TOWARDS A METHODOLOGY FOR EVALUATING THE EFFECTS OF MEASURES TAKEN TO IMPLEMENT EU ENVIRONMENTAL LEGISLATION

INTRODUCTION

At the third meeting of the REM Steering Group on 7 April 2000, the IEEP project team was asked to develop a methodological framework that Member States could use to evaluate the effects of measures taken to implement EU environmental legislation. This framework could then be 'piloted' in relation to different sectors at a workshop for officials and practitioners to be held in September.

In approaching this task, the IEEP team has been concerned not to reinvent the wheel, but to build on approaches that are already established, and in use in the EU. One important source for such approaches is the six-volume MEANS collection *Evaluating Socio-economic Programmes*, which was published in 1999 by DG Regional Policy in order to improve and promote evaluation methods, in particular in relation to the EU Structural Funds. Approaches developed by the MEANS programme have also influenced Commission guidance on evaluations in other sectors, such as EU agri-environment measures.

Two important lessons have emerged from studying the MEANS collection. The first is that evaluation methodologies are still being developed, and it is therefore unwise to be too dogmatic in seeking to identify a 'best' approach. 'Methodology is defined as the 'science' of the construction of methods…In reality, however, professional (evaluation) practice is still closer to expertise than to science. The word methodology is therefore somewhat ambitious at this stage' (1).

The second lesson is that there is no single approach that would be equally applicable to all types of environmental measures in all circumstances. Volume 3 of the MEANS collection (*Principal Evaluation Techniques and Tools*) describes no fewer than 23 different approaches to

- structuring an evaluation
- observing changes in the field
- analysing data
- making evaluation judgements.

Together, these constitute a tool box that can be used to develop a variety of methodologies tailored to suit different evaluation questions and types of environmental measures, and which can take account of different constraints in terms of budget and timescale. Section 1 of this paper suggests that a screening and scoping process is necessary to identify the most appropriate approach to evaluating the effects of different types of measure.

Nevertheless, there are a number of generally-applicable steps that can be taken to improve our knowledge of the effects of environmental measures. These relate principally to the collection of appropriate data and the identification of indicators needed to establish the direction and strength of causation - ie to improve our

knowledge of whether and how far changes in the state of the environment can be attributed to a particular policy measure. These basic requirements are discussed in Section 2.

1. TAILORING THE APPROACH TO THE CIRCUMSTANCES

> 'To know whether an evaluation tool is appropriate in the context, it is necessary to examine the object to be evaluated, the available data, the evaluation budget, and the deadline for presentation of the results. Choosing a tool without taking all these factors into account could have serious consequences: absence of conclusions, or irrelevant conclusions, deadlines that are not met, conclusions that are not credible etc.' (2)

1.1 Different evaluation questions

Although the current REM project is centrally concerned with evaluating the ***effects*** on the state of the environment of environmental measures, there are several other types of evaluation questions that can be asked - and might be in the future in the light of a number of current EU initiatives. Each has different implications for data collection and the choice of methods.

1. Assessing *effects* on the state of the environment requires the collection of potentially a very broad range of data to identify both intended and unintended consequences of the measure on environmental quality and resource use;
2. Assessing *effectiveness* in relation to the objectives of the measure requires a cl arification of objectives (which may not be explicit), but a more limited scan of the range of effects than in (1), determined by the nature of these objectives;
3. Assessing *cost-effectiveness* requires in addition to (2) information on costs directly attributable to specific policy outcomes and impacts;
4. Assessing *effects on progress towards sustainable development.* There is a developing debate on whether evaluation should cover wider impacts on social and economic as well as environmental issues. Such an approach could require the collection of a very wide range of information and the involvement of a broad range of stakeholders.

4.2 Different types of environmental measures

Environmental measures come in a wide variety of forms. They differ *inter alia* in relation to

- the nature of the environmental issues they address;
- their target sectors or geographical areas;
- the element in the DPSIR chain on which they are focused.

These differences influence the choice of an appropriate evaluation methodology, as follows:

- Where causal links are few and predictable, standard *models* may be derived from examining a small number of case studies. Sufficiently regular relationships between changes in pressure variables and state variables enable the identification

of coefficients which may be generally applicable, and which can be used to assess effects in different Member States without the need to collect empirical data in every case. Examples of such measures include:

- the impact on air quality and blood lead levels of specific reductions of lead in petrol;
- the impact on CO2 emissions of changes in vehicle technology and the pattern of new car sales;
- effects on nitrate leaching of changes in the timing and quantity of fertilizer applications.

Conversely, where implementation chains are long and policy players are numerous, the use of models will not be appropriate. In any event, the assumptions on which models are built always need to be tested and refined on the basis of initial empirical observation.

- Where the application of a measure is *differentiated* geographically or by target sector, within or between Member States (or both), comparative case studies can be used to help identify causal relationships. Examples include measures which identify designated/non-designated zones, or products; or registered/non-registered industrial sites etc.

- Where a *target sector is small* (eg farmers in nitrate vulnerable, or environmentally sensitive, zones) the effect of measures may be established by in-depth interviews. This is not possible where a policy has universal application (although focus groups or panels may be used – but these are technically difficult and expensive to establish).

- Where the *link between a policy and its impact on the environment is too diffuse* or extended – as with framework Directives or Directives which establish procedures only - a measure may not be 'evaluable' at all. Alternatively, it may be more practical to focus an evaluation on immediate outputs and outcomes as a rough proxy for ultimate impacts (see Figure 1). For example, we know that reductions in the production of ozone-depleting substances in a particular Member State (outcome) will eventually have some beneficial effect on the level of stratospheric ozone, without being able to compute exactly how much, or when.

1.3 The need for screening and scoping

In the light of these factors, it is necessary to undertake an initial process of screening and scoping, similar to that used in *ex ante* environmental assessments. This would seek to answer the following key questions:

- Can this measure be evaluated at all?
- If so, can an evaluation assess ultimate effects on the environment, or should it focus on intermediate outputs and outcomes only (see section 2.2)?
- Does the nature of the measure, or the problem it seeks to address, lend itself to modelling, or is there a need for broad collection of empirical data?

- What is the range of effects on the environment that need to be investigated?
- What available tools and methods are most appropriate, given the constraints of budget and timescale?

It is clear that such a screening and scoping process needs to involve skilled evaluators in addition to officials with specific sectoral expertise.

5. ESTABLISHING EFFECTS

This section focuses on the evaluation question on which REM is centred, that is: *What are the current and likely future effects on the state of the environment of a particular environmental measure?* The discussion relates to the evaluation of those measures where the use of models is not appropriate because of the nature of the policy and/or wide variations in the way it is implemented.

In these circumstances, establishing effects requires Member States to

a. monitor changes in the state of the environment following the implementation of the measure;
b. establish that there is in fact a causal link between the measure and any observed changes in the environment;
c. assess the *extent* to which those changes are the result of the measure, by discounting the effects of other influences.

2.1 Tracing causation

No evaluation of effects is possible in the absence of data on the baseline situation with which subsequent environmental and socio-economic changes can be compared. However, for most EU environmental measures, there is no requirement on Member States to collect and report baseline data before the policy is in place. This is a major limitation of EU reporting requirements which needs to be addressed if evaluations are to be undertaken through this mechanism.

However, even when baseline data is available, simple 'before and after' comparisons are not sufficient to establish causation. This is a weakness of the Nitrates Directive, for example. Whether and how far a particular measure can be considered responsible for subseqent environmental changes depends on several factors, which are discussed below. One key consideration is the manner in which the measure is *implemented.* In this respect there may be considerable variations between Member States.

To take account of implementation, it is helpful to refer to the diagram reproduced in Figure 1. This 'unpacks' the relationship between a policy measure and its ultimate impact on the environment into a number of key elements:

- **Inputs** - eg staff, administrative structures, resources, training etc
- **Outputs** - these are under the entire responsibility of officials and include designations; number of inspections; guidance notes; training courses offered etc
- **Outcomes** - the response to these outputs of target groups – eg reductions in emissions; increased recycling rates; shifts in the use of transport modes;

- **Impacts** - the ultimate effect of these changes in behaviour on resource use, environmental quality and biodiversity.

To help establish a causal link between a measure *as it is actually experienced on the ground* and its ultimate environmental impact, it is necessary to collect data and identify suitable indicators not just in relation to the state of the environment, but also in respect of *outputs* and *outcomes* at the appropriate geographical or sectoral level. Changes in the state of the environment which are not associated with appropriate changes in outputs and outcomes are unlikely to be the consequence of that particular environmental measure.

1.2 Contextual factors

Indicators linking outputs with outcomes and impacts can only *suggest* causal links between a measure and the state of the environment – they do not establish them. Many 'exogenous' influences unrelated to a particular policy will also have an impact. They range from changes in general economic activity and in the size and socio-economic structure of the population, to the impact of parallel policies, such as changes in support prices, taxes, subsidies, information campaigns etc. It is important in any evaluation of effects to monitor such contextual factors - particularly in relation to parallel policies (which are relatively easy to document) – and attempt to unravel their respective contributions to environmental changes. This is easier to assess if *comparisons* can be made between geographical areas or sectors where they apply differentially.

2.3 Net effects

It is also important to distinguish between the *gross* and *net* effects of a particular measure. This requires some assessment of deadweight, displacement and substitution.

- *Deadweight* describes those changes in outcomes that would have happened anyway, regardless of whether the particular measure was put in place. An example would be the wider use of cleaner technology resulting from normal capital replacement, rather than, say, an energy efficiency campaign.

- *Displacement* refers to a situation where polluting activities are simply relocated rather than reduced absolutely as a result of geographically-targeted measures. To evaluate the net effects of designating environmentally sensitive areas thus requires the monitoring of activities immediately outside their boundaries.

- *Substitution* occurs when measures focused on particular target sectors or species occur at the expense of those that are non-targeted.

Assessing deadweight may require in-depth interviews with a selected cross-section of a particular target sector. Displacement and substitution can only be assessed on the basis of wider monitoring beyond the immediately targeted area or sector.

2.4 Data needs

It follows from the above discussion that the range of data and indicators required to evaluate the environmental effects of a particular measure goes well beyond what is conventionally required by existing reporting obligations. Essential requirements include:

- *Baseline data* on the situation before the measure is put in place. This should cover not only data relating to the relevant state of the environment (S) and pressures (P), but also the existing behaviour of target sectors, and relevant parallel policies. This information should be reported at the time of formal transposition, before the stage of practical implementation.

- The identification of suitable indicators, and regular monitoring on the basis of them, of:

    - the immediate outputs and outcomes of the measure
    - changes in relevant parallel policies
    - associated effects of the measure on other relevant sectors or geographical areas (to assess displacement, substitution)
    - changes in the relevant state of the environment.

3. CONCLUSIONS AND RECOMMENDATIONS

- Environmental measures differ significantly, and no single evaluation methodology is appropriate to all of them.
- For some types of environmental measure, models may be appropriate to assess effects. In other cases, the wider collection of empirical data is necessary.
- For items of *existing* EU environmental legislation, a process of initial screening is necessary to establish
  - which measures are capable of evaluation at all;
  - whether the evaluation of environmental effects is practicable, or whether the focus should be on intermediate outputs and outcomes.
- To establish causation, data collection requirements are extensive, and include the need for baseline data, indicators of policy outputs and outcomes, in addition to indicators of state and pressures.
- For *new* EU measures, the needs of evaluation should be made integral to the design of the measure, including explicit quantified objectives and timetables; the identification of appropriate indicators and monitoring arrangements; procedures for reporting and regular review.
- Evaluation could be made easier through the increased use of pilot projects, and greater national and/or regional differentiation in the implementation of EU environmental policy, enabling comparisons.
- Screening and scoping, and the design of measures for evaluation, require a high level of technical skill not normally available in technical units. Consideration should be given to the establishment in DG Environment and national Environment Ministries of an Evaluation Unit with horizontal responsibility for evaluation.

- In view of the need for technical expertise in the design and management of evaluations, consideration should be given to whether legal reporting requirements which include assessments of effects/effectiveness should be undertaken by independent external evaluators rather than by Member States themselves.

References

1. European Commission *Evaluating socio-economic programmes* – The MEANS Collection Vol.3 p19
2. Ibid p23.