



ANNEX B – THE GOVERNANCE OF POLICY
EVALUATION: AN INTERNATIONAL REVIEW OF
POLICY EVALUATION FRAMEWORKS AND PRACTICES

Rapport van de Sociaal-Economische Raad van Vlaanderen, Wetstraat 34-36, 1040 Brussel

W www.serv.be – T +32 2 209 01 11 – E info@serv.be

Disclaimer:

Dit achtergronddocument werd opgemaakt door het SERV-secretariaat ter ondersteuning van het sociaal-economisch overleg en de beleidsadvisering door de sociale partners in de SERV. De bevindingen, interpretaties en conclusies in dit achtergronddocument vallen volledig onder de verantwoordelijkheid van het SERV-secretariaat en kunnen op geen enkele wijze toegeschreven worden aan de raad, een organisatie vertegenwoordigd in de raad of een lid van de raad.

Kennisname raad 27 juni 2022

Contactpersoon	Peter Van Humbeeck	pvhumbeeck@serv.be	02 209 01 01
	Stijn Wouters	swouters@serv.be	02 209 01 07

Deze Annex B fungeert als bijlage bij deel 4 van het achtergrondrapport van het SERV-Secretariaat bij het advies 'De Vlaamse brede heroverweging als opstap naar een structureel evaluatiebeleid'. Het gaat meer in detail in op de principes, instrumenten en internationale goede praktijken voor de verankering van ex post beleidsevaluatie

Het volledige rapport bestaat uit 5 delen en 3 bijlagen:

1. Beleidsevaluatie: wat en waarom?
2. Beleidsevaluatie in Vlaanderen: beleidsinitiatieven 1999-2021
3. Terugblik op de Vlaamse Brede Heroverweging
4. Principes en internationale goede praktijken voor verankering van ex post beleidsevaluatie
5. Toetsing van Vlaanderen aan een analyse kader voor de verankering van beleidsevaluatie
 - Bijlage A: Spending reviews en beleidsevaluatie in de federale overheid
 - Bijlage B: The governance of policy evaluation: an international review of policy evaluation frameworks and practices
 - Bijlage C: Ex post evaluation and spending reviews: OECD principles, best practices and recommendations

The Governance of Policy Evaluation

An international Review of Policy Evaluation Frameworks and Practices

Peter Van Humbeeck¹

¹ Policy Advisor at Social-Economic Council of Flanders (SERV); Research Fellow at Antwerp University, Government & Law Research Group, Public Administration & Management Research Group; Chairman of the Flemish Evaluation Society (VEP), Vice-president of the Flemish Association for Public Governance and Policy (VVBB).

Introduction

This part presents a review of policy evaluation frameworks and practices in OECD countries as a benchmark for the Flanders Region in Belgium. The focus is on the governance of evaluation (regulatory, legal, and organisational context), its implementation (capacity, transparency, and data availability), and the use of information arising from evaluation-related activities.

We build extensively on work by the OECD and the EC on policy evaluation (cf. list of references) and on discussions with the Policy Monitoring and Evaluation Team of the OECD Directorate for Public Governance (<http://www.oecd.org/gov/policy-monitoring-evaluation.htm>) when working on a Cross Country Study on Governance Insights for Policy Evaluation (published September 2021). That study was prepared in the context of a review of Spain, financed by DG REFORM (European Commission), on the availability and use of high-quality performance and evaluation information in the design and implementation of public policies (Producción y uso de evaluaciones para la toma de decisiones: Una comparativa internacional). The information for that review came from interviews with senior stakeholders and secondary research including through the OECD's databases.

The review of policy evaluation frameworks and practices in OECD countries in this report is illustrated with examples from benchmarking countries that were looked into more thoroughly in the cross country study for Spain, notably the United States of America (US), Canada (CAN), the United Kingdom (UK), Germany (GER), and the Netherlands (NLD), however with additional references to several other OECD countries, and in particular to Spain and the recommendations of the OECD for the Spanish autonomous region of Castilla y León.

The structure of this report is as follows:

- We start with a (discussion on the) definition of evaluation, the relevance of evaluation for public policy and the difference with other instruments/practices.
- The second part goes into some key governance questions: What does a supportive policy framework look like? Is a legal basis necessary and what should be in it? What institutions develop and adjust the framework and provide oversight, coordination, guidance, ...? What policies to evaluate and when? What is the role of Parliament and the Supreme Audit Institutions?
- The third part discusses questions around the evaluation process: How to conduct evaluations? Who conducts the evaluations? How to create transparency and involve stakeholders?
- Fourth, we dive into some conditions for robust evaluation systems: How to build capacity? How to ensure quality? How to increase data and information availability? How to promote use?
- We end with some general challenges for promoting a systemic view, building evaluation systems and maintaining long term support for evaluation systems.

Table of contents

- Introduction..... 5**
- Table of contents..... 6**
- Policy evaluation 8**
- Policy evaluation..... 8
- Relevance of evaluation for public policy 11
- Difference with other instruments/practices 12
- Key governance questions..... 14**
- What does a supportive policy framework look like? 14
- Which legal basis?..... 16
- Who develops and adjusts the framework, provides oversight, coordination, guidance, ...? 18
- What role and autonomy for line ministries? 22
- How to organize the governance of ex ante evaluation? 24
- What policies to evaluate when? 26
- What role for Parliament and Supreme Audit Offices? 32
- How to organise the interplay between central, sub- and international governments?..... 36
- The evaluation process..... 39**
- How to conduct evaluations?..... 39
- Who conducts the evaluations? 41
- How to create transparency and involve stakeholders?..... 43
- Conditions for robust evaluation systems 48**
- How to build capacity?..... 48
- How to ensure quality? 54
- How to increase data and information availability? 58
- How to promote impact and use of evaluations? 61
- Building and maintaining support 65**
- How to build commitment and leadership 65
- How to create a supportive community 67
- Annex: Additional Benchmarking Information..... 69**
- Comparison of evaluation frameworks in 6 OECD countries 69
- Ex post* evaluation for primary laws (2021) 72
- Ex post* evaluation for subordinate regulations (2021) 72
- Nature of legal framework for policy evaluation 74
- Features of the framework for policy evaluation..... 75
- Institutions with competences related to policy evaluation across government..... 76

Mandate of Centre of Government for policy evaluation	77
Number of Regulatory oversight bodies in EU Member States.....	78
Guidelines and methods for policy evaluation.....	79
Use of sunseting clauses and automatic evaluation requirements in EU Member States.....	83
Mechanisms in EU Member States to assess the achievement of underlying policy goals	84
Mechanisms for the promotion of quality.....	85
Quality standards included in evaluation guidelines	86
Methods used by countries in the evaluation of government-wide policy priorities	87
National Evaluation Societies in OECD countries	88
Mechanisms to promote the use of evaluations	89
Quality control of ex post evaluation in EU Member States	90
Efforts to assess and report on performance in EU Member States	90
References	91

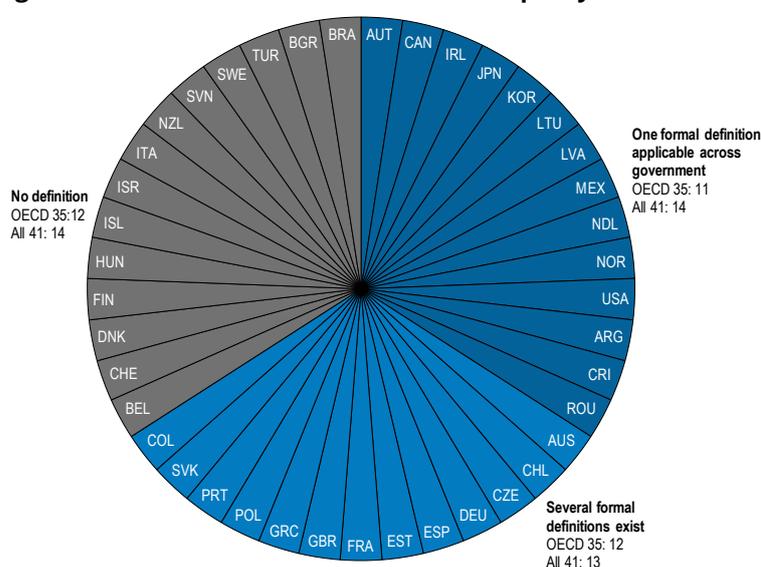
Policy evaluation

In this section we discuss definition of 'policy evaluation', the goals of policy evaluation and the differences with other instruments and practices.

Policy evaluation

Many countries have adopted a formal definition of policy evaluation. In some cases, the definition is embedded in a legal document. Some countries define evaluation in guidelines or manuals.

Figure 0. Government's formal definition of policy evaluation



Note: n=41 (35 OECD member countries). Kazakhstan answered that they do not know if there is a formal definition for policy evaluation. Answers reflect responses to the question, "Does your government have a formal definition of policy evaluation?" and "Please provide the definition/s and the reference to the relevant documents".
Source: OECD Survey on Policy Evaluation (2018).

Key conceptual terms in definitions fall across three categories:

- (1) type of public interventions evaluated, such as policies, regulations, programmes, interventions,
- (2) criteria for evaluation, such as appropriateness (relevance, rationale, policy logic), effectiveness (achievement of objectives), efficiency (cost and benefits, unnecessary costs), impacts (effects, directly and indirectly, intended and unintended, primary and long term secondary)
- (3) characteristics, such as quality attributes (systematic, objective, rigorous), time (ex-post, ex-ante), process (proportionality, evidence, user needs, alternatives, methods, ...) and actors (internal vs. external evaluation).

Often, also the goal of evaluation is an important characteristic.

The definition of evaluation developed by the United Nations Evaluation Group is that evaluation is *"an assessment, as systematic and impartial as possible, of an activity, project programme, strategy, policy, topic, theme, sector, operational area, or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability"*. The definition of evaluation further states that evaluation *"should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making processes of the organizations and stakeholders"*.

The **OECD's definition of policy evaluation** is: *"the structured and objective assessment of an ongoing or completed policy or reform initiative, its design, implementation and results. Its aim is to determine the relevance and fulfilment of objectives, efficiency, effectiveness, impact and sustainability as well as the worth or significance of a policy, with the purpose of enhancing the quality of decision-making and providing tailored advice to improve policy formulation and implementation"*.

The OECD Best Practice Principles for reviewing the stock of regulation mentions that (ex post) review of regulations should include *"an evidence-based assessment of the actual outcomes from regulations against their rationales and objectives, note any lessons and make recommendations to address any performance deficiencies"*.

Although the different elements from the OECD definition can be found in the **national definitions**, there are some differences in how countries describe and interpret evaluations.

In **CAN**, evaluation is defined as the systematic and neutral collection and analysis of evidence to judge merit, worth or value, which typically focus on programs, policies and priorities and examine questions related to relevance, effectiveness and efficiency. Depending on user needs, however, evaluations can also address other themes and issues, but in any case their purpose is to inform decision-making, improvements, innovation and accountability.

In **GER**, evaluation is defined as a procedure which compares the originally formulated expectations regarding objectives, benefit and costs with the actual effects, unintended consequences and costs. Evaluation is intended to reveal whether the new regulation has proven its worth or which corrective actions, if any, are required. It is based on a systematic methodology, a process that can easily be understood by third parties, and precise empirical data sets.

In the **NLD**, evaluations are examinations of the effectiveness and efficiency of policies.

In the **UK**, evaluation is defined as a systematic assessment of an intervention's design, implementation and outcomes, where an intervention could be either a policy, a programme or any other government activity meant to obtain a change. According to the Magenta Book (cf. infra) 'Policy evaluation uses a range of research methods to systematically investigate the effectiveness of policy interventions, implementation and processes, and to determine their merit, worth, or value in terms of improving the social and economic conditions of different stakeholders'.

In the **US**, evaluation is an individual, systematic assessment of one or more programs, policies and organisations, using systematic data collection and analysis and intended to assess their effectiveness and efficacy. Evaluations may address questions related to the implementation of a program, the effectiveness of program strategies, or factors that relate to variability in effectiveness of the program or strategies.

When **comparing these definitions**, a number of issues stand out.

- The OECD definition takes a broad view on the **type** of evaluations: It is about the design, implementation and/or results of a future, ongoing or completed policy initiative. It is about ex ante evaluation or (impact) assessment of a proposed policy and ex post evaluation of a completed policy.
- The UN's and also OECD's definition of evaluation is also broader on the **criteria** (the goal of an evaluation could be to assess the efficiency, effectiveness, impact or sustainability of a given policy) - and therefore better - than more narrow definitions in several countries that focus on one or more specific aspects (such as **NLD**: only efficiency and effectiveness).
- The UN is very clear on the **goal** of evaluation. It stresses that evaluation is not about fault-finding or judging an individual or a team. Rather, evaluation is an opportunity for internal and external stakeholders to contribute their knowledge and views about a particular intervention. At the end of the process, evaluation provides feedback by recognizing achievements that have been made. Furthermore, it identifies ways for improvement and supports evidence-based decision-making.
- A discussion might be to what extent an evaluation has to note lessons and make policy **recommendations**. Many observers believe a policy evaluation should draw lessons and contain recommendations for improvement. An evaluation is of less value if it does not either affirm that a policy or regulation is performing well and needs no change, or identify changes that would improve its performance. Therefore, often, evaluations are required to make recommendations about any changes considered beneficial. Both in **CAN** and **GER** the definition explicitly refers to the policy cycle and the fact that evaluations are linked to the decision-making process and thus policy recommendations. The OECD Draft Best Practices for Spending Reviews explicitly mention that spending reviews should conclude with clear recommendations.
- An evaluation cannot be fully '**objective**'. What is important, is that standards are met (sound, methodology, ethics, ...). A more suitable term would be neutral (cf. **CAN**). An evaluation should be conducted from an impartial position, without any personal or political factor influencing (from a researcher or policymaker) the research design and its implementation.
- Sometimes the primary **target group** is mentioned explicitly ('policymakers'). Evaluation is indeed connecting evidence and decision-making. But evaluation is also connecting 'outside' stakeholders with decision-making. Therefore, the target group or customers of evaluations are not merely policymakers, but a large variety of stakeholders. The **UK** definition of evaluation (cf. supra) explicitly adds 'in terms of improving the social and economic conditions of different stakeholders'. Transparency and stakeholder involvement indeed are an important feature of an evaluation system, especially in an era of co-production where the public sector does not detain a monopoly on policy implementation or even policy formulation and results (e.g. 'right to challenge'). As will appear further in this report, the input from non-governmental actors are an important element of a country's evaluation system.
- To complete this discussion, the practice of evaluation is not always 'structured' or 'systematic', nor fully compatible with methodological standards. Especially in smaller entities with less resources and capacity, such as subnational and local governments, policy evaluation includes all initiatives and instruments that can provide information for decision

making. Often, some form of evaluation exists, be it through the use of **less formal methods and light tools** that are easier to implement and less resource intensive (e.g. inquiries, round tables, talks, ...). These methods are sometimes more appropriate to local circumstances, for example because there is more 'proximity' of stakeholders compared to central government. This evaluation practice often remains under the radar (no outsourcing, no separate reports, internal use, ...).

Relevance of evaluation for public policy

Policy evaluation has a critical role to play in effectively design, implement and deliver public policies and services. While economic, social and environmental challenges are increasingly complex, governments are facing growing pressure to deliver more and better for less. Ensuring that policy making is informed by sound evidence on what works, why, for whom, and under what circumstances is essential to achieve key long-term objectives.

Ex post evaluations complete the regulatory cycle that begins with ex ante assessment of proposals. However not all regulations will have been rigorously assessed ex ante, and even where they have, not all effects can be known with certainty in advance and the original environment justifying the regulation may have changed. Policies should therefore be periodically reviewed to see how they have actually worked in practice. Evaluations are also feeding into ex ante assessments to ensure that laws are relevant and fit for purpose.

In some circumstances, the formal processes of ex post impact analysis may be more effective than ex ante analysis at informing ongoing policy debate. This is likely to be the case if regulations have been developed under pressure to implement a rapid response to emergency needs, such as the COVID-19 pandemic. And to some extent, all laws are experiments. There are often uncertainties about how regulations might actually affect citizens and businesses in practice. Ex post evaluation helps to assess whether laws are working as originally intended and, if not, to propose improvements.

As such, policy evaluation is a crucial element of 'Evidence informed policy making' (EIPM) which can be defined as "a process whereby multiple sources of information, including statistics, data and the best available research evidence and evaluations, are consulted before making a decision to plan, implement, and (where relevant) alter public policies, programmes and deliver quality public services".

Policy evaluation and its strategic use throughout the policy cycle can:

- (1) *provide tailored advice to improve policy and enhance policies' efficiency and effectiveness.* Efficient and effective public policy must be informed by solid evidence. Many external factors influence the attainment of policy objectives, demonstrating a need to periodically undertake checks to establish whether rules are working as intended. Evaluation generates knowledge and therefore is an important source of evidence on the links between policy interventions and their outcomes and impact.
- (2) *enhance accountability and transparency.* It can promote the 'value for money', the up to date-ness, the coherence of policies, trust in public institutions and legitimacy for the use of public funds and resources.
- (3) *promote learning* by a fostering deeper understanding of observed policy success or failure and of what works, why, for whom, and under what circumstances, and at what cost.

Throughout the evaluation process, the people engaged in the evaluation gain relevant insights in the policy that can lead to an improvement, even without a formal policy change, for example, into how changing behaviour more effectively.

- (4) *create opportunities for dialogue*. Evaluations can enhance the democratic process and create the space for stakeholders to reflect on an intervention, whether individually or in groups. By providing a useful platform, evaluations triggers stakeholders to discuss the subject of the evaluation. Evaluation processes can help to build relationships and ensure a better understanding of the different needs and interests of participants and other stakeholders. It may inspire new ideas and energy that can lead to new partnerships or new initiatives.

This however this not mean that the case for policy evaluation is straightforward. In fact, OECD data show that systems for the ex post review of regulations for example remain less developed than for ex ante assessments, with fewer countries having formalised arrangements. The reason is that there will typically be more political or bureaucratic resistance to scrutiny of regulations in place than for those in prospect. This is understandable, in light of the possibility of a review finding that certain regulations introduced previously have been unduly costly or failed to achieve their objectives. Given the weaker incentives for ex post than for ex ante assessments, it is useful to have systems in place to ensure that policy evaluations are conducted.

Difference with other instruments/practices

The focus of this report is mainly on (ex-post) evaluation. Although there are some similarities with other instruments which can reinforce the evaluation framework, it can be useful to highlight differences between evaluations and other tools.

- **Spending reviews** have a specific focus on improving the quality of public expenditure and on proposing reallocations. They are tools for developing, assessing, recommending and adopting policy options by analysing the government's existing expenditure. They link these options to the budget process. The purposes of a spending review include (1) enabling the government to manage the aggregate level of expenditure, (2) aligning expenditure according to the priorities of the government and (3) improving effectiveness within programmes and policies. Evaluations, on the other hand, can have a broader learning function and are not focused on the budget process, as they help to assess performance with regard to the policy objectives that were initially fixed. Evaluation however play a major role in ensuring the quality of public expenditure, by contributing to sound spending reviews or informing performance management. Spending reviews can also lead to broader or more in depth evaluations.
- **Performance monitoring** seeks to check progress against planned targets and can be defined as the formal and ongoing reporting of evidence to show that resources are adequately spent, outputs are successfully delivered and milestones met. As distinguished from monitoring policy evaluation seeks to analyse linkages between policy interventions and effects and include an evidence-based assessment of the actual outcomes against their rationales and objectives.
- **Audit** aims to determine whether the information collected or actual conditions correspond to established criteria, including compliance with financial or legal rules. Auditing helps to ensure that public-sector entities and public servants will perform their functions effectively, efficiently, ethically and in accordance with the applicable laws and regulations. Auditing if

more focused on the internal organisation. In some countries Supreme Audit Institutions conduct also performance audits with an evaluative approach.

- **Reviews** is the term that is often used for initiatives to assess the stock of regulations in the context of regulatory burden reduction programs. There are performed to make sure that regulations are fit for purpose, do not entail excessive costs or administrative burdens and don't have adverse incentive effects on innovation, competition, investment and efficiency. The results from such reviews may help to improve the regulatory environment for businesses and citizens but they could go further by assessing whether the regulation is working as intended. Reviews of existing regulations can also yield useful learnings about ways of improving the design and administration of new regulations, for example, to reduce compliance costs or change behaviour more effectively. In this way, ex post reviews complete the regulatory cycle that begins with ex ante assessment of proposals and proceeds to implementation and administration.

Key governance questions

Embedding policy evaluation in government requires a supportive institutional set-up and infrastructure with a clear and transparent framework for implementation and use. Such a framework has to provide answers to some basic questions we briefly discuss:

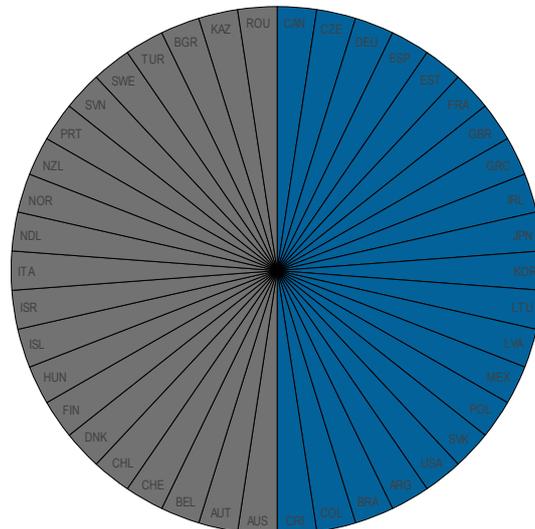
What does a supportive policy framework look like?

A **policy framework** is generally a document or set of documents that provides strategic direction, guiding principles and courses of action to the government for a specific sector or thematic area. Policy frameworks also tend to allocate institutional responsibilities for evaluation.

Features of the framework for policy evaluation can involve i.a.:

- Objectives or expected results of the evaluation policy
- Policy areas or programmes covered by the evaluation policy
- Responsibilities of government institutions concerning policy evaluation
- Requirements for government institutions to undertake regular evaluation of their policies
- Planning of evaluation activities
- Capacity building in evaluation (e.g. training to civil servants).
- Standards for ethical conduct
- Requirements related to the quality standards of evaluations
- Requirements related to stakeholder engagement
- Requirements related to transparency and evaluation reporting
- Requirements related to the use of evaluation findings into policy planning making

Figure 0. Availability of a policy document on policy evaluation across government



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Apart/under the umbrella of a legal framework, has your government developed a policy framework for organising policy evaluation across government?"
Source: OECD Survey on Policy Evaluation (2018).

It should be noted that context matters a lot. There is no single one-size-fits-all approach to the design and adoption of an evaluation framework. Systems building initiatives need to be

aware of the local political and institutional context. This implies obtaining an understanding of the often messy reality of how actual policy making occurs, and how and when to seize the opportunities for evaluation to play a role. Apart from structural components (instruments, procedures, institutions ...) cultural components (values, norms, habits ...) are essential to contribute to a culture of evidence and of policy evaluation activities are planned and carried out well and have an impact on policy practice. Size matters as well. Smaller countries or subnational and local entities often have less resources and capacity. They therefore have less opportunities to invest in (the full implementation of) a sound governance framework and use evaluation policies that are easier to implement and less resource intensive.

A recent OECD review of ex ante evaluation in **NLD** found little clear evidence that the RIA-system (IAK) is having an impact on decision making within government. One observation was that the nature of Dutch Coalition Agreements (Regeerakkoord) makes it difficult to integrate IAK into policy considerations, as specific regulatory instruments are set out in these agreements to address policy issues, meaning that there is little space to consider alternatives. This issue was raised by a number of ministries during the OECD fact-finding mission, as a crucial issue limiting the ability of the IAK to influence decision making. The OECD recommended i.a. that, in the longer term, the Government could explore how to integrate the IAK within the Coalition Agreement process e.g. through establishing goals or objectives within agreements where possible, instead of specific policy instruments, then utilising the IAK to choose the most appropriate instrument.

Furthermore, building an evaluation system requires time and resources, has to cope with challenges and opposition. This means that the quality of the system will have to gain support, grow and improve over time and that the **sequencing of implementing the building blocks** (what first?) is a difficult but important question.

CAN for example evolved from performance management to policy on evaluation (2009) and policy on results (2016).

In the **US** the evolution of the evaluation framework took more than 20-years and the 2018 Foundations for Evidence-Based Policymaking Act has a the 2½ year implementation process.

In **NLD**, despite various structural changes of the system during the past decades, stakeholders continue to see room for improvements and continuous modifications are being made, now to implement the operation 'Insight In Quality'.

A recent review by the OECD of Better Regulation Practices in EU Member States showed that some have had substantive changes to their ex post evaluation systems since 2017, such as the United States, Croatia, Greece, Italy, Latvia, Lithuania, the Netherlands, Portugal and the European Union.

In the **US** President Obama issued in 2011 Executive Order (EO) 13579 to strengthen "regulatory lookback" or "retrospective analysis" to better institutionalise ex post evaluation. This was followed by the US Foundations for Evidence-Based Policymaking Act of 2018-2019. This act establishes a whole set of measures and instruments to organise and reinforce policy evaluation.

Since 2018, **Croatia** requires ex post evaluations on primary laws two years after their enactment and policy makers are mandated to assess whether those laws are meeting their objectives. The Government Legislation Office is the oversight body in charge of reviewing the quality of the ex post evaluations and signing off on the evaluation reports. The Ministry of Economy and Sustainable Development reviews the quality of administrative burden reductions as part of the SME test processes for primary laws and subordinate regulations.

Greece introduced Law 4622 in 2019. Amongst other topics, it made periodic ex post evaluations mandatory for all primary laws and for major subordinate regulations, and it now requires all ex post evaluations to contain an assessment of costs and benefits. Evaluation techniques and oversight functions related to ex post evaluations were also strengthened.

In **Italy**, new non-binding guidance on ex post evaluation was issued in 2018. Initial steps have been taken to plan ex post evaluations when preparing RIAs for major legislation. Ministries publish a two-year plan of regulations to be evaluated.

As part of broader reforms in **Latvia**, ex post evaluations are now required for some subordinate regulations and an evaluation of all policy documents conforming to the SDGs was recently conducted.

Lithuania has introduced some general requirements to conduct monitoring and ex post reviews of existing primary laws and in 2020, it strengthened the regulatory oversight function and transparency of ex post evaluations.

The **Netherlands** saw an improvement in oversight and quality control for periodic ex post evaluation of the effectiveness and efficiency of regulations. The Budget Inspectorate is now responsible for reviewing the quality of ex post evaluations and it has developed a toolbox with guidance for officials conducting these evaluations.

Portugal's main regulatory oversight body was created in 2017 and has taken the role of coordinating ex post evaluations of subordinate regulations across the public administration and assisting officials in conducting them. Following the COVID-19 pandemic, Portugal introduced sunset clauses for some regulations.

The **European Commission** has an institutionalised approach to conducting *ex post* evaluations. Under the "evaluate first principle", the European Commission utilises evaluations to improve the existing regulatory environment and inform impact assessments. The European Commission also in some cases utilises evaluations 'back-to-back' with its impact assessment as part of the same process when it proposes legislative changes. Such an approach provides information and evidence about the existing regulatory environment, the extent to which what has occurred was originally expected, and if not (or if the originally envisaged are not currently being attained), allows the Commission to put forward new regulatory directions in the form of a new impact assessment. The European Union's ex post evaluation system now combines systematic evaluations of individual regulations with comprehensive "fitness checks" of policy sectors, inviting comment on evaluation Calls for evidence. The EU's regulatory oversight body also now provides summary ratings on evaluations that are made publicly available along with compliance statistics.

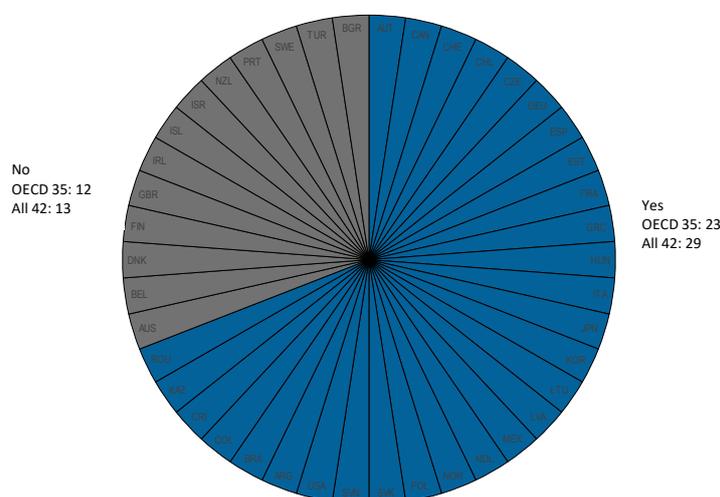
Which legal basis?

A **legal basis** to undertake policy evaluations can be helpful, but is **not necessary nor sufficient** to have a strong framework for policy evaluation nor to ensure the quality of its findings and that the results are used effectively.

The existence of a **sound legal framework can be important** to promote policy evaluation and to clarify institutional responsibilities from a legal perspective. Policy frameworks can include different legislative acts (**NLD, US, ESP**), but this is not necessarily always the case, and some countries internally adopt evaluation policies in the form of guidelines without issuing any specific regulation. Moreover, some countries who lack an overarching legal framework have the more strongest policy frameworks to promote evaluation (**UK, CAN, IRE**) and in **NLD** there is no regulation on spending reviews, but this lack of regulation has clearly has not proven to be detrimental for the conduct of policy evaluations or spending reviews. Also in **GER** there is no

federal regulation establishing an evaluation framework except for from some principles in the Constitution on efficiency and economy.

Availability of a legal framework guiding policy evaluation across government



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Is there a legal framework guiding policy evaluation across government?"

Source: OECD Survey on Policy Evaluation (2018).

Hence, a legal framework guiding policy evaluation across government is **not a condition for success**. It depends on the implementation, context and culture. Furthermore, many challenges cannot be addressed by law: low political interest in policy evaluation, scarce capacity and capabilities for policy evaluation (e.g. lack of time, lack of financial resources, lack of data, ...).

Evidence shows that there can be **flaws in both directions**:

- Without legal backing, fully embedding evaluations inside government activity requires strong anchors and structural integration in public sector processes to avoid that initiatives are washed out. If that is not the case, a legal basis can provide for a more robust environment, more difficult to encompass by government or institutions.
- On the other hand laws or decrees, in and of themselves, do not ensure the effectiveness of a policy evaluation system. In some contexts, rigid institutional set-ups can even have adverse effects. For example, formal measures to undertake and use evaluation can create a fear of sanctions, which can prevent risk-taking, experimentation and innovation. Additionally, an excessively rigorous system may turn public institutions into formalistic bureaucracies.

An important additional question is the **coverage of the legal framework**. It must balance clarity and flexibility. The evidence suggests that existing regulatory frameworks define the basic evaluation obligations of government departments and agencies, and are complemented by more detailed guidelines and planning requirements. Beyond these basic features, the nature and extent of regulation varies.

An interesting example is the **US** Foundations for Evidence-Based Policymaking Act of 2018-2019. The act mandates the Office of Management and Budget (OMB) to develop guidance and advice in

policy evaluation and includes a provision that requires agencies to submit annual evaluation plans, which shall "describe key questions for each significant evaluation study that the agency plans to begin in the next fiscal year" and to report on the results for the previous year. Central oversight is provided by OMB which will publish the various evaluation plans and results. The act also mandates government's agencies to: (1) designate a senior employee as evaluation officer whose responsibility it is to prepare and oversee a four-year evaluation plan and to undertake other tasks such as organising evaluation training for officials; (2) develop capacity assessments which "shall contain an assessment of the coverage, quality, methods, effectiveness, and independence of the statistics, evaluation, research, and analysis efforts of the agency"; (3) implement OMB guidance for programme evaluation; (4) identify "key skills and competencies, establish or update an occupational series, and establish a new career path" on programme evaluation. The 2010 Government Performance and Results Act Modernization Act (GPRAMA) contains a performance framework. There are performance reviews Annual Performance Plans (ex- ante), a Strategic Review Report (ex post), and an Annual Performance Report (ex post).

A question linked to this, is the **appropriate level of the legal framework**. Some countries have specific stipulations in their constitutions (**GER, FRA, MEX** are examples of countries where the constitution is the base for the evaluation framework) while others focus on primary or secondary legislation (with the last one as the most frequent legal basis for institutionalization, and often more detailed containing specific information concerning evaluation annual plans, timing, selection criteria, etc.). Some countries have framed evaluation as part of larger public management laws (**US**), and some have regulated policy evaluation as part of their budgetary governance framework (**GER**). Other countries have issued specific legislations on policy evaluation. Most countries included have for the majority opted to focus on secondary legislation and/or guidelines to specify the framework, allowing them to adapt it more easily. The primary legislation is mainly used to state the principles and importance of evaluations. The **UK** for instance relies heavily on guidelines to organise the evaluation framework.

Who develops and adjusts the framework, provides oversight, coordination, guidance, ...?

The identification of **institutional actors** with allocated resources and mandates to oversee or carry out evaluations is an important issue.

There is **not a single recipe** to institutionalising policy evaluation across government. There are centralized systems with one Government department or agency coordinating evaluation activities across government, other countries have very decentralised approaches (**GER, NLD**). Factors such as the political system, public administration cultures and the rationale for evaluation, shape the development and characteristics of evaluation cultures;

- A *centralised system* is more likely to mean that a common evaluation approach can be promoted across Government with different ministries adopting similar practices with regard to the planning, implementation and use of evaluations. A degree of centralised coordination is also a prerequisite for a more strategic approach to evaluation, i.e. being able to look across a number of evaluations that cover different aspects of the same policy area to arrive at overall conclusions and to identify common learning points.
- A more *decentralised approach* to organising the evaluation function makes it easier for individual ministries to customise evaluation methods and practices to their own specific requirements. However, the drawback is that this can lead to evaluation being neglected or being undertaken in a way that does not benefit from the sharing of experience and know-how. Such a fragmented approach also makes it more difficult to adopt a strategic approach

to evaluation or to aggregate evaluation results across different parts of government that share common policy objectives.

- Adopting a *'hybrid' approach* is another option, i.e. a centralised framework with flexibility for individual Government departments to customise the framework to their specific programmes and priorities.

However, there is a growing need for a **whole-of-government approach**. Because of efficiency, economies of scale and crosscutting, multi-dimensional policy challenges, coherence and policy coordination is required across ministries, agencies, levels of government and policy areas. Evaluation systems benefit from coordination mechanisms to enable greater alignment and sharing of practices across institutions.

The OECD Recommendation of the Council on Regulatory Policy and Governance (2012), the OECD Best Practice Principles for Regulatory Policy (2020) stipulate that there **need to be oversight** and accountability within government administrations to provide ongoing assurance that significant areas of policies and regulation will not be missed and that policy evaluations are conducted appropriately. The recently adopted OECD Recommendation of the Council for Agile Regulatory Governance to Harness Innovation (2021) acknowledges the critical role of oversight in addressing many emerging regulatory challenges.

Although the exact balance between centralisation and autonomy for line ministries depends on the individual country, a degree of **centralised coordination** is a prerequisite for a strategic approach to evaluation. Indeed, a sound evaluation system entails the ability to look across a number of evaluations that cover different aspects of the same policy area to draw overall conclusions and identify common learning points.

This implies establishing **evaluation champions**, that is, institutions whose mandate it is to coordinate policy evaluations across the government. Evaluation champions are indeed needed to provide incentives to line ministries and agencies to conduct the evaluations, to coordinate evaluations across institutions and to promote their quality and use across government. OECD data shows that almost all OECD countries have at least one institution with this overseeing or coordination role.

In practice, the **centre of government** (CoG such as the Presidency, Cabinet Office, Government Office or Prime Minister Office, ...) often is the main actor that provides strategic direction for policy evaluation and plays a crucial role in embedding a whole-of-government approach to policy evaluation (all case studies). (The second actor is the ministry of finance). CoG institutions can facilitate policy evaluation across government, due to the centre's role in steering and coordination. Allocating the role of principal institution in charge of policy evaluation close to political power can be interpreted as a sign of political commitment. In addition, the CoG usually has the political leverage to ensure that the findings of evaluations are subsequently used in forthcoming steps of the policy-cycle. Many governments therefore have strengthened the capacities of their CoG to lead strategic coordination efforts. In fact, governments often have established **dedicated units inside the CoG** to champion an evidence/evaluation agenda throughout government (**UK, US, CAN**).

This central institution has **several key roles** in managing the evaluation eco-system, making sure that evaluation can take place at the right time and in the right place and that it can feed into decision making. They provide oversight and accountability to ensure that key areas of regulation are not missed and evaluations are conducted appropriately. They issue guidelines,

stimulate capacity building, ensure quality of evaluations and manage and oversee an evaluation calendar. They normally do not conduct themselves policy evaluations.

The OECD Indicators of Regulatory Policy and Governance (iREG) survey however shows that as of end 2020, only approximately 30% of the Member States declared to have a mechanism (e.g. a body, unit or network) to co-ordinate ex post evaluation efforts across the public administration from a whole-of-government perspective. The OECD therefore indicated that overall progress in the development of robust and systematic oversight of ex post evaluations remains slow, and that further efforts will be required to **develop oversight**. A recent review of EU Member States showed that the largest improvements have been in oversight and quality control of ex post evaluations, since more oversight bodies now scrutinise ex post evaluations and assist officials in conducting them than in 2017 (OECD, 2022).

Benchmark Country	Main responsible actor	Location within government
United States	Office of Management and Budget (OMB), within the Executive Office of the President.	CoG
The Netherlands	Ministry of Finance. Although this is rather exceptional in the OECD context, there is a similar constellation in Ireland.	MoF
Canada	Treasury Board Secretariat (TBS): the TBS-led process tracks permanent results and reports to support accountability and effectiveness in the medium term. This process aims to clarify the financial aspects and the resources that went into programmes and what was achieved. Privy Council Office (PCO): The Results and Delivery process in the PCO focuses on shorterterm political mandates.	CoG
United Kingdom	Shared central role for both Her Majesty's Treasury (HMT) and the Cabinet Office, but the main role is for the Treasury and the budget process (spending review process) is mainly used as lever.	Both CoG and MoF
Germany	Actors with some responsibilities regarding ex-post evaluations are present both within the MoF and the CoG, but they do not play a prominent central steering role compared to the other countries.	Both CoG and MoF

In the **US** the Office of Management and Budget (OMB) within the centre of government is the main central actor for the evaluation system.

CAN has two the main central actors: the Treasury Board Secretariat (TBS) and the Privy Council Office (PCO). Each has a specific angle. While the TBS led process is designed for permanent results tracking and reporting to support accountability and effectiveness in the medium term, the Results and Delivery process in the PCO was established to focus on shorter term political mandates. The TBS led process focuses on transparency aims to be clear on the financial aspects and the resources that went into programmes and what was achieved.

In the **UK** there is a shared central role for both Her Majesty's Treasury (HMT) and the Cabinet Office, but the mail role is for the HMT since it is mainly the budget process (spending review process) that is used as lever. Within HM Treasury there are 20 spending teams that advise ministries on decisions regarding spending reviews, examine and approve submissions for new spending on projects and programmes and monitor the budgets and spending risks of government departments. HM Treasury also supports government departments in the production of Single

Department Plans (SDPs) but it is mainly the Cabinet Office that oversees the SDP's and supports and monitors delivery of policies within the SDPs. Both the Cabinet Office and HM Treasury develop guidelines related to evaluation and the use of evidence for policy making.

In **ESP**, in the region of **País Vasco** the Directorate of Economy and Planning is responsible for the evaluation and analysis of programmes and public policies developed by the Government, including the following activities: the identification, analysis and selection of the public policies to be evaluated, and the follow-up of the conclusions arising from those evaluations, in coordination with the affected departments; the development and adaptation of quantitative instruments and models to carry out structural analysis, ex-ante simulation and ex-post evaluation of the impact of public policies; the promotion of the culture of evaluation of public policies; the formulation and dissemination of evaluation methodologies. These functions are carried out in coordination with the Department of Public Governance and Self-Government.

Also for **ESP**, in its review of **Castilla y León**, the OECD found that not having a unit responsible for coordinating the implementation of an evaluation system (nor evaluation units within regional line ministries) results in the lack of an evaluation culture and in practically non-existent evaluation practices in Castilla y León. De OECD recommended creating or appointing a unit in charge of coordinating policy evaluation across government. This unit would not carry out evaluations itself but be responsible for elaborating a strategic evaluation agenda, elaborating guidelines for evaluation, designing and offering regular training courses on evaluation issues available for all public servants within Castilla y León and identifying the strategic, crosscutting and/or highly technical evaluations that would require outsourcing and carrying out the outsourcing process.

The main feature of the evaluation and monitoring framework in the **NLD** is that it is a decentralised system where each minister is responsible for the evaluation and monitoring in their remit, with the Minister of Finance responsible for supervision, guidance and coordinating. The Centre of Government has no role in this respect.

In **GER** both within the Ministry of Finance and within the Centre of Government there are actors with some responsibilities regarding ex-post evaluations, but these actors do not play a prominent central steering role such as in the other observed countries. This follows from the so-called 'department principle' (Ressortprinzip), which states that each federal minister manages its area of competence independently and under its own responsibility. As a result, there is no specific governance framework in place and no uniform approach to evaluations. The case of GER seems to illustrate that in the absence of central oversight and coordination it is difficult to build a strong evaluation record. The Court of Auditors regularly signals gaps in terms of methodology, competences and resources for evaluation and the NKR highlighted that evaluations carried out in 2017-18 did not use clear evaluation criteria nor empirical analysis, and did not provide concrete recommendations. Recent evolutions have shown that there was a need to have a certain degree of central steering. In 2013 a framework for ex-post evaluation was established for important legislative initiatives, but did not include requirements. The 2019 update includes requirements for draft bills to facilitate and improve quality of ex-post evaluations. Furthermore the Federal Government is developing a guidance document outlining the steps and methods of an evaluation. And starting from 2015, the federal government added annual spending reviews as a complement to its top-down method of preparing the federal budget. Spending reviews are carried out jointly by the Federal Ministry of Finance and the relevant line ministries. The organisational structure of a spending review rests on two levels: steering committee with deputy ministers and working groups with finance ministry and line ministries' staff.

Austria introduced systematic requirements for both ex ante and ex post assessments, and requires major regulations to be evaluated after five years. The Federal Performance Management Office is responsible for ensuring the quality of both ex ante and ex post assessments.

The Regulatory Scrutiny Board of the **European Commission** conducts reviews of ex ante impact assessments, as well as selected ex post evaluations.

From the examples above, it is clear that **Ministries of finance** also play an important role and have responsibilities in some countries (e.g. **CAN, UK, NLD**). Especially when ex post evaluation is closely linked to spending reviews, the Ministry of Finance performs a fundamental role. It is involved at all stages and is responsible for the link between a spending review and the budget process. At the same time, it is important to build close collaboration with line ministries., both to enhance the capacity for conducting spending reviews and to create trust and a sense of ownership of the spending review process, as well as to benefit from the information and expertise at their disposal.

Other countries have established **independent agencies** that set and coordinate evaluations across government.

Moreover, certain **line ministries** can also play a central role in promoting and/or co-ordinating policy evaluation across government (e.g. the Better Regulation Executive BRE within the Department for Business, Energy and Industrial Strategy in the **UK**).

Outside the executive, **Supreme Audit Institutions** often play an important role.

In any case, there is a need for a **clear-cut allocation** of responsibilities in order to ensure alignment. A specific issue to pay attention to is the division of tasks regarding the different central actors for the different parts of the evaluation framework, which can lead to a fragmented approach.

In **CAN**, the fact that there are two central entities involved creates the need to cooperation. Some steps are being taken to introduce some connection between the activities of both entities. First, under the guidance of the PCO, a new governance body called "Central Agencies and Justice" has been created. This body consists of the Ministry of Finance, the Ministry of Justice (in charge of reviewing the legal aspects of new legal proposals), the PCO, and the TBS. The objective of this new governance body is to share agendas and discuss priorities, thus establishing a starting point for high-level coordination. Secondly, the PCO and TBS are working closely to identify how the tools elaborated by the PCO could be leveraged by the TBS as conditions for funding to Departments to provide incentives that are more powerful. The COVID-crisis has highlighted even more the need to cooperate.

What role and autonomy for line ministries?

The hybrid system implies that there is a certain freedom for the line ministries/departments to organise the evaluation function, both regarding the institutional set-up as the way evaluations are conducted, although this degree of autonomy varies a lot between countries.

This means that often there is **a dual role in the administration**, with some form of coordination from the centre, either through CoG processes, or budget and resources related aspects with the Ministry of Finance, and policy evaluation to be carried out and institutionalised in all the sectoral ministries and agencies.

So, **line ministries** conduct evaluations within framework established at central level, and often they have a large autonomy in order to allow for flexibility to tailor evaluations to the context. Irrespective of the Government evaluation function set-up (centralised or decentralised), there is a strong argument for individual departments and agencies to **have their own evaluation units or officers**. This helps to ensure that evaluation can be customised to the needs of different

policy areas and stakeholders. It can also strengthen the 'ownership' and independence of the evaluation function and improve the communication of results to policymakers. As a result, there often is a **central actor within the (larger) line ministries** as well, responsible for organisation and quality. They moreover often participate in a network of evaluation functions or interdepartmental seminars for mutual learning across ministries.

Examples are **UK, CAN** and **US**, where the centralised coordination role is supported by an obligation on individual departments to appoint evaluation officers and to develop evaluation plans with in-built review mechanisms that are supervised by the coordinating entity. The recent rules introduce a more coordinated and uniform system of the evaluation function across different parts of the federal Government whereas previously evaluation approaches varied and were largely a matter to be decided by individual departments and agencies.

The **UK** offers a clear example of the 'hybrid' approach. UK Government Departments are expected to follow the central guidelines but they have the flexibility to adapt them to their specific needs as long as they comply with the basic principles. Within this overall framework, each Government department has responsibility for developing its own evaluation strategy and evaluating its own policies.

Similarly, in **CAN** Canada's Policy on Results establishes that every large Department should have its own unit for Performance Monitoring and Evaluation and is accountable for the quality of policy evaluation. These units have, on average, a diversified social science background, especially in large Departments. Large Departments must have an evaluation function, led by a Head of Evaluation and overseen by a Performance Measurement and Evaluation Committee (PMEC) chaired by the Deputy Head. On the other hand, line ministries are responsible for establishing a departmental results framework.

In the **US**, The Foundations for Evidence-Based Policymaking Act in the U.S. requires agencies to create three new positions: Chief Evaluation Officer, Statistical Official, and Chief Data Officer. It also requires the creation of a new (or enhancement of an existing) job series in the civil service for policy evaluation. In addition, the Office of Evaluation Sciences (OES) within the General Services Administration works with departments and agencies to translate behavioural insights into concrete recommendations to improve government activities and evaluate their impact. Since 2015, OES completed more than 70 evaluations in collaboration with agencies, and has also developed resources ('Evidence Act Toolkits') to support agencies' learning process and evidence-building activities. Agencies might have their own Manual and Handbook on monitoring and evaluation.

In **ESP**, all departments of the autonomous community of Cataluña are required to have at least one person in charge of coordinating evaluation within the department and acting as interlocutor with other institutions for evaluation issues. The OECD recommended for Castilla y León to ensure that each regional line ministry has a focal person/unit responsible for coordinating evaluation related activities within it. This unit should be in regular contact with the central evaluation management unit to coordinate the strategic evaluation agenda, the commissioning and supervision of external evaluations, the development of training activities, amongst other. The evaluation person/unit within each regional line ministry would be responsible for pushing forward the evaluation agenda within its regional line ministry.

In **GER**, due to the autonomy given to the Ministries to organise their internal evaluation system, the presence of evaluation structures within the public administration is not systematic: while some ministries often rely on research institutions to conduct evaluations on their behalf, other entities have their own specialised units, such as the Federal Office for Economic Affairs and Export Control (BAFA) which has an evaluation unit since 2008, and the federal ministries for Economic Affairs and Energy (BMWi), of Family, Senior Citizens, Women and Youth (BMFSFJ) and of Education and Research (BMBF).

Also in **NLD**, since it is a decentralised framework, the main actors are the line ministries. They are responsible for conducting the ex-ante evaluations, the monitoring and the ex-post evaluations. For the Spending Reviews the Ministry of Finance has the lead, but the line ministries have an important role. There is no specific regulation on the evaluation function within the ministries. It is up to the responsible minister to organise it. But each ministry has to establish a Directorate for Financial and Economic Matters (FEZ) to coordinate all activities related to the budget process and reporting, including the coordination and supervision of the spending reviews. The recent evolutions are an indication that there will be an increase in a centralised approach as a result of the operation insight in Quality, introducing new instruments such as the Strategic Evaluation Agenda and the Public Value Scan.

How to organize the governance of ex ante evaluation?

It should be noted that the mechanics above concerns ex post evaluation.

For ex ante evaluation often the same institutions are at play, albeit different units and with regulatory scrutiny often more established than for ex post evaluation. This follows i.a. from the 2012 OECD Recommendation of the Council on Regulatory Policy and Governance which calls for countries to 'establish mechanisms and institutions to actively provide oversight of regulatory policy procedures and goals, support and implement regulatory policy and thereby foster regulatory quality'. The Recommendation highlights the importance of 'a standing body charged with regulatory oversight (...) established close to the centre of government, to ensure that regulation serves whole-of-government policy' and outlines a range of institutional oversight functions and tasks to promote high quality evidence-based decision making and RIA frameworks and enhance the impact of regulatory policy.

In line with the Recommendation, a working **definition of 'regulatory oversight'** has been employed in the 2018 OECD Regulatory Policy Outlook, which adopts a mix between a functional and an institutional approach. 'Regulatory oversight' is defined as the variety of functions and tasks carried out by bodies/ entities in the executive or at arm's length from the government in order to promote high-quality evidence-based regulatory decision making.

These **functions** can be categorised in four main areas which however do not need to be carried out by a single institution / body (see e.g. **UK**):

- (1) regulatory policy development and systematic improvement: set processes, track success, ensure transparency through public monitoring and reporting, , propose changes to improve the regulatory governance framework;
- (2) co-ordination: promote a whole of government approach and coherence in the administration;
- (3) capacity building: develop guidance, advice and support, provide assistance and training;
- (4) quality control: monitor adequate compliance with guidelines, scrutinise impact assessments, challenge if deemed unsatisfactory.

The OECD moreover has found that internationally, the parts of governments responsible for regulatory oversight are **often situated in the centre of government**. There are a number of advantages for setting up regulatory oversight in the centre as it gives the function a broad overview of departmental activity and therefore might contribute to better consistency of individual policies and regulations issued by ministries and their alignment with government's policy goals.

A clear example is the **US**, where the performance and evaluation framework is complemented by regulatory impact analyses (RIAs) for significant regulatory actions, conducted by agencies and then submitted to the Office of Information and Regulatory Affairs (OIRA), located in the OMB. This Office reviews drafts of proposed and final regulations. In addition to the ex-ante RIA agencies are also required to retrospectively analyse existing regulations.

Another example is **CAN**, where TBS can challenge new proposals from the Departments, but there is no direct challenge mechanism for Departments' evaluations. For ex ante evaluation, the evaluation directorate is not directly involved as the analysis is conducted by the frontline policy units. Recently, there has been an increase in the expectation that proposals go forward based on previous suggestions and analysis. Nevertheless, there are no prescriptive guidelines on that point. Also interesting, a new governance body called "Central Agencies and Justice" has been created to bring to the same table the CoG, the Ministry of Finance and the Ministry of Justice, which is in charge of reviewing the legal aspects of new legal proposals, to share their agendas and discuss priorities.

In the **UK**, it is the Better Regulation Executive (BRE) that leads ex ante evaluation and the regulatory reform agenda across government. BRE is a unit within the Department for Business, Energy and Industrial Strategy (BEIS). This includes issuing guidance on how to operate the Better Regulation Framework (with regulatory impact assessment or RIA as the central tool), ensuring monitoring (e.g. annual report) and providing advice and support to the Better Regulation Units (BRU) in the Departments. A BRU oversees that department's processes for better regulation and advises on how to comply with the requirements. The Regulatory Policy Committee (RPC) is the Government's independent advisory body, sponsored by BEIS, set up to provide scrutiny of the evidence and analysis used by government departments of both their RIAs and post-implementation reviews (PIRs) and to ensure decisions are made on the basis of robust, evidence-based policy making in line with HM Treasury's Green Book and this guidance. The Committee is composed of seven independent experts from a range of backgrounds and are appointed after an open competition. It is supported by a secretariat of civil servants that include economists, policy advisers and operational researchers.

In **GER** Regulatory Impact Assessment is carried out for legislative proposals of the federal government by the relevant line ministries. The Ministry of Interior has a central role in the process, as it has contributed to the drafting of the guidelines (Leitfaden zur Gesetzesfolgenabschätzung) and deals with the methodological guidance of the RIAs (Arbeitshilfe zur Gesetzesfolgenabschätzung).

In **NLD** ex-ante evaluation is mainly performed by the Integral Assessment Framework (IAK or Integraal Afwegingskader) that has been introduced in 2011, following OECD reviews of Dutch regulatory policy, IAK brings together 110 different quality requirements into one integrated administrative framework. There is no regulation for the IAK, but it is an commitment of the Cabinet that all proposals that the Cabinet will present to the Parliament will be subject to an IAK. The responsible Ministry is that of Justice and Safety and (unlike ex post evaluation) not the Ministry of finance. In line with the Dutch Government tradition of ministerial autonomy however, individual ministries are responsible for guaranteeing the quality of their own regulations and for producing an adequate response to the seven IAK questions. There is no part of government with responsibility for providing guidance, advice and capacity building. This responsibility for training provision is largely left to ministries to organise internally and therefore general IAK training and specialised methodological training is not systematically provided across government. A number of informal cross-government officials and ministerial groups do exist, to facilitate the sharing of information and best practice, such as the IAK Working Group (chaired and supported by the Ministry of Justice and Safety) and the network of Regulatory Coordinators (chaired and supported by the Better Regulation unit in the Ministry of Economic Affairs and Climate Policy).

In a recent review (2020) of the Dutch IAK in **NLD**, the OECD pointed out a salient feature of the Netherland's system of government, that is that no one ministry is clearly able to exercise authority over the other and that even the Prime Minister operates with a small number of support staff and refrains from actively steering ministers. The Ministry for Justice and Security moreover has its own policy portfolio on which it needs to gain cross-government agreement (e.g. immigration policy), so their Minister will be unlikely to hold up cabinet colleagues' policies on a regular basis based on insufficient impact assessments, and also may have difficulty with driving large scale reform of the IAK, especially if this would mean imposing stricter rules on his/her colleagues. Importantly, a key recommendation of the OECD to the NLD was to strengthen regulatory oversight and supervision of the IAK to ensure that key roles and responsibilities are clearly understood, and that regulatory management tools are used effectively, for example by setting up a new regulatory oversight body which operates with a certain degree of autonomy from central government (such as in the UK or similar to the EU Commission's Regulatory Scrutiny Board) and installing a Ministerial Committee, supported by a secretariat of officials, to co-ordinate and oversee the IAK policy.

What policies to evaluate when?

There is a big difference between the countries on the **scope** of the mandatory evaluations as well as on the **periodicity** of the evaluations, varying between a strict regulated system encompassing all policies and complete autonomy by the line ministries. Based on recent developments it appears that there is a growing tendency to find an equilibrium between the need to have the relevant information available at the appropriate moment and the autonomy of the line ministries to decide on the way and the moment they will provide this information.

Few countries evaluate all policies, which can be due to an evaluation can be time consuming and entails a costly process. It is therefore important to decide upon a **selection process or criteria** for determining when and what type of evaluation is needed. So questions are: Who decides on what grounds? Who should be involved or consulted? Should we use thresholds (and which)? Rely on formal requirements, such as evaluation clauses in laws? Install a process of planning, resulting in a periodical evaluation planning (coordinated, transparent, participatory)? Ask a specific group to set priorities (cabinet, evaluation commission, parliament, ...)?

Three practices emerge:

- Setting **threshold and proportionality criteria** is something already well embedded in some countries in the Regulatory Impact Assessment process, but often difficult to implement because a lot of information (and discussion) is needed to check thresholds such as (a) wide coverage across the economy or community (b) significant impacts on citizens or organisations (c) *prima facie* evidence of a 'problem'. The result of this 'technocratic' selection moreover might not match with the available resources or priorities put forward by politicians or stakeholders. Additionally, thresholds may turn public institutions into formalistic bureaucracies and 'ticking the box' behaviour. The recent review of Better Regulation Practices in European countries OECD (2022) showed that few EU Member States have thresholds in place to determine whether ex post evaluations should be undertaken.

In **GER** policies that have a considerable financial impact for citizens or the economy (more than € 1 million costs) have to include an 'evaluation clause' which mandates an ex post evaluation within 5 years from the entry into force of the law. Besides this mandatory evaluation, it is up to the ministries to decide on the evaluations they want to conduct.

In the **UK** there are no formal thresholds, but resources and effort employed in monitoring and evaluation should be related to the scale of the proposals under consideration according to the

guidelines of the Treasury. Thus in practice, not all policies are evaluated. Government departments prioritise which policies to evaluate, including: policies that have formal requirements to be evaluated (i.e. evaluation clause introduced in the law), policies considered financially important or particularly emblematic of Government's political line, and policies with potential learning benefits. In other occasions, the evaluation of some policies/programmes is prompted by the NAO or the Parliament reports and recommendations. HM Treasury encourages the evaluation of programmes with big spending decisions (list of priorities).

In **Estonia** a threshold test is used to determine whether ex post evaluations of primary laws should be undertaken. If ex ante impact assessment shows the occurrence of significant impacts, ex post assessment is required by rule and the plan for conducting it must be presented in the explanatory letter of the draft law. If the line ministry is of the view that ex post evaluation is unnecessary, the reasons for such decision have to be shown in the explanatory letter of the draft law. However, the Legislative Quality Division can make suggestions regarding which laws and regulations should be subject to ex post assessment. A new ex post evaluation strategy is currently under development in Estonia, under which legislative proposals introduced in response to an emergency must always be subject to ex post evaluations, since they may undergo a less detailed RIA.

In **Denmark** there are no formalised threshold or other factors used to identify regulations that require an ex post evaluation. Instead, it is at the discretion of Danish Business Regulation Forum (DBRF) to decide which regulations will be assessed ex post. However, the DBRF's decision is based on the regulatory burdens perceived by Danish businesses. Ministerial officials also have the discretion to choose whether to undertake an ex post evaluation and this decision is reportedly based on the political significance of a legislation.

As has already been noted, threshold and proportionality criteria are well embedded in some countries in the *ex ante Regulatory Impact Assessment* process.

The **US** operates a quantitative test to decide to apply RIA for subordinate regulation. Executive Order 12866 requires a full RIA for economically significant regulations. The threshold for "economically significant" regulations (which are a subset of all "significant" regulations) is: "Have an annual effect on the economy of USD 100 million or more or adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or State, local, or tribal governments or communities."

In the **NLD**, for ex ante evaluation, the policy is that the detailed nature of the reply to the IAK questions must be proportional. However, ministries are given no further detailed guidance on how to ensure proportionality in their IAK analyses. It appears that the IAK in the Netherlands is not really targeted towards those policies where the analytical resources are likely to make the greatest positive impact.

CAN employs a Triage System to decide the extent of the analysis. The Triage System underscores the Cabinet Directive on Regulatory Management's principle of proportionality, in order to focus the analysis where it is most needed. The Triage Statement determines whether the proposal will require a full or expedited RIA, based on expected costs and other impacts.

In **AUS** a Preliminary Assessment determines whether a proposal requires a regulation impact statement (RIS or RIA).

MEX has a quantitative test to decide whether to require a RIA, based on a 10 questions checklist.

In the **European Commission**, a qualitative test is employed to decide whether to apply RIA for all types of regulation. Impact assessments are prepared for Commission initiatives expected to have significant economic, social or environmental impacts. The Commission Secretariat general decides

whether or not this threshold is met on the basis of reasoned proposal made by the lead service. Results are published in a roadmap.

In **ESP** Law 50/1997 (partially modified by Law 40/2015, and Royal Decree 931/2017) states that public entities must prepare a report of the regulatory impact analysis of new laws, royal legislative decrees and regulatory standards. This report must contain a series of sections, including: b) Content and legal analysis, d) Economic and budgetary impact, f) Impact due to gender, in childhood and adolescence, and in the family, h) Cost-benefit analysis, and j) ex post evaluation analyses required. In **Cataluña**, the government agreement from 2011 states that all non-regulatory initiatives with a budgetary impact above 10 million euros must have an ex-ante evaluation. However, in some cases, this evaluation is done as a tick the box exercise and information arising from it is not used as input for the design of the new initiatives. Additionally, according to article 64 of Law 26/2010, the management units must prepare a report evaluating the impact of their proposals of new regulatory provisions. This impact evaluation report must contain, as a minimum, the following reports: A budget impact report, an economic and social impact report, a regulatory impact report, and a gender impact report. In its recommendations for the region of **Castilla y León**, the OECD proposed to make ex-ante evaluation and regulatory impact assessments not compulsory for all new policies and legislations, but limiting it to strategically salient policies/legislations by developing This would allow reducing the risk of turning evaluation into a tick the box exercise and would allow regional line ministries focusing time, people and resources on where these would yield the most results.

- Some countries have **formal rules for periodic programmed evaluations**, embedded in general laws or policies reviews, or through the use of **evaluation clauses** (review requirements in the legislative/regulatory framework itself, especially for policies and regulations with potentially important impacts, innovative features or where effectiveness is uncertain) and **sunset clauses** (the automatic lapsing of regulations after a prescribed period unless they have been re-made, normally reserved for subordinate regulations rather than primary legislation, for which the cost and disruption caused by any rules being inadvertently terminated could be high.). However, the use of sunset clauses is slightly more prominent than automatic review clauses across OECD members, although both are usually implemented on an ad hoc basis. Post-implementation reviews within shorter timeframes (1-2 years) are relevant to situations in which either an emergency regulatory measure was deemed necessary, ex ante regulatory assessment was judged inadequate, or a regulation proceeded despite known deficiencies or downside risks.

In the Netherlands (**NLD**) for example, the Comptability Act (Budget Law) obliges ministers to periodically examine the effectiveness and efficiency of their policies. In the Periodic Evaluation Research (RPE) scheme, this provision is further elaborated in substantive and procedural requirements for these mandatory periodic evaluation studies. For example, the RPE specifies that a Policy Review (i.e. a synoptic document: a way to synthesize insights from evaluations at the level of a budget article) should be carried out at least every 7 years, giving a comprehensive assessment of effectiveness and efficiency at the level of budget articles. Policy reviews should be accompanied by an independent expert opinion. After delivery, the reports are sent to the House of Representatives with a cabinet response. The General Administrative Law Act moreover obliges ministers to evaluate every 5 years all subsidies that are based on a legal requirement (unless otherwise specified by law). The RPE also applies to these subsidy evaluations. The Comptability Act also prescribes that law proposals ex ante have to be accompanied by an explanation of the objectives, efficiency and effectiveness of the proposal.

In **CAN**, the framework of the Policy on Results foresees a comprehensive coverage of all programme spending. They should be evaluated approximatively every 5 years (with flexibility). In

addition, the Financial Administrative Act (FAA) stipulates the requirement to evaluate transfer programmes, and grants and contributions programmes every 5 years on a rolling basis.

In the **UK**, a statutory ex post evaluation is required of all subordinate regulations where those regulations affect either business or a voluntary or community body. The report must set out the objectives to be achieved, assess the extent to which they have been achieved, assess whether those objectives remain appropriate, and if they remain appropriate – assess the extent to which they could be achieved in a less onerous manner. The first report is due within five years of the commencement of the subordinate regulation, with subsequent reviews to be conducted within the next five years.

In **ESP**, in the region of **Castilla y León**, primary as well as secondary legislation require public institutions to regularly evaluate their budgetary policies and programmes and to evaluate public policies before, during and after implementation. In practice, in spite of the existence of these laws, evaluations are almost inexistent within regional line ministries, except for certain evaluations corresponding to programmes financed with European Funds and strategic plans.

In **Australia**, there is a general requirement to conduct an ex post review within five years for all new regulations with “a substantial or widespread economic impact”. In addition a post-implementation review (PIR) must be conducted within two years for any regulation introduced, removed or significantly changed without an adequate regulation impact statement, including where the Prime Minister has granted an exemption from regulatory impact statement (RIS) requirements because of exceptional circumstances. The Office of Best Practice Regulation maintains a public register of outstanding PIRs and determines whether agencies are complying with best practice.

Following the COVID-19 pandemic, **POR** introduced sunset clauses for some regulations.

- The third method is to install a **planning process** to define and translate ambitions into strategic and operational evaluation plans or agendas. They not only are a selection tool, but also play a crucial role in coordinating evaluation efforts, promoting quality and use (cf. the importance of the issue of timeliness of evaluation results to promote their use in decision-making, advance notice of forthcoming evaluations can trigger stakeholders to contribute, it can facilitate stakeholder preparation and avoid consultation fatigue etc.) and safeguarding public transparency. Programming of policy evaluations can moreover provide a solution to different problems such as a lack of long-term planning of policy evaluation research, a short-term vision when it comes to the choice of evaluated themes, a shortage of policy evaluations in some areas, insufficient coordination with other policy evaluations, fragmentation of activities and resources, and low predictability.

An interesting example is the **US**, where the Foundations for Evidence-Based Policymaking Act of 2018 includes a provision that requires agencies to submit annual evaluation plans, which describe key questions for each significant evaluation study that an agency plans. There is no strict regulation on the periodicity nor the scope of the evaluation. It is up to the departments to decide on the calendar and the scope, though they have to make sure to conduct evaluations and submit them as part of their Evaluations Plans. The Evidence Act introduced some innovations, such as the Learning Agenda for identifying and addressing priority questions relevant to programs, policies and regulations of an agency. However, these innovations do not engender specific regulations regarding the scope or timing, which are both to be decided by the departments and agencies.

In **CAN** each large Department is required to prepare a five-year, rolling departmental evaluation plan on the principle that all spending should be evaluated periodically and then target evaluations strategically to focus on risks, needs and priorities, providing a rationale for what will not be evaluated. These plans identify the mandatory and discretionary five-year schedule of evaluations,

evaluations of high-risk and high-priority spending programmes, the lead Department for each evaluation, extent of planned coverage, and extent of organisational spending and programmes that will not be evaluated in the planning period, the key reasons for not evaluating, and the year in which the spending or programme was last evaluated. The agenda is set internally in each Department but every fall there are consultations with both the officers and team members of large Departments and TBS. TBS suggests some aspects to be looked at, and these consultations have an impact on the final choice. Overall, the new Policy on Results allow for more flexibility in selecting evaluations and what to focus on.

Interestingly, in **NLD**, as a result of the recent operation “Insight in Quality” (IIK) the government decided to change course. The disappointing results of the existing framework were i.a. attributed to a lack of flexibility to focus on important strategic topics, the shortage of policy evaluations to draw up a Policy Review and the decoupling of the timing of the Policy Reviews from the political policy agenda: a mandatory evaluation within a clear timeline does not per definition provide the information that was needed by policymakers and the timing was not necessarily in synch with the political agenda. Therefore NLD started with a Strategic Evaluation Agenda (SEA) from the budget of 2021 onwards. Each ministry is obliged to draw up (at least) one annual plan for all types of research and to submit it to the Ministry of Finance. This is included as an appendix to the National Budget. The research is divided into three categories: (1) policy reviews (2) other (evaluation) research into effectiveness and efficiency and (3) other research. The SEA offers the opportunity for the line ministries to organise the evaluation framework on a broad thematic basis (not necessarily limited to a specific policy) and to consider when it is strategically useful. The departmental SEA are sent to the Parliament annually and look forward 3-4 years. They give an overview of the key policy themes and prioritized evaluations (ex-post as well as ex ante and ex-durante).

In **ITA**, ministries publish a two-year plan of regulations to be evaluated.

At the central level in **ESP**, Law 50/1997 (partially modified by Law 40/2015, and Royal Decree 286/2017) establishes that the Regulatory Planning and Evaluation Board must elaborate every year a Regulatory Plan listing all legislative or regulatory initiatives that ministerial departments will submit to the Council of Ministries for approval. As part of the Regulatory Plan, the different ministerial departments must identify those initiatives that must be subject of an ex post analysis. To define which initiatives should be evaluated, ministerial departments must consider the costs or savings these initiatives entail for the administration, the administrative duties imposed on the recipients of the initiative, economic and gender impact, amongst others.

In its recommendations for the region of **Castilla y León**, the OECD proposed to develop a strategic evaluation agenda identifying the policies that need to be evaluated, a timeline specifying when each of these evaluations should be carried out (i.e. during the design, during the implementation, and after the implementations of a policy), and the resources for each evaluation. The central evaluation management unit should elaborate this strategic evaluation agenda based on the inputs provided by the regional line ministries (through its evaluation person/unit) in terms of needs, scope, objectives and calendar. The strategic evaluation agenda could also include inputs provided by other stakeholders such as the Parliament of Castilla y León and the Court of Accounts.

- **Also for ex post evaluations linked to spending reviews, adequate planning is very important.** The spending review timetable should be planned so that findings are available in time for the budget formulation process so governments can compare merits of new spending proposals against changes to the composition of existing spending. A medium-term horizon (3-5 years) can enhance the effectiveness of spending reviews. Compared to a single year, longer time horizons increase the range of options governments can consider. In many instances, the findings from a spending review can only be implemented over the medium-term, such as redesigning the delivery of public services and proposing legislative changes. It

is therefore important that the recommendation of a spending review are included in the government's medium-term expenditure framework.

- Of course, policy evaluations can also be initiated **on an ad hoc basis** in response to an emergent issue or crisis, such as a natural disaster or major public health problem. They can also be established to address a more general theme or concern, such as impediments to competition, or to focus on a particular economic activity or segment of society, such as regional development.

An important feature is **how the planning is conducted**. There are benefits in reviewing policies or regulations as a group, rather than in a piecemeal fashion, where they are interactive or operate jointly to achieve related policy objectives. Care must be taken that the selection process takes this into account. This implies some form of coordination and consultation.

Departments vary in the extent to which they decided and formalised their programme for policy evaluation. Some have a **structured process** for determining the work programme, which include consulting with relevant stakeholders, consideration of the programme for government and considerations of existing and expected future resources. For example, a specialist unit for policy analysis can lead the development of the agenda, conduct a range of consultation meetings with divisions across the Department as well as with other stakeholders. Assistant Secretaries then discusses and prioritise a range of options to take forward. The unit then develops these options into a proposal for the management board to make the final decisions. These arrangements ensure that the scarce resources are being applied in a way that is well prioritised and avoids fragmentation and lack of structural attention due to constantly shifting short-term political demands.

Some countries have a process of **engaging stakeholders** in selecting policies for ex post evaluations or identifying areas that may require reform. Their input helps to identify areas for improvement and is often included in the scope of evaluations.

In the **US**, in the course of the implementation of the 2011 Executive Order (EO) 13579 on regulatory lookback and retrospective analysis, each agency was instructed to draw up design plans for ex post evaluations within 120 days and to involve the public. These plans were then made public with another opportunity for stakeholders to respond to them. A few months later, 580 initiatives were in progress. More structurally, agencies were required to continuously gather ideas from the public about rules and regulations that should be evaluated, to prioritize evaluations that could have a significant impact, and to make the planning for ex post evaluations public.

Sometimes a process of **public 'stocktakes'** can be used to provide for a periodic opportunity for stakeholders to identify current problem areas in specific sectors. 'Stocktakes' can be useful for soliciting public views about problems and priorities for evaluation. They can also be an effective means of identifying cumulative regulatory burdens or detecting adverse interactions across different regulations.

In **CAN** external stakeholders are not consulted in the planning of evaluations. In order to organise the programme of evaluation, the evaluation directorates identify the mandatory requirements for evaluation and then adds the needs identified by the managers. Given the costs of evaluation, there is also a need to identify what can be done by the programme managers and front-line units in terms of "light analysis", while the heavier evaluations are subject to a thorough process and inclusion in the 5-year rolling plan. The agenda is set internally in each Department but every fall there are consultations with both the officers and team members of large Departments and the TBS. The TBS suggest some aspects to be looked at, and these consultations are taken very

seriously. In addition, Chief Science Advisors play a role in Departments' evaluations planning in some departments. Small Departments don't need a formal evaluation plan, but must annually conduct an evaluation planning exercise to identify the Department's five-year evaluation needs by considering information needs to support management decision making, other planned Departmental oversight activities, and joint evaluations with other Departments.

What role for Parliament and Supreme Audit Offices?

Parliament is an important actor. Governments are not able to bind the actions of their successors. For this reason bipartisan support and external institutional actors are important if good practice is to be sustained. Both Parliament and the Supreme Audit Office have a role to play to nurture and maintain a good evaluation system.

Policy evaluation has reached a particularly high degree of **institutionalisation** in the parliaments of **France, Sweden** and **Switzerland**, whose evaluation function is constitutionally mandated. The Swiss Parliament's evaluation system stands out with its wide-reaching rights to obtain information from the executive and related follow-up requirements.

Parliaments can have an impact on the evaluation system in different ways: as initiator of evaluations (often in cooperation with Supreme Audit Office), as user of policy evaluations and as gatekeeper for qualitative evaluation systems.

- First, parliaments sometimes **initiate or produce evaluations**. In many countries there is a cooperation with supreme audit institutions. Parliament sometimes commissions evaluations to Government. For instance, by introducing evaluation clauses into laws.

In **GER**, the Bundestag (the lower House of Parliament) requires annually approximately 80 reports from the Federal Government regarding the evaluation of policies or regulations and measures of administrative action of the government. In some cases, parliament organises evaluations, often through hearings resulting in parliamentary resolutions.

An interesting example of a more structured approach is the system of 'rapports d'information' in the French parliament (**FRA**). The organisation of an evaluation (étude d'évaluation) and the formulation of policy conclusions are entrusted to two representatives ('rapporteurs'), one of which must always be one of the majority and one of the opposition. They always act together (in the determination of the scope, the validation of the data, the presentation of the report and the conclusions, e.g. presentation to the press, ...). A working group is set up for a specific topic (the size is not fixed, in practice it goes from 4 to 15 members), with members of the relevant committees. The rapporteurs are supported by the parliamentary services. In most cases, the evaluation itself is called upon by the Court of Auditors to ensure independence and objectivity in the analysis. The evaluation is carried out in close consultation between the Court of Auditors and the rapporteurs. In addition, the rapporteurs can also take initiatives themselves or afterwards (requesting documents, organising hearings and roundtables, conducting surveys, carrying out place visits, going on a foreign mission...). This, too, happens frequently. The rapporteurs' conclusions are part of the final report and are discussed in parliament (in France in a separate committee). The aim for these conclusions is to carry the consensus of both rapporteurs (in 90% of cases), with exceptionally divided conclusions (10% of the evaluations carried out). The report of that discussion is added to the final report (just like the Report of the Court of Auditors). The final report is then transmitted to the government which is obliged to respond in writing within 3 months of receipt. The government's response is always discussed in parliament, in the presence of the ministers concerned. Six months after the publication of the final report, the rapporteurs must submit a follow-up report to the Committee on the implementation of the conclusions. In summary, this approach in the French Parliament always aims at an in-depth analysis of high quality, within a

deadline, often with a strong involvement of the Court of Auditors and the field, on the basis of a certain pragmatism and customization, within a pluralistic cooperation across party lines and majority-opposition, with a clear responsibility and leadership role for the rapporteur-MPs, with a number of guarantees for implementation in practice, and that within the greatest possible transparency, with good results.

In **CAN** the Parliamentary Budget Office (PBO)'s mandate is to provide independent analysis to Parliament on the budget as well as matters of particular significance relating to the nation's finances or economy; and at the request of a committee or a parliamentarian, to estimate the financial cost of any proposal that relates to matters over which Parliament has jurisdiction. By providing independent and non-partisan financial and economic analysis, the PBO supports Parliament with the goal of raising the quality of parliamentary debate on questions of public money and promoting greater budget transparency and accountability. However, PBO has no specific mandate in terms of evaluation.

In the **US** the legislative branch of government is supported principally by the Government Accountability Office (GAO) and the Congressional Budget Office (CBO). The GAO acts as an independent agency with the mandate to investigate how federal dollars were spent, and provide audit, evaluation and investigative services to Congress. The great majority of GAO's work and analyst staff is focused on evaluations and performance audits, and it prepares some 600 – 700 reports on evaluations each year and makes the information available publicly on its website. GAO evaluations are initiated upon request by congressional committees, to support the oversight of executive branch activities, on the independent initiative of the Comptroller General, and/or as prescribed by law. The CBO is an independent fiscal institution that provides budgetary and economic analyses and reports to Congress to assist the federal budget process. The materials typically include evaluation and performance information. Also at the service of the Congress, the Congressional Research Service (CRS) provides objective and confidential legislative analysis of current policies and the impact of proposed policy alternatives.

- Second, Parliament is a major **user and demander of policy evaluations**. Parliaments have been instrumental in increasing evaluation use by promoting the use of evaluative evidence in the budgetary cycle and by requesting more performance data on government spending. This is done through public hearings and instruments for parliamentary scrutiny of the executive. Moreover, the direct institutional relationship in many countries with **Supreme Audit Offices** of Courts of Audit allows the Parliament to receive the Court's self-initiated evaluations in addition to performance audits, to be used in the political debate. The Parliamentary Office Bodies often facilitate the work of the Parliament by commenting in a structured way the evaluations that are produced by the ministries/departments or by scrutinising the information delivered by the government.

In the **UK** the Parliament has claimed an important role in the field of evaluations. The Parliament checks and challenges the work of the UK Government by questioning government ministries, debating, and investigating through its different committees. Through the Public Accounts Committee the Parliament uses the National Audit Office Value for Money evaluations to hold government officials accountable for the efficiency of public spending. Every year, this Committee conducts sessions in which the NAO presents the results of its reports. Government officials are called to these sessions to explain and answer questions regarding the implementation of their policies and programmes and the efficiency of public spending. These sessions are publicly broadcasted and receive great media coverage. The Public Accounts Committee is made up of a nominated group of Members of Parliaments that seat in the House of Commons, and by convention, its Chair is always a member of the opposition. Other commissions of Parliament can also use the reports carried out by NAO for accountability purposes.

In **CAN** the parliamentary committees may use the evaluation reports to accomplish their mission of making monitoring Government accountable for its activities. The Parliament is also the recipient of all program evaluations produced by Departments as well as the departmental results reports. In addition, parliamentary enquiry committees are charged with special assessments. To facilitate access by individual parliamentarians, Departments and agencies are required to list all evaluations undertaken within their Departmental Performance Reports, and include a list of all planned evaluations in their Report on Plans and Priorities.

In the **US**, over time, Congress, has requested an increasing volume of evaluation and performance information. Departments have responded by making more information available online to prevent material to Congress becoming overwhelming and to make information available on a timely basis.

In **NLD** the parliament is not actively involved in the evaluations, but the legislation stipulates that the government must inform the parliament on both the planned and executed evaluations and spending reviews, and has to provide performance monitoring information that allows the Parliament to exercise the control of the ministers. A policy review is always sent to the House of Representatives, not only when finalised, but also in its conception phase to validate the concept. The Comptability Act also prescribes that law proposals put forward by government for deliberation in parliament have to be accompanied by an explanation of the objectives, efficiency and effectiveness of the proposal. Because parliament is not pleased with the current quality of those explanations, it has decided to ask a committee of experts to propose a tool to check this quality and cooperate with experts. The tool is now being tested in a number of cases.

In **GER**, the Parliament's interest in evaluation has grown over time: committees frequently ask for scientific reports and can request impact assessments of draft bills.

In **ESP** there is little demand from the Parliament for policy evaluation, and, where available, there is also little use of its results. The Parliament has an Economic and Budgetary Analysis Unit and a Budget Office. However, both units are underdeveloped and understaffed (with two economists and one administrative staff, respectively). They are thus not able to carry out evaluation-related activities or actively assist Members of Parliament with their scrutiny and accountability activities. The Court of Accounts plays an important role in conducting performance audits verifying the efficiency and effectiveness of policies and the results of programmes, systems and procedures. However, the Parliament has not actively and systematically used the results of the auditing reports of the Court of Accounts to hold government to account and support its decision-making process.

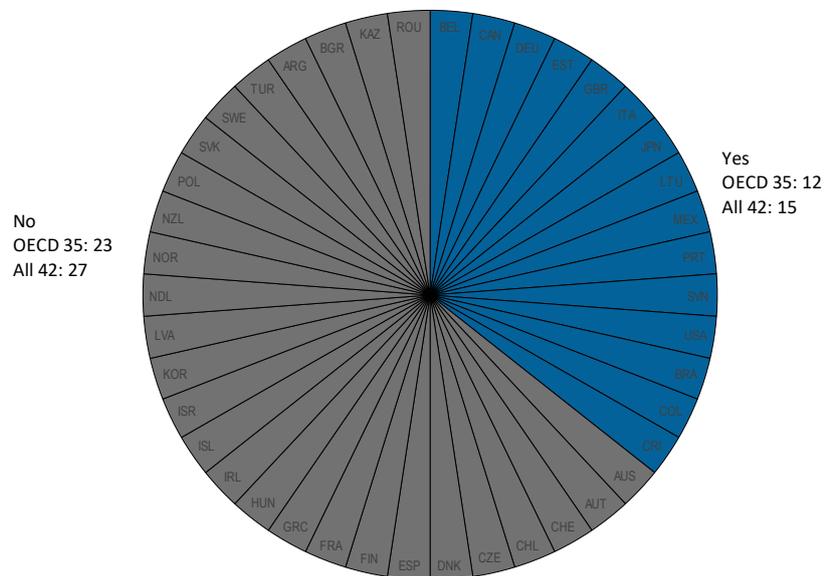
- Third, some Parliaments play a role in regulatory oversight and the scrutiny of RIAs and ex post evaluations prepared by the executive. Concerning RIA, certain parliaments focus on scrutiny of formal and procedural aspects, notably by verifying that the regulatory proposal is accompanied by a complete RIA (e.g. the **Italian** Senate and the **Slovenian** Parliament). In other cases, such as the **European Parliament**, **France** and **Norway** parliamentary bodies conduct more substantial scrutiny, such as in-depth checks on RIA quality.
- Fourth, as Parliaments contribute to ensure accountability, they often play an important role in **promoting a structured or systematic approach** to conducting evaluations. Some parliaments demand governments to submit each year a report on the policy evaluations carried out and on how the results of these evaluation were reflected in policy (either standalone documents or integrated in the budget documents).

In **NLD** for example, this is linked to an annual momentum to discuss and evaluate the results of the government policy ('verantwoordingsdag', each year on the third Wednesday of May). Also the parliament has been an important factor to stimulate the government to systematically improve the framework over the years. The changes regarding the budget framework, introduction of the IAK, and the operation Insight in Quality (IIK) have all been instigated by the Parliament.

The previous decade the **Supreme Audit Institution** of several countries have also conducted a systematic audit of the executive's evaluation system (**GER, UK** and the **US**).

In **BEL** the Court of Audit in 2018 prepared a report on the ability of federal government departments to assess public policies. The analysis showed that the majority of public services has developed evaluation practices. However, the Court underlines the lack of a central vision and strategy on evaluation within the public function. The report recommends fully incorporating evaluation in the policy cycle and budget by delivering yearly evaluation notes to Parliament. The audit also notes the lack of a clear evaluation function and advocates for the need to facilitate data access, develop analytic tools and strengthen quality assurance methods (e.g. making policies evaluable by clearly and explicitly defining the logic of intervention and collecting the necessary data). Finally, the Court characterized the use of evaluation results as insufficient because public services rarely make them publicly accessible.

Figure 1. Audit of the policy evaluation system by Supreme Audit Institutions



Note: n=42. Answers reflect responses to the question “Has your Supreme Audit Institution audited the executive's policy evaluation system in the past ten years”.

Source: OECD Survey on Policy Evaluation (2018)

Despite the fact that the Parliament is in the position to play an important role in this field, and can indeed act accordingly as some country examples show, in general the evidence suggest that discussion of evaluation findings in parliament generates less interest than possibly can be expected. There are **several ways to increase the interest and impact of the Parliament**. Supporting the members of Parliament (MP) in a very focused way is something that has been used in several countries to empower the role of the MP in the field of evaluations. Most parliaments for example have research and information services that help members of parliament order or request evaluation reports. Some conduct evidence syntheses, thus playing a knowledge brokerage function. Another response can be to provide training and expertise for MP's on policy evaluation, or the instalment of a specific committee for overseeing policy evaluation, such as the Public Accounts Committee in the UK. This provides a platform to discuss

work on policy evaluation, following the theme for evaluation and the planning that was agreed upon.

In **FRA**, for example, three days of full discussions were organised in such a committee, which resulted in the adoption of parliamentary resolutions.

In the **NLD** the Analysis and Research Department of the Parliament (Dienst Analyse en Onderzoek or DAO) monitors how the government organises and implements the evaluation function. It counts about 30 persons, half of which are knowledge brokers at the service of the standing committees in the Parliament and the other half are information specialists who assist their colleagues in retrieving the relevant information. When a policy review is transmitted to the Parliament, the DAO is available for the members of Parliament to provide the necessary information. The DAO also performs a more active intervention, providing the Members of Parliament with information that facilitates them to ask relevant questions to the Ministers regarding the policy reviews or ex-ante evaluations. It is up to the parliamentary commissions to decide how they use the information provided by the DAO. The "Insight into Quality" Plan requires the ministry of finance to monitor and ensure the application of the article associated to the revised Budget Law of January 2018 to strengthen the evaluation system. It works on a proposal to this effect, explicitly involving the parliament as recipient of the information in this process.

In the **UK** the Scrutiny Unit is part of the Committee Office in the House of Commons and provides specialist expertise to select committees for the ex-ante evaluation of texts emanating from the government.

In **CAN**, the Parliamentary Budget Office (PBO)'s does not have a specific mandate for evaluation, but by providing independent and non-partisan financial and economic analysis, it supports Parliament with the goal of raising the quality of parliamentary debate on questions of public money and promoting greater budget transparency and accountability. Most of the evaluations, performance information and results data however are not taken up by Parliament. The parliamentary environment is highly political and this leads politicians to dig into evaluations and effectiveness only if the issue raises political concern. The use of evaluations tends to remain sporadic.

In the **US** the MP are supported by the Congressional Budget Office (CBO) and the Congressional Research Service. Over time, Congress has requested an increasing volume of evaluation and performance information.

In **GER** and **BEL** a different approach is followed, namely by providing the MP additional resources to hire people to support them. It is up to the MP on how they want to organise their support.

The OECD recommended **Castilla y León** in **ESP** to raise awareness of Members of Parliament on the benefits of policy evaluations, encourage a more active participation from Parliament in the definition of the annual evaluation plan, support capacity building in Parliament to use evaluations by providing training to Members of Parliament on the use of evaluation at the beginning of each new legislature, or for new Members of Parliament or parliamentary assistants.

How to organise the interplay between central, sub- and international governments?

In the case-countries it seems that there is **not much formalized interaction between the central and subnational governments**. In most cases, subnational governments define when and how to carry out evaluation. There are no requirements or guidelines from the Central government. Each subnational government is responsible for its own approach and prioritization of the use of evidence, evaluation and performance information. Even when the federal

Government transfers funds to the provinces or regions which are then responsible for the delivery, like in **CAN**, there are minimal reporting requirements to the central government. There seems no oversight from the central government. In the **US**, however, the Pew Research Center provides information on the evaluation capacities of US States. Hence, the system and its performance can vary between governments and/or thematic areas, sometimes linked to the size which may influence resources and capacity.

Several countries use soft tools instead of formal coordination mechanisms. For example, in **CAN** the federal-provincial relationship is a negotiated model that relies on voluntary cooperation. The Federal Government organizes policy tables with the provinces to discuss the delivery issues that fall within provincial responsibility. The issues of evaluations are not discussed at the policy table but separately as part of a professional practice. This aims at creating a network of evaluators who share best practices, common knowledge and evaluation results. The network includes evaluators from the majority of Canadian provinces and ensures that evaluations are high quality and satisfy obligations from the federal Government. Both federal and provincial actors benefit from this synergy. The goal is to enable a dialogue based on expertise. Within this network, they agree which data will be collected and receive the data in program operations branch. This enables Departments to carry out rich analysis. Departments then share results from their analyses with the provinces. This creates a virtuous cycle that helps supporting the network. This virtuous information cycle and the shared objective between the federal and the provincial level of improving programs for the citizens ensure the existence and resilience of the evaluation network.

In the **Netherlands**, since 2006, every municipality is required to install either a local policy evaluation function or commission ('rekenkamerfunctie') or an independent local audit office ('rekenkamer'). The design is left to municipalities. They may also set up a common audit office with one or more other municipalities. The local audit offices carry out investigations into the effectiveness, effectiveness and legality of local policies. This research can be ex post and/or ex ante. Each audit offices determines its agenda. Often, an external research partner conducts the evaluations. The research reports (including the findings, conclusions and recommendations contained herein) are made public. Reviews of the system showed that in many municipalities, thorough research is absent and the importance of evaluations is not recognised. Small municipalities with smaller budgets to spend on research regularly use light methods (quick scans, roundtables with citizens, ...) that can also be useful. The central government's subsequent attempts to encourage municipalities to make improvements have had little effect. As long as there is the legal freedom to shape the evaluation function as they see fit, municipalities can invoke this freedom. The Dutch Government therefore decided in 2019 to require an independent audit office and thus to abolish the possibility of a function or commission. In addition, some investigatory powers were extended and the possibility of councilors being advisory members of an audit office was added to increase the involvement of the city council in the audit office. More investment is also being made in training for court researchers, and in improving cooperation and knowledge sharing between court auditors. The government is also promoting the creation of a common audit office for smaller municipalities to pool their budgets, support, knowledge and expertise.

Sometimes, like in the **UK** and **NLD**, there is some level of coordination between **NAO** and the audit offices of the devolved administrations, which allows them to exchange good practices and keep a regular communication, but the NAO cannot carry out value for money evaluations of programmes and policies implemented in the devolved administrations.

Cooperation is more developed or being developed on the subject of **collection of data and access to data**, reflecting a will or need to gather 'whole of government' data, from Federal as

well as subnational governments (**US, NLD**). But overall, data sharing between central and regional governments is a challenge, as shown again by the COVID-19 crisis.

Cooperation is also present in the form of **consultations on federal evaluations** (priorities), where requirements of agencies to consult stakeholders include States or local governments (e.g. US, on the preparation of Learning Agendas).

For the EU the 2020 Indicators of Regulatory Policy and Governance Survey sheds some light on **the interplay of EU Member States and the European Commission**. In theory, there are multiple benefits. The opportunity to learn from evaluations is not limited to the policy ministry conducting the review. Results can be widely applicable.

Yet, currently **sharing results of ex post evaluations beyond an individual jurisdiction is rare**. Despite the fact that around 85% of EU Member States undertook at least one *ex post* evaluation in the past five years, less than 15% shared their results with the European Commission where the evaluation involved areas of EU legislative competencies. Only 4 countries (**DEN, FIN, GER, ITA**) provide the European Commission with the results of their own domestic *ex post* evaluations of EU directives/regulations. 23 countries do not share the results of their *ex post* evaluations of EU directives/regulations with the European Commission.

When the European Commission undertakes *ex post* evaluations, the majority of EU Member States do not use the results to inform either their negotiation position or for the transposition of newly made EU directives.

- There are 11 Member States that report using the results of European Commission's *ex post* evaluations to inform their **national negotiating position** for the development of new or redesigned EU directives/regulations (16 countries answered 'never'). **SLO** is the only country that engages in this practice systematically. For example, the **NLD** incorporates the results of the Commission's *ex post* evaluations in the BNC-fiches that are sent to the parliament to inform the Dutch starting point for negotiations. **Latvia** uses the information gained from the European Commission's *ex post* evaluations in adopting the national positions to explain the aims and shortcomings of the new/redesigned EU legal acts.
- 9 EU Member States reported using the results of the European Commission's evaluations to inform the **transposition** of new or redesigned EU directives (19 countries answered 'never').. **SLO** is the only country that systematically utilises the findings of the European Commission's *ex post* evaluations to inform its transposition of major EU directives/regulations.
- **DEN** uses the findings of the Commission's *ex post* evaluations for its own domestic evaluations.

The evaluation process

The core of an evaluation clearly is the evaluation process. Here, **guidelines** i.a. play an important role to provide answers to the questions how to conduct evaluations, who conducts evaluations, how to involve stakeholders and how to create transparency.

How to conduct evaluations?

Guidelines, toolkits and other supporting documents on evaluation can assist policy makers in conducting policy evaluation successfully. Evidence shows that the majority of countries has guidelines to support the implementation of policy evaluation across government.

Evaluation guidelines can be described as non-binding documents or standards that provide high-level guidance on the implementation and use of evaluations at national level. 27 out of 37 OECD countries have guidelines applicable across government. They may be aimed at evaluators (whether internal or external) to provide guidance on the design of data collection methods, and advice on the identification and design of evaluation approaches; at public sector managers to assist them in better planning, commissioning, managing and following-up on evaluations; or at policy and decision-makers to give information on how to use evaluations and uptake evaluative evidence.

Evaluation guidelines usually focus on the identification and design of evaluation approaches, the design of data collection methods, stakeholder engagement in the evaluation process, the quality standards and independence of evaluations, reporting of evaluation results, and use of evaluation evidence. These guidelines are generally produced by Governments, but also by academia, evaluation societies or international organisations (although these latter's guidelines do not formally apply to evaluations done by the public sector). User groups also vary with some evaluations designed mainly for officials while others have a wider target group that includes evaluation professionals.

CAN presents a significant number of guidelines for the implementation and evaluation of policies. These guidelines are designed to support programme managers and heads of evaluations in their work (Guide to Developing Performance Measurements Strategies, 2010 or the Rapid Impact Evaluation, 2017), but can also be more theoretical, such as the Theory-Based Approaches to Evaluation: Concepts and Practices (2017). In addition to publishing these guidelines online, the Treasury Board actively promotes the use of the guidelines by delivering presentations, newsletters etc..

Also the **UK** provides detailed evaluation guidance (the 'Green Book', 'Magenta Book' and 'Public Value Framework', 'Government Social Research Professional Guidance', 'Quality in Qualitative Evaluation', 'Quality in policy impact evaluation', ...). The Green Book is rather focused on the ex-ante appraisal but also containing info on monitoring and evaluation and the Magenta book provides a comprehensive overview of evaluation in government from its scoping and early design to the use and dissemination of its findings. These manuals are not a mechanical or deterministic decision-making device, they provide thinking models and methods to support the officials doing appraisals and evaluation. These documents also provide information on practical matter such as public procurement procedures for selecting evaluators or the use of evaluation results by policymakers.

In the **US** OMB publishes guidance material on the various phases of the implementation of the Evidence Act, and thus on Learning Agendas, programme evaluation, etc. It recently updated and

consolidated its guidance on programme evaluation standards and practices as part of the implementation of the Foundations for Evidence Based Policy Making Act of 2018. The guidelines are more focused on providing detailed information on how to organise the evaluation work.

In the **NLD** the Ministry of Finance has elaborated a manual for the Policy Reviews (i.e. a synoptic document: a way to synthesize insights from evaluations at the level of a budget article). The policy review should include, at least, the following elements: a demarcation of the policy to be investigated, a justification for the policy pursued and the objectives pursued with it, a description of the policy and substantiation of the expenditure involved in the policy, an analysis and assessment of the effectiveness and efficiency of the policy pursued, including a substantiation of how the effectiveness and efficiency of the policy instruments of the relevant article (sub-chapter) have been or are being evaluated. More specific, a policy review has to provide an answer for 15 specific questions. In 2021 the Ministry of Finance also issued a toolbox in addition to these guidelines, providing information on methods and techniques for evaluation.

In **ESP**, at the central government level, the Institute for the Evaluation of Public Policies updated and published in 2020 four guides to be used by public entities in developing their evaluation-related activities across the different stages of the policy cycle: the Evaluability Guide for Public Interventions, Guide for Evaluating the Design of Public Policy, Guide for Evaluating the Implementation of Public Policy, and Guide for Evaluating the Results of Public Policies. Also País Vasco and Cataluña have developed several guidelines and training material to help public institutions and social entities to carry out the evaluation of public policies.

GER does currently not have general guidelines at the federal level on how to conduct ex-post evaluations or monitor the effectiveness of programmes. Paragraph 7 of the General administrative regulation for the Federal Budget Code (VV-BHO) contains some elements of guidance, but these are relatively limited. Some ministries have published internal guidance documents. The National Regulatory Control Council (NKR) as an independent advisory body, has developed an evaluation model for good evaluation which can be used as a possible basis for evaluation standards by the federal ministries.

Some common types of **evaluation designs** are (1) experimental designs in which intervention and control groups are determined by random assignment; at a minimum, the variable of interest is measured before and after the intervention; (2) quasi-experimental designs that involve either a time comparison (before and after the intervention) or comparison with another group, such as a similar community in a jurisdiction not affected by a policy change. Comparison groups are not determined by random assignment, which makes quasi-experimental designs typically more feasible than experimental designs; (3) non-experimental designs that examine variation without any making comparisons over time or between groups. Examples of non-experimental designs include descriptive, cross-sectional, and case study. Nonexperimental designs tend to rely heavily on qualitative methods. The focus of the design may be to provide an accurate description rather than to prove a specific hypothesis; (4) mixed-methods designs that combine multiple designs or methods to provide multiple perspectives.

Guidelines however have to respect the desired methodological pluralism and allow for the necessary **tailor-made** approach within each individual evaluation. They have to be systematically **updated** to integrate experience and new developments.

The Green Book in the **UK** for instance has been updated regularly since its apparition (last update was end 2020).

In **CAN**, before the Policy on Results of 2016, guidelines were very prescriptive while now Departments are more able to adapt how they evaluate to target exactly the information they need to gather. The TBS's guidance provides today a better balance between prescribing what

Departments need to do while leaving them flexibility to adapt to specific needs. Some programs are evaluated in all their dimensions, while for other programs evaluation is limited by a lack of data or tools. One of the challenges to carry out effective spending reviews include the short time frame for implementation. To solve this and other issues, over time, CAN moved to a more flexible, targeted approach to spending reviews. Each year, reviews involve a small number of Departments and select horizontal spending areas.

Evaluation standards can be part of the evaluation guidelines, but can also exist separately. They set out more general ethical principles applying to evaluation activities rather than detailed methodological guidance. Evaluation standards can also be produced by evaluation professionals themselves. The World Bank's standards, which are built around the three principles of independence (of the evaluation), credibility (of the research) and utility (of the results) are a good example of the sort of evaluation standards that exist. In many countries (**CAN, UK, US, CHE, ...**) evaluation societies have played the key role in developing such standards. The standards emerging from professional organisation, can be complemented with standards emerging from the government (**US**).

The **Canadian** Evaluation Society (CES) e.g. has adopted program evaluation standards and uses these standards to give trainings. These standards identify 5 standards for a good evaluation plan: utility, feasibility, propriety, accuracy and accountability, and thus are not an ethical guide.

The **German** evaluation society has also established standards according to these standards (utility, feasibility, propriety and accountability).

The **American** Evaluation Association has published a guide on Principles for Evaluators. These standards reflect the core values of the American Evaluation Association and are intended as a guide to describe the professional ethical conduct of evaluations. OMB published in the USA Program Evaluation Standards and Practices (M-20-12) to guide agencies in developing and implementing evaluation activities and in hiring and retaining qualified staff. The standards for evidence-building identified and elaborated on in the Program Evaluation Standards and Practices are related to relevance, utility, rigor, independence and objectivity, transparency and ethics.

Who conducts the evaluations?

As described in the previous chapter, line ministries are responsible for the evaluations in the 5 observed countries. However, this does not necessarily imply that officials within the ministries do all evaluations. An important choice to be made is about "who" carries out an evaluation: external evaluations or internal evaluations?

- **External** evaluation refers to an evaluation conducted by entities and/or individuals outside the government. This type of evaluation could be considered more independent, it can take a more critical look at the policy being studied and their results can be potentially more trusted. However, it can also be limited by the knowledge of the evaluator about the context and political process, as well as access to relevant data.
- **Internal** evaluation refers to an evaluation conducted by a governmental institution. Internal evaluators may have more knowledge about a public policies, provide a more accurate assessment and have easier access to inside data. Internal evaluation enhances learning from experience and facilitates putting results to use. But they can also have mixed incentives when it comes to assessing and reporting on how well a policy has performed. In the process of conducting an evaluation, internal evaluators can be under political pressure

and time constraints which can affect the validity of the findings of the evaluation and its public deliberation.

When evaluation is carried out by **external evaluators** (e.g. audit offices, research bodies, management consultants), their autonomy and evaluation skills often improve the reliability and extent of evaluation. However, external evaluators may have limited understanding of the substance and the culture of the evaluated organisation. They may offer standard or unduly theoretical evaluations. Evaluated organisations or the administration in general may be less eager to accept findings and implement possible recommendations.

Overall, **there is no evidence to determine which type of evaluation is better or preferred in all cases**. The decision about whether to do policy analysis in house or externally involves a range of considerations including the skills and expertise available in house, the scale of the project, the timeframe for completion and any budgetary constraints. There is no one size fits all solution. There is a range of risks and issues with conducting work in house and contracting it out.

In **CAN**, evaluation activities are both internally conducted and outsourced to external experts. The choice between the two is ad hoc, it can be decided by the line ministry (due to complexity, limited internal capacity, etc.).

In the **US** on occasion, Departments will use an independent consultant to look at performance or conduct an evaluation, but the majority of these activities are led in-house.

In the **UK** some departments do this using inhouse evaluators while others rely mainly on external contractors. The large variety of organizations engaged in the evaluation of the largest public policy mechanisms can be explained by the search for independent evaluations. Independent research organizations and consulting firms are considered to be more objective than the administration. This outsourcing constitutes for the press and public opinion a guarantee of the quality of the evaluations and of the credibility of their results.

In **GER** there is a common sense and usual practice that evaluations are outsourced to external stakeholders in order to ensure objectivity. Usually, these comprise scientists from universities and research institutes.

In the **NLD** also the majority of evaluations is conducted by external parties, partly because of the lack of internal capacity in most line ministries and partly to increase the objectivity. The same logic is applied in the UK where government departments outsource a considerable part of their evaluation work to external stakeholders as it is seen as guaranteeing evaluation's independence and quality.

In practice, the discussion is often more theoretical than practical, as differences between internal and external evaluations are often blurred because of **hybrid approaches**, mixing internal evaluations and external evaluation for specific or more technical aspects of a policy. Moreover, a government can commission the evaluation to an external organisation, while still ensuring that civil servants control the research questions addressed by the evaluation, or help with the gathering of data. Similarly, an internal evaluation can be accompanied by oversight or a steering group of 'outside' stakeholders and experts. Or evaluations can be performed or supervised by a central agency in government.

A **sound institutional set-up** promotes the principles of transparency and accountability in the management of evaluations. In some countries, there are also rules of conduct and ethical or integrity standards for evaluations that are helpful to avoid capture by policymakers or stakeholders and avoid (perceived) conflicts of interest (cf. supra). In any case, at a minimum, the

test should be that while the reviewer will need relevant knowledge and experience, there should be no conflicts of interest - real or perceived - or reasons for being unduly influenced by different interests.

In **NLD** it is recommended to include an independent evaluator in the process. There are several possibilities to include this independent input in the evaluation exercise, for instance through an advisory body or by involving this person in the actual evaluation exercise. It is also mandatory that, after the evaluation exercise has been concluded, an independent opinion of the policy review is made and sent to the House of Representatives. This can be done by the same person who was involved by the evaluation or by another person.

In **CAN** once the decision to conduct a review is taken, a coordinating committee of central agency senior leadership develops specific terms of reference for the review. Evaluation work undertaken on the response to Covid has been done by evaluation shops to provide speedy advice to Ministers.

In the **US** it was observed that the use of third parties does not necessarily ensure independence or neutrality. As such, the integrity and control processes for evaluations apply to outsourced evaluations as well as evaluations that are completed using in-house resources.

In **Germany** every outsourced evaluation has to be published, contrary to internal evaluations where this is not mandatory. The transparency of the outsourced evaluations can be considered as an external mechanism of checks and balances.

An important issue is to have **in-house capability** in evaluation, both in order to conduct evaluations internally as well as to oversee those commissioned externally. Consultants can usefully supplement the expertise available within government, but how they may best contribute in specific cases needs careful consideration, and they should not be over utilised to the detriment of internal capability. In some countries, every policy analysis project that is commissioned must go through the specialised unit for checks and approval. Another advantage is that the analyst can act as a broker between the policy teams and the external contractor. It also ultimately increases the likelihood of an appropriate and high quality project.

To complete, it should be mentioned that in a number of countries advisory councils, independent research institutes and universities play a significant role as suppliers of policy evaluations; often at their own initiative (notably in **NLD** and **GER**).

How to create transparency and involve stakeholders?

Transparency of the evaluation system, involving stakeholders throughout the evaluation process and public accessibility of evaluations are key elements to ensure quality, gain support and strengthen trust.

According to the OECD Best Practice Principles for Regulatory Policy, evaluations should

- be publicly announced
- allow for stakeholder input
- be made publicly available (the evidence - for critical questioning and public scrutiny – as well as the results, preferably with the government's response).

Tools en practices to promote transparency are

- Publicly available **evaluation agendas** (cf. supra).

Publicly available evaluation agendas are a standard procedure in the **Netherlands**. The agenda of the planned Policy Reviews are transmitted to the Parliament. For the newly installed Strategic

Evaluation Agenda a similar procedure will be installed, allowing the Parliament and the public to know the plans in advance in order to comment on them. There is also a repository of all planned Policy Reviews that can be consulted on the web.

The **US** Program Evaluation Standards and Practices stipulate that Federal evaluation must be transparent in the planning, implementation and reporting phase. The Foundations for Evidence-Based Policymaking Act of 2018 includes a provision that requires agencies to submit annual evaluation plans.

In **Canada** the Departments have to include in the Departmental Performance Reports a list of all planned evaluations in their Report on Plans and Priorities.

- A public **repository with previews** of evaluations (which can be integrated into evaluation agendas). For more significant reviews in areas of regulation with major impacts and/or strong community interest, it can be important that draft reports be prepared for public discussion and feedback, including on preliminary findings and recommendations.

In **NLD** a policy review is always sent to the House of Representatives, not only when finalised, but also in its conception phase to validate the concept.

- **Publication of policy evaluation reports**, together with information on how the results are used for the development of policy

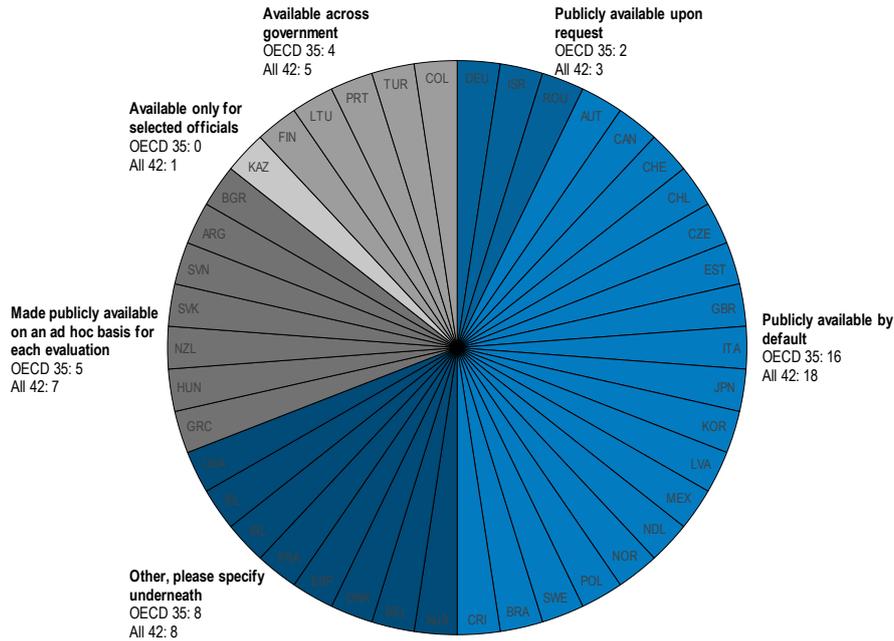
This is the case in the **Netherlands**.

In **Canada** the publication of impact evaluations requested by the administration varies according to who produces them. When the evaluations were contracted with external stakeholders, they are in principle systematically published. Conversely, when the evaluations are carried out directly by the officials of the ministries, their publication is less systematic.

The Magenta Book in the **UK** on evaluations states that there should be a presumption towards openness for all evaluation findings and materials. In practice, the publication of evaluations depends on the level of importance and government commitment of the programme/policy.

In **Germany** external conducted evaluations are published but the internal evaluations not necessarily.

Figure 0 Publicity of evaluation results



Note: n=42. Answers reflect responses to the question “The results of the evaluation are”. In “Other”, the majority of countries agreed that the public availability of evaluation results will depend on the specific agency that commissioned the evaluation, and in its organization.

Source: OECD Survey on Policy Evaluation (2018)

- Government feedback.** Governments are obviously not obliged to accept a review’s recommendations. But when they choose not to do so, the basis for this should be explained (**NLD, CAN, ...**). Lack of information about the outcome of past reviews can reduce the future willingness of stakeholders to participate in the process (devoting time and resources to preparing submissions, responding to data requests, etc.) and may erode public trust in government’s regulatory efforts generally. The government’s decision in response to a final report’s recommendations (whether acceptance or rejection) should be published within a reasonable period and best one that has been designated in advance (e.g. 2-3 months).
- Communication strategies and tools.** Tailored communication and dissemination strategies that increase access to clearly presented research findings are very important for use because, in isolation, publicity alone does not significantly improve uptake of evaluations in policy-making. Rather, the way evidence is presented should be strategic and driven by the evaluation’s purpose and the information needs of intended users. When evaluation results are well synthesised, tailored for specific users and sent directly to them, their use is facilitated. These strategies can include use of standard formats for evaluation summaries, infographics, tailored synthesis of research evidence, for example in the form of policy briefs, dissemination of ‘information nuggets’ through social media, seminars to present research findings, etc.

For example, in **CAN**, departments are diffusing evaluation findings beyond departmental websites via such platforms as Twitter and LinkedIn. Evaluations are also made publicly available on the federal Government website and TBS helps disseminate this information in a user-friendly manner

through the Government of Canada InfoBase system. The transparency of evaluation findings is promoted through the use of infographics in evaluation reports. Progress against delivering commitments is highlighted in an online Mandate Letter Tracker. This platform helps Canadians hold the Government accountable for delivering high level commitments. Departmental issue annual results reports.

- **Evaluation databases and portal sites** to centralise evaluations and evidence in one easily accessible place. Norway and Japan, among others, have launched a web service which provides links to policy evaluation data including analysis sheets and evaluation reports from all ministries. In other countries there seems no evaluation database giving an overview of all conducted evaluation, contrary to information on performance monitoring where several countries have portal sites containing references to this information in a consolidated matter. In the Netherlands the MoF keeps track of all Policy Reviews on a website, but there is no database containing other evaluations.
- **Spaces for dialogue and interaction** between different actors within government and external actors such as civil society organisations).
- Periodic publication of a **report on the status of policy evaluation** conducted by the ministries and how the results of the evaluation have been reflected in policymaking process (often to be sent to Parliament for discussion).

In the **UK**, in November 2017, a report commissioned by HM Treasury and led by Sir Michel Barber, "Delivering better outcomes for citizens: practical steps for unlocking public value" ("Public Value Report"), triggered a renewed attention on outcomes and results.

The Operation "Insight in Quality" that has been conducted the past years in the **Netherlands** was the result of an external report on the quality of budget expenditures.

Stakeholder engagement is an important way to ensure that evaluation improves policymaking. First, and most obviously, stakeholders provide a means of obtaining more complete information about impacts and responses, as well as the opportunity to test preliminary analysis and findings. Second, engaging stakeholders can help identify which areas need evaluation the most with targeting reviews at regulations or regulatory areas that might be problematic. Likewise, mechanisms that enable more continuous engagement with stakeholders can help identify problematic issues in a timely way. Third, in giving the public the opportunity to express views and make an input to proceedings, it can build trust in the evaluation process and even a sense of 'ownership' of the outcomes, making the implementation of any changes politically easier to manage. This is especially important for more sensitive or contentious areas of regulation.

Stakeholders include a **variety of actors** such as citizens, CSOs/NGOs, representatives of academia, representatives of the private sector and international organisations.

Many countries engage stakeholders during the **implementation** of the evaluation (designing evaluations, providing the data relevant to the evaluation, carrying out evaluations) and during the dissemination of the results (discussing the results of the evaluation, communicating the results of the evaluation). Some engage stakeholders also when **deciding what policies** should be evaluated (cf. supra) and in following-up on the **use of results**.

Consultations can be time-consuming and resource intensive. Therefore, the nature and coverage of consultations should be **proportionate** to the significance of the regulations and the degree of public interest or sensitivity entailed.

In the **UK** the departments are expected to consult and engage with stakeholders and the wider public to get a better understanding of the current situation and valuable insights.

In **CAN** stakeholders are not consulted in the planning of evaluations, but can be engaged when conducting evaluations. According to the Policy on Results the head of evaluations has to consider the information needs of the major stakeholders in the Departments five-years evaluation coverage.

In the **NLD** the newly installed Public Value Scan requests the ministry to take care of civil society, by indicating the support of the policy within target groups and stakeholders, forcing them to interact with civil society on these questions.

In **GER** during the ex-ante impact analysis, consultations with other ministries and stakeholders, such as Länder, communes and associations – can be held in the form of workshops, interviews or online consultations. This process can vary depending on the sector under study, the draft proposal and the available resources.

Conditions for robust evaluation systems

Measures to build capacity, ensure quality, increase data availability, promote use and build and maintain (long term) support through committed leadership and evaluation communities are equally important for evaluation systems to thrive, than having a clear legal and institutional framework and guidance.

How to build capacity?

Capacity building and embedding **learning** in the organisational culture is crucial.

The **knowledge, skills and abilities** required to conduct policy evaluation have to be available. Evaluation networks and associations have worked to establish a list of core competencies required to be an evaluator, in an effort to professionalise evaluations. They focus on professional, technical, interpersonal, management and organisational skills necessary to be an evaluator, reflecting the wide variety of competencies such a profession requires beyond technical expertise including 'soft skills'. Indeed, evaluation research often also requires skills related to participation and interaction processes, data analysis, knowledge management, use of language, presentation, empathy, etc. as well as skills to commission and contract services.

Capacity enhancement needs to be pursued through the **training** of existing staff as well as through **recruitment**, with on-the-job learning an important element.

Developing evaluation capacity is equally important for both Government officials and evaluation professionals given the importance of outsourcing of a significant part of the evaluations. Mechanisms for the development of skills and competences should not only aim at **evaluators**, managers, or senior civil servants to carry out high quality evaluations but also at decision-makers to make the best **use** of evaluations and increase demand for evaluative evidence.

Manuals are not enough as there is a wide gap between the knowledge of evaluation methods and the concrete practical performance or steering of policy evaluation research. **Capacity building methods** therefore include a variety of other activities like training courses, networking and the sharing of know-how, hands-on advise (a 'methods advisory person or team' to support evaluators in the field) and accreditation (the **CAN** Evaluation Society offers 'Credentialed Evaluator Designation'), specific requirements or job category for evaluators in the government etc. They can all play an important role in capacity building and, more generally, nurturing an evaluation culture.

- **Inhouse training** by government is common.

In **CAN** policy evaluation capacity is developed and improved mainly through training. The Results Division organizes information and learning sessions (called Drop-in Sessions) based on different results-related subjects of interest to the evaluation and performance measurement communities. They have recently started to record and post these sessions on an unlisted Results Division YouTube channel for their communities to view. The presentations and the links to the recorded videos can be accessed via the Results Portal (i.e., a sharing platform internal to the federal government) and are also mentioned in the Results Newsletters. They also share the presentations of other departments' and non-governmental organizations' workshops via the Results Portal when relevant to the evaluation and performance measurement communities. Guidance is not limited to formal guidelines, as interpretation of policies and guidelines, with concrete examples and case studies, is crucial. This is the reason for active knowledge mobilisation. Guidelines are not just

uploaded on the website, but the TBS delivers presentations, newsletters, etc., trying to communicate to people and disseminate knowledge. All of the approximately 50 analysts working in the TBS' Results Unit are encouraged to participate in training as well as on-the-job learning opportunities.

In the **UK** the central government supports capacity development of its evaluators through various ways, such as training for internal evaluators, establishment and support of a network of evaluators to share good practices, advisory panels for evaluations (Cross-Government Trial Advice Panel). Ministries carry out internal training programmes. For instance, the Ministry of Business, Innovation and Skills offers technical training on scientific assessment methods on an annual basis. Regular training is available on regulatory impact assessments via the Government Economic Service and the RPC.

In **NLD** the ministry of finance facilitates trainings, gives (practical) advice and develops guidelines. Moreover, early 2021 the ministry issued the website 'Toolbox Policy Evaluations' (<https://www.toolboxbeleidsevaluaties.nl/>). It is a dedicated website to help policy makers, evaluation coordinators and policy researchers to set up and execute a good policy, programme or project evaluation. The toolbox divides the evaluation process into five steps, and each step refers to concrete 'tools' (evaluation methods and data collection methods) and examples of evaluations. It also contains background information on policy evaluation in the Netherlands.

Also in the **US** ministries carry out internal training programmes. Within the Department of Health and Human Services for instance there is a specific unit responsible for organising seminars and training sessions to enhance the evaluation capacity within the department. In July 2019, OMB established an inter-departmental council of evaluation officers to support capacity building on evaluation as a cross-governmental function. OMB also provides technical assistance through workshops and seminars to support the implementation of its guidance memoranda.

Officials in **France** have access to a training on ex post evaluation that enables them to get familiarised with the relevant theories and methodologies. In addition, the French government has organised ad hoc training seminars on ex post evaluation of public policies, in partnership with French research institutions.

The training offered by **Austria** is specific to ex post evaluation and to the monitoring tools used for this purpose by the Austrian government. It also covers the evaluation principles as well as information and reporting requirements.

The training programme in **Greece** covers the better regulation framework as a whole, including ex post evaluation. The programme runs over several days and ex post evaluation is an integral component of the training, along with other core regulatory management tools such as stakeholder engagement and RIA.

In **Italy**, the National School of the Administration organises the training course 'How to build RIA and ex post evaluation'. The course aims to update managers and officials involved in the development of RIA and ex post evaluation. It is an operational and practical training course for policy officials, aiming at practicing techniques of consultation, policy option analysis, assessment of impacts. Lessons are rich in interaction on case studies.

Germany is an exception in a sense it is entirely left in the hands of the line ministries to organise training and take other initiatives for capacity building.

- Sometimes **specific coordination bodies or learning mechanisms** such as commissions and integrated services are set up, enable aligning and sharing practices.

In the **US**, an interagency council has been set up that regroups Evaluation Officers and is intended to serve as a forum for exchanging information and advising OMB on issues affecting the

evaluation functions such as the evaluator competencies, best practices for programme evaluation, and evaluation capacity building.

In **CAN** coordination and feedback mechanisms with branches of Government are taking place through both formal and informal structures. The Results Division has recently created two steering committees, one for evaluation and the other for performance measurement. The goal is to discuss challenges and develop concrete actions to address them. They have regular meetings (about 2-3 per year) where all the Heads of Evaluation and Heads of Performance Measurement are invited. For about 3 years, this has been their main point of contact to engage with the Heads on various issues. However, with the creation of the new steering committees, their approach is evolving. There is a sense that working with a smaller group of Heads on specific issues and then validating the approach with the broader group, will be more effective in working collaboratively on specific files and implementing concrete actions (e.g., promoting the value of the functions, improving data quality and availability, sharing best practices, recruitment & retention of staff, professional development, etc.).

In the **UK** There are also several bodies/groups which support and facilitate the development of guidance and sharing of practices related to policy evaluation. The Government Economic Service and the Government Social Research Service are the two cross-government analytical civil service professions that are most commonly tasked with undertaking policy evaluation work. They set a range of professional standards, coordinate learning and development, and provide both formal and informal mechanisms for developing standards for evaluation of policy and social research and sharing experiences. The Cross-Government Evaluation Group (CSEG) is coordinated by the Government Economic and Social Research Teams, the professional support units for the Government Economic Service and the Government Social Research Service. The CSEG is a cross-departmental and cross-disciplinary group made up of analysts and evaluation managers from government bodies that works to improve the supply of, stimulate the demand for, and encourage the use of good quality evaluation evidence in government decision making.

In the **NLD**, the Bureau of Strategic Analysis (BSA), a unit inside the Ministry Of Finance, has created an evaluation community (Interdepartmental Supervisory Committee on Performance and Evaluation or IBP), mainly consisting of people working in FEZ. They meet regularly (monthly or two-monthly) to exchange good practices and discuss relevant approaches for evaluations and policy reviews. This community has contributed to install an increased ownership of the line ministry in this field, evolving from a top-down process to more co-creating of the tools and framework. Within this group, new instruments such as Strategic Evaluation Agenda and the Public Value Scan have been discussed before their introduction. Within some ministries, an evaluation committee is installed, although there are no standard guidelines for this, and every minister can organise it according to its own insights. Typically this committee consists of senior officials responsible for a directorate, external expert(s) and a delegation of the FEZ. This committee can discuss the planning and implementation of the evaluation activities in the ministry, and can also discuss some major evaluation exercises.

- A number of countries have established **specialised units within finance ministries** to support capacity building of public servants and build institutional knowledge so lessons from one spending review can be applied to future reviews.

Latvia implemented spending reviews as an integrated part of the 2016 State Budget. Latvia has since looked at how to improve the process by carrying out deeper analysis of spending and the possible ways to reduce the administrative burden. A separate division (Budget Development Division) was established within the Budget Department in mid-2018 dedicated to spending review.

Norway established a specialised unit for spending reviews in the Ministry of Finance to build capacity and scale up the use of spending reviews. Located within the Budget

Department, the unit works closely with line ministries to reach agreements on the set of recommendations put forward to ministers for final decision.

- **Evaluation societies** sometimes offer training courses and workshops on evaluation.

In **CAN** the Canadian Evaluation Society CES has been committed since 2009 to the implementation of the title of accredited evaluator. This professional title program is built on three pillars: a code of ethics, professional standards and a set of 36 theoretical and practical skills (knowledge of basic concepts of assessment, mastery of data collection and processing methods, project management skills, interpersonal skills, etc.). The expected skills do not focus on a particular discipline but on general professional dimensions; it is deliberately not prescriptive from the point of view of the methods and techniques used (quantitative, qualitative) in order to encourage multidisciplinary evaluations.

Also in the **US** evaluation is recognised as a profession and is represented and supported by the American Evaluation Association.

In **GER** the evaluation society is the main entity which focuses on the professionalisation of evaluators and the development of a theoretical framework. It has published the Standards for Evaluation and Recommendations on education and training in evaluation. DeGEval has also established working groups to serve as a platform for specific dialogue in various fields of application and are involved in annual conferences and meetings, to which also ministries regularly participate.

- There are often Masters courses on evaluation and evaluation-related subjects being offered by several **universities (UK, GER, ...)**.

In the **UK**, NESTA with the Alliance for Useful Evidence has created 'Evidence Masterclasses', providing "an immersive learning experience" for senior decision-makers who want to become more skilled and confident users of research.

In **GER** institutes such as the Institute for Economic Research (DIW) and the Institute for Employment Research (IAB), propose doctoral trainings that are often open to administrative staff as well.

On the other hand, in **CAN** the two major schools for the initial training of civil servants, the Canada School of Public Service (CSPS) at the federal level, and ENAP in Quebec, do not offer training modules in impact evaluation in their catalogue.

- Some countries show examples of a specific **job category for evaluators** in the government and requirements on the level of skills that are needed by specific (high-level) persons involved in the evaluation.

In the **US**, the Foundations for Evidence-Based Policy Making Act requires agencies to create three new positions: evaluation officer, statistical official, and chief data officer. It also requires the creation of a new (or enhancement of an existing) job series in the civil service for program evaluation.

The **UK** Civil Service is composed of a range of professions. Two of these civil service professions are specialised in evaluation activities: the government economic service and the government social research profession. The Government social research (GSR) profession works alongside other analysts (economists, statisticians and operational researchers). GSR professionals use the core methods of social scientific enquiry, such as surveys, controlled trials, qualitative research, case studies and analysis of administrative and statistical data in order to explain and predict social and economic phenomena for policymaking.

In **IRE** the Irish Government Economic and Evaluation Service (IGEES) operates as an integrated, cross-Government service, supporting better policy formulation through economic analysis and policy evaluation.

In **Canada** the Treasury Board Secretariat has a document listing the competencies for the Heads of Evaluation.

In **NLD** in important policy departments, research coordinators – often senior policy officials – fulfil a key function regarding policy evaluations. They are relatively experienced and help forward less experienced colleagues. The appointment of a Chief Science Officer (CSO) within the Ministry of Social Affairs (and other ministries) can be mentioned as an example. The CSO is a scientist who provides support and direction to (evaluation) research and a variety of other knowledge initiatives within the ministry. He can also build a bridge between academic researchers and the ministry, aiming at more evidence-based policy development and policy-relevant research.

- An interesting approach for ministries with relatively limited staffing capacity is **evidence synthesis**, the possibility of using “meta-analysis” and systematic reviews to complement the need to conduct original analytical work for any new policy question.

The meta-analysis and systematic reviews are part of the “What Works” approach championed by the **UK**. It offers a cost effective alternative to produce analytical research through systematic scans of the international literature.

- When capacity needs to be built up from scratch, recruitment of people who are already skilled in evaluation techniques has an obvious role to play. External **experts**, whether academics or others businesses, can usefully supplement government expertise, particularly when specialised skills are called for.

In the **UK** a Cross-Government Trial Advice Panel, a panel of experts from academia and the civil service, was set up by the Cabinet Office and the Economic and Social Research Council to help ministries to implement impact evaluation and support the increased use of controlled experiments in public policy. Independent bodies support and facilitate policy evaluation, such as the What Works Network (WWN) and the Behavioural Insights Team (BIT). Research laboratories, which may depend on major British universities, consulting private companies carry out a significant part of public policy impact evaluations. Private non-profit associations or foundations, including Nesta and the Royal Statistical Society, are also involved in evidence-based policy. This large variety is a strength, but also creates challenges related to the lack of coordination mechanisms of this complex ecosystem. Indeed, WWNs, think tanks, research centres and government groups are all conducting policy evaluation and commenting on government’s performance, but the links between those institutions are not always very clear.

- Sometimes **international contacts** can be leveraged to learn about best practices and improve capacities.

In **CAN**, Health Canada, for example, has strong connections with entities in the US, UK and AUS.

- Some cases offer a **strategic approach to capacity building**.

In the **US**, in addition to its other provisions, the Evidence Act in the US places emphasis on the learning function of evaluation by calling for the development of ‘Learning Agendas’ which are at the core of the two main activities that all agencies are required to carry out - evaluation plans and capacity assessments. ‘Capacity assessments’ help agencies obtain a better understanding of their evaluation expertise and resources to undertake evidence-building activities. These assessments encourage agencies to reflect on their “coverage quality, methods, effectiveness, and independence

of their statistics, evaluation, research and analysis efforts.” ‘Learning Agendas’ are comprised of prioritised research questions and evidence-building activities, specific to each agency’s purposes and needs that will better guide their future practices and decision-making processes. Strategic objective reviews (SORs) are to be carried out on an annual basis, with the aim of informing annual planning and budget formulation but also long-term strategy. They help to identify areas where additional evaluation, analyses, skills or other capacity are needed, and to develop a culture focused on learning and improving performance.

Summary requirements for planning and evaluation for departments based on the Evidence Act

	Frequency	Content
Agency Priority Goals	2 years (quarterly updates)	Near-term results advancing progress on longer-term goals in the Strategic Plan
Evaluation Plan	Annually	Evaluation activities to be conducted in the next fiscal year
Performance Plan	Annually	Strategic objectives and performance goals for the next fiscal year
Performance Report	Annually	Progress on goals and objectives set in the Performance Plan and the Strategic Plan
Capacity Assessment	4 years	Coverage, quality and methods of statistics and evaluation efforts of the agency
Learning Agenda	4 years (updated annually)	Policy questions relevant to programs, policies and regulations of the agency
Strategic Plan ⁽¹⁾	4 years	Agency mission, long-term goals and objectives, strategies, risks and challenges
Strategic Review	Annually	Progress on the strategic objectives (noteworthy progress or areas for improvement)

In **CAN** a standard curriculum within the federal public service context has not yet been developed. Still, TBS has mapped available courses for the development of performance measurement and evaluation skills according to different levels (beginner, intermediate and advanced). Functional communities on performance measurement and evaluation are developed to exchange knowledge. The challenges and opportunities related to staffing capacities and resources for evaluations are usually reported in the annual Departmental Evaluation Plans.

In the **Netherlands** there is some room for experimentation, as proven by the different experiments that have been set up during Operation Insight in Quality. The improvement paragraph in Policy Reviews requires the ministries to look at information gaps when evaluating policies and come up with strategies to mend these gaps in the future. Similarly the Strategic Evaluation Agenda that was introduced by the Operation Insight in Quality aims to gather more relevant insights at the right moment, and the learning aspect has been highlighted when introducing this new instrument. Also noteworthy in this respect is a recent initiative of the MoF. In 2018, the MoF started awarding the prize for 'Best grant evaluation. The prize is intended to draw attention to the quality of these evaluations, to highlight the positive points and to provide tips to improve the quality. The jury includes representatives of the Court of Audit, the Netherlands Bureau for Economic Policy Analysis, the Bureau for Statistical Analysis and the MoF.

More generally in the **NLD** the operation Insight in Quality intended a shift from accountability towards more attention for the learning perspective. The introduction, inspired on the experiences in the UK, of a Public Value Scan is a good example. Rather than trying to find an answer for the question whether the policy has been efficient and effective, it focuses on the question whether everything is done to maximise the expected added value of the policy under the current circumstance, with input from various experts, relevant stakeholders and the policy officers within the ministry. The focus is not primarily on the results and outcomes achieved (product evaluation), but rather on the activities undertaken and efforts made (process evaluation). Furthermore, the emphasis is also more on the current policy and less on an ex post justification of the policy pursued up to that point. This means that the public value scan can be considered as a form of *durante* evaluation. This new instrument is very new (only one pilot has been conducted for the moment) and it has still to be determined how this relates to the existing other tools, such as the policy reviews.

In **ESP** the region of **Cataluña** provides an interesting example. It has set up the Catalan Institute of Public Policy Evaluation (IVÀLUA) in 2008, with the aim of increasing the culture of evaluation and the quality of evidence-based public policies. IVÀLUA is a public consortium comprising the Generalitat of Cataluña, the Provincial Council of Barcelona, and the Pompeu Fabra University. Its

resources come from the members of the consortium, mainly from the Generalitat of Catalunya and the Provincial Council of Barcelona. IVÀLUA carries out the evaluation of public policies and programmes either on its own initiative, or on behalf of public institutions from the autonomous community. It carries out between two and three evaluations per year mostly on social policies, as well as particular evaluations commissioned and paid by public institutions. Although IVÀLUA conducts evaluations itself, its main activities are related to the generation of general knowledge on evaluation issues. IVÀLUA has developed several guidelines and training material to help public institutions and social entities to carry out the evaluation of public policies. IVÀLUA has also developed standard specifications that departments can use as the basis to define the terms of reference to outsource specific evaluation-related activities. The purpose of these standard specifications is to facilitate the outsourcing process that departments may carry out to get their evaluations done (when they are not done internally), as well as help guarantee a minimum quality level of the future outcomes of the evaluation process. Finally, IVÀLUA offers a wide range of training courses on evaluation issues. It offers introductory assessment courses on key evaluation concepts and use of evidence, and more advanced courses on quantitative and qualitative methodologies for policy evaluation. IVÀLUA also offers tailor-made training activities in response to the specific demands and development level of the different institutions that wish to broaden the knowledge of their staff in evaluation issues. Moreover, it has conducted a search of training courses in policy evaluation of open access, and has published the links in its web site to facilitate access.

How to ensure quality?

Quality is key. The quality of evaluations is essential to ensure impact on policy-making, and thus in ensuring that evaluations actually serve as tools for learning, accountability and better decision-making. Bad quality evaluations can lead to poor evidence, which can be costly and misleading, and threatens political support to invest funds and resources for policy evaluation.

Capacity building is not enough to guarantee the quality of evaluation. For example, evaluation **guidelines** and standards are not always put into practice in a way that improves evaluation results. Establishing a direct link between the existence of guidelines and standards, and the quality and usefulness of evaluation studies is inherently problematic. This is because there are many factors in addition to the existence or otherwise of guidelines that influence how well an evaluation is conducted.

For example, according to a 2016 National Audit Office (NAO) assessment in the **UK**, fewer than half the evaluations it examined (14 out of 34) were “of a sufficient standard to give confidence in the effects attributed to policy because they had a robust counterfactual.” Several problems were highlighted: firstly, independent evaluators outside of the Government often experience difficulties accessing the necessary official data that is needed to evaluate the impact of programmes and policies. Secondly, evaluation timescales can be unrealistic which makes it more difficult to deliver high quality evaluations. Last but not least, the NAO argued that there is often a lack of demand from policymakers for evaluations with few incentives for Government departments to generate and use evaluation evidence, and few adverse consequences for failing to do so.

A variety of **mechanisms for quality control** of evaluations and **quality assurance mechanisms** can be put in place to foster quality evaluations that are technically and methodologically sound in well-governed. These mechanism can be imposed from the central level or created at a more decentral level (ministries/departments).

- Quality assurance starts from the beginning of an evaluation by addressing the **evaluability** or the extent to which a policy can be evaluated, meaning that it should be possible to

evaluate it in a credible and reliable manner. Two main factors may affect the degree to which a policy can easily be evaluated: the nature and design of the policy itself and the quality and availability of specific data. Clearly laying out the objectives of a policy and the levers to attain it will facilitate the evaluation. This also implies that the intentions of the policy should be explicit. One way to facilitate clear policy objectives is to develop a **theory of change** and **intervention logic** model, which can be done either at the stage of policy design, or when developing an evaluation. A theory of change is a set of interrelated assumptions explaining how and why an intervention is likely to produce outcomes in the target population. The quality and availability of non-evaluation specific **data** (big data, open data, statistical data, administrative data, etc.) in a format that can be readily used is also a factor in how easily a policy can be evaluated.

For example, in the **US** guidance for RIAs suggests that future regulations should be drafted in a way to facilitate evaluation of their consequences, taking into account the empirical testing of the effects of rules both in advance and retrospectively. OMB moreover is responsible for both the performance monitoring as the evaluation framework, allowing them to align both instruments. This facilitates that information that stems from the performance management can be useful for evaluation purposes.

This is also the case in the **Netherlands** where, following a new regulation in 2018, all policy proposals that imply major policy changes have to include an evaluation paragraph (evaluatieparagraaf), where the ministry indicates how the proposal will be evaluated after implementation. This is one of the outcomes of the discussion as part of the operation Insight in Quality, on the link between ex-ante and ex-post evaluation. An important role of the FEZ is to stimulate the programme officers to start thinking about the policy reviews far ahead of the timing that the policy review has to be conducted, making sure that the necessary data are available when the policy review has to start.

In the **UK** policy proposals should also envisage monitoring activities and indicators, so that corrective actions can be taken during policy implementation. For instance, single departmental plans include indicators that can be used to monitor performance on each objective. After policy implementation, government departments are expected to conduct ex-post evaluations, including project implementation review and post-evaluation review. Ex-ante and ex-post evaluations should be linked, as the policy design should identify the relevant indicators to be monitored and evaluated, and ensure that the information is collected during the policy process.

In **GER** the 2013 concept for evaluation was updated in 2019, when a decision of Secretaries of State clarified additional aspects to be included in the draft bills that would facilitate and improve quality of ex-post evaluations. When presenting a draft bill the objectives of the bill and the criteria for achieving the objectives have to be stated, as well as the steps and methods of evaluation that will be used. Furthermore the Federal Government is developing a guidance document outlining the steps and methods of an evaluation.

- During and after an evaluation, **stakeholder involvement, public consultation, advisory panels and committees, scrutiny mechanisms** and **peer review** can be used to promote the quality of evaluations. Their main aim is to provide comments and feedback throughout the different phases of the evaluation (design, data collection, synthesis, etc.). These advisory panels and committees may be composed of policy practitioners, evaluations experts and stakeholders and may be established on an ad hoc basis or systematically (e.g. the **EC** Impact Assessment Board, the Impact Assessment Committee in **BEL**, ...). Peer review is another mechanism that is used by several countries to guarantee and improve the quality of evaluations. Strong central control on the quality of the ex post evaluations done within line ministries is less common. Some countries have organised a control mechanism at the level

of the line ministries. Besides this internal controlling mechanism, external oversight by for instance the supreme audit office may also be important for the quality.

Scrutiny mechanisms exist mainly for ex ante regulatory impact assessments. In **CAN** the Privy Council Office (PCO) checks the quality of the analysis for regulatory proposals. In **GER** a Regulatory Control Council (Normenkontrollrat - NKR) checks the relevance of the estimations of their compliance costs of regulations in ex ante assessments and other aspects such as the objectives and the necessity of the proposed regulation.

In **Canada** there is a controlling mechanism installed in the larger departments (the Performance Measurement and Evaluation Committee reviews evaluation reports) but there is no institutionalised control mechanism at the central level. In the **Netherlands** there is no obligation to install an evaluation committee but some ministries have created voluntarily an evaluation committee, although there are no standard guidelines for this, and every minister is free to organise it according to its own insights. An example of the role an oversight committee can play to increase the quality, is to be found in the **UK** where the National Audit Office initiates about 70 to 80 value for money reports per year, reviewing a wide range of topics. The majority of these reviews are examined by the Parliament and discussed in the Committee of Public Accounts where government officials are held to account for their policies. In the **US** the Government Accountability Office has a similar role, providing recommendations on how to improve the efficiency, effectiveness and responsibility of government operations. It prepares 600-700 reports on evaluations each year prescribed by law, requested by congressional committees or on their own initiative. **GER**, in November 2019, adopted a requirement for independent quality assurance for all internal ex post evaluations and all ex post evaluations of legislative proposals exceeding EUR 5 million in annual compliance costs.

Peer review is well established in the **NLD** where policy reviews are to be accompanied by an independent expert opinion and sent to the House of Representatives. In **CAN** the Performance Measurement and Evaluation Committee (PMEC) in each large department reviews evaluation reports and summaries and monitors follow-up on evaluation recommendations and action plans. TBS strongly encourages peer review. In **GER** and **UK** peer reviews - both internal and external - of evaluations are a widespread practice.

In **Canada**, a variety of methods are used to **involve stakeholders** in consultations on the RIAs. They include the use of emails, phone calls, third-party facilitated sessions, roundtable meetings and online consultations. In a second step, regulatory proposals and their accompanying RIAs are pre-published in the Canada Gazette, Part I for public consultation. A great number of regulations are updated as a result of the consultation process on the draft text and the RIAs. Departments and agencies must summarise the comments received, explain how stakeholder concerns were addressed, and provide the rationale for the regulatory organisation's response (i.e., the decision to change or not change the draft regulation) in the final RIA, which is then published in the Canada Gazette, Part II together with the final regulation.

Two-thirds of **EU Member States** engage stakeholders in some ex post evaluations. **Estonia** and **Sweden** and the **European Union** report to systematically engage stakeholders in ex post evaluations. Since 2017, **Croatia** and **Latvia** involve stakeholders in evaluations. Most commonly, stakeholders are provided ongoing opportunities to submit comments, participate in interviews and meetings. Their input helps to identify areas for improvement and is often included in the scope of any evaluation.

In **Sweden**, experts from business organisations and other interest groups can be appointed as experts in a committee of inquiry established by the responsible ministry to carry out ex post evaluation of a regulation. Referral bodies and stakeholders are also invited to provide comments on the final report, which are then dealt with by the responsible ministry in the continuous work within the Government Offices.

- When **commissioning** an evaluation, guidelines for drafting the **terms of reference** (ToRs) have proven to be essential for quality evaluations. ToRs cover the background context of the evaluation, its scope, goals, methodology, team composition, stakeholders to be engaged and the evaluation budget. The ministry of national planning and economic policy (Mideplan) of **Costa Rica** e.g. has developed such guidelines.
- An important way to guarantee and improve quality is the set-up of **controlling and oversight mechanisms**. These exist mainly for ex-ante regulatory impact assessments (cf. supra). There usually is no strong central control on the quality of the (ex-post) evaluations done within line ministries. Relatively few EU Member States have an oversight body in charge of systematically reviewing the quality of either ex post evaluations. Some countries have organised a control mechanism at the level of the line ministries. Besides this internal controlling mechanism, external oversight by for instance the supreme audit office are also important for the quality.

In **Canada** there is a controlling entity in the larger departments. According to the Policy on Results the Performance Measurement and Evaluation Committee within the larger departments reviews evaluation reports. There is no institutionalised control mechanism at the central level (TBS).

Germany in November 2019 introduced a requirement for independent quality assurance of all ex post evaluations of legislative proposals exceeding EUR 5 million in annual compliance costs.

In the **Netherlands**, there is no obligation to create an evaluation committee such as in Canada. Some ministries have created voluntarily an evaluation committee, although there are no standard guidelines for this, and every minister is free to organise it according to its own insights. Typically, this committee consists of senior officials responsible for a directorate, external expert(s) and a delegation of the FEZ. This committee can discuss the planning and implementation of the evaluation activities in the ministry, and can also discuss some major evaluation exercises.

In the **USA**, the Government Accountability Office has a similar role, providing recommendations on how to improve the efficiency, effectiveness and responsibility of government operations. It prepares 600-700 reports on evaluations each year prescribed by law, requested by congressional committees or on their own initiative.

Ex post evaluations conducted by the **European Commission** are subject to quality control mechanisms of the Regulatory Scrutiny Board, which publishes its opinions on evaluation quality.

- As explained earlier, **Parliament** may play a role in the scrutiny of ex post evaluations prepared by the executive.

Furthermore ensuring the right **conditions for implementation** is essential. An integral part of an evaluation policy framework is therefore an explicit policy for:

- **Guidance** for technical quality (proper design, sound data collection, rigorous methods, ...), Standards for the quality of the evaluation process (independence, ethical conduct, ...) and self-evaluation tools and checklists
- Supporting and promoting evaluator **competences** (training, advice, academic experts, professional networks etc) (**CAN, UK, ...**)
- **Data** collection and use (**US**)
- **Expertise and resources** (e.g. evaluation units, **UK, CAN, US...**)
- **Transparency** (of the planning, the evidence, the results, ...) (cf. supra).

In this respect, the OECD encourages its members to **monitor and assess ex post evaluation policy**. Today, only a minority of EU Member States publish online reports on the performance of their *ex post* evaluation system, mostly on an *ad hoc* basis (Reporting on the performance of RIA systems is comparatively more widespread). Only **Austria** has assessed the effectiveness of ex

post evaluations in improving the regulatory stock in the past five years and published the results. Efforts by the EU's Regulatory Scrutiny Board (2019) to draw forward-looking conclusions from its scrutiny of *ex post* evaluations illustrate, however, the benefits of conducting this kind of assessment on a systematic basis. Such benefits include identifying recurrent design flaws to improve methodological approaches and helping to prevent potential biases and conflicts of interest.

How to increase data and information availability?

The quality and availability of non-evaluation specific **data** (big data, open data, statistical data, administrative data, etc.) in a format that can be readily used is a factor that influences how easily a policy can be evaluated.

A good evaluation system relies on **comprehensive, quality data**. The full implementation of an evidence-based agenda implies leveraging the data that are available for analytical purposes. As such, a high quality national statistics system is an integral part of any evaluation strategy, as well as up-to-date databases and registers that mutually communicate and disaggregate data at the desired level. Data access and re-use is critical for ensuring the efficiency and effectiveness of the policy evaluation research effort. Access to trusted, well described data, together with the software, models and workflows that are necessary for the production and analysis of this data, is necessary across the many different policy domains.

Important **data sources** for policy evaluation are:

- Statistical data: commonly used in research, it corresponds to census data or more generally to information on a given population collected through national or international surveys.
- Administrative data: this data is generally collected through administrative systems managed by government departments or ministries, and usually concerns whole sets of individuals, communities and businesses that are concerned by a particular policy. For instance, it includes housing data and tax records.
- Big data: mainly drawn from a variety of sources such as citizen inputs and the private sector, big data is most often digital and continuously generated. It has the advantage of coming in greater volume and variety.
- Evaluation data: this data is collected for the purpose of the evaluation. It can take the form of qualitative questionnaires, on-site observations, focus groups, or experimental data. See further down for a description of impact evaluation methods to collect and analyse data.

Often, the limited availability and quality of data across government agencies and departments is a **major challenge** for evaluation practices. Standard data collections within government may not have the granularity or specificity needed to evaluate all relevant impacts. Therefore, one clear message from the OECD Best Practice Principles on Reviewing the Regulatory Stock is that it is imperative to begin thinking about how regulations ought to be reviewed **at the time when they are initially designed**, particularly to identify the data that will need to be collected to assess the impacts. Data requirements for evaluations should be considered at the time a policy of regulation is being made, as part of wider consideration of the type of *ex post* evaluation that would be most appropriate. In such circumstances it may be that the data needed to assess performance has to be collected as part of the regulatory regime itself. This can be done under compliance reporting obligations and/or through survey instruments.

From a broader data perspective, many data are not **reusable** by, or useful to, different communities if they have not been sufficiently documented and contextualised or appropriately

licensed. In many countries, there is no integrated data infrastructure and access to data across departments remains difficult. The architecture often does not guarantee data interoperability. Data siloes remain within the Government itself, as Departments operate in similar sectors but cannot (or refuse to) share data. Problems remain with obtaining linked files where coverage of common identifiers to merge datasets is low. In some cases, the challenge is to understand what data and data sets currently existed within departments, and how departments could use the data for policy analysis.

Apart from technical problems, enabling the strategic use and quality of data requires **human capabilities**, especially the willingness of public entities and civil servants to share use data, and to invest in databases, data linkages and analytical tools. Political reluctance towards sharing evidence on policy impact and effectiveness may also be another barrier in accessing data.

In **European countries**, the requirements for national statistics systems and the European Directive on Public Sector Information entail a stronger institutionalisation of the evidence, statistical or data functions in government and the development of the National Data Infrastructure (NDI), which aims to improve how the government collects, manages, shares and stores data in order to make it more useful for policymaking.

At the same time, much of the data related to the pandemic is personal or sensitive, and it is critically important to protect **privacy** and ensure the security, ethical management and use of the data. On the other hand, GDPR rules and conservative practices regularly restricts access to individual administrative data, does not allow for merging statistical files and hampers the scope of data use and data linkages without having to release any confidential information outside of the boundaries where the data have been collected. Academics speak of their frustration at the complexity of the processes they are required to go through for even repeat requests to use the same data sets of multiple occasions. It also seems that existing government's IT systems are facing limitations, resulting in challenges to carry out statistical and other forms of analysis.

In several cases, e.g. in **cooperation between central and subnational Governments** has proven to be a topic for improvement, notably with regard to access to information and data. Access to data is complicated as data are gathered by different entities and there is no obligation to share data with the federal level or vice versa. Often subnational Governments enjoy full autonomy in some specific matters and this issue has become as a challenge during the COVID 19 pandemic (e.g. **CAN**).

In recognition of the limitations in infrastructure to support the use of evidence, some jurisdictions have launched **initiatives** to try to maximise the use of government's existing assets for EIPM. These include measures to avoid fragmentation and duplication of efforts (e.g. in developing separate data sharing infrastructures) and to promote public sector integration and cohesion, library facilities, data portals and clearinghouses as well as data sharing software such as the use of APIs and open data and other methods of maximising government's data assets. Some departments create a 'Data Map', setting out the relevant departmental and agency data to increase awareness of the data that is held by the policy units.

At the **intergovernmental level**, a revised OECD Recommendation on Access to Research Data from Public Funding was adopted in January 2021. At a more operational level the RDA-COVID-19 WG Recommendations and Guidelines for Data Sharing were published six months into the pandemic and provide detailed advice for different research domains. Policymakers, research funders, and research institutions around the world have been working together to implement these frameworks and guidelines, to mandate, incentivise and support actions that harmonise

and streamline the responsible and timely provision and exchange of data locally, nationally and internationally.

Fostering a **Data Driven Public Sector** culture can be a very effective way to enhance the quality of ongoing evaluations through the application of relevant data. Data-Driven Public Sector initiatives generate an environment in which data about policy interventions is available in real time to avoid waiting for monthly or quarterly updates across a wide range of policy areas because the data they need is more frequently available and accessible. As a result, governments can gain better insights into the policy process and make quick policy adjustments in the short-term, with benefits in the mid-and long-term. Policymakers can assess whether policies have had the desired effect, and if that data is published in Open Data format, so can other stakeholders. Policy evaluation can therefore become a more open, inclusive and ongoing process.

In the **US**, the Foundations for Evidence-Based Policymaking Act was designed in part to ensure that the necessary data quality and review structures were in place to support the use of administrative data in evaluations. Guidance is being prepared by OMB for Open data access and management and for Data access for statistical purposes. Agencies are i.a. required to submit an evidence-building plan (the so-called Learning Agenda) with strategies, data and methods to develop evidence and create an Advisory Committee on Data for Evidence Building to analyse and make recommendations on how to promote the use of data for evidence building. OMB established an Advisory Committee on Data for Evidence Building at the federal level, to review, analyse and make recommendations on how to promote the use of federal data for evidence building and how to facilitate data sharing and data linkage.

In **CAN** the Data Strategy Roadmap for the Public Service is intended to support a more strategic use of data while protecting citizens' privacy. The recommendations are structured around four themes: stronger governance, improved data literacy and skills, enabling infrastructure and legislation, and more focused treatment of data as a valuable asset. The goal is to set a foundation so that the Government of Canada creates more value for Canadians from the data they hold. Within this data strategy, the Central Performance and Impact Assessment unit at TBS is coordinating the collection of data on business innovation from across Departments in order to facilitate the conduct of analyses from a horizontal perspective. This unit will also be undertaking analysis with the data collated. Co-location of TBS and Statistics Canada staff was a success factor with combined teams jointly working on validation of results. Another success factor was the engagement of several actors involved in evaluations, i.e. Departments and agencies as well as the research community. As a result, Statistics Canada is developing a rich dataset that can benefit researchers and policy analysts within Government and academia, and gaining experience with an approach and methodology that can be applied to other domains. From an Open Data perspective, official data is provided by the Treasury Board through a single data-driven tool called InfoBase instead of being scattered across close to 100 annual Departmental Plans. The Canadian Government has also worked to support the "demand side" of open data and information, identifying and collaborating with stakeholders in organizations and companies that leverage open Government data and information. However, there are still several challenges, including: finding and accessing the right data, having common data definitions and terminologies, overcoming privacy-related issues, and mobilizing the resources to make the kind of longer-term investments required to develop new data linkages of good quality.

In **NLD** Policy Reviews should be a synthesis of existing information, but it appears that in most policy areas there is not enough evaluation information available to base the policy review solely on these previous evaluations. Therefore the questions have been updated recently to include an improvement paragraph: are there gaps in the available information to make a relevant evaluation of the effectiveness and efficiency of the policy and how can this gap be filled? There is however no central monitoring of the way the improvement paragraphs is being implemented by the ministries.

It is too early to evaluate if this measure will lead to a substantial improvement of the availability of data for the policy reviews.

In **Germany**, the Federal Statistical Office is the leading provider of statistical information required for developing informed opinions and decision-making processes. It provides advice to federal ministries and is contacted by the Federal Government whenever additional data are required or problems arise with the interpretation of official data. The Federal Statistical Office conducts the nationwide statistics in close cooperation with the statistical offices of the Länder. During the coronavirus crisis, ad-hoc evaluations of experimental data have contributed to illustrating the impact of the pandemic on economic development, mortality figures and mobility.

How to promote impact and use of evaluations?

A sound framework and correct implementation of evaluations is no guarantee for successful use of evaluations in the policy process. Therefore, **uptake** of policy evaluations results should be enhanced. Both demand and supply are important to an evaluation system and it is through the interaction between the two that an overall evaluation culture develops. A key driver for successful use of evaluation information is a strong demand from Parliament and external stakeholders.

Some **mechanisms to promote the use of evaluations** include:

- Conducting user-focused or **utilisation-focused evaluative processes**: participation of users and stakeholders in the evaluation process, strong cooperation between researchers and policy makers...

The Public Value framework in the **UK** is very focused on optimising the process of turning funding into policy outcomes for citizens.

The newly introduced Public Value Scan in the **Netherlands** requests each minister to indicate what value the pursued policies add to society when evaluating a policy.

- More awareness for the virtues of **'positive evaluation'**. Positive evaluation focuses more on success than failure. It can promote the willingness to evaluate, learning effects and the uptake of evaluation results.
- **Evaluation agendas** (planning, like for example in the **US** and **NLD**) to match evidence with demands of policy makers and users' needs in terms of timing and priorities. Attention for enhancing the policy impact of an evaluation should not come after the evaluation has been concluded, but much sooner. This supposes, among other things, careful consideration of the timing of policy evaluation to make sure they are properly linked to the policy and political cycle.
- **Evidence synthesis**, using standardized formats, meta-analysis and systematic reviews, ...

Here reference can be made to IZA World of Labor (**GER**), an academic-supported online platform with a comprehensive set of understandable, non-technical summaries of policy-relevant research on labour market themes. This is a initiative by a number of German scientific institutions to summarise labour market research (not just in GER). This is done by top researchers who do this for the policy and other researchers. The focus is on short, focused and understandable. The summaries ('onepager') always have the same structure: Elevatorpitch, Key findings and Author's main message. There is also always a link to a more extensive article of 10 pages;

Also in **AUS** The Policy Liaison Initiative created a tailored summary format to present the findings of reviews).

Meta-evaluations are common practice in **GER** and **UK**. This term is used for evaluations designed to aggregate findings from a series of evaluations; it can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluation.

- Increasing demand for evaluations through **competency development** tailored at users (see the OECD-JRC Skill Set for Evidence-Informed Policy-Making).
- Embedding use in the **institutional set-up**, within and outside of the executive (management response mechanisms that indicate whether policy makers agrees or disagrees with the assessment and recommendations of a policy evaluation, the reasons for agreement or disagreement are provided, and actions to be taken in response to the evaluation are described; the incorporation of policy evaluation findings into the budget cycle or discussions of findings at the highest political level and in Parliament; nominating specific individuals or a specific body to take forward the findings of evaluations, etc.;

In the **NLD** the policy reviews are sent to the House of Representatives with a cabinet response; The IAK has been supplemented with a new requirement that all proposals for the Council of Ministers have to contain information on the goals, effectiveness and efficiency as well as the instruments and the budgetary consequences of the new proposal. This new requirement has to be integrated in the IAK, focusing more on goal orientation, effectiveness and efficiency.

The policy profession standards require **UK** civil servants who belong to the policy profession, (i.e. the civil servants who are involved in the design of public policies) to use scientific data in their work. This means that they are required to base their public policy recommendations on evaluation results.

The **US** has created an Interagency Council on Evaluation Policy, co-chaired by OMB and the Department of Labor, composed of about ten high-capacity evaluation officers from government agencies, who meet on a monthly basis, to discuss evaluation results.

In **CAN** Departments prepare Departmental Results Reports (DRRs) that are submitted to parliament and are available on Departmental websites so that the public can access the information. Treasury Board Submissions requesting funding to the Treasury Board (a key financial decision stage in the budget process for the Cabinet) have to have one or more paragraphs indicating what past evaluations examined, what the findings were, what was changed. It should also indicate if new evaluations are coming up in the next cycle.

- Promoting **communication, availability and accessibility** of evaluations (tailored communication and dissemination strategies, transparency, websites, communication, smart 'lobbying' or 'expectations management' with respect to policy makers and media, ...).
- **Organising interaction** between analysts and policy makers. Beyond publishing reports on websites, some countries develop strategies to encourage interaction between analysts and policy makers. Strategic policy discussions, policy forums and presentations are some of the examples. In AUS for example, the Policy Liaison Initiative was an attempt to create an 'Evidence-Based Policy Network' to facilitate knowledge sharing between policy makers and researchers, alongside seminars by national and international researchers in field of evidence synthesis and implementation.
- Moving from knowledge management to **knowledge brokerage**, acknowledging the fact that the worlds of policy-making and research are very different, with different professional cultures, resources, language, imperatives and time frames than those of science. Knowledge brokers are organisations or individuals, who thanks to their intermediary position in the system can establish and maintain links and cooperation between knowledge producers and

users. Knowledge brokers play a key role in strengthening the relationship and collaboration between evidence producers and policymakers. A knowledge broker (interface, transfer agent, boundary spanner, intermediary ...) and can take on a variety of forms, ranging from individual professionals (such as Government chief science advisors in some countries, or ministerial advisors) over government entities to dedicated organisations. Harnessing their knowledge about both the research process and the policy cycle as well as they extended connections with representatives of these two worlds, they can increase availability of the robust knowledge and build culture and capabilities of evidence use. They can help to facilitate policymakers' access to the results of evaluations and to research evidence by helping them to navigate research material that may be unfamiliar. They can also help to articulate policymakers' needs, constraints and expectations, translating them for researchers who may be unfamiliar with the policy process.

The **UK** provides an interesting example of knowledge brokerage. A 'What Works Network' (WWN) was created in 2013 with the purpose of expanding the culture of evidence use in policy making. Today the WWN is constituted of around 14 independent What Works Centres (WWCs) that are engaged in the three activities: Improve the supply of high-quality evidence (generating evidence), create incentives to use evidence in policy making decisions and practice (translate evidence for specific user groups in a friendly-format) and build capability across government to improve the use of evidence for policy making. The What Works Team (WWT) coordinates the WWN and the Cross-Government Trial Advice Panel. Since 2015, the WWT encourages the ministries to publish the Areas of Research Interest (ARI) indicating the fields where there is a need for scientific evidence, in order to transmit them to researchers. Ministries have since joined in and issued an ARI document, which are published in the government's webpage. Furthermore, WWCs support the government with specific research and impact evaluation activities in those situations where there is not enough time or appropriate resources for departments to conduct the research themselves. In these cases, informal channels are used to encourage the coordination between the WWCs and the government departments: Sometimes the WWT links the WWCs with those teams within the government departments that have the need to understand something specific; at other times, the chief economist or head of evaluation of the government department makes use of informal channels to establish a relationship with the respective WWC. However, since the WWCs are not the only entities capable of providing support in the evaluation process, units within the government department can turn out to other entities.

- Finally, the **linkage between policy evaluation** on the one hand and performance monitoring **and spending reviews** on the other can be exploited to promote the impact and use of evaluations. The performance, evaluation and value for money is integral to the budget process. Governments are more systematically aligning planning priorities with spending priorities by ensuring greater linkages between policy and budgeting. In several countries spending reviews are used to develop a structural anchoring of policy evaluations into expenditure reviews in the policy and budgetary process. The recent OECD Performance and Results Survey on COVID-19 showed that in several countries performance budgeting and spending reviews are likely to become more important in the future to help decision-makers in future budgets and to assess the effectiveness and efficiency of public policies and services. An interesting element is that while targeted and ad hoc spending reviews may be needed in the short-term, in the long-term comprehensive and whole-of-government spending reviews may be useful to support national recovery plans.

The performance monitoring in the **USA** is regulated in the Government Performance and Results Modernisation Act (GPRAMA). The Evaluation framework is established by the Foundations for

Evidence-Based Policymaking Act. Although regulated by different Acts, the OMB is responsible for the implementation of both acts.

In **Canada** the Government combined evaluation and performance measurement in one policy: the Policy on Results (2016), owned by the Treasury Board Secretariat. Prior to this unified policy, there was an evaluation Policy and Standards (2001) as well as the Management, Resources and Results Structure Policy (2008). The motivation behind the introduction of the Policy on Results, and thus joining both pillars in one policy, is about instilling a culture of evidence-based decision-making. The objectives are to improve the achievement of results across Government and to enhance the understanding of the results Government seeks to achieve, how these results are achieved and with which resources. Moreover, the Policy is based on two mutually-reinforcing pillars: evaluation and performance measurement with a more balanced approach between the two pillars than the one in place before 2016.

In the **UK**, HM Treasury and the Cabinet Office continue to push forward the public value agenda to shift the focus from inputs to outputs and outcomes. HM Treasury and the What Works Team have been developing new informal coordination mechanisms. Together, they are working on creating incentives for the use of evidence and evaluation in the spending process, and on filling the information gaps in certain areas. As a result, the creation of a new Evaluation Task Force was announced in the November 2020 Spending Review to ensure evaluation evidence is integral to spending and operational decisions by incorporating it into HM Treasury spending processes, and supporting government departments to generate evidence in priority areas with evidence gaps.

Building and maintaining support

The **challenges** for promoting policy evaluations are plenty and relate to i.a. the coordination and establishment of a whole-of-government strategy, human resources (capacity and capabilities), political interest and demand for policy evaluation, financial resources required for policy evaluations, the availability and quality of evidence/data, the mandate of the institutions, ...

In the previous sections, building on the evidence and experience from countries, many answers have been proposed to deal with these challenges. But standalone measures are not enough if evaluation is to become a regular part of the life cycle of public policies and to be conducted in a methodologically rigorous and systematic manner. An important issue therefore is how to promote a systemic approach and build and maintain **(long term) support**.

This requires working on an **evaluation culture**. Indicators for the maturity of the evaluation culture are i.a.

- Evaluation takes place in many policy domains
- Institutional arrangements in the government for conducting evaluations and disseminating their results to decision makers
- Supply of evaluators specializing in different disciplines and mastering different methods
- A national discourse on evaluation
- A profession with its own societies
- Institutional arrangements in Parliament
- Evaluation activities within the Supreme Audit Institution
- More evaluations with a focus on outcomes

To build an evaluation culture, it is i.a. crucial to (1) ensure commitment and demonstrate leadership to endorse (and pursue) a policy or strategy at the highest political level and (2) create a supportive evaluation community.

How to build commitment and leadership

The institutionalisation process of evaluations involves a wide variety of actors, many of them subject to inertia or resistance to change. It is inevitable that there will be situations in which exemptions are sought from best practice requirements. There is the risk of a 'tick a box' approach to compliance emerging over time, in which form takes precedence over substance. Without effective guidance public organisations may fail to make coordinated decisions and agree on a common vision, mission and shared goals. Without strong political interest and demand for policy evaluation, it is also difficult to find incentives for civil servants usually busy in managing day-to-day responsibilities. Such problems have been detected at various times in most jurisdictions.

Leadership is essential to overcome opposition and maintain evaluation systems in practice. Strategic leadership requires advocates among politicians and the top of the public administration. Senior officials need to promote a culture of evaluation within their organisations and be vigilant in ensuring that good practice is actually followed 'on the ground'.

- To create leadership, **outside pressure** from international organisations, benchmarks, media, industry, etc. can play an important role. Often, policy evaluation systems can be modified following **crises**: these crises can be financial, fiscal or result from major disasters

and pandemics, with concerns about the performance in a policy domain, thus calling for better understanding of what works in this policy area.

In **CAN** e.g. the Covid-19 crisis has provided momentum for greater integration and better coordination at the federal level, following difficulties with access to data.

- Also, good concrete **examples of the usefulness of evaluations** can be important to convince policy makers of the worth of policy evaluations. Pilot projects can be utilized as game changers because when carried out according to good practices and quality standards and can 'sell' the benefits of evaluation.
- Several countries also look for incentives for policy makers to invest in an evaluation framework and culture. One promising measure is to strengthen **knowledge and competences of policy makers** to analyse and interpret evidence in order to increase the demand for evaluations and the use of evaluation results. The OECD and the Joint Research Centre of the EC recently developed a mapping of the relevant skills and competencies from a demand side perspective for evidence informed policy making (see box). This work suggests that training for Senior Civil Service leadership is aimed at increasing managers' understanding of evidence informed policy making and policy evaluation, enabling them to become champions for evidence use. Intensive skills training programmes aimed at policy makers may be more focused on interrogating and assessing evidence and on using and applying it in policy making. Through such trainings, policy makers not only learn new skill but often also have increased motivation to use evidence and many become research champions and train or mentor others. Such **trainings for Senior Civil Service** can take the form of workshops, masterclasses or seminars given by national schools of government in the context of their leadership programmes or specific training courses developed by ministries or agencies.

In **CAN** the executive training in research application (EXTRA) programme provides support and development for leaders in using research. The programme is targeted towards leaders in the healthcare field.

In the **UK**, the Alliance for Useful Evidence organises an evidence masterclass.

In **BEL**, training is provided for members of parliament at the beginning of each new 5-year term.

The OECD/JRC skillset for Evidence Informed Policy Making

1. Understanding evidence informed policy making: Policy makers will be able to understand the role of evidence and its place in the policy making cycle, as well as the challenges and opportunities which come with the use of evidence.
2. Obtaining evidence: Policy makers with this skill will be able to gather existing evidence in their own policy area. They will also be able to discuss evidence gaps and commission high quality evidence to solve these gaps.
3. Interrogating and assessing evidence: Policy makers will be able to assess the reliability and appropriateness of evidence. They will have an ability to interrogate evidence by critically assessing its quality and context, using a range of techniques to challenge assumptions and biases.
4. Using and applying evidence in the policy making: Policy makers will understand their own policy context and recognise possible uses of evidence in the policy cycle. They will be proficient

in using innovative techniques like behavioural insights, and foresight to support policy design and implementation.

5. Engaging with stakeholders in evidence informed policy making: Policy makers with this skill will have strong engagement and communication skills, including ability to provide information to different types of audiences and to engage and inspire variety of stakeholders.
6. Evaluating the success of evidence informed policy making: Policy makers with this skill will understand different evaluation approaches and tools, and that evaluation should be built in the policy cycle and should serve to inform and improve Evidence informed policy making.

How to create a supportive community

Government is by far not the only actor who is involved in policy evaluation. Many evaluations are performed 'outside', sometimes not even on request of the government (e.g. universities, think tanks, advisory councils, ...) and many actors play an active role (e.g. training, oversight by civil society, networks of practice, ...)

A supportive 'epistemic community' (informal and formal networks) within and outside the public authorities (civil servants, the executive, parliament, Court of Auditors, advisory councils, social partners, NGOs, universities and centres of excellence, the media, evaluation societies, ...) has been identified as a crucial factor for the development of policy evaluation systems. These stakeholders are influential to institutionalise policy evaluation and critical in facilitating demand for evaluation. They often help to define regulatory aspects of evaluation, monitor its implementation and provide oversight.

- **Parliament** is an important actor. Governments are not able to bind the actions of their successors, so bipartisan support for regulatory policy is highly desirable if good practice is to be sustained. This requires agreement among political leaders of different parties that, while policy ideas will always be contested, the core elements of good regulatory process will not. Parliaments also sometimes initiate evaluations, but more often provide an important link in the feedback loop by using evaluation results to help inform policymaking. (see separate part on role of parliament).
- **Courts of Auditors** also have an important role in evaluation systems. In addition to performing specific evaluations, SAI can play a useful role to offer general guidance in evaluating the evaluation system as a whole and to conduct reviews of the performance of the policy evaluation system ("evaluator of evaluators" in government by auditing the effectiveness of an evaluation system and those responsible for it, e.g. **UK, BEL, ...**).
- **National associations of evaluators** often play an important role in promoting the competencies of evaluators and the quality of evaluations (**CAN, US, GER, BEL, ...**). Evaluation societies use a variety of approaches for encouraging competencies for quality analysis in the evaluation community, by establishing guidelines and increasing awareness through workshops, trainings, webinars.

In **CAN** all program evaluators in ministerial evaluation units are members of the Canadian Evaluation Society (CES), which is responsible for establishing a professional development framework for evaluators and offers training and accreditation.

- In addition, **international organisations** and development banks have played an important role for the development of evaluation systems. European Union (EU) membership and EU Structural Funds, for instance, seem to have been crucial for the dissemination and

promotion of policy evaluation in some European countries, given the strict accountability requirements related to the use of these funds.

- **Non-governmental stakeholders** including advisory councils, social partners, NGOs, universities and centres of excellence, citizens, the media and evaluation societies have an important role in promoting an evaluation culture. They often play a significant role as suppliers of evaluation (e.g. in **NLD** and **GER**), to drive and push the issues that are part of their core mission. Equally, they are also often strong users of evaluation products that can serve their purpose. Also in the UK a whole range of independent bodies support and facilitate policy evaluation, such as the What Works Network (WWN) and the Behavioural Insights Team (BIT). Research laboratories, which may depend on major British universities, consulting private companies carry out a significant part of public policy impact evaluations. Private non-profit associations or foundations, including Nesta and the Royal Statistical Society, are also involved in evidence-based policy.
- In addition, **citizens** have a significant role to play, to not only inform and engage in evaluation processes, but also as a group to promote the use of evaluations and decisions that can make a difference in their daily lives.

In the **UK** one of the main challenges of the evaluation framework is related to the ethical use of trials in social policy, particularly in the education field. This may result in opposition from key stakeholders, it can discourage the use of certain evaluation methods or hinder the design of evaluation for certain policies/projects.

Annex: Additional Benchmarking Information

This annex gives additional benchmarking information from mainly two sources:

- The OECD Survey on Policy Evaluation of 2018;
- The indicators of Regulatory Policy and Governance Surveys (iREG) 2014, 2017 and 2021. This survey draws on responses from delegates to the OECD Regulatory Policy Committee and central government officials. In 2021, the survey was responded to by 37 OECD countries, Costa Rica and the European Union.

More information on the iREG indicators can be found at oe.cd/ireg. The OECD moreover provides access to the iREG database with countries' answers:

https://qdd.oecd.org/subject.aspx?Subject=GOV_REG

Comparison of evaluation frameworks in 6 OECD countries

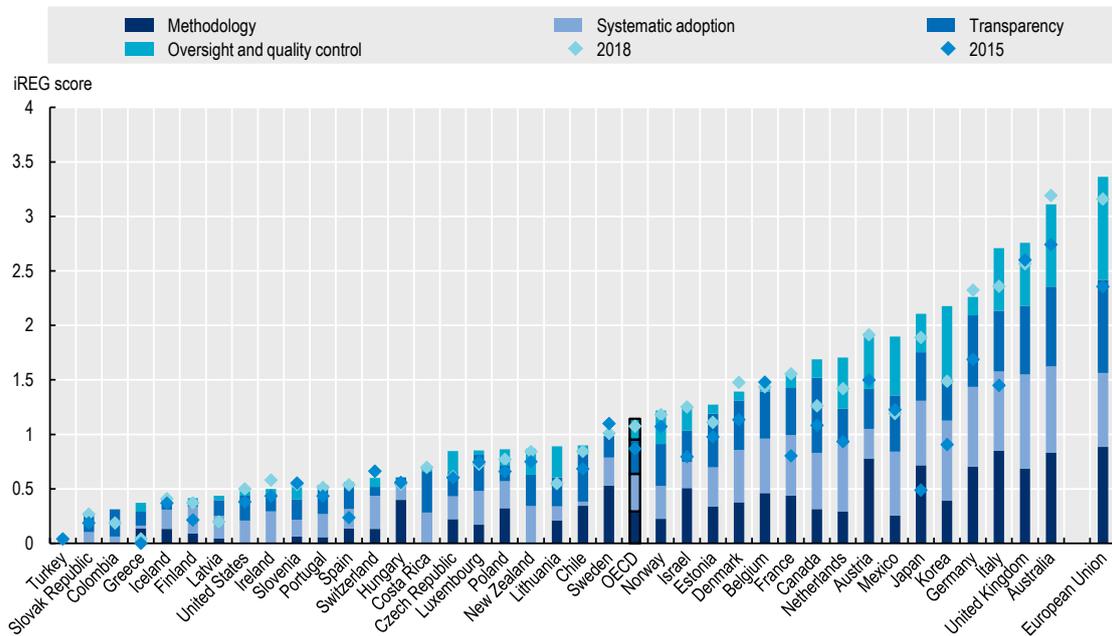
This table does not contain all relevant elements, but gives an overview of some elements that have been discussed in this report.

	Canada	Germany	Netherlands	UK	USA	Spain
Is there a formal definition of policy evaluation?	One formal definition across government	Several formal definitions	One formal definition across government	Several formal definitions	One formal definition across government	Several formal definitions
Who does the evaluations (internal/external)?	Hybrid system (combination external and internal)	Hybrid system (combination external and internal)	Hybrid system (combination external and internal)	Hybrid system (combination external and internal)	Hybrid system (combination external and internal)	Hybrid system (combination external and internal)
Are all programmes/policies evaluated within a period of 10 years or less?	Large departments submit five-year evaluation plans, selecting the policies to be evaluated, based on risks, needs and priorities. Policies defined as priorities in the government programme are evaluated.	Policies with considerable financial impact and major regulatory projects have to regularly evaluated.	Systematic coverage over a period of 5 or 7 years. This might change in the future due to the introduction of Strategic Evaluation Agenda for planning, allowing a focus on important policies.	No systematic coverage, but some policies have formal requirements to be evaluated (by insertion of an evaluation clause into laws). Proportionality principle.	Departments select the evaluation activities as part of their Evaluation Plan. Congress can mandate an evaluation for a certain programme.	No systematic coverage, but government can determine policies that have to be evaluated (cf. Spending reviews)
Evaluation guidelines available?	Yes	Not at central level	More limited	Yes	Yes	Yes

	Canada	Germany	Netherlands	UK	USA	Spain
Responsible bodies for evaluations (coordination bodies)?	Treasury Board Secretariat (TBS) has overall oversight. Privy Council Office (PCO) focuses on shorter terms political mandates.	No overarching entity responsible for all evaluations.	Each Minister is responsible within their remit for policy evaluation. The Ministry of Finance has an overview role.	No central entity responsible for an overarching framework. Her Majesty's Treasury (HMT) provides guidance and uses the spending review process as a lever.	The Office of Management and Budget (OMB) is responsible to implement the evaluation framework.	AIReF carries out evaluations within the framework of Spending reviews. The Institute for the evaluation of public policies has a mandate to create a culture of evaluation.
Responsible bodies for evaluations (level of line ministries/Departments)	Policy on Directive and Result outline responsibilities of line ministries. Every Department has their own units for evaluation. Deputy Secretaries are accountable for implementation of the framework.	Based on the subsidiarity principle (Ressortprinzip): each line ministry is responsible for evaluations in their remit. There is no uniform approach and not each Ministry has a policy evaluation unit.	The organic budget law and the Decree on Periodic Policy Evaluation devolve the responsibility for policy evaluations to the line ministries. Within the line ministries, the Directorates for Financial and Economic Affairs (FEZ) have the responsibility for implementing the evaluation framework. They do not conduct evaluations themselves.	Evaluation policy and implementation is decided at the level of individual departments.	The Evidence Act describes in detail the roles and responsibilities to be created within line ministries.	There is no clear description of the organisation of the evaluation function in the line ministries. Each Ministry is responsible for the implementation of the framework within their remit
Involvement of public institutions at arm's length or	The SAI can do performance audits. Parliament can	The National Regulatory Control Council (NKR) is the	The SAI conducts performance audits and self-initiated	The National Audit Office carries out value for money	The Government Accountability Office (GAO) and	AIReF carries out evaluations within the framework

	Canada	Germany	Netherlands	UK	USA	Spain
beyond the government?	commission evaluations.	main independent advisory body at arm's length. The SAI can conduct performance audits.	evaluations, providing the Parliament information to determine whether a policy is effective. The Planning Bureaus are an important independent body providing policies advice.	studies, which inform the Parliament. What Works Centres produce independent and practical advice in different policy areas. The independent Institute for Government provides external and independent evaluations.	Congressional Budget Office (CBO) have responsibilities relating to evaluations.	of Spending reviews required by the central government and is commissioned by Autonomous Communities to conduct evaluations. The Court of accounts produces audits (compliance, operational and performance with an evaluation approach).
Availability of evaluations?	The evaluations are publicly available by default	The evaluations are publicly available upon request.	The evaluations are publicly available by default	The evaluations are publicly available by default	Public availability depends on the agency that commissioned the evaluation.	AIReF and IEPP publish their evaluations by default.
Incorporation findings in budget cycle?	Yes	Yes	Yes	Yes	Yes	No
Is there a mechanism in place that tracks the impact of the framework?	No. There will be an evaluation of the policy in 2021. There has been an audit by the Supreme Audit Institutions the last 10 years.	No. There has been an audit by the Supreme Audit Institutions the last 10 years.	No. There has been a recent Parliamentary initiative to assess the current framework with improvement proposals.	No. There has been an audit by the Supreme Audit Institutions the last 10 years.	The U.S. Governmental Accountability Office (GAO) has the mandate to review the implementation of the GPRAMA.	No
Is there a systematic coordination between the evaluation framework and implementation at the central level and the regional level?	No	No	No	No	No	No

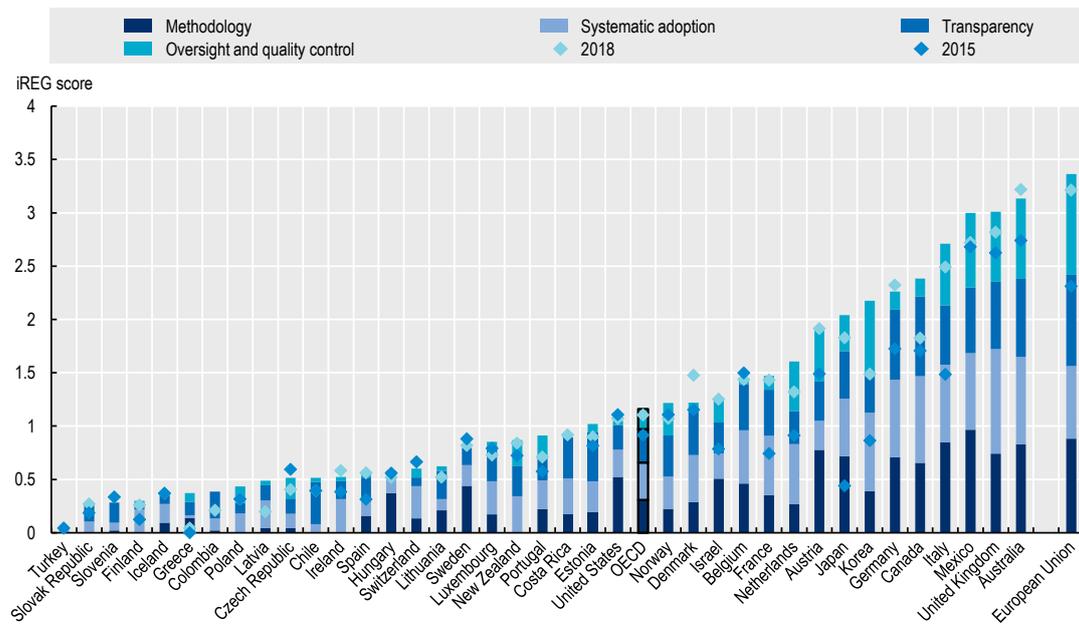
Ex post evaluation for primary laws (2021)



Note: The indicators present information on processes and practices as of 31 December 2014, 31 December 2017 and 1 January 2021 respectively, and are based on responses to the 2014, 2017 and 2021 OECD Regulatory Indicators survey, where respondents were government officials in OECD countries. The maximum score for each of the four categories is one, and the maximum aggregate score for the composite indicator is four.

Source: Indicators of Regulatory Policy and Governance Surveys 2014, 2017 and 2021.

Ex post evaluation for subordinate regulations (2021)



Note: The indicators present information on processes and practices as of 31 December 2014, 31 December 2017 and 1 January 2021 respectively, and are based on responses to the 2014, 2017 and 2021 OECD Regulatory Indicators survey, where respondents were government officials in OECD countries. The maximum score for each of the four categories is one, and the maximum aggregate score for the composite indicator is four. Source: Indicators of Regulatory Policy and Governance Surveys 2014, 2017 and 2021.

Nature of legal framework for policy evaluation

	Constitution	Primary legislation (laws or equivalent)	Secondary/subordinate legislation
Austria	○	●	○
Canada	○	●	○
Chile	○	●	○
Czech Republic	○	●	○
Germany	●	●	●
Estonia	○	●	●
France	●	●	●
Greece	○	●	●
Hungary	○	●	●
Italy	○	●	●
Japan	○	●	●
Korea	○	●	●
Latvia	○	●	●
Lithuania	○	●	●
Mexico	●	●	●
Netherlands	○	●	●
Norway	○	○	●
Poland	○	●	●
Slovakia	○	○	●
Slovenia	○	●	○
Spain	○	●	●
Switzerland	●	●	●
USA	○	●	○
OECD Total			
● Yes	4 (17%)	21 (91%)	17 (74%)
○ No	19 (83%)	2 (9%)	6 (26%)
Argentina	○	○	●
Brazil	○	○	●
Colombia	●	○	○
Costa Rica	●	●	●
Kazakhstan	○	○	●
Romania	○	○	●

Note: n=29 (23 OECD member countries). 13 countries (12 OECD members) are not included as they answered that they do not have a legal framework guiding policy evaluation across the government. Answers reflect responses to the question, "At what level is policy evaluation across government legally embedded? (Check all that apply)".

Source: OECD Survey on Policy Evaluation.

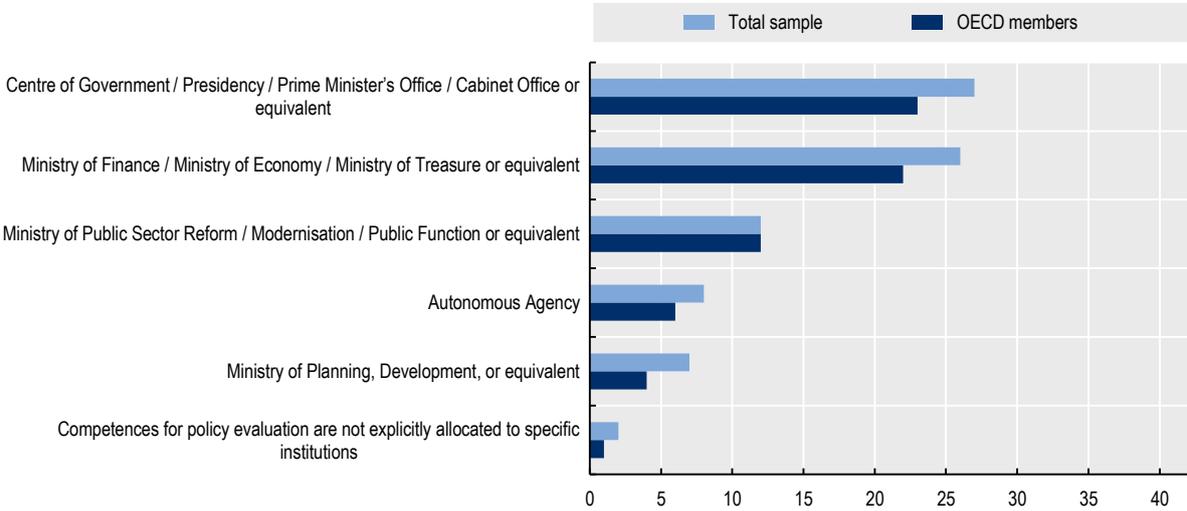
Features of the framework for policy evaluation

	Objectives or expected results of the evaluation policy	Policy areas (thematic) or programmes covered by the evaluation policy	Responsibilities of government institutions concerning policy evaluation	Requirements for government institutions to undertake regular evaluation of their policies	Standards for ethical conduct	Requirements related to the quality standards of evaluations	Requirements related to stakeholder engagement	Requirements related to evaluation reporting	Requirements related to the use of evaluation findings into policy planning making
Canada	●	○	●	●	○	○	○	●	○
Czech Republic	○	○	○	○	●	○	●	●	●
Germany	●	●	●	●	●	●	●	●	●
Estonia	●	○	●	○	●	○	●	●	○
France	●	○	●	●	○	●	○	○	○
Great Britain	●	○	●	●	●	○	○	●	●
Greece	●	●	●	●	○	●	●	●	●
Japan	●	●	●	●	○	●	○	●	●
Korea	●	●	●	●	●	●	●	●	●
Latvia	●	●	○	●	○	○	●	●	○
Lithuania	○	●	●	○	●	○	●	○	○
Mexico	●	●	●	●	●	●	●	●	●
Poland	○	●	●	●	●	●	○	●	●
Slovakia	●	●	●	○	○	●	○	○	●
Spain	●	●	●	●	●	●	○	○	○
USA	●	○	●	●	○	○	●	○	○
OECD Total									
● Yes	13 (81%)	10 (63%)	14 (88%)	12 (75%)	9 (56%)	9 (56%)	9 (56%)	11 (69%)	9 (56%)
○ No	3 (19%)	6 (37%)	2 (12%)	4 (25%)	7 (44%)	7 (44%)	7 (44%)	5 (31%)	7 (44%)
Argentina	●	●	●	●	○	○	●	●	●
Brazil	●	○	○	●	○	○	●	●	○
Colombia	●	○	●	●	○	●	●	●	●
Costa Rica	●	●	●	●	○	●	●	●	●

Note: OECD Survey on Policy Evaluation (2018).

Source: n=20 (16 OECD member countries). 21 countries (18 OECD member countries) are not included as they answered that they do not have a policy framework for organizing policy evaluation across government. Data is not available for Ireland. Answers reflect responses to the question, "Which elements do(es) the document/s referred to under Q4 and Q5 cover concerning policy evaluation across government? (Check all that apply)". The documents referred to under Q4 and Q5 are the ones stipulating a policy framework organising policy evaluation across government. The option "Other" is not included.

Institutions with competences related to policy evaluation across government



Note: n=42 (35 OECD member countries). Source: OECD (2018) Survey on Policy Evaluation

Mandate of Centre of Government for policy evaluation

	Defining and updating the evaluation	Developing guideline(s)	Providing incentives for carrying out policy evaluations	Undertaking policy evaluations	Requiring government institutions to undertake specific policy evaluations	Defining course of action for commissioning evaluations	Developing skills, competences and/or qualifications of evaluators	Developing standards for ethical conduct	Ensuring quality standards of evaluations	Promoting stakeholder engagement in evaluations	Overseeing the evaluation calendar and reporting	Promoting the use of evaluation	Serving as a knowledge centre and providing a platform for exchange	Following up on evaluation reports
Australia	○	○	○	●	●	●	○	○	○	○	○	●	○	●
Canada	●	●	○	●	●	○	●	●	○	○	●	●	●	●
Estonia	●	●	●	●	○	○	●	○	○	●	○	●	○	●
Finland	○	●	●	○	○	●	○	●	●	○	●	●	●	●
France	●	●	●	●	●	○	●	○	●	●	●	●	○	○
Germany	●	●	●	●	○	○	●	○	●	●	●	●	●	●
Great Britain	.	f.
Greece	●	●	○	○	●	●	●	○	●	●	●	●	○	●
Hungary	●	●	●	●	●	○	○	○	○	○	○	●	●	●
Iceland	○	○	○	○	○	○	○	○	○	○	○	○	●	○
Israel	●	●	●	○	●	●	●	○	●	●	●	●	●	○
Italy	●	●	●	●	●	●	○	●	○	●	○	●	●	○
Korea	●	●	●	●	●	○	●	●	●	●	●	●	●	●
Latvia	●	●	●	●	●	○	●	○	●	●	○	●	●	●
Lithuania	●	●	●	○	●	○	●	○	○	●	○	●	●	○
Netherlands	○	○	○	○	○	○	○	○	○	○	○	●	○	○
New Zealand	○	○	○	○	○	○	○	○	○	○	○	●	○	○
Portugal	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Slovakia	○	○	●	●	○	○	●	○	○	○	○	●	●	○
Slovenia	●	○	●	○	●	●	○	○	●	○	●	●	○	●
Spain	●	○	○	○	○	○	○	○	○	○	○	○	○	○
Turkey	●	○	○	●	●	○	○	○	○	○	○	○	○	○
United States	●	●	●	○	●	○	●	○	●	●	○	●	●	○
OECD Total														
● Yes	16	14	14	12	14	7	12	5	10	11	9	19	13	11
○ No	6	8	8	10	8	15	10	17	12	11	13	3	9	11

Note: n=23 OECD member countries). Answers reflect responses to the question, "Please list the duties and responsibilities of this/these institution/s related to policy evaluation across government" for the Centre of Government / Presidency / Prime Minister's Office / Cabinet Office or equivalent. Answer option "Other" is not included. Note: The UK approach to policy evaluation splits these responsibilities amongst the Cabinet Office, the Treasury, and professional analysts across government (e.g. the Government Economic Service and the Government Social Research Service), with most institutions also developing their own supplementary guidance and some form of ministerial/management response to the results. Source: OECD (2018) Survey on Policy Evaluation

Number of Regulatory oversight bodies in EU Member States

EU Member State	Number of regulatory oversight bodies in each EU Member State				
	1	2	3	4	5
Austria		✓			
Belgium		✓			
Bulgaria	✓				
Croatia		✓			
Cyprus	✓				
Czech Republic			✓		
Denmark					✓
Estonia		✓			
Finland	✓				
France			✓		
Germany			✓		
Greece		✓			
Hungary	✓				
Ireland			✓		
Italy			✓		
Latvia		✓			
Lithuania				✓	
Luxembourg	✓				
Malta			✓		
Netherlands				✓	
Poland				✓	
Portugal	✓				
Romania	✓				
Slovak Republic	✓				
Slovenia				✓	
Spain			✓		
Sweden		✓			
European Union			✓		

Note: Data is based on 27 EU Member States and the European Union.

Source: Indicators of Regulatory Policy and Governance (iREG) Survey 2021.

Guidelines and methods for policy evaluation

Country	Year	Author	Title
Australia	2014	Department of Finance	Resource Management Guidance for the Public Governance, Performance and Accountability Act 2013 RMG 131 Developing Good Performance Information
	2013	Department of the Prime Minister and Cabinet	Monitoring, review and evaluation (Cabinet Implementation Unit Toolkit)
Austria	2013	Federal Chancellery	Handbook for Performance Management
	2013	Federal Chancellery	Handbook for Performance Management
Canada	2017	Treasury Board Secretariat	Guide to Rapid Impact Evaluation
	2013	Treasury Board Secretariat	Assessing Program Resource Utilization When Evaluating Federal Programs
	2012	Treasury Board Secretariat	Theory-Based Approaches to Evaluation: Concepts and Practices
	2010	Treasury Board Secretariat	Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies
	2019	Treasury Board Secretariat	Integrating Gender-Based Analysis Plus into Evaluation: A Primer
Czech Republic	2016	Min.Regional Development	Methodological guidance for evaluations in the 2014-2020 programming period
Estonia	2012	Government of Estonia	Methodology of Impact Assessment
	2012	Estonian Evaluation Association	Good Public Evaluation Code of Practice
	2011	Government of Estonia	Good Public Engagement Code of Practice
Finland	Annually	Council of State/PMO	Government's Annual Plan for research, foresight and evaluation
France	2017	France Stratégie	Guide de l'évaluation socio-économique des investissements publics
	2016	France Stratégie	Comment évaluer l'impact des politiques publiques : un guide l'usage des décideurs et des praticiens
	2010	INSEE	Méthodes économétriques pour l'évaluation des politiques publiques
		Youth Experimentation Fund	Guide méthodologique relative aux évaluations du FEJ
Germany	2016	Federal Ministry of the Interior	Handbuch für Organisationsuntersuchungen und Personalbedarfsermittlung
	2011	Federal Ministry of Finance	Arbeitsanleitung Einführung in Wirtschaftlichkeitsuntersuchungen
	2007	Federal Ministry of the Interior	Empfehlungen für interne Revisionen in der Bundesverwaltung
	2000	Federal Ministry for Family Affairs	Zielgeführte Evaluation von Programmen
Great Britain	2018	HM Treasury	Guide to developing the project business case
	2018	HM Treasury	Guide to developing the programme business case
	2018	Better Regulation Executive	Better regulation framework
	2015	Government Social	Government Social Research Publication Protocol

Country	Year	Author	Title
		Research Service (HM Treasury)	
Greece	2018	Secretariat General of the Government	Manual of Inter-Ministerial Coordination
	2015	European Commission	Better Regulation Toolbox
	2015	European Commission	Better Regulation Guidelines
Ireland	2018	Department of Public Expenditure & Reform	Public Spending Code
Italy	2018	PCM	Guidelines RIA
	2017	NUVAP	Guidelines for ex post and ongoing evaluations: requesting and using evaluations
	2017	NUVAP	GL ex post & ongoing evaluations
	2017	Decree President Council of Ministers	Guidelines for ex-ante and ex-post impact analysis of regulatory acts
	2015	NUVAP	GL Evaluation Plans
	2015	NUVAP	Evaluation Plans 2014-2020: general orientation and a short guide on available guidance
Japan	2005	Ministry of Internal Affairs and Communications	Basic Guidelines for Implementing Policy Evaluation
	2005	Ministry of Internal Affairs and Communications	Policy Evaluation Implementation Guidelines
	2007	Ministry of Internal Affairs and Communications	Implementation Guidelines for Policy Evaluation of Regulations
	2010	Ministry of Internal Affairs and Communications	Guidelines for Publication of Information on Policy Evaluation
	2010	Ministry of Internal Affairs and Communications	Implementation Guidelines for Policy Evaluation Pertaining to Special Taxation Measures
	2013	Ministry of Internal Affairs and Communications	Target Management-based Policy Evaluation Implementation Guidelines
Korea	2017	Office for Government Policy Coordination	Government Performance Evaluation Manual
Lithuania	2011	Ministry of Finance	Recommendations on Implementation of Programs Evaluation Methodology
	2010	Ministry of Finance	Evaluation of EU structural assistance: Methodological guidance
Latvia	2018	Ministry of Finance	Instruction on Analysis of the Execution of State Budget
	2016	Cross-sectoral Coordination Centre	Manual on Policy Making
Mexico	2007	Ministry of Finance,	General Guidelines for the Evaluation of Federal Programs

Country	Year	Author	Title
		Ministry of Public Administration, National Council for the Evaluation of Social Development Policy	
Norway	2018	DFO	Strategic and systematic use of evaluation in management/governance
	2009	Ministry of Justice	Evaluation of laws
	2007	DFO	Evaluation of central governmental grants
	2005	Ministry of Finance	Guidelines to carry out evaluations
New Zealand	2018	Superu	Making sense of evidence: A guide to using evidence in policy
	2015	Superu	Evaluation Standards for People Commissioning, Using, Participating in, or Conducting Evaluations
Poland	2015	Ministry of Economic Development	Wytyczne w zakresie ewaluacji polityki spjnosci
Portugal	2018	Juris App	Manual
Spain	2015	AEVAL	Practical guide for the design and implementation of public policy evaluations
	2007	AECID (Spanish Agency for International Cooperation for Development)	Spanish Cooperation Evaluation Management Manual
Slovakia	2016	Ministry of Finance	Value for money
	Forthcoming	Ministry of Economy	RIA 2020
Switzerland	2015	Federal office of justice	Planifier une évaluation, en assurer le suivi et en valoriser les résultats
	2013	State secretariat for economic affairs	Analyse d'impact de la réglementation - Manuel
	2012	Federal office of justice	Recommandations de l'Office fédéral de la justice pour la formulation des clauses d'évaluation
	2005	Federal office of justice	Guide de l'évaluation de l'efficacité la Confédération
United States	2018	OMB	A-11 section 200: an overview of the Federal Performance Framework
	2018	OMB	M-18-04: Monitoring and Evaluation guidelines for agencies that administer foreign assistance
	2019	OMB	M-19-23: guidelines for the implementation of the Foundations for Evidence-Based Policymaking Act of 2018
	2020	OMB	M-20-12 Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices
Argentina	2018	National Council for the Coordination of Social Policies	Resolución No.310 Lineamientos de MyE
	2018	Cabinet Office	Resolución No.212/18 Plan Anual de MyE
Brazil	2018	Civil House, Ministry of Finance, Ministry of Planning, Ministry	Public Policies Evaluation: practical guide for ex ante analysis

Country	Year	Author	Title
		of Transparency and Comptroller General	
	2017	Ministry of Social Development	"How to promote impact evaluation in social programs"
	2015	Ministry of Transparency and Comptroller General	Methodology Manual for Evaluating Government Programs Execution
	2014	Federal Court of Auditors	Referential for Governance Evaluation in Public Policies
Colombia	2018	Departamento Administrativo de Planeación Nacional	Guide for the evaluation of public policies
Costa Rica	2018	Ministry of National Planning and Economic Policy	Guide for the use of evaluations: guidelines for its implementation and follow-up on recommendations.
	2017	Ministry of National Planning and Economic Policy	Manual of evaluation for public interventions
	2017	Ministry of National Planning and Economic Policy	Guide on the approach of gender equality and human rights in evaluation: guidelines for its incorporation into the evaluation process.
	2017	Ministry of National Planning and Economic Policy	Guide of evaluability: methodological guidelines for the evaluability of public interventions.
Kazakhstan	2017	Ministry of National Economy	State planning system

Source: OECD Survey on Policy Evaluation (2018).

Use of sunseting clauses and automatic evaluation requirements in EU Member States

	Do regulations include 'sunseting' clauses?		Do regulations include automatic evaluation requirements?	
	Primary laws	Subordinate regulations	Primary laws	Subordinate regulations
Austria	For some primary laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations
Belgium	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Bulgaria	Never	Never	Never	Never
Croatia	Never	Never	Never	Never
Cyprus	Never	Never	Never	Never
Czech Republic	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Denmark	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Estonia	Never	Never	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Finland	For some primary Laws/ subordinate regulations	Never	Never	Never
France	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Germany	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations
Greece	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	Never	Never
Hungary	Never	Never	For all primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations
Ireland	Never	For some primary Laws/ subordinate regulations	Never	Never
Italy	Never	Never	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations
Latvia	Never	Never	Never	Never
Lithuania	For some primary Laws/ subordinate regulations	Never	For some primary Laws/ subordinate regulations	Never
Luxembourg	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	Never	Never
Malta	Never	Never	Never	Never
Netherlands	For some primary Laws/ subordinate regulations	Never	For some primary Laws/ subordinate regulations	Never
Poland	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	Never
Portugal	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	Never	Never
Romania	Never	Never	Never	Never
Slovak Republic	Never	Never	Never	Never
Slovenia	Never	Never	Never	Never
Spain	For some primary Laws/ subordinate regulations	Never	Never	Never
Sweden	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	Never	For some primary Laws/ subordinate regulations
European Union	For some primary Laws/ subordinate regulations	For some primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations	For all primary Laws/ subordinate regulations

- For all primary Laws/ subordinate regulations
- For major primary Laws/ subordinate regulations
- For some primary Laws/ subordinate regulations
- Never

Note: Data is based on 27 EU Member States and the European Union.
 Source: Indicators of Regulatory Policy and Governance (iREG) Survey 2021.

Mechanisms in EU Member States to assess the achievement of underlying policy goals

	When designing laws, policy makers have processes in place to identify the achievement of a regulation's goals.		Do <i>ex post</i> evaluations contain by default an assessment of whether the underlying policy goals of regulation have been achieved?	
	Primary laws	Subordinate regulations	Primary laws	Subordinate regulations
Austria	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations
Belgium	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations
Bulgaria	Never	Never	For some regulations/ Some <i>ex post</i> evaluations	Never
Croatia	Never	Never	For some regulations/ Some <i>ex post</i> evaluations	Never
Cyprus	Never	Never	Never	Never
Czech Republic	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	Never
Denmark	Never	Never	For some regulations/ Some <i>ex post</i> evaluations	Never
Estonia	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	Never	For some regulations/ Some <i>ex post</i> evaluations	Never
Finland	For some regulations/ Some <i>ex post</i> evaluations	Never	Never	Never
France	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations
Germany	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations
Greece	Never	Never	Never	Never
Hungary	Never	Never	Never	Never
Ireland	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	Never	Never
Italy	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations
Latvia	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	Never	Never
Lithuania	For some regulations/ Some <i>ex post</i> evaluations	Never	Never	Never
Luxembourg	Never	Never	Never	Never
Malta	Never	For some regulations/ Some <i>ex post</i> evaluations	Never	Never
Netherlands	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations
Poland	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	Never
Portugal	Never	Never	Never	Never
Romania	Never	Never	Never	Never
Slovak Republic	For some regulations/ Some <i>ex post</i> evaluations	For some regulations/ Some <i>ex post</i> evaluations	Never	Never
Slovenia	For all primary regulations/ All <i>ex post</i> evaluations	Never	Never	Never
Spain	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations	Never	Never
Sweden	Never	Never	For some regulations/ Some <i>ex post</i> evaluations	Never
European Union	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For major regulations/ <i>Ex post</i> evaluations regarding major regulations	For all primary regulations/ All <i>ex post</i> evaluations	For all primary regulations/ All <i>ex post</i> evaluations

- For all primary regulations/ All *ex post* evaluations
- For major regulations/ *Ex post* evaluations regarding major regulations
- For some regulations/ Some *ex post* evaluations
- Never

Note: Data is based on 27 EU Member States and the European Union.

Source: Indicators of Regulatory Policy and Governance (iREG) Survey 2021.

Mechanisms for the promotion of quality

	Provisions expressed in a policy/legal framework	Guidelines for policy evaluation across government	Competence requirements for evaluators	Peer review (internal/external) of evaluations	Systematic and meta-evaluations	Other
Australia	○	●	○	○	○	●
Austria	●	●	●	●	○	●
Belgium	○	○	○	○	○	○
Canada	●	●	●	●	○	●
Chile	●	○	●	○	○	○
Czech Republic	○	●	○	○	○	○
Denmark	○	○	○	○	○	●
Estonia	●	●	●	○	○	○
Finland	●	●	●	●	○	○
France	●	●	●	●	○	○
Germany	●	●	○	●	●	●
Great Britain	○	●	●	●	●	○
Greece	●	●	●	○	○	○
Hungary	○	○	○	○	○	●
Iceland	○	○	○	○	○	○
Ireland	○	●	○	○	○	●
Israel	○	○	○	○	○	○
Italy	○	●	○	○	○	●
Japan	●	●	●	●	○	○
Korea	●	●	●	○	○	○
Latvia	●	●	○	○	○	○
Lithuania	○	●	○	○	○	○
Mexico	●	●	●	●	●	○
Netherlands	○	●	○	●	○	○
New Zealand	○	●	○	○	○	○
Norway	○	●	○	○	○	○
Poland	●	●	○	○	●	●
Portugal	○	●	○	○	●	○
Slovakia	○	●	●	○	○	○
Slovenia	○	○	○	○	○	○
Spain	○	●	○	●	○	●
Sweden	○	○	○	○	○	○
Switzerland	○	●	○	○	○	○
Turkey	○	○	○	○	○	○
United States	●	●	●	○	○	○
OECD Total						
● Yes	14	26	13	10	5	10
○ No	21	9	22	25	30	25

Note: n=35 OECD member countries. 12 OECD member countries answered that there are no mechanisms to ensure the quality of evaluations across government. Answers reflect responses to the questions "How does your government ensure the quality of evaluations across government?" and "Are there guidelines available to support the implementation of policy evaluation across government?". Systematic and meta-evaluations refer to the evaluations designed to aggregate findings from a series of evaluations. In Germany, regular exchange take place within the network of evaluation units of development cooperation agencies and externally through the OECD DAC evalnet. Hungary has a consultation process to review the evaluations, In Ireland each Accounting Officer is responsible for ensuring compliance with the Public Spending Code in their Department/Office. Italy has different mechanisms to improve the quality of the evaluations as part of the National Evaluation system such as steering groups. Poland has a system of assessment of quality of conducted evaluations in the policy evaluation guidelines.

Source: OECD Survey on Policy Evaluation (2018)

Quality standards included in evaluation guidelines

	Technical Quality of evaluations						Good Governance of evaluations		None Of The Above
	Identification and design of evaluation approaches	Course of action for commissioning evaluations	Establishment of a calendar for policy evaluation	Identification of human and financial resources	Design of data collection methods	Quality standards of evaluations	Independence of the evaluations	Ethical conduct of evaluations	
Australia	○	○	○	●	○	○	○	○	○
Austria	○	○	○	●	●	○	○	○	○
Canada	●	○	○	○	●	●	●	●	○
Czech Republic	●	○	○	○	●	●	●	●	○
Estonia	●	●	○	●	●	●	●	●	○
Finland	○	●	●	○	○	●	●	●	○
France	●	○	○	○	○	○	○	○	○
Germany	●	●	●	●	●	●	●	●	○
Great Britain	●	○	●	●	●	●	●	●	○
Greece	●	●	●	●	●	●	●	○	○
Ireland	●	○	○	○	●	●	●	○	○
Italy	○	●	○	●	○	○	●	○	○
Japan	●	○	●	○	●	●	○	○	○
Korea	●	○	●	○	●	●	○	○	○
Latvia	●	●	●	●	●	●	○	○	○
Lithuania	●	○	○	●	●	○	●	○	○
Mexico	●	●	●	○	○	●	●	●	○
Netherlands	○	○	○	○	○	●	○	○	○
New Zealand	●	●	○	●	●	●	●	●	○
Norway	●	○	○	●	●	○	○	○	○
Poland	○	○	●	○	●	●	●	○	○
Portugal	○	●	○	○	○	○	○	○	○
Slovakia	●	○	○	○	○	●	●	○	○
Spain	●	●	○	●	○	●	●	●	○
Switzerland	○	○	●	●	●	●	●	●	○
United States	●	○	●	●	●	●	●	●	○
OECD Total									
● Yes	18	10	11	14	17	19	17	11	0
○ No	8	16	15	12	9	7	9	15	26

Note: n=26 OECD member countries. 9 OECD member countries answered that they do not have guidelines to support the implementation of policy evaluation across government. Answers reflect responses to the question, "Do the guidelines contain specific guidance related to the: (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018)

Methods used by countries in the evaluation of government-wide policy priorities

	Regression/econometrics/structural equation modelling	Randomised controlled trials	Qualitative Comparative Analysis	Contribution analysis	(Comparative) case studies	Process tracing	Theory-based evaluation
Australia	○	○	○	○	○	○	○
Austria	○	○	○	○	○	●	○
Canada	●	○	●	●	●	○	●
Chile	○	○	●	○	●	○	○
Estonia	●	○	●	●	○	○	○
Finland	○	○	●	●	●	●	○
France	●	●	●	○	●	○	○
Germany	●	●	●	●	●	●	●
Great Britain	●	●	●	●	●	●	●
Greece	●	○	●	●	●	●	○
Hungary	○	○	●	○	○	●	○
Ireland	●	●	●	○	●	○	○
Israel	○	○	○	○	○	●	○
Italy	○	●	○	○	●	○	○
Japan	○	○	○	○	●	○	●
Korea	○	○	○	○	○	○	○
Lithuania	●○	○	●	○	●	○	●○
Latvia	●	○	●	○	○	○	●
Mexico	●	●	●	●	○	●	●
Poland	●	●	●	●	●	●	●
Portugal	○	○	○	○	○	●	○
Slovakia	●	○	○	○	○	○	●
Spain	●	○	●	○	○	●	○
Sweden	●	○	○	○	●	○	●
OECD Total							
● Yes	13	7	15	8	13	11	9
○ No	11	17	9	16	11	13	15
Argentina	○	○	○	○	○	○	○
Brazil	●	○	●	○	●	●	●
Colombia	●	●	●	○	●	○	○
Costa Rica	●	○	●	●	●	●	●
Romania	●	○	○	○	○	○	○

Note: For the main institution on government-wide policy priorities n=29. 4 countries answered that they do not have government-wide policy priorities. Moreover, 9 countries answered that they do not evaluate their government-wide policy priorities. Answers reflect responses to the question, "Which quantitative or qualitative methods of impact evaluation have been used over the past three years for the evaluation of government wide policy priorities? (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018)

National Evaluation Societies in OECD countries

Country	Name of the Society Network	Website
Australia	Australian Evaluation Society	https://www.aes.asn.au/
Austria	Austrian-German Evaluation Association	https://www.degeval.org/home/
Belgium	Flemish Evaluation Platform	http://www.evaluatieplatform.be/VEP/index.htm
Canada	Canadian Evaluation Society	https://evaluationcanada.ca/
Chile	Red Chilena de Evaluación	http://www.evaluacionpoliticaspUBLICAS.com/
Czech Republic	Czech Evaluation Society	https://czecheval.cz/
Denmark	Danish Evaluation Society	http://danskevalueringsselskab.dk/
Estonia	Estonian Evaluation Society	http://www.praxis.ee/vana/index.php-id=1029.html
Finland	Finnish Evaluation Society	http://www.sayfes.fi/in-english/
France	French Évaluation Society (SFE)	http://www.sfe-asso.fr/
Germany	German Evaluation Society (DeGEval)	https://www.degeval.org/en/home/
Greece	Hellenic Evaluation Society	http://www.hellenicevaluation.org/index.php/el/
Hungary	Hungarian Evaluation Society	https://www.europeanevaluation.org/content/hungarian-evaluation-society
Iceland		
Ireland	Irish Evaluation Network	https://www.dcu.ie/eqi/ien/index.shtml
Israel	Israeli Association for Program Evaluation	http://www.iape.org.il/en_index.asp
Italy	Italian Evaluation Association	http://valutazioneitaliana.eu/
Japan	Japan Evaluation Society (JES)	http://evaluationjp.org/english/index.html
Korea	Korean Evaluation Association	http://www.valuation.or.kr/
Latvia	Latvian Evaluation Society (LATES)	http://www.izvertesana.lv/en/about-us/
Lithuania		
Luxembourg	Luxembourg Evaluation and Foresight Society	http://solep.lu/
Mexico	Academia Nacional de Evaluadores Mexicanos (ACEVAL)	http://aceval.org/
Netherlands	Dutch Evaluation Society (VIDE)	https://www.videnet.nl/
New Zealand	Aotearoa New Zealand Evaluation Association (ANZEA)	https://www.anzea.org.nz/
Norway	Norwegian Evaluation Society	http://norskevalueringssforening.no/
Poland	Polish Evaluation Society	http://pte.org.pl/
Portugal	Portugal Evaluation Association (AvalPortugal)	https://avalportugal.wordpress.com/
Slovak Republic	Slovak Evaluation Society (SES)	http://www.evaluacia.sk/en/
Slovenia	Slovenian Evaluation Society	https://www.sdeval.si/eng
Spain	Spanish Public Policy Evaluation Society (SEE)	http://www.sociedadevaluacion.org/
Sweden	Swedish Evaluation Society	http://svuf.nu/
Switzerland	Swiss Evaluation Society (SEVAL)	https://www.seval.ch/
Turkey	Turkish Monitoring and Evaluation Society (TMES)	https://www.ived.org.tr/english
United Kingdom	UK Evaluation Society	https://www.evaluation.org.uk/about-us/
United States	American Evaluation Association	https://www.eval.org/

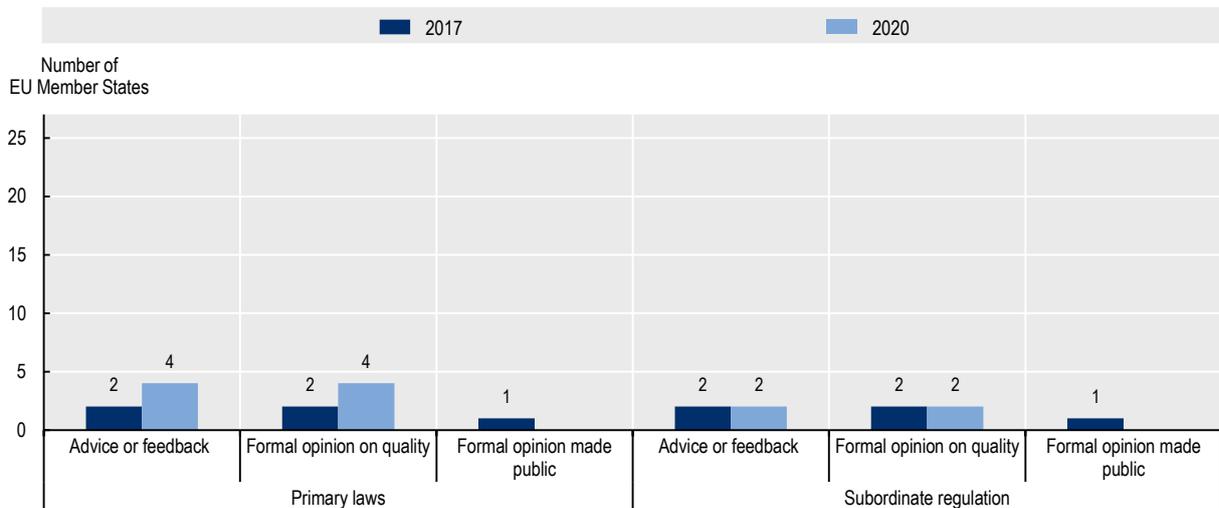
Mechanisms to promote the use of evaluations

	Management response mechanism in place	Incorporation of findings into the budget cycle		A rating / grading system	Coordination platform to promote use of evidence	Discussions of findings at the Council of Ministers	No specific initiatives in place
Australia	○	●		○	○	○	○
Austria	●	●		○	○	○	○
Belgium	○	○		○	○	○	●
Canada	●	●		○	●	○	○
Chile	○	●		○	○	●	○
Czech Republic	○	○		○	○	○	●
Denmark	○	○		○	○	○	○
Estonia	●	●		○	●	○	○
Finland	○	●		○	●	●	○
France	○	●		○	○	○	●
Germany	○	●		○	●	●	○
Great Britain	○	●		○	●	○	○
Greece	●	●		○	○	●	○
Hungary	○	○		○	○	●	○
Iceland	○	○		○	○	○	●
Ireland	○	●		○	●	○	○
Israel	○	●		○	○	○	○
Italy	○	○		○	○	○	○
Japan	●	●		●	●	●	○
Korea	●	○		○	○	●	○
Latvia	●	●		○	●	●	○
Lithuania	○	●		○	○	●	○
Mexico	●	●		●	●	●	○
Netherlands	○	●		○	○	○	○
New Zealand	●	○		○	○	○	○
Norway	○	○		○	●	○	○
Poland	○	○		●	●	○	○
Portugal	○	●		○	○	●	○
Slovakia	●	●		○	○	○	○
Slovenia	○	○		○	○	○	●
Spain	○	○		○	○	○	●
Sweden	○	●		○	○	○	○
Switzerland	○	○		○	○	○	●
Turkey	○	○		○	○	○	●
United States	○	●		○	●	○	○
OECD Total							
● Yes	10	21		3	12	11	8
○ No	25	14		32	23	24	27

Note: n=35 OECD member countries. Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations". The option "Other" is not included. A rating/grading system refers to classify the robustness of evidence provided and recommendations derived from the policy evaluations exists.

Source: OECD Survey on Policy Evaluation (2018)

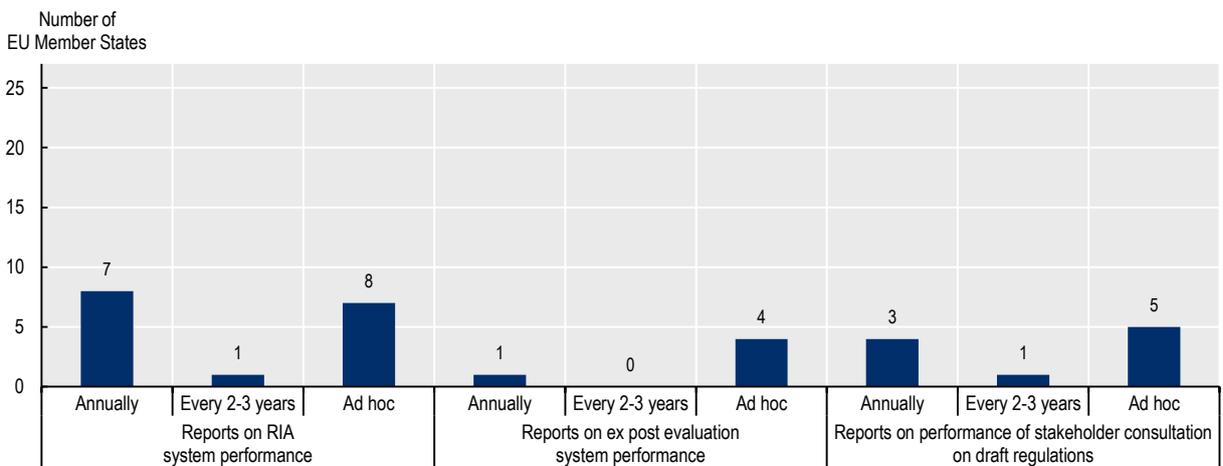
Quality control of ex post evaluation in EU Member States



Note: Data is based on 27 EU Member States.

Source: Indicators of Regulatory Policy and Governance (iREG) Surveys 2017 and 2021.

Efforts to assess and report on performance in EU Member States



Note: Data is based on 27 EU Member States.

Source: Indicators of Regulatory Policy and Governance (iREG) Survey 2021.

References

- Cabinet Office UK (2018), What Works Network Membership criteria, http://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/747965/WW-membership-paper_Oct2018.pdf .
- Canada Treasury Board (2016), Policy on Results, <http://www.tbs-sct.gc.ca/pol/doceng.aspx?id=31300> .
- CEE (2010), Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies, <http://www.canada.ca/en/treasury-board-secretariat/services/auditevaluation/centre-excellence-evaluation/guide-developing-performance-measurementstrategies.html> .
- CEE (2017), Guide to Rapid Impact Evaluation, <http://www.canada.ca/en/treasury-boardsecretariat/services/audit-evaluation/centre-excellence-evaluation/guide-rapid-impacetevaluation.html>
- CEE (2017), Theory-Based Approaches to Evaluation: Concepts and Practices, <http://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centreexcellence-evaluation/theory-based-approaches-evaluation-concepts-practices.html> .
- European Commission (2021), Better Regulation Guidelines, https://ec.europa.eu/info/sites/default/files/swd2021_305_en.pdf
- European Commission (2021), Better Regulation Toolbox, https://ec.europa.eu/info/sites/default/files/br_toolbox-nov_2021_en_0.pdf.
- European Parliamentary Research Service (2020), *Better Regulation practices in national parliaments*, <http://dx.doi.org/10.2861/06573>
- HM Treasury (2018), Guidance The Green Book: appraisal and evaluation in central government, <http://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-incentral-government> .
- HM Treasury (2020), The Green Book: appraisal and evaluation in central government, <http://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-incentral> -
- HM Treasury (2020), The Magenta Book, <http://www.gov.uk/government/publications/themagenta-book>.
- Ministerie van Financiën Nederland (2021), Strategische Evaluatie Agenda, http://www.rijksbegroting.nl/2021/voorbereiding/begroting,kst282764_10.html.
- NKR (2018), Germany: Less Bureaucracy, More Digital Services, Better Regulation. Let's get toit! - Annual Report 2018, <http://www.normenkontrollrat.bund.de/resource/blob/656764/1548226/a53ca395512296087f96a45d0b839d44/2018-11-09-jahresbericht-englisch-data.pdf?download=1>.
- OECD (1997). Best practice guidelines for evaluation. [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PUMA/PAC\(97\)2&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PUMA/PAC(97)2&docLanguage=En)

OECD (1997). Promoting the use of programme evaluation. [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PUMA/PAC\(97\)3&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PUMA/PAC(97)3&docLanguage=En)

OECD (2005). OECD Guiding Principles for Regulatory Quality and Performance, Paris. <https://www.oecd.org/competition/sectors/37318586.pdf>

OECD (2008). Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0362>

OECD (2009) Better Regulation in Europe: an OECD assessment of Regulatory Capacity in the 15 Original Member States of the EU, Paris, OECD. <https://www.oecd.org/gov/regulatory-policy/better-regulation-in-europe-the-eu-15-project.htm>

OECD (2012), Recommendation of the Council on Regulatory Policy and Governance, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264209022-en>

OECD (2014) Ex post evaluation of regulation: an overview of the notion and of its international practices, Paris, OECD. GOV/RPC(2014)14/ANN5

OECD (2014), Centre Stage Driving Better Policies from the Centre of Government, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm\(2014\)3&docLanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm(2014)3&docLanguage=en)

OECD (2015), Recommendation of the Council on Budgetary Governance. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0410>

OECD (2016), Supreme Audit Institutions and Good Governance: Oversight, Insight and Foresight, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/9789264263871-en> .

OECD (2017), Making policy evaluation work. <https://doi.org/10.1787/c9bb055f-en>

OECD (2017). Survey on Policy Monitoring and Evaluation. [https://one.oecd.org/document/GOV/PGC\(2017\)29/ANN1/en/pdf](https://one.oecd.org/document/GOV/PGC(2017)29/ANN1/en/pdf)

OECD (2018), Building Capacity for Evidence Informed Policy Making: Towards a Baseline Skill Set, <http://www.oecd.org/gov/building-capacity-for-evidence-informed-policymaking.pdf>

OECD (2018), Centre Stage 2- The organisation and functions of the centre of government in OECD countries, <https://www.oecd.org/gov/centre-stage-2.pdf>.

OECD (2018), Ex-post assessment of regulation: Practices and lessons from OECD countries, OECD Publishing, Paris. [https://www.oecd.org/centrodemexico/publicaciones/OECD%20\(2018\)%20Expost%20assessment.pdf](https://www.oecd.org/centrodemexico/publicaciones/OECD%20(2018)%20Expost%20assessment.pdf)

OECD (2018), OECD Regulatory Policy Outlook 2018, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264303072-en> .

OECD (2018), Policy Framework on Sound Public Governance, <http://www.oecd.org/gov/draft-policy-framework-on-sound-public-governance.pdf>

OECD (2019), Better Regulation Practices across the European Union, OECD Publishing, Paris. <https://doi.org/10.1787/9789264311732-en>

OECD (2019), Budgeting and Public Expenditures in OECD Countries 2019, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264307957-en>.

OECD (2019), OECD Good Practices for Performance Budgeting, OECD Publishing, Paris, <https://doi.org/10.1787/c90b0305-en>

OECD (2019), The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>

OECD (2020), Building Capacity for Evidence-Informed Policy-Making: Lessons from Country Experiences, <http://doi.org/10.1787/86331250-en> .

OECD (2020), Ex Ante Regulatory Impact Assessment: Netherlands, <http://www.oecd.org/gov/regulatory-policy/RIA-Netherlands.pdf> .

OECD (2020), Improving Governance with Policy Evaluation: Lessons From Country Experiences, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/89b1577d-en>.

OECD (2020), Reviewing the Stock of Regulation, OECD Best Practice Principles for Regulatory Policy, OECD Publishing, Paris, <https://doi.org/10.1787/1a8f33bc-en>.

OECD (2020), The Irish Government Economic and Evaluation Service: Using Evidence-Informed Policy Making to Improve Performance, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/cdda3cb0-en>.

OECD (2020), Toward robust policy evaluation, OECD Publishing, <http://doi.org/10.1787/c03e01b3-en> .

OECD (2020). Policy Framework on Sound Public Governance, <https://doi.org/10.1787/c03e01b3-en>

OECD (2021), Evaluation framework and practices: a comparative analysis of 5 OECD countries.

OECD (2021), OECD Regulatory Policy Outlook 2021, OECD Publishing, Paris, <https://dx.doi.org/10.1787/38b0fdb1-en>.

OECD (2021). Draft OECD Best Practices for Spending Reviews. GOV/SBO(2021)9

OECD (2021). Enhancing access to research data during crises: lessons learned from the COVID-19 pandemic. Paris, OECD. [https://one.oecd.org/document/DSTI/STP/GSF\(2021\)13/FINAL/en/pdf](https://one.oecd.org/document/DSTI/STP/GSF(2021)13/FINAL/en/pdf)

OECD (2021). Recommendation of the Council concerning Access to Research Data from Public Funding, OECD/LEGAL/0347 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0467>

OECD (2021). Recommendation of the Council for Agile Regulatory Governance to Harness Innovation. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0464>

OECD (2021). Recommendation of the Council on Enhancing Access to and Sharing of Data <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>

OECD (2022). Better Regulation Practices across the European Union GOV/RPC(2021)23/REV2

OECD (2022). Draft Recommendation on Public Policy Evaluation. COM/GOV/PGC/SBO(2021)2

OMB (2020), Circular No. A-11, <http://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf>.

OMB (2020), Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices,
<http://www.whitehouse.gov/wpcontent/uploads/2020/03/M-20-12.pdf>

RDA (2020). COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: <https://doi.org/10.15497/RDA00052>

SERV (2010). Rapport Wetgevingsprocedures, -structuren en -instrumenten in Vlaanderen: een evaluatie van 10 jaar wetgevingsbeleid in Vlaanderen, Brussel, SERV, 10 maart 2010

SERV (2015). Achtergronddocument Ex post decreetsevaluatie. Brussel, 21 september 2015.
https://www.serv.be/sites/default/files/documenten/SERV_20150921_decreetsevaluatie_ADV_i.pdf

SERV, VLIR (2021). Betere overheidsdata voor sterk beleidsonderzoek en goed bestuur. Brussel, 13 september 2021.
https://www.serv.be/sites/default/files/documenten/SERV_20210913_SERV_VLIR_oproep_betere_overheidsdata_ADV.pdf

Stockmann, R., W. Meyer and L. Taube (eds.) (2020), Spain, Palgrave Macmillan,
<http://doi.org/10.1007/978-3-030-32284-7> .

Tweede Kamer (2019), Tweede Kamer, Brief van de Vaste Commissie voor Financiën voor de verbetering verantwoording en begroting (Vergaderjaar 2019-2020, nr. 176),
<http://www.tweedekamer.nl>.

US Congress (2019), Foundations for Evidence-Based Policymaking Act of 2018,
<http://www.congress.gov/bill/115th-congress/house-bill/4174/text> .

Van der Knaap, Peter en Rudi Turksema (2016). Oog voor succes: een pleidooi voor positieve beleidsevaluatie. Boom Bestuurskunde, 2016, nr. 2.
<https://tijdschriften.boombestuurskunde.nl/tijdschrift/bk/2016/2>

Van der Knaap, Peter en Rudi Turksema (2019). Positieve beleidsevaluatie: hoe evaluatieonderzoek kan bijdragen aan beter beleid. Boom Bestuurskunde, 2019, nr. 3.
<https://tijdschriften.boombestuurskunde.nl/tijdschrift/bk/2019/3>

Van der Knaap, Peter, Valérie Pattyn en Dick Hanemaayer (2020), Beleidsevaluatie in theorie en praktijk, Den Haag: Boom bestuurskunde.
<https://www.boomdenhaag.nl/webshop/beleidsevaluatie-in-theorie-en-praktijk>