

Chapter 6 – How to carry out an evaluation and fitness check

TOOL #45.	What is an evaluation and when it is required	382
TOOL #46.	Designing the evaluation.....	392
TOOL #47.	Evaluation criteria and questions	408
TOOL #48.	Conducting the evaluation.....	420
TOOL #49.	Format of the evaluation report.....	426
TOOL #50.	'Back-to-back' evaluations and impact assessments	444

TOOL #45. WHAT IS AN EVALUATION AND WHEN IT IS REQUIRED

1. INTRODUCTION

Commission evaluations^{618,619} and fitness checks assess the performance of existing policies, programmes and legislation. This tool provides guidance to Commission officials on the application of the 'better regulation' guidelines, the definition of evaluation and fitness checks, the obligations to evaluate and the evaluation planning.

Box 1. What is an evaluation?

Evaluation is an **evidence-based judgement** of the extent to which an existing intervention is:

- effective in fulfilling expectations and meeting its objectives;
- efficient in terms of cost-effectiveness and proportionality of actual costs to benefits;
- relevant to current and emerging needs;
- coherent both internally and externally (with other EU interventions or international agreements); and
- has EU added value i.e. produces results beyond what would have been achieved by Member States acting alone.

Evaluation uses evidence to judge how well the intervention has performed so far compared to earlier expectations prior to implementation or compared to earlier projections made in the context of an impact assessment.

Evaluation goes beyond a factual assessment of *what* has happened it considers *why* something has happened; *how much change* can be attributed to the EU intervention and *to what extent this change meets original expectations/projections*. Evaluation **thus aims (where possible) to draw conclusions about the causal effects of the EU intervention on the actual outcomes/results⁶²⁰. Evaluation should also look at the wider perspective, seeking to identify (and learn from) any unintended or unexpected effects, whether positive or negative, which were caused by the EU intervention but not anticipated, for example in the impact assessment or in the act agreed by the EU co-legislators. Evaluation should provide an evidence-based assessment of whether the EU intervention continues to be justified or where**

⁶¹⁸ Throughout the 'toolbox' and 'guidelines', the term 'evaluation' is used to describe evaluations of programmes – often classified as interim, final, ex-post; evaluations of policies based on legal instruments or soft law – generally classified as interim because few policies have a set end-date; and fitness checks – evaluations of a group of interventions. If special consideration is required, this is mentioned in the text.

⁶¹⁹ EU decentralised agencies (e.g. EFSA, European Food Safety Authority) have their own legal personality are charged with a particular task (e.g. food safety). They are distinct from the Union Institutions (European Parliament, Council, European Commission, etc.) and have their own legal personality. Evaluations conducted by EU decentralised agencies follow the [Evaluation handbook for Agencies](#). EU executive agencies, on the other hand, are established by implementing decision; the European Commission delegates power for the implementation on its behalf and under its responsibility of Union programmes. Evaluations are governed by the Financial Regulation and the 'better regulation' guidelines.

⁶²⁰ As such, evaluation goes further than typical monitoring or audit activities. 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). Performance audit takes a broader look at the overall conduct of the work and its results.

lessons can be learned for improving. As such, evaluation forms the basis for a possible future revision of the intervention which will be further developed in an impact assessment.

Evaluation and fitness check

Evaluations can cover a single EU intervention or a group of interventions which have some relationship to each other (normally a common set of objectives or specific procedures, e.g. reporting) justifying a joint analysis. The latter kind of evaluation is called a **fitness check**. There are no set criteria to identify the scope of a fitness check – rather, the scope should bring together EU interventions whose evaluation as a group will contribute to a better understanding of the role played by the EU in achieving their related objectives and reacting to broader policy concerns. Although fitness checks to date have mainly considered groups of related legislative actions, it is possible to include within the scope of a fitness check any type of EU intervention, e.g. spending programmes, strategies, cross-cutting evaluations of EU decentralised agencies and other EU bodies. It is advisable to discuss the scope of the fitness check early in the preparatory process in an upstream meeting with the Regulatory Scrutiny Board.

A broad fitness check can provide a helicopter view on related policy interventions, assessing in particular coherence and opportunities for streamlining. However, it may prove to be challenging to establish and investigate the intervention logic, to identify good evaluation questions for certain evaluation criteria and find the appropriate point(s) of comparison. All fitness checks should pay particular attention to identifying and quantifying (to the extent possible) synergies (e.g. improved performance, lower costs, reduced burdens, simplification, automation/digitalisation) or inefficiencies (e.g. excessive burdens, overlaps, gaps, inconsistencies, implementation problems or obsolete measures) within the group of measures and help to identify the cumulative impact of the group of interventions, covering both costs and benefits.

Fitness checks can provide a more global picture of the burdens carried by businesses, citizens, and public administrations by looking at the cumulative incurred impacts of several legal acts. As such, they can help identify future simplification opportunities and quantify potential burden reductions, including by increasing digitalisation. When conducting a fitness check in a policy area, it is important to also consider the coherence in roles and tasks of relevant EU agencies and other EU bodies operating in that policy area. As such, a fitness check can help identify synergies including services sharing or reducing bureaucracy and can help derive lessons from the evaluation findings to inform decisions on potential mergers or closures of EU agencies (or other EU bodies) operating in this area.

Evaluation and the 'one in, one out' approach

In the implementation of the **'one in, one out' approach**⁶²¹ regarding the cost of EU legislation, evaluations and fitness checks will be important sources to identify possibilities for simplification and burden reduction without jeopardising net benefits. This REFIT⁶²² aspect should play a prominent role in the evaluation work, as far as possible based on quantification of costs and cost savings potential, applying the EU Standard Cost Model⁶²³ or other cost models. In the context of the 'one in, one out' approach, evaluations will verify initially estimated costs and benefits against actual outcomes, following co-legislators'

⁶²¹ See Tool #59 (*Cost estimates and the 'one in, one out' approach*)

⁶²² See Tool #2 (*The regulatory fitness programme (REFIT) and the Fit for Future Platform*)

⁶²³ See Tool #58 (*The EU Standard Cost Model*)

amendments and national implementation. Fitness checks of entire policy areas (rather than evaluations of specific legislative acts) are particularly useful in this regard. Apart from assessing the extent to which a policy initiative is achieving its objectives, they look at the cumulative impacts of legislation, overlaps and inconsistencies, and so give a more complete picture of the benefits brought to and burdens borne by businesses, individuals, and public administrations. Fitness checks can be tailored to look at specific elements, such as digitalisation opportunities, potential for removing unnecessary or overlapping reporting requirements, for simplifying authorisation systems.

Fitness checks of horizontal issues, such as reporting obligations have already been conducted in a limited number of policy areas (e.g. environment, agriculture, financial services) and they have identified significant potential to simplify the requirements and reduce reporting costs. The [fitness check of reporting and monitoring of EU environment policy](#) is a good example, the [fitness check of EU supervisory reporting requirements in EU financial services legislation](#) is another one.

Objectivity and independence

DGs may commission supporting studies from external contractors to help with certain aspects of the evaluation. The precise content of such outsourced work depends on the needs of each evaluation and the resources available. Studies generally present important information – but may only cover part of the overall picture. A study on its own cannot be considered as an evaluation that meets the Commission's standards. Even where the scope of an external study covers the full scope of an evaluation, it does not represent the evidence-based judgements of the Commission services. An evaluation report in the form of a staff working document is required to bring together all the information and present the position of the lead DG⁶²⁴.

Box 2. Principles of objectivity and independence

An evaluation can be carried out internally by the Commission services or with the help of external contractors via a supporting study. Evaluation work whether undertaken in-house or outsourced to external parties – should respect the principles of objectivity and independence. The analysis in an evaluation can be considered objective and independent if it is based on all relevant information, it is conducted without influence or pressure by third parties and reports transparently on the positive and negative elements of the analysis.

The 'better regulation' function within the lead DG together with the interservice coordination group have a role to play in avoiding bias in evaluation reports. The 'better regulation' function within the lead DG should provide guidance to those conducting the evaluation and together with the interservice coordination group, they should ensure that no important or critical findings have been left out of the evaluation without explanation and that the conclusions of the evaluation are based on an objective reading of the data, clearly indicating the weaknesses of the evidence collected.

⁶²⁴ See Tool #49 (*Format of the evaluation report*)

2. WHAT ARE THE REQUIREMENTS TO EVALUATE?

Evaluation is an essential tool to help manage and inform revisions of existing EU legislation, policies, and programmes. The Commission is committed to regularly evaluate in a proportionate way all EU spending and non-spending activities with an impact on society or the economy. In line with the '**evaluate first**' principle⁶²⁵, evaluation of EU intervention precedes work (i.e. the impact assessment) dealing with the revision of that (or related) intervention. In cases where political urgency may require a Commission proposal at short notice, evaluations may be carried out back-to-back with an impact assessment⁶²⁶.

When planning an evaluation, it is important to make a preliminary assessment of data needs and data availability⁶²⁷ and to consider how long the EU intervention has been operating. This will allow making clear from the start what the analysis will be able to deliver. Consideration also needs to be given to the time required to carry out the actual evaluation work⁶²⁸ with DGs taking a pro-active role in identifying evaluation needs and planning accordingly.

Many evaluations are mandatory by evaluation or assessment clauses in EU legislation⁶²⁹. Evaluations of spending activities need to comply with the requirements of the Financial Regulation. Evaluations may also be necessary due to the application of a sunset clause⁶³⁰.

Review clauses are also common⁶³¹ – requiring work with backward and forward-looking elements. For example, a review clause may require that, by a certain date, the Commission produces an assessment of certain or all elements of an intervention together with, if appropriate, proposals for change. Often such clauses relate to very specific articles, or to the possibility of introducing new elements (e.g. broadening scope of application). Where sufficient evidence to carry out a meaningful evaluation is not available and no revision of the intervention is envisaged, services should consider carrying out a different exercise e.g. an implementation report instead.

For spending programmes, the **Financial Regulation**⁶³² requires an evaluation of all programmes and activities that entail significant overall spending⁶³³. These rules also apply in full to evaluations of EU agencies conducted or contracted out by the Commission. The evaluation requirements for projects or programmes financed by the (11th) European Development Fund (EDF) budget are laid out in [Council Regulation \(EU\) 2018/1877](#).

⁶²⁵ See section II.3 of the [European Commission 2019-2024 Working Methods](#)

⁶²⁶ See Tool #50 ('*Back-to-back*' evaluations and impact assessments) for more information on the specific processes to follow.

⁶²⁷ Section 9 of any associated impact assessment should include indications of monitoring and evaluation arrangements; See Tool #11 (*Format of the impact assessment report*) for further information.

⁶²⁸ See Tool #46 (*Designing the evaluation*)

⁶²⁹ See Tool #44 (*Legal provisions on monitoring and evaluation*)

⁶³⁰ See for instance as regards EU decentralised agencies point 60 of the [Joint Statement and Common Approach \(Parliament, Council & Commission\)](#), 2012

⁶³¹ See for instance Article 29 of the EU Regulation (2019/715) of 18 December 2018 on the framework financial regulation for the bodies set up under the TFEU and Euratom Treaty and referred to in Article 70 of Regulation (EU, Euratom) 2018/1046 of the European Parliament and of the Council.

⁶³² See the EU's [Financial Regulation](#), particularly Article 34.

⁶³³ While the Financial Regulation does not define the level of the 'significant spending', the 'better regulation' guidelines define it indicatively as at least EUR 5 million.

In addition, for spending programmes financed by the EU budget, an obligation to evaluate is 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. Evaluation findings should also be properly reflected in the programme statements accompanying the draft budget proposal each year.

Further sector-specific evaluation requirements are also explicitly included in the EU Treaties in the area of justice, freedom and security; common security and defence policy; research, technological development and space; industry; employment; social policy and public health.

Box 3. Activities which need not necessarily be evaluated in the standard way

It may not be necessary to apply the guidelines fully when evaluating:

- Individual (pilot) projects, groups of projects or sub-activities where their findings will feed into an overarching evaluation. This is particularly relevant for (spending) programmes where there may be many projects or sub-activities that require some degree of assessment that has a narrower scope than evaluation as defined in these guidelines. It is also the case for external programmes where findings coming from evaluations of country programmes, specific delivery methods, tools or elements of certain themes, feed into larger or overarching evaluations including of legal instruments;
- EU executive agencies, where aspects of the agency's performance will be evaluated within the context of the associated programme⁶³⁴;
- EU decentralised agencies, where evaluation of the agency's performance is part of the evaluation of the policy area concerned (fitness check)⁶³⁵;
- A limited set of actions within an EU intervention which are not expected to lead to changes to the wider intervention e.g. a directive which contains a clause requesting the Commission to evaluate, review or assess the definition of X after one year and decide if it is appropriate;
- Performance at an early point in the implementation of an intervention, when information on the longer-term changes (results and impacts) is not yet available (in this case an implementation report is more appropriate);
- The internal administrative policies of the Commission (translation, interpretation, human resources and security, the Publications Office and certain areas of Eurostat).

Such work, which would not generally constitute an evaluation, should nonetheless broadly follow the concepts and principles of evaluation presented here. In the above cases, a more proportionate approach should be applied; in general, a separate Decide planning entry, a 'call for evidence' or evaluation report in the form of a staff working document (SWD) might not be required and/or the consultation strategy and evaluation criteria could cover a lesser scope. **Where a Directorate-General has doubts about the degree of application and the steps that should be followed, they should discuss the approach with the Secretariat-General**, preferably during the annual discussions establishing the evaluation

⁶³⁴ Legal provisions on evaluation of executive agencies are set out in [Council Regulation \(EC\) No 58/2003](#) of 19 December 2002 laying down the statute for executive agencies to be entrusted with certain tasks in the management of Community programmes (Article 25).

In addition, Commission Decision C(2014) 9109 final sets out guidelines on executive agencies

⁶³⁵ Without prejudice to the individual evaluation clauses inserted in most of the agencies' founding acts and Article 29 of Regulation (2019/715).

planning.

Proportionality of evaluation

The 'evaluate first' principle applies to all EU interventions (policies, programmes, and legislation) prior to revising. Nevertheless, the scope and depth of the analysis should be **proportionate** to the importance of the EU intervention. Setting the appropriate depth and scope of the analysis implies deciding on:

- resources and time allocated to the overall evaluation process, including data collection, preparing the 'call for evidence', organising consultation activities, conducting external studies, drafting the evaluation report or fitness check in the form of a SWD;
- the importance and priority given to the intervention (e.g. political guidelines, Commission work programme, operational and strategic decision-making needs, contribution to the sustainable development goals);
- specific factors, such as: the type of intervention e.g. regulation, directive, decision, recommendation, communication; the timespan of implementation and associated data availability; its magnitude and complexity; the significance and nature of the intervention's expected or observed impacts.

It is the **responsibility of the lead DG, in cooperation with the interservice coordination group**⁶³⁶, to determine the level of analysis, considering all relevant factors as well as any constraint in the availability of time, resources and data. Setting the right level of analysis is likely to be an **iterative process**. First indications should be provided in the 'call for evidence'⁶³⁷.

Evaluation analyses and findings should be fully exploited in related impact assessments conform with the 'evaluate first' principle. Furthermore, fitness checks and relevant evaluations should be included in reporting exercises, e.g. in the DG's annual activity reports, strategic and management plans, programme statements, the Annual Management and Performance Report.

3. EVALUATION PLANNING & THE FIVE-YEAR ROLLING EVALUATION PLAN

Good evaluation planning is a critical step to ensure the availability of evaluation results that will feed back into the policymaking cycle. Evaluation activities should be planned in a transparent and consistent way so that the relevant evaluation findings are available in due time for operational and strategic decision-making. They are also an important tool in improving organisational learning, ensuring transparency, and demonstrating the accountability of the Commission.

⁶³⁶ See Tool #48 (*Conducting the evaluation*) describing the role of the interservice coordination group.

⁶³⁷ See Tool #51 (*Consulting stakeholders*)

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

Each Directorate-General should ensure that:

- A periodic evaluation of EU interventions or EU decentralised agencies or other EU bodies under its responsibility is done. Under the '**evaluate first**' principle, evaluation findings and lessons learned should feed into the decision-making cycle, backing up proposals for change and revision. Evaluation findings should, where available, be included in the 'call for evidence' for new initiatives.
- Appropriate monitoring and evaluation activity is planned at the time of adoption of each initiative⁶³⁸ to develop or amend EU action (in general such arrangements are first set out in the relevant impact assessments). It is strongly recommended that the associated work is then entered into the evaluation planning, ensuring that deadlines set in legislation are met.
- Relevant evaluation results are available to feed into the Annual Management and Performance Report⁶³⁹. These results are presented in the Annual Activity Reports drawn up by Directorates-General and in the annual Programme Statements prepared to justify resource allocation in the draft budget, both of which provide a key source of information for the Annual Management and Performance Report.

Regular screening of responsibilities

It is recommended that each Directorate-General screens regularly the legislation falling under its responsibility to identify:

- legal obligations to carry out an evaluation (including under the EU Financial Regulation) or any kind of requirement to review or produce a report⁶⁴⁰, considering whether a full evaluation would be timely and appropriate⁶⁴¹;
- when individual interventions were evaluated last and ensure evaluations are available in a timely manner to feed into the next steps in the policymaking cycle;
- the potential to carry out fitness checks (i.e. cross-cutting evaluations);
- the potential for burden reduction, cost savings and simplification in the context of the REFIT programme and the 'one in, one out' approach⁶⁴²;
- other feedback provided on interventions, including any recommendations from the Fit for Future platform, complaints, infringements or SOLVIT cases⁶⁴³ which might imply problems with implementation, application, or performance.

⁶³⁸ The 2016 Interinstitutional Agreement on Better Law-making envisages a systematic approach to evaluation and monitoring in basic acts: See Tool #44 (*Legal provisions on monitoring and evaluation*).

⁶³⁹ See Financial Regulation, Article 247 Integrated financial and accountability reporting

⁶⁴⁰ For example, report from the Commission to the European Parliament and the Council, implementation report, etc.

⁶⁴¹ The European Parliament produces a [rolling check list of review clauses in EU legislation](#).

⁶⁴² The general idea of the 'one in, one out' approach is to systematically and proactively seek to reduce burdens imposed by existing legislation when introducing new burdens. The role of evaluations will be to verify initially estimated costs and benefits against actual outcomes.

Rolling evaluation plan

The planning of evaluation activities of individual Directorates-General takes the form of a (minimum) five-year indicative rolling programme, where the plan is broadly fixed for the first two years and stays more indicative for later ones, providing an overview of the structure and coverage of the evaluation policy. The plan is submitted through the [interinstitutional studies database](#) and should respect the relevant instructions.

The update of the interinstitutional studies database is under the responsibility of each Directorate-General. In the management plan, the Director-General certifies that the studies database is up to date and includes all evaluations and studies on a 5-year rolling base. But the database is a living tool to follow evaluations or any other study in its life cycle, from the launch to the completion and needs to be kept up to date. It therefore requires regular updating to be effective, in particular regarding the links to deliverables published by the Publications Office of the EU or Decide references of the documents.

All Directorate-General's policy areas, including both spending and non-spending interventions, should over time be proportionally covered in the rolling evaluation plan.

The Secretariat-General compiles the evaluation plans of individual Directorates-General into a single Commission evaluation plan which is published on a central website⁶⁴⁴. This ensures transparency, enhancing stakeholder access by publicising what is being and will be evaluated when and thus further enabling their participation.

Each Directorate-General should indicate a central evaluation contact (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. The evaluation plan will be a key input to the Regulatory Scrutiny Board's decision on which evaluations it will review.

Timing of evaluations in the policy cycle

While the evaluation plan lists all the evaluations scheduled by a Directorate-General during a (minimum) five-year period, the timing of the evaluation (i.e. when it starts and ends) is a key element of the overall plan. Timing is a crucial factor affecting the use of evaluation in the decision-making process; it should enable the results to feed into further preparations for the design, renewal, modification, or suspension of activities. Both planning and timing need to be aligned with political priorities and while there may always be emergencies, good management of these elements helps to ensure that reasonably up-to-date evaluation material is always available for part of the EU acquis.

Evaluation planning should consider the life cycle of the intervention, legal obligations and the operational and strategic decision-making needs of the DG. Often, evaluation planning starts by looking at when the findings are required (e.g. to meet a legal obligation, or to feed into a policy proposal) or when reporting by Member States is due.

Planning for evaluations of *spending initiatives* covered by the multi-annual financial framework (MFF) or the European Development Fund (EDF)⁶⁴⁵ is generally defined in the

⁶⁴³ See www.SOLVIT.eu or contact solvit@ec.europa.eu

⁶⁴⁴ See http://ec.europa.eu/smart-regulation/evaluation/index_en.htm [to upd]

⁶⁴⁵ A different set up applies for the European Structural and Investment Funds.

legal basis, predefining the timing of the different evaluations within the seven-year funding cycle as well as the nature and number of contributions and reports that are expected. To allow the interim evaluation findings to feed into ex-ante evaluations (or impact assessments)⁶⁴⁶ for the subsequent programming period, the interim evaluation results are generally scheduled for delivery around the fourth year of the programming period. Most (ex-post) evaluations are expected to take place within four years of the end of the programming period.

Although the timing of the interim evaluations might potentially be too early to capture the full results and impacts, carrying out the evaluations later would not allow the evaluation findings to be used in impact assessments for the preparation of the subsequent programming period. Equally, there would be insufficient time for any improvements to the current programme to produce significant benefits (although the interim evaluation would allow to correct some implementation issues). To address this, bringing together the final evaluation of the preceding programme and the interim of the ongoing programme, provides a longer data set against which to evaluate performance⁶⁴⁷.

It is much harder to apply a uniform timetable to the evaluation of **non-spending activities** which follow a wide range of formats (e.g. regulation, directive, decision, recommendation, communication, trade agreement). Regulatory activities in particular, have very different policy cycles. For different reasons, Member States can be given relatively long or short periods to complete the transposition of an EU law into national legislation. In other instances, different policy areas and different interventions within one policy area take effect at different moments in time, with a corresponding delay in the arrival of change, etc.

Ideally, when planning for an evaluation of a non-financial intervention, sufficient time should have passed since its implementation to ensure **at least three years' worth of sufficiently full set of data**, meaning that the evaluation cannot be produced before the fourth year following implementation. This makes evaluation planning difficult if a given intervention is revised very frequently, for instance every three years, especially if new objectives are introduced or old ones refined. Even in such cases it is unlikely that the full range of actions are amended each time, so it should still be possible to schedule an evaluation after a given period. However, the scope of the evaluation may need to limit the analysis of certain aspects, reflecting the fact that they have been in place for a shorter duration.

The planning should also consider the need for public/targeted consultations under the 'call for evidence', as well as the duration of relevant administrative procedures, including public procurement of external studies⁶⁴⁸, validation by hierarchy, interservice consultation, preparation of submission/resubmission to the Regulatory Scrutiny Board. If the Regulatory Scrutiny Board decides to scrutinise an evaluation⁶⁴⁹, a minimum of four weeks should be added to the timing of that evaluation.

Resourcing

⁶⁴⁶ See Tool #9 (*Spending programmes, financial instruments, and budgetary guarantee*)

⁶⁴⁷ This works particularly well in cases where there is continuity in objectives and implementation between one funding cycle to the next.

⁶⁴⁸ Detailed guidelines to public procurement can be found on the [Commission's Internal Financial Website](#).

⁶⁴⁹ See Tool #3 (*Role of the Regulatory Scrutiny Board*).

Every operational DG should have a designated evaluation function which can take different shapes (ranging from one full-time equivalent responsible for evaluation to a central unit in the DG and all forms in between). Depending on the organisational model chosen by a Directorate-General, individual evaluations can be run by the evaluation unit, by the operational unit responsible for a particular intervention, or in the form of a joint collaboration between the DG's evaluation function and the operational unit(s). **Whether external contractors are used or not, evaluations require significant input from Commission staff** – both from the lead Directorate-General as from other Directorates-General represented on the interservice coordination group. The decision whether to use external contractors or not depends, amongst others, on the data available as well as resources and relevant expertise in-house.

Typical activities that can be contracted out to external contractors comprise data collection, (including all or some consultation activities), desk research, literature review, modelling, data analysis, synthesizing information, answering the evaluation questions.

The resources, both internal and external, allocated to an evaluation should be proportionate to the perceived importance and expected effects of an intervention, both in terms of its costs and the changes (benefits) it generated.

TOOL #46. DESIGNING THE EVALUATION

1. INTRODUCTION

Designing the evaluation means identifying the nature and sequence of tasks, assessing data and data collection methods and the range of analytical methods to be used to deliver the evaluation. Good design starts by identifying the purpose of an evaluation and identifying what is in scope (e.g. which interventions, which policy objectives and results, over what time period and for what geographical coverage) and what is not. This will influence the evaluation questions set and as a result, the methodology applied, and the data and research undertaken to answer robustly these questions. This will then affect who is assigned which tasks within the evaluation process.

As with any project, good planning and preparation is necessary to produce a high-quality final deliverable. Evaluation requires a critical, evidence-based assessment – using robust and reliable data drawn from a range of sources and analysed in an appropriate manner. Evaluations need to present a clear chain of logic between the data, analysis and conclusions and highlight any particular strengths or weaknesses. This does not happen by accident – but by design.

Other factors which will affect the design include: the political priority of the intervention(s) being evaluated; the timing of the evaluation within the intervention(s)' policy cycle; financial resources and personnel available. It is important to understand how such factors may influence the evaluation process to manage expectations about what the evaluation will realistically be able to deliver⁶⁵⁰.

The 'call for evidence'⁶⁵¹ should provide a first description of the context, purpose and scope of the evaluation and the proposed methodological approach. However, the level of detail provided should simply set the scene. Further detail should be developed during the evaluation process, allowing the appropriate work to be done and enabling the evaluation to meet its purpose.

For fitness checks and evaluations that are selected by the RSB for scrutiny, the Board is available to have an early and informal bilateral discussion with the lead DG(s) based on the 'call for evidence' in an upstream meeting⁶⁵². **The selection of evaluations for scrutiny is communicated to DGs in the second quarter of the year (T) and concern evaluations and fitness checks to be finalised in the next year (T+1).** Therefore, when planning the work, it is advisable to assume that the evaluation may be selected for scrutiny as this adds at least four weeks to the process. Since the Board scrutinises all fitness checks, their planning should factor in this process upfront⁶⁵³.

2. DESIGNING THE EVALUATION

When designing an evaluation, it is important to:

- (1) Clarify the **purpose** of the evaluation: deciding and clearly describing what the evaluation will deliver and how its findings will be used.

⁶⁵⁰ See also Tool #44 (*Legal provisions on monitoring and evaluation*)

⁶⁵¹ See Tool #51 (*Consulting stakeholders*)

⁶⁵² See Tool #3 (*Role of the Regulatory Scrutiny Board*)

⁶⁵³ Ibid

- (2) Define the **scope**: setting out clearly what will be evaluated. This may be in terms of interventions, measures or legal articles, time period, geographical coverage, particular effects or any other relevant aspect. The reader should understand what will be covered by the evaluation and what will not, and the reason why.
- (3) Explain the **intervention logic**: summarising how the intervention was expected to work (i.e. at the time of adoption by the Commission or later by the co-legislators, or at the time of implementation), including the underlying assumptions. The intervention logic can draw from any prior impact assessment or other documentation such as the explanatory memorandum, which justified the initial policy action.
- (4) Draft good **evaluation questions**: they should address the five evaluation criteria and any other aspect as relevant, also considering the feedback on the 'call for evidence' to the extent possible. Questions should cover all issues that are known to be of interest to the stakeholders.
- (5) Identify the **appropriate point(s) of comparison**: the different evaluation questions should be answered against relevant starting points (e.g. benchmarks) or expected achievements for example, comparisons may be drawn against the changes projected or expected under the preferred policy option (as in the impact assessment or changes since measurements made at the start of an intervention. Before doing such analysis, it may be necessary to update the original points of comparison to take on-board changes introduced by the co-legislators to the Commission's proposal.
- (6) Consider **appropriate data collection and analytical methods**: it is important that the evaluation is set up to collect and analyse a range of different data, using the appropriate data and methodologies to fill existing data gaps and to answer robustly the evaluation questions.

It is essential to clearly present the key design elements of the planned evaluation in the 'call for evidence' as this is the first public communication with interested stakeholders who may provide input to the evaluation work or wish to use the evaluation findings. Stakeholders may also provide feedback on the 'call for evidence' which could affect, in some instances, the evaluation design⁶⁵⁴.

2.1. Purpose of the evaluation

It is important to be clear from the start and state in the 'call for evidence' the purpose of an evaluation – to explain why the evaluation is being carried out, the sort of findings it is expected to provide and how these findings may be used.

Evaluations may be launched for specific reasons, such as:

- to ensure compliance with an article in the legal base or in the Treaties. This applies to many policies and is for instance mandatory for programmes carried out under the Multi-annual Financial Framework (MFF);
- to ensure compliance with the EU Financial Regulation;

⁶⁵⁴ Issues relating to the purpose or scope of an evaluation (for example concerns raised by stakeholders, input from other DGs or the RSB, lack of data, timetable) must be addressed by the lead DG early on, to prevent miscommunication and to ensure appropriate evaluation design. In the extreme case, it may be necessary to consider delaying the evaluation or agreeing that a different kind of analysis is acceptable.

- to identify efficiency gains, synergies and opportunities for more policy coherence; to identify potential for simplification and burden reduction in a specific area (fitness checks);
- to assess the role of existing EU interventions before bringing forward possible changes ('evaluate first') and to identify the problems that revisions will tackle;
- to identify opportunities for more streamlined/enhanced implementation; and
- to keep the EU law fit for purpose particularly when there are signs that the legislation is not performing as it should (e.g. high number of infringements or complaints).

All evaluations are required to explore the potential for simplification and burden reduction.

2.2. Scope

The scope of an evaluation also has a significant impact on the design and subsequent conduct of the evaluation. It is important to know from an early stage what will be in the scope of the evaluation and what will not. Any limitations to the scope identified upfront must be clearly justified in the 'call for evidence'.

Key factors affecting the scope relate to the coverage of:

- **an individual intervention or group of interventions:** the scope should be defined in terms of the EU intervention – the legislative act(s) and/or related programme(s), associated implementing and delegated acts, policy(ies), soft law (action plans, strategies, communications etc.) covered. Usually, all aspects of a given piece of legislation or an intervention will fall within the scope of an evaluation. There may be circumstances when particular aspects are excluded, e.g. due to lack of experience, something not having been tested perhaps due to a later than expected implementation date or because a given article is conditional on other activities that are not yet complete. Any significant constraints to the scope should be explained in the 'call for evidence' and reflected in the design. In the most extreme case, factors affecting scope may raise questions about whether the evaluation should be delayed or a different analysis undertaken.
- **geography:** all Member States, a subset, the wider EEA, third countries (incl. developing countries) etc.
- **period of time:** this could be the full period since the intervention was implemented or a shorter period depending on the purpose/timing of the evaluation such as from the end of the period covered by the last evaluation to the current day.
- **particular effects anticipated:** this may be affected by the time period covered as some effects may not have been expected to materialise at the time of the evaluation. It could also reflect the significance of the different effects expected and availability of data or the particular political and/or policy context.

The scope may be also influenced by any feedback on the (perceived) performance of EU action e.g. from stakeholders or monitoring. This may lead to the inclusion of action(s) where there is evidence of synergies (e.g. interventions working together to complement each other

or where data is re-used, simplifying demands on stakeholders) or of problems (e.g. targets not being met, low transposition or compliance, complaints). However, it may also be interesting to include in the scope action(s) where there is a theoretical or expected link but no apparent evidence indicating problems or successes. Just because 'all is quiet' does not necessarily mean 'all is working as expected'; it may, for example, in the worst case mean that the intervention is irrelevant or that there are significant loopholes.

When considering the scope of an evaluation, due attention should be paid to both the political and policy context of the intervention(s):

- Political aspects relating to context may reflect, for example, a particular emergency or a 'hot topic', which could significantly reduce the time available to conduct the evaluation, the depth or scope of analysis.
- Policy considerations may refer to the intervention being evaluated within the logic of the wider relevant policy framework or any relationship to other actions. Sometimes related actions will be considered under coherence – for example, an evaluation of EU environmental action relating to water may consider coherence with another EU environmental action relating to waste, or with a particular EU business related action. In other cases, the interactions and their coherence may be so important and integral to the policy that a fitness check is needed. For example, the Fitness Check of EU Freshwater Policy⁶⁵⁵ looked at six EU environmental actions in the area of fresh water, with further consideration under coherence of wider environmental issues such as waste and emissions controls, but also the common agricultural policy and regional policy.

There is a clear link between the purpose of an evaluation and its scope. If the scope of an evaluation is limited in some way, then it might not be able to fulfil its purpose and meet the expectations of decision makers and stakeholders.

Box 1. Considerations affecting the decision to do a fitness check

- There are no set criteria for defining the scope of a fitness check – consideration should be given to a range of factors including its purpose, context, timing, maturity of actions being considered for inclusion, data availability and resources.
- It is critical to define an acceptable/appropriate scope early. The EU interventions, which fall within the scope of a fitness check should interact together or have common objectives. Where this is not the case, it is likely that the work undertaken will deliver several evaluations of the separate interventions, rather than the desired synergy of a joined-up evaluation of the EU policy and its framework.
- When considering which EU actions to include in scope, careful consideration needs to be given to:
 - the interaction between purpose and scope. Increasing the scope by adding more EU actions might increase the time and resources required, but could deliver synergies, providing more useful findings and a more comprehensive overview of the EU's role in that policy area. However, expanding the scope and including more loosely connected actions may not provide much additional or important information. For

⁶⁵⁵ https://ec.europa.eu/environment/water/blueprint/fitness_en.htm

example, in the chemicals area, there is certain key central EU legislation (REACH, CLP regulation) relating to chemicals, but EU actions on chemicals also includes legislation on pesticides, medical devices, toys' safety and many others, which may or may not need to be included in the scope. Often, looking at EU framework directives and related legislation together delivers a good picture about the legislative framework in place.

- how long the various EU actions have existed and hence operated or not operated together.
 - the depth of analysis required – whether the interventions need to be assessed together (a fitness check) against all (five) criteria, or whether the focus would be on the criterion of coherence.
- Sectoral fitness checks look at how EU interventions affect a specific sector. They look at the same five criteria, but from the sectoral perspective. When designing a sectoral fitness check, it is crucial to consider the use of its possible results (conclusions and lessons learned) beforehand.

Many of the legal acts assessed in a sectoral fitness check apply to a wide range of sectors, not just the specific sector identified for a given sectoral fitness check. It is beyond the scope of a sectoral fitness check to perform a full evaluation of those acts. However, some qualitative (and to the extent possible, quantitative) data of the wider performance of those acts should generally be considered, to provide the appropriate context for the sectoral fitness check. For example, often a certain sector bears a higher proportion of the overall costs, whilst the benefits are to the wider society.

Sectoral fitness checks are likely to require a higher level of coordination across the different Commission services due to the varying legislation in the scope.

- In some cases, fitness checks can be targeted at specific aspects of the legislation in question, for example: reporting requirements, supervisory activities, processes. These fitness checks have a more prominent role in burden reduction and simplification.

2.3. The intervention logic

The intervention logic provides a (narrative) description and usually a diagram summarising how the intervention was expected to work. Put another way, it describes the expected logic of the intervention or chain of events that should lead to the intended change. An intervention is expected to be a solution to a problem or need – the intervention logic is a tool which helps to explain (and often visualise) the different steps and actors involved in the intervention, and their dependencies – thus presenting the expected 'cause and effect' relationships. The intervention logic is useful both as:

- a communication tool – facilitating discussion of the intervention with different parties, helping to identify differences in understanding or to clarify particular details;
- an analytical tool – identifying relationships and dependencies that were expected, based on certain assumptions made in the impact assessment (where one exists). For example, were certain activities expected to occur in parallel or sequentially? Were all activities expected to generate outputs or just some? Who was expected to do something? Was the expected 'output' from one person or entity an 'input' for someone else? A well-constructed intervention logic will help to identify the relevant evaluation questions.

Constructing the intervention logic means considering how different actors were expected to react, what actions were expected to be triggered by the EU intervention, how both actors and actions were expected to interact to deliver the promised changes over time and ultimately achieve the objectives of the EU intervention being evaluated. Given that an evaluation delivers a judgement on why and how the EU intervention has actually worked, compared to what was expected (i.e. at the time of adoption by the Commission or later by the co-legislators, or at the time of implementation), it is important to have this understanding from the start.

In practice, there is a link between the intervention logic of an evaluation and the intervention logic/problem tree of any preceding impact assessment. In the impact assessment, the problem tree typically describes (a) the problem and its drivers, (b) the general and specific objectives, and (c) all potential solutions (options) and how they are supposed to work. In the evaluation the intervention logic typically describes (a) the need for the intervention and (b) the expected outputs, results, and impacts of the intervention and (c) how they are supposed to be achieved. The intervention logic also considers external factors, which may influence both the expected performance of the EU intervention or generate some type of effects.

There are many possible formats or approaches to describe the logic of the intervention⁶⁵⁶ and the 'better regulation' guidelines do not mandate a particular approach.

In practice, a traditional starting point to describe the logic of the intervention is to consider the categories presented in Figure 1 below (i.e. needs, objectives, inputs, activities, expected outputs, results, impacts, external factors, other EU policies) and assess those using the five evaluation criteria⁶⁵⁷.

Whatever approach is used to describe the logic of the intervention, thought should be given to the following elements being reflected:

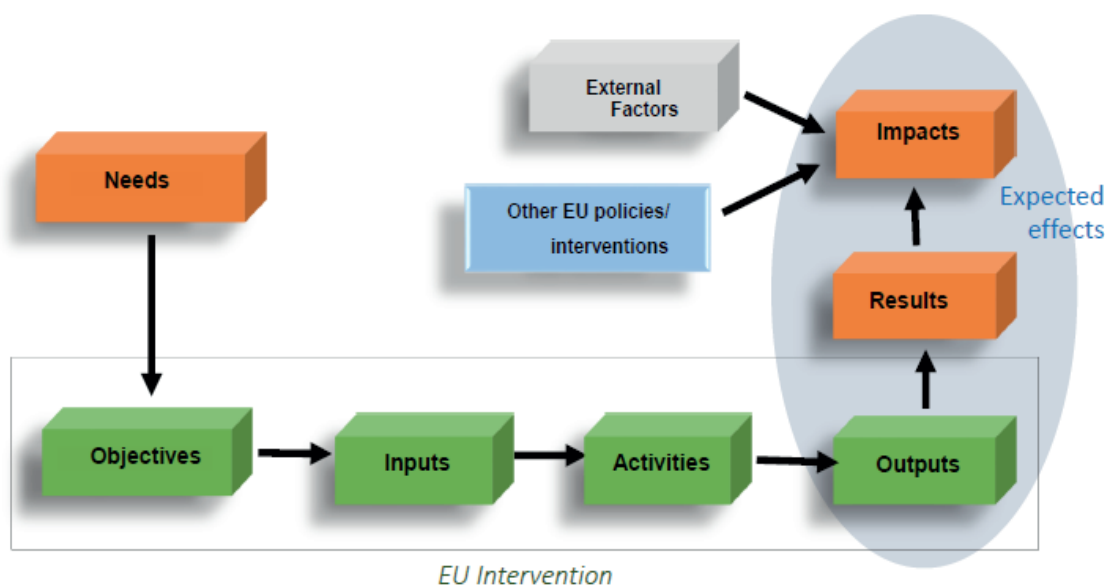
- What was the rationale for the intervention? What were the '**problems**' or '**needs**' that the intervention was meant to solve? → **Tip:** Look at relevant background documents to understand the context – these may include a previous impact assessment, associated studies, legal text (particularly the recitals) and associated explanatory memorandum. These may also be helpful in identifying key deadlines, milestones, and deliverables.
- How were the **objectives** expected to be achieved? What was the 'positive desired situation'? What were the expected changes that the EU wanted to achieve?
- How were these changes to be achieved? What **inputs** were expected to be used? → **Tip:** Inputs can be a very encompassing term, covering for example resources such as staff, time, and equipment as well as the legal act. Which events (**activities**) were expected to happen?
- What was expected to be achieved in the short, medium, and long term (**expected outputs, results, impacts**)?

⁶⁵⁶ In evaluations, the concept of intervention logic combines elements from the Logframe approach and the Theory of Change. Both approaches are used to describe causal pathways in interventions and the mechanisms that enable them. The differences between the Logframe approach and the Theory of Change are not clear-cut and they are used in various definitions. In application, their meaning often overlaps.

⁶⁵⁷ See Tool #47 (*Evaluation criteria and questions*)

- Can any **external factors** be identified which may have influenced the performance of the EU intervention, or generated the same type of effects? Reality is complex and many other players and factors can intervene and influence a situation.

Figure 1: Simplified intervention logic

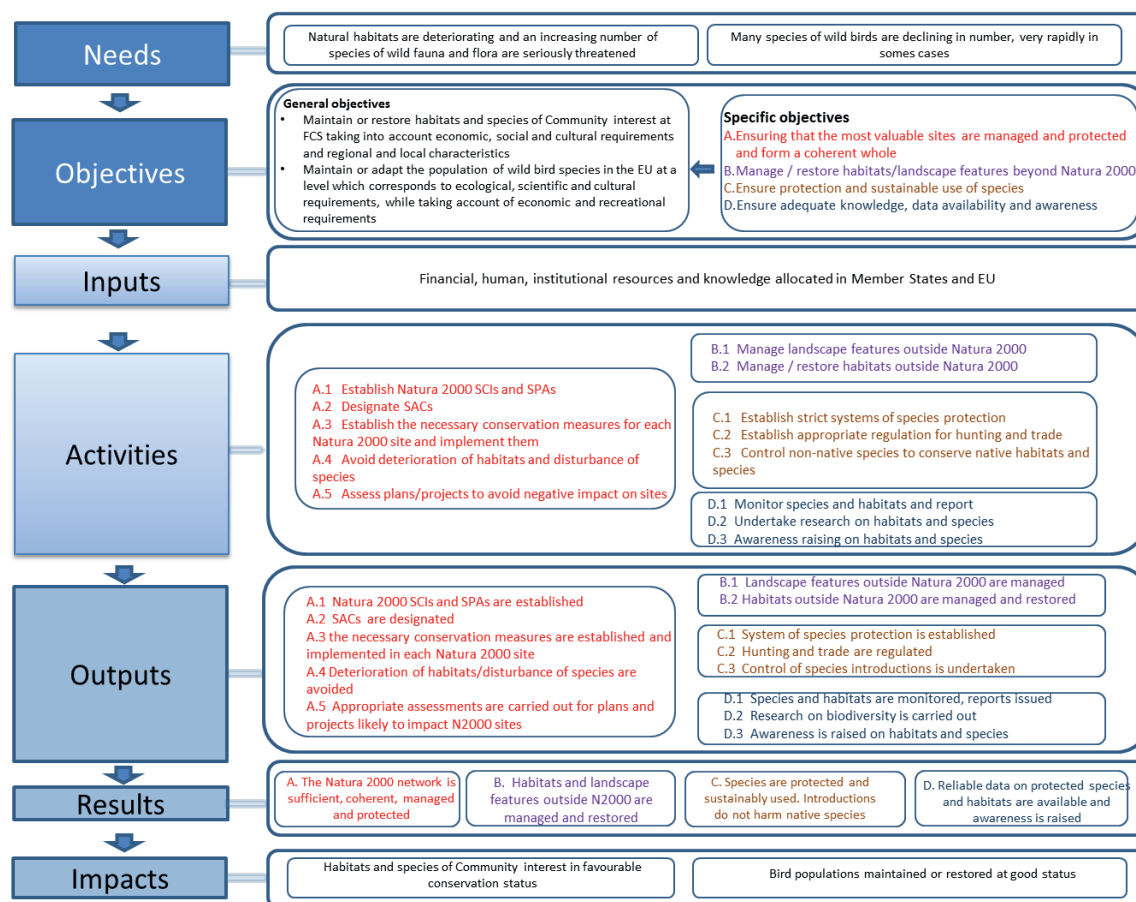


Moreover, the intervention logic (narrative part) could further describe:

- Who was expected to be involved? How were these entities expected to be involved – were they going to be affected by the intervention (positively or negatively)? Were they responsible for taking a particular action? Activities are often more tangible or visible.
- When was something supposed to happen? What was the expected order of activities or events? What was expected to happen at the same time (in parallel)? What was supposed to happen before or after something else (sequentially)? What changes were expected to be positive (benefits / cost savings) or negative (costs) and for whom? This consideration of changes over time (outputs/results/impacts) may then be linked to the hierarchy of objectives in the impact assessment.
- What obligations were set or what provisions were expected to be put in place? For the evaluation of legislative actions, many of the required actions are identified in the articles of the legal act. These physical ‘inputs’ are often translated into monetary values, leading to a broader consideration of what has been needed to achieve objectives and possibly to considerations of costs and benefits. For example, costs and benefits related to changes in employment practices made by an employer to comply with an EU law, or costs incurred by a Member State due to new reporting obligations may start with a consideration of training needs, time taken to train, new systems bought, etc.

A typical example of an intervention logic included in a Commission evaluation staff working document is presented below.

Intervention logic for the Birds and Habitats directives:



It is important to check the draft intervention logic with the interservice coordination group, to see whether it 'flows' and is truly 'logical', describing the appropriate causal pathways. It is also important to check whether appropriate attention has been given to all the elements which influenced the design of the intervention, including changes made during the adoption process.

In terms of the process, a first (rough) draft intervention logic (design phase) should be developed by the lead DG and discussed with the interservice coordination group ideally **at its first meeting**. It can also be useful to test it out both with colleagues who understand the intervention, and with people who have little familiarity with the subject.

It is recommended that a first (rough) draft intervention logic is included in the terms of reference for external work, and contractors are asked to either (a) demonstrate their understanding and evaluation skills by providing a more elaborated version, and/or (b) develop and test with various stakeholders the shared understanding of the draft intervention logic, and/or (c) develop the final intervention logic which provides, among other things, for the actual effects (i.e. actual outputs, results, impacts).

A good understanding on how the intervention was expected to work, including the underlying assumptions is essential for identifying good evaluation questions.

2.4. Drafting good evaluation questions

Establishing the intervention logic is useful in identifying specific and robust evaluation questions linked to the initial expectations of the policy intervention. Robust evaluation questions encourage critical analysis. By defining and sharing the questions early in the process, the Commission services clarify what they intend to analyse and invite interested parties to provide relevant contributions. These questions will then influence the design and conduct of the subsequent steps, and their answers will feed the narrative of the evaluation report.

Questions should be worded in a way that forces the evaluator to provide a complete, evidence-based answer that improves understanding of the performance of the EU intervention against the five evaluation criteria. Answers should go beyond providing a yes/no answer based on simple description and to look at what the links were between the changes observed and the original intention of the EU intervention(s). This tends to mean that evaluation questions are 'causal' or 'normative', but they also need to be 'critical'. Moreover, evaluation questions should draw from the intervention logic, assessing for example whether the expected outputs, results and/or impacts were achieved effectively and efficiently, or whether needs and objectives are still relevant, etc⁶⁵⁸.

Broad, generic questions based on the criteria themselves, such as *How effective has the EU action been?*, should be approached with a number of more detailed, specific questions e.g. *What factors linked to the Directive have most influenced progress?*, *To what extent have the arrangements set out in Article 7 been able to influence progress towards the objectives of the Directive?* Although general questions are important as they allow the analysis to follow the collected evidence, which can often unearth unexpected or unintended changes, this is only true if these general questions are made more specific.

Specific questions allow issues raised during adoption or implementation to be further investigated, e.g. in relation to detailed arrangements that caused strong debate and a compromise solution to be adopted, or based on feedback from stakeholders that a given article is problematic or a procedure is difficult. It is important that such questions reflect the level of progress and performance expected at the time of the evaluation. For example, if at the time of the evaluation no businesses have completed a certification process specified in the respective EU legislation, there should be no evaluation questions about how well the specified process has worked⁶⁵⁹; instead, if not obvious, the evaluation question should be on the cause of this outcome.

When considering the questions to ask it is also important to think about the usefulness of an answer, and the feasibility of obtaining an answer. This again links back to considerations of scope and purpose. The availability of data to provide evidence to answer the question is also a relevant consideration.

2.5. Point(s) of comparison

The evaluation aims to capture the change that an intervention has brought over time. To do this, it needs to compare actual performance against one or more points of comparison.

⁶⁵⁸ See Tool #47 (*Evaluation criteria and questions*), Figure 1 on a simplified view of the intervention logic and the 5 evaluation criteria

⁶⁵⁹ See Tool #47 (*Evaluation criteria and questions*).

Where there is a prior impact assessment, the expected results and impacts of the preferred policy option should be the preferred point(s) of comparison. To the extent possible and proportionate, these may need to be adapted to create an evaluation point of comparison, **reflecting any relevant changes since the IA was published** (such as significant changes during the legislative process). Ideally, such point(s) of comparison are already quantified in the impact assessment⁶⁶⁰ or can be quantified (e.g. expected level over time of pollution or emission of certain substances; level of employment). As a minimum, a qualitative description of what was expected to happen should be given (e.g. description of expected trends in nature, levels of education in society etc).

Where no prior impact assessment exists, the evaluation will usually have to build its own point(s) of comparison for example based on existing data measuring the situation at or around the time the EU action started. Where there is no existing data measuring the situation at or around the time the EU action started, this may have to be a qualitative description.

In other instances, there may be one or more figures which do provide some information relevant to the situation (e.g. level of a certain chemical before EU intervention; level of employment). A reasonable starting point for constructing a point of comparison for the evaluation would then be to take this information and reflect on how it might have continued over the period being evaluated, had there been no policy change. For example, is it realistic to assume a constant level of performance over this period (number does not change)? Or was some level of increase or decrease more likely? How big an increase or decrease? This might allow a reasonable point of comparison to be created. In doing this, it will be important to state clearly the assumptions being made (e.g. constant performance, decline of 10% a year in line with global trend, etc).

Other useful points of comparison that could be investigated for relevance and applicability include:

- Another scenario or benchmark, such as the situation prior to implementation, an earlier programme, or systems in countries outside EU.
- Different performance ('policy-on') or counterfactual ('policy-off') scenarios may be used, based on clearly stated assumptions. For instance, a common counterfactual scenario would be to consider the situation where the Union does not act⁶⁶¹.
- Other benchmarks may also be identified or exist for the policy context. They may be used for comparison, to provide additional context or as an aid to interpreting the change identified. For example, there may be benchmarks for certain types of cost, durations of processes which can be drawn from industry or from other countries. Often this type of benchmarks are concrete, actual figures or levels of performance which have been achieved – rather than aspirations or predictions.

⁶⁶⁰ The predictions from the impact assessment are generally the estimated costs and benefits. This will provide an insight into how actual developments compare to what was expected at the time. The original impact assessment predictions may need to be updated to address changes introduced during the discussions with the co-legislators.

⁶⁶¹ As EU policies operate in a complex environment influenced by a wide range of factors falling outside the scope of the EU intervention, it requires effort and resources to identify a robust counterfactual scenario. For more information on counterfactual scenarios see Chapter VIII – Tool #68 (*Methods for evaluating causal effects*)

Evaluations may have several points of comparison. This depends on the evaluation criterion to be assessed. For example:

- When assessing relevance, the suitable points of comparison are the original needs and objectives behind the EU intervention and new needs arising from changing policy objectives or upcoming technological, social, environmental, or economic changes and the use of strategic foresight information.
- When assessing coherence, suitable points of comparison could include looking at the changes in coherence between the start and end of the period being evaluated or, if applicable, information from the impact assessment (i.e. assessment of coherence of preferred policy option). They could also include looking at existing or new national or international actions.
- When assessing the EU added value, the comparison involves consideration of performance against a projection of how the situation was expected to evolve without the EU intervention (a defined counterfactual, or some estimate of the cost of the Union not acting – ‘the cost of non-Europe’). Often such analysis assesses whether the subsidiarity arguments put forward before the intervention (as presented in a prior impact assessment, or other accompanying documents such as the explanatory memorandum) were valid and whether the expected changes resulting from EU action were delivered.
- When assessing effectiveness and efficiency, suitable points of comparison could include looking at what was expected to have happened at this point, including comparison with the stated specific and operational objectives (and targets if applicable) as in the impact assessment, explanatory memorandum etc; or another scenario or benchmark, such as the situation prior to implementation, an earlier intervention, groups that did not participate or benefit from the intervention, or similar systems in countries outside the EU.

For consistency purposes, **the source for identifying points of comparison** should ideally stay the same for addressing all evaluation criteria (e.g. the prior impact assessment, the situation at or around the time the EU action started, etc).

When deciding on the appropriate point(s) of comparison possible, consideration should be given to:

- **availability of data:** Does the impact assessment contain information reflecting the adopted intervention? Is data available to judge if these projections need to/can be updated? Are there other ‘benchmarks’ from EU/non-EU countries or organisations that can be used to assess performance? Could the UN sustainable development goals and targets be used as benchmarks? Are other relevant data available that can describe the situation before the intervention? Is it feasible to collect them? Has monitoring been conducted since the intervention started (allowing a robust assessment of actual change over time)?
- **relevance of the points of comparison:** different sources for extracting comparison points may be more useful, depending on the different evaluation criteria and/or the exact evaluation question posed. However, different sources for extracting comparison points may be very difficult to manage and can create some arbitrary conclusions.

After identifying the appropriate point(s) of comparison, it is necessary to consider what type of analysis is possible. At the very least, comparison should enable the evaluator to judge whether change has occurred and to decide whether it is a change in the desired direction (e.g. increase or decrease) and to what extent as a consequence of the intervention. When causal analysis⁶⁶² is not possible or only at disproportionate cost in terms of data collection and resources, EU evaluations have to rely on qualitative, reasoned arguments (backed by the appropriate quantitative and qualitative evidence) about the likely role/contribution of an EU intervention to the changes observed.

There is growing expectation to find ways to reliably quantify the impacts of EU interventions and efforts should be made to aim for high quality causal evidence⁶⁶³.

2.6. Methodologies and data

The methodology of an evaluation should consist of a combination of tools and techniques assembled to provide answers to its evaluation questions.

Data availability and quality of data will play a key role in deciding which analytical methods can reliably be applied. It will also influence consideration on who will be assigned which tasks – what will be done by Commission staff and what will be contracted out.

The time and budget allocated to the evaluation will have a significant influence on both the methods chosen and the data collected. The lead DG may also wish to contact the JRC who can advise on data sources and methods that might be used for the evaluation. In particular, the JRC is developing knowledge-management services bringing together data collected during earlier evaluations, impact assessments and studies⁶⁶⁴. It can also help in reviewing existing evidence in international specialised literature.

It is also important that evaluations clearly state the challenges that have been encountered and resulting limitations in the certainty or accuracy of the findings, particularly as these vary greatly from case to case.

When considering human resources, it should be kept in mind that desk officers in the Commission involved in evaluation are not expected to become experts in the many tools and techniques available for use in evaluation. Such expertise is generally available in the DG's evaluation / 'better regulation' function, economic analysis function, from the JRC and/or from external contractors. However, it is still necessary to have a general understanding of various approaches to data collection and analytical methods, in order to:

- ensure the focus is only on relevant data, which helps to answer the evaluation questions;
- recognise the strengths and limitations of the methods proposed;

⁶⁶² Causality, in the context of evaluation, checks the plausibility of the expected chain of events whereby the EU intervention was expected to alter behaviours and create the expected changes (e.g. as identified in a preceding impact assessment) or any other unintended or unexpected changes. It seeks to establish a relationship between an intervention and the observed changes in the issues which the intervention addressed. See also Tool #68 (*Methods for evaluating causal effects*)

⁶⁶³ See also Tool #4 (*Evidence-informed policymaking*)

⁶⁶⁴ https://knowledge4policy.ec.europa.eu/microeconomic-evaluation_en

- understand which methods can be combined, to ensure that the data is based on several sources of information and that the analysis is drawn from several perspectives (triangulation);
- estimate whether a contractor's offer to use a certain set of methodologies in an evaluation is realistic in view of the reliable data, time, and budget available.

When performing an evaluation not supported by any external (contracted out) work, it is equally important to have a reasonable overview of such aspects, before acquiring a deeper understanding of the methodologies selected for use.

It is important to consider ways to design the evaluation **so that it is possible to triangulate data and modelling results from different sources**. When thinking about the data needed for an evaluation, **it is necessary to look first at what is already available** – for example, from existing monitoring and reporting arrangements or studies which have been conducted, **including any prior impact assessment**, explanatory memorandums or staff working documents accompanying the policy measure under evaluation. A key source of information, which is often overlooked, is to **check what evidence-based complaints (if any) have been received from stakeholders and interested parties**. For evaluations involving legislative interventions, it is also important to **check whether any infringements have been detected** and to investigate the underlying reasons for the alleged violation. Both these sources of information can provide significant insight into the actual and perceived performance of the intervention being evaluated. Equally, in policy areas related to the four freedoms of the single market, the SOLVIT database might provide additional information.

Once this stock-take of existing data has been carried out, it will be easier to identify new data which will need to be collected and to consider the mix of objective (quantitative) and subjective (qualitative) data. Reduction to either only objective or only subjective data for the whole evaluation work is not optimal. Objective data often comes from statistical reports, monitoring, or modelling (although the degree of objectivity will be affected by the assumptions underpinning the model), while subjective data is generally opinion based. All evaluations require a consultation strategy presenting consultation scope and objectives, identification of stakeholders, envisaged consultation activities, their timing and language regime⁶⁶⁵.

Efforts should be made to improve the degree and quality of quantification and modelling provided in evaluations and fitness checks and to put a number (or a range) on the costs, benefits and value delivered by the EU intervention⁶⁶⁶.

Whether evidence is quantitative or qualitative, every evaluation should include an assessment of limitations and uncertainties of the evidence (e.g. due to poor data availability, assumptions or modelling issues), as well as mitigation measures used to remedy such limitations. Building on this assessment, the evaluation should also conclude on the robustness of findings and conclusions made.

The information on data collection and analytical approach is brought together in an 'evaluation matrix'. This identifies for each evaluation question per criterion:

⁶⁶⁵ See Chapter VII on Stakeholder consultation

⁶⁶⁶ See Tool #57 (*Methods to assess costs and benefits*)

- the 'success' or 'judgement' criteria (i.e. points of comparison) against which the answer to the question can be assessed;
- associated indicators and/or targets, which may be both quantitative and qualitative;
- data necessary to fill the indicators and provide information for the success/judgement criteria – this should consider not just the type of data, but which stakeholders it involves, whether it is already available or needs to be collected (and how this might happen), time required / feasibility to collect. Where questions have the same sources of data, this may suggest a way to cluster them; where questions require specialised data, the resource implications should be considered;
- analytical methods for turning data into necessary indicators or information – again considering how this can be done, who will do it, time required / feasibility to do etc. This may then lead to consideration of data collection and analysis methods (and can also be used to inform decisions on the work to be contracted out and its possible value).

Example of an evaluation matrix:

To what extent do the provisions of Directive 92/83/EEC ensure proper functioning of the internal market?

Question	Sub-question	Judgement criteria	Indicator	Data sources
1.1 To what extent does the Directive ensure legal certainty and clarity with regards to the classification of alcohol and alcoholic beverages for excise purposes?	1.1a Which products are difficult to classify (e.g. because they could, arguably fall within several excise categories)?	The applicable rules result in difficult and/or unclear classifications of alcoholic beverages	Classification of alcoholic products that do not fall into one clear category, such as mixtures of fermented beverages and spirits, alcopops containing cleaned-up alcohol, cream liqueurs, mead, by the Member States. Other reported instances of alcoholic products whose excise classification was difficult	Survey to national tax authorities Survey to economic operators Reported statistics Studies /papers
	1.1b Do the ambiguities post Siebrand (C-150/08) still cause problems in this area?	Degree to which classification of concerned products follows the criteria laid down in the judgement	Classification of products containing a mixture of fermented and distilled alcohol products by the Member States The interpretation of the 'essential character' or particular products	Survey to national tax authorities Survey to economic operators Reports and studies

Chapter 8 provides information on available methodologies for evaluation. The European Commission's Competence Centre on Microeconomic Evaluation (JRC) can provide advice on appropriate evaluation methods and data, where needed.

2.7. The evaluation design checklist

The following step-by-step approach can be used to assist the lead DG and the interservice coordination group in designing the key elements of the evaluation / fitness check and monitor the level of completion for each step. For evaluations and fitness checks selected for scrutiny, the checklist may also be used to structure the discussion with the Regulatory Scrutiny Board in the upstream meeting⁶⁶⁷.

⁶⁶⁷ See Tool #3 (Role of the Regulatory Scrutiny Board)

Evaluation design – a step by step approach (‘evaluation design checklist’)



TOOL #47. EVALUATION CRITERIA AND QUESTIONS

1. INTRODUCTION

All evaluations and fitness checks should base their analysis on the evaluation criteria of **effectiveness, efficiency, coherence, relevance, and EU added value of the intervention**, or provide due justification why this is not the case⁶⁶⁸. Additional criteria beyond these five can be added, if at all necessary. This also needs to be duly justified.

The evaluation criteria should structure the analysis and ensure it is comprehensive. The evaluation needs to report on them⁶⁶⁹, but the purpose of the analysis should be to draw operational conclusions and lessons learned that can feed into future initiatives.

Evaluations and fitness checks should always assess the economic, social, and environmental impacts of EU interventions (expected or unexpected) with, where relevant, particular emphasis on those impacts identified in a previous impact assessment.

The degree of analysis conducted for each criterion should be proportionate: it depends on the **intervention** being evaluated, the **timing** of the evaluation and the adequacy and reliability of the **data**. Often this will mean that for some criteria new data will need to be collected, analysed, and compared with other findings; while for others, a short summary can be presented based on existing reports and information.

For example, at an 'early' stage in the intervention's lifecycle, it may not be necessary to judge the relevance criterion in any depth. If it is only a short time since the intervention has been implemented, it may be fair to assume the continued relevance of the action and hence simply restating previous arguments should suffice; alternatively, stakeholder feedback may be the only indicator of whether needs have changed and some summary presentation of their (unchanged) opinions may also be sufficient. Equally, EU added value may be difficult to judge in the early years, particularly if the intervention concerns setting up new EU decentralised agencies or other EU bodies or putting in place a framework. In these cases, confirming the validity of the (theoretical) EU added value may be as much as is reasonably possible at that time.

Establishing the intervention logic is usually helpful in identifying specific evaluation questions, based on the evaluation criteria. As mentioned in the tool on designing the evaluation (Tool #46), there is a need to ensure that all evaluation questions focus on providing useful information such as information on the changes the intervention sought to achieve, investigating particular intervention characteristics or factors, which have (not) worked. It is also advisable not to have too long a list of evaluation questions at the start of an evaluation as it may be too constricting and prevent the analysis from 'going where the data leads'. Whilst evaluation sub-questions can be developed early in the evaluation process, e.g. to help define a particular question, or drill down on specific areas, this may also happen at a later stage in response to evidence collected. In this regard, it is important to give the

⁶⁶⁸ The evaluation of a single intervention may on an exceptional basis omit one or two of the five evaluation criteria. An exception must be granted, and clear justification for such omission must then be provided in the evaluation roadmap and repeated in the final evaluation report. Fitness checks always consider the five criteria.

⁶⁶⁹ [Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-making](#)

evaluator the flexibility to develop additional questions or sub-questions later in the process as they see fit.

Depending on the specific evaluation there may be overlaps between the criteria – for example, for relevance and EU added value. **It is generally more important to identify a good set of evaluation questions, than to be too concerned about how to categorise the question as long as the different criteria are addressed.**

It is advised to discuss the evaluation questions with the interservice coordination group. The evaluation questions can either be set at the time that the 'call for evidence' is published or discussed at the first interservice coordination group meeting and defined shortly thereafter. These questions will then influence further the design and conduct of the subsequent steps; they should be reported in Annex III of the evaluation report that takes the form of a staff-working document⁶⁷⁰.

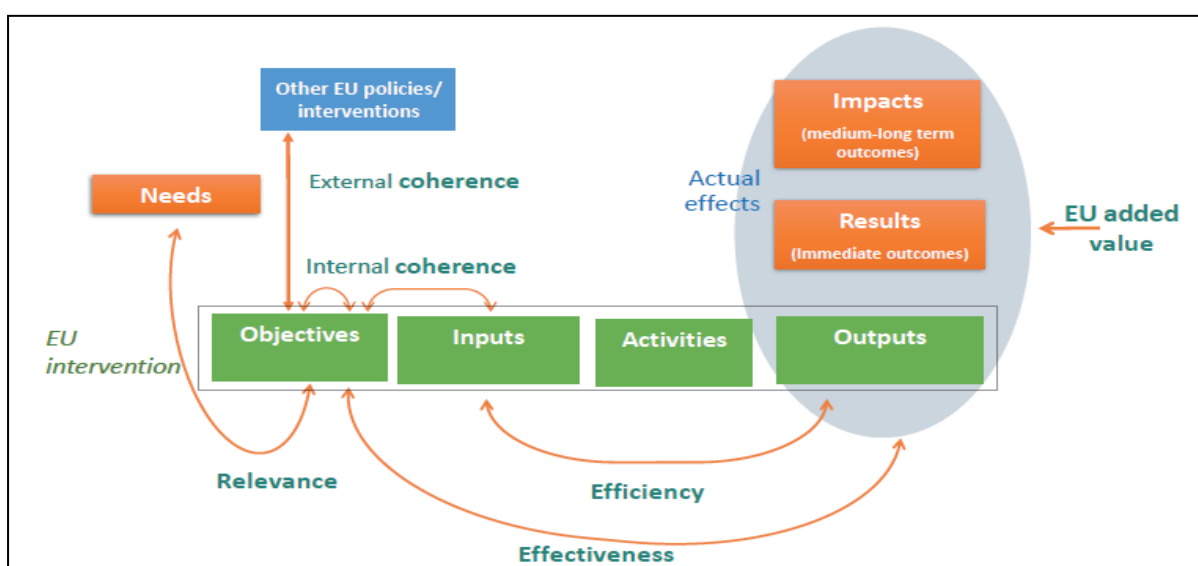


Figure 1: Simplified view of the intervention and the 5 key evaluation criteria

2. EFFECTIVENESS

Effectiveness analysis considers how successful EU action has been in achieving or progressing towards its objectives. The evaluation should form an opinion on the progress made to date and the role of the EU action in delivering the observed changes. If the objectives have not been achieved, or things are not on track, an assessment should be made of the *extent* to which progress has fallen short of the target and what factors have influenced *why* something has not been successful or *why* it has not yet been achieved. To this end, the effectiveness analysis should seek to identify the factors driving or hindering progress and how they are linked (or not) to the EU intervention.

The **effectiveness analysis should look closely at the benefits of the EU intervention as they accrue to different stakeholders**⁶⁷¹. It should identify what factors are driving these benefits and how these factors relate to the EU intervention.

⁶⁷⁰ See Tool #49 (*Format of the evaluation report*)

⁶⁷¹ See also Chapter VIII *Methods, models and costs and benefits*.

Consideration should also be given to whether the objectives can still be achieved on time or with what delay. The analysis should also try to identify if any *unexpected* or *unintended* effects have occurred. The analysis needs to consider how the observed changes may be linked to the actions triggered by the EU intervention. Ultimately, the analysis of effectiveness should provide an answer on whether the EU action has successfully achieved its objectives.

Box 1: Typical examples of questions about effectiveness

- How successful has the EU intervention been in achieving (or progressing towards) its objectives? To what extent were the expected changes resulting from EU action delivered?
- What have the quantitative and qualitative effects of the intervention been? Are they as expected when introducing the intervention?
- What external factors have affected progress towards the objectives and how are they linked to the EU intervention?
- If the objectives have not been achieved yet, can the objectives still be achieved in time? If not, with what delay?
- Are there any unexpected or unintended effects that have occurred, and which drove or hindered progress? What can explain these effects?

Points of comparison to consider

In many cases, performance can be identified from monitoring data covering the relevant period. This can then be compared to a relevant point of comparison⁶⁷² such as:

- What was expected to have happened at this point (based on the adopted proposal, drawing on the appropriate information from a prior impact assessment or other relevant documents, such as the explanatory memorandum), including comparison with the stated specific and operational objectives (and targets if applicable); or
- Another scenario or benchmark, such as the situation prior to implementation, an earlier intervention, groups that did not participate or benefit from the intervention, or similar systems in countries outside the EU.

3. EFFICIENCY

Efficiency considers the resources used by an intervention for the given changes generated by the intervention (which may be positive or negative). Differences in the way an intervention is approached and conducted can have a significant influence on the effects, making it interesting to consider whether other choices (e.g. as demonstrated via different Member States) achieved the same benefits at less cost (or greater benefits at the same cost).

Efficiency analysis should look closely at the costs of the EU intervention as they accrue to different stakeholders⁶⁷³. It should identify what factors are driving these costs and how these factors relate to the EU intervention. The purpose of the assessment of efficiency is to

⁶⁷² For more information on various points of comparison, see Tool #46 (*Designing the evaluation*)

⁶⁷³ See also Chapter VIII on *Methods, models and costs and benefits*

show that resources are used to their best and therefore that the costs generated are strictly necessary to reach the policy objectives. If this is not the case, the potential for simplification is to be highlighted in this analysis.

The efficiency analysis should also compare the identified costs with the benefits that were identified under the effectiveness criterion. As a rule, the benefits of EU interventions are expected to exceed the costs they generate. However, in specific cases, costs might exceed benefits in the short term, with net benefits occurring with a lag. Furthermore, those who bear the costs do not always reap the benefits. This is often the case for safety, health, environment, climate, or consumer protection policies. The evaluation should take into consideration such lags and identify such distributional effects and assess whether they were expected or not. Efforts should be made to address in quantitative terms benefits and costs arising from the intervention.

All evaluations are required to explore the potential for simplification and burden reduction. Typical efficiency analysis will measure administrative and regulatory burden and look at aspects of simplification, which is important for ALL evaluations. Where appropriate, evaluation findings should pinpoint areas where there is potential to reduce inefficiencies, particularly unnecessary regulatory costs, and simplify the design and implementation of the intervention. The full efforts to support and perform an intervention can be broken into different categories such as adjustment costs, administrative costs, fixed costs, running costs, etc⁶⁷⁴.

'Better regulation' and particularly the REFIT programme (commitment on simplification) place a strong emphasis on identifying and where possible measuring (i.e. if possible, quantifying or monetising) the costs and benefits of EU interventions⁶⁷⁵.

Tailored fitness checks, for example those that look at specific reporting requirements, have a strong simplification and burden reduction potential. The efficiency assessment should look particularly at the angle of reducing overlaps, inconsistencies but also at alternative ways of performing the required action(s), such as the scope for simplification.

A cumulative cost assessment (CCA)⁶⁷⁶, although providing inputs into the evaluation process, is not sufficient on its own to provide the required full picture in terms of the efficiency of the EU intervention. As such, it cannot be "the sole basis for policy recommendations"⁶⁷⁷. To serve as an instrument for the policymakers, CCAs need to be put in the context and CCAs need to be supplemented by the analysis of the corresponding benefits arising from the EU legislation⁶⁷⁸.

Assessing costs and benefits may be (methodologically) easier for spending programmes which have well defined stakeholders, systems, etc. Doing this with precision at EU level can be difficult since obtaining robust, good quality data to use in the evaluation of costs and benefits may be challenging, particularly across all Member States which may have

⁶⁷⁴ See Tool #56 (*Typology of costs and benefits*)

⁶⁷⁵ See Tool #2 (*The Regulatory fitness programme (REFIT) and the Fit for Future Platform*); and Tool #57 (*Methods to assess costs and benefits*)

⁶⁷⁶ CCA are studies that aim to estimate the overall regulatory burden on a particular sector. In the EU context, the CCAs will look at the costs arising from the EU regulations.

⁶⁷⁷ Page 15, [Regulatory Fitness and Performance: State of Play and Outlook; COM \(2014\) 368 final](#)

⁶⁷⁸ Further discussion on costs and benefits analysis in the context of CCA is provided in the Tool #57 (*Methods to assess costs and benefits*)

implemented legislation in a variety of different ways and at different points in time. However, sufficient efforts should be dedicated to this task, given its importance.

Table 1: Approach towards efficiency analysis in specific cases

	Costs				Benefits					Benefits and costs have same scope and are directly comparable?
	Specific to the sector	To all affected sectors	General welfare/ Costs to society	Overall cost of a policy	Specific to the sector	To all affected sectors	General welfare / Benefits to society	Of a specific policy	Overall benefit of a policy	
Evaluation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fitness check	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sectoral fitness check	Yes	No	No	No	Yes	No	No	Partial	Partial	Possibly
Cumulative cost assessment	Yes	Possibly	Possibly	Possibly	No	No	No	No	No	N/A

Box 2. Typical examples of questions about efficiency

- What are the costs and benefits for different stakeholders? Were the costs and benefits of the intervention distributed as expected among them? What could explain the differences (if any)?
- To what extent are the administrative and/or adjustment costs of the intervention justified, given the changes/effects it has achieved?
- Have the changes/effects of the intervention been achieved at initially expected costs or were these costs different for whatever reason? What could explain the differences (if any)?
- To what extent do factors linked to the intervention, influence the efficiency with which the observed achievements were attained? What other factors influence the costs and benefits?
- Have any inefficiencies been identified? Could the intervention have been done in a more efficient way? What is the simplification and cost reduction potential of the intervention?
- If there are significant differences in costs (and benefits) between Member States, sectors, etc., what is causing them? Were they expected?
- How timely and efficient is the intervention's administrative process (e.g. for reporting and monitoring)?
- For spending programmes, how efficient was the governance of the programme?
- For decentralised agencies, to what extent has the governance structure of the agency supported its ability to perform its tasks, having regard to its size, composition, organisation and working processes? To what extent were their activities effectively resourced?

Points of comparison to consider

Points of comparison for efficiency are usually quantitative. In many cases, performance can be identified from monitoring data covering the relevant period. This can then be compared to a relevant point of comparison⁶⁷⁹ such as:

- What was expected to have happened at this point (based on the adopted proposal, drawing on the appropriate information from a prior impact assessment or other relevant documents, such as the explanatory memorandum), including comparison with the stated specific and operational objectives (and targets if applicable); or
- Another scenario or benchmark, such as the situation prior to implementation, an earlier intervention, groups that did not participate or benefit from the intervention, or similar systems in countries outside the EU.

4. RELEVANCE

Relevance looks at the relationship between the needs and problems at the time of introducing the intervention and during its implementation. **Relevance should also look at the relationship between the current and future needs and problems in the EU and the objectives of the intervention.** Relevance analysis requires a consideration of how the objectives of an EU intervention (legislative or spending measure) correspond to wider EU policy goals and priorities. Analysis should identify if there is any mismatch between the objectives of the intervention and the current and the foreseeable future needs or problems. For example, 'problem drivers'⁶⁸⁰ identified in the impact assessment may now be obsolete; circumstances may have changed, and the needs and problems now may not be the same as the ones looked at when the intervention was designed.

Relevance analysis is very important – because if an intervention does not help address the current and the likely future needs or problems (on the basis of new objectives) then it does not matter how effective, efficient or coherent it is – it may no longer be appropriate (this is why relevance is sometimes called the 'kill' criterion!). The relevance analysis may also consider future needs or problems identified on the basis of trends and foresight analysis. This is key information that will assist policy makers in deciding whether to continue, change or stop an intervention and also explains the strong link between relevance and the criterion of EU added value – which assesses whether action continues to be justified at the EU level.

Box 3. Typical examples of questions about relevance

- To what extent did the scope and objectives of the intervention remain relevant over the implementation period?
- How did the objectives of the intervention (legislative or spending measure) correspond to wider EU policy goals and priorities?
- To what extent is the intervention still relevant in view of the objectives? How well do

⁶⁷⁹ For more information on various points of comparison, see Tool #46 (*Designing the evaluation*)

⁶⁸⁰ See Tool #13 (*How to analyse problems*)

the (original) objectives of the intervention still correspond to the needs within the EU?

- To what extent do the needs/problems addressed by the intervention continue to require action at EU level?
- How well adapted is the intervention to the technological or scientific progress that has appeared since its introduction? How well adapted is the intervention to social, environmental changes or developments related to implementation, reporting and compliance?
- Have the circumstances changed in the meantime so much that the intervention has to change/adapt to them over time?
- How relevant is the intervention to EU citizens? Will it continue to be relevant for them in the near future?
- For decentralised agencies, how have the agency's activities contributed to the achievement of the overall EU policy objectives and priorities, and how well have they been able to respond to the needs of their stakeholders?

Points of comparison⁶⁸¹ for relevance are usually more qualitative. It is necessary to think about what the needs and objectives behind the EU intervention were and compare them to the current situation. It is also advisable to consider how the situation is likely to develop in the future (e.g. based on new political priorities and objectives, anticipated or upcoming technological, social, environmental or economic changes, and the use of strategic foresight information). In addressing the question of continued relevance, it may be useful to look at how megatrends (such as demographic change or digitalisation of services) may impact the intervention. Information on foresight approach and related instruments can be found in Tool #21 (*Strategic foresight for impact assessments and evaluations*).

5. COHERENCE

The evaluation of coherence involves looking at how well (or not) different interventions, EU/international policies or national/regional/local policy elements work together. It may highlight areas where there are synergies which improve overall performance or which were perhaps not possible if introduced at national level; or it may point to tensions e.g. objectives which are potentially contradictory, overlapping or approaches which are causing inefficiencies.

Checking '**internal**' coherence means looking at how the various components of the same EU intervention operate together to achieve its objectives, e.g. the different articles of a piece of legislation, different actions under an action plan. Checking '**external**' coherence means that similar checks can be conducted in relation to other ('external') interventions, at different levels: for example, between EU interventions within the same policy field (e.g. a specific intervention on drinking water and wider EU water policy) or in areas which may have to work together (e.g. water policy and chemicals policy, or chemicals and health and safety). Where relevant, analysis of coherence may involve checking whether interventions are in line with the objectives of the European Green Deal, or whether the intervention is consistent with the overarching environmental goals (such as the Climate Law) or other policies targeting the environment⁶⁸². At its widest, external coherence should also look at compliance with

⁶⁸¹ See Tool #46 (*Designing the evaluation*)

⁶⁸² See Tool #36 (*Environmental impacts*)

national policies or international agreements/declarations (for example EU labour market interventions might be looking into coherence with ILO conventions), in particular the UN sustainable development goals or EU interventions in developing countries.

The focus on coherence may vary depending on the type of evaluation **and is particularly important in fitness checks**, where coherence analysis will look for evidence of synergies or inconsistencies between policies in a related field that are expected to work together. Even when evaluating an individual intervention, it is important to check coherence with other interventions which have the same or similar objectives.

When assessing coherence, comparison with other scenarios is predominantly qualitative. Suitable points of comparison⁶⁸³ could include looking at changes in coherence between the start and end of the period being evaluated or, if applicable, information from the impact assessment (i.e. assessment of coherence of preferred policy option). They might also include looking at existing or new national or international actions, including the UN sustainable development goals that have been adopted in 2015 and are now at the heart of the Commission's policymaking. The level of coherence being evaluated (e.g. internal to a given EU action, within a given policy field, wider EU policy or wider global context) could also affect the comparison point and degree of analysis possible.

Box 4. Typical examples of coherence questions

- To what extent is this intervention coherent with other EU and national interventions that have similar objectives⁶⁸⁴?
- Does the intervention comply with the green oath to 'do no significant harm', and could it help respond better to significant harm?
- To what extent are the various elements of intervention coherent with one another?
- To what extent have the various elements of intervention generated synergies and/or compensated possible trade-offs among them?
- To what extent is the intervention coherent with (current) wider EU policies and priorities (e.g. Commission policy priorities)?
- To what extent is the intervention coherent with international obligations, including the SDGs?
- For decentralised agencies, to what extent has the agency coordinated its work and made best use of existing resources with other relevant bodies, governmental institutions, and agencies active at EU, national and international level carrying out similar tasks? To what extent were inappropriate overlaps of mandates avoided?

6. EU ADDED VALUE

EU added value⁶⁸⁵ looks for changes that are due to the EU intervention, over and above what could reasonably have been expected from national actions by the Member States. In many

⁶⁸³ See Tool #36 (*Environmental impacts*)

⁶⁸⁴ Outline these objectives to make the question less abstract.

⁶⁸⁵ For further information see [SEC\(2011\) 867 final "The added value of the EU budget"](#).

ways, the evaluation of EU added value brings together the findings of the other criteria, presenting the arguments on causality and drawing conclusions, based on available evidence, about the performance of the EU intervention.

Under the principle of subsidiarity (Article 5 Treaty on European Union), and in areas of non-exclusive competence, the EU should only act when the objectives can be better achieved by Union action rather than action by the Member States. It requires consideration of the added value of EU action compared to that of other actors⁶⁸⁶. **EU added value analysis should, where applicable, respond to the subsidiarity analysis conducted in any related IA.**

In practice, the EU added value in evaluations should assess retrospectively whether the subsidiarity analysis was valid, and the policy remains fit for purpose. The EU added value assessment provides the verification of compliance with the subsidiarity principle, based on information and analysis of impacts that actually occurred.

The sources and nature of the EU added value vary from intervention to intervention. It is useful to distinguish the European added value of an EU policy measure in general (such as an EU regulation to foster the single market) and that of an EU spending programme per se. In both cases, EU added value may be the result of different factors: coordination gains, legal certainty, greater effectiveness or efficiency gains, complementarities, synergies, etc. In all cases, concluding on the continued need for the intervention at EU level may be difficult as the measurement of EU added value is challenging.

In areas where the EU has exclusive competence⁶⁸⁷, the appropriate answer to the question of EU added value may simply involve re-stating the reasons why the EU has exclusive competence⁶⁸⁸ or may already be answered by the efficiency and effectiveness analysis.

When assessing the EU added value, the comparison⁶⁸⁹ involves consideration of performance against a projection of how the situation was expected to evolve without the EU intervention (a defined counterfactual⁶⁹⁰, or some estimate of the cost of the Union not acting – ‘the cost of non-Europe’). Often such analysis is qualitative, analysing whether the subsidiarity arguments put forward before the intervention (as presented in a prior impact assessment, or other accompanying documents such as the explanatory memorandum) were valid and whether the expected changes resulting from EU action were delivered. It may also be appropriate to analyse whether any contextual change or other factors affected the assumption that such change could only be generated by EU level action.

Key steps for assessing EU added value in an evaluation:

- Check whether the explanatory memorandum or impact assessment accompanying the Commission’s legislative proposal contain adequate justification regarding compliance with the principle of subsidiarity.

⁶⁸⁶ See also Tool #5 (*Legal basis, subsidiarity and proportionality*)

⁶⁸⁷ Article 3 TFEU of the Lisbon Treaty defines the following areas as exclusive EU competences: the competition rules within the internal market, the customs union, the common commercial policy, monetary policy for the Euro countries, the conservation of marine biological resources under the common fishing policy and the conclusion of international agreements.

⁶⁸⁸ See Tool #5 (*Legal basis, subsidiarity and proportionality*)

⁶⁸⁹ See Tool #46 (*Designing the evaluation*)

⁶⁹⁰ See Tool #68 (*Methods for evaluating causal effects*)

- Formulate appropriate questions to check whether the subsidiarity justification in the IA or explanatory memorandum is still valid and whether the expected results have materialised. The subsidiarity grid that supports a proposal may be also used to identify relevant questions. Indicative examples of this set of questions are also included in Box 5 below.

Box 5. Typical examples of questions on EU added value

- Are the subsidiarity arguments put forward in a prior impact assessment, or other accompanying documents such as the explanatory memorandum or subsidiarity grid, still valid?
 - Could the objectives of the policy have been achieved sufficiently by the Member States acting alone?
 - Would national action or the absence of EU level action significantly damage the interests of other Member States?
 - In the absence of EU level action, to what extent would Member States have had the ability or possibility to put in place appropriate measures?
 - In case the initial problem and its causes (e.g. negative externalities, spill over effects) varied across the national, regional and local levels, did the EU level action help establish a level playing field?
 - Were there significant/appreciable transnational/cross-border aspects being tackled? Could these be quantified?
 - Were there clear benefits from EU level action?
 - Were there economies of scale and services sharing? Were the objectives met more efficiently at EU level than they would have been met by Member States acting individually (larger benefits per unit cost)?
 - Did the functioning of the internal market improve?
 - Were there benefits in replacing different national policies and rules with a more homogenous policy approach?
 - Did the benefits of EU-level action outweigh the loss of competence of the Member States and the local and regional authorities (beyond the costs and benefits of acting at national, local, and regional levels)?
 - Were some Member States more affected than others? If so, to what extent have these differences been offset, or have they created negative net trade-offs between the Member States?
 - Was there improved legal clarity from implementing EU legislation?
- Could the identified results/outputs/impacts have been achieved without EU intervention?
- Is it still valid to assume that the objectives of the intervention can best be met by action at EU level?
- What would be the most likely consequences of stopping or withdrawing EU intervention?
- What is the additional value resulting from the EU intervention(s), compared to what could reasonably have been achieved (in terms of effectiveness and efficiency) by Member States acting at national and/or regional levels?

7. THE USE OF ADDITIONAL EVALUATION CRITERIA

There are cases where the legal basis (or the specific nature of the intervention) requires the **use of additional evaluation criteria**. The most common additional criteria evaluated by the Commission are listed below. Most of these can be addressed through evaluation questions under the existing five compulsory criteria. Therefore, unless the legal basis of an intervention explicitly requires reporting on an additional criterion, every effort should be made to address additional criteria as part of the five compulsory ones. For example:

- **utility:** To what extent do the changes or effects of an intervention satisfy (or not) stakeholders' needs? How much does the degree of satisfaction differ across different stakeholder groups? Utility could be addressed through evaluation questions under the relevance and/or EU added value criteria.
- **complementarity:** To what extent do EU policies and interventions support and usefully supplement other policies (in particular those pursued by the Member States)? Complementarity could be addressed through evaluation questions under the coherence criterion.
- **coordination:** To what extent are **interventions** organised to maximise their joint effects to better achieve objectives, e.g. by mobilising resources combined with harmonising measures? Coordination could be addressed through evaluation questions under the coherence and/or efficiency criteria.
- **equity:** How fairly are the different effects distributed across the different stakeholders/regions/genders / social groups⁶⁹¹? Equity could be addressed through evaluation questions under the effectiveness criterion.
- **sustainability:** *How likely are the economic and/or social and/or environmental effects to last after the intervention ends?* It is often hoped that the changes caused by an intervention will be long-lasting. It can be important to test this expectation for interventions, which have a finite duration, such as particular programmes. Sustainability could be defined as an objective and addressed through evaluation questions under the effectiveness criterion, given that it is related to the durability of the effects. Sustainability could also be a crosscutting criterion not limited to durability of the policy effects. For example, sustainability of the economy, the environment or the social fabric when implementing the intervention. In this case too, sustainability could be defined as an objective and addressed through evaluation questions under the effectiveness criterion (e.g. to what extent have the intervention fostered a sustainable use of the natural resources while achieving its objectives?).
- **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? Acceptability could be addressed through evaluation questions under the effectiveness criterion, as it is related to the degree of acceptance of the effects.

Care needs to be used to avoid the multiplication of criteria, which may have limited added value and make the evaluation analysis repetitive. **The use of additional criteria needs to be duly justified.**

⁶⁹¹ Such as for example low-income groups, persons with a minority ethnic or racial background, persons with disabilities etc.

8. GOOD PRACTICE TIPS

- Evaluation questions should be worded in a way that forces the evaluator to go beyond an answer based on simple description. Questions that start with 'how', 'why', 'to what extent' are more likely to ensure that the answer provided looks at what the links were between the changes observed and the EU intervention(s). Questions that start with verbs such as 'do' the directives...? 'Are' the directives providing...? should be avoided as they tend to provoke yes/no answers, or they should be accompanied with a request to explain.
- If necessary, use clarifications or sub-questions to define specific elements for the evaluator to consider in answering the evaluation questions (e.g. in answering the question, special attention should be paid to ...)
- Dare to ask the difficult question, including those that may challenge current policy.
- Try not to have too many evaluation questions. Sometimes it is necessary to have very specific questions, other times it is better to have a more generic set and see where the data/analysis leads. There is always a trade-off between the number of questions and the depth of analysis that can be conducted, especially across all Member States.
- Check any prior impact assessment to see what issues were addressed and what expectations were presented. Where necessary compare the proposal accompanying the impact assessment to the final actions adopted/introduced and try to identify where amendments to the Commission proposal may have changed the intervention logic described in the impact assessment.
- Encourage consideration of the 'end-user' perspective. End users are most affected by actions triggered due to EU interventions – they have practical experience of what has happened on the ground and may have a different perspective from policy makers, governments, NGOs, etc.
- Evaluation questions, which guide the work of the evaluator, are often worded in a technical way, using terminology that makes sense within the Commission, but which requires further explanation (simplification) to many stakeholders. **It is strongly recommended to avoid using the evaluation criteria as such in formulating questions** that look for input from stakeholders. Instead, 'translate' the criteria into more manageable and understandable concepts, particularly for consultation, data gathering exercises and for drafting the report. For example, rather than referring in an abstract manner to the 'objectives of the policy' describe the objective in more common terms – so ask 'What progress has been made towards increasing the availability of funding to small and medium sized businesses'?

TOOL #48. CONDUCTING THE EVALUATION

1. INTRODUCTION

When the planning and designing is over, the actual evaluation work starts. **The final product of this process is the evaluation report, which takes the form of a staff working document written by the lead Directorate General(s) for all evaluations and fitness checks⁶⁹².**

The support work for an evaluation can be outsourced to external contractors and/or draw on the (internal) work of Commission services including services offered by the JRC. External support work may involve one or several studies (contracts). The evaluation report brings together all work⁶⁹³ carried out during the evaluation process.

For all work conducted as part of the evaluation, the evaluation manager⁶⁹⁴ and the interservice coordination group (ISCG) have a particular role to steer the project and improve its quality at all key steps. Where there is external work, this may include advising contractors, supervising their work (and hence its quality), and enforcing the timetable.

It is important to constantly check the quality of the work being undertaken, ensuring that it is evidence-based and free from bias. Thorough, robust and reliable research, data collection and analysis, are core activities to conducting high-quality evaluation and drawing appropriate evaluation findings and conclusions. Robust and reliable results can be delivered only by objective evaluations. Every effort should be made to ensure the transparency of the evaluation – both in terms of how it progresses (e.g. involvement of ISCG, working group, stakeholders) and when reporting (e.g. in terms of collection and use of data, analysis and results). Any limitations to the method applied or the data collected should be clearly discussed over the course of the evaluation, addressed where possible and described in the final report.

Equally, care should be taken to spot weaknesses in:

- the data: e.g. do they come from a reliable source? Have enough respondents replied? Do we not overly rely on only one data source?;
- 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?
- the timing of the evaluation: e.g. have expected impacts been reasonably materialised?

⁶⁹² See Tool #49 (*Format of the evaluation report*)

⁶⁹³ The vast majority of evaluations involve external contractors in some way. They can be commissioned for all or some tasks such as:

- collecting and analysing the relevant evidence (including consultation work);
- developing analytical models and running them;
- providing 'first' answers to some/all evaluation questions;
- presenting evidence-based conclusions.

The terms of reference (ToR)/Technical Specifications are written by the Commission services and set out the work that a contractor is required to do.

⁶⁹⁴ The policy officer responsible for the evaluation and/or the evaluation function/unit of the lead DG.

Box 1. Conducting an evaluation – key principles

- The ISCG and the evaluation manager play a key role in assuring the quality of the overall evaluation.
- A public consultation for an evaluation is not always necessary. It is at the discretion of a Directorate-General whether public consultation is needed to support an evaluation. In all cases, appropriate targeted consultation activities should be carried out. For evaluations of broad public interest and for fitness checks, a public consultation is highly recommended
- The evaluation methodology should follow that identified in the design phase (including any commitments in associated Terms of Reference/Technical Specifications) or explain why this has not been possible.
- All evaluations should consider the evidence base built up over earlier parts of the policy cycle and in particular any prior impact assessment, where appropriate.
- All evaluations should make credible efforts to obtain data from a wide range of qualitative and quantitative sources and distinguish between the opinion of, or data from, vested interests and independent sources. Where possible, it is recommended that the supporting data substantiating the evaluation assessment is available in an easily accessible format.
- Particular effort should be made to access and reuse data that is already collected, for instance by ESTAT, or by public services in Member States for administrative purposes (admin data).
- Proportionate effort should be made to quantify costs and benefits, reflecting the role of the intervention being evaluated and external expectations or discussions (e.g. where there has been significant debate or disagreement between stakeholders in relation to actual costs or benefits delivered). Where this is not possible, a clear explanation of the efforts made, and the restrictions encountered should be given.
- Ethics and integrity concepts should be respected. Any conflicts of interest should be reported to the appropriate actor in the Directorate-General and Secretariat-General.
- Evaluations should be evidence based and adhere to the principles of objectivity and independence⁶⁹⁵.

2. THE ROLE OF THE EVALUATION INTERSERVICE COORDINATION GROUP

Interservice coordination group (ISCG)

The ISCG must be involved in all key steps of the evaluation following validation until the launch of the interservice consultation on the evaluation report and if applicable, the associated executive summary. It may be consulted on any associated report/communication to the European Parliament and Council. When applicable, the group should discuss the draft evaluation report prior to its submission to the Regulatory Scrutiny Board (RSB). The minutes from this discussion are submitted to the Board⁶⁹⁶.

⁶⁹⁵ See Tool #45 (What is an evaluation and when it is required)

⁶⁹⁶ See Tool #3 (Role of the Regulatory Scrutiny Board) for the list of documents submitted to the Board

Who?	The group is led by the relevant lead DG or service. A representative of their evaluation function / 'better regulation' support unit must also be included in the ISCG.
	DGs with policies linked to the subject of the evaluation or likely to be affected by the evaluation should be invited to participate. A representative from the Secretariat-General must be invited.
	By definition, the ISCG is an internal Commission group, made up of representatives from DGs. However, this does not prevent consultation with other groups in consideration of their knowledge and expertise, which include non-Commission staff (e.g. from EU decentralised agencies and other EU bodies, Member States, academia) which could also provide advice and information. This is particularly relevant when a partner DG is evaluating an EU decentralised agency. In such a case, it is important to involve representatives of the agency in the evaluation process, but they will not be a member of the ISCG. One way to do this would be to create an ISCG, but to discuss relevant aspects of the work in an additional, wider group (identified at the discretion of the lead DG), which includes both the ISCG members and agency representatives. Further examples of such groups include Member State advisory committees, academic or reference panels.
	In addition, DGs with core expertise in specific areas, such as economic analysis (e.g. ECFIN), scientific research and analytical models (e.g. JRC, RTD), social impacts (e.g. EMPL), SMEs, competitiveness (e.g. GROW), environment (e.g. ENV), fundamental rights (JUST), development cooperation (INTPA), innovation (RTD), digital/ICT (CNECT and DIGIT)⁶⁹⁷, should also participate where appropriate to ensure that the evaluation calls upon all relevant expertise in the Commission.
	The invitation to set up an ISCG should take the form of a note from a senior manager of the lead DG to those of the identified DGs asking to nominate a representative.
	Existing ISGs can be used to steer the evaluation work particularly where such a group has been used to conduct a prior impact assessment or to provide advice to policy development.
	Consultants may be invited to make presentations regarding supporting studies or contracts but should leave the meeting when substantive discussions take place between ISCG members. The lead DG should make sure the confidential nature of internal ISCG discussion remains protected.
Why?	An effective ISCG 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, helps to ensure coherence with other policies and demonstrates an open and transparent approach to critically analysing performance and delivery. Involving other services should also anticipate (and solve) problems that could emerge later in the process (e.g. during interservice consultation). Colleagues from other areas or with diverse professional profiles are a good test of whether your arguments are clear and easy to follow.

⁶⁹⁷ See Tool #28 (Digital-ready policymaking)

	The ISCG encourages a pooling of knowledge and brings together a range of different perspectives. The mix of policy/operational experience and evaluation expertise should prevent biases from taking over the evaluation and encourage critical judgement.
	The ISCG is the best way to ensure that a DG's views are considered by the lead DG. It is important, therefore, to plan participation well in advance and, to the extent possible, for members to participate pro-actively. As far as possible, make sure the position expressed in an ISCG is representative of the position your DG is likely to take during ISC.
When?	An ISCG can be established/convened as soon as the evaluation has been included and validated in Decide⁶⁹⁸. The ISCG should meet as many times as needed to cover the important elements of the evaluation. When the evaluation is to be submitted for scrutiny, the ISCG should also discuss the final draft prior to its submission to the Board. More meetings or consultation of the ISCG in writing can also be envisaged, particularly in the case of complex evaluations running over a long period. Meetings may also follow the timing of other milestones such as an external study or a stakeholder consultation. It is important to remain mindful of the workload this generates for other services.
How?	The first meeting of the ISCG should discuss the intervention logic, evaluation questions and other elements affecting the design/conduct of the evaluation including the consultation strategy. It is recommended that the 'call for evidence' is also discussed in the first meeting, together with the consultation strategy. Ideally, the evaluation questions will be finalised shortly after this meeting. It is good practice to share information and record decisions on these key elements in written format.
	The ISCG must steer the evaluation through subsequent key phases (e.g. design /conduct, including consultation, studies and evaluation report that takes the form of a staff working document), providing input and information and ensuring the quality, impartiality and usefulness of the final product.
	The ISCG should also be involved in the preparation of terms of reference/technical specifications for external studies and, as part of this, drawing up the scope of possible modelling work. The ISCG <u>should be involved</u> in the design of stakeholder consultation strategy and any consultation documents. It should discuss any feedback received from stakeholders on the 'call for evidence'. It should always discuss intermediate findings, e.g. deliverables and quality of modelling work or supporting studies, and drafts of the evaluation report. The ISCG is not involved in any strict procurement-related activity in the context of the evaluation work.
	Meetings should be well prepared with <u>invitations and documents being circulated in principle at least one week in advance</u>. Similarly, ISCG members should be given <u>at least one week since they received them</u> to provide written comments on the documents. Minutes of meetings should be prepared which

⁶⁹⁸ See Tool #6 (*Planning and validation of initiatives*)

	record transparently and accurately the views of the ISCG members. The minutes of the last ISCG meeting should be attached to the covering note if the draft evaluation report is submitted to the RSB.
	The lead DG is recommended to establish a collaborative workspace for sharing documents which facilitates more flexible participation by DGs.

3. KEY PROCEDURAL STEPS

The time needed to prepare an evaluation will vary from case to case. Sufficient time needs to be allocated to ensure that the evaluation can be conducted according to these guidelines and, where necessary, the Commission can report to the European Parliament and Council by the date set in the legal base. Where an evaluation is linked to a (review) clause that invites the Commission to present new proposals by a certain date, care must be taken to ensure that the planning allocates sufficient time for the evaluation and the impact assessment. Good planning also implies taking account of the time needed to meet the various procedural requirements, including scrutiny by the Regulatory Scrutiny Board, political validation to launch the interservice consultation and the time needed for translations.

Box 2. The key steps in an evaluation

1. **Political validation:** generally evaluations to be conducted by a Directorate-General are approved during the management plan process. At an appropriate point in time, each individual evaluation is introduced and validated in Decide⁶⁹⁹;
2. **Establish an interservice coordination group (ISCG)** to steer the evaluation;
3. Draft the '**call for evidence**', for consultation with the ISCG and agreement with the Secretariat-General. Finalise the consultation strategy with the ISCG, including the 12-week internet-based public consultation, where applicable;
4. **Steer and conduct the evaluation work** which may involve managing supporting studies by contractors. Involve the ISCG as appropriate throughout the conduct of the evaluation work;
5. Prepare the **evaluation report**⁷⁰⁰ that responds to the issues in the 'call for evidence', presents the analysis, and answers the evaluation questions. Where applicable, prepare an **executive summary** presenting the findings of the evaluation report⁷⁰¹;
6. For those **evaluations selected for scrutiny**, submit the draft evaluation report together with the supporting documents to the Regulatory Scrutiny Board⁷⁰². **Address the Board's comments** and incorporate them into a revised version prior to launching the interservice consultation. Where the opinion is negative, the DG has an option to resubmit an amended report to the Board;
7. Launch the **interservice consultation**;
8. **Publish** the evaluation report and any supporting contractors' (final) study. **Where required by the basic legal act**, transmit the evaluation to the Parliament and the

⁶⁹⁹ See Tool #6 (*Planning and validation of initiatives*)

⁷⁰⁰ The evaluation report takes the form of staff working document outlined in Tool #49

⁷⁰¹ For exceptions to the preparation of the executive summary, see Tool #49.

⁷⁰² See Tool #3 (*Role of the Regulatory Scrutiny Board*) for the list of documents submitted to the Board

Council together with a short Commission report/communication;

9. **Disseminate the evaluation findings** as appropriate and identify any **appropriate follow-up actions** to put into practice the lessons learned and capitalise on the evaluation findings. Feed the evaluation findings into the next step of the decision-making cycle (Annual Activity Reports, Annual Management Plans, impact assessments for revisions, etc.)

TOOL #49. FORMAT OF THE EVALUATION REPORT

1. INTRODUCTION

The evaluation report takes the form of a staff working document (SWD)⁷⁰³. The report is the final output and key deliverable of the evaluation process, presenting the evidence-based judgement of an EU intervention or set of EU interventions for fitness checks.

The lead DG should present the conclusions of the evaluation in a way that is useful to policymakers and that can serve as a basis for future policy development. The report also serves to communicate to stakeholders the methods, evidence base and analyses applied when evaluating the EU intervention. The drafting of the evaluation report in the form of a staff working document falls under the exclusive responsibility of the European Commission services.

The report should tell an evidence-based story of the EU intervention. Like the impact assessment report, the evaluation report should answer a set of key questions in separate sections. Any limitations to the robustness of the process and evaluation findings must be analysed in the report. The process and methodology used to undertake the evaluation should be described and detailed. Throughout the report all evidence should be clearly presented, referenced and, if possible, hyperlinked. DGs must use the standard format described below for the report, which will ensure consistency across the Commission.

Evaluations may draw on a range of sources and methods, including legal analysis, statistics, expert opinions, external supporting studies, surveys, consultations, case studies, analytical models⁷⁰⁴. Where evaluations are largely based on the work of external contractors (e.g. supporting studies), the evaluation report should draw on this work, but it is the lead DG that takes ownership of the findings and conclusions of the evaluation. If there are reasons why the lead DG thinks there are different answers or draws different conclusions to those of the external contractor, this should be brought out in the report, together with the necessary supporting justification – either by showing why they interpret the evidence differently, or by bringing in additional information.

2. WHY IS THE EVALUATION REPORT IMPORTANT AND HOW IS IT STRUCTURED?

The evaluation report is the key document that will inform stakeholders and policymakers on the outcome of the evaluation, presenting the judgements and lessons learned. The evaluation report is expected to tell the story of what has been achieved with a particular EU intervention and how it was achieved (or not) in narrative form. **It should be written by the lead DG** irrespective of whether it draws or not on an external supporting study. It is the response to the issues raised in the 'call for evidence' and answers the questions of the evaluation (see Box 1 below). It is also the basis for any follow-up action, such as revision of legislation. It can provide an indirect feedback mechanism acknowledging the contributions that stakeholders and experts have made throughout the process. With the evaluation report, Commission services take ownership of the evaluation work.

⁷⁰³ The evaluation report or fitness check in the form of a SWD is hereafter shortened as 'the report' unless specified otherwise.

⁷⁰⁴ See Tool #4 (*Evidence-informed policymaking*).

The report should be a self-standing document that follows the standard structure set out below, to ensure consistency across Commission services. It should be written using non-technical language with non-expert readers in mind and should provide the reader with a complete picture of the main issues and findings. More detailed information or explanations, including on the methodologies used, should be provided in the relevant annexes.

An evaluation report should be accompanied by an executive summary as a stand-alone document. An executive summary is not needed when the Commission provides a report to the European Parliament and the Council.

Box 1. Mandatory elements of the evaluation report

- The evaluation report should contain:
 - A critical, evidence-based judgement of the EU intervention, based on a range of data and analysis. It should be presented in narrative form ('story') guided by the evaluation questions which the evaluation intended to address.
 - The answers to the following questions that make up the intervention's assessment 'story':
 - What was the expected outcome of the intervention?
 - How has the situation evolved over the evaluation period?
 - Evaluation findings:
 - To what extent was the intervention successful and why?
 - How did the EU intervention make a difference and to whom?
 - Is the intervention still relevant?
 - What are the conclusions and lessons learned?
 - A clear chain of logic between the evidence, analysis and findings, the answers to the questions and the conclusions drawn.
 - Clear conclusions based on the evidence collected which generates useful information for future policy decisions and which helps the organisation to learn.
 - Annexes on (I) procedural information, (II) methodology and analytical models used, (III) evaluation matrix and answers to the evaluation questions, (IV) overview of benefits and costs and, where relevant, table on simplification and burden reduction and (V) stakeholder consultation-synopsis report.
 - A summary of the changes introduced following the opinion of the Regulatory Scrutiny Board (RSB) if the evaluation report has been scrutinised (to be included in Annex I on procedural information).
- All contractors' final studies (plus other relevant interim deliverables from external work) and the evaluation report should follow the appropriate corporate publication requirements⁷⁰⁵.
- Non-confidential data used in the evaluation or supporting studies should be publicly available.

⁷⁰⁵ See section 7 below and GoPro for additional information on publication requirements.

The evaluation report presents in a self-standing and non-technical manner the process, evidence, analysis, conclusions, and lessons learned. **The evaluation report should not be longer than 50 pages of text accompanied by annexes as appropriate.** Even where the full body of work described in the 'call for evidence' has been outsourced to contractors, who have written up their process and findings as a separate study, the evaluation report must provide sufficient detail, enabling the reader to follow the evidence and logic and understand the answers and conclusions without having to read the contractors' report or any other supporting materials.

Underlying data, statistics, information, expert contributions, and stakeholder views should all be referenced, particularly where choices are made, or conclusions are drawn based on them. Whenever possible, direct hypertext internet links should be provided.

Stakeholder views should be integrated throughout the text of the evaluation report. A description of the views of the different stakeholder groups should be included and any differences within or across such groups should be highlighted. A separate Annex should contain the analysis of the stakeholder consultation.

The evaluation report template and executive summary cover page can be downloaded from [GoPro](#).

3. DETAILED STRUCTURE AND CONTENT OF THE EVALUATION REPORT AND ITS EXECUTIVE SUMMARY

The 'better regulation' criteria of effectiveness, efficiency, relevance, coherence, and EU added value serve to organise the analytical part of the evaluation work and collect the factual evidence on these criteria. The results of the data collection feed into the evaluation report and help draft the 'story' of EU intervention. The evaluation report should tell what was achieved and how the EU intervention helped this happen, but also what was not achieved, why so and what lessons were learned.

To the extent possible, the report should ensure a strong link between the retrospective evaluation and any prior impact assessment. Evidence on what was achieved and not achieved and how EU intervention made this happen is crucial information feeding into the problem definition and baseline option of any future impact assessment. Conversely, the discussion on what success was expected to look like (as in a prior impact assessment, explanatory memorandum, etc.) provide the starting point for the evaluation.

The evaluation report should follow the structure presented below. Each section provides further guidance on the issues to be covered.

Section 1 – Introduction

Section 1 explains the purpose and scope of the evaluation / fitness check. Issues to be addressed in this section are:

Purpose and scope of the evaluation / fitness check

- *What is the purpose of the evaluation / fitness check (refer to the legal basis of the EU intervention being evaluated, TFEU article, etc.), what will the evaluation / fitness*

check deliver and how will its results be used (e.g. to fulfil a legal obligation, provide the basis for a possible future IA, to improve application)?

- *What are the main issues the evaluation addresses (check legal obligations besides 'better regulation' requirements; check consistency with the text in the 'call for evidence')?*
- *What evaluation criteria are applied (noting that all evaluations should investigate effectiveness, efficiency, relevance, coherence, and EU added value of the EU intervention)? To what extent are the five compulsory criteria covered?*
- *A few, concise sentences informing about the methodology applied, its robustness and its limitations. Methodological information on how the evaluation was conducted is to be detailed in the Annex II to the evaluation report.*
- *What time period does the evaluation cover (from the start of the intervention until now; when covering a different time period, explain why)?*
- *What geographical scope does the evaluation cover? (if not all Member States and if other countries, explain why).*
- *Which related implementing / delegated acts does the evaluation cover? If they are not covered, explain why.*

Section 2 - What was the expected outcome of the intervention?

Section 2 explains the rationale for the intervention at the time it was prepared/adopted, i.e. the problem or the needs the EU was trying to address and its underlying causes, what it expected to achieve and how that achievement was to be assessed (the intervention logic). It should draw to the extent possible on the impact assessment report (if available), the explanatory memorandum and the preamble of the final legal text. It should also include the points of comparison against which the intervention will be assessed.

Issues to be addressed in this section are:

Description of the intervention and its objectives

- *Explain the logic of the intervention being evaluated and establish the link with the preceding impact assessment (if any). Clarify to what extent the analysis in the impact assessment is still relevant (depending on what was finally adopted).*
- *Provide a brief description of how the intervention fitted in the wider policy framework (in the past).*
- *Include, wherever possible, a reference to the UN sustainable development goals (SDGs) that the intervention aimed to address and sketch which SDG-related insights the evaluation can provide⁷⁰⁶.*
- *Provide a brief description of the problems and/or needs the intervention was intended to solve.*
- *Provide a brief description of the objectives that were agreed upon to solve/address the problems/needs.*

⁷⁰⁶ See Tool #19 for more guidance on sustainable development goals and references to the available toolkit.

- *Provide a brief description of the expected achievements at the time of the evaluation in terms of outputs, results, and impacts.*
- *Provide a brief description how the actions of the intervention were ultimately expected to deliver on the objectives (i.e. what success was expected to look like).*

Summarise this information in a descriptive text, preferably supported by an intervention logic diagram. It should connect the needs (i.e. problem definition in the impact assessment or in the explanatory memorandum), with the objectives, the intervention's actions and the expected achievements in terms of outputs, results and impacts. It is generally helpful to use a graphical representation (visual aid) illustrating how the different components were expected to fit together⁷⁰⁷. This should also refer to potential external factors influencing the expected achievements.

Point(s) of comparison

- *Describe what the situation was like before the intervention started, how it was expected to develop and/or any other relevant points of comparison.*
- *Use all relevant information from the preceding impact assessment (if any), analytical documents accompanying the legal proposal and/or the explanatory memorandum.*
- *Describe the point(s) of comparison against which you will assess your intervention. You may need to use different points of comparison for assessing the different evaluation criteria and should clearly explain the choice made.*
- *Use tables / graphs / other visual aids as necessary to help the reader understand the text.*

This discussion should draw on the prior impact assessment if available (but updated to reflect changes during the adoption process) and/or the explanatory memorandum. The discussion should cover in particular the situation linked to the problems/needs the intervention was intended to solve (a quantitative description should be provided as far as possible). This should facilitate a comparison with the current situation and should therefore cover as far as possible the same parameters/indicators that are used to describe the state of play in the following section (Section 3).

Section 3 - How has the situation evolved over the evaluation period?

Section 3 explains the state of play in implementing the EU intervention (both legally and on the ground) and presents what has happened so far. The presentation in this section should be factual as in Section 2 ('What was expected to be achieved?'). Any judgements on the findings should not be presented here but in Section 4 ('Evaluation findings').

Issues to be addressed in this section are:

Current state of play

- *Explain how the intervention has been implemented, summarising which Member States have done what and what problems/infringements have been identified.*

⁷⁰⁷ See Tool #46 (Designing the evaluation)

- *In case of delays in implementation, explain what happened and consider the impact of delays on the implementation.*
- *Describe the current situation from the start of the period under evaluation: what has happened in quantitative and qualitative terms. Consider the monitoring arrangements put in place and use the different indicators to report.*
- *If unexpected or unintended changes have been identified, explain what they are and whether there have been 'knock-ons' in other areas due to this intervention.*
- *Outline any external factors or market developments that have impacted the implementation of the intervention.*

It is not necessary to present in this section all the evidence collected; this can be done in an annex. However, clear references and hyperlinks should be added, signposting where further details/information can be found.

Use tables / graphs / other visual aids as necessary to illustrate the current situation.

Be aware that there is a need to balance the data presented here (and earlier in Section 2) in a descriptive format, with later analysis in Section 4 ('evaluation findings'). Cross-referencing may assist in avoiding unnecessary repetition.

By the time readers reach the end of this section, they should be able to understand what the intervention was expected to achieve and by when; how that was expected to happen (intervention logic); and what the situation is now (factual information). They should also understand the general approach taken for the evaluation and any limits that have been identified, providing them with a context for the subsequent analysis and managing their expectations about the level of detail and sophistication that will be provided.

Section 4 - Evaluation findings (analytical part)

4.1. To what extent was the intervention successful and why? *[Related criteria to assess: effectiveness, efficiency, coherence]*

4.2. How did the EU intervention make a difference and to whom? *[Related criterion to assess: EU added value]*

4.3. Is the intervention still relevant? *[Related criterion to assess: relevance]*

Section 4 deducts a critical, unbiased, and evidence-based judgement on the success (or failure) of the intervention and reports on it. It also describes the critical success (and/or failure) factors, which may feed in the discussion on lessons learned and suggestions for improvement for future policy design.

Section 4 should be analytical, and the narrative should **clearly encompass the five evaluation criteria**. All three questions above should be answered.

The evidence and argumentation leading to an answer should be clearly presented and their association to the evaluation criteria clearly referenced. Text and arguments should be self-standing and accessible to non-expert readers. Reference to the relevant evaluation criterion (or criteria) should be explicit in the narrative.

Section 4 should compare the expected situation (described in Section 2 – ‘*What was expected to be achieved?*’) with the actual situation (described in Section 3 ‘*How has the situation evolved during the evaluation period?*’). The evidence gathered on the questions addressing the ‘better regulation’ criteria⁷⁰⁸ should feed into this section to substantiate the discussion with the necessary evidence base.

Question 4.1 – To what extent was the intervention successful and why? [*Related criteria to assess: effectiveness, efficiency, coherence*]

‘**Success**’ is assessed in terms of the extent to which an intervention achieves its objectives⁷⁰⁹:

- **effectively**;
- **efficiently**; and
- in **coherent** way.

The evidence gathered to answer the questions addressing these three evaluation criteria⁷¹⁰ should feed into the narrative here. Arguments supporting the narrative should be clearly presented and their association to any of the three evaluation criteria clearly referenced.

More detailed analysis by criterion could be documented in Annex III (*Evaluation matrix and answers to the evaluation questions*) to the evaluation report.

Under the effectiveness and efficiency discussion, efforts should be made to address in quantitative terms benefits and costs arising from the intervention.

In case the evaluated intervention concerns an EU initiative that was considered as ‘relevant’ or ‘highly relevant’ for SMEs at the impact assessment phase (see section 2 of the ‘SME test’ - Tool #23) or these impacts emerged afterwards, the evaluation should also cover the effectiveness and efficiency of the intervention for SMEs.

The analysis of efficiency should cover administrative and adjustment costs, and aspects of simplification – these are important for all evaluations. Where appropriate, evaluation findings should pinpoint areas where there is potential to reduce inefficiencies (particularly unnecessary costs) and simplify the intervention by considering e.g. the use of digital solutions.

The proportionality of costs and benefits should be assessed. Actual costs and benefits should be outlined in Annex IV, in the ‘Overview of benefits and costs’ table and, where relevant, the separate table on ‘Simplification and burden reduction’.

Indicative questions for addressing each of these three criteria are provided in Tool #47 (*Evaluation criteria and questions*)

⁷⁰⁸ See Tool #47 for indicative questions by criterion. All relevant analysis and answers to such questions should be documented in Annex IV (‘Evaluation matrix and answers to the evaluation questions by evaluation criterion’)

⁷⁰⁹ As outlined in the intervention’s legal act and/or any prior impact assessment. In the latter case, if co-legislators’ amendments affect the initial objectives presented in the impact assessment, the initial objectives should be updated accordingly.

⁷¹⁰ See Tool #47 (*Evaluation questions and criteria*).

Question 4.2 - How did the EU intervention make a difference and to whom? *[Related criterion to assess: EU added value]*

Answering question 4.2 ('How did the EU intervention make a difference and to whom?') goes beyond re-stating the legal reasons/objectives presented in section 1 ('Purpose and scope of the evaluation') and should strive to present an ex-post evidence-based assessment of the EU intervention in the policy under evaluation. There are two broad approaches on **EU added value**:

- In areas not falling within its exclusive competence, the Union shall only act if the objectives being pursued cannot be sufficiently achieved at Member State policy level (COM(2018)703 final). This implies there is a 'cost of non-Europe' due to the scale, scope, costs, and efforts, organisational or transnational effects of the envisaged action. The evaluation could investigate to what extent the 'cost of non-Europe' can be determined in the intervention area. A counterfactual evaluation methodology⁷¹¹ is one possibility.
- Making an evidence-based case for EU intervention can also be approached by assessing the intervention's 'EU added value'. EU added value is the value that results from an EU intervention which is additional ('above and beyond') to the value that would have been otherwise created by Member States individually. The changes should be reasonably argued, and if possible, factually demonstrated to have arisen from EU intervention. The focus is not so much on the costs (of 'no Europe'), but on the gains (of 'having Europe').

The narrative in this part should **explain clearly to whom the EU intervention made a difference** (e.g. EU people, businesses etc.). Where relevant it should provide an assessment of the EU added value in different territories (rural, urban, cross-border, outermost regions).

Indicative questions for addressing the EU added value are provided in Tool #47 (*Evaluation criteria and questions*)

Question 4.3 - Is the intervention still relevant? *[Related criterion to assess: relevance]*

The evaluation should look at the objectives of the intervention and see how well they reflected and most importantly, they still reflect current and future needs (continuing relevance). This is key information that will help policymakers decide whether to maintain, adjust or terminate the intervention.

The narrative here should highlight any mismatch between the original objectives and current and future needs and problems (also considering elements of foresight) and the reasons for it.

To facilitate answering question 4.3 ('Is the intervention still relevant?') the analysis could draw from the questions provided in Tool #47 under the **relevance** criterion. In addressing the question of continued relevance, it may be useful to look at how megatrends (such as demographic change, digitalisation) may impact the intervention. Information on foresight approach and related instruments can be found in Tool #20 (*Strategic foresight for impact assessments and evaluations*).

Practical tips: Use the information collected to analyse how far the outputs and outcomes observed match the expectations stated when the intervention was adopted, referencing the

⁷¹¹ See Chapter VIII Tool #68 (*Methods for evaluating causal effects*)

intervention logic as appropriate and showing whether the logic has been followed as expected or not. Consider the impact of delays in implementation. Bring together different sources of data (clearly referenced so that the reader can investigate further if they wish) and assess what/how new developments and external factors might influence the future relevance of the intervention. Ensure triangulation of data (i.e. its verification by comparing several sources).

Presentation should be clear, concise, and understandable to the non-expert reader. If there is insufficient data or evidence to answer one or more questions related to the evaluation criteria, this should be clearly stated and linked to the limitations under Section 1 above (brief reference) and to the Annex II on methodology (more comprehensive reference).

Answers to all evaluation questions that were agreed with ISCG should explicitly feed in the relevant parts of this section. The level of analysis should not be compromised; section 4 of the evaluation report should present the findings derived from the questions related to the five criteria⁷¹² and provide appropriate substantiation throughout. The evaluation matrix and, where relevant, more details on answers to evaluation questions by criterion should be presented in Annex III.

Question 4.1 '*To what extent was the intervention successful?*' requires discussion of three evaluation criteria (effectiveness, efficiency, coherence) in one narrative. The building blocks of this narrative could be by criterion, discussing efficiency and effectiveness together (cost-effectiveness), discussing coherence more prominently in the case of a fitness check, or providing a more comprehensive narrative that blends the three criteria. In case a more comprehensive narrative is selected and there is no clear distinction among the three criteria, then the answers to their specific evaluation questions should be provided in Annex III ('*Evaluation matrix and answers to evaluation questions by criterion*'), to provide the reader with all relevant detail on the evidence base informing this narrative.

Section 5 - What are the conclusions and the lessons learned?

An important purpose of evaluations is to enable the organisation to learn and to raise critical policy issues for the attention of the political level. Information reported in the previous sections serves to feed this section to present the conclusions of the evaluation results. Section 5 should focus at least on the following three areas:

Describe what elements of the **EU intervention are working or not** and why.

- Did the intervention achieve its objectives? Did its actual performance match the expectations? Did the intervention generate unintended effects? Were these positive or negative effects, and for whom? Was the intervention more/less costly than expected, and if so, why? Etc.

Summarise what has been **specific about the EU role** in this particular intervention that could not have been achieved without EU involvement.

- Did the EU bring actors together to work on shared solutions that would not have happened otherwise? Did the efficiency of public finances increase because of

⁷¹² See Tool #47 (*Evaluation criteria and questions*)

achieved economies of scale and scope? Did the EU involvement help safeguard key EU interests? Etc.

Present credible **evidence-based lessons** and, where possible, suggested areas for improvement.

- What are the lessons that can be derived from this primarily backward-looking exercise? Do these lessons indicate a continuation of the problems/needs that triggered EU intervention in the past? Do these lessons indicate a lack of coherence that hinders the performance irrespective of a sound logic of intervention? Did the intervention trigger changes in the target group that need to be considered for future policy design? Do these lessons need to be addressed or will resolve over time? Etc.

Given that all evaluations need to investigate how to simplify and cut burdens, clear reference should be made to **lessons relating to (REFIT) issues** such as regulatory or unnecessary burden, simplicity/complexity, identification of efficiencies/inefficiencies, achievement of objectives at low/high (appropriate/reasonable) costs.

The **conclusions** should be written in such a way that policy makers can use them as a basis for future policy development but respecting the limits of a staff working document. For example, the evaluation report **should not make any commitment for future action or direction of action**. It is important to present the lessons learned and include a systematic screening of the evidence, indicating which findings match expectations, which findings are too preliminary to conclude (wait and see) and what does not work.

Finally:

- There should be a clear and logical progression between the description of what has been achieved presented in section 3, the answers to the questions in section 4 and the conclusions being drawn in section 5. No new detail or issue should be presented in the conclusions section – such information should always be presented in the preceding sections first. Related to this, avoid confusion by taking care to use consistent terminology throughout the evaluation report.
- Where relevant, references should be added and/or comments inserted from the stakeholder consultations (public consultations, targeted consultations, workshops with stakeholders, etc.).

Annexes that must be included in the evaluation report

Annexes are used to present additional technical material particularly to support the information presented in the main body of the report (e.g. a more detailed description of the concerned market or monitoring indicators). Annexes should not be excessively long, be restricted to information which is relevant and contain references and hypertext links to external information sources wherever possible (rather than reproducing the material in the report itself).

The following annexes are required:

- Annex I. Procedural information
- Annex II. Methodology and analytical models used

- Annex III. Evaluation matrix and answers to the evaluation questions (by criterion)
- Annex IV. Overview of benefits and costs and, where relevant, table on simplification and burden reduction
- Annex V. Stakeholders consultation - synopsis report

Annex I. Procedural information

- Identify the lead DG; the Decide reference and if relevant the Commission work programme reference;
- Describe any exceptions from the usual procedural requirements of the 'better regulation' guidelines together with an appropriate justification;
- Organisation and timing: provide the general chronology of the evaluation or fitness check and specify which DGs participated in the steering group (ISCG) and how many meetings of the group were held;
- **Consultation of the Regulatory Scrutiny Board (if relevant).** Briefly explain how the Board's recommendations have led to changes compared to the earlier draft. This should be presented in tabular format – the first column identifying the Board's recommendation and the second column how the report has been modified in response;
- Explain which **evidence** has been used in the evaluation or fitness check together with sources and any issues regarding its quality (i.e. has the information been quality assured?);
- **External expertise.** Describe how expert advice has been used in the process, including scientific expertise and/or use of Commission expert groups or expertise from the EU decentralised agencies and other EU bodies. Outline any studies/work carried out by external contractors, with references and internet links.

Annex II. Methodology and analytical models used

- Provide a transparent account of what has been done, by whom (external contractor, Commission), any changes from the original plan (set out in the 'call for evidence') and any mitigating measures taken.
- List any known limitations, e.g. data, timing, etc. and explain the mitigating measures taken. Provide an overall analysis of the reliability of the available data.
- Provide a critical assessment of the work carried out by the external contractor, which allows an understanding why you agreed or disagreed with their conclusions.
- Provide a more elaborate description of the process as well as details relating to the methodologies applied (e.g. studies carried out/used, sources of data, models, stakeholder consultation).
- Explain any uncertainty in the analytical results and the robustness of the results to changes in underlying assumptions or data inputs. It should also be clear how such uncertainty has been addressed or minimised in the analytical work.
- Explain the steps taken to assure the quality of the analytical results presented in the evaluation.

- When the evaluation relies on modelling or other analytical techniques, this Annex should include the following additional information, for any model used:
 - brief description of model structure and modelling approach with any key assumptions, limitations and simplifications;
 - intended field of application and appropriateness for the specific analysis presented;
 - model validation and peer review with relevant references; This may include the extent to which the model/technique and input data / results have been discussed with external experts;
 - citation of input data following good practices for data citation for maximum transparency;
 - explanation of the likely uncertainty in the model results and the likely robustness of model results to changes in underlying assumptions or data inputs;
 - explanation as to how uncertainty has been addressed or minimised in the modelling exercise with respect to the policy conclusions;
 - the steps taken to assure the quality of the modelling results presented in the report.

A concise description of the point(s) of comparison used in any modelling exercise in terms of the key assumptions, key sources of macroeconomic and socio-economic data, the policies and measures they contain and any assumptions about these policies and measures (such as the extent to which they are deemed implemented by the Member States, or their estimated impact following implementation).

Where the point(s) of comparison is not the preferred option identified in a prior impact assessment, the reasons for this should be clearly explained, including any related changes introduced during the adoption process.

Annex III. Evaluation matrix and, where relevant, details on answers to the evaluation questions (by criterion)

The evaluation matrix serves to help organise the evaluation work by:

- translating each of the five 'better regulation' criteria into evaluation questions about the EU intervention under investigation;
- describing per question the data sources that will be consulted to answer the evaluation questions;
- indicating per question the success criteria (i.e. points of comparison) to enable the judgment whether the intervention was successful;
- defining per success criterion the indicator(s) and describe what the indicator(s) will measure (either quantitatively or qualitatively).

The annex should include the evaluation matrix that serves as the organising framework of the evaluation work and factual answers to the questions by evaluation criterion agreed with the ISCG. All the evaluation criteria – effectiveness, efficiency, relevance, coherence, and EU added value – should be addressed in the evaluation matrix, unless a substantiated reason is provided. The analysis and evidence in this Annex provide the main points substantiating the assessment in section 4 – Evaluation findings.

Questions and their respective evidence-based answers could be presented one by one. It is recommended that this is done separately for each evaluation criterion – effectiveness, efficiency, relevance, coherence, and EU added value. **The coverage allocated to each criterion will vary depending on its importance and the depth of the evidence/analysis presented in the main text of the evaluation report.**

For example, in case there is no clear distinction among the criteria of effectiveness, efficiency, coherence when answering Question 4.1 '*To what extent was the intervention successful?*', then the answers to their specific evaluation questions should be provided in this Annex, to provide the reader with all relevant detail on the evidence base informing this narrative.

When external support studies exist, answers to the questions in the evaluation matrix can be concise but factual with reference to these studies for reporting in full. For other cases, answers to the questions in the evaluation matrix must provide sufficient detail, enabling the reader to follow the evidence and logic and understand the answers.

Any differences between the actual evaluation matrix used and the one created at the start of the evaluation (design phase) should be used to inform Annex II (section on limitations).

Annex IV. Overview of benefits and costs and, where relevant, table on simplification and burden reduction

Annex IV is a record of the resources used by an intervention and the changes generated by it (i.e. an assessment of costs and benefits). All costs and benefits that can be linked to the intervention, as identified by the evaluation, should be summarised clearly in a tabular format. An indicative structure for this table is provided below. The structure of the table can be adjusted as the evaluation sees fit, but in all cases, **costs should be classified according to the EU Standard Cost Model**⁷¹³.

⁷¹³ See Tool #58 (*EU Standard Cost Model*)

Table 1. Overview of costs and benefits identified in the evaluation ¹									
	Citizens/Consumers		Businesses		Administrations		[Other...] _ specify		
	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment	
[Cost or Benefit description]:									
Mark the type of cost/benefit, each on a separate line: Costs: Direct compliance costs (adjustment costs, administrative costs, regulatory charges) Enforcement costs: (costs associated with activities linked to the implementation of an initiative such as monitoring, inspections and adjudication/infraction) Indirect costs (indirect compliance costs or other indirect costs such as transaction costs) Benefits: Direct benefits (such as improved well being: changes in pollution levels, safety, health, employment; market efficiency) Indirect benefits (such as wider economic benefits, macroeconomic benefits, social impacts, environmental impacts)	Type: Choose one-off or recurrent	Provide the monetary value	Where no quantification is possible, please provide ranges or explain the reasons why	Provide the monetary value	Where no quantification is possible, please provide ranges or explain the reasons why	Provide the monetary value	Where no quantification is possible, please provide ranges or explain the reasons why	Provide the monetary value	Where no quantification is possible, please provide ranges or explain the reasons why

In case the evaluation has identified (measurable) potential for simplification and burden reduction, this should be summarised in a separate table. An indicative structure for this table is provided below. The structure of the table can be adjusted as the evaluation sees fit, but in all cases, costs should be classified according to the EU Standard Cost Model.

TABLE 2: Simplification and burden reduction (savings already achieved)								
Report any simplification, burden reduction and cost savings achieved already by the intervention evaluated, including the points of comparison/ baseline where available (e.g. REFIT savings predicted in the IA or other sources).								
	Citizens/Consumers/Workers		Businesses		Administrations		[Other...] _ specify	
	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment
Title ¹ :...								
Type: One-off / recurrent (select)	Provide the estimated monetary and quantitative value (point value or range)	If no monetisation/quantification is possible, please explain here the reasons. Qualitative analysis on simplification benefits provides important information and should be inserted here.						
PART II: II Potential simplification and burden reduction (savings)								
Identify further potential simplification and savings that could be achieved with a view to make the initiative more effective and efficient without prejudice to its policy objectives ² .								
	Citizens/Consumers/Workers		Businesses		Administrations		[Other...] _ specify	
	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment	Quantitative	Comment
Description:...								
Type: One-off / recurrent (select)								

Annex IV should be filled in as far as possible, to provide a systematic presentation of the costs and benefits which have been identified and assessed during the evaluation process. As a minimum, Annex IV should include an explanation of the cost/benefit and a qualitative summary of the information collected. As far as available, Annex IV should include both:

quantitative data (e.g. time taken, person days, number of records/equipment/staff etc. affected or involved) and monetisation (in euro) of the costs/benefits.

For all information presented, it should be clear whether it relates to all Member States or is drawn from a subset. Cross-references to the source and assumptions behind any figures should also be provided.

Where there is a prior impact assessment, the table should contain, as a minimum, the same costs/benefit categories identified in it, to enable straightforward comparison between expected (as in the impact assessment) and actual (as in the evaluation) costs and benefits.

Annex V. Stakeholder consultation – synopsis report

This annex summarises all stakeholder consultation activities undertaken for the evaluation or fitness check. The content and format of this annex is outlined in Tool #54 (*Analysing data and informing policymaking*).

The Executive Summary

An evaluation report should be accompanied by an executive summary – a reader-friendly **stand-alone** staff-working **document**. Thus, a non-technical style should be applied, providing the full picture of the evaluation and any technical terminology and jargon should either be adapted or explained for a non-expert reader. The executive summary may follow the structure of the evaluation report.

The executive summary should be short (max 4-5 pages) and be available in English, French and German.

An executive summary is not necessary when the Commission provides a report to the European Parliament and the Council.

4. GOOD PRACTICE TIPS FOR DRAFTING THE EVALUATION REPORT

- Given the importance of providing a good evidence base, all data and analysis should be clearly sourced and where necessary further detail provided in an annex.
- To be credible, evaluations need to state the findings clearly and objectively, i.e. avoid picturing too rosy a picture or not avoid being critical where relevant. The evaluation is a primarily backward-looking exercise. Therefore, findings and conclusions must be phrased so that it is clear what has been achieved and what is lacking. Avoid replacing this backward-looking angle by forward looking recommendation for future inclusion, e.g. recommending a specific improvement when a lack of something has been observed. Care needs to be taken so that the phrasing of conclusions, lessons learned and suggested areas for improvement do not go beyond the limits of an evaluation report.
- The executive summary and the conclusions section of the evaluation report should both contain clear statements on the robustness and reliability of the data and analysis which form the basis of the evaluation, to reflect the common practice of reading either of them first.
- Compare what is being delivered in the final evaluation to what was agreed in the 'call for evidence'. It is easy to promise everything at the start of an evaluation and then find

that it is not possible to deliver. Such limitations or variances from the plan should be clearly written up in the evaluation report.

- In cases where several evaluations of a repetitive nature with very similar content and structure are carried out (e.g. in case of certain funding instruments) it may be possible to cover them in a single evaluation report. This approach would need to be agreed in advance with the SG on a case-by-case basis.

5. THE EVALUATION REPORT AND ANY ASSOCIATED REPORT/COMMUNICATION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL

The evaluation of legislation may be based on a reporting/review/evaluation clause in a legal act which obliges the Commission to review or evaluate the legislation after a certain time and to provide the European Parliament and the Council with a report (or Communication – hereafter, just ‘Commission report’)⁷¹⁴.

In cases where the legislation requires such a Commission report (i.e. one adopted by the College), the evaluation report should be linked to and support the Commission report. This should help keep the main text of the Commission report concise (maximum 10-15 pages is recommended). Where the Commission reports formally to the European Parliament and the Council, it is sufficient to organise one single interservice consultation covering both the Commission report and the evaluation report.

Where there is no legal obligation for the Commission to report formally to the co-legislator, the lead DG can still decide to do so because of the dissemination value and use made by many stakeholders of such documents. It is sufficient to organise one single interservice consultation covering both the Commission report and the evaluation report.

The evaluation report describes the Commission services’ approach, analysis, and conclusions to the evaluation. Such an evaluation report is an analytical document, drawing conclusions underpinned by factual information and analysis.

In contrast to the evaluation report, a formal report to the co-legislator can also set out any political message about the evaluation and indications of next steps. This Commission report can be a shorter, self-standing document, referring to more detailed indications in the evaluation report. It should provide clear indications on all key aspects of the evaluation including, if appropriate, political responses to the conclusions of the evaluation. This Commission report is not submitted to the RSB⁷¹⁵.

6. ‘BACK-TO-BACK’ EVALUATIONS AND IMPACT ASSESSMENTS

For ‘back-to-back’ evaluations and impact assessments, the evaluation findings can be presented as an annex to the impact assessment report, unless the evaluation has been selected for separate scrutiny by the RSB⁷¹⁶.

⁷¹⁴ See Tool #44 (*Legal provisions on monitoring and evaluation*)

⁷¹⁵ The RSB provides an objective assessment of the quality of the work of the Commission services in relation to the evaluation, and hence this is based on the Commission services’ evaluation report or fitness check (in the form of a staff working document).

⁷¹⁶ See Tool #50 (*‘Back-to-back’ evaluations and impact assessments*)

7. COMMUNICATING THE FINDINGS FROM THE EVALUATION REPORT AND THE EVALUATION EXERCISE TO THE PUBLIC

Communicating and promoting the evaluation exercise and its findings to the widest possible audience is of paramount importance.

A dissemination plan could be drawn up from the early stages of the design. It should list the different interested audiences as well as identify the appropriate messages you want to convey to suit what your different audiences would be most interested to know. Commission services are also encouraged to seek guidance from their respective communication units, if applicable.

To maximise transparency and access, all relevant supporting deliverables from any external work and the evaluation report should be widely disseminated. The following files related to the evaluation must be published on the '*Have Your Say*' web portal:

- the 'call for evidence';
- the evaluation report with annexes and (if applicable) its executive summary (in English, French and German);
- the factual summary report of the public consultation (if applicable);
- the opinion(s) of the Regulatory Scrutiny Board (if applicable).

In addition to the publication on '*Have Your Say*', DGs can establish a dedicated webpage for each evaluation or fitness check within their associated general policy pages. Such pages can act as a communication tool, bringing together information and providing an update of progress both during the evaluation and after.

Apart from the publication of the evaluation report, the supporting documents, the related material and of any news items or press releases of the evaluation exercise, DGs should make use of a vast array of communication means to better disseminate the evaluation findings to their different audiences. For example, DGs are encouraged to widely disseminate the evaluation results in the social media and to not rely only on the reports and documents published on '*Have Your Say*' or on their own dedicated webpage.

The terms of reference (or technical specifications), all contractors' final reports (and other relevant interim deliverables from external work) and evaluation reports/fitness checks in the form of staff working documents should be published in a manner compliant with [corporate guidelines](#).

Where contractors are involved, it may be desirable to ask them to provide findings in different formats (e.g. PowerPoint presentations or videos, leaflets, different documents, using social media) to facilitate dissemination of evaluation exercise and/or its findings.

8. FEEDING EVALUATION FINDINGS INTO THE DECISION-MAKING PROCESS

For an evaluation to fulfil its purpose, results must feed into the decision-making process, ensuring that lessons are learned and communicated.

Further to the dissemination of the evaluation report and the evaluation exercise, the evaluation results and lessons learned should feed into the DGs Annual Activity Reports.

- Related follow-up actions (such as decisions to undertake an impact assessment, improve guidance) should be identified in the Annual Management Plans of the Commission services.
- As a part of ensuring transparency and accountability for EU actions, it is recommended that the Directorate-General(s) concerned convene, within a reasonable time of the completion of the evaluation report, upon any follow-up actions with senior management. It is good practice that the unit in charge of the evaluation provides an (annual) progress update against the follow up actions identified. If necessary, the REFIT Scoreboard should be updated.
- Follow-up action plans may be drafted to set out actions planned resulting from the evaluation and their indicative timetable. It is equally important to be clear where no action is envisaged and explain why that is the case. The document should distinguish between policy follow-up (major policy changes requiring Commissioner/College endorsement) and administrative/organisational follow-up. Where it takes the form of a staff working document, no commitments should be taken for the Commission. The degree of commitment presented in the follow-up document will also depend on several factors including the timing of its publication in respect to the Commission's strategic planning and programming cycle (e.g. commitments made in management plans, Commission work programme). It should in no case pre-empt results of a possible impact assessment.

TOOL #50. 'BACK-TO-BACK' EVALUATIONS AND IMPACT ASSESSMENTS

1. INTRODUCTION

Usually, evaluations and impact assessments are conducted sequentially so that the results of the evaluation can be fully used in the subsequent impact assessment. This requires appropriate advance planning and may not always be possible. When necessary, evaluations and impact assessments may be launched at the same time and carried out (in a so-called 'back-to-back' manner) as a single process.

The intention to conduct a 'back-to-back' evaluation / impact assessment must be clearly specified when the initiative is presented for political validation.

2. THE 'BACK-TO-BACK' PROCESS

(1) 'Call for evidence'

The 'call for evidence' will cover elements of the evaluation and of the impact assessment⁷¹⁷. The 'call for evidence' template for 'back-to-back' evaluations / impact assessment is available on [GoPro](#).

(2) Interservice coordination group

A single interservice coordination group (ISCG) should be set up. This is chaired by the Secretariat-General for important or politically sensitive initiatives or by the lead DG or service.

(3) Conduct

The evaluation and impact assessment work will follow the usual steps. Any external work conducted to support the evaluation and/or impact assessment will be steered and discussed with the ISCG.

(4) Stakeholder consultation

A single consultation strategy⁷¹⁸ can be prepared. This strategy should be revised and adapted throughout the process to ensure that the necessary work is conducted to gather information from all the identified stakeholder categories to meet the evaluation and impact assessment requirements.

As usual, the consultation strategy should include a range of appropriate consultation activities. One single public consultation can be used to support the 'back-to-back' evaluation and impact assessment. The public consultation should have a good mix of backward-looking and forward-looking questions that address existing performance and the design of the new initiative.

⁷¹⁷ See Tool #51 (*Consulting stakeholders*)

⁷¹⁸ See Tool #52 (*The consultation strategy*)

Only one synopsis report⁷¹⁹ needs to be drafted as part of the impact assessment report. If the Regulatory Scrutiny Board has selected the evaluation for scrutiny, the same synopsis report should be attached to the evaluation.

(5) 'Back-to-back' reports

For 'back-to-back' evaluations and impact assessments, the results of the evaluation will usually be presented as an annex to the impact assessment report. This annex will follow the full structure of an evaluation report. Compulsory annexes accompanying the evaluation, when overlapping with those of the impact assessment (e.g. procedural information, methodology, stakeholder consultation) as well as the executive summary should be reported once. The introductory part of the evaluation could explain why a 'back-to-back' approach was taken and identify any limits or issues caused by overlaps in conducting the evaluation and impact assessment.

(6) Regulatory Scrutiny Board

For 'back-to-back' evaluations with an impact assessment, in which the evaluation has been selected for separate scrutiny by the RSB, the Board will issue a separate opinion on the evaluation. In such cases, a separate evaluation report is always required as well as a separate slot in the Board's meetings plan. If there are imperative and well-justified timing constraints, the RSB may consider the evaluation report and the impact assessment report at the same meeting but in two separate slots.

For back-to-back evaluations, in which the evaluation has not been selected for separate scrutiny by the RSB, the evaluation will be annexed to the impact assessment report, or, at the discretion of the lead DG, it may be presented as a self-standing report. In either case, the findings of the evaluation will be scrutinised by the RSB as part of its scrutiny of the impact assessment report.

3. POINTS OF ATTENTION

A 'back-to-back' approach allows the evaluation and impact assessment of an EU intervention to be carried out in a single process, quite often with support of a single external study contract, which allows for efficient use of time and resources. Services should however not be tempted to run the two parts in parallel. This would prejudge the outcomes of the evaluation and could lead to low(er) quality of analysis, most often because the problem(s) would be ill-defined.

It is imperative to allow for sufficient time for the evaluation part to thoroughly examine the intervention according to the five evaluation criteria. This should lead to a proper diagnosis of the areas and problems to be tackled in the subsequent revision of the EU intervention. A thorough assessment as part of the evaluation stage leads to conclude on lessons learned that provide a solid basis for the problem definition on which to base the different steps of the impact assessment stage. Therefore, the problem definition in the impact assessment cannot be finalised until at least tentative evaluation conclusions are available.

The handling of the evaluation and impact assessment in one single process, if done properly as explained in the previous paragraph, will enhance the consistency of the two steps and has

⁷¹⁹ See Tool #54 (*Analysing data and informing policymaking*)

the advantage that it is easier to handle external communication on the different steps and the public consultation. Stakeholders and the public will easier understand the whole process and will be able to contribute at the same time on what worked (less) well and where they see room for improvement for the future.