



**EUROPEAN
COMMISSION**

DG MARKT Guide to Evaluating Legislation

Brussels
March 2008

Acknowledgements

This evaluation guide was prepared in Directorate-General for Internal Market and Services under the direction of Werner Stengg, Head of Unit B2 - Impact Assessment, Evaluation and Economic Analysis. The main contributors were Teresa Fitzpatrick, Sara Piller and Joanna Poczowska from Unit B2.

The editing team would like to express their sincere thanks to the following people for their valuable comments, feedback and other contributions:

- colleagues in Unit B2 - Impact Assessment, Evaluation and Economic Analysis (DG MARKT);
- Mauro Miranda, Unit E1 - Services (DG MARKT);
- Alan Zoric, Evaluation Unit (DG EMPL);
- Commission Working Group on evaluation of legislation and non-spending activities (DG BUDG);
- Kevin Williams – OECD;
- Rick Davies, for his presentations on social network analysis (European Evaluation Society Training);

We would particularly like to thank participants to Training Sessions: "Evaluating Legislation the MARKET way" and other colleagues in DG MARKT directorates whose comments and queries were invaluable for our final editing.

Comments on the evaluation guide would be gratefully received and should be sent to:

Directorate-General for Internal Market and Services
Unit B2: Impact Assessment, Evaluation and Economic Analysis
European Commission
B-1049 Brussels

Or by email to: werner.stengg@ec.europa.eu

Table of contents

- 1. INTRODUCTION 5**
 - 1.1. PURPOSE OF THIS EVALUATION GUIDE..... 5**
 - 1.2. WHAT IS AN EVALUATION?..... 6**
 - 1.3. WHY DO EVALUATION? 7**
 - 1.4. WHAT HAS TO BE EVALUATED AND WHEN? 8**
 - 1.5. WHO DOES AN EVALUATION? 9**
 - 1.6. WHAT AN EVALUATION IS NOT 9**
 - 1.7. HOW TO DO AN EVALUATION..... 10**
- 2. EVALUATING LEGISLATION..... 11**
- 3. DOING THE EVALUATION..... 17**
 - 3.1. PLANNING 18**
 - 3.2. DESIGN 20**
 - 3.2.1. Intervention Logic 21*
 - 3.2.2. Evaluation questions 27*
 - 3.2.3. Identifying data sources and types 33*
 - 3.2.4. Identifying stakeholders 35*
 - 3.2.5. Timetables, deadlines & expected outputs (reports)..... 36*
 - 3.2.6. Quality criteria 36*
 - 3.2.7. Technical specifications 37*
 - 3.3. CONDUCTING THE EVALUATION..... 38**
 - 3.3.1. Key phases..... 39*
 - 3.3.2. Tools and techniques 42*
 - 3.3.3. How to develop each phase of the evaluation in practice 46*
 - 3.3.4. Quality Assessment..... 54*
 - 3.4. DISSEMINATION STRATEGY..... 55**
 - 3.5. EXPLOITING THE RESULTS 57**
 - 3.5.1. Policy proposal 58*
 - 3.5.2. ABM – Activity Based Management 58*
 - 3.6. FOLLOW-UP ACTION 58**
- 4. FURTHER INFORMATION & ASSISTANCE 60**
- 5. ANNEXES..... 60**

Figures

Figure 1 : What does an evaluation look at?	6
Figure 2 : "Standard" cause-effect logic for evaluation	7
Figure 3 : Evaluation within the policy cycle.....	8
Figure 4 : Scope of the different analytical studies	9
Figure 5 : Key stages of evaluation and main actors involved.....	10
Figure 6 : Tackling the problem of plastic shopping / carrier bags – different ways of 'getting there' when evaluating legislation	12
Figure 7 : Cause-effect logic for evaluating legislation	13
Figure 8 : Applying the cause-effect logic to plastic bag example	14
Figure 9 : Evaluating legislation must consider <i>all</i> the effects... even the unexpected!.....	15
Figure 10 : Introduction of the smoking ban in HORECA businesses in Ireland – target groups and other parties affected	16
Figure 11 : Six key stages of evaluation.....	17
Figure 12 : Example of a project included in the multi-annual evaluation program.....	19
Figure 13 : What the intervention logic should depict	22
Figure 14 : Focus of the intervention logic	22
Figure 15 : Intervention logic model for legislation.....	24
Figure 16 : Categories of evaluation questions linked to cause-effect logic of evaluating legislation	28
Figure 17 : Stage 3 - conducting the evaluation	38
Figure 18 : Four key phases of evaluation.....	39
Figure 19 : Different evaluation tools and their relevance to phases of the evaluation.....	42
Figure 20 : Stage 4 – dissemination strategy.....	55
Figure 21 : Stage 5 – exploiting the results	57
Figure 22 : Example extract from the follow-up Action Plan of the Citizens Signpost Service evaluation	59

Boxes

Box 1 : Case Study - introduction to the Services Directive	18
Box 2 : Case Study - intervention logic for authorisation schemes.....	26
Box 3 : Case Study - evaluation questions for authorisation scheme.....	31
Box 4 : Case Study - evaluation question success criteria and indicators for authorisation scheme.....	33
Box 5 : Case Study - data sources for authorisation scheme.....	34
Box 6 : Extract from Services Directive mandate: stakeholders and steering group	35
Box 7 : Extract from Services Directive mandate: timetable, deadlines and deliverables	36
Box 8 : Case Study - possible structuring phase for evaluating the authorisation schemes of the Services Directive	46
Box 9 : Case Study - possible data collection phase for evaluating the authorisation schemes of the Services Directive.....	49
Box 10 : Case Study - possible analysis and validation phase for evaluating the authorisation schemes of the Services Directive.....	52
Box 11 : Case Study - concluding the evaluation of the Services Directive	53

Tables

Table 1 : Example of time planning for an external evaluation	20
Table 2 : Overview of procurement procedures and indicative deadlines.....	37
Table 3 : Deliverables (reports) at each evaluation phase	41
Table 4 : Description of evaluation tools and techniques highly recommended for the structuring phase.....	43
Table 5 : Description of evaluation tools and techniques for other phases	44

1. Introduction

The evaluation of activities, particularly spending programmes¹, is a well established discipline at the Commission. However, evaluation of policies and legislation - the main type of evaluation carried out by DG MARKT - is a relatively new and developing field, gaining more and more importance in the era of better regulation. Whilst the **new Communication on evaluation** (SEC[2007]213) sets out a revised framework and quality standards for all evaluations, the available guidance is very general and most relevant to activities such as spending programmes. Given that there is no central Commission guidance tailored to evaluating legislation², this guide has been written to provide specific information aimed at facilitating the evaluation of INTERNAL MARKET & SERVICES legislation (and policies³).

As Internal Control Standard N°14 (Evaluation of Activities) requires evaluations, including those of legislation, to be performed in accordance with the revised framework and quality standards, the guide also attempts to ensure these are adequately met. Internal Control Standard N°14 and the specific standards are provided in Annex 1.

One of the first steps within the Commission's new evaluation framework is that each DG must establish an EVALUATION CHARTER, formally setting out the basic procedural responsibilities and organisation for conducting evaluations. Before reading this guide, you are advised to read the **Evaluation Charter for DG MARKT**, which can be found in Annex 2.

1.1. Purpose of this evaluation guide

This guide is intended to help desk officers in DG MARKT who are involved in evaluating legislation. **Its purpose is twofold:**

1. to explain what an evaluation of legislation is; and
2. to provide practical step-by-step guidance on how to do such an evaluation.

The guidance provided here covers the widest possible choice applicable to as many scenarios. Therefore, it should be used **proportionately**, selecting the tools and techniques appropriate to the needs of each individual evaluation project.

¹ Spending programmes are specifically defined measures that aim to fund, from the EU budget, development across the EU. Typical 'spending DGs' such as AGRI, REGIO and RTD have long traditions in evaluating whether the funds have been spent wisely through these programmes.

² Usually **spending programmes** are well defined with regard to their objectives and resources available, have more tangible and measurable actions and results, have easily identifiable beneficiaries and affected parties, and usually assess whether money is being spent wisely. **Legislation**, however, which often deals with "concepts" or "principles", is more complex in that there are multiple layers of interaction that must be taken into account, but which are interlinked, making them difficult to capture with traditional evaluation models.

³ Even though this guide focuses on legislation, it is also relevant for other policy instruments that are non-spending.

A range of examples are used to illustrate specific points and issues⁴, and a MARKET case study, based on the Services Directive, has been developed to provide more concrete assistance.

The remainder of Chapter 1 provides some basic information regarding **what** evaluation of legislation is **and why** it is important. Chapter 2 sets out **what is different** about evaluating legislation and the elements that must be considered, while Chapter 3 is divided into 6 sections (mirroring the 6 stages set out in the Charter) and deals with **how** to do an evaluation. Chapter 4 provides a brief summary of further information sources. In an attempt to be as practical / accessible as possible, the more detailed information on the background and theory of evaluation is extracted from this guide and provided in Annex 3.

We have also used colour coding throughout the guide to make it easier to see how the various components of the figures and relevant text are linked.

1.2. What is an evaluation?

An evaluation is an **evidence-based assessment** of how well legislation has done (or is doing) what it set out to achieve. It looks at legislation in terms of:

- What has changed - is it what we wanted to change?
- Why it has changed - is it because of the action we took? or due to something else?
- How it has changed - did things get better, worse or stay the same?
- Who was affected - which people/groups were affected by the action we took?

In other words, evaluation looks at:

Figure 1 : What does an evaluation look at?



An evaluation should look at all aspects of legislation and its process over a given time-frame, ideally covering adoption, implementation and impacts. It should consider not just whether the legislation did what it was expected to do, but also what other effects may have happened as a result. By measuring these changes against the situation at the time of adoption⁵, the evaluation can then judge just how well legislation has met its targets in the real world. It is important to note that these changes may be negative as well as positive.

⁴ The examples are purely for pedagogical reasons and are not taken from any existing evaluation source.

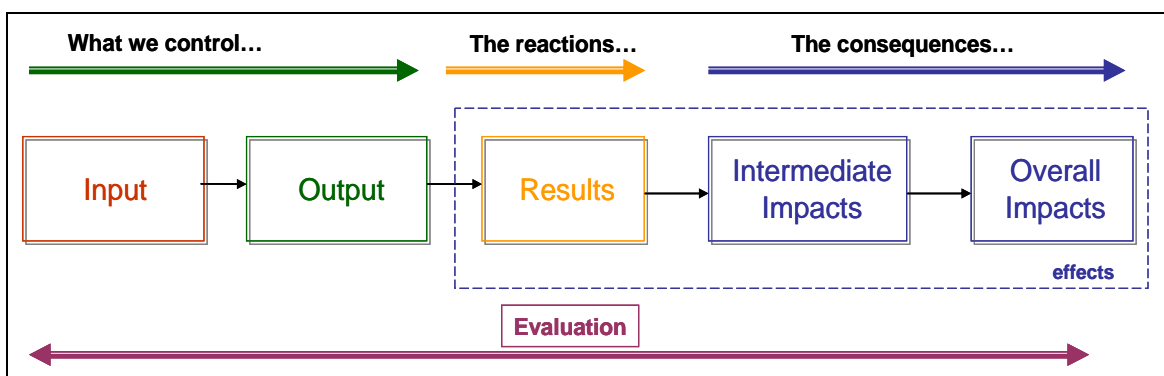
⁵ This is known as the counterfactual: the point in the life-cycle of the legislation against which progress can be assessed. Since many Member States may have taken preparatory action in the period between when the "baseline" was established in the Impact Assessment and prior to actual adoption, evaluating legislation needs to use the "counterfactual" rather than the "baseline" for a more stable starting point. This will also allow, where desirable, to verify the predictions of the Impact Assessment.

In short, an evaluation gathers and analyses all the facts associated with the chain of events in Figure 1 in order to determine the effects of legislation within the wider policy context. Traditionally, evaluation verifies the "cause-effect" links between:

- **inputs** (resources) – what was used, e.g. *staff needed to draft a Commission proposal*;
- **outputs** (action) – what we did, e.g. *the legislative instrument*;
- **results** – what happened, e.g. *national measures resulting from implementation*;
- **intermediate impacts** – what we wanted to achieve, e.g. *more cross-border integration*;
- **overall impacts** (general objectives) – the major change strived for, e.g. *increased productivity and competitiveness*.

This evaluation cause-effect logic is illustrated in Figure 2.

Figure 2 : "Standard" cause-effect logic for evaluation



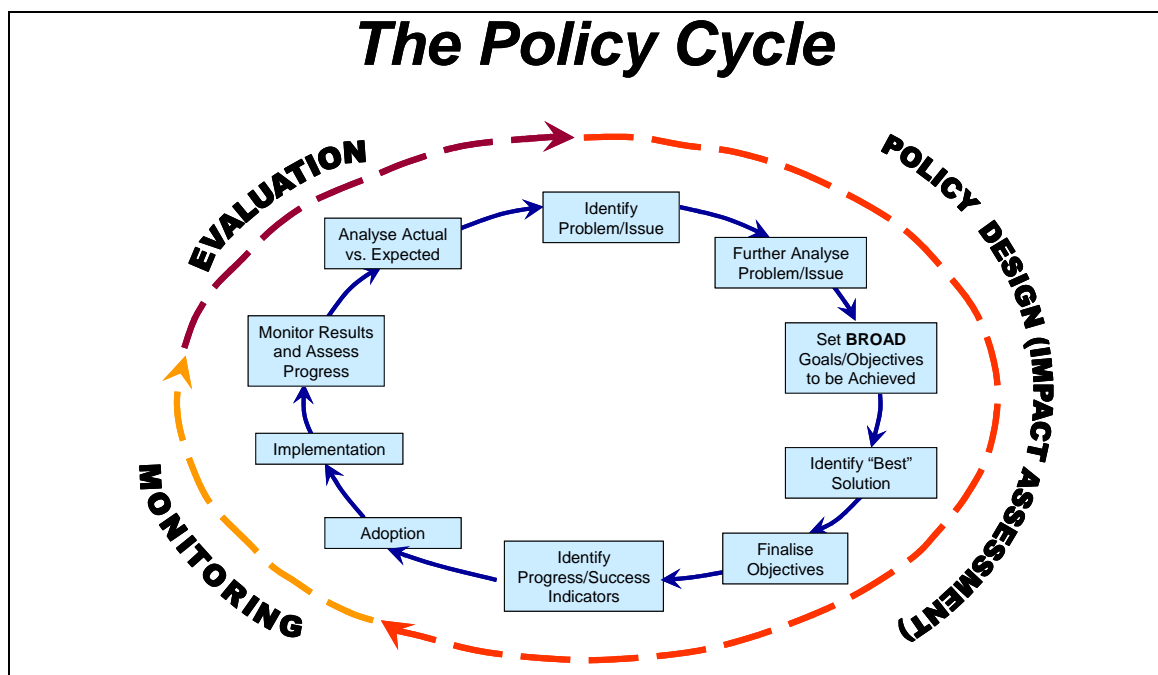
1.3. Why do evaluation?

As an essential learning and development tool, evaluation provides key information and solid evidence so that we can adapt our policies according to the needs of ever-changing political situations, allowing us to ensure they contribute to an effective regulatory environment and demonstrate to the citizen the benefits of Europe (and specifically the benefits of the Single Market).

The combination of the new Communication (SEC[2007]213), the revised Internal Control Standards and the increased introduction of evaluation into the more recent legal bases, emphasise the important role of evaluation – it is an obligatory tool for achieving better regulation within the policy cycle.

Figure 3 below shows where evaluation sits in the "policy cycle".

Figure 3 : Evaluation within the policy cycle



1.4. What has to be evaluated and when?

All activities addressed to external parties (i.e. citizens, business) must be evaluated periodically (the Financial Regulation sets this timeframe as every 6 years for spending programmes)⁶. The DG multi-annual evaluation plan which is coordinated by MARKT/B.2⁷, ensures that all major policies are regularly evaluated. Certain practical issues especially as regards timing and the level at which the evaluation takes place are also taken into account. The evaluation can be of:

- an individual piece of legislation; or
- part(s) of a piece of legislation, if its scope is large; or
- several pieces of legislation within the same theme or field (thematic evaluation).

Individual evaluations must always bear in mind the complete legal framework and policy context in which the legislation is embedded.

As a result of the introduction of the Impact Assessment System⁸, recent legislation should include provisions relating to evaluation requirements. Where there is no provision related to evaluation, a general rule of thumb is that the first evaluation can only be expected to take place a reasonable period after the date of transposition (e.g. 3 to 5 years after transposition, provided that most Member States have met the deadline). Subsequent evaluations can take place as and when appropriate. Most specifically, evaluation should probably be conducted ahead of any review/revision of legislation, thereby ensuring that the evaluation results contribute towards improved policy-making.

⁶ SEC(2007)213

⁷ The evaluation plan is an integral part of the Annual Management Plan. Hence Unit B.2 works together with Unit A.2 to produce this document.

⁸ All Impact Assessments should include a section outlining clearly the proposed monitoring and evaluation requirements.

1.5. Who does an evaluation?

The Operational Unit is responsible for managing an evaluation, working in close cooperation with B.2. The evaluation team may include external assistance for some or all of the work. A Steering Group, chaired by the Operational Unit, made up of interested parties should provide assistance where necessary and guidance on the quality of the work being carried out.

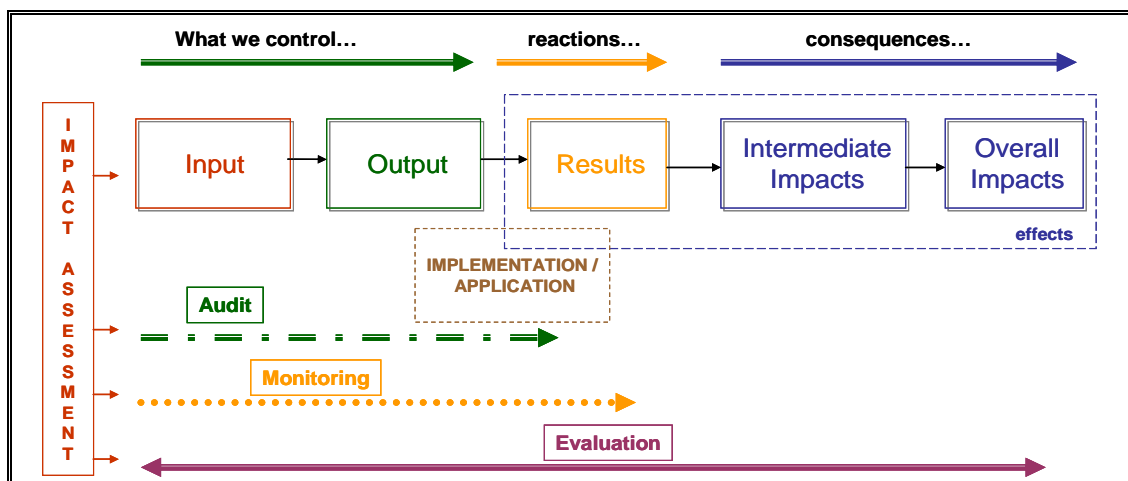
1.6. What an evaluation is NOT

The following types of reports are often confused with an evaluation report. In general, they are more limited in scope, and even though none of them can replace an evaluation, they can provide useful input:

- an **implementation / application report** = report on whether and how a legislative instrument has been implemented across the EU, sometimes including key elements of how the legislation has been applied as well as detailed market trends;
- a **monitoring report** = regular data collection and review of progress related to actions; focuses on immediate results rather than analysing why certain things are happening;
- an **audit** = examination of correct procedures, reliability & integrity of information, compliance with policies, adequacy of internal control systems and implementation performance.

The figure below shows the scope of these reports as compared to an evaluation:

Figure 4 : Scope of the different analytical studies

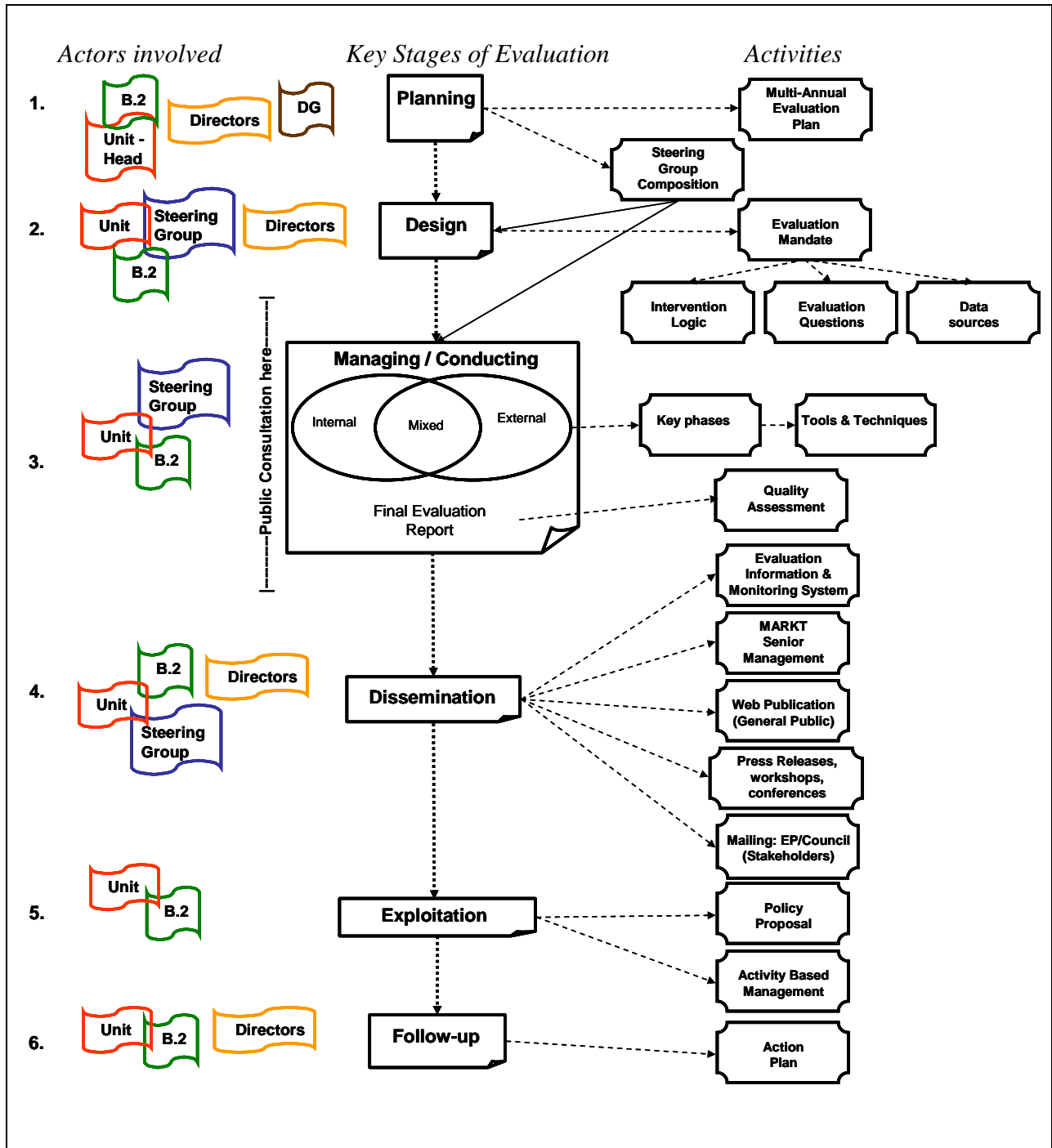


Sometimes legislation contains a "Review Clause" instead of or alongside a reference to "Evaluation". Depending on the scope and purpose of the Review, it may be more pertinent to carry out an evaluation or an Impact Assessment. If in doubt, please contact B.2 for appropriate advice.

1.7. How to do an evaluation

An evaluation is spread out over six key stages and involves a variety of activities and actors. The Evaluation Charter (Annex 2) sets out exactly how each of these actors is involved over the course of an evaluation. The work flow chart below provides a summary of who is involved in each stage of the evaluation and the activities contained in each stage.

Figure 5 : Key stages of evaluation and main actors involved



2. Evaluating Legislation

Having established that "traditional" evaluation usually verifies the "cause-effect" links as set out in Figure 2, one can ask why we cannot keep the same logic when evaluating legislation. There are a number of differences between spending programmes and legislation which makes evaluating the latter a more complex task. Evaluation of legislation must take into account that there are:

- **multiple actors** i.e. EU, Member State, regional/local government;
- **multiple types of legislation** i.e. Regulations, Directives, "soft-law" instruments, all with varying degrees of binding nature;
- **a number of ways for implementation and transposition to occur** (i.e. direct transposition; gold-plating; use of options and/or derogations; delayed, incorrect or incomplete transposition etc.)⁹;
- **various trade-offs** between those who benefit (winners) and those who are adversely affected (losers) amongst **targeted groups and other parties**;
- **positive and negative effects** and any attempts to maximise and minimise them respectively;
- **a range of indirect as well as direct impacts and unexpected as well as expected effects.**

In other words, there is rarely just one way to achieve the aims of legislation – in general there are many alternative routes, different methods of getting there and possible diversions along the way. When evaluating legislation these different possibilities must be taken into account in order to understand what the *actual* impacts have been. Equally, an evaluation must also remember to look at what changes have occurred due to other circumstances.

Figure 6 below provides a basic picture showing how the introduction of a Directive¹⁰ which aimed at reducing packaging waste e.g. to reduce the use of plastic bags may lead to two separate measures - a tax and an incentive - taken by two different Member States. Although both achieve the main objective of the EU legislation (fewer plastic bags for landfill disposal), they also generate other very significant effects, such as other types of waste. Technological developments (e.g. the biodegradable carrier bag), trends or behavioural changes also effect legislation.

⁹ For example, Directives :

- may be simply transposed into national law, somehow "sitting on top" of the rules that existed previously (*implies direct word for word transposition, no more, no less*),
- or may be integrated into the existing legislation, forming one coherent body of text (*implies that e.g. MS may add extra explanation, amend or drop other national laws to make consistent – not really extending the policy in any way, just intending to make it operational and efficient*),
- may contain options or derogations that Member States can avail of,
- or Member States may also choose to add further obligations, sometimes known as "gold-plating" (*so could take either of first two types of action, and then extend the policy to make it more stringent to achieve their own objectives*).

¹⁰ Directive 94/62/EC.

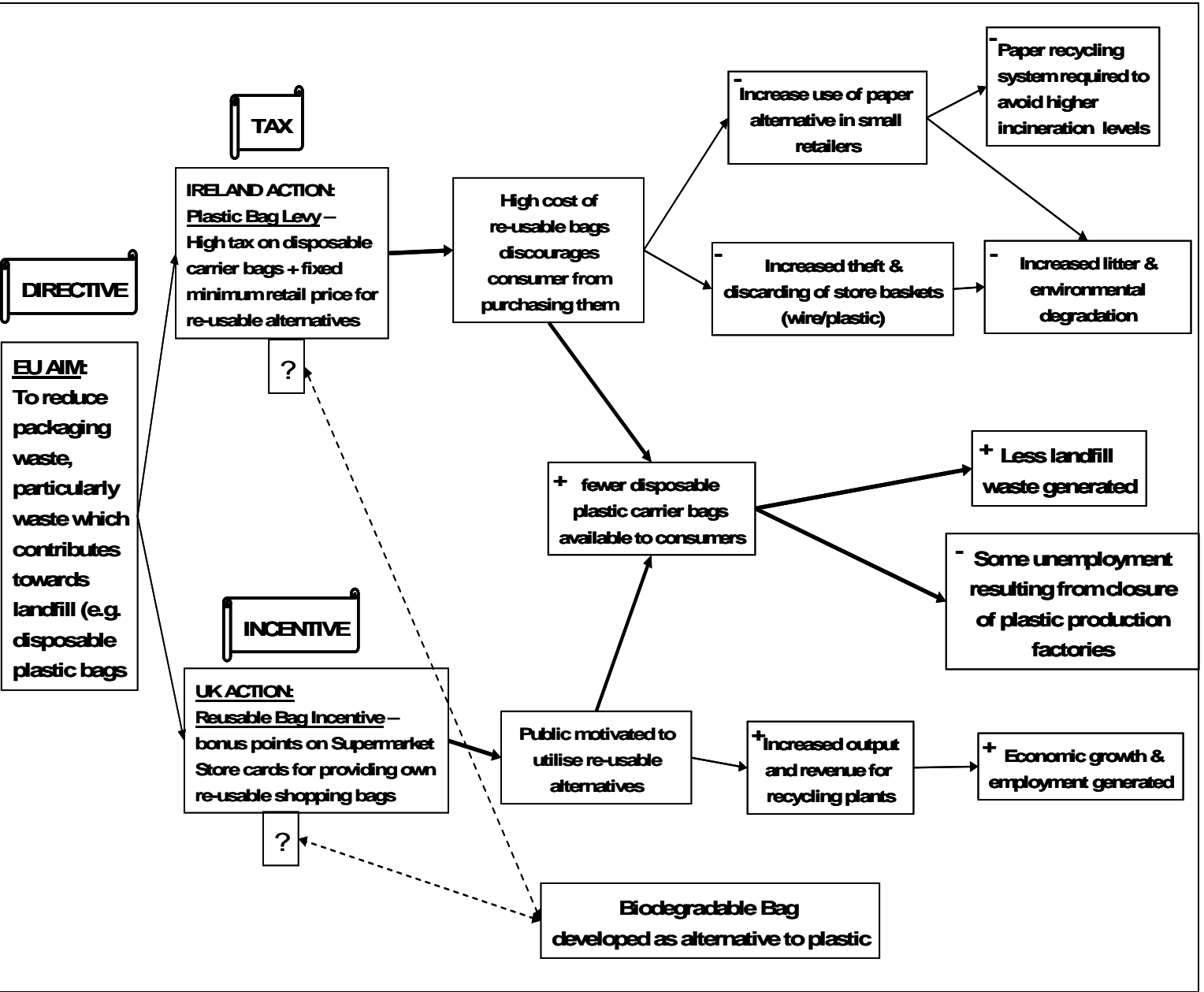
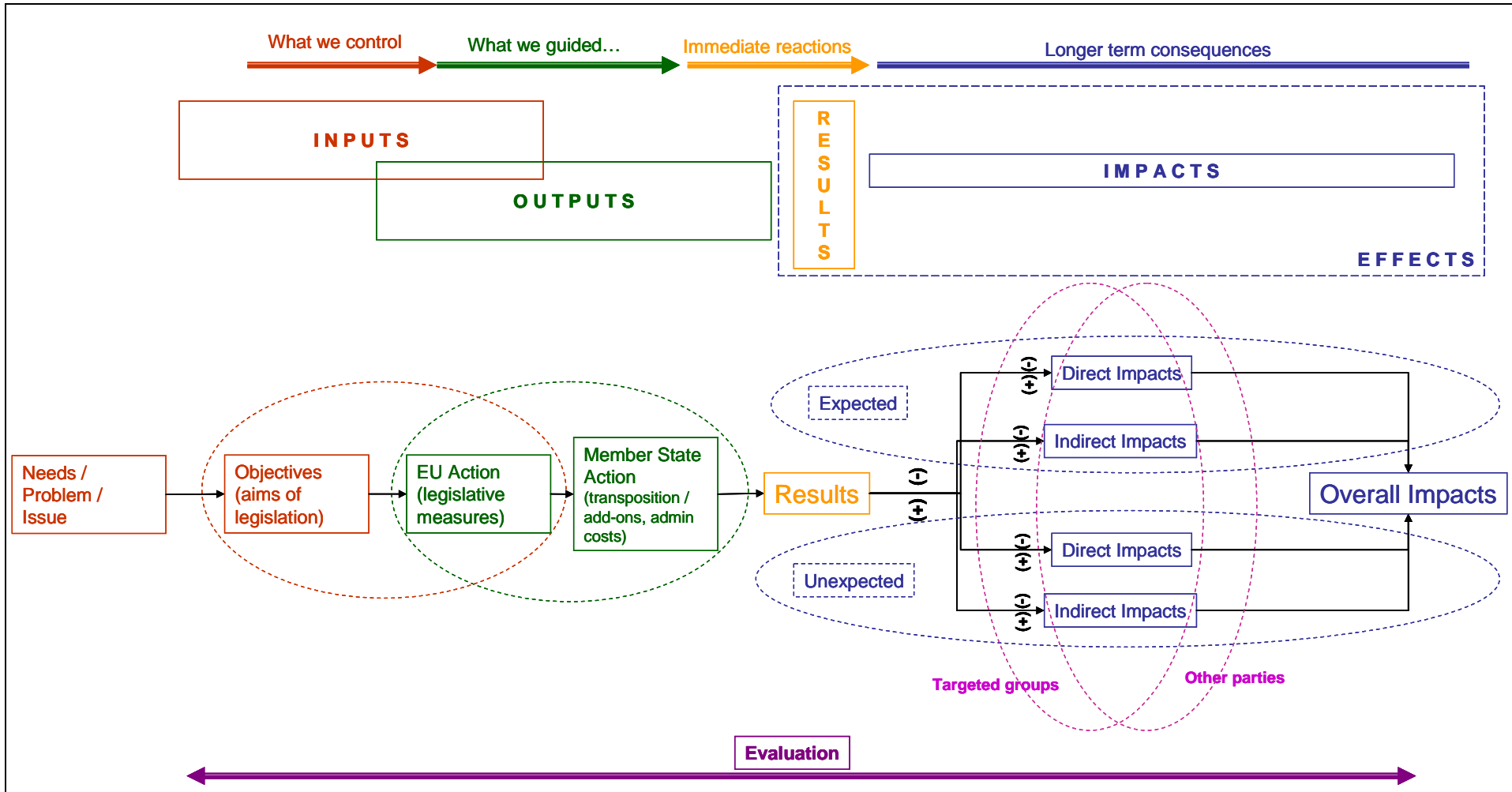


Figure 6 : Tackling the problem of plastic shopping / carrier bags – different ways of 'getting there' when evaluating legislation

Figure 7 : Cause-effect logic for evaluating legislation



One of the initial difficulties in using the standard evaluation logic model for evaluating legislation is the duality of the EU legislation itself: not only can it be considered an *input* (a resource that the EU puts into the mix), but also as an *output* (an action that requires something to happen). This creates an overlap.

Next, as the example of the Directive intended to reduce packing waste illustrated (Figure 6), EU legislation is implemented by Member States. Whilst it could be argued that their actions are the results of the EU legislation, it is **not** the role of the Commission to evaluate the Member States' actions. And yet, their intervention cannot be ignored when evaluating EU legislation, and so for this reason Member State action must be considered together with the EU legislation as an *output*.

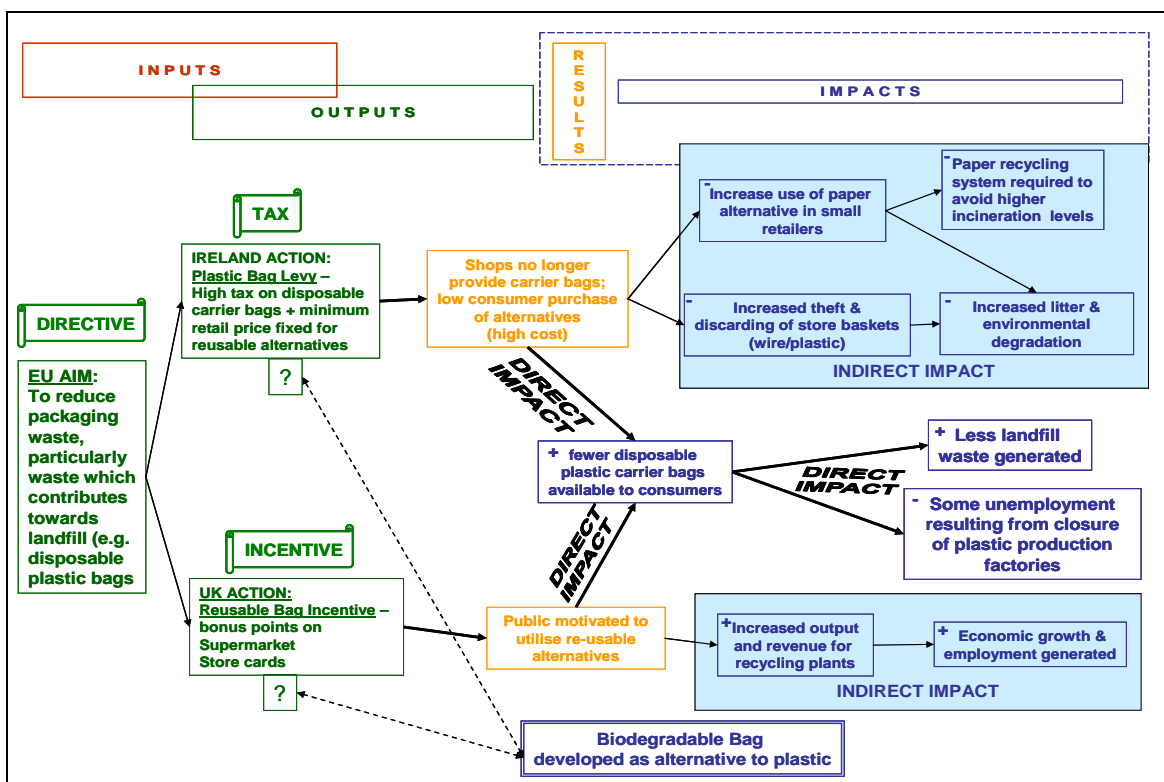
In this way, the reactions to the combined EU and Member State action can be determined as the *results*. The evaluation will verify whether these are in fact the same as those expected at the time of adoption of the EU legislation and will, quite probably, uncover several unexpected effects.

Lastly, due to the very nature of legislation, which can have consequences reaching far beyond the initial objectives, the evaluation will have a greater focus on the *impacts*: not only the direct ones, but it must also examine the indirect impacts, and by definition, not only on those groups that the legislation was aimed at, but on others affected inadvertently.

We will return to Figure 7 later in this guide, when we explain how to analyse these cause-effect relationships through pertinent evaluation questions.

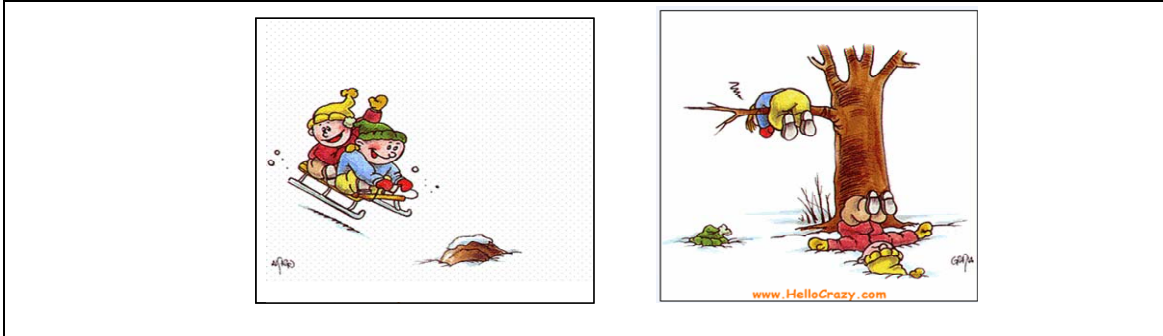
To see how important it is for an evaluation of legislation to consider *all* the effects in order to provide a sound analysis, in Figure 8 we have re-used the plastic bag example presented in Figure 6 and colour-coded it according to the cause-effect logic model above (Figure 7).

Figure 8 : Applying the cause-effect logic to plastic bag example



As can be seen from Figure 8, the action taken by Ireland has reduced the problems arising from widespread use of disposable plastic carrier bags (**direct impact**). However, it has also, inadvertently generated other waste problems (**indirect impact**), possibly requiring a review of the mechanisms existing to deal with the increase in paper waste, for instance. The (**unexpected**) technological development of a biodegradable bag may (or may not) also have effects on the chosen measures.

Figure 9 : Evaluating legislation must consider *all* the effects... even the unexpected!



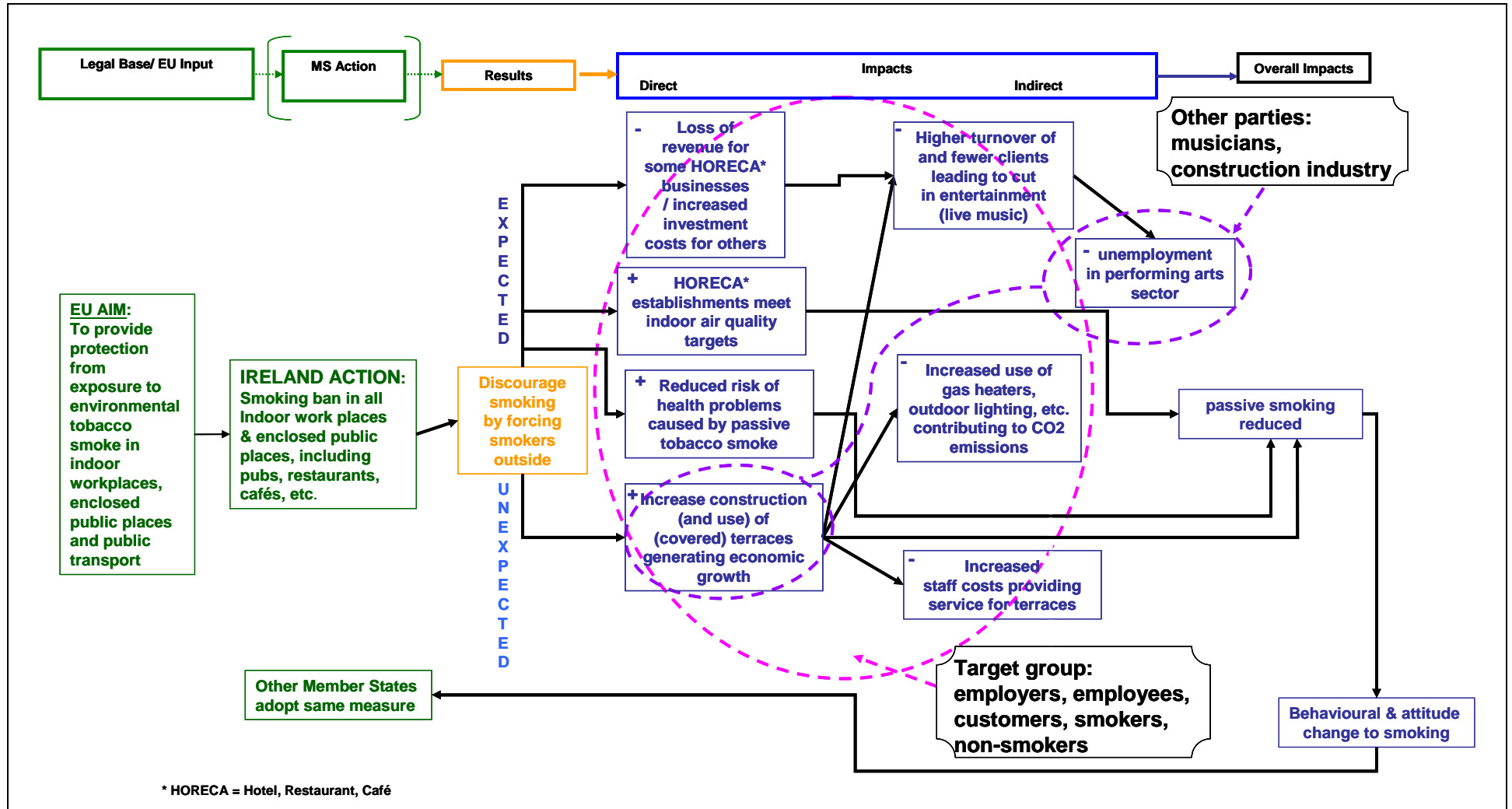
Evaluation will need to include all these issues in its analysis if it is to appropriately assess the success of the legislation in its full context. It also needs to take into account effects on other parties not just the targeted groups.

To illustrate the idea of target groups and other parties better, let's consider some of the changes which have occurred as a consequence of Council Recommendation 2003/54/EC relating to protection from exposure to environmental tobacco smoke. In Ireland, one of the first Member States to adopt related legislation, this led to a ban on smoking in pubs, restaurants and cafés.

The legislation introduced has potentially had direct and indirect impacts on the identified **target groups**, e.g. non-smoker employees may now be less at risk from passive smoking. However, some of the indirect impacts have adversely affected **other parties** not necessarily considered at the time when the legislation was planned/introduced. For example, as a large number of customers, smoking & non-smoking, now sit outside, it is unprofitable to have live bands playing music inside. As a result, some musicians have lost jobs because of the introduction of a smoking ban (see Figure 10).

A further, unexpected indirect impact on other parties may also be attributed to this measure. Following the perceived success of this measure and the change in public attitude to smoking, other Member States have since gone on to adopt similar measures.

Figure 10 : Introduction of the smoking ban in HORECA businesses in Ireland – target groups and other parties affected

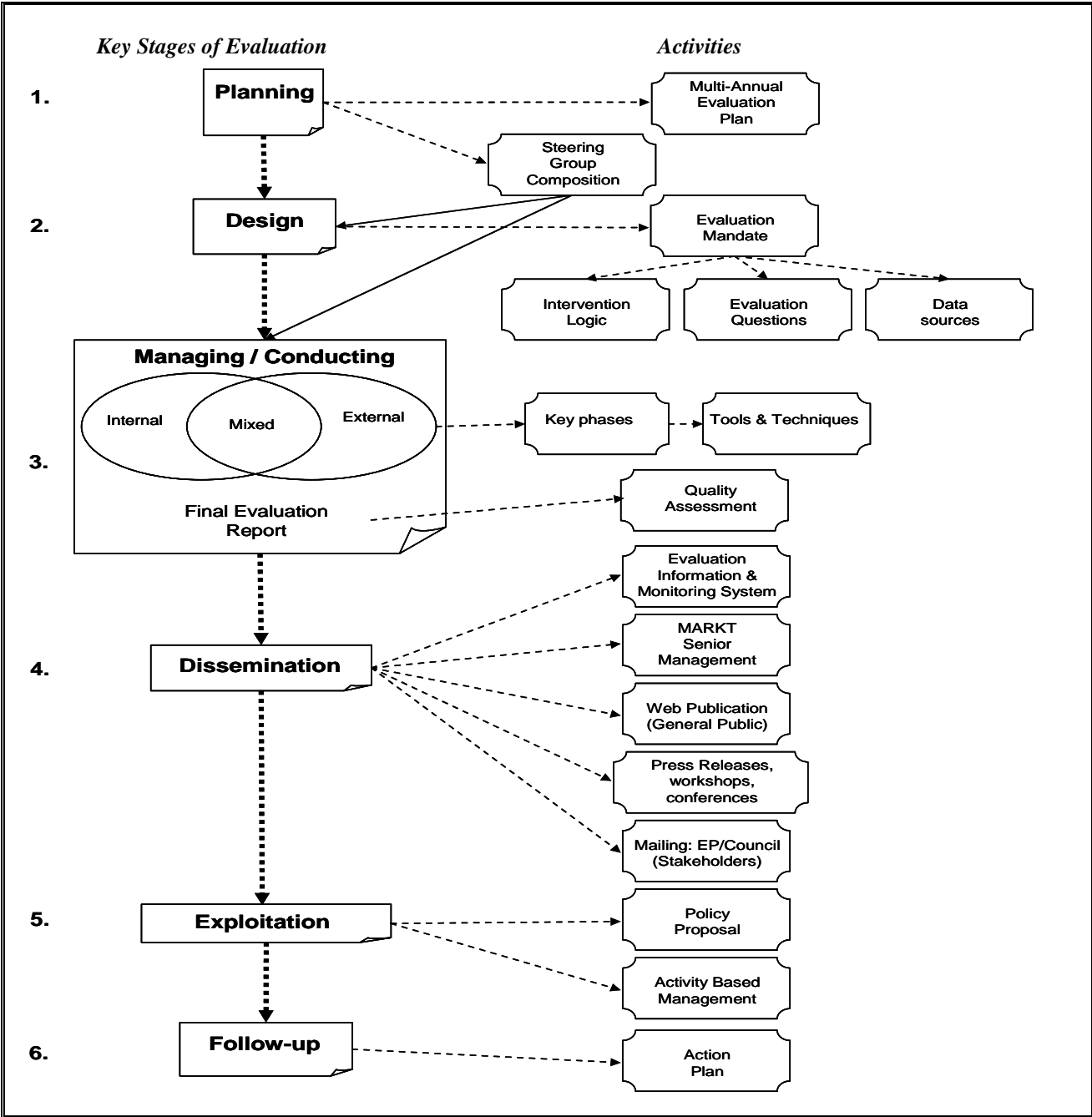


3. Doing the Evaluation

The DG MARKT Evaluation Charter divides an evaluation into six key stages that contain specific steps to ensure evaluation results of the highest quality. This next section is divided according to these six key stages, providing particular guidance on the various elements that are associated with each stage. Figure 11 provides a summary flowchart of these key stages and their associated elements.¹¹

B.2 is closely involved in all key stages of an evaluation, so remember to make use of the advice available!

Figure 11 : Six key stages of evaluation



¹¹ This flowchart will be used for some of the stages described below. Each time this happens, the stage being discussed and the activities involved will be shown in bold.

To show how to complete the key stages and actions required for an evaluation of legislation, a DG MARKT case study based on the Services Directive, has been developed to demonstrate specific points. Box 1 provides an introduction to the Services Directive case study.

In using this case study, it should be noted, that due to the complexity and broad scope of application of the Services Directive, the scope of the case study has been limited to the issue of Authorisation Schemes. Similarly, it is for illustrative purposes only – that is to say, the content is based on assumptions¹².

Box 1 : Case Study - introduction to the Services Directive

The Services Directive was adopted in December 2006. It is designed to address the need to make the internal market for services fully operational, whilst preserving the European social model.

The main objectives of the Services Directive are to facilitate the ability of service providers to exercise the freedom of establishment and the freedom to provide services. After it has been implemented, service providers should find it easier to set up a business in the EU or to provide services across Member State borders.

We will focus on the issue of Authorisations (section 1 Articles 9-13; Chapter III "Freedom of establishment for providers"). These provisions apply to all cases where a company wants to start a new business or to open a new establishment like a subsidiary or a branch. It concerns establishment in the provider's own Member State or another Member State and thus is not related to a cross-border provision of a service.

Within the issue of authorisation schemes, further focus will be put on 3 major sectors: construction and real estate; legal advice and consultation; tourism and leisure.

3.1. Planning

It is important to ensure that there are adequate resources (financial and human) to undertake the evaluation and that this request is done far enough in advance of when the results will be required. All evaluations included in the **multi-annual evaluation programme** (co-ordinated by B.2 every autumn) must have an indication of the required resources. This evaluation programme **is approved by the Director General** as part of the Annual Management Plan, which clearly signals that useful results for decision-making are expected. An example of a project included in the evaluation programme can be found in Figure 12.

The legislation itself probably gives some indication as to when evaluation results are expected to be available for decision-making purposes, i.e. review clause, evaluation article. The Monitoring & Evaluation section of the impact assessment – if one was carried out - acts as a good starting point for providing information on what was expected to be evaluated.

TIP: PLAN AHEAD! AN EVALUATION MAY TAKE MUCH LONGER THAN YOU THINK!

¹² The case study was chosen as it is an issue likely to be familiar to the majority of readers and provides a good example of the complexity of evaluating legislation. However, the authors take full responsibility for any misrepresentations or oversimplifications in the content.

Figure 12 : Example of a project included in the multi-annual evaluation program

Unit	Subject/Title	Reason, objective(s) and intended use of the evaluation	ABB Heading(s) concerned	Timing ¹³		Type of evaluation ¹⁴		Resources involved/ estimated ¹⁵	
				Start	End	prospective ("P"); retrospective ("R"); prospective and retrospective ("P&R")	external ("E"); internal ("I"); internal with external support ("I&E")	Human (Man/months)	Financial (in €)
B2	Evaluation of the methodology used to assess the performance of network industries providing SGEI	The Communication COM (2002) 331-Methodological Note for the horizontal evaluation of network industries providing services of general interest - foresees that the methodology will be reviewed and evaluated in 2006. The purpose of this evaluation is to assess the Methodological Note and its practical application in the three horizontal assessments reports of 2004, 2005 & 2006. It will also elaborate on potential improvements and future orientations of future horizontal assessments of the performance of network industries	120201 Policy Strategy and Co-ordination	10/2006	07/2007	P&R	E	6	120000

¹³ Starting with design, ending with follow-up action.

¹⁴ Evaluation can be forward-looking (e.g. Impact Assessment), backward-looking (e.g. for accountability), or both (e.g. for a review of legislation).

¹⁵ Even totally external evaluations will have internal Human Resource implications (e.g. a desk officer to manage the contract).

Planning doesn't end with the inclusion of the evaluation in the multi-annual programme. The evaluation needs to start well enough in advance so that the results are available on time and can be used in the decision-making process. In practical terms, this means planning a schedule for the evaluation project that can deliver results BEFORE the date by when they are required. Adequate time is needed for: designing the evaluation; drafting any technical specifications and launching the procurement procedure itself (if relevant); carrying out the actual evaluation; and preparing the appropriate dissemination of findings.

Table 1 below indicates how much time each stage of an evaluation can take. The timings given are typical for external evaluations; it will be different for internal or mixed evaluations. This is best done by working backwards from the date by which you need the information from the evaluation.

Table 1 : Example of time planning for an external evaluation

Project Stage	Duration (+ / -)	Deadline
Date evaluation results are needed by (report to EP/Council, start of work on a new proposal, ...)	---	June 2012
approval of report, dissemination of findings, follow-up action	1 month	May 2012
duration of the evaluation work (or contract)	10 months	July 2011
preparation of technical specifications & tendering procedures (if relevant)	9 months	October 2010
design & launch of evaluation	4 months	June 2010
planning		Autumn 2009

For more information, please contact B.2, who can advise you according to your particular evaluation needs.

3.2. Design

Every evaluation must have an **evaluation mandate** at the start of the process, which sets out the framework of what, why and how the evaluation is to be carried out¹⁶. It covers the complete design of the evaluation, setting out the issues to be examined and where to find the evidence to be analysed. The evaluation mandate governs the process of the evaluation; it is an essential tool for improving how an evaluation is designed and hence ensuring high quality results are produced. A blank copy of the DG MARKET evaluation mandate is provided in Annex 4 and a completed mandate for the case study is in Annex 5¹⁷.

¹⁶ SEC(2007)213, Standard C.2 (Annex 1).

¹⁷ Some significant parts of the case study mandate are also duplicated in the relevant sections as illustrative examples.

A steering group¹⁸ also needs to be set up early on, to assist with elaborating the evaluation process and to help facilitate the work of the evaluation team, e.g. provision of contact points, data research sources, etc. The steering group should play an active role in drafting the mandate and any associated technical specifications at the design stage.

The first parts of the evaluation mandate will already have been drafted, at the initial planning stage (i.e. when feeding planned projects into the multi-annual evaluation programme), making it easier to request appropriate **resources and set a timeframe**. Please ensure that the first parts of the mandate reflect the information originally submitted for the multi-annual evaluation programme. Where any significant changes have been made, they should be clearly identified and discussed with B.2. The initial parts of the mandate cover:

- **organisational aspects** (evaluation type, resources, timing, assigned desk officer(s), ...)
- **purpose** (what will the results of the evaluation be used for – review, state of play, ...)
- **objective** (what kind of information you expect it to be able to give you for the above purpose)
- **justification** (why are you doing the evaluation - requirements in the legal base, political demand,...)
- a **description of the action** to be evaluated
- **scope** (how broad should it be in terms of geography, timescale, parts).

The remaining elements of the evaluation mandate (items 1 to 7 below) can only really be completed when work starts on the evaluation itself, but are the key elements of the design of the evaluation project.

1. A diagram of how the legislation was expected to work (**intervention logic**)
2. The **evaluation** (criteria and) **questions**
3. **Data sources**
4. **Stakeholders and Steering Group**
5. **Timing, deadlines & expected outputs**
6. **Quality criteria**
7. If appropriate, the **Technical Specifications** for any external work

Some further explanations of these items are provided below in chapters 3.2.1. – 3.2.7. In addition, as the intervention logic and the evaluation questions are the foundations of the evaluation design, B.2 also offers to deliver a workshop on these two issues tailored to specific evaluation projects at the early stage in the process for units who wish to avail of it.

3.2.1. *Intervention Logic*

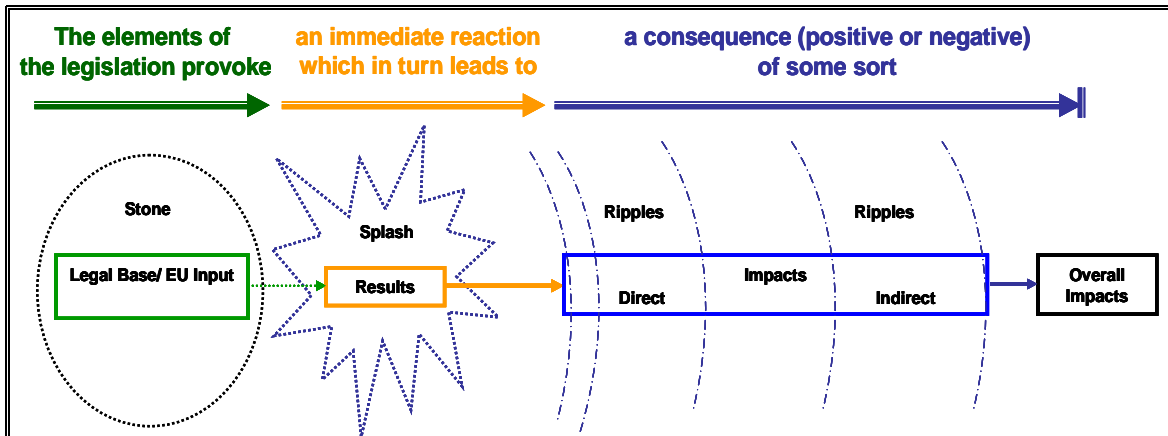
The Intervention Logic is a model within which the originally anticipated "cause and effect" of introducing the legislation are set out. It should be a visual representation of what was expected to happen due to the legislation, at the time of its adoption/design. This is done by examining the objectives of the legislation and establishing what it aimed to achieve and how. The Impact Assessment, if it exists, is a good starting point

¹⁸ For more information about the Steering Group, please refer to the section 3.2.4. 'Identifying stakeholders'.

since the problem statement is intended to present and analyse the "Needs/Problem/Issue" at that point in time, the desired "Objectives" to improve the situation are set out and ultimately the preferred course of action for the EU ("the EU inputs") is recommended. However, do bear in mind that legislation may have been modified on passage through the European Parliament and Council - the Intervention Logic will need to reflect this.

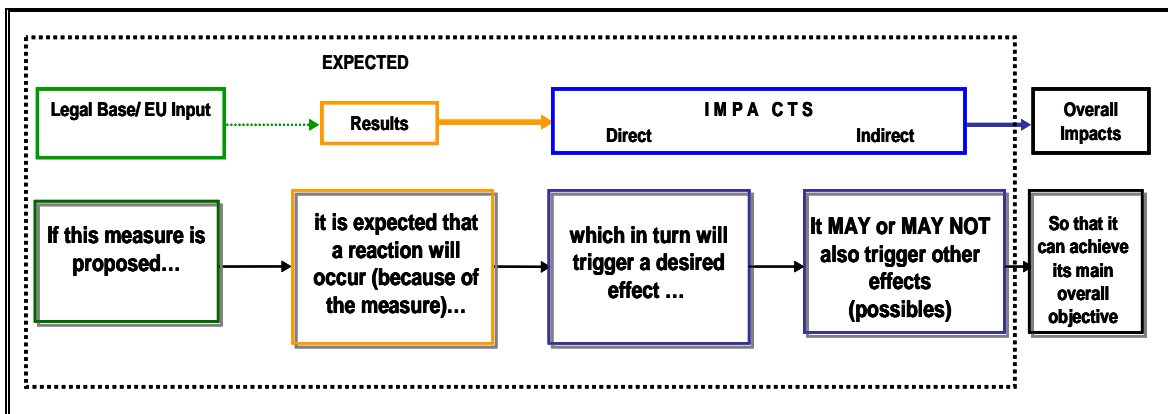
Think of what happens if you drop a stone into a pond – it is expected to make a splash and for the water to ripple out. Figure 13 overlays this cause-effect idea to show how to interpret what the intervention logic should include.

Figure 13 : What the intervention logic should depict



As illustrated in Figure 14, developing the intervention logic should start by focusing attention first on the expected results and impacts of the legislation being evaluated, by listing each measure, result and impact in appropriate columns and plotting the expected cause-effect links from the EU legislative measures to these, as intended when the legislation was adopted.

Figure 14 : Focus of the intervention logic



Mapping out the relationships between the legislation's objectives (i.e. the problems it aimed to solve), the measures taken and the anticipated impacts will help to establish pertinent evaluation questions (see section 2. below). Although the evaluation will need to examine the complete cause-effect logic of the legislation as per the model in Figure 7 the intervention logic provides a simpler starting point. To see how the cause-effect logic for evaluating legislation is translated into an intervention logic model, look at the animated slide on our website¹⁹.

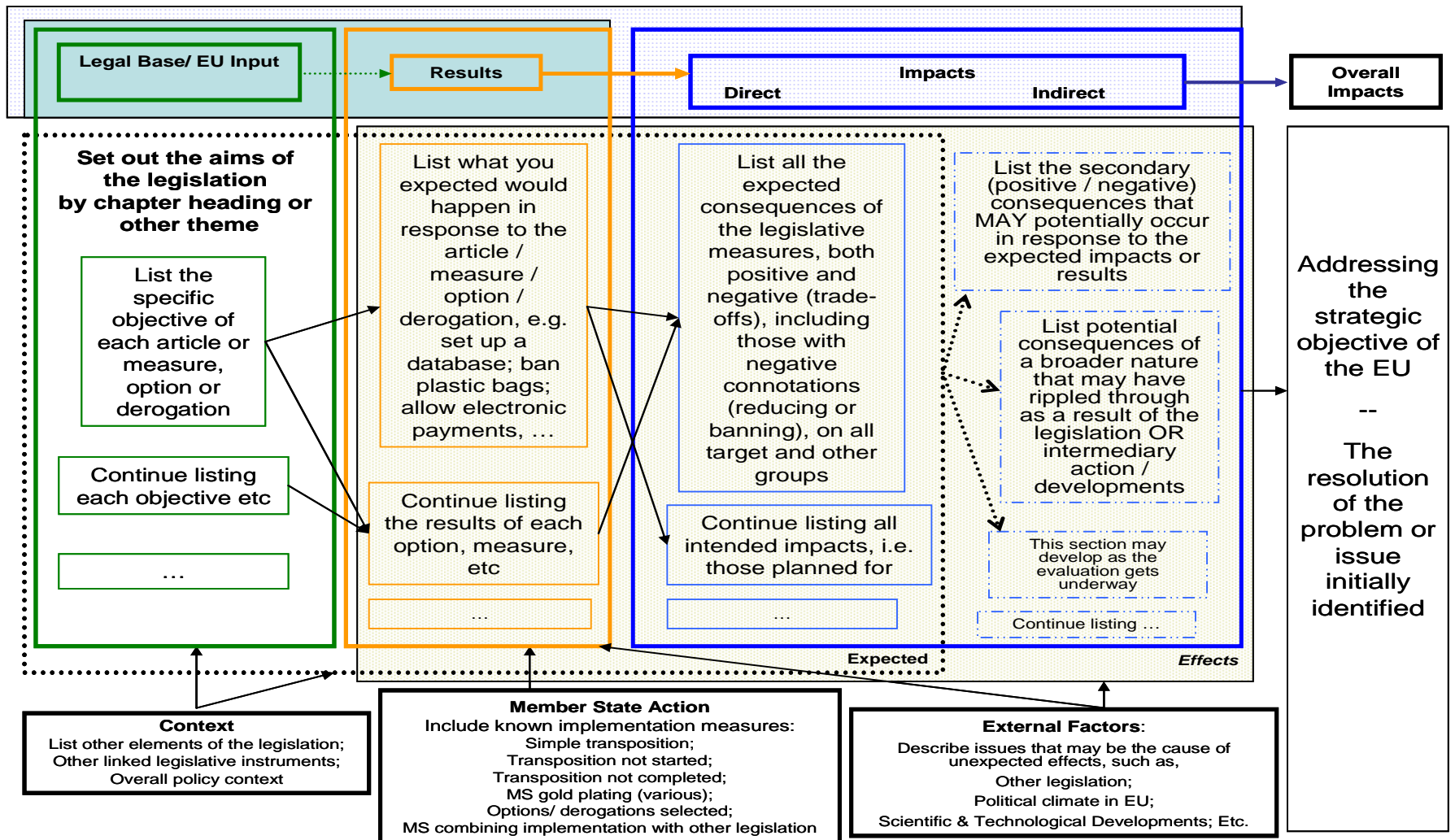
¹⁹ http://www.cc.cec/markt/intranet/index.cfm?action=domain&id_domain=1707 .

In order to keep the intervention logic simple at the design stage, Member State Action (what Member States actually did), external factors (unexpected causes/effects) and context (the original issue being addressed) are temporarily removed from the cause-effect picture (if they are actually known at this point). Although target groups and other parties also tend to be less visible within the intervention logic, it is important to ensure that the effects on all parties are considered in the evaluation.

However, an intervention logic is a dynamic tool: at this design stage it helps to establish a base picture against which the expected and actual effects can be measured in the evaluation. In the later stages of evaluation as more information is gathered, further more complete versions of the logic model can be drawn up to include Member State Action, unexpected effects and contextual issues. As Member State Action can be so varied, it would be impossible to map out in the intervention logic how each measure has been implemented without becoming unmanageable. Annexes 7 & 8 provide suggestions on how to deal with this.

Figure 15 below shows you how to build up the expected cause-effect links within the intervention logic in more detail.

Figure 15 : Intervention logic model for legislation



Understanding these relationships will help formulate evaluation questions specific to the legislation, not only establishing whether the representation is reality, but providing an indication of whether and/or how this reality is being achieved.

Some information for recreating the logic of the legislation (at the time of its inception) can be taken from the original Impact Assessment, if such a document is available²⁰. If the legislation pre-dates the requirements of Impact Assessment then a good starting point is to look at the explanatory memorandum, the objectives written in to the legislation itself, or any feasibility, scoping or other ex ante study that may provide similar insight. Information from implementation reports should help identify some of the Member State actions and also the related results.

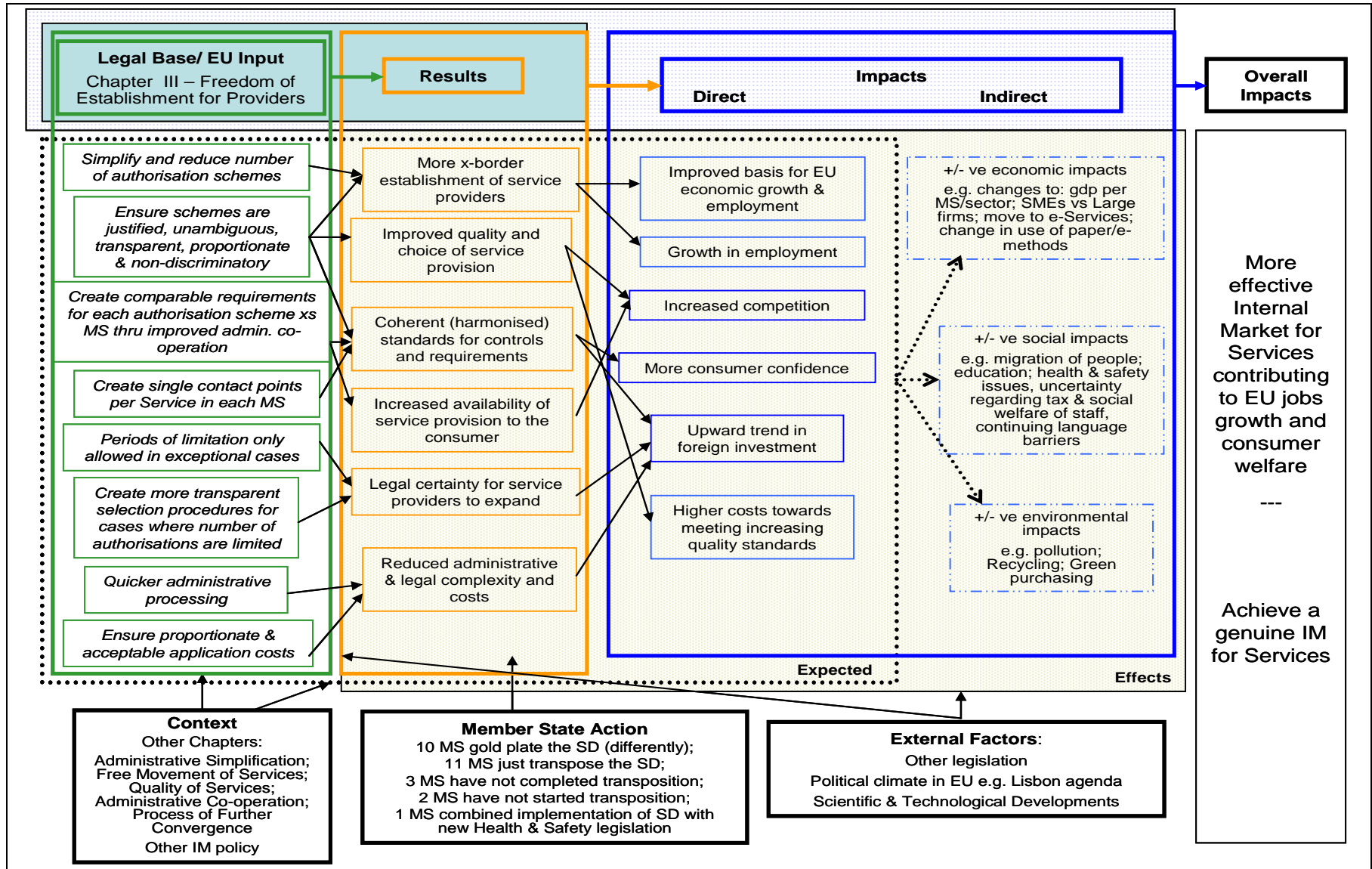
A concrete example of an intervention logic created for the authorisation schemes part of the Services Directive can be found in Box 2 below²¹.

TIP: IT IS WORTH SPENDING TIME DOING THIS. A SOLID INTERVENTION LOGIC IS A CRUCIAL ELEMENT IN DESIGNING AN EVALUATION THAT WILL PRODUCE QUALITY RESULTS.

²⁰ When using information from such studies, or indeed, if you are lucky enough to have an Impact Assessment that contains the Intervention Logic of the proposal, do remember to adjust the Intervention Logic to the actual legislation which was adopted.

²¹ Extracted from the mandate example in Annex 5.

Box 2 : Case Study - intervention logic for authorisation schemes



3.2.2. Evaluation questions

To assess to what extent the intervention logic reflects reality, a series of specific questions that are formulated in such a way as to provoke enquiry are needed. These *evaluation questions* aim to gather the relevant information required to examine how the cause-effect relationships of the legislation have in fact happened.

In other words, YES / NO type of questions should be avoided as they provide very little evidence for evaluation, for example:

- *Has improved choice of services contributed to increased EU economic growth? (YES)*
- *Is reduced administrative complexity strengthening consumer rights? (NO)*

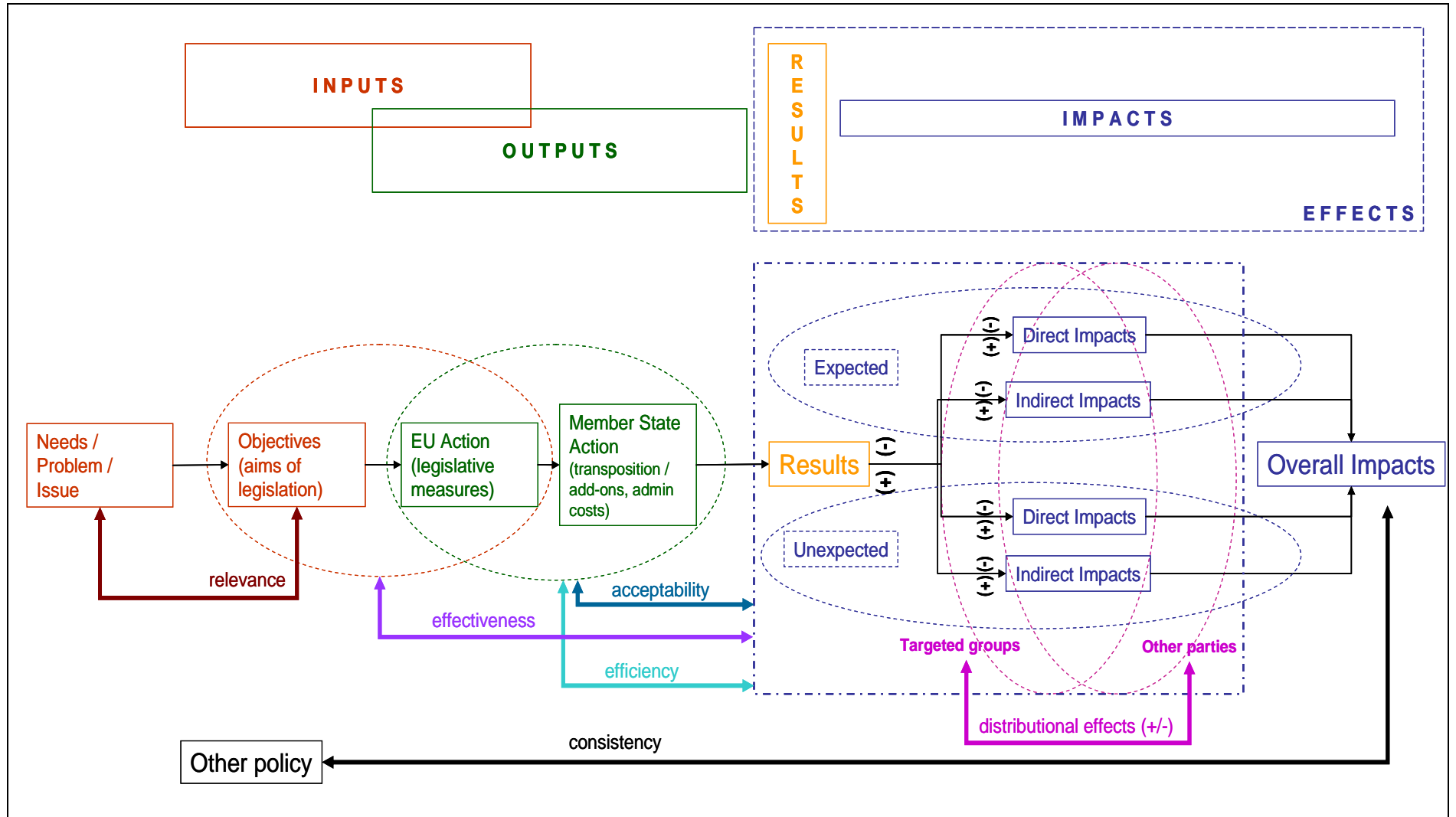
However, **correctly formulated, tailor-made evaluation questions** will lead to the collection of detailed evidence for a broader analysis of the legislation's achievements in the appropriate context. For example:

- *To what extent has the choice of services been improved and how has this contributed to increased economic growth?*
- *In what ways has administrative complexity been reduced? Which, if any, have contributed to strengthened consumer rights, and how?*

In this guide, we use six different categories of evaluation questions to examine the main cause-effect relationships: **relevance, effectiveness, efficiency, distributional effects, acceptability and consistency**. Usually a notion of **added value** is also analysed. In general, evaluation questions should be constructed for each of these categories to ensure a fully rounded analysis. Of course, more focus may be placed on some types of questions, depending on the purpose of the evaluation.

The graphic in Figure 16 below illustrates these categories of evaluation question and shows the cause-effect relationships that they are intended to measure/examine.

Figure 16 : Categories of evaluation questions linked to cause-effect logic of evaluating legislation



Some basic, general questions are provided below as examples of the style of question that should be considered for each category. They could serve as the basis for questions for any given evaluation, but would need to be tailored to meet the specific requirements. As a rule, general questions will not provide suitably detailed evidence relating to the context of the legislation being evaluated - specific evaluation questions based on the information set out in the intervention logic are necessary.

Questions should be formulated in such a way that they enquire:

➤ **whether the objectives of the EU inputs are still *relevant* to the problem as it is today.**

Advances in technological and scientific development, political will, and citizen's agendas continually influence the policy environment. In the light of any such changes it is paramount to check whether the problem is still valid or has evolved.

Questions :

- *in what way has the initial problem evolved ?*
- *to what extent are the objectives still relevant to the evolved problem or do they need to be reviewed?*
- *to what extent does the scope of the legislation still match the current needs or the problem?*

➤ **whether the EU legislation has been *effective* in meeting, or moving towards, the desired outcomes.**

If the EU legislative measures have been appropriate for achieving the objectives and have been presented in a clear, direct and unambiguous manner, then they should be successful in achieving the desired effects. If this is not the case, it may be possible to see how things such as delays and infringements have prevented progress. Equally Member States may affect the outcome as they can decide to take additional action possibly to address perceived/actual transposition failures or to strengthen measures they feel require extra impetus in their political arena.

Questions :

- *to what extent have the objectives of the legislation been achieved?*
- *to what extent have the selected measures been appropriate in moving towards achieving the objectives?*
- *what are the reasons for any infringement activity resulting from this EU legislation, e.g. unclear rules, cumbersome measures, political opposition ... ?*
- *what additional measures have been introduced by MS and what effects can be attributed to them?*

➤ **whether the EU legislation has delivered its results *efficiently* in terms of the resources used to obtain the actual effects.**

Resources can be broken down into a number of categories: time, staff, transposition cost, enforcement costs, administrative costs, compliance costs, etc. EU legislative (or national) measures that introduce disproportionate burdens or complexity in relation to the problem they are trying to address are poor performers in terms of efficiency. This is of particular interest in the current climate of better regulation, reducing administrative cost/burden and simplification.

Questions :

- *what were the costs (transposition, enforcement, administrative, compliance) of this EU legislation and how do they compare to the benefits?*
- *could the same effects have been achieved at lower costs?*
- *what were the costs of any additional measures introduced by MS and how did they contribute to the benefits?*
- *to what extent could this EU legislation be further simplified?*

➤ **whether there are *distributional effects* of the legislation across different groups.**

The evaluation should explore how the benefits (and costs) of the legislation have been distributed across multiple groups, for example: winners vs. losers; large business vs. SMEs; business vs. consumer; business vs. employee; business vs. environment; incumbents vs. new entrants; between sectors; between government levels; between geographical regions; etc.

Some distributional effects will be deliberate. However, in addressing a problematic issue for one set of target groups it is quite possible that an intervention may have an unintentional effect on other groups.

Questions :

- *what positive & negative effects have impacted on which groups?*
- *how are positive and negative effects spread across different groups?*
- *what measures have been introduced to combat any undesired effects (in general or on a particular group)?*

➤ **whether the legislation itself (including any additional national measures) and effects were *acceptable* to the stakeholders involved.**

Questions :

- *what has been the level of resistance/acceptance by each stakeholder group and for what reasons?*

➤ **whether the actual effects (negative as well as positive ones) of this legislation are *consistent* with the strategic objectives of the DG and with the economic, social, environmental and overall objectives of Commission policies in general.**

It is important to establish that Internal Market legislation is coherent with policy objectives in these (and other) fields.

Questions :

- *to what extent are the impacts consistent with Single Market policy objectives, such as a) creating cross-border opportunities for enterprises & professionals in the EU, b) delivering measurable benefits to consumers in the Internal Market?*
- *what economic, social and environmental impacts (positive & negative) have arisen from this legislation?*
- *what measures, if any, are in place to maximise/minimise the impacts?*
- *to what extent are the impacts consistent with trade and other policy outside the EU?*

- An additional question category relates to assessing the ***EU added-value*** and **understanding whether the results could have been achieved without EU intervention**. If there was an Impact Assessment accompanying the original proposal, this should have concluded that EU action was **expected** to be necessary. The evaluation may wish to consider if this expectation was justified, or whether other means of achieving the same (or better) outcomes were actually feasible.

Questions :

- *to what extent has addressing the problem/issue at EU level gone beyond what could have been achieved by national, global or other measures?*
- *what have been the (extra) benefits/advantages of acting at EU level?*

Given the complex nature of evaluating legislation and the scarce resources available, there is a real need to focus on the important issues which can be of most benefit to future policy development. The evaluation will not serve its purpose if it explores areas (albeit interesting ones) unrelated to those where evidence-based findings are required for decision-making. It is better to ask a smaller number of questions and learn something quite detailed, than to set a large number of questions which only permit a superficial analysis of what has happened and why.

Some concrete sample questions for the case study are presented in Box 3²².

Box 3 : Case Study - evaluation questions for authorisation scheme

Case Study: Evaluation of Authorisation Schemes (part of Services Directive)
Key questions for the evaluation:
<i>Relevance; Effectiveness; Efficiency; Distributional effects; Acceptability; Consistency; EU added-value</i>
Relevance:
<ul style="list-style-type: none"> • Are the principles set out for authorisation schemes still appropriate today? • Should other principles be defined?
Effectiveness:
<ul style="list-style-type: none"> • Which authorisation schemes have been changed, abolished or maintained as a result of the implementation of the directive? • What different measures have been implemented at national/sector level and how have these affected results? • To what extent have the above changes (positive or negative) contributed to the objective of the Services Directive to facilitate the Freedom of Establishment? • If there are still problems with authorisation procedures, what are they and to what extent are these problems related to how the Directive has been interpreted or implemented?
Efficiency:
<ul style="list-style-type: none"> • Could improvements have been made to authorisation schemes by any other measure – was legislation in this form necessary? • What were the costs of screening the existing authorisation schemes? • What were the costs of changing authorisation schemes?

²² Extracted from the example mandate in Annex 5.

Case Study: Evaluation of Authorisation Schemes (part of Services Directive)

- How proportionate are the above costs to the benefits obtained?
- To what extent are the current schemes simpler, quicker and cheaper to administer and apply for? And what is the potential for, and cost of, further simplification?

Distributional effects:

- How have the observed costs/benefits been spread (equally/proportionally) across the various groups involved, in particular, service providers, consumers, new market entrants, expanding businesses, employees, in the three sectors: construction and real estate; legal advice and consultation; tourism and leisure?
- What measures have been introduced to combat any undesirable effects?

Acceptability:

- What has been the reception from stakeholders – how does it differ between different stakeholders?
- Are there differences according to sector (with particular focus on the three selected sectors)?

Consistency:

- Have the changes to authorisation schemes, both positive and negative, contributed to the Single Market objectives of creating cross-border opportunities without conflicting with the specific objectives of the authorisation schemes themselves?
- In what way have such changes influenced the global objectives of authorisation schemes?
- What steps have been taken to maximise synergies for authorisation schemes or minimise the impacts of conflicting objectives?

EU Added Value:

- To what extent could the changes brought about by the Services Directive have been achieved by national measures only?

TIP: KEEP IN MIND WHY YOU ARE DOING THE EVALUATION – WHAT DO YOU REALLY NEED TO KNOW FOR FUTURE DECISION-MAKING?

Success Criteria / Indicators:

In essence, success criteria are developed based on the expected results or impacts of most relevance to the evaluation questions. They allow the evaluation team to measure how well actions (i.e. legislation) have led to progress towards the objectives. To do this, indicators that can measure progress in some way need to be established, enabling the evaluation team to draw appropriate fact-based conclusions²³. Consequently success criteria and indicators provide the link between the intervention logic and the evaluation questions.

Developing success criteria and indicators can be a quite useful first step in identifying what material/data is already available. This is a task that external consultants can often be asked to undertake.

Box 4 gives an example of success criteria and indicators for the case study.

²³ At this point it is worth considering the indicators defined at the early stages of drafting legislation (e.g. see Monitoring and Evaluation section of IA where appropriate). Such indicators are meant to help measure how well the objectives have been met.

Box 4 : Case Study - evaluation question success criteria and indicators for authorisation scheme

Evaluation Question	Success Criteria	Indicator
Which authorisation schemes have been changed, abolished or maintained as a result of the implementation of the directive?	<ul style="list-style-type: none"> • Fewer authorisation schemes • Changes resulting in quicker schemes • Simpler, more consistent authorisation schemes across Member States 	<ul style="list-style-type: none"> • 50% reduction in number of authorisation schemes by 2010 • Change in time taken to receive authorisation (from 2007 to 2010?) • Number of MS where authorisation scheme X is same

3.2.3. Identifying data sources and types

Simply asking the right questions is no guarantee of a successful evaluation; some indication of where to find the answers and the likely availability of such information in the first place is also necessary.

The evaluation mandate should indicate for each evaluation question:

- **what kind** of information is needed

The sort of data potentially required for answering the evaluation questions is particularly important for understanding not only the resources that will be required for collecting the data, but also which analytical tools will be most appropriate. Data can be numerical, such as statistics (e.g. proportion of market share), economic (e.g. Gross Domestic Product), costs (quantitative) or non-numerical, such as market trends, opinions, attitudes, sector reactions (qualitative).

- **where/from whom** to get this information

The identification of primary (to be collected) and secondary (already available) data sources is important, particularly when considering the time-frame required to undertake the evaluation. Primary sources include the target groups and other parties (as identified earlier, see Figure 10) administrations, representative associations, non-governmental organisations, etc. - basically any source which the evaluation team will have to address directly to obtain data. Secondary sources can be previous evaluation reports, reviews, case law, reports, readily available statistical data (Eurostat), other statistical analyses, academic literature, etc.

This section of the mandate should be as detailed and as exhaustive as possible to ensure the appropriate mix of evaluation tools later on. The case study example in Box 5 below²⁴ provides a flavour (but is not complete) of the type of information expected. This table could be expanded to include indicators, either in the design stage or at a later point.

TIP: TRY TO ACHIEVE A BALANCE BETWEEN QUALITATIVE AND QUANTITATIVE DATA AND PRIMARY AND SECONDARY SOURCES. BE SURE TO IDENTIFY THE TARGET GROUPS AND OTHER PARTIES.

²⁴ Extracted from the mandate in Annex 5.

Box 5 : Case Study - data sources for authorisation scheme

Category	Evaluation Questions	Data Type	Source
Relevance	Are the principles set out for authorisation schemes still appropriate today? Should other principles be defined?	<ul style="list-style-type: none"> An overview of the situation in the EU at this point in time? – a "state of play"; A summary of the situation as was at time of drafting legislation 	<ul style="list-style-type: none"> Impact Assessment; Implementation Reports; Academic Literature; case law Member State authorities; Infringement unit; National judiciaries; European Court of Justice
	Which authorisation schemes have been changed, abolished or maintained as a result of the implementation of the directive?	<ul style="list-style-type: none"> Statistics on the number of authorisation schemes per Member State and per sector Qualitative data on authorisation schemes 	<ul style="list-style-type: none"> relevant authorities
Effectiveness	What different measures have been implemented at national/sector level and how have these affected results?	<ul style="list-style-type: none"> Legislation put in place in each Member State 	<ul style="list-style-type: none"> Member State authorities; Representative associations; National judiciaries;
	To what extent have the above changes (positive or negative) contributed to the objective of the Services Directive to facilitate the Freedom of Establishment?	----	---
	If there are still problems ...?
	Which ...?	
	What different measures ...?		
Efficiency			
Distributional effects			
Acceptability			
Consistency			
EU value-added			

3.2.4. Identifying stakeholders

It is also very important to establish at an early stage **who the interested parties are**, i.e. the key people who are expecting to **use** the evaluation results. This will help keep the focus of the results on the needs of decision-makers and can be useful when planning the data collection. Such stakeholders might possibly include the European Parliament, Council, Member States, other DGs, amongst the usual MARKET hierarchy and Cabinet.

Other stakeholders such as NGOs, trade associations, business, consumer associations, citizens, ..., will also be interested in the outcome of the evaluation and will often form the target base for your consultations early in the evaluation work and be the main contributors in the data collection phase.

The Evaluation mandate should be quite explicit in identifying the Stakeholders, naming a contact, if possible, as well as the entity they represent, e.g. Mr. Jack O'Faltrades, DG EMPL; Ms. Bree Collage, Association of Builders.

This is also a good moment to reflect again on the composition of the **Steering Group** to ensure that it is appropriate to the specific needs of the evaluation. It is worth remembering that the role of the steering group is to facilitate the design, the work and dissemination of the evaluation results, so it is advisable to include other units and services. It may also be worth considering including key external stakeholders. Again, be explicit in identifying who the members of the steering group are and consult B.2 if you are unsure of the appropriate size and composition.

Box 6 shows the section of the mandate where this information must be provided.

Box 6 : Extract from Services Directive mandate: stakeholders and steering group

Stakeholders:	
Within Commission:	DGs represented in the Steering Group
Member State authorities:	IMAC members responsible for implementing the Services Directive
Business, trade, NGOs, etc:	<p><u>Business organisations from the EU and national levels – general ones as well as sectoral where authorisations schemes play particularly important role, e.g.:</u></p> <ul style="list-style-type: none"> • UNICE - Confederation of European Business • EuroCommerce • ETOA – European Tour Operators Association • Eurochambers • FIEC - European Construction Industry Federation • CIAA - Confederation of the food and drink industries of the EU • RICS - Royal Institution of Chartered Surveyors (real estate) <p><u>Consumer organisations, e.g.:</u></p> <ul style="list-style-type: none"> • EUROCOOP - European Community of Consumer Cooperatives • ANEC - European Consumer Voice in Standardisation • BEUC - The European Consumers Organisation <p><u>NGOs:</u></p> <ul style="list-style-type: none"> • European Citizen Action Service, ECAS <p><u>Trade unions, e.g.:</u></p> <ul style="list-style-type: none"> • UNI-Europa • ETUC/CES - European Trade Union Confederation

Other:
Steering Group:
<input type="checkbox"/> DG internal steering: <input checked="" type="checkbox"/> With participation of other Commission services (recommended): COMP, EMPL, ENTR, MARKT, SANCO, SECGEN <input type="checkbox"/> Steering including Member State's representatives (optional): <input type="checkbox"/> Steering including external stakeholders (optional):

TIP: TRY TO IDENTIFY AS BROAD A STAKEHOLDER BASE AS POSSIBLE AND INVOLVE KEY INTERESTED PARTIES IN THE STEERING GROUP FOR A THOROUGH EVALUATION.

3.2.5. Timetables, deadlines & expected outputs (reports)

The mandate should include an indicative timetable of the key milestones or deadlines which the evaluation should meet. This will help structure the work and provide a rough planning for both the Steering Group and the evaluation team, particularly where some or all of the work is expected to be carried out externally.

Box 7 shows the section of the mandate dealing with information about milestones and deadlines. It has been filled in using dates for the Services Directive case study.

Box 7 : Extract from Services Directive mandate: timetable, deadlines and deliverables

Milestones, deadlines:			
a) Designation of the Steering Group	01/01/2011	e) Quality Assessment	31/05/2012
b) Validation of the Mandate (including ToR)	28/02/2011	f) Key Findings & Recommendations	15/06/2012
c) Selection of Proposals (if external evaluation)	15/06/2011	g) Dissemination Strategy	30/06/2012
d) (Draft) Final Report	30/05/2012	h) Follow-up Action Plan	15/09/2012

TIP: BE SURE TO START THE EVALUATION WORK SUFFICIENTLY IN ADVANCE OF WHEN THE FINDINGS ARE REQUIRED BY THE DECISION-MAKING PROCESS.

3.2.6. Quality criteria

The Evaluation Standards (Annex 1) place particular emphasis on assessing the evaluation work against established **quality criteria** and require these criteria to be an integral part of the evaluation mandate. The criteria are: relevant scope, appropriate methods, reliable data, sound analysis, credible results, valuable conclusions, clarity of deliverables. To ensure from the start that all parties involved are aware of these criteria and how they will be applied, the (blank) standard **Quality Assessment Form** (in Annex 6) must be attached to each evaluation mandate. This form must be completed at the end of the evaluation.

While the formal Quality Assessment is to be carried out on the (draft) Final Evaluation Report, regular quality control based on the same criteria must be carried out at key stages of the evaluation AND most importantly where any external work has been requested. More detailed information related to carrying out a Quality Assessment is provided under 3.3.4.

3.2.7. Technical specifications

If it is decided to contract out some or all of the work, the appropriate Technical Specifications need to be drafted, setting out what work is expected of the contractor. This would generally build on the information already set out in the evaluation mandate, which must then be amended so that the **specific tasks allocated to a contractor are clearly identified** in each section, as appropriate.

Remember that the normal procurement rules apply, so make sure that you allow enough time for this in your overall planning and start drafting Technical Specifications well in advance of the intended launch date to avoid any delays²⁵.

Table 2 below provides a brief overview of the most common procurement procedures and gives indicative times for the length of the process, from the moment when the first draft of the terms of reference is ready through to signing the contract with a contractor.

Table 2 : Overview of procurement procedures and indicative deadlines

Procedure	Indicative duration
Open procedure	6-7 months
Specific contract following a Call for Expression of Interest (AMI list)	2-6 months
Framework contract of DG MARKT	Around 3 months
Framework contract of DG BUDG	Around 3 months

Note: indicative duration of open and AMI procedures is based on the average time to prepare tenders inside DG MARKT (based on 2007 tenders). It does NOT include the time required to prepare the terms of reference which depends on the resources available in the operational unit and the complexity of the subject. For the AMI procedure, the duration will additionally depend on the number of applicants invited.²⁶

Unit B.2 must be consulted on any such document and it is advisable to get the Steering Group involved in this process as well. Ultimately the Operational Unit is responsible for managing the contract.

TIP: DRAFTING TECHNICAL SPECIFICATIONS ALWAYS TAKES LONGER THAN YOU THINK – START DESIGNING THE EVALUATION AND DRAFTING ANY DOCUMENTS WELL IN ADVANCE.

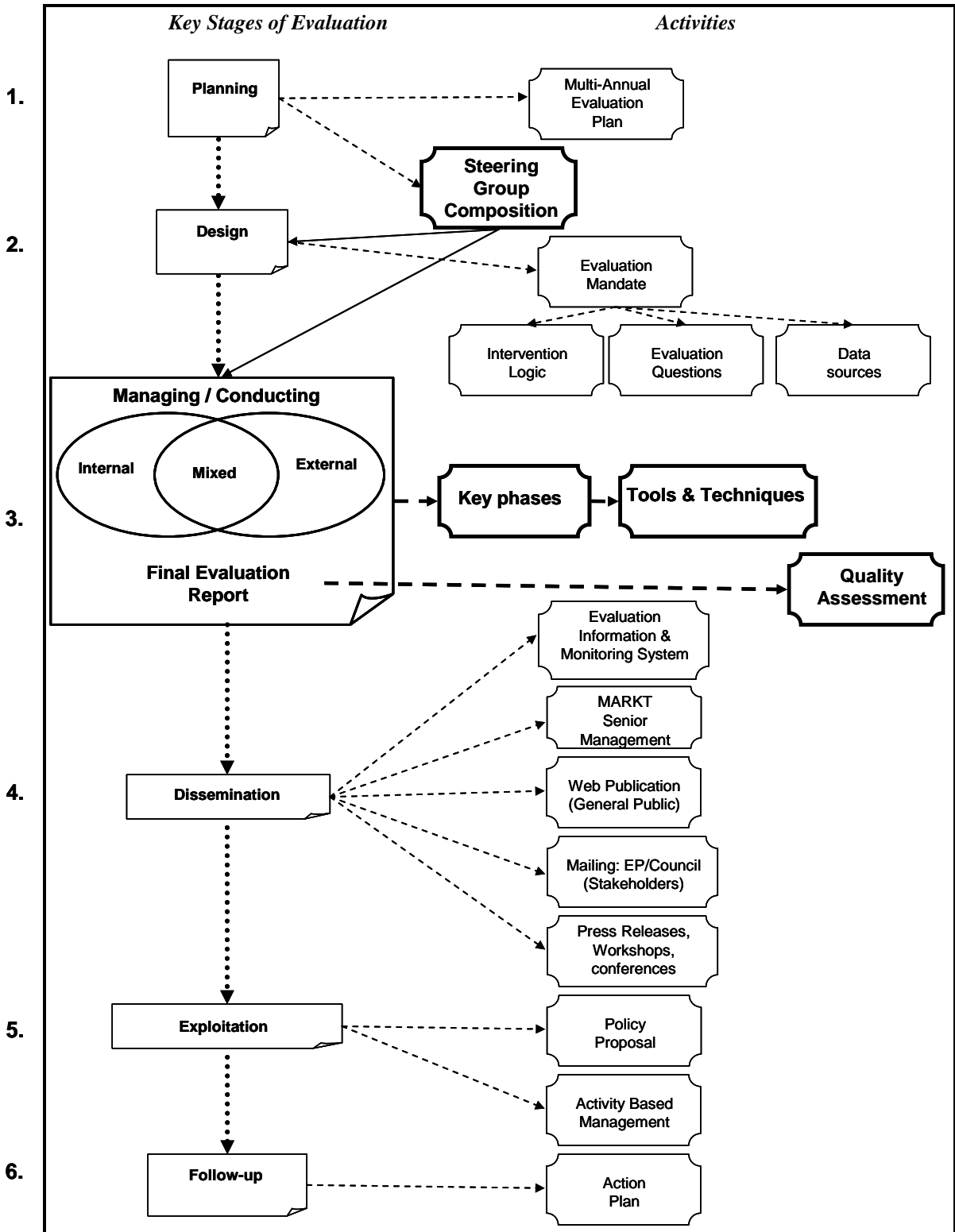
²⁵ See DG MARKT intranet > Horizontal Services > budget > tenders for Technical Specification template.

²⁶ For the more detailed information on the AMI and open procedures and the valid thresholds please consult the Budget intranet site: http://www.cc.cec/markt/intranet/index.cfm?action=domain&id_domain=702 or contact Unit A1/Budget; for information on the framework contract of DG MARKT or BUDG please contact Unit B2, or please see DG MARKT intranet > Evaluation and IA > Contractual facilities > DG MARKT – FWC.

3.3. Conducting the evaluation

Equipped with a well-designed evaluation mandate and having obtained the appropriate resources, the evaluation work should be launched according to how it has been planned and resourced. Figure 17 below shows the main stages in conducting evaluation.

Figure 17 : Stage 3 - conducting the evaluation



The evaluation may be conducted:

- entirely by an external consultant;
- entirely by an internal team made up of desk officers directly and indirectly concerned by the legislation;
- a mixture of both, e.g. contracting out the data collection and analysis, but retaining internal control over the report drafting, recommendations, etc.

Where the situation calls for a truly independent assessment to satisfy stakeholder interests, e.g. the Single Market Review, a completely external evaluation is often the most desirable approach and can add to the credibility of the findings.

However, there are instances where purely internal evaluation is more appropriate, either because of the highly specialised nature of the legislation or due to political sensitivity or quite simply for organisational reasons.

Historically, the preferred choice in DG MARKT has been to opt for a mixture of both external and internal expertise. Data collection and analysis can be quite resource intensive and it often makes sense to contract this out to evaluation experts who are more familiar with the different tools and techniques suited to this task, have the resources to cover 27 different countries, etc. On the other hand, the operational unit is more likely to have the necessary background knowledge for determining the pertinent findings to present to decision-makers.

What follows in this section is equally important regardless of how the evaluation is organised.

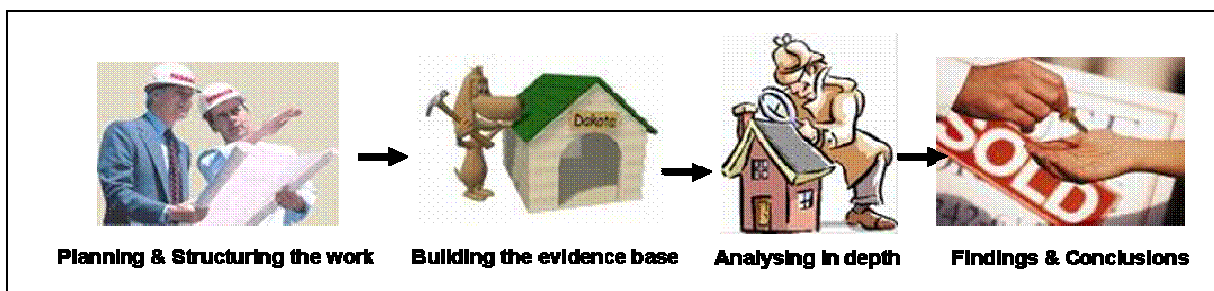
3.3.1. Key phases

To conduct a good evaluation, there are **four phases** which must be completed:

- clear **structuring** of the tasks ahead;
- comprehensive **data collection**;
- thorough and complete **analysis and validation** of findings; and
- reaching **evidence-based conclusions** and, if appropriate, recommendations.

It is like a building project: you need plans and design, so that you can source the appropriate material for the building; the work is assessed and interesting features noted that allow the estate agent to conclude an appropriately valued sale (see Figure 18).

Figure 18 : Four key phases of evaluation



Structuring expands on the design and involves preliminary analyses of readily available data in order to set out: a clear plan of what issues are of most importance for more in-depth enquiry; the proposed methodology for data collection, analysis and validation; a concrete time-table; and who will be responsible for the various aspects.

The *data collection phase* is the most time-consuming and resource-intensive part of an evaluation as this is where the fact-finding and evidence gathering takes place. There is a certain overlap with the analysis phase as on-going basic analysis of the data gathered will already be taking place. Nonetheless, the data collection phase only ends once all/enough information has been received.

Although some analysis will have already been done in the previous phase, the next phase is where the evidence gathered is *analysed and validated* in-depth with a view to answering the evaluation questions and drawing some concrete findings that are reliable and credible.

The final phase of the evaluation pin-points some clear judgments based on earlier findings. These lead to *evidence-based conclusions* and, as appropriate, proposed recommendations for future action.

In order to ensure that the evaluation stays on track, each of the four key phases should end with the production of a report showing the achievement of the associated tasks. Although it is not strictly necessary to prepare formal reports for evaluations carried out **internally**, as each section builds on the previous one, it can be useful to at least follow the format even if any documents are not subject to formal sign off. However, for **external** evaluations, these reports become a more formal requirement (as they are generally linked to the contractual obligations). Table 3 shows expected content of each report type.

Using these reports, the **quality** of the evaluation throughout its entire process²⁷ can be **monitored** by the Operational Unit, B.2 and the Steering Group. Each report should build on the previous submission, ensuring that a comprehensive final report can be delivered²⁸.

²⁷ Evaluation Standard D.5, SEC(2007)213.

²⁸ Evaluation Standard D.4: "The final evaluation reports must as a minimum set out the purpose, context, objectives, questions, information sources, methods used, evidence and conclusions".

Table 3 : Deliverables (reports) at each evaluation phase

Evaluation Phase	Expected Content	Report Type
Structuring	<ul style="list-style-type: none"> • Intervention Logic and evaluation questions; • Proposed methodology for answering evaluation questions; • Appropriateness and potential availability of data; • Preliminary analysis including typologies, review of literature, outcome of mapping exercises, Network Analysis, timetable / Gantt chart²⁹. 	Inception Report
Data Collection	<p>An initial analysis of the data collected covering:</p> <ul style="list-style-type: none"> • gaps or difficulties in accessing data; • assessment of response rates and view on balance across the representative data groups; proposal/progress for validation methods (case studies, focus groups, etc); • balance between qualitative and quantitative data; • presentation of primary and secondary data sources. 	Interim Report
Analysis & Validation of Findings	<ul style="list-style-type: none"> • A succinct and sound analysis of all the data in relation to the evaluation questions set out in the evaluation mandate; • clear evidence-based findings and preliminary conclusions. 	Draft Final Report
Concluding	<p>Taking into account the results of the quality assessment, but preserving the evaluation team's independent valid findings, the final report should encompass the entire evaluation work and may include, as appropriate, recommendations for future action.</p>	Final Report

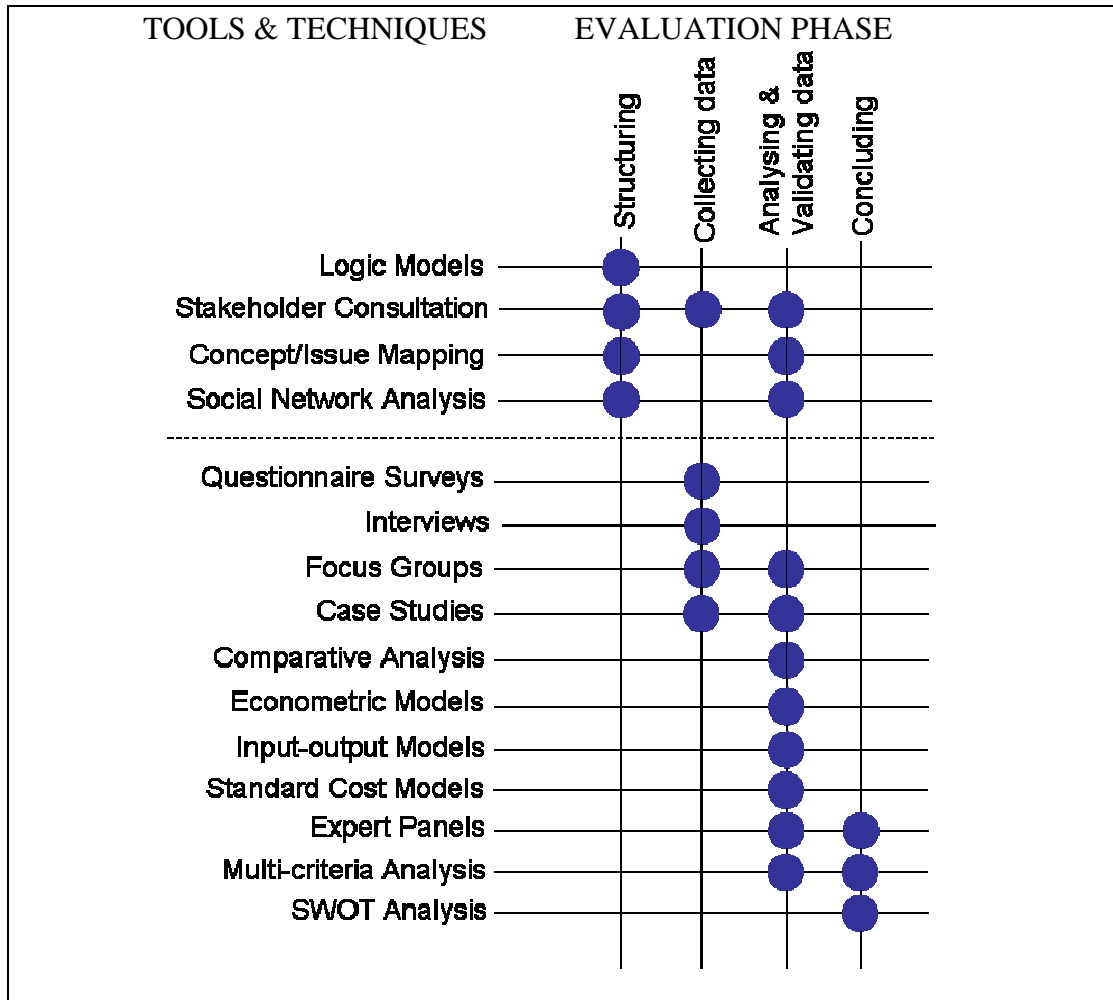
Organising regular meetings between the evaluation team (internal or external) and the Steering Group - as a minimum for the end of each phase - will allow the evaluation team to check their work is continuing in the correct direction and to request help in facilitating contact with wider stakeholder groups and/or information sources.

²⁹ Bar-code style project schedule.

3.3.2. *Tools and techniques*

The work for these four key phases is generally done using specific tools and techniques suited to evaluation. Figure 19 gives an over-view of the most commonly used tools and their relevance to each of the four phases:

Figure 19 : Different evaluation tools and their relevance to phases of the evaluation



Adapted from the EVALSED Guide³⁰

The first four tools listed for the structuring phase are highly recommended and if used, should ensure a thorough preliminary understanding of existing or basic information (mainly obtained through desk research). Executing this phase well is particularly important when evaluating Directives, Recommendations and "soft" legislation, where there are so many different aspects to consider. Using all four of these recommended tools can provide a more complete over-view of the multi-dimensional interactions that make legislation complex to evaluate, helping to identify areas which require further clarification. As a result of such analysis, the evaluation team should be better placed to select key issues that (if necessary/appropriate) the evaluation can more usefully focus on.

³⁰ <http://www.evaled.info/>.

The evaluation team can decide to use **any combination** of the tools and techniques, as they see fit and according to their suitability to help provide answers to the evaluation questions. The choice of tools clearly depends on the time planning, financial and human resources that are available for an evaluation. It should be stressed, however, that even the most sophisticated tools have some methodological limits and simply using a given tool does not give absolute certainty that the cause-effect relationships identified are 100% correct. However, using a combination of these tools can reduce uncertainty and therefore improve the quality of evaluation.

Table 4 and Table 5 provide a brief explanation of these tools/techniques and what they aim to do:

Table 4 : Description of evaluation tools and techniques highly recommended for the structuring phase

Tool/Technique	Description
Logic Models	A tool that helps to establish the expected cause-effect relationships on which the design of the legislation was based. It provides a valuable qualitative structure for defining appropriate evaluation questions. The intervention logic is one kind of logic model.
Stakeholder Consultation	Stakeholders, particularly those in the public domain, can play an important role throughout the evaluation; (see surveys, interviews, focus groups, case studies below). At the early stages of an evaluation it can be very important to obtain their perspectives and insights as to how the legislation has been implemented; this can help obtain a consensus of key issues and an understanding of how various factors have influenced progress towards success (or not). For more detailed guidance, see: http://www.cc.cec/home/dgserv/sg/stakeholder/index.cfm?lang=en
Concept/Issue Mapping	A particularly useful tool for establishing similarities/differences between impacts. During the process different types and clusters of impacts, groups and outcomes are identified, which can help focus the evaluation on specific areas. This can be particularly helpful when the design of the legislation is complex. For practical guidance on using this tool see Annex 7.
Social Network Analysis	Expanding on the logic models and using the outcome of the mapping techniques, this tool can be very effective when trying to understand the complex causal relationships that Member State action can have on legislation and the wide ranging impacts (positive & negative) can have on various groups, whilst also taking external factors into account. It can help categorise and rate/rank the causal links making it easier to establish a keen understanding of their nature and the degree to which their interaction has contributed to the cause-effect. Again, this can be useful when considering the choice of case studies or where to focus further in-depth enquiry. Practical guidance on using this tool is provided in Annex 8.

It is vital that a logic model (for example an intervention logic) is prepared early on, to assist in the formulation of the evaluation questions. However, if Issue Mapping and Social Network Analysis can also be conducted at this early stage, the evaluation questions could be significantly more focused. If the necessary data to perform this analysis is not available at an early point, it is strongly recommended that these two techniques are used during the later part of the structuring phase. Such analysis could, equally, form part of the initial tasks given to contractors. More detailed examples of what these tools involve can be found in Annexes 7 & 8.

Table 5 : Description of evaluation tools and techniques for other phases

Tool/Technique	Description
Questionnaire Surveys	When addressed to the appropriate groups, this can be an effective tool for collecting facts and opinions in a structured format. Depending on the type of questions used, different types of data can be collected – for example, closed questions allow the respondent to choose from a set of pre-defined responses; open questions permit any thoughts/views to be collected. Good question design, a practical administrative approach and some knowledge of the target population (e.g. to ensure adequate sampling for extrapolation) are necessary for a successful survey to be conducted. Within the Commission, IPM ³¹ and EBTP ³² are available and facilitate the use of this technique.
Interviews	Interviews are a way to obtain in-depth information from selected stakeholders and can be used to expand on qualitative data already obtained through other sources and/or to develop other research tools. They often provide validation of data collected and tend to be (semi-) structured, i.e. based on predefined questions formulated in an interview guide. They can be conducted face-to-face or by telephone.
Focus Groups (Workshops)	This technique involves gathering several (homogenous) groups of 6-8 to discuss and draw consensus on key data. A larger variation is the workshop which can involve a much larger number of participants. This can be a useful and fairly quick technique for validating key data and initial findings and/or obtaining further information on pertinent issues.
Case Studies	This allows in-depth study of how events actually took place in a specific context. Case studies must be selected for their relevance. They are used to uncover further data/findings and can be well suited to the analysis of implementation and impacts.
Comparative Analysis	A quantitative estimation of the difference between the situation just prior to a policy being introduced ³³ and the current situation to establish the changes which have occurred. This is particularly useful in assessing impacts on target and unknown groups, analysing before/after trends, pin-pointing the gap between winners/losers and other sectors or groups.

³¹ Interactive Policy Making tool: http://www.cc.cec/home/dgserv/digit/corporate_ict/info_systems/ecomm_projects/ipm/index_en.htm

³² European Business Test Panel: http://www.cc.cec/markt/intranet/index.cfm?action=domain&id_domain=1315

³³ The situation just prior to a policy being introduced is sometimes called the counterfactual.

Tool/Technique	Description
Econometric Models	Using economic/statistical data, such models can help to quantitatively evaluate the net effects in areas such as growth and employment. This analysis can be useful in validating the evidence for causal relationships and can be used for forecasting and responding to "what if...?" questions (sensitivity analysis).
Input-output Models	Input-output models are used in the analysis of economic activity across several sectoral impacts. Readily available statistical data may be available from EUROSTAT for use in these models. ³⁴
Standard Cost Model	The Standard Cost Model has been designed as a model for quantifying the costs of Administrative Burden in a uniform and transparent manner. It provides insight and allows for comparison of costs on national level relating to how EU Directives have been translated into national law. ³⁵
Expert Panels	In an expert panel, a group of specialists chosen for their expertise in the field are asked to critically assess all the data and analyses available and to formulate a value judgement on the evidence provided. This can be particularly useful where data for analysis has been inconclusive, or incomparable.
Multi-criteria Analysis	A technique which allows qualitative and quantitative data to be compared so that an overall picture can be established, allowing multiple stakeholders to draw comparisons according to their preferences, or several criteria at once (e.g. "Which" or "Test-Achat" reports). It allows integrated conclusions to be drawn.
SWOT Analysis	SWOT analysis assesses the S trengths, W eaknesses, O pportunities, T hreats of a given situation, and relates them to the desired objectives. Although often used in the structuring phase of a prospective evaluation, this may also usefully be applied in the concluding phase of an evaluation of legislation. A SWOT analysis is an obvious tool to use if the evaluation being conducted precedes a review.

For more detailed information on any of these tools, consult the EVALSED guide³⁰.

³⁴ http://www.cc.cec/home/dgserv/estat/index_en.html .

³⁵ For more information: http://ec.europa.eu/enterprise/admin-burdens-reduction/admin_burdens_en.htm .

3.3.3. How to develop each phase of the evaluation in practice

In this section we will underline the main elements related to each phase of an evaluation. We will also show you how some of the tools described above may be used in practice drawing on the example of the Services Directive.

Box 8 to 10 present suggestions for each phase of the evaluation of our case study, setting out the potential tasks and some appropriate tools that could be selected for use in this example evaluation. Important issues related to the use of the proposed tools and techniques are also highlighted.

⇒ **Structuring Phase**

A lot of evaluations fail because not enough attention is paid to the preparation and structuring phase. It is imperative that everyone involved in the evaluation – desk officers, Heads of Unit, managers, research team, steering group, consultants – all have a **common understanding** of the key elements being evaluated. This will help determine that the best/most appropriate combination of tools to find answers to the evaluation questions are selected.

It is very important during this early stage of the evaluation to start making all potential stakeholders and data sources aware of the evaluation. This can be done by making sure that the evaluation is discussed (briefly) at any relevant stakeholder events, committee meetings, etc.

TIP: LETTING PEOPLE KNOW IN ADVANCE THAT WORK WILL BE CARRIED OUT IN A GIVEN AREA ALLOWS THEM TO THINK ABOUT WHAT POSSIBLE CONTRIBUTIONS THEY MIGHT MAKE!

Box 8 : Case Study - possible structuring phase for evaluating the authorisation schemes of the Services Directive

1. Establish a proposed methodology:

- Building on the design set out in the mandate, the first step in the structuring phase was to set out a **project plan** outlining the tasks in more detail. Using the Evaluation mandate table on data sources, an overview of what data was readily available for preliminary analysis was established and the potential availability of data required for answering the evaluation questions was determined. Potential sources of data were reviewed and classified (e.g. qualitative, quantitative). This made it easier to **decide on the appropriate tools** to use for the first round of data collection (IPM questionnaire, phone interviews, desk research). It also ensured that there was enough **time available** to complete each phase of the evaluation (bearing in mind that there would be some overlapping of phases). Since emphasis had been placed on three sectors (construction and real estate, legal advice and consultation, tourism and leisure) additions were made to the evaluation team to ensure that there was some expertise in each of these areas. The stakeholder lists were amended to include committees and organisations in these sectors in the expectation that they would be able to provide useful inputs / data / should be included in consultation etc.

2. Desk research:

- The starting point for the preliminary analysis was a **review of existing literature** (studies, reports, etc.), summarising any useful data/views and determining gaps, areas that need complementary data or verification. All members of the unit were asked to recommend potentially interesting literature, as well as possible further sources of information. We looked

to see whether other studies had been conducted by particular Member States, academics or institutions/organisations linked to a particular aspect (e.g. one sector such as construction; or in relation to technological developments in e-commerce). One quick way to start doing this was to enter key phrases such as "studies on" + "Services Directive" or "authorisation schemes" or "construction and real estate"; etc into internet search engines.

- **Contact points** responsible for the transposition of the legislation of the Services Directive in every Member State were identified. The relevant ministries were then contacted, and informed about our intention to launch a short on-line survey to gather information about transposition and early impacts. This survey was intended to help us establish any necessary typologies (common groupings) for subsequent analysis. Some of the larger organisations/European federations were also able to nominate a contact point for issues relating to the Services Directive. The contacts were also asked to recommend any relevant reports/studies that they were aware of.

3. Preliminary data-gathering:

- We set up a **short on-line questionnaire (e.g. using IPM) to send to the contact point identified in every Member State** asking about all the ways in which they transposed the authorisation scheme legislation, including all gold-plating measures. Questions aimed to fill in any gaps or add to the overview information taken from Implementation Reports analysed in the literature review. At the same time we asked what potential impacts this legislation had had e.g. Question 1)"What, in your opinion, are the three main impacts from the Services Directive? Please rate their importance on a scale from 1 to 5 where 1=low importance and 5 = high importance"; Question 2)"What, in your opinion, are the three main impacts from the legislation relating to authorisation schemes? Please rate their importance on a scale from 1 to 5 where 1=low importance and 5 = high importance". The questions relating to potential impacts were sent to contact points in organisations/federations, but they were not asked to provide **views** about the ways in which transposition of the authorisation scheme legislation had occurred at this point. (We considered Member State information about transposition to be factual rather than opinion-based.)
- Given that some of the contact points were not able / willing to answer on-line, and also as a means of chasing those who had not answered within four weeks, we arranged a series of **telephone interviews with the contact person** permitting immediate answers to be collected. When phoning to arrange the interviews, we asked again whether they would like to reply on-line. We also sent email copies of the questions out at the same time as we confirmed the appointment, giving respondents time to think about the answers, look at specific legislation, think about some examples of where there were particular problems related to authorisation schemes in some sectors, examples of gold-plating, some sector-specific information, etc.

4. Preliminary analysis – establishing typologies:

- Information obtained from survey/telephone calls was cross-referenced to other information already available (e.g. transposition reports in the initial overview analysis).
- After assessing the initial data quality, we used the data from the survey and literature review to carry out **Concept/Issue Mapping**. This allowed us to group and estimate the importance of the various impacts identified (see Annex 7 for more detailed description and example). Doing this for the question about the impacts of the Services Directive provided some broad contextual information and background. Conducting this analysis in relation to the question on the impacts of the authorisation schemes helped us to identify three particular areas of interest: the market, migratory employment and foreign competition. Following this we re-visited our planning to re-view the timing and check whether we had sources/contacts for these issues.
- Using **Social Network Analysis** on transposition data (and/or typologies of measures), we examined the different ways that Member States had transposed the various measures relating to authorisation schemes. This highlighted that three measures (the conditions for granting authorisation; conditions for establishing authorisation schemes and selection from amongst several candidates) had been gold-plated and meant that we wanted to pay particular attention

in the following work to how this had happened and consider how this gold-plating had affected the anticipated results. The measures relating to prohibited requirements and the ability to permit limited duration of authorisation schemes only in exceptional circumstances had in general been transposed as written, so we only needed to focus on verifying whether they produced the expected results. The measures relating to authorisation procedures had been equally transposed or gold-plated making this area a possible candidate for case studies (see Annex 8 for more detailed description and example).

- Drawing on all available data/information, we updated the **Intervention Logic** used in the design phase to create a more in-depth picture.

WHAT TO WATCH OUT FOR:

1. **Make sure your stakeholders are aware of the evaluation ahead of its launch.**
2. **Timing of consultations** – for every consultation done by the Commission there are some minimum requirements that have to be fulfilled e.g. clear content, time limits for answering, feedback, etc.³⁶ – be aware of them! In the above case relating to a short questionnaire, it might be possible to reduce the time given for answering, especially if contact points have been pre-warned by telephone or been given the possibility of answering by telephone.

Think also about the need for reminders and follow-up calls in case your contact points do not answer within the deadline.

3. **Languages** – think about which languages the survey needs to be written in (and if necessary, plan the time and resources for translation) and in which languages the respondents can answer. The best practice is to publish the survey and allow answers in all 23 EU official languages – this is obligatory if you publish an open public consultation, where it is desirable that citizens in all EU 27 Member States participate. Sometimes depending on the target group (e.g. representatives of the Member States coming for Internal Market Advisory Committee (IMAC) meetings in Brussels, some lobbyists based in Brussels, associations, etc.) it might be enough to have the questionnaire available in 3 languages: EN, FR and DE. But even then, you have to allow answers in any of the 23 languages.

⇒ **Data Collection**

One of the key issues to watch out for regarding data collection is that the right kind of data for answering the questions is collected. This may seem a little obvious, but data collection can use up significant resources and this could all be wasted if the work is not focused on getting the information required. **Specific data** (e.g. number of hours per employee spent on implementing the legislation in question) rather than general data (e.g. total number of employees) is more likely to answer the evaluation questions.

A good evaluation looks for an **adequate balance of quantitative and qualitative data**. Some evaluation questions tend to require more of one type of data than another, whilst others rely on a mix of both data types. In general, a combination of qualitative and quantitative data provides a substantial evidence base for evaluative analysis as data from one source complements or confirms data from the other. For example, a question relating to *efficiency* will probably require quantitative data relating to costs in terms of money/time spent of implementing the legislation whereas questions on *acceptability* will rely more on the opinions of the stakeholders (qualitative data).

³⁶ For more information on consultation go to: <http://www.cc.cec/home/dgserv/sg/stakeholder/index.cfm?lang=en> .

One of the biggest problems is often that **appropriate data is not available**. In this case, caution is called for when using alternative data, which needs to be more carefully considered and verified. For example, there are certain implications for the robustness of results depending on whether an evaluation uses a guesstimate of costs provided by a survey or recorded data on the actual costing. When using guesstimates or proxies, the evaluation report should explain the weaknesses accordingly.

It is crucial that the **reliability of the data** being provided for analysis is assessed if you want to produce a good evaluation. No amount of data, qualitative or quantitative, is useful if it does not come from reliable and verified sources. It is important to ensure that data comes **from more than one source** or "population" group and that these sources are sufficiently large enough to be **representative** of the stakeholders and/or identified target groups. For example, if the legislation aims at encouraging small specialist retailers to supply products beyond their national borders and you have no data from these sources, but rather large amounts of data from major general stockists or supermarkets, the information may be considered unreliable and unrepresentative.

This "triangulation" method (i.e. the use of multiple sources) of ensuring reliability is also important in **avoiding bias** in any one given direction which can happen if data is not taken from a sufficiently wide range of sources. Similarly, careful attention is needed when deciding whether it is appropriate to **extrapolate data** (i.e. extend your findings from a smaller group, to the wider population) and particularly whether there is adequate reliable data to do so. When data is extrapolated, the assumptions should be clearly explained in the evaluation report.

Box 9 : Case Study - possible data collection phase for evaluating the authorisation schemes of the Services Directive

Having learnt from the preliminary analysis how the legislation relating to the authorisation schemes has been implemented across the Member States and also in some of the different sectors, it was necessary to find out about the results and impacts. Moving from the general picture built up in the structuring phase, data collection in this phase was more focused on the areas or threads identified from the preliminary analysis as the most important.

1. General data collection:

- The issue analysis and social network analysis had highlighted some particular elements to **focus** on. However the evaluation team were also influenced by the political situation (the Commission had recently been very pro-active in the field of equal opportunity) and the unexpected availability of some additional budget for case studies. We realised that whilst in some areas it might still be possible to collect data for all Member States, there were particular countries which might merit more detailed investigation e.g. countries where four or more measures had not been transposed or where there was particular evidence of very different practises in construction and real estate. From our initial analysis it also appeared that countries could be grouped into three different categories in relation to the issue of foreign competition. Due to limited resources, it was decided to investigate this issue in two countries from each of these groups more closely and then to test any findings from this more detailed analysis against the larger group. If it appeared that these detailed results were representative of the larger group it would then be possible to extrapolate to the wider EU.
- We conducted a large on-line **questionnaire survey amongst stakeholders who had been affected by the authorisation schemes legislation** in the three major sectors. Working in cooperation with our contact points in each Member States we tried to reach a variety of (national) organisations and through them, a selection of stakeholders (e.g. companies).

There were problems obtaining sufficient email addresses for some sectors in some countries. Also we realised (eventually!) that we had launched our questionnaire during a holiday period which significantly reduced the initial response rate and resulted in an extension of the consultation period, more reminder emails etc. Respondents were asked to provide their views on the practical implementation of the provisions and what had changed for them – for example: in terms of transparency; administrative procedures they had to go through to get an authorisation; time of waiting for authorisation; perceived changes in the level of competition and where it came from; changes to their customer base etc.).

- The **EBTP (European Business Test Panel)** – a tool available at the Commission – was also used to reach companies potentially affected by this legislation. EBTP was used to consult on the broader issues relating to authorisation schemes, since it provides a representative sample, consisting of around 3000 businesses of different sizes and from all sectors of the EU economy (i.e. nationality, sector and size). It also tends to have a higher response rate than in open consultations.³⁷ The main subject of this consultation was the issue of migratory employment.
- **Interviews were also conducted with a wide range of industry associations** at the EU level. At the national level we concentrated more on the three sectors which had been given particular emphasis. In this way we were able to obtain detailed information on many sector specific issues. Each interview started with a common set of questions, designed to obtain some core data, before moving on to more sector specific issues.

2. Focused data collection:

The information obtained from the earlier consultation and research identified a couple of additional issues / problems that merited further attention. The following tools were then used:

- **Focus groups** - the on-line consultation highlighted that in the majority of Member States discriminatory authorisation schemes targeting hoteliers and restaurateurs remained in place. We set up a focus group composed of various stakeholders concerned by these schemes, e.g. national (or regional) authorities, national hotel association, etc. to provide more targeted, detailed information. Further focus groups were set up to look at issues affecting SMEs in the construction sector, and tourism and leisure industries.
- **Case study** - as data collection progressed and a more robust cause-effect picture became apparent, several particular issues emerged that could be considered to describe to a larger or lesser extent how the legislation had progressed. Case studies were carried out to look in more detail at particular cause-effect "threads" that stood out (e.g. discrimination in the construction sector; measures relating to authorisation procedures, comparing countries who had simply transposed to others where gold-plating had occurred). Although in general, the simplification of authorisation schemes seemed to be linked to increased cross-border establishment, there were indications that in two or three Member States the figures were actually reducing. Since this cause-effect thread was contrary to the majority it was also included as a case study.

This second part of the data collection phase overlapped to a certain extent with the analysis phase. The focus groups and case studies were used simultaneously to provide a platform for validating data collected by other means.

WHAT TO WATCH OUT FOR:

1. **Concentrate data collection on what you really need** – it can be very tempting to ask for lots of information that in practise does not help answer the evaluation questions and hence was unnecessary/distracting.
2. **Try to ensure a balance of qualitative and quantitative data** – opinions need to be backed up by facts.

³⁷ For more info on EBTP please go to: <http://ec.europa.eu/yourvoice/ebtp/> .

3. **Ask the right questions** – and avoid ambiguous, vague questions that may be misinterpreted; **of the right people** - a zookeeper may be able to tell you how many visitors they had, but probably not how many ice-creams were bought!
4. **Make sure the sample of respondents is appropriate** – if there are 300 visitors and only two provide data, this will not be enough to provide a representative/accurate picture.
5. **Give careful thought to the selection of case studies** – they should provide robust balanced examples of how the legislation as a whole worked (examples of areas for improvement as well as good practice).

⇒ **Analysis and Validation of Findings**

For a **sound analysis** based on reliable data, there are some key elements that can ensure information has been properly scrutinised and is presented in an understandable and useful format. Data should be organised so that **comparisons** are carried out within the various information categories in order to produce an acceptable, suitably wide-ranging set of results. **Cross-analysis** of the quantitative and qualitative data is also necessary when identifying any significant patterns – at the same time, this can simultaneously confirm interesting facts discovered in earlier research.

Of course, depending on the purpose of the evaluation and the focus of research areas, there are a number of analytical tools that should **demonstrate clearly how the data collected can be understood**. These should produce the results required to answer the evaluation questions.

There are some tools, however, that are most commonly used in the analytical phase of an evaluation: Case studies, Comparative analysis and Multi-criteria analysis – all of which have caveats attached.

- **Case studies** are a particularly useful tool in examining closer any significant patterns while verifying initial findings, provided that the cases chosen are **typical of the pattern** or **area of focus** for the evaluation.
- **Comparative analysis** generally explores the net effects of trade-offs between different target groups, where it is crucial to make sure that **like is being compared with like** (i.e. NOT apples & oranges, where the only common fact is that they both grow on trees!). It is equally useful when comparing the current situation to the counter-factual data.
- **Multi-criteria analysis** – an alternative to cost-benefit analysis - will pull the analytical phase together by **synthesising results** in a manner that supports the judgements required in answering the evaluation questions. It is necessary here to **check that the weighting of criteria is undertaken without bias**, for instance by testing how sensitive the results are to minor/major changes in the weighting used.

Finally, the **findings should be sufficiently validated** during the course of this phase through corroboration with other research, with appropriate sources and/or through expert panels. Where the credibility of results is questionable i.e. results are imprecise or tentative due to issues such as the unavailability of appropriate data, this must be very clearly set out in the evaluation report.

Box 10 : Case Study - possible analysis and validation phase for evaluating the authorisation schemes of the Services Directive

As mentioned before, some aspects of this phase clearly overlapped with the data collection phase, particularly in terms of validation, where we wanted to use more in-depth data collection tools to try and corroborate data collected by other means. Analysis could only be undertaken once a sufficient amount of data had been collected and in some instances (e.g. findings in the construction sector) it was revised as further information became available.

- We looked at **Trade Statistics** for a range of Service sectors using **Input/Output Models** to investigate/confirm indications of increased cross border activity. Most effort was concentrated in the area of the 3 main identified sectors (construction and real estate; legal advice and consultation; tourism and leisure).
- Bringing together information from the on-line questionnaire and earlier interviews, we obtained sufficient data to estimate the average time taken to obtain an authorisation under the old and new systems in five sectors for 12 Member States. We then used the **Standard Cost Model** methodology to establish and compare administrative costs of the provisions for authorisation schemes across the EU. (After some discussion, the average salary data produced by Eurostat was used. The salary figures obtained from the questionnaires were not used, as within the responses provided by each country, the range was too large.)
- We conducted several **Comparative Analyses** to establish differences across sectors (e.g. hotel industry benefiting more than leisure complexes) and also whether the actual effects of the implementation of the provisions for authorisation schemes are considerably different from the initial baseline expectations. Differences between Member States were also explored, in particular in relation to the degree of foreign competition. Although we had intended to compare details for a limited number of schemes against the systems which exist in non-EU countries e.g. USA, this analysis was not conducted due to lack of available data.
- Using the results of the sectoral comparative analyses, case-studies, surveys and other data collection methods used so far, a **Multi-Criteria Analysis** across the three major sectors (tourism and leisure, legal advice and consultation, and construction and real estate) was conducted. By choosing a number of qualitative and quantitative criteria for each sector, (sectoral growth in cross-border competition, consumer confidence and quality of service provision) this analysis established the overall picture highlighting common issues across the sectors, identifying elements of success and pointing towards areas for improvement.
- **All findings were summarised** in relation to the evaluation questions.
- A significant amount of time was dedicated to **validating the emerging findings** with contact points, stakeholders and Member State committees. Three separate full-day workshops were held with stakeholders from the three key sectors. The feedback mechanisms of the consultation processes were also used, allowing respondents to see how their views had been taken into account and giving them a final opportunity to provide input.

WHAT TO WATCH OUT FOR:

- **Make sure that data has been cross-checked through more than one source** – analysing unsubstantiated data can undermine the findings.

⇒ Concluding an evaluation

1. The **conclusions** of the evaluation should be **clear, unbiased and visibly supported by the strength of the evidence** previously analysed. These conclusions are not simply the summary of all the answers to the evaluation questions, but rather the pertinent issues arising from the evaluation as a whole.

This might include presenting:

- **any lessons to be learned emerging from the findings, such as where improvements in legislation or implementation may be necessary;**
 - **outcomes of analyses of trade-offs, costs/benefits and opportunities that should/could be built on, e.g. areas for simplification, particularly where evaluation precedes a review; and**
 - **any mediating factors that can still have implications for the impact of the legislation, e.g. outstanding issues to be resolved, technological developments, market trends, etc.**
2. Any **recommendations** for future action that arise from the conclusions must also be **clear, comprehensible**, and above, all **practical**. Recommendations may also be requested from external consultants, especially where they have been contracted for the entire evaluation. However, where DG MARKT intends to issue its own recommendations based on the evaluation work and conclusions of an external consultant, it must be made quite clear who has made each recommendation.

Where an evaluation is truly mixed, with external consultants only providing part of the work, this must be made explicit in the report and the study provided by the consultant must be supplied, in full and without modification, as an annex to the evaluation report.

Box 11 : Case Study - concluding the evaluation of the Services Directive

- Some of the feedback from the validation stage relating to findings in the legal advice sector was confusing. **To test our findings further, we set up an expert panel** made up of five well known and respected legal practitioners and three leading academics. They were given access to the data and analysis available and asked to provide their assessment. In general, their findings were in line with the emerging results of the evaluation, although their suggestions did provide new and clarifying insight e.g. into reasons why reducing the legal complexity does not always lead to lower costs.
- **To assist in drawing up conclusions**, all the evidence and analysis was "allocated" to various evaluation questions allowing us to see the answers, highlighting lessons to be learned and permitting us to draw conclusions about the trade-offs that occurred alongside costs and benefits.
- We also created a **SWOT analysis**, setting out the various Strengths and Weaknesses and looking at future Opportunities and Threats. This helped us to identify what we hope are robust recommendations for future areas of work. 😊

3. The **Final Evaluation Report** should have a clear structure, be understandable to the uninformed / general reader, with a logical flow. It should set out as a minimum:
- the purpose of the evaluation;
 - what was evaluated (objectives, context);
 - how the evaluation was designed and conducted (key questions, information sources, methods used);
 - what evidence was found;
 - what conclusions were drawn.

3.3.4. Quality Assessment

The quality of the evaluation must be assessed throughout the process³⁸, culminating in a final Quality Assessment (QA) of the Final Evaluation Report. To achieve this in practice it is best to relate the appropriate sections of the QA to the corresponding phases based on the reports submitted to the Steering Group. In this way the QA is progressively built up until all its 8 sections of the QA have been fully completed.

The QA has several purposes. Its principal aim is to ensure that evaluations continually meet minimum standards, but it can also act as a useful tool in ensuring the evaluation team remain focused and on track with their research. Its second largest use is that the final QA, as an integral addition to the final report, presents the reader/potential user of the results with a considered view of the limitations within which the findings can be valued and contextualised. In turn, this can encourage evaluation teams, particularly external ones, to pay attention to the quality of the work they produce.

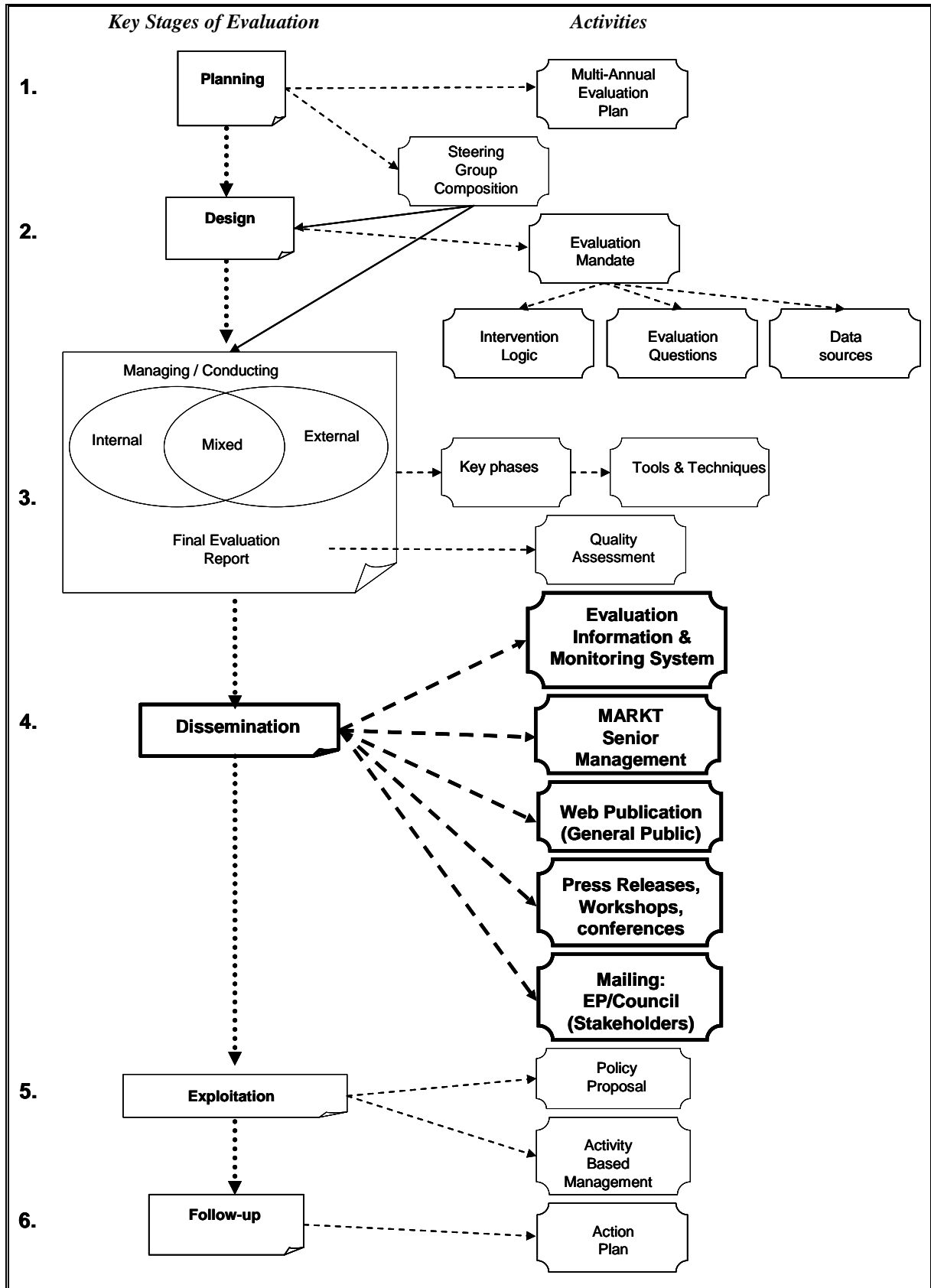
Since the QA should reflect an objective viewpoint of the evaluation work, it is important that the Steering Group contribute to (or at least comment on) these assessment phases. Nevertheless, the evaluation managers (Operational Unit and B.2) maintain the discretionary authorship for the final QA. Where the QA relates to external work the operational unit is responsible for the final assessment; for internal work, responsibility reverts to B.2 – see charter (Annex 2).

Instructions on how to apply the quality criteria are provided within the QA form/template which can be found in Annex 6.

³⁸ SEC(2007)213, Standard D.5 (Annex 1).

3.4. Dissemination Strategy

Figure 20 : Stage 4 – dissemination strategy



It is important that the **evaluation findings are made available and communicated** efficiently to the various interested parties including the decision makers (DG MARKT Senior Management, European Parliament and Council), stakeholders and the general public³⁹. Hence a coherent dissemination strategy is called for.

The dissemination strategy should identify the different audiences, the best manner in which to communicate the results to them, and include summaries of findings deemed appropriate for these various target groups.

The essential steps include ensuring that the **evaluation report and its Quality Assessment are made accessible to the general public**, usually by posting them on the Europa pages (done by B.2). Where external consultants have been involved in any part of the evaluation, such as data collection and analysis, their final report should also be published as background material.

The report and summary of findings must also be **introduced into EIMS** (Evaluation Information Management System), a repository database centrally managed by DG BUDGET to which all the Commission services submit information on all evaluations, starting at the planning stage and ending with the follow-up. Staff from B.2 will be responsible for introducing the relevant evaluation data into EIMS at each phase, based on summaries provided by/agreed with the operational unit.

Specific mailings (can be electronic) of the final evaluation report, or at least its executive summary, should be made **to key stakeholders**. A tailored **summary of results** that avoids the use of technical jargon should be prepared in the form of a **briefing** that can be used by decision-makers, particularly Cabinet and Senior Management in DG MARKT, and colleagues attending workshops or conferences where such information may be useful.

Finally, the key findings of the evaluation should be extracted into an intelligible **press release**. Other ad hoc situations may also occur for key findings to be presented, such as academic articles, responses to European Parliament queries, etc. and these should also be adapted according to their audience.

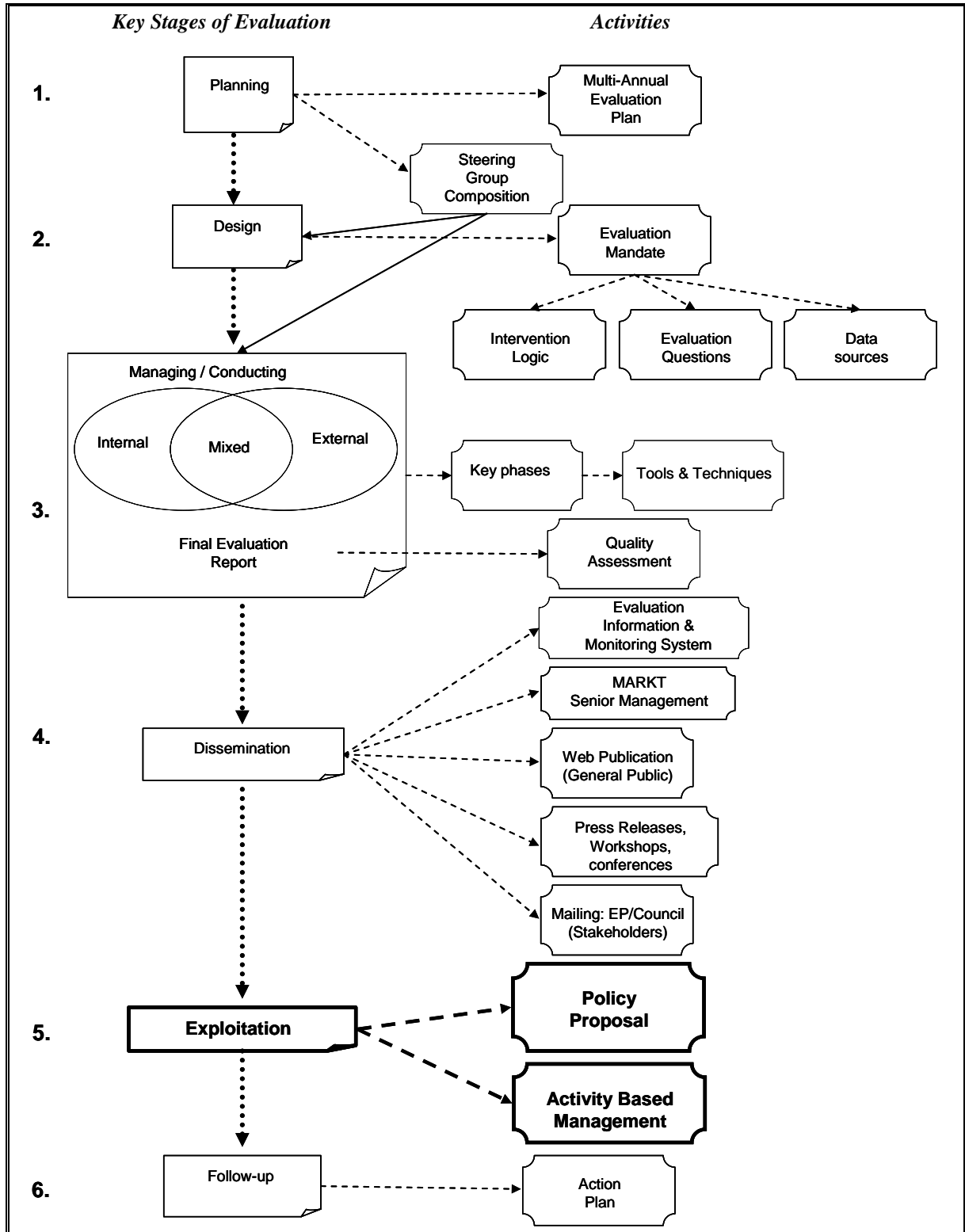
A dissemination strategy checklist can be found in Annex 9.

³⁹ SEC(2007)213, Standards E.2 and E.3 (Annex 1).

3.5. Exploiting the results

Evaluation results must be available for **use** by a variety of decision-makers in different contexts and it is very important that the key findings and recommendations are **presented** to these audiences **in a way which they will understand**. Not only should the results feed back into policy making, but in a broader sense, they should also feed into the strategic planning and programming cycle at the appropriate juncture.

Figure 21 : Stage 5 – exploiting the results



3.5.1. Policy proposal

The most obvious use of evaluation results is in formulating a new/adjusted policy proposal. The purpose of the evaluation itself may have been to establish which (if any) parts of the legislation need to undergo revision. The decisions presented in the proposal for revision can be underpinned by evidence-based findings directly or by using them in the relevant impact assessment.

Some of the evaluation results may feed directly into the impact assessment of another policy proposal related to the legislation evaluated. Where findings can be generalised, it might even be possible for them to provide input to the impact assessment of a similar proposal in a completely different policy area.

3.5.2. ABM – Activity Based Management

Throughout the various decision-making steps in Activity Based Management, evaluation results should provide the evidence-based justifications for priority setting (Annual Policy Strategy), budgetary resources (Activity Statements), attribution of resources (Annual Management Plan) and achievements (Annual Activity Report).

3.6. Follow-up Action

Even though the **Follow-up Action Plan** is the concluding stage of the evaluation process it is not a static document. It is not simply a response to the findings and recommendations of the evaluation report, but is a reporting tool showing how the action considered necessary for improving / clarifying / adapting the legislation is being carried out. As such, the operational unit must provide an update to B.2 of the Action Plan every 6 months following its approval until completion. This update will feed into the report to Directors on evaluation in the DG, helping to identify/publicise the progress being made.

The Follow-up Action Plan should **contain the key findings** appropriately organised to match the **recommendations**, with the **intended action** (or reasoned justification for not taking up the recommendation) and **provisional deadline** set out accordingly. A template is provided in Annex 10. Figure 22 below shows an example of the Action Plan done for one of the past evaluations in DG MARKT.

Figure 22 : Example extract from the follow-up Action Plan of the Citizens Signpost Service evaluation

<u>ACTION PLAN of the Citizens Signpost Service Evaluation¹</u>			
STRATEGIC ISSUES			
<u>Findings</u>	<u>Recommendations</u>	<u>Concrete actions to be taken and provisional calendar</u>	<u>Follow up (after 1 year)</u> <i>(Were planned actions implemented? If yes, when? If not, why not? ...)</i>
The operational objective to provide citizens with practical information in response to enquiries is too general. The lack of targets and long-term goals weaken the development of the service.	Establish a strategic plan including statement, goals, objectives, targets and indicators that are SMART to guide the next steps of CSS development.	Certain targets might be difficult to pre-establish, as quantitative target indicators depend on factors we can not control entirely. After agreement with hierarchy on some key strategic issues, SMART strategic plan to be drafted and agreed (<i>December 2006</i>)	
Currently not maximising the objective of CSS helping the Commission to form a better understanding of the operation of the Internal Market in practice.			

¹ Certain aspects will have to be considered over the longer term, after the expiry of the current contract in 2007. The evaluation also stresses that the potential outlined by the recommendations can only be achieved with a significant increase in human and financial resources allocated to CSS within DG MARKT.

4. Further Information & Assistance

For readers interested in obtaining a broader perspective of analytical tools for policy making, B.2's **website** contains links to wider information on evaluation in general, the Communication on Evaluation (SEC[2007]213), evaluation programmes, etc. as well as associated information on Impact Assessment.

For anyone wishing to follow **training** on evaluation in general, DG ADMIN and BUDGET offer several courses that provide an overview of the Commission's evaluation framework and general evaluation instruction.

B.2 also runs periodic **training on evaluating legislation**, which provides some practical exercises on the key elements set out in this guide. If you are likely to be involved in an evaluation in the forthcoming year, you may wish to enrol in one of these training sessions.

Of course, B.2 team members can always be contacted with any queries you may have and we would particularly like to encourage you to **contact us** at the very **early** stages of planning and designing your evaluation so we can provide you with any help you may need from the beginning.

AND FINALLY...

Happy Evaluating! – B.2 looks forward to working with you on your evaluation projects.

5. Annexes