



EUROPEAN COMMISSION
SECRETARIAT-GENERAL

Public consultation on Commission Guidelines for Evaluation

November 2013

All comments and suggestions for improvement should be sent to:

SG-EVALUATION-AND-SIMPLIFICATION@ec.europa.eu

Table of contents

Introduction.....	3
How to use this document:.....	4
Part I: Evaluation: key concepts and requirements	5
I.1 Why we evaluate	5
I.2 Legal commitments to evaluate	6
I.3 Evaluation: its role in the Smart Regulation and policy cycle	6
I.4 Evaluation defined.....	7
I.5 Evaluation criteria	9
I.6 Building up the evidence base.....	11
I.7 The different forms of evaluation	15
I.8 Quality assurance	17
I.9 Evaluation: what is required.....	21
Part II: Practical steps to conduct an evaluation	23
Introduction.....	23
II.1. Who does what in evaluations, screening and planning.....	25
II.2. Establishing a steering group	29
II.3. Establishing an evaluation mandate	30
II.4. Designing the evaluation and defining Terms of Reference	32
II.5. Research and analysis.....	40
II.6. Final report	41
II.7. Quality Assessment form	42
II.8. Dissemination plan	43
II.9. Follow-up action plan	44

Introduction

Through its Smart Regulation¹ agenda, the European Commission has committed to design, deliver and support the implementation of interventions of the highest possible quality. Evaluation is a key tool within Smart Regulation, providing a **critical, evidence-based judgement** of whether an intervention has met the **needs** it aimed to satisfy and **actually** achieved its **expected** effects. It goes beyond an assessment of whether something happened or not, and looks at **causality** – whether the action taken by a given party altered behaviours and led to the expected changes.

A key commitment in the "Smart Regulation" Communication of 2010 was the undertaking to increase the use of evaluation and the establishment of the "evaluate first" principle. This ensures that:

"all significant proposals for new or revised legislation are in principle based on an evaluation of what is already in place."²

This statement provides the cornerstone for the strengthened approach to evaluation presented in these guidelines. Providing objective, credible and critical evidence, collected and presented in an impartial and transparent manner, at a timely point in the decision making process will improve the design, delivery and implementation of EU interventions.

The following text sets out the key concepts and processes for European Commission staff involved in retrospective (backward looking) evaluations of all types of EU interventions. More detailed information and references will be provided on a revised evaluation website³. Throughout the text "intervention" and "action" are used as umbrella terms to describe a wide range of EU activities including: expenditure and non-expenditure measures, legislation, action plans, networks and communication activities.

These guidelines replace the Commission guidelines adopted in 2004⁴ and apply to all retrospective evaluations (including Fitness Checks⁵, final, ex-post and interim evaluations) of all EU interventions⁶.

¹ See COM (2010) 543 final – Smart Regulation in the European Union

² Page 6 COM (2010) 543 final

³ [website]

⁴ The 2004 guidelines can be found at:

http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf

⁵ A Fitness Check is a comprehensive evaluation of a policy area. It assesses if the regulatory framework for a policy area is fit for purpose and, if not, what should be changed. It identifies excessive burdens, inconsistencies and obsolete or ineffective measures and helps to identify the cumulative impact of legislation.

⁶ It is not necessary to carry out an evaluation of each individual project or sub-activity under an intervention, unless the sectoral regulation specifies otherwise. Even where separate evaluation is not warranted, such projects or sub-activities may, on a case by case basis, be considered as part of evaluation studies embracing larger parts of the intervention.

How to use this document:

The text is split into:

- Part I presents the key requirements and concepts underpinning evaluation at the European Commission. It is intended as a general introduction to evaluation.
- Part II provides practical assistance, explaining the different steps required to conduct an evaluation and produce a good evaluation report.
- Annexes⁷.

These new guidelines set out what needs to be done to comply with the Commission's Evaluation Standards. The latest version of these Standards is provided in Annex 1⁸. The Evaluation Standards apply to the full evaluation process: from resourcing, organisation and planning of evaluation activities to design of an individual evaluation, the expected conduct when doing an evaluation and ultimately the dissemination and use of evaluation results. They commit the Commission to conducting high quality evaluations, contributing both to Smart Regulation and better integration of evaluation findings in the Commission's Strategic Planning and Programming Cycle.

⁷ In the final version of these guidelines, the annexes will also contain templates, glossary, bibliography etc.

⁸ The previous, binding standards are presented in the Commission Communication "Responding to Strategic Needs: Reinforcing the use of evaluation" – SEC(2007)213.

Part I: Evaluation: key concepts and requirements

1.1 Why we evaluate

Evaluation is a tool to help the Commission assess whether its policies are delivering the desired effects and are ultimately improving the economic, social and environmental conditions for European businesses and citizens and contributing to the EU's global role. It is an opportunity to take stock and assess achievements systematically, to take a critical look at whether EU activities are fit for purpose and to verify stakeholder feedback, criticisms and requests for change.

Evaluation is **not just an assessment of what** has happened; it **also considers why** something has occurred **and how much has changed as a consequence**. It should add further value by providing an independent confirmation of the situation and by looking at a broader perspective.

Evaluation serves four main purposes.

- A. **To provide timely and relevant advice to decision-making and input to political priority-setting:** Evaluation is a decision-support tool. It does not replace, but aids decision-making, both at a strategic (planning) level, and at the level of the design of a new intervention. It aims to raise the quality of the debate, reinforcing the principles of Smart Regulation and administrative simplification.
- B. **Organisational learning:** The results of an evaluation should be used to improve the quality of an on-going intervention and to prepare future ones. Evaluation can take a wider look at an intervention and its environment and identify not just the areas for improvement but also the positive practices and achievements which should be widely shared and if possible duplicated in other areas. Evaluation also provides the opportunity to look for things which may have happened as a result of an intervention that were never predicted to happen – the "unintended" and/or "unexpected" effects.
- C. **Transparency and accountability:** EU stakeholders and citizens have a right to ask the Commission to give an account of what was done and achieved, not least because tax payers' money is being used to develop and fund the various interventions. This entitles citizens, stakeholders and parliamentarians to hold the administration to account and to see more clearly whether previous promises have materialised and if not what the likely reasons were and what aspects deserve special attention. Transparency can also help to increase trust, as institutions that are transparent and self-critical tend to be more trusted than institutions which do not produce realistic and objective, detailed and full assessments of the performance of their actions. By publishing evaluation findings, the Commission is publicly taking responsibility for its actions, acknowledging how an intervention is performing and inviting further feedback.
- D. **Efficient resource allocation:** Resources are limited and allocation between interventions or even between the separate elements of an intervention should be based on prioritisation of unmet societal/stakeholder needs. The final allocation may be influenced both by the estimated expectations and by any previous experience in running the same or a similar activity.

1.2 Legal commitments to evaluate

The Commission's Internal Control Standard number 14⁹ commits the Commission to evaluate all the different types of interventions it undertakes. It requires that:

"Evaluations of expenditure programmes, legislation and other non-spending activities are performed to assess the results, impacts and needs that these activities aim to achieve and satisfy."

Internal Control Standard 14 is further supported by other commitments to evaluation:

- The internal control standards state that evaluations should be performed "in accordance with the guiding principles of the Commission's evaluation standards"¹⁰. This can be audited by the DGs' audit function and the Commission-wide Internal Audit Service.
- The process and requirements for evaluation are further elaborated in the Financial Regulation¹¹ and associated rules of application¹². In particular, evaluations are required for interventions where spending is over €5 million.
- A commitment to evaluation is also included in Article 318 of the Treaty on the Functioning of the European Union (TFEU), which requires the Commission to establish an annual evaluation report of the Union's finances based on results achieved.

1.3 Evaluation: its role in the Smart Regulation and policy cycle

"Smart regulation" means delivering EU policies and laws that bring the greatest possible benefits to people and businesses in the most effective way. It covers the whole policy cycle – design, implementation, enforcement, evaluation and revision. All EU interventions – legislative or non-legislative, spending programmes or other measures - aim to achieve certain objectives through one or several means, in line with the goals and responsibilities set by the EU Treaty. To achieve the best possible quality of intervention, it is important to ensure that political choices are informed by the best available evidence on the issues to be addressed and the means to do so. In general, this is done through impact assessments when designing an intervention and through (retrospective) evaluations whilst the intervention is on-going or after it has finished.

Broadly speaking, impact assessments¹³ use evidence to assess whether a specific intervention is justified and how it should work to achieve certain objectives. Evaluations use evidence to assess whether a specific intervention was justified and whether it worked (or is working) as expected, achieving its objectives. Evaluations also analyse why something happened (or did not) and look for unintended or unexpected effects which were not predicted at the time of the impact assessment.

⁹ The internal control standards are available at:

http://ec.europa.eu/budget/library/biblio/documents/control/sec_2007_1341_annexes_en.pdf

¹⁰ Internal Control Standards, Appendix II, page 30

¹¹ See Commission's financial regulations, particularly Chapter 7, Article 30.4. The Financial regulations are available at:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:298:0001:0096:EN:PDF>

¹² See Commission's rules of application, in particular Article 18.3. The rules of application are available at:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:362:FULL:EN:PDF>

¹³ Guidance for impact assessments is available at: http://ec.europa.eu/governance/impact/index_en.htm

Impact assessments and evaluations complement each other around the policy cycle (see below). Evaluations rely on impact assessments to identify why and how an intervention was expected to work. Impact assessments rely on evaluations to identify why an intervention may not have worked as expected and whether and how it should be revised. Consultations support both tools, opening policy making to stakeholders and helping to put into practice the provisions of the Lisbon Treaty on participatory democracy¹⁴.

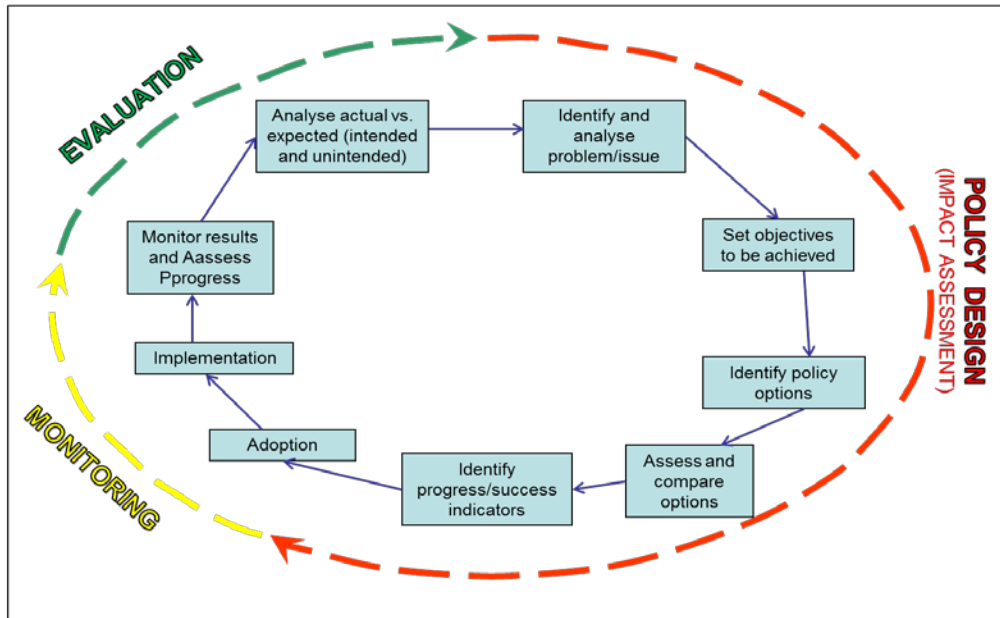


Figure 1: Smart Regulation and the Policy Cycle

1.4 Evaluation defined

Retrospective Evaluation is defined in the European Commission as: a critical, evidence-based judgement of whether an intervention has met the **needs** it aimed to satisfy and **actually** achieved its **expected** effects. It goes beyond an assessment of whether something happened or not, and looks at **causality** – whether the action taken by a given party altered behaviours and led to the expected changes and/or any other unintended or unexpected changes. As a minimum, an evaluation should look at the effectiveness, efficiency, coherence, EU-added value and relevance of an intervention or justify why this is not the case.

Looking more closely at some of the key terms in the definition:

Critical judgement: Evaluations must take a neutral point of view. They must be independent and impartial. After collecting and analysing a wide range of data, from a diverse and appropriate range of sources, the evaluator must be able to draw conclusions which identify both the good and bad consequences of an intervention. The judgement should be as specific as possible.

Evidence-based: Evaluations are not mere speculation, or modelling of the future – they are based on facts and real-life chains of events. A word commonly used in this context is « robust »: every conclusion can be traced back to evidence.

¹⁴ See Article 11 of the Treaty on European Union.

Needs: Evaluation is a public policy concept. Its focus is on what public interventions can do to meet the needs/problems of society and to improve on the current situation. EU interventions are limited by the subsidiarity principle to addressing the subset of needs which fall within EU competences. The objectives of an intervention should reflect the needs which the intervention was intended to address and as such are linked to the problems and opportunities which existed before the intervention was introduced.

Actual vs. expected: As part of the Smart Regulation process, the Commission conducts an impact assessment identifying the *expected* impacts of an intervention, providing evidence for political decision-makers on the advantages and disadvantages of possible policy options by estimating or predicting their potential future impacts. At a later point in time, an evaluation is conducted. It looks back at the performance of (completed or on-going) intervention and **critically** compares what has *actually* happened to the earlier estimates and assumptions of what was expected to happen.

Effects: Interventions are expected to cause changes – these were what was **intended** to happen. Such changes may be positive or negative. It should not be assumed that all changes are due to the EU intervention - the observed change may be caused by a range of factors.

Observed change = contribution of EU intervention (**effect**) + contribution of other factors

Interventions can sometimes have **unintended**¹⁵ and/or **unexpected**¹⁶ effects (which may also be positive or negative). A good evaluation will attempt to identify both the expected/intended and unexpected/unintended effects.

Some changes take longer to occur than others. Within evaluation, a whole range of terms are used to describe these different changes over time (e.g. outputs, results, impacts, effects, outcomes). Within this guide the term "effects" is used as a collective noun for the various changes over time.

Within the Commission, the terms (most commonly) used to describe the "changes over time" are defined below. However it should be noted that this is not a precise science; for example, often the decision to label something an output rather than result is linked to how the hierarchy of objectives has been constructed for the intervention. Differences in classification may also occur due to the specific nature of the intervention – defining the outputs and results of a legislative intervention may be more difficult than for a programme intervention. The decision whether to include the outputs of an intervention needs to be taken on a case by case basis, depending on what is being evaluated and when the evaluation takes place.

Outputs are what is directly produced or supplied through the EU intervention. They often relate to the expected deliverables of the intervention and are identified based on its operational objectives. Sometimes the output falls under the influence of the Commission Services, sometimes outputs are shared with Member States and sometimes they are completely under Member State control.

Results capture the immediate/short term changes in a situation and are defined based on the specific objectives of the intervention.

Impacts broadly define the changes over a longer period of time and are defined based on the general objectives of the intervention.

¹⁵ Unintended effects would include changes which it would not have been reasonable to predict might occur as a consequence of the intervention.

¹⁶ Unexpected effects would include changes which it could have been reasonable to predict might occur as a consequence of the intervention at the time the intervention was being designed.

Figure 2 illustrates how these different components are linked together to present the logic of the intervention. It provides a simplified model of how an intervention is triggered by a certain set of needs or problems occurring within a certain context (the cause) and how it is designed to produce the intended changes (the effects).

Based on the needs/problems identified and their corresponding objectives, the Commission decides to make or propose an intervention, intended to address those needs/problems and achieve the objectives. The Commission, Member States and other parties can be involved; they may provide inputs to the intervention or be involved in producing the outputs which are caused by the intervention. Gradually, the intervention should start causing the intended effects, improving the situation and meeting the original needs/addressing the problems identified. However, external factors may also generate the same type of effects, making it more difficult to judge whether the changes are due to a given intervention. Evaluations should try to account for these external factors, explaining any difficulties faced or resulting uncertainty.

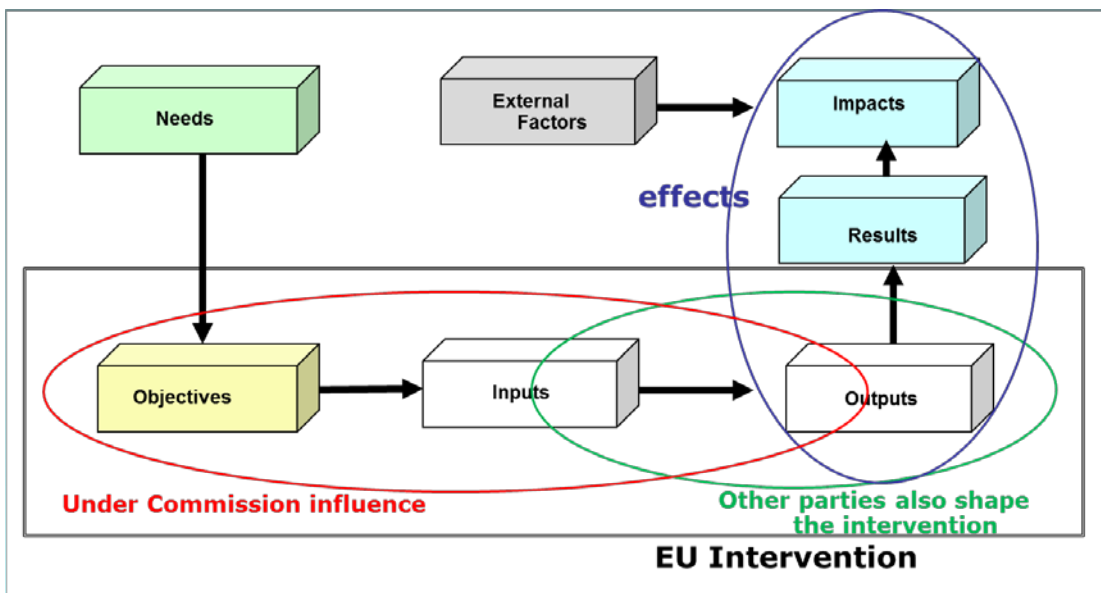


Figure 2: The different components of an intervention or the "intervention logic" (simplified)

1.5 Evaluation criteria

An EU intervention is a public policy response intended to address the needs of EU citizens and stakeholders and ultimately, to improve the status quo. Figure 3 shows the relationships between the different components which constitute an intervention. These relationships are used to frame the scope and focus of an evaluation. Often they are referred to as the "evaluation criteria" and phrased as a set of "evaluation questions" which the evaluation is designed to answer.

Unless duly justified, all evaluations must contain an assessment of the effectiveness, efficiency, coherence, relevance and EU added value of the intervention.

The scope of an evaluation must be tailored to the particular intervention and is influenced by the type of intervention and the point the intervention has reached in its own development cycle. **Where it is decided not to assess one of these five criteria in a given evaluation, a clear justification**

for this decision needs to be provided at the start of the process (in the project plan or mandate) and repeated in the final report.

The degree of analysis conducted for each criterion will depend on the timing and data availability at this point (proportionality). In some cases new data will need to be collected, analysed and compared with other findings; in others, an assessment can be presented based on existing information. The evaluation can serve as a useful tool to bring together data from a range of sources, providing a central and comprehensive point of reference for stakeholders.

The mandatory evaluation criteria considered within Commission evaluations are identified below and discussed further in Part II. For Commission evaluations it is often useful to compare differences in these criteria across the Member States. This is particularly true when considering aspects of efficiency.

The 5 mandatory evaluation criteria:

Relevance: *To what extent do the (original) objectives (still) correspond to the needs within the EU?*

Effectiveness: *To what extent did the intervention cause the observed changes/effects? To what extent can these changes/effects be credited to the intervention? To what extent do the observed effects correspond to the objectives?*

Efficiency: *Were the costs involved justified, given the changes/effects which have been achieved? What factors influenced the achievements observed?*

EU-added value¹⁷: *What is the additional value resulting from the EU intervention(s), compared to what could be achieved by Member States at national and/or regional levels?*

Coherence: *To what extent is this intervention coherent with other interventions which have similar objectives? To what extent is the intervention coherent internally?*

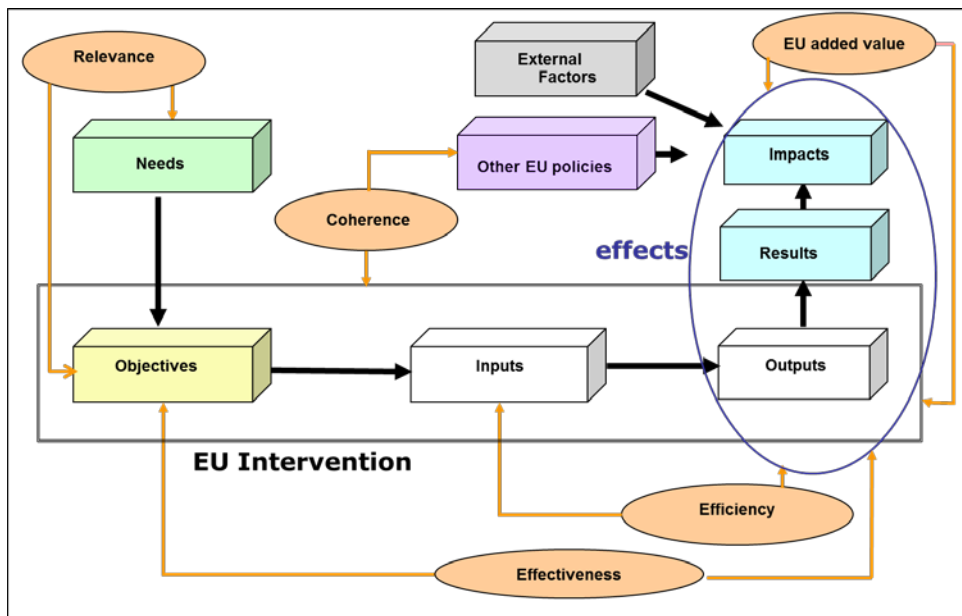


Figure 3: The simplified intervention logic and key evaluation criteria

¹⁷ For further information on this criterion see SEC(2011) 867 final "The added value of the EU budget"

Other evaluation criteria:

There are also several further evaluation criteria which it may be appropriate to consider, depending on the type of intervention and the timing of the evaluation. The most common additional criteria evaluated at the Commission are discussed in Part II.

1.6 Building up the evidence base

All interventions follow their own cycle and the desired effects can take time to manifest. Often trade-offs need to be made between when to produce the evaluation results (so that the results can feed into the decision making process) and the amount of data available (needed to provide a robust analysis and evidence base). The costs related to data collection also need to be taken into account.

Such constraints need to be recognised when planning and designing an evaluation, allowing appropriate methodologies to be used and better management of the expectations of what the evaluation can deliver. They also need to be accounted for in evaluation reports which should indicate clearly what aspects have been foregone or limited and for what reasons and with what consequences for the credibility of the evaluation.

An evaluation should be able to draw on a wide range of existing information that has been built up over time, since the start of the intervention. As soon as an intervention begins, mechanisms should be put in place to monitor the intervention and collect the appropriate data. These are generally based on earlier proposals for monitoring and evaluation (for example, the final chapter of the preceding impact assessment, if one exists). The main sources of data which may feed into an evaluation (depending on availability) are presented below. Whilst useful, it is unlikely that any of these sources will contain sufficient information to qualify as an evaluation.

1.6.1 Monitoring and Audit

Evaluation is distinct from monitoring and audit. These are in fact complementary but different exercises and serve different purposes. Whereas monitoring and audit are used primarily for the purpose of internal management and the procedure for the discharge of spending, evaluation is first and foremost an aid to decision-making.

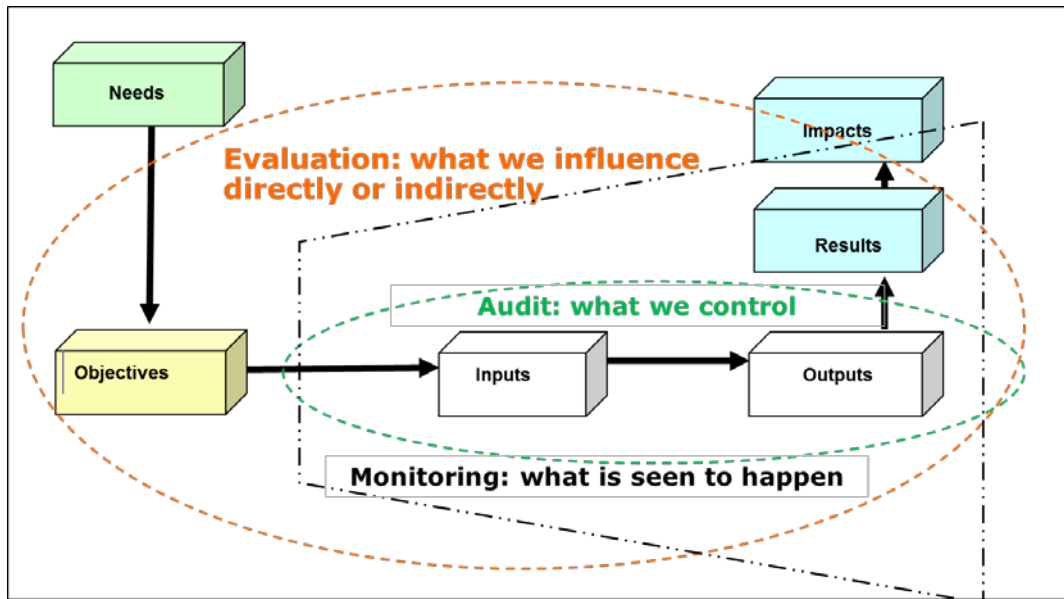


Figure 4: *Monitoring and audit – related but different*

Monitoring is a continuous and systematic process carried out during an intervention. To be most useful, monitoring (and future evaluation arrangements) should be defined at the start of an intervention¹⁸. Monitoring records observations, which can then be used to steer the intervention for internal management purposes. It enables those managing the intervention to correct deviations from the operational objectives and improve implementation. Monitoring is an essential first step in an evaluation as it provides important factual data. Note, however, that the data collected will reflect all changes – those caused by the EU intervention AND those caused by other factors.

Most **audits** in the public sector focus on the accuracy of **financial** reporting and/or **compliance** with the rules and therefore verify the effectiveness of the management and internal control systems in ensuring compliance. In the case of a **financial audit**, this involves providing an opinion on the reliability of the financial statements and verifying the legality and regularity with which funds have been received and spent in compliance with the rules. In the case of **compliance** audit, only compliance with the rules would be verified. Just like monitoring, financial/compliance audit therefore tends to be focused more on the implementation of an intervention and its inputs and immediate outputs, whilst evaluation is centered mainly on an intervention's performance in terms of its longer-term effects (causality).

In practice however, the distinction between audit and evaluation has become increasingly blurred. The Court of Auditors, the Commission's internal auditors and to some extent the Commission's programme auditors are now increasingly conducting **performance audits**, which go beyond the

¹⁸ Within Smart Regulation, the final step of the impact assessment process is to identify future monitoring and evaluation requirements. For further information, see the impact assessment guidelines available at http://ec.europa.eu/governance/impact/commission_guidelines/commission_guidelines_en.htm

traditional legality and regularity questions to analyse the "sound financial management" (i.e. the economy¹⁹, efficiency and effectiveness) of the intervention.

Performance audit is similar to evaluation. Both activities involve the examination of policy design, implementation processes and their consequences to provide an assessment of economy, efficiency and effectiveness on an entity or activity. They require similar knowledge, skills and experience and involve similar methods for collecting and analysing data. The main difference is the context in which these activities take place and the purpose of each. The auditor has a stronger focus on methodology, looking into the economy, efficiency and effectiveness of how the work was done to achieve the objectives. There tends to be less of a focus on the policy impact (e.g. economic, social, and environmental) of an intervention or comparison of the impact of policy alternatives.

In the case of evaluation, the analysis is not just an assessment of what has happened; it also considers why something has occurred and how much has changed as a consequence. Evaluation should add further value by providing an independent confirmation of the situation and by looking at a broader perspective.

To say it in a simplified manner: monitoring looks at **"what"** (what has occurred; what has been the output of the intervention), compliance audit looks at **"how"** (how the internal control systems have functioned and how resources have been used at the implementation level), whereas performance audit takes a broader look at the overall conduct of the work and its results. Evaluation looks at **"whether"** (whether the changes and movement towards the set objectives are due, at least in part, to the intervention) and **"why"** (why an intervention has been more or less successful).

In Figure 4 monitoring and audit are more concerned with the boxes (inputs, outputs etc); evaluation looks at the boxes and the arrows (causal links) which join them.

I.6.2 Studies

The Commission conducts a large number of studies many of which look back over a period of time and consider what has happened in a particular policy area. However, most studies do not contain all the critical analysis and judgement required for an evaluation (see I.9). Often they are restricted to "state-of-play" analysis, sometimes only addressing a limited number of specific aspects, discussing in detail what the current situation is.

In addition, the process for conducting these studies is often less structured and transparent than that required for an evaluation. Often study results present only part of the overall picture and serve a more limited purpose. This may, at least in the eyes of the outside reader, make the findings seem less well-founded, independent and impartial, making them less useful as a basis for drawing any broad conclusions.

I.6.3 Legally required reports

Many legal measures supporting interventions (particularly those relating to policy or regulatory) contain provisions requiring the production of different documents on the performance of an

¹⁹ Economy considers whether the inputs used for the intervention were made available in due time, in appropriate quantity and quality and at the best price.

intervention at a given point in time. The specific requirements vary from intervention to intervention. For example, Member States may have to report on what they have done in accordance with the policy or regulation, or the Commission may conduct its own assessment of these Member State actions.

Although the focus of such reports is often on national activity, they may contain information that can play a significant role in an EU evaluation, helping to identify whether the actual situation is caused by Member State actions or by the EU intervention, or a combination of both. Other information may come from reports summarising complaints received or infringement proceedings launched (or directly from the complaints received/ or infringement proceedings where available).

It is important to maximise the usefulness of all these reporting arrangements by:

- Building in an early assessment of key performance issues to start to build the broader picture;
- Ensuring the use of such reports both in on-going monitoring and by contributing to strengthening the evidence base of future evaluations.
- Ensuring a reasonable combination and timing in different forms of reporting linked to a date for a full evaluation or the final steps to be taken to complete an evaluation.

Compliance reporting (or conformity assessment)

For directives, it is important to check the compliance of national law with EU law²⁰. Earlier reports may focus on checking the completeness and correctness of national measures enacting EU law (compliance or conformity assessment reports). Although at first glance the evaluative content of compliance reports may not be obvious, they actually should provide key inputs to evaluation, providing information relating to:

- Delays in adoption and implementation which are likely to have consequences for the timely achievement of the objectives of the EU intervention;
- The form taken by the national measures introduced to comply with EU law in one area which contradicts those taken in another area, adds to other requirements so as to overburden national authorities or stakeholders with regulatory requirements, or causes other problems for the implementation of measures in another EU area (coherence in a broader sense).

Highlighting potential problems of this nature at an early stage can lead to quick and timely changes by Member States, which may even lead to earlier improvement in the effectiveness of the measure, rather than waiting for the findings of an evaluation.

²⁰ In general only directives require national transposition measures and hence compliance assessment.

Regulations can sometimes require certain national implementation measures but these are not the subject of systematic checks by the EU Commission.

Implementation reports:

Implementation reports are generally focussed on national activity and how it complies with the objectives of the EU intervention, so often having a wider policy dimension than a more legal compliance report. As such, it would not be appropriate to substitute the requirement for an implementation report by an evaluation. The first implementation report is usually requested 3-5 years after the adoption of the EU intervention – which may be too early in the policy cycle for sufficient data to be available for a full evaluation, or may constitute an initial assessment which would justify pushing back the date envisaged for a full evaluation. However, implementation reports are sometimes repeated after a further given interval [3-5 years] and **from the second report onwards, it may be possible to identify opportunities to transform the implementation report into an evaluation or link them.**

Review or assessment clauses:

Sometimes legislation contains an explicit requirement that the Commission conduct an evaluation; in many instances there is a less specific requirement that the Commission should "review" or "assess" the performance of the intervention and, if appropriate, propose changes to the intervention. These are not terms of art with a specific definition of what has to be done. Such analysis can involve both a retrospective assessment of the current intervention and a prospective assessment of possible changes – or, in Smart Regulation terms, an evaluation of the current intervention followed, if appropriate, by an impact assessment leading to future proposals for change. To avoid confusion, and to ensure a consistent approach, existing planned reviews can retain the scope defined when they started but must clearly indicate if they constitute evaluations, or a less extensive forms of reporting, based on a justification linked to the particular stage reached in the intervention's cycle. Planned reviews may also be expanded into evaluations.

1.7 The different forms of evaluation

There are different types of retrospective evaluation.

In terms of the **timing** of the evaluation, ***interim/mid-term, final and ex-post evaluations*** can be distinguished. ***Interim*** (sometimes called mid-term evaluations) are conducted particularly whilst the intervention which has an end-date is on-going (i.e. the majority of evaluations conducted on EU legislation are interim). Where an intervention has a completion date (e.g. a spending programme), it is possible to distinguish between ***final*** evaluations, which are conducted on the conclusion of the intervention, and ***ex-post*** evaluations which may take place several years after the intervention finished. Where an action, such as a policy or regulatory measure, has no end-date and is therefore on-going unless and until some further action is taken, a retrospective evaluation – looking back at how the measure has performed compared with the objectives set – takes the form of an interim evaluation.

Depending on who is producing the final evaluation report, evaluations can be classified as ***internal, external or mixed***. ***Internal*** evaluations are completely conducted by Commission Services. ***External*** evaluations are delivered by contractors external to the Commission. ***Mixed*** evaluations combine both approaches – the final report is usually written by the Commission services based on support work delivered by external contractors.

The internal evaluations defined above should not be confused with the ***evaluation of internal services of the Commission***. This latter evaluation relates to a special form of evaluation conducted by Directorates General responsible for internal policies concerning areas such as Translation, Interpretation, Human Resources and Security, the Publications Office and Eurostat. They can have a significant overlap with the content of performance audits.

Reflecting the subject of an evaluation, there are ***process and impact evaluations***. ***Process*** evaluation is concerned with how a policy was developed and delivered, assessing the processes/activities involved in its implementation. Process evaluations also have significant overlaps with performance audits. ***Impact*** evaluation focuses on what differences a policy made, looking further than the simple outputs, to see whether the desired changes over time occurred. Many evaluations contain both process and impact elements.

Another common form of evaluation is ex-ante evaluation. As stated earlier, this is a form of prospective (i.e. forward looking), rather than retrospective (i.e. backward looking) evaluation and hence is not covered by this guide. In the Commission ex-ante evaluation is now predominately conducted in the form of impact assessment.²¹

Traditionally, Commission evaluations have been conducted on *individual* interventions, which may themselves have involved a range of projects or actions.

Wider experience with policy or legislative evaluation has led to the creation of ***the Fitness Check***²² – the comprehensive evaluation of a policy area (i.e. the evaluation of *a group of related interventions* that are linked by a common set of objectives). A Fitness Check should assess whether the regulatory framework for a policy sector is fit for purpose and provide the basis for policy conclusions on the future of the relevant regulatory framework. It should identify any excessive regulatory burdens, overlaps, gaps, inconsistencies and/or obsolete measures which may have appeared over time, and to help to identify the cumulative impact of legislation. A Fitness Check is therefore like an evaluation of an individual measure except that it covers a group of measures which have some relationship between them which justifies their being evaluated together.

The evaluation of individual interventions and Fitness Checks of policy areas are complementary and mutually reinforcing tools. Evaluations of individual initiatives, although indispensable, do not always show the full picture and a more strategic and global view is often required: Fitness Checks respond to this need.

The lessons learned during the pilot phase²³ testing this form of evaluation have confirmed the concept of the Fitness Check as a type of evaluation and indicate the valuable role that they can play within policy and decision making.

²¹ These are covered by the Impact Assessment Guidelines

http://ec.europa.eu/governance/impact/commission_guidelines/commission_guidelines_en.htm and the Ex-Ante evaluation: A Practical Guide for Preparing Proposals for Expenditure Programmes

http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/ex_ante_guide_2001_en.pdf

²² The concept of a Fitness Check was introduced in COM (2010) 543 final – Smart Regulation in the European Union.

²³ Pilot Fitness Checks were carried out during the period 2010 – 2013 on policy areas falling under the remit of five Directorate Generals (ENV, EMPL, MOVE, SANCO and ENTR). Further information can be found at

http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm

The Commission is also committed to undertaking joint evaluations together with interested Member States.²⁴ A series of pilots for these *"joint" evaluations* are planned to be launched during 2013²⁵.

I.8 Quality assurance

Continuous quality assurance is necessary throughout the evaluation process to:

- provide relevant and timely evaluation results of high quality;
- ensure accessibility of evaluation results via clear and transparent communication to decision-makers and other relevant stakeholders;
- guarantee the best possible use of evaluation results;
- deliver consistency in evaluation across the Commission.

The following sections set out how quality assurance is provided within the Commission evaluation system.

I.8.1 Transparency

Transparency can help to increase trust in the workings of the EU Commission. Institutions that are transparent and self-critical tend to be more trusted than institutions which do not produce realistic, objective, detailed and full assessments of the performance of their interventions. To improve the transparency of its evaluation activity, the Commission will publish systematically and centrally:

- its 5-year evaluation planning by Directorate General;
- the evaluation mandate, terms of reference (where relevant), final report and quality assessment of the final evaluation report for each individual evaluation.

I.8.2 Clear rules: the evaluation standards

The Commission Evaluation Standards (see Annex 1) set out the governing principles which apply to all evaluations and provide basis for the quality assurance along the evaluation process.

²⁴ There is a tradition of joint evaluations for spending programmes with co-management between Commission and other relevant involved parties, e.g. external aid projects.

²⁵ See the 2012 Communication on Regulatory Fitness COM (2012) 746 final

The standards are grouped into five categories, reflecting the key stages of evaluation:

Standards	Governing Principle
A. Resourcing and organisation of activities	Evaluation activities must be appropriately organised and resourced to meet their purposes.
B. Planning	Evaluation activities must be planned in a transparent and consistent way so that relevant evaluation results are available in due time for operational and strategic decision-making and reporting needs.
C. Design	Evaluation design must provide clear and specific objectives, and appropriate methods and means for managing the evaluation process and its results.
D. Conduct	Evaluation activities must be conducted to provide reliable, robust and complete results.
E. Dissemination and use of results	Evaluation results must be published and followed up to ensure maximum use of the results and that they meet the needs and interests of decision-makers and stakeholders.

The Evaluation Standards provide the quality assurance through the following measures.

I.8.3 Evaluation function within each DG

The Commission's organisational framework for evaluation is decentralised. Individual Directorates General are responsible for the evaluation of their activities and can structure their exact evaluation arrangements to reflect the Directorate General's needs and requirements. Each DG must have an evaluation function with a clearly defined responsibility for co-ordinating and monitoring evaluation activities of the DG (from the planning of evaluations until their dissemination and use), promoting quality evaluation and organisational learning and assisting the central services in the implementation of the Commission Evaluation Policy. Within a DG, the evaluation function is the guardian of evaluation quality assurance.

The evaluation function provides the first support/check in ensuring that the DG's evaluation framework is appropriately designed to meet DG's needs while respecting the Evaluation Standards. It should contribute to clear definition of the tasks, responsibilities, organisation and procedures for all actors involved in planning, designing and conducting evaluations and disseminating and using evaluation results. Through early involvement in the evaluation process, the evaluation function has to ensure that the required Evaluation Standards and rules are met for every evaluation carried out at a DG.

I.8.4 Steering group

At the beginning of each evaluation (internal, external and mixed), a steering group must be set up. The group should have at least three participants drawn both from the operational unit responsible for the area under evaluation and from other units. It must always include a representative from the DG's evaluation function. Where the evaluation touches competences of several DGs, it is important that they are invited to participate in the steering group too. This is particularly important for Fitness Checks, where the scope can often interest more than one DG.

The role of the steering group is to steer the project, provide input and information and to improve the quality, impartiality and usefulness of the evaluation. By overseeing a minimum set of mandatory tasks, the steering group should make a strong contribution to the quality of an evaluation, providing input at all the key steps in the process. The steering group encourages a pooling of knowledge and - being composed of members from different functions - brings together a range of different perspectives. This mix of experience should prevent particular biases from overtaking the evaluation and encourage critical judgements to be made.

To assure the quality and independence of the evaluation, the steering group must be involved particularly at the stage of evaluation mandate, terms of reference and all deliverables, especially the final evaluation report. In terms of quality assurance, one of the most important tasks of the steering group is to contribute to the quality assessment of the evaluation final report.

I.8.5 Quality Assessment of the final evaluation report

According to the evaluation standards, the quality of an evaluation must be assessed on the basis of pre-established criteria throughout the process. To ensure a consistent approach to quality assessment of evaluation reports, a minimum set of quality criteria have been identified (see II.7.3 for the list of criteria).

The Steering Group contributes to Quality Assessment (QA) of the final report. The QA provides a judgement of whether key aspects of an evaluation meet the required standards and provides any related comments. Sometimes, a first draft QA is provided on the draft final report, allowing the Commission staff/external contractor to address certain weaknesses identified (at the discretion of the report author) while maintaining their independence.

The QA of the final evaluation report is published alongside the evaluation report to (i) inform the public of the opinion of the Commission services about the quality of the final report, (ii) to encourage contractors to deliver a high quality final deliverable.

I.8.6 Fostering independence

Robust and reliable results can be delivered only by independent evaluations. An evaluation can be considered as independent when evaluators

- (i) can carry out their tasks without influence or pressure from the organization,
- (ii) are given full access to all relevant information required, and
- (iii) have full autonomy in conducting and reporting their findings.

I.8.7 Review panel

A cross-DG Quality Review panel could be set up, to comment on the quality of finalised evaluations and contribute to the improvement of future practices. It could meet twice a year and report on a random selection of finalised evaluations, identifying strengths, weaknesses and lessons learned²⁶.

²⁶ For other possible suggestions, see COM (2013) 686 final: Strengthening the foundation of Smart Regulation – improving evaluation

I.9 Evaluation: what is required

To build on existing good practices and satisfy the "evaluate first" principle, it would be required that the following steps are met to qualify as a Commission evaluation:

A. The evaluation is followed by a **Steering Group** with a minimum of three members, including at least one from the Directorate General's evaluation function. Where the intervention clearly overlaps with the responsibilities of other Directorates General, they must be invited to participate in the steering group. The steering group will be consulted on the mandate, terms of reference (where applicable) and all draft reports. It will contribute to the quality of all deliverables and provide input to the quality assessment of the final report, based on a minimum set of pre-determined criteria.

B. All evaluations **assess the evaluation criteria of: effectiveness, efficiency, coherence, relevance and EU added value** of the intervention, or provide due justification why this is not the case. Other evaluation criteria may be added, as appropriate. The exact level of analysis provided for each criterion will differ according to the intervention being evaluated and the point reached in the intervention's lifecycle.

C. An **evaluation mandate** is published centrally, early in the process, indicating:

- i. The purpose and scope of the evaluation;
- ii. The background and objectives of the intervention being evaluated (including a short description of how these were expected to be achieved);
- iii. The expected start and publication date of the final report;
- iv. Contact details, allowing stakeholders to provide feedback.

D. Where external contractors are involved, the **Terms of Reference** for their work are published centrally on the same website as the evaluation mandate.

E. The evaluation follows a clearly defined **methodology** which addresses the evaluation criteria which have been selected. Where a prior (Commission) impact assessment exists, the same broad categories of impacts must be analysed. Credible efforts must be made to obtain data from a wide range of qualitative and quantitative sources. Any consultation undertaken must comply with the Commission's minimum standards for consultation.

F. The **final report** contains:

- i. An executive summary of no more than (6) pages;
- ii. Critical judgement, based on a range of data, of the evaluation criteria which the evaluation intended to address;
- iii. A clear summary of the methodology followed and a final assessment of the limitations of both the approach taken and the data used;
- iv. A clear chain of logic between the analysis and findings presented, the answers to the evaluation questions and the conclusions drawn.

G. As a minimum, the evaluation findings are **disseminated** via the central publication of the final report, alongside the evaluation mandate and quality assessment.

H. The Commission Services in charge of the evaluation draw up a short document outlining their responses to the conclusions and recommendations made in the final evaluation report. They also identify appropriate **follow-up** actions within six months of the completion of the final evaluation report. As useful, regular progress updates must be provided to senior management.

Additional requirements for a Fitness Check:

As a type of evaluation, a **Fitness Check** is subject to basically the same requirements identified above. However, since a Fitness Check is a comprehensive policy examination, it must:

- * **assess the five evaluation criteria** (effectiveness, efficiency, relevance, coherence, EU added value) **without exceptions;**
- * be based on a **more comprehensive evaluation mandate** which includes both the evaluation questions and a working description of the intervention logic;
- * be followed by a **steering group which includes members from the Secretariat General** and other Directorates General active in the same policy area;
- * produce a **final report written by the Services of the Commission**, taking a minimum format of a staff working document.

Part II: Practical steps to conduct an evaluation

Introduction

The key purpose when doing evaluations is to learn about how EU policies have made a difference on the ground and what the EU institutions should improve. The "end-user" perspective should therefore be central to the analysis.

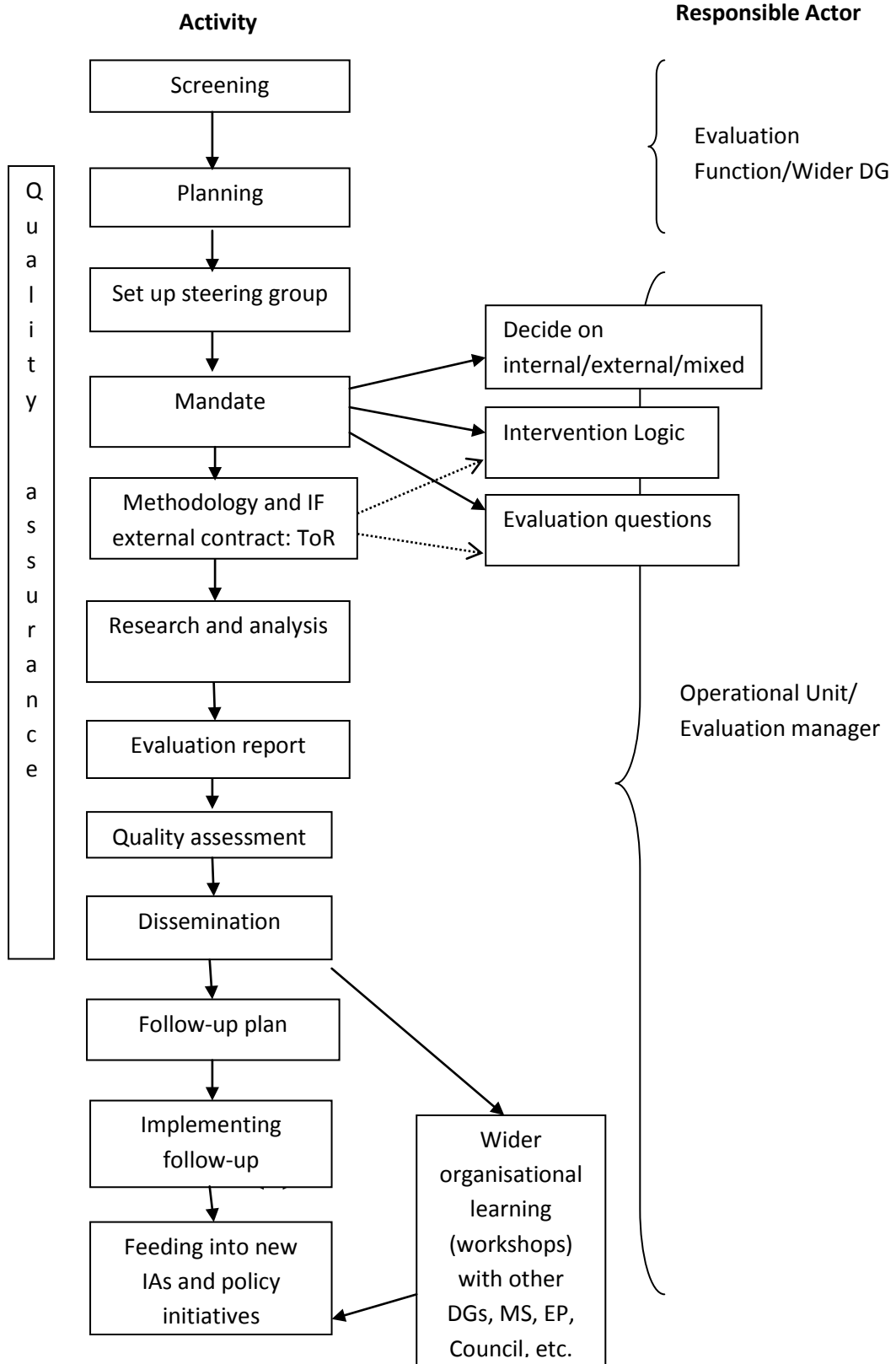
A good final result is likely to be achieved when the planning is transparent, minimum criteria for the research questions and methodology are fulfilled and quality assurance occurs throughout the project. The aim is to get an impartial and objective assessment with constructive criticism from evaluators. The final report should be easily readable and the findings be disseminated and followed up by concrete actions.

The following sections contain boxes with the **minimum requirements** on how evaluations must be done but also provide **further recommendations**.

The document provides guidance through the main steps of an evaluation, following the chronological order of actions needed to organise and conduct an evaluation:

1. Who does what in evaluations, screening and planning
2. Establishing a Steering Group
3. Establishing an Evaluation Mandate
4. Defining Terms of Reference
5. Research and analysis
6. Final Report
7. Quality Assessment form
8. Dissemination Plan
9. Follow-up action plan
10. Feeding information back into new impact assessments

Figure 1: Process of an evaluation



II.1. Who does what in evaluations, screening and planning

Before an evaluation can be conducted, a number of preconditions should be in place: clarification of who does what during evaluations, careful screening and planning of what to evaluate when. The following sections explain these aspects.

II.1.1. Who does what in evaluations?

II.1.1.1. Who manages and coordinates evaluation work?

The Commission's organisational framework for evaluation is decentralised, making individual Directorates General responsible for the evaluation of their activities. The choice of structure is at the discretion of each Directorate General, reflecting its needs and requirements.

- In line with the Commission's evaluation standards, ***evaluation functions*** are set up within the Directorates General to co-ordinate and support their evaluation activities and ensure the independence of evaluation. The staffing level of an evaluation function needs to be sufficient to enable the function to participate fully in each main stage of the process of each individual evaluation.
- ***Individual evaluations*** are run either by the evaluation function or by the operational unit responsible for a particular intervention.
- Further support is provided by the ***Secretariat General*** which is responsible for supporting measures, including provision of general guidance and training on evaluation and ensuring that the evaluation standards are respected.
- There is also a Commission-wide ***evaluation network*** which meets several times a year to discuss evaluation issues and share information.

II.1.1.2. Who does the research and writes the Final Evaluation Report?

- The vast majority of evaluations are ***out-sourced*** i.e. the Commission hires consultants to collect and analyse the relevant data and answer a set of evaluation questions. If the consultants do all the work to answer the questions and their final report is the evaluation document, this is called ***external evaluation***. The work that the consultant is required to do is set out in the *Terms of Reference* written by the Commission services. The evaluation project is overseen by a *Steering Group*.
- Occasionally Commission Services write their own evaluation report based on the input from the consultants and possibly other work they have undertaken (this is known as a ***mixed evaluation*** as it relies both on external and internal inputs).

- On rare occasions, no consultants are involved and the evaluation is run and delivered completely by Commission Services (this is known as an *internal evaluation*).²⁷

II.1.1.3. Who contributes information to an evaluation?

Member States, stakeholders, academics, citizens and a wide range of other parties are involved by providing data and opinion about interventions and wider policies. By contributing to, reading and reacting to evaluation reports, they provide further direct input to the decision making system. They play an important part in testing findings and driving independent and impartial evaluations.

II.1.2. Screening

In order to comply with the Commission evaluation requirements²⁸ and ensure important inputs for the evidence base of future Impact Assessments and proposals for new initiatives, **valid evaluation information should be continuously available for all Directorate General's interventions/policy areas.**

To ensure that all initiatives are covered, the Director General should ensure screening of all initiatives taking place in DG's remit to identify:

- legal obligations to carry out an evaluation,
- when the individual initiatives were evaluated last time (and ensure data are continuously available),
- check the potential to carry out Fitness Checks on initiatives with common objectives instead of running separate evaluations on individual initiatives.

Mapping carried out under the REFIT programme²⁹ (first conducted at the start of 2013) should serve as a basis for this exercise. However, the screening must go beyond the relevant acquis and has to cover also non-legislative initiatives of the DG.

II.1.3. Planning

Good evaluation planning is a critical step to ensure the availability of evaluation results that could feed back into the policy making cycle. Evaluation activities must be planned in a transparent and consistent way so that relevant evaluation findings are available in due time

²⁷ It should be noted that the term "internal evaluation" in these guidelines means that the data collection, research and writing of the Final Evaluation Report is *done by Commission staff*, without help from an external consultant. The term "internal evaluation" should therefore not be confused with an evaluation of Commission activities that provide a service *within* the Commission (to other Commission departments, for instance translation services, human resource services, etc) which can also be evaluated.

²⁸ Evaluation Standards of 2007:

http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf

²⁹ The REFIT Communication COM(2012) 746 final requires systematic mapping and planning:

http://ec.europa.eu/governance/better_regulation/documents/1_EN_ACT_part1_v8.pdf

for operational and strategic decision-making and reporting needs.³⁰ Each Directorate General will ensure that it has monitoring and evaluation planned at the time of adoption by the Commission of each significant initiative to develop or amend EU action.

II.1.3.1. Rolling evaluation plan

The planning of evaluation activities of individual Directorates General takes the form of (minimum) 5-year indicative rolling programme, where the plan is broadly fixed for the first 2 years and stays more indicative for later ones, providing an overview of the structure and coverage of the evaluation policy. The plan needs to be annually updated (but may also be updated in year, if necessary).

In principle, all policy and programme Directorates General should be conducting at least one evaluation (or Fitness Check) each year.

In practice, the evaluation plans are integrated in the Management Plans of individual Directorates General according to the format specified in each year Annual Management Plans circular. The Secretariat General compiles the individual Directorates General's evaluation plans into a Commission evaluation plan made available via a central website. This ensures transparency of the Commission's evaluation activities, enhancing stakeholder access, allowing them to know what is being evaluated when and further enabling their participation. In addition, each Directorate General should indicate a central evaluation contact point within their plan (for example a functional mailbox address), which is responsible for answering questions from stakeholders on the planning, timing and progress of work on evaluations all year round.

II.1.3.2. Coverage

As explained in the section about screening above, all of a Directorate General's policy areas, including both spending and non-spending interventions, must be proportionally covered in the rolling evaluation plan. Valid evaluation information should be continuously available for all DG's policy areas.

Next to the evaluations planned systematically based on the screening of initiatives in a DG's remit, thematic or ad-hoc evaluations might need to be carried out as a result of new strategic decisions, implementation problems, indications coming from monitoring results, complaints from stakeholders, infringements procedures, audit reports etc. Such additions should be added to the evaluation plan as they occur.

II.1.3.3. Timing

Timing is a crucial factor for fostering the use of evaluation in the decision-making process; it should enable the results to be fed into decisions on the design, renewal, modification or suspension of activities. Evaluation planning should take into account the life cycle of the intervention and the operational and strategic decision-making needs of the DG.

³⁰ Evaluation Standard B.

Planning for evaluations of *spending initiatives* covered by the Multi-annual financial framework is generally a relatively straight forward exercise³¹. The Multi-annual financial framework covers a seven year period, which predefines the timing of evaluations (the timing, nature and number of contributions and reports expected are normally defined in the legal basis of the spending programmes). In order to allow the mid-term (or interim) evaluations results to feed into ex-ante evaluations (impact assessments) for the subsequent programming period, the mid-term evaluation results need to be delivered around the fifth year of the programming period. Most ex-post evaluations are expected to take place a few years after the programming period ends.³² Although the timing of the mid-term evaluations might not be ideal (i.e. potentially insufficient time has elapsed since the beginning of the programming period to allow for materialisation of the intervention's results), carrying out the evaluations later would not allow the evaluation results to be used in Impact Assessments for the preparation of the subsequent programming period. Equally there would be no time to make improvements to the current programme.

It is much harder to apply a uniform timetable to the evaluation of *non-spending activities* which take such a wide range of formats (e.g. regulation, directive, decision, recommendation, opinion, action plan, communication, trade agreement etc.). Regulatory activities in particular, have very different cycles. Member States can be given relatively long or short periods of time to complete the transposition of the EU law into national legislation; sometimes varying conditions in Member States allow for derogations or different schedules. In other instances, not all the components of the regulation take effect at the same time with a corresponding delay in the arrival of change etc.

The 'evaluate first principle' has to be taken into account when planning evaluations. The 'evaluate first principle' requires that a revision of an intervention which is expected to require an impact assessment (particularly those expected to be included into the Commission Work Programme) in year "n" should be able to rely on an evaluation of the performance of EU action over recent years as a basis on which options for the future can be considered. Such evaluation should be finalised in the year "n-1" at the latest, given that the evaluation results would feed into the subsequent prospective impact assessment. For instance, if you foresee that a legal act should be revised/adopted in 2015, an impact assessment would need at least to be started in 2014 and so a retrospective evaluation must be finalised at the latest early 2014.

Studies

Coordination has to be ensured when it comes to planning studies and evaluation. The Directorate General should avoid unnecessary overlaps of studies and evaluations. Note: Studies cannot be used as a substitute for evaluation (see I.6.2).

³¹ Different set up applies for Structural Funds.

³² In some cases, the mid-term evaluations of the current programming period cover also ex-post evaluations of the previous programming period.

II.2. Establishing a steering group

II.2.1. What is a steering group?

A steering group is a group of people who work in the same or related areas as the subject of the evaluation, who meet to discuss all aspects of the evaluation – from setting it up, through to accepting the final report and identifying the follow-up actions.

II.2.2. Why is a steering group important?

The role of the steering group is to steer the project, provide input and information, ensure quality, impartiality and usefulness of the evaluation. As well as ensuring policy coherence, it provides a constant quality control, overseeing all the key steps in the process and not just providing feedback at the end, when it might be too late to improve the evaluation. The steering group encourages a pooling of knowledge and brings together a range of different perspectives. This mix of experience should prevent particular biases from over-taking the evaluation and encourage critical judgements to be made.

A good steering group providing strong quality assurance can significantly increase the credibility of an evaluation. By ensuring that different perspectives are discussed, analysed and reported it improves the quality of the evaluation and demonstrates an open and transparent approach to critically analysing performance and delivery.

II.2.3. Minimum obligations

- The evaluation is followed by a Steering Group with a minimum of **three members**, including at least one from the DG's evaluation function.
- Where the intervention clearly overlaps with the responsibilities of **other DGs**, they must be invited to participate in the steering group.
- The steering group will be **consulted** on the mandate, terms of reference (where applicable) and all draft reports.
- It will contribute to the quality of all deliverables and provide input to the quality assessment of the final report, based on a minimum set of pre-determined criteria.

II.3. Establishing an evaluation mandate

II.3.1. What is an evaluation mandate?

The mandate is the "project plan" for the evaluation. The mandate for an evaluation plays a similar role to the roadmap for a Commission initiative³³ – it sets out the purpose of the evaluation and provides key information on the scope, timing, data and analysis planned to be used.

II.3.2 Why is an evaluation mandate important?

The mandate **will be published** on the Europa website, allowing stakeholders and interested parties to know from an early point in the process, what will be investigated, when they will have the opportunity to participate and when to expect the results.

An additional practical reason for drawing up a careful mandate is that most evaluation results should at some future stage feed into future impact assessments. It is therefore important to ensure from the beginning that all evaluations will, amongst others, produce the key information needed for further policy and decision making.

II.3.3. Minimum obligations

1. An evaluation mandate (project plan) must be published centrally, early in the process, indicating³⁴:

- i. the purpose and scope of the evaluation;
- ii. the background and objectives of the intervention being evaluated (including a short description of how these were expected to be achieved);
- iii. the expected start and publication date of the final report;
- iv. Contact details (e.g. functional mailbox), allowing stakeholders to provide feedback.

2. Scope of the evaluation:

All evaluations must **assess the evaluation criteria of: effectiveness, efficiency, coherence, relevance and EU added value of the intervention**, or provide due justification why this is not the case (in the mandate). Other evaluation criteria may be added, as appropriate. The exact level of analysis provided for each criterion will differ according to the intervention being evaluated and the point reached in the intervention's lifecycle.

Please note that the five minimum criteria to be covered in all evaluations (effectiveness, efficiency, coherence, relevance and EU added value) still allow DGs to perform any kind of **other studies** that do not address these criteria. Studies and other kinds of reports can also constitute the building blocks for a later full evaluation, only perhaps needing some updating

³³ http://ec.europa.eu/governance/impact/planned_ia/planned_ia_en.htm

³⁴ It is expected that a standard evaluation mandate template will be developed.

or confirmation as part of the final process. But such other studies then cannot be called "evaluations" and cannot be used to comply with the "evaluate first principle".

II.4. Designing the evaluation and defining Terms of Reference

II.4.1. What is meant by design?

Designing the evaluation means identifying the sequence of tasks to be done and range of methods to be used during an evaluation.

- Where Commission outsources the work to *external consultants*, the Terms of Reference is the document which contains the description of the work on offer to potential contractors. It identifies the aims and scope of the project, evaluation questions and the key deliverables, setting out key dates and forms the basis upon which contractors present their offers.
- Where the Commission performs the evaluation *internally*, such information should already be available at an early planning stage and should be included in the mandate rather than creating a separate document.

II.4.2. Why is the design/are the Terms of Reference important?

The vast majority of evaluations are ***out-sourced*** i.e. the Commission hires consultants to collect and analyse the relevant data and answer a set of evaluation questions. The work that the consultant is required to do is set out in the *Terms of Reference* written by the Commission services. It becomes part of the contract and is enforceable and entails penalties if the contract is not performed as required.

If an evaluation is conducted purely ***internally*** (without the help of an external contractor), no Terms of Reference are needed. In this case also, the evaluation mandate and the steering group are very important. There must be a project plan with a list of tasks to be performed, be it in the mandate or in addition to the mandate, to keep the evaluation project on track and ensure the quality of deliverables. It is also essential that the desk officer performing the evaluation participates at least in Commission Evaluation Training Modules 1 and 3.

N.B. the level of detail and specificity of a Terms of Reference can depend on the different procurement procedures.

II.4.3. Minimum obligations

Where external contractors are involved, the Terms of Reference for their work must be **published centrally** on the same website as the evaluation mandate.

This should ensure that the final Evaluation Questions and further relevant information, e.g. a draft of the intervention logic are made available to the public and are easy to find alongside the mandate.

II.4.4. How to produce the Terms of Reference in practice

Key aspects are:

- Defining the scope (which legal acts, actions, projects to look at, which period to look at, which countries to look at)
- Explaining the intervention logic
- Drafting good evaluation questions.
- Designing the evaluation so that a range of different data is collected and analysed and different methodologies used. The need to cross-check/double-check all information should be emphasised. An eye should be kept on matching the available budget and the expectations towards research that can realistically be done by the contractors.

II.4.4.1. Defining the scope

The scope of the evaluation should be defined in terms of:

- Legal instrument(s) (e.g. Directive X plus its three implementing acts)/EU initiatives covered
- Period covered (e.g. 2006-2012)
- Geographic area in which impacts will be observed (e.g. all Member States or just a few, third countries or not)

II.4.4.2. Explaining the intervention logic

If the intervention logic has not already been explained in the mandate at least a draft version should be included in the Terms of Reference. Contractors could then be asked to refine it either as part of their bid (thus demonstrating their knowledge of the subject) or during the early stages of their research and to test it during the evaluation project.³⁵

II.4.4.3. Drafting good evaluation questions

Good evaluation questions encourage critical analysis. By defining the questions at the start of the process, the Commission clarifies what it intends to analyse and invites interested parties to provide relevant contributions.

All evaluations must assess the evaluation criteria of: **effectiveness, efficiency, coherence, relevance and EU added value of the intervention**, or provide due justification why this is not the case (in the mandate). Additional criteria beyond these five can be added. For instance, when evaluating development assistance programmes, the criterion of 'sustainability' may also need to be considered.

The scope of an evaluation must be tailored to the particular intervention and is influenced by the type of intervention and the point the intervention has reached in its own development cycle. Where it is decided not to assess one of these five criteria in a given

³⁵ In purely internal evaluations (done by Commission staff without the help of external consultants and therefore without Terms of Reference), the intervention logic will have to be developed by the operational unit itself and verified by the steering group.

evaluation, a clear justification for this decision needs to be provided at the start of the process (i.e. in the mandate) and also repeated in the final report.

The degree of analysis conducted for each criterion will also depend on the timing and purpose of the evaluation (proportionality). In some cases new data will need to be collected, analysed and compared with other findings; in others, an assessment can be presented based on existing information. The evaluation can serve as a useful tool to bring together data from a range of sources, providing a central and comprehensive point of reference for stakeholders.

For example, at an "early" stage in the intervention's lifecycle, it may not be possible to judge the criteria relevance in any depth. If it is only a few years since the intervention has been made, it may be fair to assume the continued relevance of the action; alternatively, stakeholder feedback may be the only indicator of whether needs have changed. Equally, EU-added value may be difficult to judge in the early years, particularly if the early changes are related to setting up new organisations or putting in place a framework.

The **five mandatory evaluation criteria** to be considered within Commission evaluations are discussed below:

For Commission evaluations it is often useful to compare differences in these criteria across the Member States. This is particularly true when considering 'efficiency' findings.

Relevance: *To what extent do the (original) objectives (still) correspond to the needs within the EU?* Since needs are continually evolving and changing, consideration of relevance can be important when it is necessary to form a judgement of whether the intervention is still needed, or whether it needs to be altered in some way. An intervention may not achieve the expected level of success for a range of reasons. For example, the wrong "problem drivers" may have been identified during the impact assessment; incorrect assumptions may have been made about the cause and effect relationships; circumstances may have changed and the needs/problems now are not the same as the ones looked at when the intervention was designed.

Effectiveness: *To what extent did the intervention cause the observed changes/effects? To what extent can these changes/effects be credited to the intervention? To what extent do the observed effects correspond to the objectives?* The intervention was expected to deliver a certain set of changes/improvements. Evaluation should form an opinion on the progress made to date and see if the objectives have been met. If the objectives have not been achieved an assessment should be made of the *extent* to which progress has fallen short of the target. It should be assessed whether the objectives can still be achieved and, if so, with what delay. It should also identify the factors influencing *why* something was successful or *why* it has not yet been achieved. The analysis should also try to identify if any *unexpected* or *unintended* effects have occurred.

Efficiency: *Were the costs involved justified, given the changes/effects which have been achieved? What factors influenced the achievements observed?* The full efforts to support

and perform an intervention can be broken into different categories such as: staff, purchases made, time and/or money spent, fixed costs, running costs, etc. These costs can be associated to different aspects of an intervention and judged against the changes achieved.

Differences in the way the intervention is approached and conducted can have a significant influence on both the cost of making the intervention and on the changes that happen as a result, making it interesting to consider whether the use of other policy instruments might have achieved the same effects at less cost (or greater effects at the same cost). Efficiency analysis should also consider particular Smart Regulation concepts such as simplification potential and administrative burden. ***N.B. efficiency looks at the ratio of inputs (or "costs") to outputs/effects (or "benefits").*** Although an analysis of costs can be very important, in terms of efficiency analysis, it only provides half the story.

EU-added value: *What is the additional value resulting from the EU intervention(s), compared to what could be achieved by Member States at national and/or regional levels?* Under the principle of subsidiarity (Article 5 Treaty on European Union), the EU should only act when the objectives can be better achieved by Union action rather than by potentially varying action by Member States. It requires consideration of the value and improvements which are caused by the EU rather than another party taking action.

The sources and nature of this additional value vary from intervention to intervention. It is, in particular, useful to distinguish the European added value of an EU policy measure in general (like an EU regulation to foster the single market) and that of an EU spending programme per se. In both cases, European added value may be the results of different factors: coordination gains, legal certainty, greater effectiveness, complementarities etc. In all cases, measurement is a challenge and the final judgement on whether expected added value would justify an EU programme is ultimately the result of a political process. This point has been stressed on several occasions by the European Institutions³⁶.

Coherence: *To what extent is this intervention coherent with other interventions which have similar objectives? To what extent is the intervention coherent internally?* Many interventions share a set of common objectives and it is important to check how well they work together. Coherence can be internal i.e. between the objectives of an intervention, and external i.e. with other policy areas. An example of internal coherence would be to consider how well the intervention's components combine to achieve growth which is both smart and sustainable. External coherence can be very far reaching – for example EU water policy contains several measures designed to improve the overall quality of freshwater. This is affected by actions taken under the Common Agricultural Policy (CAP) and the responsible DGs have to interact to ensure that the two policies remain coherent.

³⁶ Text in this paragraph was taken from p.3 of SEC(2011) 867 final " The added value of the EU budget".

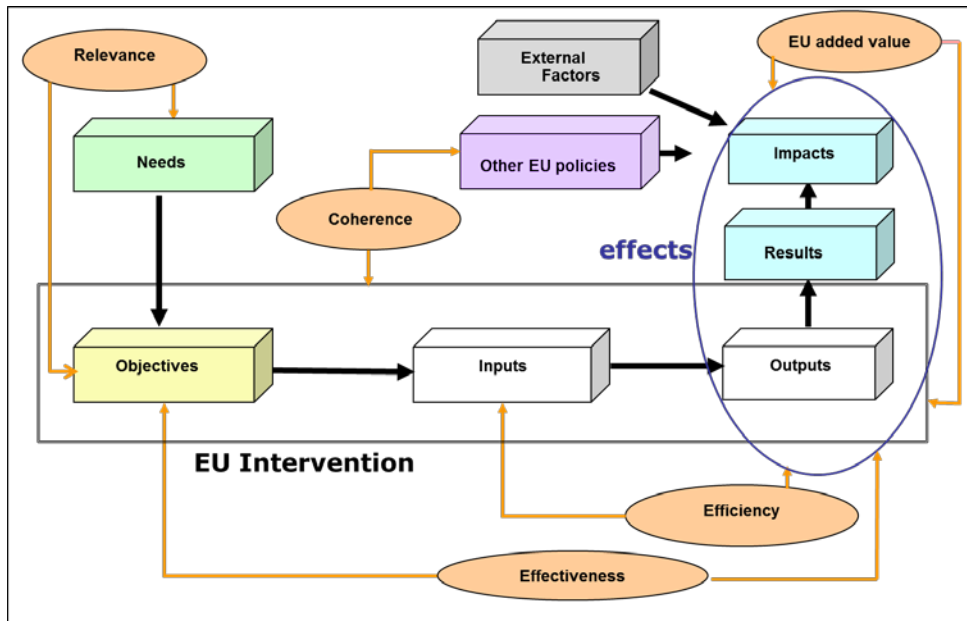


Figure 2: The simplified intervention logic and key evaluation criteria

Other evaluation criteria:

There are also several further evaluation criteria which it may be appropriate to consider, depending on the type of intervention and the timing of the evaluation. The most common additional criteria evaluated at the Commission are shown below.

Utility: *To what extent do the changes/effects of an intervention satisfy (or not) stakeholders' needs? How much does the degree of satisfaction differ according to the different stakeholder groups?*

Complementarity: *To what extent do EU policies and interventions support and usefully supplement other policies (in particular those pursued by the Member States)?*

Coordination: *To what extent are **interventions** organised to maximise their joint effects, e.g. by mobilising resources combined with harmonising measures?*

Equity: *how fairly are the different effects distributed across the different stakeholders / regions? / genders?*

Sustainability: *How likely are the effects to last after the intervention ends? It is often hoped that the changes caused by an intervention are permanent. It can be important to test this expectation for interventions which have a finite duration, such as particular programmes.*

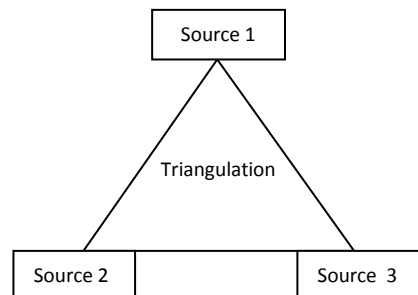
Acceptability: *To what extent can we observe changes in the perception of the intervention (positive or negative) by the targeted stakeholders and/or by the general public?*

II.4.5. Cross-checking information to identify robust evidence

Good evaluation studies are based on data (facts) and arguments from *several* sources and analysed with *several* methods, to avoid a one-sided approach and biased results. If several unrelated sources point to the same direction, this increases confidence that the findings are robust. Ideally, the analysis should be based on three different sources (sometimes called "tri-angulation"). In simpler terms, it means that arguments should not just be reported, but double-checked against other sources of information and the results are recorded in the evaluation report.

It is important to understand that evaluation should **not be based only on stakeholder views**. It always implies a careful analysis of stakeholders' arguments and a double-checking against the arguments of other stakeholder groups and, where possible, against information from independent third parties or official statistics.

To capture the "end-user perspective" in an evaluation, it needs to be carefully checked whether to rely on data and arguments by *organised* stakeholder groups at national or European level is sufficient, or whether it would be better to reach out directly to final beneficiaries or end-users (by interviewing a **representative sample** of individuals - consumers, farmers, travellers, students, business owners, etc.) which we hoped would benefit from a policy or have had to bear its cost. Likewise, quantitative data should always be complemented with and double-checked against qualitative information from other sources (interviews, etc.). This means for instance that an evaluation should not just present an econometric model, its results and limitations, but should always seek to get further confirmation by asking stakeholders in how far and why the results make sense to them.



For example:

Source 1: EU wide stakeholder survey

Source 2: National monitoring data

Source 3: Study comparing EU and US, based on academic research, stakeholder survey, etc.

On the one hand, it is important to encourage good quantitative data collection. On the other hand, the emphasis should be on the careful interpretation of quantitative data, not simply on including more quantitative data in evaluation reports.

II.4.6. Considering evaluation design/methodologies

The methodology of an evaluation is usually composed of a combination of tools and techniques assembled and implemented in order to provide answers to the specific questions posed within that evaluation, with due regard for the time, budgetary and data constraints of the project.

As desk officer in the Commission, you will not be expected to become an expert in evaluation methodology because such expertise is normally easily available from external consultants. However, for several reasons it is still necessary to have a general understanding of various methods:

- To estimate whether consultants' offer to use a certain number of methodologies in an evaluation study is realistic in view of the time and budget available.
- To recognise the strengths and weaknesses of various methods in your particular case.
- To get a feeling of which methods can be well combined to ensure that the data is based on several sources of information and the analysis is based on several perspectives (triangulation).
- If you are ever asked to perform an internal evaluation, you will need a reasonable overview, before having to acquire a deeper understanding of the particular methods you want to use.

Overview of methods

Certain techniques are particularly useful at a particular stage during an evaluation. Annex XYZ [to be added later] contains a detailed list of methods (techniques) that is suitable in each of the four stages of an evaluation:

- (1) Structuring the project,
- (2) Observing what has happened,
- (3) Analysing how and why it has happened
- (4) Judging whether the intervention achieved its objectives well and coming up with ideas for improvements in the future.

The terms of reference should leave the choice of methods to external contractors and rather concentrate on defining good evaluation questions. The contractors should be asked to explain in their bid the advantages, limitations and risks involved in using the proposed technique.

N.B. A contractor should be able to explain even the most complex method to a non-expert. If they are not able to do this, the method should be rejected as too uncertain to be likely to form a sound basis for a judgement which should be well-understood by stakeholders and the general public.

A critical review of methods that can help put contractors' proposals into context is included in 'Evalsed' Sourcebook.³⁷

Specific rules for evaluations of EU regulation

- While the scope of any evaluation needs to be defined early in the project, this is particularly important in the case where 'regulation' is to be evaluated. Regulations are often related to other regulations as parts of systems. Therefore the definition of the

³⁷ The Sourcebook is available at:

http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm [Note: under revision]

regulation to be evaluated should be put into context of its relation to other rules and its role within the logic and contours of the relevant regulatory framework.

- Where appropriate, evaluation of regulation should include an assessment of administrative burden, simplification potential, impacts on small and medium-sized enterprises (SMEs) and the global competitiveness of EU business as part of the analysis of "efficiency". As explained above, "efficiency" in an evaluation context addresses the question whether the impacts have been achieved at reasonable cost.
- For **administrative burden assessment**, a definition and relatively simple EU Standard Cost Model has been developed by the Commission and agreed with the Member States. The Standard Cost Model can be found in Annex 10 of the Commission's Impact Assessment Guidelines. An administrative burden calculator is available at: <http://adminburden.sg.cec.eu.int/default.aspx>. A considerable amount of data is available that can inform new evaluations.

- For the **SME aspects**, it is important to consider how easy it has been for SMEs to comply with the legislation and whether they incurred disproportionate costs in comparison with their limited staff and turnover. For example, it makes a big difference whether a reporting obligation that creates 5 person-days of work per year needs to be implemented by a multinational enterprise with thousands of employees which may absorb this cost relatively easily, or whether it needs to be implemented by a micro-enterprise that consists only of the entrepreneur plus one employee.

If micro-enterprises were already exempted from legislation in the past, it needs to be evaluated whether this resulted in clear benefits for them.

In all discussions about costs, it should be remembered that the key evaluation criterion " efficiency " <i>implies a ratio of costs to benefits</i> (or in other words: inputs to intended effects) which in turn requires an assessment of both costs <i>and benefits</i> .
--

Key data sources about infringements and complaints

In particular for regulatory evaluations, it is important to verify whether any infringements have been detected or complaints have been received from stakeholders and what they say about the overall performance of the measure.

II.5. Research and analysis

II.5.1. What is meant by research and analysis?

During the monitoring process or research phase of an evaluation, data should be collected from a wide range of sources on all the relevant aspects of the evaluation. This data is then analysed to provide the evidence base for the evaluation. The analysis may be quite complicated (e.g. statistical models) or simple (e.g. stakeholder opinions based on number of people saying yes or no).

II.5.2. Why is this important?

Monitoring results, research and analysis are core activities to conducting an evaluation and later to be able to summarise the findings and draw conclusions. It is important to be constantly checking for weaknesses in both the data e.g. Do they come from a reliable source? Have enough respondents replied? and the analysis e.g. Are the survey questions clear and simple? Do they cover a sufficient time period and identify any trends? Can the modelling be repeated?

In external evaluations, the research and analysis is normally done by contractors. The role of the evaluation manager and the steering group is to steer the project and advise contractors, as well as supervise their work, ensure the quality of the work and enforce the timetable.

II.5.3. Minimum obligations

- The evaluation must follow a clearly defined **methodology** which addresses the evaluation criteria which have been selected. The Methodology chapter of the final report should contain a section summarising the methodology followed and a final assessment of the limitations of both the approach taken and the data used.
- Where a prior (Commission) **impact assessment** exists, the same broad categories of impacts must be analysed.
- Credible efforts must be made to obtain data from a **wide range of qualitative and quantitative sources**. It should be accepted that single sources of conclusions are not as strong as conclusions confirmed by multiple sources. Opinions of interested parties need to be assessed for their objectivity. The setting of reasonable limits to the collection of data and opinion should be explained.
- Any **consultation** undertaken must comply with the Commission's minimum standards for consultation.
- While evaluations should be carried out in a participatory manner, evaluators should protect **the anonymity and confidentiality of individual information** and ensure the dignity of those interviewed.
- Evaluators need to be given enough **freedom to exercise their professional judgement** at the end of an evaluation project, based on the evidence gathered during the evaluation. No pressure should be exerted on evaluators to change their draft conclusions and recommendations, unless the latter contain obvious errors.

II.6. Final report

II.6.1. What is the final report?

The final report is the key deliverable of the evaluation process, presenting the critical judgements and answers to the evaluation questions. It summarises the evaluation and presents the research, analysis, findings and conclusions/recommendations of the evaluation, providing input to the next round of decision making.

II.6.2. Why is the final report so important?

The final report is the key document that stakeholders will see (and possibly comment upon). It is also the basis for the follow-up plans. Last but not least, it serves as indirect feedback mechanism to acknowledge the contributions that stakeholders and experts have made throughout the process.

II.6.3. Minimum obligations

The Final Report **must contain**:

- i. An **executive summary** of no more than 6 pages;
- ii. **Critical judgement** , based on a range of data, of the evaluation criteria which the evaluation intended to address;
- iii. A clear summary of the **methodology** followed and a final assessment of the limitations of both the approach taken and the data used;
- iv. A **clear chain of logic** between the analysis and findings presented, the answers to the evaluation questions and the conclusions drawn.

II.7. Quality Assessment form

II.7.1. What is the Quality Assessment form?

The Quality Assessment form is the official "sign-off" of the work done by the Steering Group. It judges whether key aspects of an evaluation meet the required standards and provides any related comments.

II.7.2. Why do we need one?

If evaluations are to increase transparency, accountability and organisational learning, the Commission needs to strive for highest quality evaluation reports, based on robust methodologies.

- During the project, we need quality assurance because experience shows that research projects can easily miss their targets or go off track. Contractors have an incentive to minimise the workload for certain contracts which can lead to suboptimal quality of their reports. The Commission needs to keep up the practical support and motivation of contractors, but also keep a certain pressure on them to deliver high quality results.
- According to the evaluation standards (D5), the quality of an evaluation must be assessed on the basis of pre-established criteria throughout the process. To ensure consistency across the Commission, it makes sense for all DGs to use a minimum common list of criteria, which they can then expand if they wish.

II.7.3. Minimum obligations

The minimum criteria to be covered by the quality assessment form are:

- Are the final report and its executive summary well structured, clear, frank and easy to understand?
- Is the scope of the evaluation performed in line with the announcements about scope in the mandate and/or terms of reference?
- Is the methodology robust? Are the assumptions and limitations clearly explained?
- Were sufficient data and arguments collected and double-checked/triangulated?
- Is the "end-user perspective" adequately addressed?
- Is there a clear and logical chain between evaluation questions, evidence, findings, conclusions and recommendations?
- Are the recommendations (if provided) useful?

II.8. Dissemination plan

II.8.1. What is the dissemination plan?

The dissemination plan lists the different audiences for the evaluation findings and identifies if different summaries of the final report need to be written, tailored to the needs of the different groups.

II.8.2. Why is dissemination important?

The purpose of evaluations, namely to promote accountability/transparency and organisational learning, can only be achieved if the information produced by such evaluations reaches those to whom we are accountable to (general public, parliaments, etc.) or certain intermediaries (journalists) and those who should learn from the results. All evaluation reports of high quality should therefore be disseminated in a manner suited to the different audiences. Active discussion and debate on these findings should be encouraged.

The European Parliament has asked for immediate access to all completed evaluations financed by the EU budget. Studies, including evaluations, are to be communicated to the EU Bookshop which will maintain a database and publish those not subject to restriction.

II.8.3. Minimum obligations

- Dissemination of the evaluation findings via the **central publication** of the final report, alongside the evaluation mandate and quality assessment.
- The Commission Services in charge of the evaluation draw up a short document outlining their responses to the conclusions and recommendations made in the final evaluation report.

II.9. Follow-up action plan

II.9.1. What is a follow-up plan?

A follow-up action plan identifies the actions which it has been decided to take as a result of the evaluation. Each follow-up action should be assigned to a particular unit/individual and a target completion date agreed. There is no set format for such a plan – for example, it can be set up as a simple Excel table or word document.

II.9.2. Why is a follow-up action plan important?

The follow-up action plan shows in a transparent manner, how the evaluation is contributing to future decision making. It should ensure that the evaluation findings are used and not simply ignored. Where no action is to be taken, the follow-up plan should clearly state this and explain why that is the case. It also presents an expected timetable, allowing stakeholders to know when to expect certain actions.

The follow-up action plan is also important in ensuring that the DG/Commission take ownership of the final evaluation results. This may mean accepting all the findings and deciding on suitable follow-up action. It may also mean that some conclusions/recommendations are rejected or not acted upon. In the latter case, the DG/Commission will provide a justification in the follow-up action plan.

It has to be acknowledged that for policy evaluations, the action plan can often not be completely decided at service level because major policy changes require Commissioner/College endorsement. Therefore a distinction between policy recommendations and administrative-organisational recommendations might be needed where mainly the administrative part can be published immediately.

II.9.3. Minimum obligations

- The Commission Services identify appropriate follow-up actions within six months of the completion of the final evaluation report. As useful, regular progress updates must be provided to senior management.