most countries in two areas. First, many have issued their own general guidance, based on the Commission's requirements but reflecting national priorities. Second, they have been involved in capacity building with programme managers through the provision of technical advice and sources of methodological and organisational good practice.
- Lastly, the lead organisation in the mid-term evaluations in most countries is the **programme level**. Their tasks include: drawing up the terms of reference; managing the selection procedure; working with the evaluator in defining methodologies and data sources; ensuring the timely delivery of quality outputs; and following up on recommendations.

The involvement of wider partnerships and stakeholders, primarily through **steering groups**, has been a common approach in the past and is recommended by the Commission for the current round. These groups are usually a sub-set of the Monitoring Committee and have differing levels of involvement in the overall process in different countries. Their contribution to evaluation quality can come through effectively steering a highly time-constrained process, and by providing bridges between the evaluation and the wider programme and between those centrally involved and the wider partnership.

An indicative **timetable** for the mid-term evaluations has been set out in the Commission guidance with a final delivery deadline of 31 December 2003. Planning has begun in every partner region, although progress has varied. Nevertheless, by the time the mid-term evaluations are undertaken, enough programme activity will have taken place for most evaluations to be worthwhile. Further, the longer programming period means that there will generally be enough remaining time for recommendations to have a real impact.

In terms of **budget**, the Commission has recommended that the average cost of the evaluations should exceed 0.1 percent of the total annual cost of the programme (the average of the last period), except for very large programmes. Such levels are considered necessary to fulfil the numerous tasks and obtain quality evaluation results. While an overview of budgets is not possible at this stage, initial feedback indicates that budgets are being set higher for this

evaluation round and that the Commission appears satisfied with the financial proposals to date.

In most regions, the **tendering process** for the selection of evaluators has only just begun. Evaluators will usually be chosen through an open call procedure and, in the light of the contract threshold limits for several evaluations, through the EU's Official Journal. The range of applicants will inevitably be restricted by practical considerations such as language and geographical proximity. The availability of skilled evaluators for the mid-terms may be difficult given the significant number of parallel evaluation exercises which are taking place. However, the ongoing process of capacity building in this area has enlarged the evaluator pool over time and most Member States have no serious concerns. The issue of whether evaluators can meet the additional demands involved in this round of evaluations is of more concern, particularly for the treatment of the horizontal themes. Possible solutions include the formation of evaluator partnerships or the further development of in-house expertise by the selected evaluators.

### **Designing and undertaking the mid-term evaluations**

As all the mid-term evaluations are being shaped by the same regulatory requirements and DG Regio guidance, they are likely to have a core of common elements. It must be emphasised, though, that each of the studies will be unique: these programme evaluations will not be standard, routine exercises. Given the stage which most programmes have reached, it is not yet possible to discuss the actual tools and methods which will in practice be used in the evaluations. Nonetheless, there is a range of issues around the definition and execution of the evaluations which can be discussed. The most important of these are:

- i. how and when the methodological choices will be made for the evaluations, and by whom, illustrating the broad involvement in this process;
- ii. whether and how the evaluations fit into a wider context, contrasting programmes which undertake relatively self-contained, periodic studies according to the EC timetable with those which have a more ongoing vision of evaluation;
- iii. favoured styles and processes of evaluation, including whether the evaluator is seen as an expert providing a simple report, for instance, or as a facilitator of a wider learning process;
- iv. the anticipated content of the evaluations; and
- v. the evaluation tools employed.

Each of these issues is addressed below, opening out some of the choices facing those about to commission and undertake evaluations. A recurring theme in this section is how the overall conception of evaluation is likely to influence outcomes in terms of scope and organisation. The key distinction lies in the balance between viewing evaluation as a reporting function, satisfying aims to do with accountability, justification and transparency, and seeing it as a learning tool (and, indeed, as a learning process) undertaken to

serve the quality, efficiency and effectiveness objectives of the programme itself.

First, **decision-making about methodologies** has multiple stages and involves many actors, either directly or indirectly. Indeed, the evaluator is only one player among many, each of whom has a particular contribution to make in defining the scope and conduct of the evaluation. While in many cases the evaluator will be expected to propose an evaluation approach and methodology, the expertise of the evaluator will best shine when it complements the understanding and practical knowledge of programme managers and partners. Detailed prior thinking could be of value here, particularly in alerting evaluators to any data problems. Knowing this at the bidding stage and bringing it to the attention of potential evaluators saves wasted time by ensuring bidders are working on the basis of meaningful assumptions and only propose feasible methods.

Second, one of the most interesting recent debates on evaluation has been about **how individual evaluation studies are conceptualised** and, in particular, the potential merits of moving from treating evaluation as self-contained, periodic exercises to seeing it as an ongoing discipline. An ongoing or continuous approach to evaluation would potentially span the lifetime of a programme. While the distinction between programmes which explicitly embrace ongoing evaluation and those which do not may in practice be small, there is scope for combining the virtues of both approaches more actively than hitherto.

Third, over recent years, Structural Fund evaluation has more widely been interpreted as **a learning opportunity** rather than simply a reporting exercise. Part of the rationale for this is that Structural Fund programmes are designed and delivered on a partnership basis, relying on the contribution and expertise of a wide vertical and horizontal partnership. This poses challenges in terms of how best to facilitate a learning process and ensure that evaluation findings reach those for whom they are most relevant, and lead to constructive change. Programmes face crucial choices here. At one extreme is 'expert-oriented' evaluation, where evaluators are contracted to scrutinise a programme, its activities and systems in a detached way and deliver an appropriate range of conclusions and recommendations. At the other end of the continuum is 'participative' evaluation, where the key emphasis is placed on the expertise and knowledge of programme actors rather than the evaluator.

Fourth, decisions on **the content of evaluation** will have a major role in determining the methodology to be used. The overall aim of the mid-term evaluations, as summarised in Working Paper 8, is *'to assess the initial results of the various forms of assistance and to make recommendations for any changes needed to ensure that they achieve their objectives'*. This disaggregates into two questions.

- Are programme strategies still relevant, or are adjustments now required?
- How well are programmes performing in terms of effectiveness and efficiency?

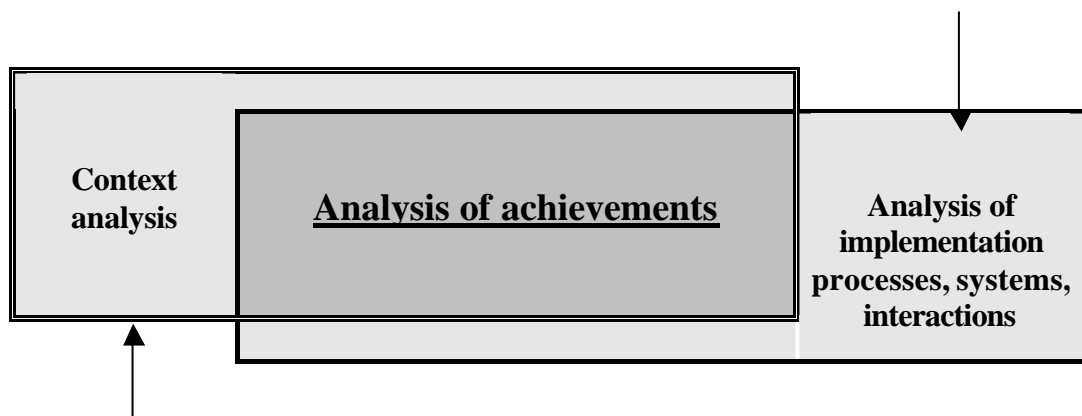
The two questions can be tackled through three forms of analysis:

- i. achievement analysis (ie. determining the effects of the programme);

- ii. context analysis (ie. identifying the wider economic trends and the institutional and policy changes of the programme); and
- iii. process and systems analysis (ie. examining the way in which the programme is delivered).

These three dimensions of analysis and the relationship between them can be illustrated schematically.

**Overall Aim 2: Understanding patterns of effectiveness and efficiency**



**Overall Aim 1: Verifying continuing strategic relevance and enriching the understanding of achievements**

The diagram emphasises the idea that the **analysis of achievements** is central to the evaluation studies, even at the mid-term stage when relatively little progress may have been made towards final economic impacts. The primary focus and core of most of the mid-term evaluations will be to quantify programme achievements to date. This arguably needs to be addressed more seriously now than in the past, particularly in the context of using the evaluations in allocating the performance reserve. However, it may not be realistic or appropriate for impacts to be assessed directly and in detail. A pragmatic compromise, and one which is methodologically feasible, is to reflect on the realism and likelihood of attaining the impact targets.

Analysing contextual change or processes and systems are not ends in themselves but complement and inform interpretation of programme achievements. Consequently, **context analysis** is undertaken: to provide insights into how a programme has performed by placing achievements against wider trends; and to verify the continuing relevance of a programme in terms of its economic and policy and institutional contexts. Similarly, **analysis of the programme's processes and systems** benefits from being undertaken as a complementary dimension, which informs and illuminates understanding of the programme's economic development achievements. For process evaluation to be seen as useful, a Structural Fund programme needs to be configured in such a way that it is seen by its stakeholders as a clear entity, with distinctive policy objectives and delivery structures – in other words, as 'something to evaluate'. In this context, using a categorisation familiar to IQ-Net partners, 'differentiated' Structural Fund systems may lend themselves more readily to process-related evaluation than 'subsumed' ones.

Lastly, in undertaking their analyses, evaluators can make use of **a range of tools and techniques**: analysis of monitoring data; project surveys; stakeholder interviews; macroeconomic modelling; and context analysis. Overall, there is widespread optimism that the improvements to monitoring systems should benefit evaluation. However, the EC is anticipating that monitoring data will be supplemented by stakeholder interviews and sample surveys of projects.

### Using the findings

As in previous mid-term evaluations, the key outputs of the evaluation exercises are reports, delivered at the end of the study. These will contain the findings and recommendations on the issues specified in the terms of reference. However, other, more interactive outputs from evaluation are possible, some of which may be delivered before the study is completed. **Dissemination** will thus depend on the balance between whether the findings are bound up in a single report and whether the results will be fed back to programme participants during the course of the evaluation (and beyond).

Overall, the two principal uses of the evaluation findings will be in allocating the performance reserve and recommending improvements to the programmes. First, **distribution of the performance reserve** and the mid-term evaluations are closely linked, though separate exercises. Evaluation results will be used in deciding how well programmes are performing on certain of their performance reserve criteria. Nevertheless, while the Commission has given clear attention to the issue in the mid-term evaluation guidance and consultation, the importance of the performance reserve in the evaluations has been mixed among Member States.

Second, and perhaps most importantly, the results of mid-term evaluations will be used in **refining individual programmes**. Programme managers will be looking to the evaluations to produce:

- an independent assessment of their activities as an external confirmation of performance;
- a deeper understanding of recognised problems and the identification of hidden or future problems facing the programme;
- inputs of new ideas into continuing discussion over how such problems could be solved;
- empirical backing to the case for certain changes to programmes, especially virement between priorities and/or changes to the scope of priorities and measures; and
- analysis which is beyond the scope of Annual Implementation Reports and monitoring, such as confirming the direction of evolution of relevant baseline indicators.

Finally, to enhance this use of the evaluation results, some programmes are aiming to compare their evaluations with those of other programmes. They plan to **benchmark their performance** by comparing the results of their evaluations in several key areas, such as the horizontal themes.



# Mid-Term Evaluation of the 2000-06 Structural Fund Programmes

## 1. INTRODUCTION

IQ-Net last addressed evaluation as a central theme in 1997, looking at experience with the interim evaluations of the 1994-96 Objective 2 programmes and the 1995-99 new Member State programmes.<sup>1</sup> Since then, considerable additional experience of evaluation has been accumulated, including a further phase of mid-term evaluations for the 1997-99 Objective 2 programmes, *ex ante* evaluations of all 2000-06 programmes, and, in some cases, *ex post* evaluations of the 1994-99 programmes (including the Objective 2 1994-96 and 1997-99 programmes) and *ad hoc* thematic studies. Preparations are now underway for the mid-term evaluations of the 2000-06 Structural Fund programmes.

Included as a requirement in the Structural Fund regulations,<sup>2</sup> mid-term evaluation is a more formalised and integral part of Structural Fund programming than in the past.<sup>3</sup> For the first time, it will be a universal exercise, undertaken to a strict deadline and having to meet specified quality thresholds. Because of the length of the current programming cycle, it will be able to feed into the existing programmes more effectively than before. A further innovation is that selected outputs of the exercise are tied in with the allocation of the performance reserve.

The mid-term evaluation also presents an opportunity for programmes to take stock, and generate information on programme performance, enabling adjustments to be made that can maximise benefits from the remaining years of Structural Funding. In this context, the following paper is intended to contribute to programme efforts to find the optimal means of maximising their evaluations. It has been drafted at a stage when preparations are still underway, most evaluations having not yet begun, (in most cases, terms of reference have not been finalised nor contracts advertised). It discusses critical issues and challenges to feed into and inform the forthcoming evaluation process. Specifically, the paper provides the following elements:

- **an overview of the preparations for mid-term evaluation:** how evaluations will be configured, how IQ-Net partners are preparing for them, what they see as the main challenges and the lessons of the previous interim evaluations; and

---

<sup>1</sup> Bachtler J, Michie R and Downes R (1997) *The Interim Evaluation of Objective 2 Programmes*, IQ-Net Thematic Paper 2(1), European Policies Research Centre, University of Strathclyde, Glasgow. <http://www.eprc.strath.ac.uk/iqnet/iq-net/downloads/reports/2.1InterimEvaluation.pdf>.

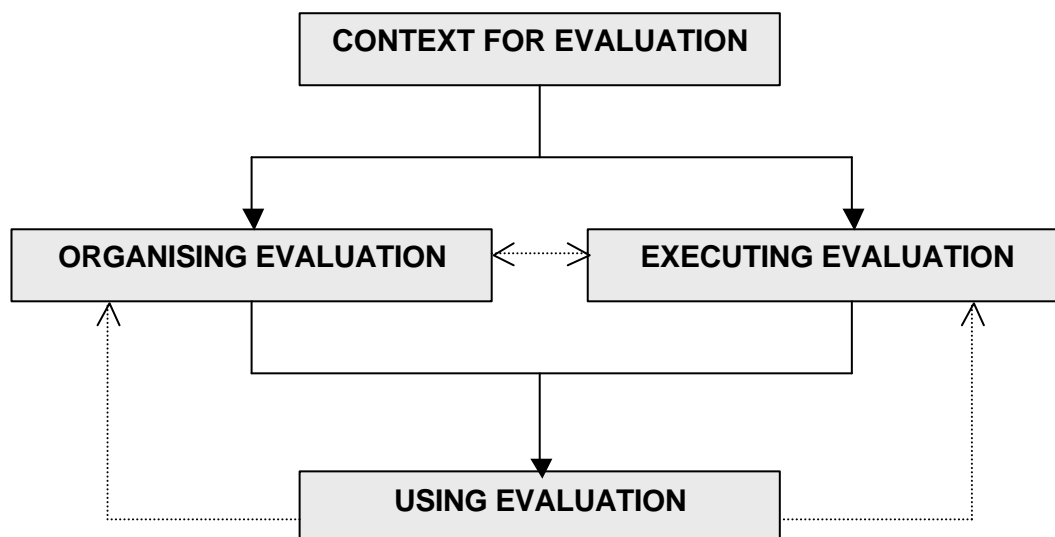
<sup>2</sup> CEC (1999) Council Regulation (EC) No 1260/99 of June 1999 laying down general provisions on the Structural Funds, OJEC L161/1 26.6.1999.

<sup>3</sup> In 1994-99, the obligation was added to SPDs after the regulations were finalised, as part of their standard clauses.

- **useful approaches to addressing the challenges of the current mid-term evaluations:** what current ideas are being explored in the regions, and on what other evaluation experience will evaluators and programme actors draw, to address the new requirements in the current mid-term evaluations?

The paper is broadly structured around four key groups of issues. They can be loosely thought of as a progression through the different stages of the evaluation life-cycle. Each stage involves distinctive opportunities and require specific challenges to be met, as described in the diagram below. At the same time, there is extensive overlap between the different elements: the organisation and execution of evaluation are processes that not only take place in parallel, but are mutually influential. In addition, the use of evaluation findings is not confined to the period after studies have been completed: often, learning processes are already underway even while an evaluation is being undertaken.

Figure 1.1: Structure of paper



Each 'stage' raises key issues which must be addressed by stakeholders and evaluators:

- **the context of evaluation:** the main challenges in the current round of mid-term evaluations, particularly the differences with previous rounds;
- **the organisation of evaluation:** the ways in which the evaluations are carried out, the organisational structures put in place and the roles and responsibilities of the Commission, the Member States and actors within individual programmes;
- **the design and execution of evaluation:** the methodological requirements and opportunities that affect the content of individual evaluations and the ways in which they are executed; and
- **the uses of evaluation:** the nature of evaluation outputs, the plans for disseminating results, and the use of the results in allocating the performance reserve and providing insights into how programming can be improved.



A central tension which the paper will explore is between evaluation as an accountability and reporting mechanism, and as a learning tool. It is argued that changes in the regulatory environment may be increasing pressures for the former to predominate, even though the latter arguably has the potential to deliver greater benefits at programme level.

## 2. CONTEXT OF THE 2000-06 MID-TERM EVALUATIONS

The 1999 regulations<sup>4</sup> brought a clear shift in the approach to monitoring and evaluation within the Structural Funds. While highlighting the importance of evaluation as part of the programming cycle, the regulations made clear that the responsibility for organising the mid-term evaluations lay with the Managing Authorities. As part of the new programming round, the clearest expression of the purpose of the current round of mid-term evaluations came in the DG Regio Working Paper<sup>5</sup> on mid-term evaluation: “*the mid term evaluation is not an end in itself but a means to improve the quality and relevance of programming.*” It should be regarded as part of a chain of evaluations – beginning with the *ex-ante* evaluations of the 2000-06 programmes, continuing with the planned follow-up mid-programme evaluations in 2005 and finishing with the subsequent *ex-post* evaluation – and has been designed to ensure that the current programmes can adjust to changing conditions during the course of the programming period.

Why do the mid-term evaluations matter? In fact, different actors have different objectives which need to be fulfilled. Indeed, the requirements are arguably so extensive and diverse that it may prove difficult for evaluators successfully to meet all of them in every study. Balancing all these interests will be one of the main challenges associated with the evaluations this time around. In general, there are three sets of separate interests that need to be considered for the evaluation, summarised in the table below: the EC; the Member States; and individual programmes.

Table 2.1: *Interests in the evaluations at different levels*

|                           |                                                    |                |
|---------------------------|----------------------------------------------------|----------------|
| <b>EC level</b>           | Informational use                                  | Accountability |
|                           | Analytical use                                     |                |
|                           | Operational use                                    |                |
| <b>Member State level</b> | Performance reserve                                |                |
|                           | Feeding into debates of future of Structural Funds |                |
| <b>Programme level</b>    | Adjusting programmes                               |                |
|                           | Programme management                               |                |
|                           | Regulatory requirements                            |                |
|                           | Durable legacies                                   |                |

<sup>4</sup> CEC (1999) *op. cit.*

<sup>5</sup> DG Regio (2000) *Working Paper 8: The Mid Term Evaluation of Structural Fund Interventions*, Commission of the European Communities, Brussels, 5 December 2000.

At European level, the EC has specific responsibilities and objectives with respect to the evaluations, in particular for ensuring the quality of evaluation reports. In addition, it has specific uses for the information which will be generated by the evaluations.

- **Informational use.** Under the more decentralised management system for the Structural Funds, the EC is heavily dependent on 'secondary information sources', including annual reports/reviews, as well as the evaluations. These are needed to monitor the efficiency and effectiveness of programming and find out whether the promised targets are being achieved.
- **Analytical use.** The evaluations are needed to feed into the analytical sections of the Third Cohesion Report (along with the global *ex post* evaluations of the 1994-99 period which the EC has commissioned).
- **Operational use.** If there are any operational adjustments needed, the evaluations will provide the evidence supporting any necessary re-programming (such as virement between priorities or measures) and be part of the evidence presented to the EC for the allocation of the performance reserve between and within programmes.

At the same time, the Member States also have a mix of obligations and ambitions for the mid-term evaluations. They look to the evaluations to:

- **inform the process of allocating the performance reserve:** especially where this is being awarded on a competitive basis between different programmes; and
- **generate information for national policy debates on the future of the Funds:** one of the preoccupations here is in ensuring that the results can be aggregated (eg. France is concerned to compile an overview of the achievements of Structural Fund programmes so it can assess the added value of these strategies).

Lastly, at the level of the individual programmes, there are several key motivations surrounding the evaluations:

- informing and enabling the process of **adjusting the programmes**, including virement between priorities and measures, based on a changing economic context and progress to date in terms of implementation;
- ensuring effective and efficient **programme management** and addressing any identified weaknesses in structures and practices;
- meeting the **regulatory requirements**; and
- helping the programmes to have **durable legacies**, especially as this may be the last time many regions will have programmes of this scale.

With multiple interests to be addressed, there has been considerable investment in organising the evaluations effectively. The Commission has played a constructive role by providing clear, consistent guidance to the Member States and individual programmes, participating on Steering Groups and providing advice through seminars and *ad hoc* consultations. Given that the evaluation process has a stronger grounding than ever before, there is considerable optimism regarding the implementation of the evaluation

exercise and the value of the results in meeting the needs of different participants.

Nevertheless, despite the well-founded start to the evaluation exercise, numerous challenges remain to be met by Member States and individual programmes. Many of these are associated with the various new features that distinguish this round of evaluations from previous ones. Overall, the current evaluations are anticipated to be a much more wide-ranging and demanding exercise. DG Regio's Evaluation Unit has highlighted the following areas as distinctive features of the current round:

- **regulatory requirements:** the regulations governing the evaluations contain more detailed and complex requirements;
- **implications of the outcome:** the evaluations are a major component in determining the allocation of the performance reserve;
- **scale of resources being dedicated to the exercise:** the budgets for the evaluations are larger than before; and
- **comprehensiveness:** in previous mid-term evaluation rounds, particularly for Objective 2 programmes, not all Member States and/or regions undertook the exercise.

Partners are in a better position to undertake successful evaluations than ever before. Indeed, an examination of the reported weaknesses of previous evaluations shows that many are being addressed in the current process.

*Table 2.2: Key changes in the current round of evaluations*

| Perceived weaknesses of previous evaluations                                                | Potential strengths of the current evaluations                                                   |
|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Benefits of evaluation to learning not fully acknowledged.                                  | → Widespread acceptance of the multiple values of evaluation as a part of programming.           |
| Weaknesses in the evaluation capacity of some regions/Member States.                        | → Embedding of evaluation culture and skills across the EU.                                      |
| Evaluability of the programmes                                                              | → Clearer aims, improved targets and better monitoring systems and data                          |
| Lack of clearly-defined and universal incentives for fully engaging in evaluation exercise. | → Strong incentives to manage the evaluations effectively and maximise the use of their results. |

The first change has been in the general attitude towards evaluation. Whereas, in the past, evaluations were sometimes greeted with reluctance (and perceived as Commission-driven exercises), this time the overall attitude is predominantly positive and engaged. In many cases, regions and Member States have been developing their experience of evaluation, often in a Structural Fund context. As a result, there are generally high expectations from this round of evaluations, with individuals at all levels wanting them to be genuinely useful and contribute to effective programming. This is reflected in the more universal and stronger link between evaluation and management, in which partners more readily appreciate the learning component of evaluation. Few now regard evaluation as simply an obligation with little or no

day-to-day relevance to programming (although the requirements of evaluation will still be onerous to many).

The second change concerns the capacity of the partners to carry out evaluations. Boosted by the accumulated experience of Structural Fund evaluation, actors are not only more familiar with the operational requirements for conducting evaluations, but have access to larger pools of evaluator skills. There is less need for partners to 'gear up' this time, particularly as the programmes have been designed with evaluation in mind.

The third change is the greater 'evaluability' of the programmes. Aims are more easily quantified, targets are more clearly and consistently defined, and data gathering has been greatly enhanced, with a step change in the quality of monitoring and the resulting data available in many programmes. In many cases, evaluators should find it easier to access and exploit programme data.

Lastly, powerful incentives have encouraged programmes to pursue robust, professional evaluations. Externally, the requirement on the evaluations to contribute to allocating the performance reserve has meant that the evaluation process has a clearer timetable than hitherto. Internally, the longer programming period makes the studies feel more purposeful. In most contexts there has been sufficient progress to make evaluation worthwhile by the time the studies take place and there will be sufficient time remaining in the programming period for recommendations, if applied, to make genuine differences. For some programmes, it will be the first time that the mid-term evaluations will be able to feed fully into the programmes as they continue.

In many respects, the mid-term evaluations may seem highly prescribed activities. As seen later, there is considerable continuity in the approaches to some aspects of evaluation. Moreover, many of the tasks of the evaluation have been set down in detail. Yet the evaluations will not be standard or routine exercises. Each evaluation will face its own portfolio of problems and opportunities and will need to develop its own approach. While many of the tools for undertaking the evaluations will be common to all programmes, it is important to stress that each programme will have its own set of options for different components of the evaluation. Along with the pressure to ensure that the evaluations are high-quality exercises, this will be the principal challenge facing the partners. The following sections of the paper discuss the range of options that apply to different phases of evaluation and the factors shaping these choices.

### **3. ORGANISING THE MID-TERM EVALUATIONS**

#### **3.1 Configuration of the proposed evaluations**

Member States have considerable flexibility in deciding how to organise the mid-term evaluations. Since the Commission's main source of advice on the evaluations – Working Paper 8<sup>6</sup> – does not specify how they should be configured, there is little specific guidance on the issue. The complexity of the

---

<sup>6</sup> DG Regio (2000) *op. cit.*

current evaluations, and the fact that some programmes did not undertake mid-term evaluations previously, means that there are challenging tasks in this round.

Member States have choices over how to configure their evaluations. Two main options have been employed.

- **Separate evaluations for each programme**, with the relevant Managing Authority being ultimately responsible. This is the case in Austria, France, Germany, Italy, Spain and the UK.
- **A single evaluation covering multiple programmes**, led by a national or territorial authority. This applies in Flanders (covering four Objective 2 programmes and an Urban programme) and Sweden (where the Objective 1 and 2 programmes will be covered by two separate evaluations).

The latter category also includes Denmark, although the Danish mid-term evaluation is distinctive. While the country only operates a single Objective 2 SPD, separate socio-economic analyses and thematic sub-studies will be carried out for the individual regional sub-programmes.

The two main approaches have distinctive strengths and weaknesses, as outlined in the table below.

*Table 3.1: Strengths and weaknesses of different ways of organising evaluation*

|                            | Strengths                                                                                                                                                                                                                         | Weaknesses                                                                                                                                                                                                                                                                                              |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Multiple programmes</b> | <ul style="list-style-type: none"> <li>• Economies of scale reduce the aggregate cost of evaluation</li> <li>• Consistency and comparability of results across several programmes</li> </ul>                                      | <ul style="list-style-type: none"> <li>• Risk of imposing ‘one size fits all’ approach to evaluations</li> <li>• Diminished engagement of individual programmes in evaluations and risk of insufficient time given to each programme</li> </ul>                                                         |
| <b>Single programme</b>    | <ul style="list-style-type: none"> <li>• More effective ‘customisation’ of evaluation to individual programmes</li> <li>• Stronger engagement of – and sense of ownership by – relevant programme actors in evaluation</li> </ul> | <ul style="list-style-type: none"> <li>• Higher costs incurred, not just financially, but also in organisational terms with respect to coordinating the evaluations</li> <li>• Risk of key aspects of evaluation not receiving sufficient attention because they may be perceived as ‘minor’</li> </ul> |

In summary, the advantages of a single evaluation covering multiple programmes are largely cost-driven. They are:

- **less costly in financial terms**, through the operation of economies of scale in centralising the design, commissioning and management of the evaluations; and
- **less costly in organisational terms**, through standardising the issues covered by the evaluations and the form in which the results are produced (particularly with respect to making informed decisions on how to allocate the performance reserve).

In contrast, the advantage of a ‘one programme, one evaluation’ approach is mainly the desire to maximise the impact that evaluations can have on individual programmes:

- **by tailoring evaluations to programme needs**, allowing programmes to introduce themes and issues of particular relevance to them; and
- **by encouraging greater engagement of individual programmes**, both in the design of the evaluation and, subsequently, in their response to the evaluation results.

The majority of Member States are opting for the ‘one programme, one evaluation’ approach, but often in conjunction with a central coordination role. Central government organisations have often given important guidance to individual programmes in their evaluation design by providing core terms of reference and acting as a source of advice.

**Where countries have decided on a single evaluation to cover multiple programmes, local circumstances have been a significant factor in the decision.** For example, the Flemish single evaluation of its Objective 2 programmes is partly motivated by the relatively small scale of the programmes. Similarly, in Sweden, cost, precedent and the desire to have a comparative framework for interpreting the evaluation results were key considerations in opting for a single evaluation for each Objective. However, Sweden recognises the importance of including programme-specific themes in its evaluation terms of reference.

**The ‘one evaluation, one programme’ approach presents new challenges for programmes.** For example, in Spain, this will be the first time that multi-fund studies have been undertaken (partly under pressure from the Commission). In previous programming rounds, separate evaluations were undertaken of the ERDF and ESF elements of Objective 2 programmes, whereas, on this occasion, both will be brought together. Some German Objective 1 regions will be following suit, after the Commission rejected requests to undertake Fund-specific evaluations.

Moreover, as part of the configuration of evaluations, **some Member States are planning meta-evaluations to bring together the lessons to be drawn from the individual mid-term evaluations.** This was exemplified by the Swedish meta-evaluation of the mid-term evaluations in the previous programming period, as described in the box below.<sup>7</sup> Similarly, in this round, DATAR intends to compare French programme performance and experiences through a synthesis of the individual evaluations, a major challenge given the large number of Objective 2 programmes. Others are considering the use of ongoing evaluation (see Section 4.2).

---

<sup>7</sup> The material for this case study is drawn from Bachtler J, Polverari L, Taylor S, Ashcroft B and Swales K (2000) *Methodologies Used in Evaluating the Effectiveness of European Structural Funds: A Comparative Review*, European Policies Research Centre and Fraser of Allander Institute, University of Strathclyde, Report to the Scottish Executive, Edinburgh.

### SWEDEN: MAXIMISING EVALUATION RESULTS WITH META-EVALUATION

On behalf of the Swedish Government, the Swedish Agency for Administrative Development (SAFAD) undertook a study to assess the interim evaluations of the 1994-99 Structural Fund programmes in terms of organisation, resources and processes. The research also had to cover the procurement of the evaluations and to include a comparative analysis, encompassing the experiences of other countries. The conclusions of the research, of which an executive summary is available in English, reported that the quality of the Swedish mid-term evaluations was relatively high, although there appeared to be a general lack of consideration of the role of domestic policies in the areas concerned and an absence of discussion on the methods or the basic data used in the evaluations.

#### 3.1.1 Coordination of the mid-term evaluations with other evaluations

**To increase the value of the mid-term evaluations, many regions are planning to link them to other evaluations running in parallel.** These can be either other Structural Funds evaluations – both other Objectives being currently evaluated and the *ex post* evaluations of the previous round - or those relating to domestic policies, particularly regional strategies, which may have influenced the drafting of SPDs. The table below summarises the advantages of making these linkages.

Table 3.2: Advantages of coordinating the mid-term evaluations with other evaluations

| Type of evaluation                | Advantages                                                                                                                                                                                                                                         |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Linkage with other SF evaluations | <b>Shared use of data:</b> eg. ESF data in Objective 2 areas<br><b>Common recommendations:</b> eg. monitoring systems                                                                                                                              |
| Linkage with domestic evaluations | <b>Shared tasks:</b> eg. socio-economic analyses for both Structural Fund and domestic regional strategy evaluations<br><b>Enhanced policy interactions:</b> eg. recommendations can suggest ways of improving co-financing and project generation |

Although the Commission is dissuading Member States from linking Structural Fund evaluations, there is scope in some programmes for joint activity. For example, the Austrian authorities have the opportunity to make use of common linkages between the Funds: the evaluation of the Austrian Objective 3 programme will separate out the ESF data collected for Objective 2 areas, allowing the mid-term evaluations of Objective 2 programmes to access the data where appropriate. In Sweden, the Objective 2 evaluation will be used in conjunction with other evaluations in progress. Echoing the previous meta-evaluation, NUTEK plan to draw together the main results of the Objective 2 mid-term evaluations and the Commission-managed *ex post* evaluation of the previous programmes (which involves a sample of Swedish Objective 2 regions).

With regard to domestic evaluations, consideration has been given in France to combining aspects of the mid-term evaluations for the Objective 2 programmes and the *Contrats de Plan Etat-Régions* (CPER). The rationale for a linkage is that the two sets of programme documents are already explicitly linked (SPDs are closely co-ordinated with the CPER strategies). The box below indicates the potential and implications of this overlap.

#### FRANCE:

##### RELATING DOMESTIC AND STRUCTURAL FUND EVALUATIONS

The French SPDs are strongly inter-linked with the key instrument to coordinate regional level expenditure, the CPER. Both run over the same period, are shaped around a common strategy and require a mid-term evaluation to be conducted. In addition, the main community, national and regional priorities inform both documents, and the CPER is the main source of co-finance for the Structural Fund programmes.

Given this close linkage, a key question in France is whether the mid-term evaluations of the SPDs and the CPERs should be linked, if not actually conducted simultaneously. However, aspirations and practicalities are having to diverge. There will be regional elections in Spring 2003, and the *Préfets de Région*, the representatives of the state in the regions, are politically unable to negotiate on the CPER mid-term evaluations with the Presidents of the regional councils near these elections. To address the problem of timing, the revision and renegotiation of the CPERs has been brought forward by six months and their evaluation exercise has also been anticipated. The consequence is that there will not be real and thorough mid-term evaluations of the CPERs, but only partial elements.

There will be other differences between the mid-term evaluations of the Structural Fund programmes and of the CPERs. The former will aim to assess the entire programme, including in its internal and external coherence, while in the latter the CPER document itself is not evaluated as a whole but is divided into thematic elements for scrutiny. This is because the CPER is too large a strategy to be evaluated as a whole, and it makes more sense to assess its main component parts (eg. urban policy or regional transport) separately. In addition, the decision-making process to conduct the mid-term evaluations of the CPERs is not the same of the Structural Fund programmes: it goes through different channels, even though it is common to see the same evaluators working on both.

#### 3.1.2 Treatment of thematic and cross-cutting evaluations

**While most regions will be undertaking programme-based evaluations, the opportunity remains to pursue more specific evaluations around particular themes.** There are several reasons for looking at certain aspects of the programmes in more detail. They may be important issues which merit separate treatment from the main programme evaluations, perhaps because of methodological considerations or out of a desire to highlight the issues. Where those themes are examined on a cross-cutting basis over several programmes, arguments of economies of scale come into play again, not just in terms of the cost of the evaluations, but also, as will be discussed later in the paper, the expertise available to undertake such evaluations effectively.



The current round of programming has introduced a range of new strategic and operational issues for Managing Authorities to address, and the mid-term evaluations provide a natural point for assessing the progress of programmes. Theoretically, the number of potential ‘themes’ for separate evaluation is large, but, in practical terms, several of the main ones have already been identified. The Commission itself has gone some way to influencing the agenda by presenting clear regulatory obligations for addressing certain issues. In Working Paper 8, it is stated:

*“A further concern for the evaluation of impact at mid term stage is related to the evaluation of expected impacts on a limited number of fundamental priorities, notably those concerning the EU’s own policies: impact on environment and equality of opportunity between men and women in particular, but also SMEs, competitiveness and innovation, the information society, local development and employment with regard to the European Employment Strategy as well as rural development.”*

Of these issues, the horizontal themes – especially environmental sustainability and equal opportunities – are pre-eminent. They are explicitly identified in one of the six objectives for the mid-term evaluations: *“to assess the extent to which the horizontal priorities – equal opportunities and the environment in particular – have been integrated into the forms of assistance”*.

Overall, **the importance placed on evaluating the horizontal themes is one of the most distinctive features of the mid-term evaluations this time.** A review of mainstreaming the horizontal themes was presented in a previous IQ-Net paper.<sup>8</sup> From the perspective of organising the mid-term evaluations, the key issue is whether the horizontal themes should be evaluated separately or as an integral part of the programme evaluations. The following table, drawn from the IQ-Net horizontal themes paper, summarises the main benefits and cost of both approaches.

---

<sup>8</sup> For further detail, see: Taylor S, Polverari L and Raines P (2001) *Mainstreaming the Horizontal Themes into Structural Fund Programming*, IQ-Net Thematic Paper 10(2), European Policies Research Centre, University of Strathclyde, Glasgow.

Table 3.3: Costs and benefits of integrating and separating the evaluation of the Horizontal Themes (HTs)

|          | Separate evaluation of HTs                                                                                                                                                                                                                                                                                                                                                                                                                       | Integral evaluation of HTs                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Benefits | <ul style="list-style-type: none"> <li>• Ensures themes receive significant attention</li> <li>• Facilitates the selection of thematic experts as evaluators</li> <li>• Useful where HT issues are new and a focused approach would help to bring a step change in awareness and ability</li> <li>• Where they are undertaken across multiple programmes, they can help promote an active exchange of practice and draw wider lessons</li> </ul> | <ul style="list-style-type: none"> <li>• Ensures continuity as the HTs will be addressed in all stages of evaluation (<i>ex ante</i>, interim and <i>ex post</i>)</li> <li>• Useful where HT issues are already well understood</li> <li>• Can reinforce the perception of the HTs as horizontal elements, relevant to all aspects and phases of programming, and at the heart of progressive, integrated economic development approaches</li> <li>• Potentially cost effective compared to commissioning separate studies</li> </ul> |
| Costs    | <ul style="list-style-type: none"> <li>• Cost implications of additional separate studies</li> <li>• Risk of perpetuating the perception that the HTs are a stand-alone issue</li> <li>• Risk of lack of continuity</li> </ul>                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>• Risk of superficial treatment of HTs, perceived as just one of many elements to be dealt with - and not the most important</li> <li>• Possibility that HT aspects will only be addressed qualitatively rather than quantitatively</li> </ul>                                                                                                                                                                                                                                                 |

Overall, **most programmes appear to favour treating the horizontal themes as part of the mainstream mid-term evaluations.** The decision to opt for an integral approach derives from Commission guidance and partly follows from the programme-based logic of how the mid-term evaluations are being conducted as a whole. In some cases, nevertheless, there has been debate over whether to undertake separate evaluations for the horizontal themes. For example, in Scotland, the issue was intensively discussed within the Monitoring and Evaluation Working Group of the Scottish Executive when the overall organisation of the mid-term evaluations for the Scottish programmes was being determined. One decisive factor in favour of the integral approach was that singling out the horizontal themes for special treatment could be against the principle of mainstreaming. In other cases, the integral approach was prompted by the feeling that the horizontal themes were not sufficiently important aspects of the programmes to warrant separate evaluations.

In general, **separate thematic evaluations do not appear to be favoured. However, there has still been a universal effort to ensure that the mid-terms will not only look at core issues but specific thematic concerns as well.** It has been widely recognised that, if the mid-term evaluations are to assist programme managers in improving their programmes, the evaluations need to focus on issues of special interest to the programmes.

This is occurring in two ways. First, several regions are considering separate evaluations for particular themes. For example, in Wales, programme managers hope that the mid-term evaluations will identify a series of issues which will be evaluated on an *ad hoc* basis in the coming years. Second, the terms of reference for evaluations typically include a focus on certain thematic issues. In Denmark, the sub-programmes are expected to raise particular issues

to be addressed in the course of the national Objective 2 mid-term evaluation. In Sweden, within the multiple-programme evaluation for Objective 2, special themes of relevance to all the programmes – both in terms of the content and the implementation of the Funds – are being defined for inclusion in the terms of reference.

### 3.2 Dividing the different roles in the mid-term evaluations

One of the most visible features of the mid-term evaluations is **the clear identification of particular roles for different levels**. The Commission, the Member States and the individual programmes all have specific tasks in the mid-terms, and these tasks have been more closely defined than on previous occasions. While roles will differ depending on how the Structural Funds are organised in individual Member States, what is striking is the degree of similarity in the responsibilities allocated to each level of governance.

The clearer articulation of tasks and responsibilities for mid-term evaluation is broadly the result of two factors. First, **the cycle of evaluations for different Structural Fund programming periods has embedded an evaluation culture and deepened evaluation capacity in all Member States**. At national and regional levels, officials involved in the Structural Funds have a strong understanding of what is required in the different evaluations of the programme cycle and, for most programmes, evaluations have become relatively routine – although not stale – exercises. To a large extent, tasks and responsibilities have been institutionalised.

Second, **the complexity of the mid-term evaluations in this round requires greater definition of responsibilities than hitherto**. The requirements of the evaluations are more detailed and specific than in the previous programming round. With the longer programming period, the incentive to make more active use of evaluation results at different levels has increased. As a result, it has been seen as important that different bodies have clearly-defined roles in the different stages of the evaluation exercise.

*Table 3.4: Division of evaluation responsibilities between different levels*

|                                          | European Commission | National/territorial | Programme |
|------------------------------------------|---------------------|----------------------|-----------|
| Guidance and advisory services           | 0                   | 0                    |           |
| Determining overall evaluation framework | 0                   | 0                    |           |
| Issuing terms of reference               |                     | (√)                  | 0         |
| Selecting evaluators                     |                     | (√)                  | 0         |
| Managing evaluations                     |                     | (√)                  | 0         |
| Ensuring completion of evaluations       |                     | 0                    | 0         |
| Responding to the results of evaluation  | 0                   | 0                    | 0         |

The tasks are distributed between three levels: the Commission, the Member States and the individual programmes. The motivations of each set of actors differ, as do their roles. Broadly, the Commission and the Member States set the terms of the overall framework for evaluation and provide support and

advice, while the individual programmes undertake the actual running of the evaluations. In some cases, the Member State and programme level co-exist, as in Denmark. The table above illustrates the main division of responsibilities between the different levels.

### 3.2.1 *The roles of the Commission*

The Commission has two key roles in the mid-term evaluations:

- **a regulatory obligation to be consulted by Member States** on their proposals for the mid-term evaluations and thus to fulfil a quality control or insurance function; and
- **a less formal and more administrative set of functions** in facilitating the preparations for the evaluations.

With respect to its regulatory role, the Commission's main activities have been based on consultation with the Member States. Specifically, the EC has determined that there is a regulatory requirement for consultation to take place on the mid-term evaluations in five areas.

- **Terms of reference.** It has reviewed the draft terms of reference of the evaluations to ensure all issues are covered in the guidance and there is consistency across the Member States.
- **Budgets.** It has advised on budgets set aside for the mid-term evaluations so that they are adequately resourced.
- **Tendering procedures.** Where appropriate, it has commented on the procedures in place for selecting evaluators (although it has not been directly involved in the tendering process).
- **Methods.** It intends to go through the evaluation inception reports to make a judgement on the appropriateness of the methods suggested.
- **Reports.** It wants to be consulted on the reports, in part so that MEANS criteria are used in their assessment.

Several DGs are involved in the mid-term evaluation exercise as a result of their responsibility for different Funds (eg. DG Employment for ESF, DG Agriculture for EAGGF). However, for the evaluation of Objective 1 and 2 programmes, the Evaluation Unit of DG Regio has been the most active part of the Commission. Indeed, it has arguably been more active than ever before in assisting Member States and regions in preparing for an effective evaluation exercise. DG Regio has done this in several ways.

- **Provision of guidance papers.** The core guidance for the evaluations is DG Regio Working Paper 8, which has been the starting point for all Member States in structuring their responses to the mid-term evaluation requirements. In its role as a reference document, the guidance has influenced all the main participants in the evaluation exercise, including not just Member States and programme managers, but the steering groups and the evaluators themselves. At all levels, the guidance has been judged to be practical and comprehensive, providing a logical, robust structure which covers most issues of importance to the evaluation. It has accelerated the process of responding to the evaluation obligations by

relieving authorities of the need to interpret these for themselves and provided additional certainty that, as long as it is followed, the studies undertaken should satisfy EC requirements. As a Working Paper, it has been taken up by choice rather than obligation by those authorities who are using it. It has been supplemented by the other Methodological Working Papers, including papers on the performance reserve, monitoring indicators and the evaluation criteria for Community added value. Documents are also provided by other DGs, such as DG Employment's *Note on the Evaluation of the Contribution of the Structural Fund to the European Employment Strategy* and various guides on the evaluation of rural development programmes from DG Agriculture.

- **Provision of advice.** DG Regio has been supplying advice to the Member States on meeting the evaluation obligations, notably through involvement in technical groups at central and regional levels (as in Spain), support in developing national-level guidance (as in Italy) and through *ad hoc* meetings with territorial administrations, either singly or collectively (eg. Portugal and the UK).

The Commission's pro-active stance on the evaluations has been driven by their own requirements. In addition to being a critical part of determining the allocation of the performance reserve (see section 5.3.1), the Commission also need the evaluation results to feed into the Third Cohesion Report. It is in their interests to expedite the evaluations as smoothly as possible.

For the most part, **the Commission's activities have been interpreted by all sides as constructive, particularly in providing the main guidance on evaluation requirements.** However, a few areas of contention have emerged, where Member States have proposed approaches which have not been accepted by the Commission (eg. the use of Fund-specific evaluations in some German Objective 1 regions). In a few cases, the EC has potentially been too actively involved in some evaluation steering groups, their role being seen as overly prescriptive.

### 3.2.2 *The roles of national level authorities*

The roles of the Member States themselves have changed since the last set of mid-term evaluations. The key change has been **clearer areas of responsibility for national level authorities in coordinating the evaluations.** In some cases, the change has been particularly pronounced, as in Denmark, where the shift from a series of Objective 2 programmes to a single programme has led to a definite placing of evaluation accountability at the national level.

While Member States generally have an important coordinating role, the division of responsibilities between different levels within Member States varies significantly across the EU. Three separate levels of responsibility are relevant here:

- **the national level,** principally through the main central government department(s) charged with overseeing Structural Fund affairs;

- **the territorial level**, especially in countries where territorial administrations have significant Structural Fund roles (eg. Austria, Belgium, Germany and the UK); and
- **the programme level**.

The table below summarises how evaluation tasks have been divided between these different levels. These tasks are threefold:

- **the design of the evaluation**, especially the overall evaluation framework and the terms of reference for individual programmes;
- **the management of the evaluation** in terms of the selection of evaluators and the day-to-day conduct of the evaluations; and
- **the formal response to the results of evaluation**, whether with regard to the performance reserve or changes to be implemented within individual programmes.

*Table 3.5: Responsibility for designing, managing and responding to evaluations*

|                           | <b>Design</b>         | <b>Management</b> | <b>Response</b>       |
|---------------------------|-----------------------|-------------------|-----------------------|
| <b>Austria</b>            | National/programme    | Programme         | National/programme    |
| <b>Belgium (Flanders)</b> | Territorial           | Territorial       | Territorial/programme |
| <b>Denmark</b>            | National              | National          | National              |
| <b>France</b>             | National/programme    | Programme         | National/programme    |
| <b>Germany</b>            | Programme             | Programme         | Programme             |
| <b>Italy</b>              | National/programme    | Programme         | National/programme    |
| <b>Spain</b>              | National/programme    | Programme         | National/programme    |
| <b>Sweden</b>             | National              | National          | National/programme    |
| <b>UK</b>                 | Territorial/programme | Programme         | Territorial/programme |

As the table shows, **there are very few Member States where the national level does not have a role in the mid-term evaluations**. In countries with strong territorial government structures, national-level tasks are often carried out within the territory (eg. Scotland and Wales in the UK). The national/territorial level tends to be involved in tasks associated with the design of the evaluations, while the actual management of the evaluations is left wholly to the programmes themselves (although sometimes the two levels are indistinguishable). Only in countries where a single evaluation of all the Objective 2 programmes is being conducted does the national level tend to take a more active operational role (Denmark, Sweden).

For the most part, the ‘response’ tasks of the national/territorial level are restricted to using the mid-term evaluation results in allocating the performance reserve (see section 5.3.1). The main tasks undertaken by national/territorial authorities lie in the design of the evaluation, whether directly setting the terms of reference or providing support and advice to the programmes in producing their own terms of reference. These can be considered under two sets of headings: issuing guidance, and capacity building.

**Several national authorities have issued their own general guidance for the mid-term studies, typically based on the structure and content of Working Paper 8.** Essentially, national governments have been responsible

for presenting standardised interpretations of the Commission's evaluation requirements, often resulting in core terms of reference which individual programmes are advised to adopt with some programme-specific tailoring. In addition, they often supply guidance on operational aspects of the evaluations, such as timetables.

This has been prompted by several major concerns on the part of Member States:

- an increasing preoccupation with comparing the performance of different Objective 2 programmes within any given Member State, mainly because of the need to allocate the performance reserve;
- a desire to compile a national overview of Structural Fund achievements with a view to informing wider policy debates; and
- an interest in examining specific policy innovations.

Overall, it is leading to more uniform evaluation efforts within countries. National (or territorial) guidance has been issued in several Member States in order to standardise the way evaluations are conducted by individual programmes. This can take the form of issuing core terms of reference, as in the UK (by the Department of Transport, Local Government and the Regions for England and the Scottish Executive in Scotland). It can also involve more general guidance, as in Italy, where the Ministry of Economy and Finances, which is the Managing Authority for the Objective 1 CSF, provided Objective 1 guidance and templates for terms of reference that were also circulated to Objective 2 programmes.

One risk of the provision of guidance is the potential for the actors targeted to respond passively rather than interpret the guidance in relation to individual programmes needs. In some cases, the temptation may be simply to follow Commission and national proposals 'blindly' rather than verify their appropriateness to the given context and, if necessary, make appropriate changes in approach or emphasis.

Perhaps sensitive to this risk, **the second key evaluation design task undertaken by national and territorial authorities has been capacity building.** In several Member States, workshops and consultations with individual programmes have been held with several goals in mind:

- to impart to programme managers the requirements of the mid-term evaluations;
- to provide technical advice and commentary in drawing up appropriate terms of reference; and
- to act as a source of methodological and organisational good practice upon which programmes can draw.

Hence, in France, DATAR has organised training seminars for the Managing Authorities, in part to encourage the adoption of common evaluation questions (eg. with respect to the horizontal themes). One of the most active examples of capacity building has taken place in Austria, through the designation of a central resource to provide advice and assistance in designing the evaluations: *the Coordination and Work Platform* (described in the box below).

**AUSTRIA:****A COORDINATED APPROACH TO DISTRIBUTING ROLES**

The decision to take an ongoing approach to evaluation in Austria has influenced the way evaluations will be organised.

- The main responsibility for evaluations (both the on-going and mid-term elements) will lie with the Managing Authorities in the nine *Länder*. They will be responsible for drafting and issuing the terms of reference, the selection of programme-specific themes or foci for inclusion in evaluations, the choice of an evaluator and the steering of the process.
- A *Coordination and Work Platform* will be established in ÖROK (the federal level coordination body for regional policy), designed to undertake a series of tasks. These include: promoting coordination and exchange of experience between the managing authorities and the evaluators; providing input for areas where there is overlap (eg. the production of a core text for the terms of reference covering the EC requirements to be supplemented by the individual programme-specific themes chosen by the *Länder*); and providing information on themes which will appear in all the evaluations (eg. the monitoring system, federal level structures, the paying authority etc) to avoid duplicated work.
- The Coordination Platform will be able to commission thematic evaluations on cross-cutting themes such as the horizontal priorities or themes of common interest to all the programmes which could have a useful comparative element. Decisions on whether or not to do this, and what themes would be selected, would be discussed in the Platform. One other option could be to look at process issues in areas such as the horizontal themes. While the entire Structural Fund system could not be analysed from a process-perspective because of its complexity, a horizontal approach could highlight case studies, potentially of good and more problematic programmes, and try to draw out process factors which point to conditions under which successful implementation can occur.

### 3.2.3 Programme level roles

In the majority of cases, **the lead organisation in evaluations is at the programme level in the form of the relevant Managing Authority.** Where programme-level actors are not the lead authority, it is their task to cooperate with the wider evaluation study which is being led by another organisation. For the most part, though, individual programmes are responsible for the management of the evaluations. As with previous evaluations, this involves several tasks:

- drawing up the evaluation terms of reference (often based on core guidance provided by national/territorial authorities);
- managing the evaluation selection procedure;
- working with the evaluator in the development of an appropriate methodology, provision of relevant monitoring data and assistance with sampling for project surveys (where appropriate);



- ensuring the delivery of the required outputs according to the agreed timetable and that the evaluation meets quality thresholds; and
- following up the recommendations with appropriate actions.

In performing these tasks, wider partnerships and stakeholders have frequently been involved, most immediately through steering groups. Such groups can, of course, be formed at either national/territorial or programme level, depending on the overall organisation of evaluations within the Member States.<sup>9</sup> Evaluation working or steering groups were a frequent feature of mid-term evaluations in 1994-99. In previous programming rounds, they were commonly employed in Belgium, Denmark, Finland, France and the UK, whereas in Member States like Germany and Spain, evaluations have routinely been directed by programme managers and/or secretariats. At their most developed, steering groups have been formalised as permanent structures, addressing monitoring and evaluation issues throughout the programming cycle.

In the current round, **evaluation steering groups will be even more widespread than before, and potentially even universal.** In Working Paper 8 (p.9), the Commission recommended:

*“Each mid term evaluation should be guided by a Steering Group representative of the monitoring committee for the form of assistance. The Steering Group’s role is largely technical. It will develop the terms of reference for the evaluation, select the evaluators, guide the evaluation, give feedback on the first draft and approve it for quality on completion.”*

As a result, steering groups have been newly introduced in Spain at the programme level and within the *Land Managing Authorities* in Germany. Such structures involve a representative group of stakeholders directing the evaluation process *in itinere*. Usually, the evaluation steering group is a subset of the Monitoring Committee, consisting of representatives of the latter, national/territorial agencies and the Commission, as well as external experts and, in a few cases, social partners. It is constituted as a special initiative of the Monitoring Committee and works under and reports to this structure (particularly where major decisions have to be taken, such as approving the terms of reference and the selected evaluator).

The tasks of the steering group can be seen in a case-study description of how such a structure operates in Toscana.

---

<sup>9</sup> In the case of Austria and the UK, steering groups have been formed at national/territorial and programme levels.

### THE TUSCAN EVALUATION STEERING GROUP

The Tuscan Evaluation Steering Group will comprise six members:

- the Manager responsible for the evaluation of the programme, from the Programme Secretariat;
- one representative of the regional Environmental Authority;
- one representative of the Equal Opportunities Regional Committee;
- one representative from the national Ministry of Economy and Finances;
- one representative from DG Regio; and, lastly,
- one representative from the Regional Union of Tuscan Provinces (URPT, *Unione Regionale delle Province Toscane*).

The main objective of the Steering Group is to ensure the quality of the evaluation system. The Steering Group will undertake the following activities:

- outlining evaluation activities and methodologies for the mainstreamed evaluations of the programme;
- monitoring the activities of and developing an on-going dialogue with the evaluator/s;
- verifying the interim evaluation report and its quality for the allocation of the performance reserve;
- defining content and procedures for the definition of thematic evaluations and the identification of the evaluators covering them; and
- cooperating with the regional Evaluation Unit (NURV *Nucleo Regionale di Valutazione degli Investimenti Pubblici*) for those areas under its competence.

In some cases, the steering groups are to be involved in all stages of the evaluation, from the drafting of the initial terms of reference to following up recommendations. For others, the groups only become operational later in the evaluation cycle. In Flanders, for example, the preparation process has so far been relatively restricted: the Provinces and programme management units have not yet been consulted, and, while a steering group will be created, this will only happen later in the process.

Steering groups can improve the quality of evaluations in several ways.

- Ultimately, these groups are a quality control mechanism. They meet more frequently than the Monitoring Committee and thus can undertake more effective steering of a time-dependent process.
- In terms of function, they are an important bridge between the specific exercise to be undertaken and the wider programme (with its ongoing priorities and calendar of work).
- They also bridge the gap between those most centrally involved in the evaluation study and the wider partnership (through participants representing a diverse range of stakeholders). Thus, they provide a conduit for relevant information and ideas to pass to – and from – particular groups. This legitimises the evaluation by ensuring that a wider group has realistic expectations of evaluation and understand some of the difficulties

inherent in this practice. As a result, it helps to increase the prospect of studies being endorsed and recommendations acted on.

### 3.3 The timetables for evaluation

An indicative and detailed timetable for organising the mid-term evaluations is set out in Working Paper 8, covering each stage of the evaluation cycle. The Commission recommended that the process of planning the evaluations should be underway by June 2002 with the aim of awarding contracts by August.

This timing will suit many, but not all. The Italian regions are the most concerned about timing in that they believe that delays to the start of their programmes mean there will have been insufficient activity to gain benefit from evaluating in time for the deadline of 31 December 2003 (by April 2002, no commitments had yet been made). In France, DATAR is also concerned, albeit to a lesser extent, that French Objective 2 programmes started slowly for a range of reasons (both external and internal) and there may not be a sufficient volume of activity for the evaluations to generate entirely reliable conclusions. However, the key lesson here will be to treat any conclusions with caution.

In line with the Commission timetable, **planning has already begun for mid-term evaluation in every partner region, although this process has advanced to varying degrees in different contexts.** Delays in France mean that contracts are unlikely to be awarded until September 2002, while, in some Italian regions, evaluators have already been selected. Nevertheless, as the comparative timetable below illustrates, all Member States have planned to ensure that they meet the Commission's deadline for submitting the results.

Table 3.6: Member State timetable for mid-term evaluations

|                           | June<br>02 | Sept<br>02 | Jan<br>03 | Apr<br>03 | Jul<br>03 | Sept<br>03 | Dec<br>03 |
|---------------------------|------------|------------|-----------|-----------|-----------|------------|-----------|
| <b>Austria</b>            | —          |            |           |           |           | →          | •         |
| <b>Belgium (Flanders)</b> | —          |            |           |           |           | →          | •         |
| <b>Denmark</b>            | —          |            |           |           | →         |            | •         |
| <b>France</b>             |            | —          |           |           |           | →          | •         |
| <b>Germany</b>            | -----      |            |           |           |           | →          | •         |
| <b>Italy</b>              | —          |            |           |           |           | →          | •         |
| <b>Spain</b>              |            | —          |           |           | →         | •          |           |
| <b>Sweden</b>             |            | —          |           |           | →         | •          |           |
| <b>UK</b>                 | —          |            |           |           | →         | •          |           |

→ Period between award of evaluation contract and completion of draft Final Report.

• Availability of Final Report.

----- Range of dates for launch of German evaluations (eg. June in Bremen, later in the autumn in Nordrhein-Westfalen and Saarland).

By the time the studies take place, enough programme activity will have taken place for most evaluations to be worthwhile. With the longer programming period this round, there will also be enough time remaining for recommendations to have a real impact. Moreover, the update study scheduled for 2005 potentially provides a form of 'safety net', allowing follow-up where the 2003 study has been too early in the programming period.

### 3.4 Evaluation budgets

While not wanting to be overly prescriptive in advising on the resources which should be set aside for the evaluations, the Commission has been clear on one issue. In Working Paper 8, it stated that:

*"For the mid term evaluation of the 1994-1999 period, the average cost was less 0.1% of the total annual cost of the programmes. In order to obtain quality evaluations, the cost should be higher than this in general, except in the case of very large programmes."*

The Commission has been keen to ensure that sufficient resources are allocated to the evaluations, given the greater number of tasks they are meant to achieve in this round. Budgets should be proportional to the size of the programme, except for particularly large programmes. Rather than set firm limits, DG Regio has discussed budgets with Member States on a case-by-case basis. Overall, **the Commission has been satisfied with the financial levels being set**, evidence that Member States are taking the evaluations as seriously as the Commission would wish.

An overview of budgets cannot be compiled at this stage. There are several reasons for this. Some programmes are an early stage, perhaps without a decision having been taken on maximum available resources, while more generally, there is sensitivity to releasing this information at this point in the tender process. Nevertheless, feedback does indicate that the budgets for the mid-term evaluations will in many cases be higher than they have been in previous rounds of evaluation. It also suggests that there is variation in how much will be allocated to evaluations.

### 3.5 Selecting evaluators

In most cases, **by May 2002, tendering processes had only just begun**, and the evaluators would undertake the mid-term studies had yet to be chosen. There were exceptions, notably Toscana, where an evaluator has already been selected to cover the evaluation needs of the whole programming period. The timetable for most regions can be seen in the case of Flanders, which is one of the furthest on in the process. Flemish authorities issued a call for expressions of interest in the relevant national media as well as the Official Journal in February 2002. In March, selected respondents were invited to submit a full bid for the work. Four responses were received by the April 2002 deadline and the final selection decision was scheduled to be taken in June by the Steering Group and the Monitoring Committee.

Evaluators are usually being selected through an open call procedure. In practice, though, no matter how widely contracts are advertised, the range of evaluators who will bid is likely to be restricted by practical considerations, notably languages spoken and proximity to the area in which the programme is

being run. Nevertheless, as many programmes have had to place calls in the EU's Official Journal because of contract threshold limits, there is scope for a potentially wider pool of evaluators to be reached.

The local capacity of the evaluator pool has been a common concern with evaluations in the past. Such concerns have diminished with the current round of mid-terms, but have not been fully removed. Two sets of challenges in selecting an evaluator have emerged:

- the question of whether there are enough evaluators to undertake the number of simultaneous evaluations required; and
- the question about whether evaluators have the sufficient skills to meet the requirements of the current round.

**Availability of skilled evaluators is potentially difficult with the mid-terms.** Bottlenecks can occur because of the significant number of evaluations being run in parallel, placing a strain on the evaluator pool. Still, most Member States have fewer problems with evaluator capacity than hitherto as they reap the benefits of capacity building to date: several cycles of Structural Fund evaluation; a series of EC-sponsored conferences developing Structural Fund specific expertise and knowledge; and, increasingly widespread evaluation societies active in publishing and organising conferences and exchanges.

Consequently, few have expressed concerns that there may not be sufficient evaluators to go round. For example, Denmark has experienced definite improvements in evaluator capacity in the current round, with a rising number of firms involved in evaluation studies and the emergence of a sizeable Danish section of the European Evaluation Society. This has been reflected in a sufficient response to calls made so far.

**Perhaps more critical is whether evaluators can meet the additional demands involved in this round of evaluations.** Of particular relevance are the horizontal themes, as evaluators will need to demonstrate their expertise in this area. Given the prominence of the themes in many terms of reference, programmes are concerned that evaluators can demonstrate they have the necessary skills to consider environmental sustainability and equal opportunities issues. For some regions, the evaluations will present other new challenges. For example, in Spain, the decision to undertake multi-fund evaluations has implications for potential evaluators experienced with previous programmes.

Several solutions have been put forward to the additional demands, among them:

- **that evaluators form partnerships** by assembling a team with the appropriate range of skills and experience - for example, this has been formalised in the Italian *Associazioni Temporanee di Impresa* where consortia of evaluators bid together; and
- **that evaluators develop the required expertise in-house**, recognising the importance of possessing the wider range of skills for future evaluations.

## 4. DESIGNING AND UNDERTAKING THE MID-TERM EVALUATIONS

As described above, the mid-term evaluations of 2000-06 Structural Fund programmes are shaped by the same regulatory requirements and DG Regio guidance and are likely to have a core of common elements. Nonetheless, each of the studies will be unique: **programme evaluations are not a standard off-the-shelf purchase.** This is not only because every programme, its policies, its institutional and political context and the people involved are distinctive, but also because there is no standardised optimal solution to the design and execution of these studies. Mid-term evaluations have multiple aims, and different emphases may be placed on each. Even where the same aims are being pursued, they can be achieved by different means. In addition, there are choices in terms of the style of evaluation.

Given the stage which most programmes had reached by April 2002, it was not possible at the time of drafting this paper to discuss the actual tools and methods which would in practice be used in the evaluations. Nonetheless, it the paper **explores a range of issues around the definition and execution of the evaluations**, notably:

- how and when the methodological choices will be made for the evaluations, and by whom, illustrating the broad involvement in this process;
- whether and how the evaluations fit into a wider context, contrasting programmes which undertake relatively self-contained studies to the EC timetable with those which have a more ongoing vision of evaluation;
- favoured styles and processes of evaluation, including whether the evaluator is seen as an expert, for instance, or as a facilitator;
- the anticipated content of the evaluations; and
- the evaluation tools employed.

Each of these issues is addressed below, opening out some of the choices facing those about to commission and undertake evaluations. A recurring theme in this section is how **the overall conception of evaluation is likely to influence outcomes in terms of scope and organisation.** The key distinction lies in the balance between viewing evaluation as a **control function**, satisfying aims to do with accountability, justification and transparency, and seeing it as a **learning tool** (and, indeed, as a learning process) undertaken to serve the quality, efficiency and effectiveness objectives of the programme itself.

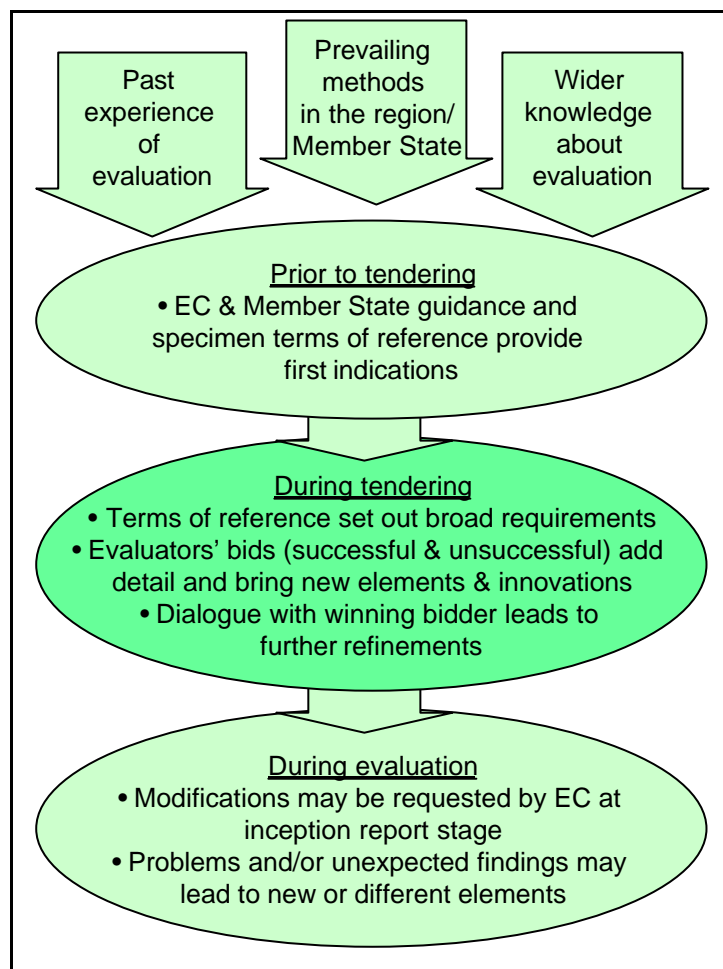
### 4.1 Overall approach: defining evaluation methods

In understanding the diversity which could arise in the mid-term evaluations, it is useful to reflect on how and when methodological choices are made, and who and what influences these outcomes.

**Decision-making about methodologies** is often considered to be confined to three phases: the elaboration of the terms of reference; the preparation of bids in response; and a subsequent brief phase of negotiation to adjust the offer in

line with requirements. However, in reality, **the process has multiple stages and involves many actors, either directly or indirectly.** This is illustrated by Figure 4.1, which summarises the stages at which methodological inputs are made, and Table 4.1, which highlights the diversity of actors who contribute to this process. Both illustrate that, even once decisions have been taken about methodological choices, the approach can still change, as a result of problems encountered, initial findings and the influence of the EC which has overall responsibility for ensuring the quality of the evaluations.

*Figure 4.1: Defining an evaluation methodology: a progressive process*



A key message from these illustrations is that **the evaluator is only one player among many, each of whom has a particular contribution to make** in defining the scope and conduct of the evaluation, based on their experience, particular knowledge and priorities. In particular, the expertise of the evaluator can best shine when it complements the understanding and practical knowledge of programme managers and partners.

Table 4.1: Contributors to methodological definition

|                   | Methodological input                                                                                                                                                   | Contributors                         |        |                          |                      |                                        |            |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|--------|--------------------------|----------------------|----------------------------------------|------------|
|                   |                                                                                                                                                                        | Evaluation experts & advisory bodies | The EC | Member State authorities | Programme management | Programme partnership (Steering Group) | Evaluators |
| Context factors   | Past experience of Structural Fund evaluation                                                                                                                          | Ö                                    | Ö      | Ö                        | Ö                    | Ö                                      | Ö          |
|                   | Prevailing methods of evaluation in the parallel domestic policy environment                                                                                           | Ö                                    |        | Ö                        |                      | Ö                                      | Ö          |
|                   | The body of knowledge on Structural Fund evaluation, including MEANS guidance and diverse other materials                                                              | Ö                                    | Ö      | Ö                        | Ö                    | Ö                                      | Ö          |
|                   | Guidance and specimen terms of reference for the mid-term evaluations                                                                                                  |                                      | Ö      | Ö                        |                      |                                        |            |
| During tendering  | Programme-specific terms of reference for a mid-term evaluation                                                                                                        | Ö                                    |        | Ö                        | Ö                    |                                        |            |
|                   | The evaluators bidding in response to the terms of reference (any evaluator who bids can influence the final methodology chosen, even if they do not win the contract) |                                      |        |                          |                      |                                        | Ö          |
|                   | Assessment of the bids received (drawing out new ideas or useful practices, not necessarily from the winning bid)                                                      | Ö                                    |        | Ö*                       | Ö                    | Ö                                      |            |
| During evaluation | Dialogue between the client and the winning bidder, following award of the bid                                                                                         |                                      |        | Ö*                       | Ö                    | Ö                                      | Ö          |
|                   | At inception report stage, when DG Regio has reserved the right to make a judgement on the appropriateness of methods                                                  | Ö                                    |        |                          |                      |                                        |            |
|                   | During the evaluation process, when initial findings or data problems may lead to additional and/or different elements of evaluation becoming desirable                |                                      |        | Ö*                       | Ö                    | Ö                                      | Ö          |

\* - where national evaluations are in operation (eg. Sweden).

A frequent pattern among partner regions is for the scope and aims of the proposed evaluations to be set out broadly in the terms of reference, but then evaluators to be invited to suggest how they should be addressed. Indeed, this is often one of the key assessment criteria for bids. **It is a delicate balance to strike, ensuring that programme actors and evaluators make an optimal contribution to defining the methodology.** For example, the terms of reference could leave the choice of methods entirely open, or could specify methods broadly (eg. desk analysis of monitoring data and a more detailed analysis of a sample of projects) and leave it up to the evaluator to respond



with the methodological detail (eg. sampling methods and the potential coverage of questionnaires).

Leaving some issues open in the terms of reference lets evaluators bring in their own ideas and propose innovations. At the same time, though, even if programme actors do not prescribe their requirements in the terms of reference, they need to have thought them through with sufficient clarity and detail to be able to appraise and compare the bids which are submitted. Detailed prior thinking could also help to alert them to any data problems which might constrain certain methodological options. Knowing this at the bidding stage and bringing it to the attention of potential evaluators saves wasted time by ensuring bidders are working on the basis of meaningful assumptions and only propose feasible methods.

## 4.2 Context: how the mid-term evaluations are conceptualised

One of the most interesting recent debates on evaluation, which has implications for the design of the mid-term studies, has been about how individual evaluation studies are conceptualised. In particular, there is the issue of the **potential merits of moving from considering evaluation studies as self-contained to seeing evaluation as an ongoing or continuous discipline**. MEANS defines on-going evaluation as:

*“Evaluation which extends throughout the period of implementation of an intervention. This form of evaluation accompanies the monitoring of outputs and results. It is too often confused with monitoring”.*<sup>10</sup>

Stripping out the overlaps between programming periods, which would complicate the picture, Figure 4.2 illustrates the four evaluations which are required by the Structural Fund regulations for the 2000-06 programmes. This is the minimum activity which will be undertaken for all programmes. There is no obligation to interpret these studies as parts of a coherent and incremental process of evaluation. The bottom row in the table, however, highlights the potential scope of an ongoing approach which could supplement the core activities.

Figure 4.2: Periodic and ongoing evaluation

| Year                                     | 1999           | 2000                                                                                                                                                                                                                            | 2001 | 2002 | 2003            | 2004 | 2005                   | 2006 | 2007 | 2008           |
|------------------------------------------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|-----------------|------|------------------------|------|------|----------------|
|                                          |                | 2000-06 Programme                                                                                                                                                                                                               |      |      |                 |      |                        |      |      |                |
| ‘Periodic’ Evaluation                    | <i>Ex ante</i> |                                                                                                                                                                                                                                 |      |      | <b>Mid-term</b> |      | <b>Mid-term update</b> |      |      | <i>Ex post</i> |
| Potential elements of ongoing evaluation |                | <b>Ad hoc thematic studies</b> commissioned as and when required and/or<br><b>Ongoing dialogue</b> with evaluators about programme progress and/or<br>Evaluation <b>steering group maintained</b> throughout programming period |      |      |                 |      |                        |      |      |                |

<sup>10</sup> MEANS (1999) *Evaluating socio-economic programmes, Volume 6: Glossary of 300 concepts and technical terms*, Commission of the European Communities, Luxembourg.

There is **no consensus about what constitutes an ongoing evaluation approach** and how the standard menu would need to be supplemented if such an approach were being followed. However, **a continuous or ongoing approach to evaluation would arguably span the lifetime of a programme. It would bring additional elements and activities to the core menu**, as and when required, with decisions being taken during advance planning and/or programme implementation. These decisions could be driven by the programme itself and/or in appropriate circumstances, by an authority empowered to take decisions on behalf of more than one programme. Among the countries where there is interest or experience in ongoing evaluation are Italy, Austria, Ireland and the UK. Where such approaches have been considered, however, there has not been unanimous support (see boxes below exploring existing practice, then new proposals, in Austria).

### **EXPERIENCE OF ONGOING EVALUATION IN AUSTRIA: OBJECTIVE 3**

#### **Scope of Austrian Objective 3 ongoing evaluation**

Building on experience in 1994-99, the evaluation of the 2000-06 Austrian Objective 3 programme will take an ongoing approach.

#### **How is ongoing evaluation organised, and what will be done?**

An evaluation contract has been let covering the whole programming period.

- The chosen evaluator will deliver annual reports on implementation to the Monitoring Committee.
- In addition, they will undertake a series of forward-looking thematic evaluations, covering the main themes in the programme.
- The mid-term evaluation will be primarily based on the annual reports delivered to that point together with the thematic studies.
- Under a single contract, two-thirds of the evaluation (including in financial terms) will be dedicated to the Objective 3 programme and the remaining third to the activities of the domestic Labour Market Service (AMS). In effect, the Structural Fund evaluation will be used to examine national activities as well. This should give an overview of labour market support activities as a whole and provide a better context for the operation of the Objective 3 programme.
- Finally, additional, labour market components will be provided.

## **PROPOSED APPLICATION OF AN ONGOING APPROACH TO OBJECTIVE 2 EVALUATION IN AUSTRIA**

### **What would ongoing evaluation consist of?**

There has been significant debate in Austria about the potential merits of adopting an ongoing approach to the evaluation of regional programmes (which would be distinctive from that used for Objective 3). This would run from 2002 until 2005 when the mid-term update studies are due. This approach views the mid-term evaluations as part of an integrated whole and recognises the inter-relationship between them and the Annual Implementation Reports. It would encourage an ongoing process of communication and learning about evaluation among the various separate programmes which would be involved in pursuing this approach, co-ordinated by ÖROK, the federal level body for coordinating regional policy, through a 'Coordination and Work Platform'.

### **How has the proposal been received?**

The proposal is supported by organisations at federal level and a range of evaluators. However, some uncertainties have been voiced by the Managing Authorities:

- The scope and benefits of ongoing evaluation are unclear. Evaluation is still seen by many programme actors within the framework of EC requirements as a reporting and control exercise.
- It would go beyond the strict regulatory requirements of the mid-term evaluations. Some *Länder* have been willing to explore options, but not all. The additional cost and effort would need to bring clear additional benefits.
- There has been particular concern about going further in terms of 'process' evaluation (this is discussed further in 4.4.3 below).

It is likely the approach will go ahead on an experimental basis, but clear that it will require very effective supportive input and steering from the 'Coordination and Work Platform' at federal level.

What specific arrangements characterise those programmes which have more deliberately opted for an ongoing approach to evaluation?

### ➤ **An evaluation plan is drawn up for a whole programming period**

This would present evaluation as parallel to, and supportive of, effective and efficient programme delivery on an ongoing basis, and draw out explicit connections between separate evaluation activities. Such an approach was taken in Scotland in 1994-99, and is being considered in Austria.

The evaluation plan would support the delivery of programmes throughout their lifecycle. Here, it would not be the evaluator but instead the programme managers and selected partners who maintain the coherent overview of evaluation.

### ➤ **Evaluation is kept on the management agenda on an ongoing basis**

One way to do this is by encouraging the Evaluation Steering Group to continue to meet throughout the programming period, and not simply to dissolve once each study has been completed and re-form in time for the next.

This provides continuity and ensures there is always a forum for discussion and exchange - eg. about the need for additional evaluations - or ensuring that existing studies are followed up. It does not have to be time consuming: the frequency and/or duration of meetings can reduce when there is limited business to address.

➤ **Evaluators are contracted for the whole programming period**

Examples of this approach are found in Italy (see box). It arguably delivers the primary advantage of ongoing approaches as highlighted by MEANS, allowing for “*good collaboration between the evaluation team and programme managers, which in turn favours a better appropriation of conclusions and recommendations*”. In addition, it can give coherence between the core studies and enable a rapid reaction when the need arises for *ad hoc* studies.

➤ **Maintaining the option to commission *ad hoc* thematic studies as and when required, in addition to the main evaluation menu**

This approach is seen in Ireland, where the National Development Plan Evaluation Unit has drafted and is implementing a programme of ongoing evaluation covering all Irish Operational Programmes and the whole National Development Plan. This consists of a series of thematic studies aiming to reinforce wider activity. It could also be one of the ways in which the commitment to ongoing evaluation in evaluation plans could be given expression.

### ONGOING EVALUATION IN ITALY

#### Ongoing evaluation in 1994-99

An ongoing approach to evaluation was adopted by the Ministry of Industry (now the Ministry of Productive Activities) in 1994-99 for the Objective 1 Multi-regional Operational Programme for Industry and the Retex, Resider, Konver and SME Community Initiatives.

#### How did the approach arise?

In 1994-99, there was no regulatory requirement to undertake mid-term evaluations. This exercise was first introduced to the Objective 1 CSF Monitoring Committee by the EC in 1998. The programme managers accepted the challenge and commissioned belated mid-term reviews. These proved useful in spite of data deficiencies and led to substantial modification of some financial and even strategic profiles.

As a result, the Ministry of Industry let a single evaluation contract to cover the programme up to the end of June 2002 (programme closure), requiring the evaluator to prepare periodic reports identifying management problems, bottlenecks and possible solutions. This approach was effective in integrating evaluation into programme management and made evaluation an *ongoing management tool*.

**How will ongoing evaluation be organised in 2000-06?**

For the equivalent 2000-06 programme (the National Operational Programme for Local Entrepreneurial Development), a similar approach has been adopted. Mid-term evaluation will be unpacked into a series of preliminary outputs, covering different and, at times, overlapping issues, defined in a detailed evaluation plan prepared by the chosen evaluators in consultation with the programme's secretariat. These preliminary outputs, updated as necessary, will be brought together in the final mid-term evaluation report that will be delivered to the EC by the end of 2003.

**Embedding the approach: relationship building between the evaluators and programme actors**

Significantly, the evaluation process is backed up by a structured programme of dissemination and interaction. The evaluators will:

- meet regularly with the Managing Authority to present the results of their ongoing evaluation activities;
- participate in meetings with the measure managers;
- attend Monitoring Committee meetings and present evaluation progress and results; and
- following submission of the interim report, will meet with the social partners and the regions.

In practice, it could be argued that ongoing evaluation is much more widespread than is commonly thought. **The distinction between programmes which explicitly embrace ongoing evaluation and those which do not may in practice be small** – ie. different points on the same relatively narrow continuum. There is a standard menu of four evaluations: the *ex ante*, mid-term, mid-term update and *ex post*. In addition, the Annual Implementation Reports provide a further regular opportunity for reflection. Given this context, every programme is in fact generating a significant quantity of analytical information. If coherent connections are made between these endeavours, then, even if none of the additional elements listed above are pursued, then any programme could arguably be considered to be taking an ongoing approach to evaluation. As much as anything, the distinction lies in attitude and interpretation rather than additional activities.

Table 4.2 highlights some of the benefits which are driving the interpretation of Structural Fund evaluation as an ongoing activity rather than as a series of self-contained exercises, and highlights some of the risks which need to be carefully managed.

Clearly, a careful decision may need to be taken about whether the more elaborate dimensions of ongoing evaluation are appropriate for any given programme. However, as stated above, this conception of evaluation can be pursued by all programmes, even those which confine themselves to the core studies, as long as actors see these studies as part of a coherent programme of investigation, and understand them as interacting on an ongoing basis with the programme, supporting and informing effective programming.

Table 4.2: The benefits and risks of ongoing evaluation approaches

|                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Benefits of ongoing evaluation</b>                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ➤                                                         | <b>Ownership and learning.</b> There is greater ownership at programme level of the evaluation process and the rationale for it. EC evaluation requirements are not experienced as an end in themselves, imposed from outside, but as a means through which programmes can raise and address questions of importance to them. What this implies is that the learning dimension of evaluation can become more established. Where the only evaluation activities are those required by the Commission, then the control and accountability functions of evaluation can more easily dominate perceptions. |
| ➤                                                         | <b>Agile problem-solving capacity.</b> Programmes have a prior agreement that <i>ad hoc</i> evaluations are a legitimate and useful way to address new issues. As a result, they are better equipped to respond quickly, not needing to wait until the next obligatory study will be done.                                                                                                                                                                                                                                                                                                             |
| <b>Benefits of ongoing evaluation using one evaluator</b> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ➤                                                         | <b>Depth of understanding.</b> The evaluator gains an in-depth understanding of the programme. This enhances the quality of insights which are possible.                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ➤                                                         | <b>Efficiency gains.</b> No time needs to be lost contracting evaluators for each separate study or familiarising them each time with the strategy, implementation mechanisms, data collection systems, etc.                                                                                                                                                                                                                                                                                                                                                                                           |
| ➤                                                         | <b>Trust.</b> A productive working environment emerges from continuity of co-operation between the programme and the evaluators. A relationship of trust grows which facilitates frank exchanges.                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Risks of ongoing evaluation</b>                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ➤                                                         | <b>High financial and resource costs.</b> Continuous evaluation may be more expensive than limiting activity to the four core studies, and might require too much input from programme actors who experience 'evaluation fatigue', reducing their willingness to participate fully in future studies.                                                                                                                                                                                                                                                                                                  |
| ➤                                                         | <b>Diversion from core evaluation requirements.</b> Evaluation could become so driven by programme level priorities and objectives, and more innovative methodologies, that those activities which are undertaken to meet the EC's core evaluation requirements may fail to meet EC expectations.                                                                                                                                                                                                                                                                                                      |
| ➤                                                         | <b>Diminishing returns.</b> There may be diminishing returns from the additional evaluation efforts, if they deliver little that is not already understood from the core studies. The larger, more complex, more innovative and/or less successful programmes might benefit most from additional evaluation effort.                                                                                                                                                                                                                                                                                    |
| <b>Risks of ongoing evaluation using one evaluator</b>    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ➤                                                         | <b>Loss of objectivity.</b> Where the same evaluator is working closely with a programme for a prolonged period, the perceived and actual independence of the evaluator may be at risk, undermining any studies.                                                                                                                                                                                                                                                                                                                                                                                       |
| ➤                                                         | <b>Lack of fresh ideas or specialist insights.</b> Using the same evaluator repeatedly can become 'stale'. More new ideas or specialist expertise might come from involving a range of evaluators.                                                                                                                                                                                                                                                                                                                                                                                                     |

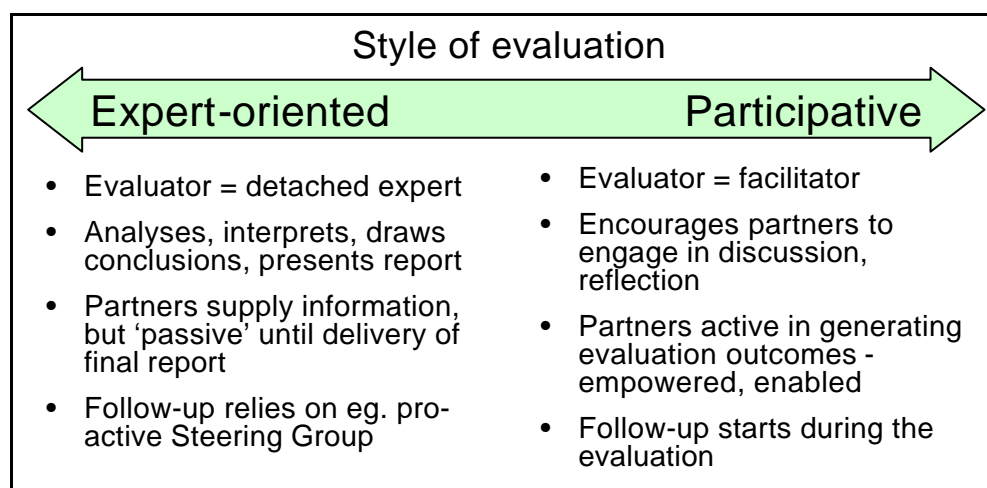
### 4.3 Style: choices about how to undertake evaluation

Over recent years, **Structural Fund evaluation has increasingly widely been interpreted as a learning opportunity**, not just an accountability and verification exercise. It is significant, for example, that the EC's synthesis report on the interim evaluations of the 1994-99 Objective 1 and 6 programmes was entitled: *Better management through evaluation*.<sup>11</sup> The study emphasised the potential for interim evaluations to be operationally and strategically oriented, stating that: "In many Member States, programme managers and policy makers...asked, on the basis of the evaluation findings, what kind of adjustments needed to be made, particularly with regard to the current interventions." In a further sign of its commitment, the EC selected 'Evaluation for quality' as the theme of its fourth international Structural Fund evaluation conference held in Edinburgh in 2000, which once again sought to disseminate lessons from international evaluation experience.

Part of the rationale for viewing evaluation as a learning experience is that Structural Fund programmes are **designed and delivered on a partnership basis, relying on the contribution and expertise of a wide vertical and horizontal partnership**. This poses challenges in terms of how best to facilitate a learning process and **ensure that evaluation findings reach those for whom they are most relevant**.

A further continuum can be imagined here, this time between what could be called 'expert-oriented' evaluation, and the increasingly widely discussed and practiced 'participative' evaluation (Figure 4.3).

Figure 4.3: A continuum of evaluation styles



At one extreme of **expert-oriented evaluation, evaluators are contracted to scrutinise a programme, its activities and systems in a detached way and to deliver appropriate conclusions** and recommendations. Programme actors, their systems and the projects which have been supported are the passive objects of the evaluation, and the expert is the locus of interpretation

<sup>11</sup> CEC (1999) *Better Management through Evaluation - Mid-term Review of Structural Fund Programmes: Objectives 1 and 6 (1994-99)*, p.5, Brussels.

and idea generation. The evaluator interprets evidence and draws conclusions, issuing a report about findings. It is then up to the programme to digest and follow up on the report, often after the departure of the expert, whose contractual obligation was limited to delivering the study.

At the other end of the continuum is **participative evaluation, where the key emphasis is placed on the expertise and knowledge of programme actors rather than the evaluator**. Here, the evaluator plays the role of facilitator rather than detached expert. Programme actors are encouraged, both singly and in groups, to *participate* in critical reflection about systems, behaviours and relationships. The dialogue about strategy, achievements and processes not only helps to understand programming achievements and processes more deeply, but also directly engages those actors who are best able to interpret patterns and initiate and implement change. Under this view of evaluation, the processes of interaction, reflection and learning are seen as equal in importance to the final evaluation report, and should facilitate self-discovery and self-diagnosis among programme partners.

**Between these two approaches lie several hybrids**. For example, towards the participative end of the continuum, there may be approaches which are participative to an extent, but do not empower actors in the way that a fully fledged participative evaluation would. More towards the expert-oriented extreme are approaches which rely on evaluators as experts, but where there are arrangements which ensure that the evaluation study is not detached from the programme. For example, there may be actors within the programme (eg. on a Steering Group) who are concerned with involving programme partners in: defining the issues the evaluator will address; verifying the expert's findings and establishing their relevance to the programme; and deciding what action should be taken as a result. Here, the learning is driven by programme actors, with the evaluator informing rather than facilitating the process.

It remains to be seen where programmes will fall this time on the continuum between expert-oriented and participative evaluation. However, it is possible that changes to the overall environment in which the evaluations are taking place this time will mitigate against the more participative, free-flowing approaches. **The profile of evaluation as a control mechanism has risen**. The accountability/transparency dimension of evaluation is more prominent, notably because of the performance reserve. In addition, the Commission has more influence than before over what the evaluations need to address and how this should be done. This may mean that the more flexible, responsive, learning dimension of evaluations, primarily benefiting the programme, may come to have less prominence. In particular, the control function may be at odds with a learning-oriented process of evaluation which asks programme actors to be open and honest in their responses, and to expose the difficulties and tensions which have been encountered. The overall outcome may be that, in aggregate, **the emphasis placed on evaluators as facilitators and evaluation as learning may decline**.

For those programmes interested in a participative approach, however, it may be useful to consider the **specific organisational requirements of this option**, since it is less familiar than the expert-oriented mode of operation. There are three pieces of key advice.



- **Build in sufficient time.** Interactive processes take longer: key stages include the evaluator becoming familiar with programme delivery mechanisms and partnership organisation, engaging in intensive dialogue, verifying impressions and finally disseminating conclusions. Time for consultation is critical. If all constituencies are not consulted, or if there is no time for those who are criticised to respond, then relationships may be damaged and conclusions perceived as ill-founded, counteracting any benefits.
- **Select facilitator-type evaluators.** Not all evaluators will be able to undertake a participative study. The 'style' of evaluators is influential in establishing trust and openness. In an audit role, the evaluator may need to work as a detached expert, processing quantitative data to generate irrefutable conclusions; in a participative study, a more effective role for an evaluator is as *facilitator*, involved in an active process of communication in which programme actors will in many cases be helped to generate their own insights and agree their own solutions. In this context, communication skills, including the ability to listen, and a highly developed awareness of the complexities of partnership-working, are as important as specific technical capabilities. Evaluators need to engage actors, inspire their trust and encourage reflection, and, at the same time, be seen as responsive yet impartial. In the Aquitaine 1994-96 Objective 2 interim evaluation, the measured and sensitive approach of the evaluators enabled actors to engage very quickly in a new process of dialogue and reflection, setting a positive precedent for subsequent evaluations, and encouraging additional thematic analyses to be undertaken.
- **Ensure programme actors understand the approach being taken.** Partners and stakeholders need to be aware that the aim of the evaluators is to facilitate a process of analysis, discovery and learning in which they are active participants, not just to create a product - a report - which imposes conclusions. Under this style of evaluation, it is counter-intuitive that the more successful the evaluator, the more programme actors will feel that they, and not the evaluator, have generated the successful outputs. This can lead to scepticism and, once again, undermine useful work.

#### 4.4 Scope of the evaluations: the aims to be addressed

In the current round, **the stated aims of evaluation studies are predominantly those presented in Working Paper 8**, unsurprisingly since a universal concern is to meet EC expectations. The overall aim of the mid-term evaluations, as summarised in the Working Paper, is *"to assess the initial results of the various forms of assistance and to make recommendations for any changes needed to ensure that they achieve their objectives"*.<sup>12</sup> It further disaggregates this into six more specific areas of activity. These are listed in the box below, organised under the two primary questions which can be derived from distilling these six aims.

---

<sup>12</sup> DG Regio (2000) *Working Paper 8: The Mid Term Evaluation of Structural Fund Interventions*, Commission of the European Communities, Brussels, 5 December 2000.

### AIMS OF THE MID-TERM EVALUATIONS

#### **Overall aim 1: Are programme strategies still relevant, or are adjustments now required?**

Rationale: Providing the evidence needed to make adjustments to the programme in order to ensure that the remainder of the programme maximises its economic effectiveness and efficiency.

##### *Specific aims*

- Verifying the continuing appropriateness of assistance relative to changes in the external environment.
- Reviewing the coherence and continuing relevance of strategic axes, priorities and objectives. Quantifying progress towards achieving them and assessing the likelihood of doing so.
- Assessing the quantification of objectives and the extent to which they have facilitated monitoring and evaluation.

#### **Overall aim 2: How well are programmes performing in terms of effectiveness and efficiency?**

Rationale: Fulfilling accountability and transparency functions, and informing current and future policy debates.

##### *Specific aims*

- Presentation of results, including against the indicators agreed for the performance reserve.
- Assessing integration of the horizontal priorities into forms of assistance
- Analysis of the adequacy of implementation and monitoring arrangements: have they facilitated or impeded performance?

Nevertheless, some programmes are pursuing additional, more *ad hoc* aims, for example identifying the Community added value of participation in Structural Fund programmes (see box). In addition, there is likely to be a small number of other *ad hoc* objectives, including evaluation to inform the assessment of particular domestic policy directions. Examples are the ‘volet territorial’ and rural development in France.<sup>13</sup>

### EVALUATING THE ADDED VALUE OF THE STRUCTURAL FUNDS

Evaluating Community added value is mentioned in Working Paper 8 as an element to be investigated. This involves identifying the added value brought by: (i) the additional regional development funding supplied by the Structural Funds; and (ii) the particular operational approaches encouraged by the Funds (including programming and implementation in partnership). This is intended to assist the EC, whose objective is to synthesise views on a macro level about the

<sup>13</sup> The ‘volet territorial’ is one of three major components of the *Contrats de Plan Etat-Régions* (CPER), alongside the regional and inter-regional components: it is a major innovation of the new 2000-06 CPER programming period.

significance and contribution of the Structural Funds, including for future cohesion reports. Underlining the importance of this ambition, DG Regio provided further detailed guidance on this area in the form of a Working Paper including definitions of Community added value, operational advice on how to address this issue and some proposed criteria and indicators to facilitate and perhaps enhance the comparability of the results of this process.<sup>14</sup>

Note that identifying Community added value would directly benefit not just the EC but also the Member States. This information would be invaluable and timely in helping to inform debates about the future of EC regional policy, ensuring that these are more firmly evidence-based. Depending on the perspective of the Member State, there may be several important outcomes:

- to provide evidence for the continuation of even relatively modestly sized Structural Fund programmes, where this is believed to be appropriate;
- to ensure that lessons are captured from Structural Fund experience and transferred into domestic economic development policy, should Structural Fund programmes not continue in many regions beyond 2006; and
- to inform the wider debate about simplification of the Funds.

France is among the Member States actively seeking to identify the positive contributions of community policies to regional policy delivery, including to decision-making processes (eg. partnership working). It is hoping that the evaluations will be able to enrich the debate and make arguments more concrete. This strand of evaluation is also of interest in Denmark, where there is a will to draw lessons from Structural Fund experience and adjust national policy instruments, where appropriate, to retain the beneficial aspects of the Structural Funds as they are phased out. Lastly, the UK is exploring options of undertaking a national evaluation of the value added of all Structural Funding, though this has not been directly linked to the mid-term evaluations.

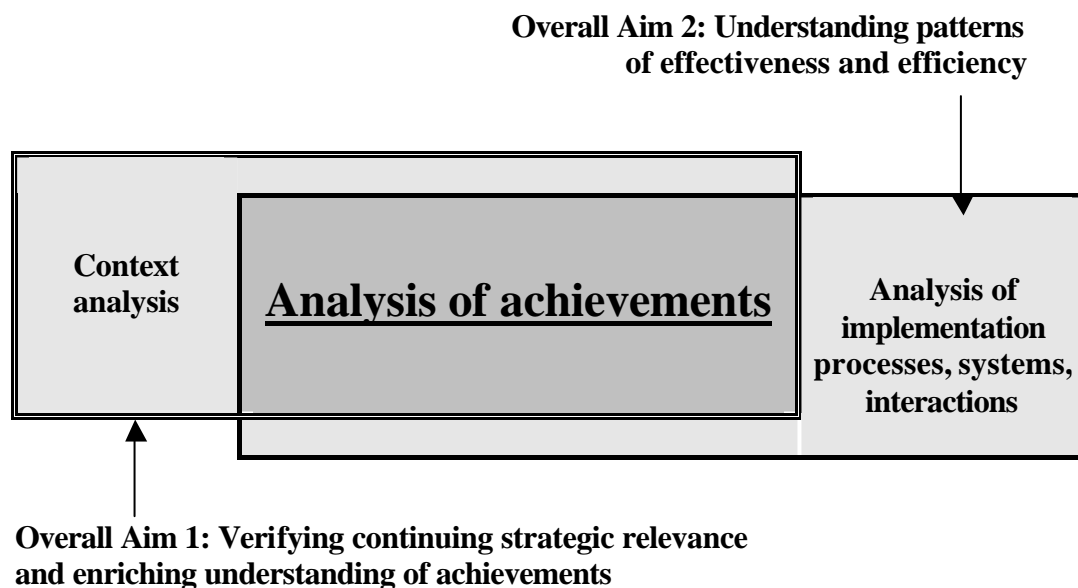
The questions which the evaluation studies aim to address are tackled through multiple streams of analysis. There are three main distinct but inter-connected streams involved:

- **achievement analysis**, at this stage referring to 'achievements' in the broadest sense - specifically, identifying outputs and results achieved to date and using this information to consider likely future impacts and make informed deductions about likely programme effectiveness;
- **context analysis**, identifying not only wider economic trends, but also institutional and policy changes; and
- **process and systems analysis**.

These three dimensions of analysis and the relationship between them are illustrated schematically in the diagram below.

<sup>14</sup> DG Regio (2002) *Community Added Value in the Context of Structural Policies: Definition and Evaluation Criteria*, Working Paper, Commission of the European Communities, Brussels, January.

Figure 4.4: The three elements of analysis and their relevance to the evaluation aims



The diagram emphasises the idea that **the analysis of achievements is central to the evaluation studies**, even at the mid-term stage when relatively little progress may have been made towards economic impacts. Achievements are not made in a vacuum, however, and, in order to draw meaningful conclusions about a programme (is it still relevant, is it efficient and effective?), it is necessary to consider it in terms of the context in which it is taking place and the processes and systems which are being used to deliver it. The diagram illustrates these two dimensions as being complementary to analysis of achievements. In addition, it underlines the idea that **analysing contextual change or processes and systems is not an end in itself, but complements and informs interpretation of programme achievements**.

In verifying the continuing strategic relevance of the programme, analysis of achievements is arguably usefully supplemented by context analysis. Assessing programme performance in terms of efficiency and effectiveness is, in turn, arguably most enriched by undertaking a critical analysis of the processes delivering these achievements. This deepens understanding beyond a simple analysis of figures, providing causal insights.

In practice, of course, the picture is more complex: the main aims of evaluations are divided into a range of more specific objectives and questions. Exploring this level in a little more detail, the matrix below provides one interpretation of which of the three streams of evaluation activity might contribute most to which of the evaluation aims set out in Working Paper 8. Where addressing a given issue involves more than one stream of analysis, there would be a process of integrating relevant insights to reach rounded conclusions.

The matrix is purely an illustrative interpretation. Different approaches to evaluation might place more or less emphasis, for example, on the potential of process and system analysis to contribute to the varied aims of evaluation.

Note, however, that addressing the horizontal themes is one of the few aims to which all three streams of analysis need to contribute (4.4.4).

Figure 4.5: The three elements of analysis and their relevance to the more detailed aims of evaluation<sup>15</sup>

|                                                                                                                                                          | Streams of analysis          |                      |                           |                                                                             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|----------------------|---------------------------|-----------------------------------------------------------------------------|
| Aims of mid-term evaluation                                                                                                                              | Context & strategic analysis | Achievement analysis | Process & system analysis |                                                                             |
| Core aims set out in Working Paper 8                                                                                                                     |                              |                      |                           | Integrating streams of analysis to generate conclusions and recommendations |
| ➤ Verifying continuing appropriateness of assistance relative to changes in external environment.                                                        | ↔                            |                      |                           |                                                                             |
| ➤ Reviewing coherence and continuing relevance of strategic axes, priorities and objectives. Progress towards achieving them and likelihood of doing so. | ↔                            | ↔                    | ↔                         |                                                                             |
| ➤ Quantifying objectives - quality of indicator system - effective, appropriate?                                                                         |                              | ↔                    | ↔                         |                                                                             |
| ➤ Assessing integration of the horizontal priorities into forms of assistance.                                                                           | ↔                            | ↔                    | ↔                         |                                                                             |
| ➤ Analysing the adequacy of implementation and monitoring arrangements.                                                                                  |                              |                      | ↔                         |                                                                             |
| ➤ Presenting results against indicators agreed for performance reserve.                                                                                  |                              | ↔                    |                           |                                                                             |
| Additional ad hoc aims                                                                                                                                   |                              |                      |                           |                                                                             |
| ➤ Assessing ‘Community Added Value’.                                                                                                                     |                              | ↔                    | ↔                         |                                                                             |
| ➤ Evaluating particular domestic policy directions                                                                                                       | ↔                            | ↔                    | ↔                         |                                                                             |

By tailoring a visual tool like this to a given programme, it could provide a useful aid to facilitate the assessment of bids from evaluators, for example. It would help to demonstrate how key aims were going to be addressed, where the focus of activity would be and that all the most important issues would be addressed appropriately in a coherent and balanced way.

The following sections examine the three key streams of evaluation activity in more detail, starting with achievement analysis, then considering context and process analysis.

<sup>15</sup> This diagram was informed by a more customised framework contained in the following document: ATI Nova - Ismeri Europa (2002) *Il disegno della valutazione del PON 'Sviluppo Imprenditoriale Locale' - Regioni dell'Obiettivo 1 (2000-2006)*.

#### 4.4.1 *Achievement analysis - the primary focus*

The **primary focus and core of most mid-term evaluations will be to quantify programme achievements** to date. Most are in a better position to establish this picture because of improved monitoring systems.<sup>16</sup>

In previous rounds of mid-term evaluation, insufficient attention may have been paid to achievement analysis in some studies, partly because programmes had not advanced sufficiently to have many outputs or results to report, let alone impacts. This is perhaps a weak response, and problematic in terms of fulfilling the accountability and transparency functions of evaluation. **Analysis of achievements arguably needs to be addressed more seriously now than in the past**, in particular to provide reliable evidence for performance reserve allocation, to feed into regional, national and EC aspirations to establish an overview of Structural Fund effects for political and accountability reasons, and to inform the process of ensuring that the maximum impact can be achieved from the current programmes, which, for many regions (especially Objective 2), may represent the last major tranches of Structural Funding.

**More ambition and focus is needed**, and many Member States are responding accordingly. There will be a **stronger emphasis on quantitative elements** in Denmark and France, for example, with greater efforts to quantify what has been achieved - even if not yet in terms of final impacts. This greater emphasis on quantification is not just for the benefit of individual programmes, but also beneficial where the national level is seeking to compile an overview. A meta-evaluation of the French mid-term evaluations of Objective 2 programmes was unsuccessful in the last round because of the lack of core quantitative elements in the studies produced.

Of course, different programmes have progressed to a different extent in terms of committing resources to projects and implementing them. Nonetheless, it will be **possible for them all to summarise *anticipated inputs (finance), outputs, results and impacts*** from projects which have been approved to date. **It may also be possible to identify *actual inputs, outputs and results*** to date, and to derive insights by comparing these actual achievements with programme commitments and overall targets.

It will be too soon at the interim stage to expect final impacts to be measurable (this applies to gross impacts, and even more so to net ones). Therefore, **it will not be realistic or appropriate to require impacts to be assessed directly and in detail**. A pragmatic compromise, and one which is methodologically feasible, is to **reflect on the realism and likelihood of attaining the impact targets**. As such, the Evaluation Unit of DG Regio is requesting an assessment of how likely it is - using available evidence at the mid-term stage - that programme impact targets will be met.

In establishing programme achievements, three broad types of quantitative information may be exploited by evaluators. The main one will be monitoring

---

<sup>16</sup> Bearing in mind, however, that presentation of monitoring data is not in itself sufficient to constitute an evaluation.

data (see Table 4.3). In addition, context indicators, and new data generated by questionnaires and surveys will be used in many cases.

Table 4.3: Scope of mid-term achievement analysis

|                                                                                | Financial inputs | Physical outputs | Intermediate results | Gross Impacts |
|--------------------------------------------------------------------------------|------------------|------------------|----------------------|---------------|
| Initial Targets                                                                | Ö                | Ö                | Ö                    | Ö             |
| Commitments made by approved projects                                          | Ö                | Ö                | Ö                    | Ö             |
| Actual progress to date (eg. from claims data)                                 | Ö                | Ö                | Ö                    |               |
| Estimate of current progress, based on other available data                    |                  |                  |                      | Ö             |
| Future projections based on progress so far                                    | Ö                | Ö                | Ö                    | Ö             |
| Verification and enriching of monitoring data through eg. a survey of projects |                  | Ö                | Ö                    | Ö             |

The exact range of data needed by evaluators depends on the methods they propose. Indeed, the assumption has been made here that it is bottom-up rather than macroeconomic evaluation techniques which will be used. It is arguably helpful if limitations of data are understood at the tendering stage, so that evaluators can propose realistic, costed solutions, rather than having to ‘fire-fight’ once the contract has been awarded. Sachsen Anhalt is working in this way, recognising that monitoring data may not be sufficiently complete to rely on entirely.

#### 4.4.2 Context analysis

**Context analysis** has two main aims. First, it is **undertaken to provide insights into how a programme has performed by placing achievements against wider trends** which may have influenced them (and attempting to establish how the programme may have influenced these trends). Understanding wider patterns is especially important where a programme has under-performed because of negative changes in the wider economic environment.

The second aim of context analysis is to help to **verify the continuing relevance of a programme**. This involves **examination of both changes in the economic context** (as above), **and the policy and institutional context**. It enables issues to be identified which mean selected policy areas merit more or less intervention by the programme.

Analysis of changes in the wider policy and institutional context, and the relevance of these to the programme, is relatively straightforward in many cases. It is especially important where there has been difficulty in committing resources because of a lack of available co-finance among partner institutions (requiring either a shift in domestic or Structural Fund priorities) and/or where elections have taken place which mean that spending priorities and

domestic economic development programmes have changed and the programme needs to adapt.

Economic context analysis, in turn, involves identifying and interpreting the evolution in baseline statistics (or ‘context indicators’) which describe relevant aspects of the economy in the eligible area. Any such analysis needs to go on to consider the potential impact on the programme of the trends identified. It is often challenging because of data deficiencies which impede the process of establishing what change has taken place over the recent past. In addition, establishing the nature and magnitude of any cause and effect relationships between the programme’s activities and the trends identified can be almost impossible, especially for the smallest programmes.

#### 4.4.3 *Analysis of processes and systems*

Depending on the issues identified, aspects to cover in an analysis of processes and systems could include, for example, the project selection process (including the criteria used, and the consistency and speed of decision-making and the ability to make strategic choices), the effectiveness and representativeness of committee structures, the functioning of programme-level publicity and communication plans and progress in adjusting systems to the new financial regulations. **Analysis of the effectiveness and efficiency of a programme’s processes and systems can be beneficial not just as an end in itself, but also because it informs understanding of a programme’s economic development achievements.** In particular, it would potentially provide insights into factors which influenced the effectiveness of the programme, and help to show how efficient actors had been in delivering those effects.

One of the mid-term evaluation aims set out in Working Paper 8 is “*analysis of the adequacy of implementation and monitoring arrangements*”. From this, it could be inferred that an analysis of systems and processes might be required of *all* programmes in the current round. However, this may not necessarily be the case, if demonstrating that achievements are on track could act as a proxy proving that implementation systems are working and that monitoring systems are adequate. The implication of this is that, in some contexts, evaluating processes and systems may potentially only be embraced when problems are identified which cannot be explained or addressed by other means (eg. attributed to contextual changes). In reality, however, there can be few programmes which are performing so well that giving no attention to this area is prudent. In addition, analysis of processes and systems potentially brings benefits in its own right.

Reviewing the plans of partner programmes, it is notable that, as in previous rounds of evaluation, the degree of emphasis placed on this dimension of evaluation varies significantly between them. Addressing process issues has in the past been relatively prominent in many UK, Swedish and French evaluations, for example, and played a lesser role in many Spanish and German studies. In the current round, the emphasis on considering process has grown in Italy, where first indications are that a primary role of interim evaluation is being seen as facilitating learning and informing adjustment of implementation processes. In Denmark, in contrast, there may be a declining



emphasis on considering process issues compared with previous rounds of evaluation. Given fixed budgets, some more resource-demanding qualitative approaches may be crowded out by the greater emphasis on quantification, even though qualitative aspects have traditionally been one of the strong points in the Danish context.

As long as they fulfil the required aims of the evaluations they undertake, **programmes have a choice about how much emphasis to place on system and process analysis** and how this should be undertaken. The amount of emphasis depends on a range of factors, three of which are explored here. **First, it is influenced by the prevailing vision of the purpose and place of evaluation.** At the most straightforward level, are actors open to this type of evaluation? Does it have a place in the established evaluation culture? Often, there will be more interest in processes and systems where a more participative style of evaluation is embraced. **The second influential factor is the institutional configuration** of Structural Fund programmes and, specifically, whether they have distinctive systems and processes to evaluate. **The third is whether programme systems are mature or whether there are innovative elements which could benefit from a ‘health check’.** The latter two factors are discussed in more detail below.

*a. Process evaluation: more likely where there are dedicated Structural Fund systems*

For process evaluation to be seen as useful, a Structural Fund programme needs to be seen by its stakeholders as **a clear entity, with distinctive policy objectives and delivery structures** – as ‘something to evaluate’. There also needs to be **scope for actors to change policies and/or modes of delivery in the light of evaluation findings**. These conditions are not met in all contexts.

To explore this, it can be helpful to return to the Structural Fund management continuum introduced previously in IQ-Net.<sup>17</sup> At one extreme are those **programmes delivered through dedicated systems** established solely and specifically for deciding on Structural Fund co-finance (which can be termed ‘differentiated’). Examples include programmes in Belgium, Denmark, the Netherlands, Sweden and the UK. Here, project appraisal is carried out by Structural Fund secretariats or other competent organisations or committees, while final project decision-making is typically the responsibility of a committee representing the Structural Fund programme partnership.

At the other extreme are **programmes using existing, domestic policy decision-making channels for Structural Fund resource allocation (characterised as ‘subsumed’)**, eg. Austria, Germany and Spain. Here, projects are appraised and selected using established domestic resource allocation routes, with Structural Fund budgets integrated alongside domestic ones, and decision-making often delegated to single competent agencies rather than reserved for dedicated Structural Fund committees. Falling between these two extremes, several Member States (eg. Finland, France and Italy) have

---

<sup>17</sup> Taylor S, Bachtler J and Rooney ML (2001) *Implementing the New Generation of Programmes: Project Development, Appraisal and Selection*, IQ-Net Thematic Paper 7(2), European Policies Research Centre, University of Strathclyde, Glasgow.

‘composite’ systems relying on domestic appraisal and decision-making systems but with some additional, Structural Fund-specific elements.

It is argued that **differentiated Structural Fund systems may lend themselves more readily to process-related evaluation than subsumed ones**. Since there are Structural Fund-specific policies, structures and channels, the programme to be evaluated has a clearer profile, and a range of distinct, dedicated structures and systems whose performance can legitimately be assessed and modified on the basis of evaluation findings.

**Evaluation of processes and systems fits less well with subsumed systems**, although it may still take place (as some Austrian, German and Spanish examples illustrate). Here, the interventions that constitute the programme belong to a complex, wider picture of domestic policy-making. There is no strong ‘programme identity’. **It is difficult to disentangle Structural Fund elements from the core operating systems of partner organisations, which are not within the remit of the Structural Fund study nor the Managing Authority to evaluate** or call into question. A further complicating factor here is that any criticism which might be raised about specific agencies could be highly political, and not even worthwhile in operational terms, since changing structures or shifting the relative weight of resources between funding agencies is not straightforward. In these contexts, one model would be for evaluation of Structural Fund-related operational issues to be confined to those parts of the system which are Structural Fund-specific. The exception would be where serious performance-related issues have been identified, when the Structural Fund evaluation might have more of a mandate to analyse and comment on specific operational issues in the wider sphere.

*b. Process evaluation: more likely where there has been system or process innovation*

Where **implementation mechanisms and instruments have been evaluated and improved in previous rounds of evaluation**, and could therefore be considered mature, it is **less likely that a detailed evaluation of processes and systems will generate significant benefits**. This is the case in Kempen, for example, where the programme management believe that past evaluations have already enabled the main management and implementation weaknesses to be identified and addressed.

In identifying **areas of operational innovation**, however, there have been several which have been widespread and may merit examination, including the new division between the Managing and Payment Authorities. Another more all-encompassing change has been mainstreaming the horizontal themes across Structural Fund programmes. It will be important to assess whether the themes have been successfully mainstreamed into the systems and processes through which projects are generated, selected and subsequently supported, the effect this has had, and the need and potential for further development.

At a more specific level, the introduction of **global grants in France** in the 2000-06 Objective 2 programmes provides a further specific example of an innovation which might benefit from examination. These instruments will be administered by specific agencies and organisations which have not previously had responsibility for managing Objective 2 Structural Funds. The novelty of

this arrangement involving delegated responsibilities means that some process-related evaluation is merited, looking at whether these delegated bodies have made sufficient progress with using the Funds, and whether they have been able to take decisions smoothly, consistently and rapidly.

Likewise, in various **Italian regions, the introduction of new instruments and procedures in the 2000-06 programmes** (some widespread and some region-specific) has led to a wish to ensure that their performance is reviewed in order that timely adjustments can be made where necessary. Among the diverse examples are evaluating the Integrated Programmes for Local Development (*Programmi Integrati per lo Sviluppo Locale*) and, in Lombardia, the Infrastructure Fund and a measure to improve the programming capacity of local authorities.

In Sweden, while initial implementation problems, including personnel constraints, delayed approvals, new regionalised systems, and the high volume of applications in the first year, have now largely been addressed, programme managers want these difficulties to be recorded by the evaluation. This is particularly because the problems may have some impact on programme achievements.

#### 4.4.4 *Horizontal themes - methodological choices*

The mid-term evaluations will provide an **important opportunity to take stock of progress in pursuing the horizontal themes**. This will be done with different aims in mind. In Sweden, for example, there is an ambition to identify best practice, in a domestic context, and potentially to generate some new ideas as well. In another programme, however, the aim will be to critically review the position of horizontal theme considerations in economic development programmes, and thereby to inform a debate about the extent to which they should be prioritised by the programme.

It has already been seen above (3.1.2) that Working Paper 8 has recommended a **horizontal approach to evaluating the horizontal themes** - ie. addressing them as an integral element of mainstream evaluation activity rather than as a separate, self-contained issue. This is the approach which most, if not all, programmes have decided to take.

Going beyond this organisational choice, **few programmes appear to have reflected much on how they will manage this element of the evaluation**. In many cases, it appears that programmes will be looking to their evaluators to provide solutions and guidance. However, while experience of evaluating environmental aspects of programmes is increasingly widespread, it is arguably the case that relatively few evaluators have appropriate direct experience of addressing the mainstreaming of gender equality.

Among the few known cases where the difficulties of evaluating gender equality aspects of programmes have already been acknowledged explicitly is Ireland, where a study has been commissioned to guide equality-related components of the various evaluations of the different components of the National Development Plan and Objective 1 CSF.<sup>18</sup> The study is driven by the

---

<sup>18</sup> Polverari L and Fitzgerald R (forthcoming) *Mainstreaming Gender Equality in Mid-term Evaluations: A Guide*, Research for the NDP Equality Unit, Ireland.

belief that a consistent and good quality response to this area is more likely to emerge if some preparatory work is done to provide signposting and methodological orientations.

More modest, but along similar lines, is the DATAR plan to compile a series of horizontal theme questions to be addressed by the evaluators in each French region. It is anticipated that this will generate a more consistent, better quality and more comparable response than leaving the issue up to each region to resolve independently. The focus of these efforts will be strongly aligned with the overall French mid-term evaluation objective of verifying the overall ongoing relevance and well-foundedness of the programmes at the mid-way stage.

The most ambitious approach to addressing the horizontal themes is currently found in a partnership between Scotland and Nordrhein-Westfalen, who will be comparing their horizontal theme analyses in order to compare their approaches and achievements and generate new ideas and insights (5.3.3).

There are arguably **three important elements to horizontal theme evaluation**, as underlined already in Figure 4.5 above:

- to identify changes in the wider **economic, institutional and policy context** which are relevant to the horizontal themes;
- to establish what **effect the programme has had** or is likely to have on the environment and gender equality (and, in the process, to review the appropriateness, completeness and relevance of the data which has been collected, given that some elements are novel); and
- to undertake a more **qualitative process and system evaluation** to establish whether the horizontal themes have been embedded into programming processes sufficiently to facilitate mainstreamed responses from applicants throughout the programming cycle: this final element is a strong reason for evaluations to review implementation systems, even where they have already been evaluated in the past at a more general level.

#### 4.5 Tools for evaluation

In undertaking their analyses, evaluators will be calling on one or more of a range of tools and techniques. The selection of methods depends on many factors, including: the aims to be fulfilled by the evaluation; the available budget; the size of the programme being evaluated (with macro-economic modelling being a more appropriate option for the largest programmes, for example); the information and data available; past experience with different evaluation techniques; the institutional context; and the desired style of evaluation (eg. participative or expert-oriented).

Among the key tools employed by mid-term evaluators are: analysis of monitoring data; project surveys; stakeholder interviews; macroeconomic modelling; and context analysis. Inter-programme benchmarking is also an interesting analytical tool, but involves comparisons which would tend to take place once evaluations had been completed, rather than being done as an integral part of an evaluation. Table 4.4 highlights the tools listed, showing which of the three streams of analysis discussed earlier (achievements, context and processes and systems) they arguably contribute to.

Table 4.4: Evaluation tools and techniques and their relative contributions

|                           | Analysis of monitoring data | Project surveys | Stakeholder interviews | Macro modelling | Context analysis | Inter-programme bench-marking |
|---------------------------|-----------------------------|-----------------|------------------------|-----------------|------------------|-------------------------------|
| Achievement analysis      | ↔                           | ↔               | ↔                      | ↔               | ↔                | ↔                             |
| Context analysis          |                             | ↔               | ↔                      | ↔               | ↔                |                               |
| Process & system analysis |                             | ↔               | ↔                      |                 |                  | ↔                             |

It is not the intention of this section to address evaluation tools comprehensively. However, given that it will be the task of programme actors to consider these tools in assessing bids from evaluators, some observations are made about several of them below, highlighting opportunities and difficulties presented by analysis of monitoring data, stakeholder interviews and surveys, and context analysis.

#### 4.5.1 Analysis of achievements: the role of monitoring data

##### **Deficiencies in monitoring data weakened many evaluations in 1994-99.**

Problems included incomplete and/or unverified data and absent or inconsistently applied indicator definitions. These deficiencies meant that significant evaluation resources sometimes had to be consumed getting to the starting point for evaluation by compiling what should have been basic background data.

As seen in previous IQ-Net research on monitoring, undertaken relatively early in the process of establishing new systems,<sup>19</sup> arrangements are being improved in diverse ways which should have made the ensuing monitoring data more robust and comprehensive than ever before. There is **widespread optimism that the improvements to monitoring systems should benefit evaluation**, allowing evaluators to build on a foundation of more reliable and complete basic summary data which programmes are compiling and maintaining, and thereby spend more time on value added tasks.

In particular, the mid-term evaluations will be an important test for the French PRÉSAGE monitoring system. The system should by now have bedded in, and be able to provide appropriate financial and physical data for programme-level evaluations. An important innovation of the system has been harmonising the definitions used in the collection and classification of data in different regions, potentially making it more aggregable and comparable.

The **main improvements to monitoring systems** relevant to evaluation, and which were discussed in more detail in the IQ-Net monitoring paper, included:

<sup>19</sup> Taylor S, Bachtler J and Polverari L (2001) *Information into Intelligence: Monitoring for Effective Structural Fund Programming*, IQ-Net Thematic Paper 8(2), European Policies Research Centre, University of Strathclyde, Glasgow.

- databases which make figures more accessible and manipulable;
- extension of the collection of monitoring data to cover both financial and physical data;
- fewer but more meaningful output, result and impact indicators, prioritising those which cast light on progress towards the main programme objectives (eg. in Germany);
- establishing clearer and more consistent definitions in advance to improve the reliability, aggregability and comparability of data collected (eg. the UK);
- verifying that indicators are quantifiable through *ex-ante* feasibility checks, and planning how and when data will be accumulated; and
- more vigorous data collection and inputting during programme implementation.

Fieldwork this time reflected confidence that systems had, in practice, improved significantly. However, it also showed that there are continuing difficulties in some contexts which might have adverse effects on the evaluation exercise. Some of these are set out here.

➤ **Technical difficulties associated with monitoring databases**

In Spain, while the ERDF database, FONDOS 2000, should bring benefits for the Spanish evaluations, there are reportedly still some teething problems at the regional and local levels.

Meanwhile, in Scotland, the single claims database, proposing to cover all Scottish Structural Fund programmes and collect monitoring information generated during project implementation, has been delayed. As a result, an interim claims database is being operated in Western Scotland, but, given its temporary character, it may not have the necessary functionality to provide optimal support to the evaluators.

In Flanders, while the Flemish financial monitoring system is already up and running, a new standard system for physical monitoring (procured from Wallonia) still has to be installed, which means that individual programmes are currently still compiling their own data and, indeed, in some cases working with paper-based systems due to technical problems.

At a more general level, the configuration of databases may complicate evaluation in some contexts, for example, where applications data is retained separately from claims (and monitoring) data, or where ERDF data is stored on a different system than ESF. Separating ERDF and ESF databases often provides the best technical solution given the distinctive nature of each Fund, but certainly means that evaluators have to familiarise themselves with more than one source of information.

➤ **Comprehensiveness of data available**

In Denmark, while the monitoring system held at the National Agency for Enterprise and Housing (NAEH) works well, physical monitoring data is only submitted by projects once they have completed implementation. This means that evaluators may have too little output and result data

available and may have to 'satisfice', or address the deficiencies using other techniques (eg. surveys of ongoing projects).

➤ **Robustness of data definitions**

Indicators in the UK have been defined more carefully this time, and steps taken to publicise the relevant guidance and resources, so that they can be applied by those implementing and reporting on projects. However, reporting problems persist. It is difficult to be certain of the reasons, but they may include personnel changes among project implementers, the complexity of projects, the delegation of claims preparation to staff who may be junior, inexperienced or only peripherally involved in the project, and problems of interpretation - illustrating that definitions cannot cover every eventuality. It may also be that the claims process has become too onerous for implementing organisations to dedicate sufficient time to it. Organisations do not always even have time to learn how to undertake the process effectively - even where this could lead to subsequent benefits.

➤ **There are few safeguards to protect against multiple counting of outputs, results and impacts**

Those processing Structural Fund claims aim to identify where multiple counting may be occurring, but data collection is not sufficiently sophisticated to identify all possible instances. Avoiding double counting is especially difficult where business development projects are involved. Several agencies with Structural Fund projects may be working with the same firm, and all reporting changes in the firm as impacts of their own intervention.

➤ **The accuracy and reliability of monitoring data has not necessarily been verified**

Claims checks look for internal coherence in what is being reported - but verification at this stage does not go much beyond this. Monitoring visits provide an opportunity to verify data, but they only take place on a sample of projects.

All of the weaknesses above present potential risks for evaluations whose core elements will be derived from monitoring databases. It may be beneficial for Managing Authorities to consider carefully the risks posed to the success of their evaluations by such deficiencies, and to explore in partnership with their chosen evaluators what could be done to address them. Of course, while it is important, monitoring data is not the sole source of information needed for any evaluation. It is necessary to generate insights going beyond reported outputs and results, for example through additional quantitative and qualitative survey and questionnaire work.

#### 4.5.2 *Additional survey work and stakeholder interviews*

In the past, monitoring data has provided the large majority of information used by evaluators in some contexts. In the current studies, **the EC is anticipating that monitoring data will be supplemented by stakeholder interviews and sample surveys of projects** (potentially addressing not only project implementers, but also final beneficiaries, where these are not the same, for example, SMEs or trainees).

The stakeholders are those who have an interest in the successful outcome of the programme. They include, but are not confined to, the members of key programme committees, notably the Monitoring Committee. Interviews with these actors capitalise on the understanding which they have of the programme and its systems, and their ability to draw on detailed knowledge to interpret patterns which are emerging from other aspects of the evaluation, whether they concern effects or processes.

Surveys of project implementers, in turn, potentially provide an important means to verify and supplement monitoring data. Of course, this may be especially valuable where there are deficiencies in the monitoring data (as described in section 4.5.1). In spite of considerable improvements, monitoring data deficiencies are likely to remain a challenge for many evaluators, and surveys are a primary instrument to address this. In addition, however, surveys have an important contribution to make in **generating information which is supplementary to and distinct from monitoring data**. This can serve several functions.

- **Verifying the reliability of monitoring data.** This is important from the point of view of demonstrating the reliability of evidence used in the performance reserve process. However, it is valuable more generally in ensuring robust results of evaluation, and, in particular, in assessing how realistic the targets set by individual projects are believed to be once projects are underway.
- **Addressing more complex evaluation questions.** Judgements are difficult to make about likely displacement effects of supported activity, their additionality, deadweight, etc, simply by exploiting monitoring data. A deeper understanding can only be gained by gathering more detailed information from a sample of implementers and beneficiaries, undertaking additional analysis and extrapolating findings across the programme's activities as appropriate.
- **Understanding patterns identified in monitoring data.** As important as identifying problems or examples of under-performance is understanding the reasons for them. Some of this information can be gathered through project implementer surveys.

There are various choices for how stakeholder interviews and wider project surveys can be organised to address process, activity, context or achievement-related questions. Face-to-face interviews based on semi-structured questionnaires arguably inspire spontaneous reflection more flexibly than a postal questionnaire, although they also introduce the challenges of greater interpersonal dynamics and have considerable resource implications if large numbers are to be undertaken. For programme stakeholders, this may be an appropriate investment. However, postal or telephone surveys may be more suitable where the aim is to gather sizeable samples of additional project-related information, especially where the emphasis is on obtaining quantitative information and/or qualitative insights gathered in a quantitative form using scoring or ranking systems. Telephone surveys again offer the interactive benefits of verbal interviewing and the opportunity to follow up on issues in a flexible way.



Interviews imply partners providing information individually to the evaluators who remain the privileged interpreters and analysts. A stimulating and cost-effective complementary approach, particularly for stakeholder interviews, is to exploit participative evaluation techniques including the organisation of workshops and focus groups to address specific issues.<sup>20</sup> This can raise the profile of priority themes and encourage the sharing of information, perspectives and ideas, leading to preconceptions being challenged and new insights emerging. The value of the opinions and ideas expressed by participants is judged by partners rather than by the more remote evaluator.

In the Western Scotland Objective 2 programme, the 1994-96 interim evaluation included a survey of programme participants which addressed programme management and delivery issues alongside questions about the effectiveness of interventions. The subsequent interim evaluation of the 1997-99 programme, undertaken by the same evaluators, built on this experience, supplementing the survey approach with working groups to identify ways forward on priority issues including the horizontal themes.

#### 4.5.3 Context analysis

We have seen before<sup>21</sup> that, with a few exceptions, including the Spanish regions, **establishing economic baseline indicators at the SPD preparation stage was challenging**. This was for various reasons, including difficulties in obtaining:

- statistics pertaining to eligible areas, where they are not coherent with standard statistical areas;
- data disaggregated in the required ways, for example, by sector or sex; and
- up-to-date statistics, with long time lags in some cases, eg. where data is derived from census results.

In some cases, special exercises were undertaken to fill statistical gaps at the programme development stage, eg. in some UK regions, primary research was commissioned to establish basic data sets.

It will be important to appreciate that updating baseline information in order to identify and interpret trends will be equally challenging. Some programmes and Member States have already responded in order to ensure that this part of the exercise can be undertaken successfully. For example, in France, the national statistical agency, INSEE, improved the quality of baseline economic and social data for eligible areas during the programme development phase. For the mid-term evaluation, given that most baseline data relates to the last census in 1999, INSEE will be computing adjustments for some data in an effort to provide an updated overview.

---

<sup>20</sup> Taylor S, Bachtler J and Polverari L (2001) Structural Fund Evaluation as a Programme Management Tool: Comparative Assessment and Reflections on Germany, *Informationen zur Raumentwicklung*, 6/7, pp341-57.

<sup>21</sup> Taylor S, Bachtler J and Polverari L (2001) *op. cit.*

## 5. USE OF FINDINGS

### 5.1 What are the outputs of the evaluation process?

As in previous mid-term evaluations, the key outputs of the evaluation exercises are reports, delivered at the end of the study. These will contain the findings and recommendations on the issues specified in the terms of reference. However, **other, more interactive outputs from evaluation are possible, some of which may be delivered before the study is completed.**

Under the participative approach to evaluation (as discussed above), one of the outputs of evaluation can be participant-driven learning, in which participants are informed of the continuing results of the study on an ongoing, interactive basis. At its fullest extent, such an approach can mean that the main function of the final report is to summarise what has already become known in the programme and, ideally, what the programme has decided to do about it, rather than to reveal conclusions for the first time.

The tension between accountability/justification and the learning aspirations of evaluations has been discussed above. An inherent danger is that the prominence of accountability functions - demonstrating what has been achieved (eg. for the performance reserve allocation) - can lead to the evaluation being seen as simply a report for the benefit of higher authorities. It can constrain open debate for fear of exposure before these authorities, thereby diminishing the learning potential at the programme level.

One way to minimise this tension is to generate different outputs for different audiences and purposes. One approach to this was used for the Graz Urban programme (Steiermark), where the evaluation drew a clear distinction between its learning and reporting functions. It was decided that open discussion could take place within the Steering Group – which in itself is a form of output - but that the only outcomes which would be reported formally to the Commission would be those which were of key importance to satisfying EC requirements.

### 5.2 How evaluation findings can be disseminated

In Working Paper 8, the Commission only recommends in non-specific terms *“that details of the outcome of the mid term review... should also be made public”* (although it does suggest using websites). Nevertheless, given the multiple objectives of the mid-term evaluations in this round and the large number of participants involved, a strong case can be made for wide distribution of the evaluation findings. In particular, if the evaluations are to result in real changes to the programmes, it may be important for dissemination strategies to be devised.

At this stage in the evaluation exercise, **only a few programmes have begun to consider how the evaluation findings will be disseminated. Nonetheless, it is already clear that dissemination responses will be varied.** For example, in Spain, final studies will be made available to those who request them, but there are no plans for publicising them more actively. In Italy, Lombardia is planning to disseminate its evaluation report through the internet and aims to

organise conferences and ‘open days’ to disseminate the evaluation results to the public. Similar techniques are being considered in the German *Länder*.

Dissemination will depend on the balance between ‘expert’ and ‘participative’ approaches to evaluation. In participative evaluation, a single report will not be sufficient if learning is being promoted within the programme. One useful approach for disseminating the lessons from evaluations among a particular community of interest has been actively pursued in France (see box).

#### **FRANCE: LESSON-LEARNING FROM EVALUATION**

Following each round of Structural Fund evaluations in France, a ‘day of exchange’ is held to draw lessons from the experience gained. The days of exchange form part of a wider structured programme of guidance and capacity building, funded under the National Technical Assistance Programme for Objectives 1 and 2. The events are coordinated by DATAR, which takes the lead in coordinating evaluation efforts.

An example of these events was the day of exchange held in 1997, based on the 1996/7 interim evaluations of Objective 1, 2 and 5b programmes. The day involved national experts, regional programme managers and evaluation experts and evaluators themselves discussing the most recent studies. The aims were to take stock of results, promote future evaluation activities and learn lessons.

In preparation for the day, a survey of those commissioning the evaluations was carried out to enable experiences to be addressed in a systematic manner. This addressed evaluation processes and the use made of results. In addition to the survey results, the published proceedings also contained a list of the evaluators used, with details of their experience and expertise.

The proceedings provided an overview of the evaluation activity undertaken, enabling each region to place its own practices in context, and to identify new ideas. Among the sections covered were:

- allocation of the evaluation function in each programme management team;
- how evaluators were selected and what studies cost;
- how studies were managed, including partnership structures used;
- the contribution of monitoring information to evaluability; and
- the realism of objectives and the impact of this on their results and reception.

Drawing on discussions during the day of exchange, the proceedings set out an explicit work programme to improve Structural Fund evaluation in France. Among the resolutions were:

- to undertake a study of the employment effects of Structural Fund interventions (now complete);
- to distribute a short guide to Structural Fund evaluation (done);
- to establish an Intranet exchange forum (underway); and
- to organise inter-regional workshops to support programme managers in writing their terms of reference for final evaluations.

Among the perceived benefits of such events are the following:

- a partnership-based approach, where all the main actors involved in evaluation meet to draw lessons and establish recommendations to move forward, ensures that resolutions reflect the interests of all key actors and have their commitment;
- those procuring evaluations can meet a wider selection of active evaluators with relevant expertise, and discuss their work with organisations which have already used their services;
- evaluation comes to be seen as a part of a wider process, in which follow-up, feedback and lesson-learning are integral parts; and
- all regions benefit from the lessons learned in individual regions.

Building on this experience, a series of training seminars has been scheduled up until June 2004 to assist those commissioning and steering evaluations of the 2000-06 programmes.

The need to consider dissemination may be more pressing where a single evaluation is covering multiple programmes. For example, in Flanders, where a single study, steered by a group with restricted membership, will cover several separate programmes, active dissemination to programme partners and administrators will be important. Indeed, the Kempen programme management unit (responsible for one of the programmes) views dissemination as one of the most critical parts of the evaluation process.

### 5.3 How the evaluation findings can be used

As already discussed, the mid-term evaluations are more complex this time around because they must fulfil a larger number of goals. The two key uses of the evaluation findings are in allocating the performance reserve and recommending improvements to the programmes. In addition, though, it is worthwhile highlighting how evaluation results are being used by programmes in comparative benchmarking.

#### 5.3.1 The evaluations and the performance reserve

**Distribution of the performance reserve and the mid-term evaluations are closely linked, but separate exercises.** The Commission has made the link explicit:

*“The mid term evaluation forms a key input into the decision on the allocation of the performance reserve. It provides the most up to date information available against the indicators agreed for effectiveness, management and financial implementation.”*

The point was also emphasised in Working Paper 4<sup>22</sup>, where it was stated:

*“It should be noted that the two processes, the performance reserve and the mid-term evaluation, are designed as separate exercises but underpin one another... The meeting point between [the two] will be the evaluator’s assessment in particular of the level of the targets*

<sup>22</sup> DG Regio (2001) *Implementation of the performance reserve*, Working Paper 4, European Commission, Brussels.

*attained at mid-term and especially those linked to the performance reserve, thereby contributing to the identification of successful programmes. In addition, the recommendations made in the mid-term evaluation report could direct the actual appropriation of the amount of the reserve allocated to these programmes towards potentially more successful programmes.*

Indeed, the quality of the evaluations themselves is one of the management indicators which will be considered in allocating the performance reserve. As a whole, in its consultation with the Member States, the Commission has reinforced the importance of how the mid-term results feed into the performance reserve.

Arrangements for allocating the performance reserve have been discussed in a previous IQ-Net paper to which partners are directed for more detail.<sup>23</sup> In brief, under the performance reserve, four percent of total financial appropriations are held back for distribution when the performance of each programme is reviewed in the first quarter of 2004. Performance will be assessed against three bundles of indicators - as above, programme effectiveness, management and financial implementation - which have been laid out in the Programme Complements with appropriate targets for the programme to achieve at mid-term (although many of these targets have been subsequently revised in advance of the evaluations and in consultation with the Commission). Member States are responsible for deciding on frameworks which will determine:

- how 'success' is to be defined in relation to the performance of the programmes against these criteria; and
- how (or whether) resources should be distributed between successful and 'unsuccessful' programmes and priorities.

These frameworks are to be discussed with the Commission before final decisions are reached on how the performance reserve is distributed.

On the first point, the Commission has argued in Working Paper 4 that programmes should be considered successful at mid-term if, under each of the three sets of criteria, they have achieved a 'value' of 75 percent or more of their corresponding standards. Many Member States have not yet set out their own procedures for how 'success' can be judged from the evaluation results, although some have already proposed their own approaches. For example, in Italy, it is proposed that the performance reserve will be given to programmes which have achieved 100 percent in the targets for at least five out of six indicators and satisfied at least one indicator under each broad category.

On the issues of allocation systems, to date, **nearly all of the Member States have made decisions on the overall systems for distributing the performance reserve** (see table below). In general, the mid-term evaluations are expected to provide the evidence for measuring all three sets of the performance indicators. Among the roles of the mid-term evaluations is to support decision-making about the allocation of the performance reserve by

---

<sup>23</sup> Bachtler J and Michie R (2001) *Programme Review: November 2000 – May 2001*, IQ-Net Thematic Report 9(1), European Policies Research Centre, University of Strathclyde, Glasgow..

providing (and verifying) relevant information about programme progress and determining whether the performance reserve indicators give an accurate impression of the management and delivery of the programmes. It is not the remit of the evaluations to make recommendations or judgements about the allocation of the performance reserve itself. Instead, this responsibility falls to the Member States, which will make proposals to the Commission which will then be discussed in inter-services consultation, before being agreed in negotiation with the Member States.

*Table 5.1: Methods for distributing the performance reserve*

|                           |                                                                                                                             |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| <b>Austria</b>            | Distributed within programmes.                                                                                              |
| <b>Belgium (Flanders)</b> | Distributed between programmes.                                                                                             |
| <b>Denmark</b>            | Distributed between measures within individual regions.                                                                     |
| <b>France</b>             | Distributed between programmes.                                                                                             |
| <b>Germany</b>            | Distributed within programmes.                                                                                              |
| <b>Italy</b>              | Distributed between programmes.                                                                                             |
| <b>Spain</b>              | Distributed between programmes.                                                                                             |
| <b>Sweden</b>             | Distributed between programmes.                                                                                             |
| <b>UK</b>                 | Distributed within territories – but decision has not yet been taken as to whether it will be within or between programmes. |

Another link between the evaluation and the performance reserve is the Commission's suggestion that any group established to set the framework for evaluation should also consider the distribution of the performance reserve – ie. there should be clear organisational continuity between the evaluations and the performance reserve. It recommends that the group should assess progress under all the performance reserve indicators and make recommendations on distributing the performance reserve on the basis of the evaluations and other analyses, which will then be discussed in consultation with the Commission.

Nonetheless, while the Commission has given clear attention to the issue in the mid-term evaluation guidance and consultation, **the importance of the performance reserve in Member State deliberations on the evaluations has been mixed.** For some Member States, the issue has been seen as of central significance in the design and delivery of the mid-term evaluations. This is particularly the case in the UK, where the contribution of the evaluations to determining the performance reserve allocation is interpreted as one of the primary goals of the evaluations. Similarly, the performance reserve means that Italian Managing Authorities are focusing more attention on the mid-term evaluations than previously – and that the exercise has become more relevant to a larger number of actors.

Nevertheless, for the most part, the performance reserve requirements of the evaluation have not dominated Member State debates over how they should be conducted. Indeed, the links between the performance reserve and the mid-term evaluation have been downplayed in some cases. The Spanish example illustrates the limitations of the evaluations to the performance reserve exercise. The main evaluation task of relevance to the performance reserve exercise is assessing programme performance in terms of how far programme

outputs, results and impacts have progressed relative to the targets that programmes set themselves. However, this information will be just part of the wider performance reserve assessment. A separate exercise led by national authorities will look at the broader picture of performance reserve indicators (including consideration of the quality of monitoring systems) to decide how resources should be allocated. To ensure true comparability of results, it is believed to be important for the performance reserve that the same organisation undertakes the assessment for all programmes. The mid-term evaluations, however, will be conducted on a programme-by-programme basis, as separate but parallel exercises.

Similarly, in Sweden, there has been little debate or unease about the performance reserve for several reasons. First, there has been considerable pressure in other areas, particularly in the light of the very high commitment levels in all the programmes, and thus the performance reserve has not been high on the agenda. Second, no real concern has been expressed that the programmes will not meet the stated targets. This is particularly true for the output targets which, given the high level of project commitments over the first two years, are already more than fulfilled in most programmes.

Regardless of the importance placed on the performance reserve in different Member States, all will have to undertake common, key tasks. These have been laid out by the Commission in Working Paper 8. They should:

- present the results on the agreed indicators and compare them with the targets originally set out in the Programme Complement (or recently revised);
- assess the quality and accuracy of the indicator data;
- comment on any changes in the targets since the programme was initiated;
- judge whether the effectiveness indicators for the performance reserve still cover at least half of programme expenditure; and
- decide whether the results for the indicators to date are a fair representation of programme performance as a whole.

All of these are relatively straightforward tasks, most of which would be required in addressing the other goals of evaluation. However, central to the completion of these tasks are decisions on which data to use in measuring progress towards effectiveness-related targets (equating to the 'analysis of achievements' task of evaluation, as discussed in section 4.4.1). Decisions on which data sets to employ (project application, claims or survey data) will affect the likelihood of programmes being able to meet their performance reserve criteria. An illustration of how this debate has been conducted can be seen in the case of the UK, as shown in the box below.

### UK: DATA SOURCES FOR PERFORMANCE RESERVE INDICATORS

Within the different territories of the UK, there has been discussion on what sources of data to use when assessing progress on the performance reserve effectiveness indicators. Evaluators have a choice of three sources: figures cited on the application forms; claims data stored on the monitoring databases; and actual data, as reported in the survey of sample projects during the course of the evaluations.

The choices have important implications. On the one hand, there is a 'rule of thumb' relating to the overstating of figures: application data are likely to have higher estimates of outputs and results than claims data, which in turn are likely to have higher figures than actual, survey-based data. In terms of the robustness of the data though, it is the survey data which are likely to be the most accurate. Consequently, in designing the evaluations, steering groups could face a trade-off between accuracy and ease of meeting performance reserve targets.

The choices made differ across the UK. For example, in England, it has been decided to measure progress on the basis of returns from the claims data and the anticipated benefits assessed at the project application stage. Performance against these targets are to be assessed on the basis of monitoring data up to the end of June 2003 with the evaluators making final updates on data up to the end of December 2003.

In Scotland though, consideration was initially given to whether estimates could be made on the basis of the *actual* data coming from the sample projects examined by the evaluators. These could be used to make estimates on progress towards the targets rather than claims data. This was prompted by a real desire among programme managers to measure the 'real' effects of the programme. However, there are clear risks in this approach, as 'real' data are likely to show lesser results than claims data. Moreover, the fieldwork involved means that the latest data that can be used will be Spring 2003: unlike the English regions, no updating was originally envisaged for the end of 2003. As a result, a final decision has been taken that performance reserve figures will be taken from claims data in September 2003. The mid-term evaluation will still have a role to play in assessing the reasonableness and methodology used in setting performance reserve targets and undertaking a sample check of claims data against survey data to verify the robustness of claims reporting procedures.

#### 5.3.2 *The evaluations and recommendations for improved programming*

In many ways, perhaps **the most important application of the results of mid-term evaluations will be in the impact that they have on the individual programmes.** As emphasised in the discussion of the aims of the evaluations (see section 4.4), the mid-term evaluations have greater scope in the current round to influence the direction of programmes. Any recommendations relating to necessary adjustments to the programmes (whether in terms of policy orientations or the distribution of budgets between headings) will be followed up. As a result, programme managers will be looking to the evaluations to produce the following:

- **an independent assessment of their activities as an external confirmation of performance:** for example, in Nordrhein-Westfalen, it is



hoped that the evaluation will provide a more accurate picture of the programme's contribution to the region's economic situation;

- **a deeper understanding of recognised problems and the identification of hidden or future problems facing the programme:** for example, in the West of Scotland, it is hoped that the evaluation will shed light on the causes of low levels of project applications under certain priorities;
- **inputs of new ideas into continuing discussion over how such problems could be solved:** for example, the Lombardia evaluation will review the 'fund for public infrastructure', set up as a solution to the difficulties in finding co-financing for public infrastructure projects, and assess its contribution to increasing project generation;
- **empirical backing for certain changes to programmes, especially virement between priorities and/or changes to the scope of priorities and measures:** for example, West Flanders is experiencing problems with its ESF allocation, while Turnhout is having difficulty in absorbing funding allocated to infrastructure – both wish the issues to be flagged up in the evaluation; and
- **analysis which is beyond the scope of Annual Implementation Reports and monitoring, such as confirming the direction of evolution of relevant baseline indicators:** for example, in the Swedish evaluation of its Objective 2 programmes, there have been discussions over whether to assess the baseline indicators.

The Annual Implementation Reports (AIRs) which will be compiled after the 2003 mid-term evaluations provide a formal mechanism to track the way in which evaluation results have been followed up. This is the intended approach in Spain (as set out in a resolution on the AIRs issued in March 2002 by the Spanish DG for Community Funds and Territorial Finance).

More widely, the mid-terms will result in lasting impacts if a more mature and realistic view of evaluation is taken which accepts two principles governing the use of evaluation results.

- **First, evaluations are unlikely to generate wholly new information or unexpected insights.** In earlier mid-term evaluations, partnerships were sometimes disappointed by the fact that evaluations told them things they felt they knew already. However, this is an important and useful function of evaluation - confirming impressions, formalising them and potentially providing better understanding of the causes of problems and their possible solutions.
- **Second, it should be recognised that some issues of importance to the programmes are beyond the scope of a mid-term evaluation.** They are limited exercises with a range of complex and often highly specified requirements. Some key issues for programmes may require more detailed analysis than is possible given the time and resources available, as for example, the interaction of state aids regulations and the Structural Funds. Nevertheless, the evaluations can be useful in at least identifying issues for future discussion and perhaps later analysis.

### 5.3.3 *The evaluations and comparative benchmarking*

One final use that can be made of the evaluation results is to compare performance between different programmes. Programmes can benchmark their performance against that of other programmes by comparing the results of their evaluations against several criteria. This approach has several benefits for programme managers:

- **relative performance:** it highlights where individual programmes are performing relatively well or have room for improvement;
- **source of good practice:** it draws out a larger range of good practice from which programmes can learn;
- **common problems:** it identifies common problems in the new programming round for which joint solutions can be found; and
- **foundation for future cooperation:** it can form the basis of longer-term sharing of experience between different programmes.

Benchmarking will affect all stages of the evaluation. At the start of such an exercise, if it is to be carried out effectively, consideration will need to be given by the participating programmes to designing evaluations that are comparable. For the mid-terms, evaluations are not being explicitly linked across different regions except where a single evaluation is covering multiple programmes. However, at the end of an evaluation, some programmes are planning joint conferences and seminars to discuss the results and the implications for the separate programmes.

Examples of inter-regional benchmarking are limited, but are worth highlighting, as they offer ways of maximising the value of the evaluations. For example, the national evaluation of the Danish Objective 2 programme is planning for the first time to benchmark the five eligible sub-regions. At an international level, the outstanding example is the cooperation planned between Nordrhein-Westfalen and the Scottish Objective 2 programmes, as described below.

## **NORDRHEIN-WESTFALEN AND SCOTLAND: BENCHMARKING EVALUATION RESULTS**

### **The background to cooperating on evaluations**

As an outgrowth of cooperation within the IQ-Net consortium, Scotland and Nordrhein-Westfalen (NRW) have been exchanging information on Structural Fund issues. This informal cooperation was given greater impetus by comparative research by Scottish and NRW officials on a range of economic development issues, including: land renewal and environmental reclamation; sustainable development and environmental technologies; and use and management of the Structural Funds. With the prospect of the mid-term evaluations, it was decided to expand on the latter area. Both territories realised they could benefit from a more structured exchange of experience in Objective 2, using the evaluation results of the one German and the three Scottish programmes.

### **Deciding how to take joint evaluation forward**

Linking the two evaluation exercises together had considerable appeal to both territories. The evaluations could achieve a deeper understanding of their respective issues through collaboration. Such cooperation would be enhanced by the differences between the territories (eg. Scotland and NRW have different approaches to implementing the Funds) as much by their similarities (eg. in terms of programme budgets).

Several models of cooperation were considered. It was suggested that both regions could work towards a close alignment of their evaluations, even having a common evaluation with a common set of contractors. Eventually, a less ambitious proposal was selected. Two workshops are now planned as part of the mutual evaluation exercise for the evaluators and steering groups in all the regions. The first will take place at the start of the evaluation process to agree on a work programme and joint evaluation methodology (scheduled for the autumn 2002). The second will be held a year later for the presentation and discussion of the results.

### **Giving focus to cooperation by concentrating on the horizontal themes**

Evaluation results will be shared on a variety of topics. It is planned that a synthesis chapter will be prepared, drawing out the main findings from comparing the programmes. Suggestions for further thematic concentration were made, including: community economic development; SME innovation; and capital access for SMEs. However, it was decided that the particular focus should be the horizontal themes of sustainable development and equal opportunities. The evaluators in Scotland and Nordrhein-Westfalen will work together to develop a common methodology to enable comparison between the different programmes, drawing out the strengths and weaknesses of the approaches taken to the horizontal themes in each programme.

Rather than being an *ad hoc* addition to the evaluations, it has been placed explicitly within the terms of reference for the relevant programme evaluations. This is underlined by the fact that evaluators are expected to quote costs for undertaking cooperation with the evaluators of the other programmes.

## 6. CONCLUSIONS

The mid-term evaluations of the 2000-06 programmes are due to be delivered at the end of 2003. **Over a year and a half before the deadline, these studies were already preoccupying Structural Fund actors across the EU.** This - perhaps exceptional - preparedness is not in practice surprising, since, as highlighted by this paper, **there is a great deal at stake** for every programme this time. The outcomes of the mid-term studies have real resource implications, feeding into the allocation of the performance reserve fund, which will have clear financial - and, potentially, prestige-related - impacts.

In addition, the evaluation task has been **more specifically and tightly defined than before through the Structural Fund regulations** and additional EC guidance, such that there are strict parameters within which they must be delivered, and which should help to deliver more consistent outputs. Notably, they now have to be completed by an immovable deadline and satisfy specific quality thresholds which will be verified by the EC. They have also become more extensive in terms of the aims to be achieved.

An additional contextual factor is that the studies do not in practice have a single set of aims, but are responding to the multiple agendas of different stakeholders, from the EC through Member State authorities to the diversity of actors at the programme level. The implication is that **they will have to be relatively extensive, methodologically rigorous, carefully co-ordinated and professionally managed exercises.**

This may be the round of evaluations where core parameters have most clearly been laid out in advance. Nonetheless, as the overview of preparations so far has demonstrated, **the mid-term evaluations cannot be seen as either a routine or standardised exercise.** They are not routine, in the sense that **they constitute 'special operations' which are additional to, although closely linked with, the ongoing functioning of the programmes.** They will require mobilisation in order to establish the necessary structures and design, manage, undertake and respond to an appropriate set of tasks.

Even though the formal requirements of evaluation have been more tightly defined than before, the studies also cannot be described as standard, in that **the responses to core expectations will be shaped by a range of unique parameters.** These include: the programme to be evaluated (including not just its policy mix but also its magnitude); the institutional and governmental system in which it is implemented; the economic and policy context in which it operates; the background, experience and approach of the individuals involved; the available resources; and, relating more specifically to evaluation itself, the available evaluator pool; the selected evaluators; prevailing philosophies of evaluation (including those reinforced through previous Structural Fund evaluation activity); the balance between the multiple aims of evaluation; and, available programme data and the influence of this on methodological choices.

Although they have made good progress in their preparations, most programmes are in reality at an early stage in the evaluation process, still planning the exercise and going through the formal stages to contract evaluators and establish structures. At the time of drafting this paper, **many**

**decisions about the evaluations had still to be taken - and, of course, the actual execution and follow-up of the studies also lay ahead.** To enable contributors to the mid-term evaluation exercise to position their efforts, **this paper has highlighted the distinctive features and requirements of the current studies, and reported on the preparations which have been made so far to undertake them.** The review has addressed the motivations and contributions of actors at EC, Member State and programme levels, and drawn out commonalities and contrasts in terms of the organisational choices which have been made.

Among the key observations from the review is that, while planning is going well at an outline level, and programmes are potentially in an excellent position to succeed in the evaluation challenge which awaits them, **it may be that deeper consideration of some of the tasks and choices which lie ahead - and some of the risks which might present themselves during the exercise - could now be beneficial.** Taking just one example, steering groups have been established to help to improve the quality of evaluation exercises, but further planning now could help to optimise their contribution. Exploring their role quickly starts to unfold a whole range of questions to be addressed, and potential risks which, at worst, might impede the ability to meet mid-term obligations:

- What will their tasks actually consist of and how exactly will they undertake them?
- How will they compare bids from evaluators in a systematic way to choose the best option in terms of value, quality and feasibility?
- How will they judge that the studies finally submitted have satisfied the relatively qualitative MEANS criteria? And what will be done if the study fails to meet expectations?
- What mechanisms will they in practice use to communicate progress in the study to the wider partnership?
- What will happen if there are differences of opinion on the steering group, or if the working relationship with the chosen contractor breaks down?

Given that programmes are still in the process of making choices, it may be useful at this stage, rather than unpacking the detail of the choices available to programmes and the steps to be taken to ensure good management, to draw the paper together by making sense of the rich tapestry of obligations, aspirations, constraints and opportunities which will be shaping the design and conduct of the mid-term evaluations. To do this, **a set of diagrams can be used to explore how partners will make their evaluation choices** by linking together:

- the overall vision of evaluation;
- the constraints and obligations which influence this vision; and
- the evaluation choices which partners can make as a result.

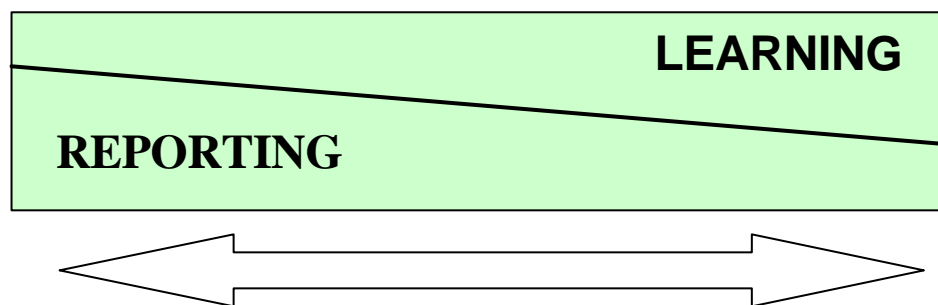
At the heart of preparing for the mid-term evaluations is a vision about what the evaluation is expected to achieve. As this paper has suggested at different

points, there is a continuum (illustrated in Figure 6.1) between two prominent, co-existing **visions of evaluation**: ‘reporting’ and ‘learning’.

- Where **reporting** is more important, the principal goals of evaluation will be to fulfil accountability and transparency requirements and provide information to a range of other actors (such as the Commission). The aims of evaluation, as well as its outputs, become focused on external, regulatory requirements with few additional aims contributed by the programmes themselves.
- Where **learning** is the main task, evaluation can allow key programme issues to be opened up and provides opportunities for exchange and development within the programmes. By concentrating on the benefits which programmes themselves derive, evaluation can be concerned not just with generating results, but the ways in which those results can feed back into programme operation and development.

The potential tension between these two functions has been a recurrent theme throughout the paper. Different programmes occupy different positions on the continuum, but wherever they are, the nature of the current framework means that all programmes are pursuing both reporting and learning aims: the difference lies in the emphasis placed on each and the balance between them.

*Figure 6.1: Visions of evaluation*

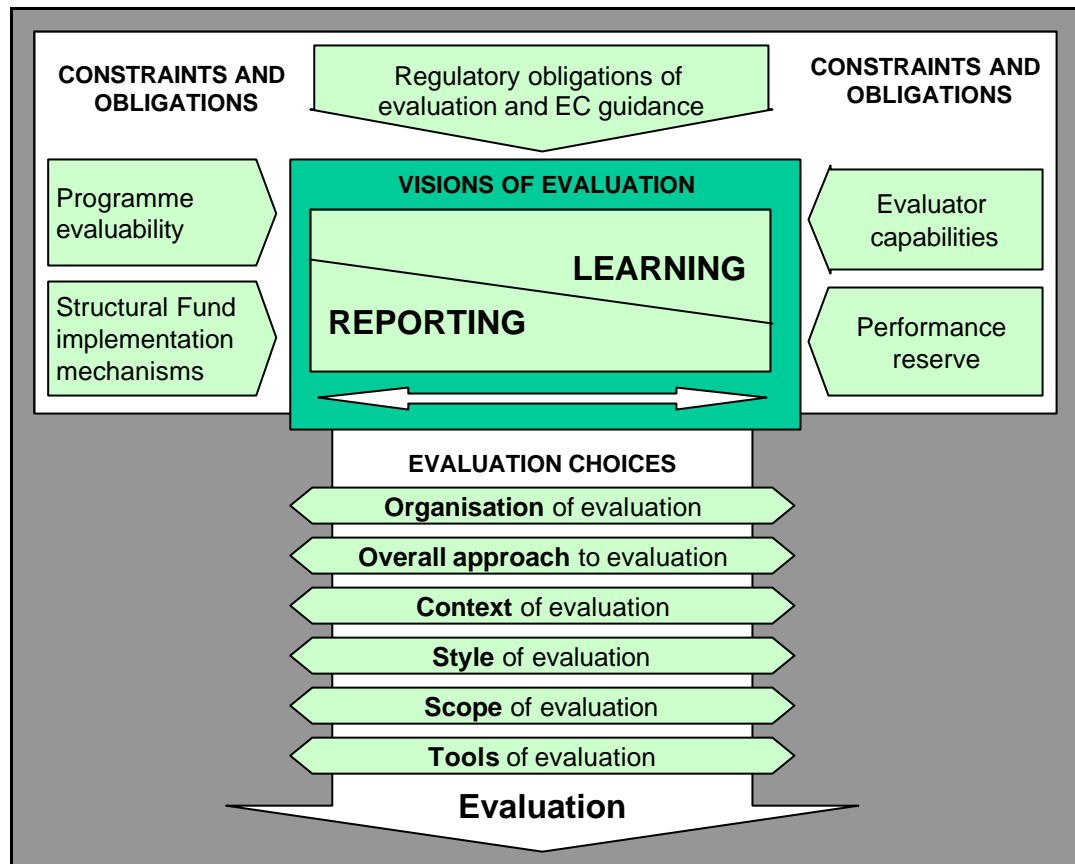


This balance between the reporting and learning functions of evaluation is shaped by the different issues covered by the paper, including the overall context of the mid-term review, how it will be organised in different Member States, the design and undertaking of evaluation, and the ways in which the findings will be used. Moreover, a programme’s position on this continuum can, of course, change over time, and will at all times be the aggregate view of those with an influence on the process, rather than necessarily a unanimous view. (For example, recent lively debates in Austria about the role of Structural Fund evaluation have helped to place Austria’s position and show how a single State can contain a variety of different attitudes.)

The course of the mid-term review in each country will be governed by this balance of functions. Taking the continuum illustrated above and placing it at the heart of a diagram describing how evaluation choices can be made (Figure 6.2 below), the different options for meeting the challenges and exploiting the opportunities in evaluations become clear. At the top of the diagram, the attitudes to evaluation held by programmes are influenced by a series of constraints and obligations, acting on all programmes but in a variety of ways.

These in turn will shape the way the programmes respond to the specific evaluation choices available to them.

Figure 6.2: Factors affecting evaluation choices



How do prevailing attitudes to evaluation emerge? Exploring the diagram, it can be seen that a range of factors influence the relative importance of reporting and learning, either by *limiting* particular responses, or by *encouraging* them: ie. the **constraints and limitations** of the programme. These include:

- the regulatory context of the evaluations and the role of EC guidance (see Section 2);
- programme ‘evaluability’ (such as the quality and robustness of monitoring systems and data gathering, as explained in Section 2);
- Structural Fund implementation mechanisms (particularly whether they are ‘differentiated’ or ‘subsumed’, described in 4.4.3);
- evaluator capabilities (3.5); and
- the performance reserve requirements (5.3.1).

The table below illustrates how these factors can influence the relative emphasis placed on reporting and learning functions in evaluation.

*Table 6.1: Factors affecting the balance between the learning and reporting dimensions of evaluation*

| Reporting                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Learning                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>EC guidance:</b> where EC expectations of evaluation are extensive, then simply demonstrating to the EC that all of these have been met may become the first priority, and consume all available resources.</p> <p><b>‘Subsumed’ Structural Fund implementation mechanisms:</b> where programme delivery is largely ‘subsumed’ into the mainstream operations of a range of funding agencies and government departments, rather than being undertaken through dedicated structures, then there is only a limited mandate for learning-oriented evaluation.</p> <p><b>The performance reserve:</b> the introduction of the performance reserve mechanism may mean that greater emphasis now has to be placed on the reporting function of evaluations, even in systems which might otherwise have been positively disposed towards learning-oriented evaluation.</p> | <p><b>Evaluability:</b> where programme targets are lacking or monitoring data poor or insufficient, then it may be necessary to substitute more robust analysis with work to improve processes and systems and thus increase programme evaluability.</p> <p><b>‘Differentiated’ Structural Fund implementation mechanisms:</b> where programme delivery is undertaken through dedicated structures, then the evaluations may provide a key opportunity to examine the functioning of these structures and make improvements.</p> <p><b>Evaluator capabilities:</b> in some contexts, the learning dimension of evaluation has grown as a result of precedents set by evaluation studies undertaken by enthusiastic proponents of interactive, learning-oriented evaluation.</p> |

Finally, the **evaluation choices** made by programmes will be influenced by the prevailing vision of evaluation, as shaped by the constraints and obligations under which they are undertaken. As the diagram shows, it can be seen that there is scope to make choices about a range of different aspects of evaluation:

- the **organisation** of evaluation as a whole: as, for example, in terms of whether a ‘one evaluation, one programme’ approach is being taken, or if multiple programmes are being covered by a single evaluation (as considered in 3.1);
- the **overall approach** taken by programmes: particularly the distinctive contributions of different actors to the definition of the scope and methods of evaluation (4.1);
- the **context** of evaluation: especially whether the evaluations are being conducted as self-contained, periodic exercises or part of more responsive continuous evaluations which can span the lifetime of the programme (4.2);
- the **style** with which evaluation is carried out: such as whether the evaluations are ‘expert-oriented’ or involve more ‘participative’ approaches (4.3);



- the **scope** and content of evaluation: with respect to the mixture of aims and objectives in the programme, the importance of thematic or cross-cutting issues (3.1.2 and 4.4) and the balance between an analysis of programme context, programme achievements and programme processes and systems (4.4) ; and, lastly,
- the **tools**, techniques and methodologies used for evaluation (4.5).

Taking the more specific example of the *style* of evaluations, a more reporting-oriented context might favour an ‘expert-oriented’ evaluation, where evaluators scrutinise a programme, its activities and systems in a detached way to deliver a report containing conclusions and recommendations. In contrast, a learning-oriented context might be more likely to favour ‘participative’ evaluation, where the evaluators act as facilitators, encouraging the engagement of programme actors in an active process of debate and reflection throughout the evaluation process.

In conclusion, more than on previous occasions, the outcome of this evaluation will be instrumental in influencing the allocation of resources, the content and form of programmes and the processes by which they are delivered. Given the investments that many regions have made in preparing for the evaluations, the mid-term review will be a good opportunity to judge the value and relevance of the new programmes. This paper has drawn attention to the range of choices available to programmes at every stage of the process, the complexities of the different tasks to be undertaken, and the real opportunities for creative engagement and maximising the potential of the evaluations. It has also highlighted tensions, which must be managed by programmes, between reporting and accountability functions of evaluation, which appear to be rising in prominence, and the function of evaluation as an active learning tool at the programme level.