Getting Serious About Evaluation Systems

Thomas Delahais, tdelahais@quadrant-conseil.fr

This is a position paper written in the context of the ECA Cohesion Conference 2023: Evaluating EU Cohesion policy – challenges and opportunities *(23 October 2023). The views expressed are the author's own. They are based on the author's experience as an evaluator of cohesion programmes and in support of managing authorities, including in collaboration with DG REGIO's Helpdesk experts.*

Discussions about how to make evaluation useful for policymaking very often start with considerations of process, methodology and data. What indicators would be meaningful, how to ensure that sufficient data is available, and so on. We argue that this approach, which focuses on delivery, may not be able to address issues of use. Indeed, use does not depend so much on the availability of evaluations, as on a shared culture of evidence; on institutional processes and policy-making tools that require the availability of evidence; on the capacity of stakeholders to deal with evidence, and so on. The concept of 'evaluation system' however (Rist & Stame 2006), though sometimes used as a synonym for 'evaluation function', offers many opportunities to think about monitoring and evaluation in terms of use first, at all levels of the system (EU, managing authorities, beneficiaries, etc.). This means reframing evaluation as being primarily a knowledge brokering activity in the policy sphere. Our focus will be on evaluation systems at the level of managing authorities, which is the foundation on which the whole edifice of the evaluation of the EU cohesion policy rests, to identify some of the key challenges that they are facing today as they move towards more use-oriented evaluation systems and make some suggestions in this regard.

MAs have delivered evaluations, now where is use?

There is a broad consensus, in European cohesion policy circles, that there is much to learn from evaluations on the success or failure of policy interventions, and that what is learned from evaluations has an important role to play in better design and implementation of cohesion programmes. An important feature of the EU cohesion policy is indeed the requirement for evaluation, which has come with incentives, procedures, guidance, control, and so on. In many countries, this requirement has played an important role in the institutionalisation of evaluation, particularly in the countries that joined the EU in the 2000s (Stern 2009), leading to the cohesion policy being probably one of the most evaluated in the world (Naldini 2018).

In previous programming periods (2000-2006, 2007-2013), managing authorities have often been busy putting in place the processes to carry out the evaluations required by the Commission. This is not an easy task. They have largely dealt with this task in a very rational, bureaucratic manner, developing an 'evaluation function' within their organisation. This function has grown processes to prioritize, plan, organise, conduct and disseminate evaluations. These processes now exist and have delivered: since 2015, almost three thousand evaluations, commissioned by managing authorities (MAs), have been published, including a thousand impact evaluations (EC 2023).

As evaluation functions in MAs have matured, however, they have turned more and more their attention to the question of use. And on this aspect, the outlook is less positive. Many of the evaluation officers who participate in the Cohesion Policy Evaluation Network, who attend the DG REGIO summer schools, who ask for support from the helpdesk - these are the people I've met and talked to - are frustrated by the contrast between efforts devoted to evaluation and actual use – in terms of stakeholders learning about the intervention and its context, changing some of their views on policy issues, relying on the generated evidence in an array of policy-related activities, and perhaps in terms of changes, small or big, in how policies are designed, monitored, implemented and so on.

Is it a cohesion issue? Bachtler and Wren have, almost 20 years ago, insisted on the intrinsic complexities of the evaluation of cohesion policy, which are as many barriers to use (2006): it has to serve too many objectives at once (those of MAs, of sectoral officers, of local partners, of all European institutions). The cohesion policy is subject to intense debates about what constitute its rationale, its effects, and so on. It is probably one of the most complex policies which exist, in which common orientations are delivered through a large variety of institutional arrangements in many different settings. In the end, these features also mean that actual use depends on each of these arrangements and settings. Despite a general concern expressed here, you can find "pockets of good evaluation practice" (Stern 2009) everywhere, and it is the same for evaluation use. Alternatively, there are places where evaluation is mere administrative annoyance. It is probably a chimera to have evaluations being used everywhere, but it is a battle that can be won place by place.

Is it a quality issue? It sure is[1]. But thanks to the Helpdesk, which assesses all published evaluations since 2015, we know that some evaluations are good, and some are bad. Impact evaluations, which have been made mandatory in the previous programming cycle, tend to be of better methodological quality, and their policy implications are clearer. Yet, even with these, even with the good ones in general, we are all too often not seeing the kind of effects on policy-making and policies themselves that could be expected. (EP 2020) Some evaluation functions, very skilled, very proactive (recently in the Czech Republic, in Estonia, in many regions all across Europe...), have found that evaluation could be useful in their current institutional setting, but at the price of considerable energy, often beyond what they can afford on a daily basis, given the amount of time and work they have to do. Where evaluation is used more routinely, this is the result of a shared culture of evaluation (or evidence) that has existed before and beyond cohesion policy - and which cannot be easily replicated by applying procedures.

Let's not make the cohesion policy an exception: evaluators have always been worried about the usefulness of their work, and they are probably more vocal about their concerns than other professions. We very often hear that evaluations would be more useful if they were of better quality

[1] A. Naldini points at several issues, such as a bureaucratic approach to evaluation, inadequate administrative processes to launch, conduct and follow up evaluations; poor methodological design; insufficient capacity both on the demand and supply side; insufficient independence; and so on (2018).

– if they were choosing the right indicators, if sufficient data was available, and so on. Yet, together with C. Radaelli and C. Dunlop, we argue that in the effort to push evaluation use, "there is a tendency to focus on the design of evaluation at the expense of learning: get the evaluation 'right' and learning will follow" (2023). We would like here to take a different perspective, explain why use is, in general, more unlikely than likely in the current situation; and explore approaches which start from use and move upstream towards what evaluation and evaluation systems can do to foster it.

There are some 'structural' obstacles to evaluation use

Evaluation is generally not a 'natural' practice in organisations, and many MAs have grown an evaluation function in response to the external EC pressure. Of course, evaluation functions are shaped by this demand. This means, that, for instance, they tend to be set up as independent structure with *de facto* limited relations with the rest of the institution (organised in sectoral silos), in order to be able to provide credible feedback to the EC. In that regard, many MAs equate 'using evaluation' with 'complying with European requirements' and 'displaying accountability'. **What matters first is that an evaluation function exists and that evaluation reports are produced**. Evaluation reports also tend to focus on questions of execution and delivery (how much of the budget has been programmed, whether the roads have been built, whether SMEs received a grant...), which are those on which MAs are under pressure to show results and have developed mechanisms of adaptation (this is an example of "*Evaluation as a chip in the negotiating game*", Stern 2009). Despite many attempts to the contrary, evaluations focus less on questions of actual implementation (or reception by stakeholders), impact, or EU added value, which, in theory, would be of more interest to policy stakeholders interested in improving policies[2].

Therefore, evaluation functions have often been geared towards delivery first, and they did so by planning evaluation work. Although they can try to make evaluations timely – by guessing when evaluations might be useful – they also have to allocate evaluation work over a certain span of time. In practice, this means that, some will come much too early to estimate any impact (especially infrastructures), some too late for decision-making cycles, and so on. There is only so many days in a year... In the last programming period, evaluation functions have had to cope with increasing demands: that all financial instruments be subject to ex ante evaluation; that impact evaluations be carried out to assess how the Structural Funds have contributed to the objectives of each priority[3]; that additional objects, such as the Youth Employment Initiative and, more recently, the Just Transition Mechanism, be subject to their own evaluation. Evaluation is also required to contribute to the new performance-based funding mechanisms, linking EU financial support to pre-defined outputs and results (ECA 2021). And let's not mention the COVID evaluations! In these conditions, **what matters most is to cope with the required delivery rhythm**.

The other 'structural' consequence of evaluation as a response to external pressure is that evaluators are more likely to be seen by their colleagues as 'inquisitors' or 'spies' than as 'comrades' (Coulson 1988)[4]. **Evaluation officers often tell how their relationship with operational departments is one-sided**: they have to reach for them, ask for how they could be useful to them, offer to involve them in evaluation governance, communicate results – to raise only mild interest in many

[2] *"Evaluations shall be carried out to improve the quality of the design and implementation of programmes" (Art. 54(1) CPR) but MAs are not held accountable for such improvements, which are difficult to track. Though accountability on financial matters works as a mechanism (MAs have to prove that they have resolved pointed issues), on use, it is expected to work "as a virtue" (Bovens 2010).*

[3] *This was abandoned in the new programming cycle, to the benefit of a more specific, use-oriented approach.*

[4] *When the contribution of EU funding is marginal, they might as well be ignored.*

cases. Colleagues genuinely requesting help from the Evaluation department (something which is not always allowed…) is actually good evidence of a developing evaluation culture in the organisation! This has enormous consequences, not only in terms of use, but on all facets of evaluation. A typical example comes with monitoring systems. Identifying good indicators 'in theory' is not hard. But in practice, the evaluation function depends on programme managers to discuss about their meaningfulness and possible use; integrate data collection in operations (in administrative forms, in contractualisation process, in feedback reports…); keep doing it in the long term; include the discussion of indicators in the programme governance; provide information in usable format when needed; and so on. This is mission: impossible without a relationship of trust and a shared culture of evidence.

A third structural problem lies in evaluation capacity. Though it is possible to have an independent team within the organisation in charge of doing the evaluation, in practice the evaluation of cohesion programmes is mostly a market activity. This is by relying on external service providers that MAs are supposed to evaluate such a wide range of interventions, at different moments of the policy cycle, with different methods, etc. Yet, evaluation officers all across Europe keep complaining about the lack of adapted skills on the market, and their lack of choice: smaller countries may witness situations of quasi-monopoly that reduce emulation among consultants. This is particularly visible on the question of impact, where the European Commission has emitted several methodological recommendations over the years. But in many countries, service providers have adopted the new language of impact, while keeping an essentially business-as-usual approach; they have promised counterfactual studies without considering seriously the related data requirements, or the adequacy of the approach to the expressed needs; they have said that they would evaluate along a theory of change, without being familiar with the concept, and so on. As mentioned before, there are good and bad evaluation reports of cohesion policies, like everywhere. But MAs can too rarely count on the market to provide the kind of excellent evaluations that could move the lines in terms of use.

Genuine evaluation systems are focused on use rather than on delivery

These problems, though we are seeing them in the framework of the EU cohesion policy, can be seen elsewhere too. As we've said before, evaluation use is a relentless question. And "conventional" evaluation functions, in reality, should not be judged on use, but primarily on their coverage of programme and the quality of evaluation products, because that is what they aim for.

There are, in general, many obstacles to the use of explicit knowledge in organisations[5]. Proponents of evidence-based policy making often form the implicit assumption that the policy sphere and the individuals that populate it are rational – but they're not (Allison 1971). Actually, it is fair to say that (cohesion) policy-making is unlikely to be *based* on evidence, which is but one of many ingredients in this process (Weiss 1980). But we can expect that evidence (and in particular, evaluations) can *inform* decisions, illuminate debates, and that, given the right conditions, stakeholders who are aware of policy evidence can be influenced by it. The question is: when these conditions do not already exist, how can they be created?

[5] *"Explicit knowledge" covers different forms of knowledge which are the process of a transparent and systematic process and usually take a formal form. This includes academic research, evaluation, expertise, technical database, inter alia (Fillol et al. 2004). It is opposed to tacit knowledge. For an overview of the different perspectives on the relationship between knowledge and public action, see Delahais & Devaux-Spatarakis 2022.*

In the 2000s, R. Rist and N. Stame (2006) popularised the concept of evaluation systems[6]. This was not just a fancy new name for evaluation functions. Thinking in systems, evaluation is necessarily part of the systems it seeks to observe, measure, and serve, and interacts with them. Therefore, **the question of what is a *good* evaluation system stems from what it is supposed to do, from its expected impact, rather than from what it delivers**.

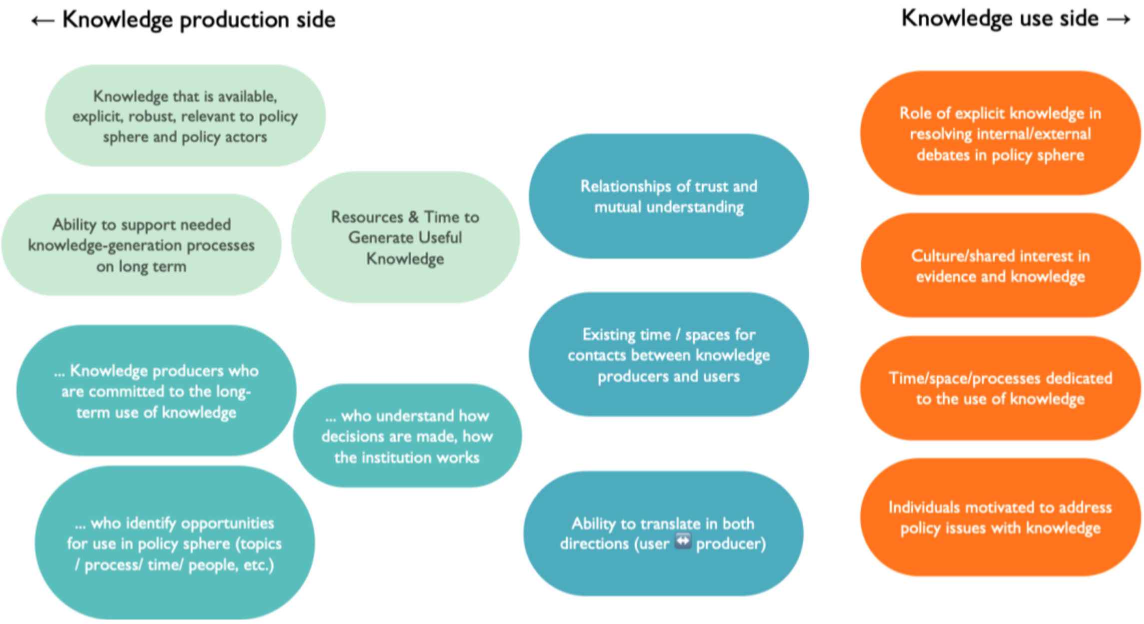
In this view, the key initial question for setting up an evaluation system is: "What are the steps of the policy cycle, what are the governance, policy-making, decision-making processes, debates, stakeholders... that could benefit from evaluation knowledge and evaluation processes? The second question is: "What needs to be done to address and strengthen these uses?" The answer given by Rist and Stame was that one-off, individual evaluations were unlikely to be influential. Of course, the process and the results can help to see problems differently and can provide helpful ideas for solving observed problems. But in the long run, knowledge production and decision-making processes live on different timelines, and without specific action, their encounter is the exception rather than the rule. Rather, organisations and people learn from the accumulation of information and from their ability to retrieve it at the right time for the right needs. They therefore require permanent streams of information that combine monitoring data, evaluations or studies, that are planned for long-term use, and that can be shaped, distributed, reinforced or curtailed according to anticipated needs and responses received.

We should add, however, that use doesn't depend primarily on the availability or timeliness of information, and not even on methodological quality, as traditionally defined in terms of validity or robustness[7]; but on the existence of processes in which evidence is or could be used. What matters then is to contribute to the frequent and long-term interactions between potential uses and users on the one hand, and information producers on the other. We suggest that a further consequence of thinking in terms of evaluation systems is therefore **to frame them as processes of knowledge management and knowledge brokering** (Ingold & Monaghan 2016, Dagenais & Ridde 2020), where important efforts are devoted to investigating decision systems to identify where evidence is or could be used; what evidence would be useful, for whom, to do what; to convincing stakeholders of the value of evaluations and other sources of explicit knowledge over other sources of knowledge (such as tacit knowledge or anecdotal information in support of their view, etc.); to filling and curating in the long term, with multiple evaluations, a "reservoir of knowledge" (Hanney et al. 2004) that is likely to be needed to support reflection, debate or decision-making; to creating conditions for knowledge producers and users to meet and trust each other; and to synthesising, reframing and translating evidence at the right time to meet current needs (see Fig. below).

[6] *F. Leeuw and J-E. Furubo described further the notion of evaluation systems, through 4 criteria (2008): a distinctive epistemological perspective; organisational responsibility; permanence; and a focus on the intended use of evaluations.*

[7] *'Good' evaluations not being more used than 'bad' ones is a widespread factor of frustration among evaluation officers, who tend to believe in quality as a factor of use. Rather, quality should be seen as a necessary but insufficient condition for use leading to positive outcomes. On the other extreme, there are example of dysfunctional evaluation systems which lead to significant use of poor quality evaluation products (Delahais & Lacouette 2019).*

FIGURE 1: PRECONDITIONS TO USE (BASED ON [CAPANO & MALANDRINO 2022](#))



In this view, monitoring and evaluation needs (as well as expectations in terms of robustness) ultimately derive from the expected uses, rather than from 'objective' characteristics of the interventions being evaluated. For instance, a Head of department can be interested in SME response to calls for projects (who, to what extent, why) because they are concerned with their region's economic dynamism. The Chamber of Commerce, on the other hand, can be interested in foresight (what do we learn from the evaluation about resilience of the industrial fabric to future global shocks) rather than in the disbursement rate or in reaching the 'official' objectives of a programme. It is the role of evaluation officers to collect and reframe these potential uses to address them, based on their understanding of the institution, ongoing dialogue with stakeholders, measurement capabilities, and the availability of capacity within the organisation and in the market.

Going towards evaluation systems will come with compromises

We can now see that the notion of evaluation systems has deep implications... and also that there is no simple solution to implement them. Of course, it is important to provide help in improving singular evaluation processes. Evaluations can be made more relevant and credible to users by fostering a continuous learning process within the evaluation, by investigating important issues in the political agenda or in the local governance, by scoping the evaluation accordingly; by multiplying evaluation methods and tools to integrate multiple perspectives, by better considering policy implications, and so on. Evaluation officers and policy makers already have access to many resources in support of this continuous effort towards increased quality. This includes guidelines, contacts with geographical or evaluation EC officers, the evaluation network, the interact programme, and the helpdesk, inter alia.

However, **reinforcing evaluation use implies, on the long run, a whole-institution approach**[8], including political and administrative support, management and governance changes, engagement of policy stakeholders, long-term planning, and so on. Evaluation processes need to fit into the organisation, which means that 1) there is a need to depart from evaluation as a reaction to

[8] *The whole-institution approach comes from the world of sustainability studies. See UNESCO-UNEVOC, 2017.*

external pressure; and 2) that the right institutional architecture is the one that is coherent with the organisation and its ways of policy-making – there is no 'one size fits all' approach. The consequences are also important in terms of capacity. A good evaluation officer in a conventional evaluation function might be a lawyer (good at procurement) or an evaluator (good at assessing methodological quality on paper); in an evaluation system it might rather be a "networker" (someone who has contacts at all levels of the administration, among policy stakeholders and in the market), an information curator, a facilitator, an educator, a sensemaker, etc. These are profiles that are harder to find, but are actually those dearly needed for 'future-proof public management' (Radaelli 2021).

In practice, this will come with compromises. T. Kupiec, D. Woktowicz & K. Olejniczak, in a series of papers, have identified how Central and Eastern European countries managing authorities (2019, 2023) have evolved very different structures to cater to different needs: a central unit targeted at strategic needs and/or external accountability; or decentralized units, with or without a coordination unit, closer to operational needs. They both come with pros and cons. For instance, a centralised function has access to better skills, with more full-time specialists. It usually delivers less evaluations, but with a potential to be of better quality. This function can serve strategic goals, which can be to make better decisions, but also to shield the administration from external criticism, for instance. But there is a risk that this function will be cut from the field and the actual information needs of those in charge of designing and implementing policies. On the other hand, decentralised units can be much closer from the operational services and serve many needs; but they are often composed of part-time agents, less skilled in evaluations and subject to many demands that crowd out the evaluation function. There would be a collective benefit to discuss these different options and what they are good at.

Some suggestions in the short and long term

An evidence culture, a shared willingness to learn from evaluation and to improve policies accordingly, is not something that can be decided on a whim - and evaluation alone, isolated and understaffed as it often is, cannot create it on its own. However, the new programming period and the evaluation plans are an opportunity to lay the foundations for such a culture and reap the benefits in the future. Let's not miss it! Here are some suggestions for using evaluation systems to this end, especially where this culture is currently lacking.

Ensure that the main ingredients for having a chance of being useful are here. **Capacity is number one on the list**: expertise and experience in evaluation processes and methodologies come to mind first, but there is only so much an evaluator can do to foster use. What is also needed is knowledge of the subject matter, knowledge of the policy-making system, networks within the institution, and capacity in knowledge brokerage. Capacity needs to be considered at an overall level, both inside and outside the institution. Continuously train colleagues in operational departments in evaluation and evaluative thinking (Quadrant Conseil 2023). Train public and private partners; and train service providers (or make sure they do). And learn from the conduct of evaluation themselves.

Establish and maintain an interface with potential users, within and perhaps outside the institution. Policy actors usually do not know at the outset how an evaluation could be useful to them, except in very general terms. Even with continuous dialogue, it may take a long time for a good topic to emerge, and even longer for the timing to be right. The good news is that evaluation is a perfect excuse to communicate frequently with the stakeholders. This interface needs to be ongoing (before, during and after evaluations), and where such a place doesn't exist yet it can become the setting in which evidence is discussed in the institution – an arguably desirable role for evaluation systems.

Identify favourable conditions and contexts. C. Dunlop and C. Radaelli have identified specific contexts in which evaluations are most likely to trigger learning (op.cit): 1) contexts dominated by experts where highly technical knowledge can be used to reduce uncertainty; 2) contexts in which contentious issues are addressed, where there is a need to highlight different perspectives and have them dialogue; 3) contexts in which different actors and interests have concerns that the evaluation can help voice, and which can ultimately shape the policy differently; 4) and hierarchical contexts where evaluation can sanction poor performance. Use the process and governance of evaluation planning to identify such contexts, and make the best of it.

At a very basic level, **finding potential users should be a starting point** for most evaluation processes. In practice, this could be managers looking for feedback from beneficiaries, officers in charge of innovative programmes where there is pressure to show that things work, sensitive issues that would benefit from a participatory process supported by empirical evidence... Stakeholders are more likely to be interested by an evaluation of the policies on which they work (which may be partly funded by the cohesion policy) than of a funding (one among others for them): Expanding the scope of evaluations, Looking at policies through transversal issues such as equity, discrimination, gender, which people care about, are some of the strategies that evaluators can use to raise interest. More broadly, evaluation efforts in cohesion policy would benefit from a better consideration of what is on top of the political agenda, and what is coming next: How is the current strategy still relevant when looking at the climate emergency, the new geopolitical context, etc.?

Concentrate efforts where conditions are most favourable. Because of the EC requirements to evaluate against all policy objectives, we know in advance that some evaluations will have a very low potential for use. But there are also contexts in which the stars might align for learning and other uses. These contexts deserve extra effort, with a view to showcase evaluation usefulness. Carefully compose the governance of these evaluations, bring in additional experts, ensure quality, plan for communicating about the evaluation results, but also the evaluation process. In the end, these evaluations can serve as a 'proof of concept', towards colleagues or partners, and may lead to some of them wanting to replicate the process on their own topic.

To seize opportunities as they arise, **evaluation plans need to become adaptive**. Initial planning should be seen not so much as a compliance tool or a to-do list, as it is still too often (EP 2020), but as a simulation of what might be done. Although it is often mentioned that the provisions of the plan are indicative and can be modified later (which is a good thing), it is important to make explicit what the arrangements will be that will allow these modifications to be guided by issues of relevance of evaluations and their usefulness. But what matters most is how the function plans to identify uses and adapt to them: What is the governance, what are the arrangements that will help the evaluation function identify and react to new needs, new possible uses? If they have the capacity, evaluation functions should also be prepared to venture outside their usual remit, for example to provide strategic advice or very operational solutions, to engage in the design process, to provide lessons for new programmes, and so on.

Similarly, flexibility is needed in evaluation processes. It is common to say that quality starts with good terms of reference. **Commissioning evaluability studies** (or doing them in-house) is a good way to improve ToRs. They are used to develop a theory of change for the evaluation, to identify needs, the range of methods to address them, and what data is available to do what. This avoids the all-too-common consultant excuse ("sorry, we promised the world, but no data!") and helps make the best decisions in terms of evaluative approach. It should be possible this way, for example, to choose an evaluative approach that fits expected uses rather than 'schools of evaluative thoughts' (Delahais et al. 2022).

But it doesn't stop there! **It is essential to invest in the early stages of the evaluation process**. The real learning potential of an evaluation very often emerges after data collection has begun; it should still be possible at this stage to reassess priorities and identify ways in which the evaluation could be useful in practice, for example through additional scoping or identifying the right unit of analysis. For instance, an evaluation which focuses on a national or regional policy, rather than uniquely on investments funded by cohesion policy; or that investigates the effects of infrastructures funding under the previous programming period, rather than the current one; or that considers adaptation to future challenges rather than past ones, is often more likely to be used than the conventional ESIF evaluation.

As a consequence, **evaluation is unlikely to be useful unless it is a collaborative process**. We have known for a long time that people often learn more from the process of evaluation than from the results stricto sensu. This is because the process exposes them to other perspectives than their own; makes them reframe and understand the problems they are trying to address differently; refocuses them on issues of implementation and impact rather than delivery; makes them make judgements about what is and is not a good result that they would not have made otherwise; and starts them thinking about how to address problems as they arise, rather than as a final step in the evaluation. Collaboration is easier said than done (Delahais 2022), but it is a cornerstone of evaluation uptake.

Evaluations, however, do not begin or end with the production of reports. Through their interface with users, through the evaluation plan, through opportunities as they arise, evaluation functions need to constantly identify policy issues that they could help address with evidence. They may respond to these needs with new reports or studies, but more often it will be with existing knowledge (from previous evaluations, from research, etc.). One of the key functions of evaluation systems should be to **maintain a body of existing, relevant knowledge and use it when needed**, rather than to try to harmonise evaluation processes and decisions. There are many ways to deliver knowledge outside of evaluation reports, through in-person formal and informal meetings, training, through notes, and even using internal and external social networks.

In parallel, evaluation functions need to work in the long term and follow their own agenda. They can develop and maintain monitoring systems with departments or units with which they have developed a relationship of trust (otherwise this is usually very difficult). They can check whether important issues identified in one evaluation are also found in others. They can develop knowledge products from the information they collect, which are not evaluations but are still useful evidence.

Conclusion: the renewed need for a shared culture of evidence

In the almost 30 years since evaluation became a requirement of cohesion policy, MAs across Europe have learned to carry out evaluations effectively and efficiently. And yet here we are with the unrelenting question of use. We have argued that there is only so much that individual evaluations can do; where the culture of using evidence in policy-making is absent, evaluations are unlikely to be useful. But we could also claim that policy stakeholders using evidence can be found everywhere; there are just unevenly distributed. If evaluations are to "improve the quality of programme design and implementation" (as per the CPR), then a change of approach is needed, starting from those users and uses and expanding from there. In practice, this will be a 'policy of small steps', with a lot of 'muddling through' (Lindblom op.cit.); but one that is more likely to be successful than a policy aiming at making evaluations used against hostile context.

In the new programming period, several of evaluation requirements (including ex ante evaluations, systematic impact evaluations, programme specific indicators in particular) and some others have been introduced (more frequent data transmission, mandatory use of Better Regulation criteria)[9]. It is unclear whether these changes may affect use at the member state level. On the other hand, 2014-2020 innovations such as the innovation plan, the experience gained in diversifying evaluation approaches, the capacity & networking tools which have been deployed, could now show their full effect. There is much to be expected, for instance, of the nascent "community of practice" around the evaluation of cohesion policy (EP 2020), in which evaluation stakeholders can share their experience and progressively raise their game in terms of use.

Beyond the regulatory and capacity support, the Commission has a role to play in detecting and showcasing useful evaluation processes that are happening today; in experimenting in innovative approaches fostering use; or in fostering coordinated approaches to test and learn from real case scenarios. In the two last programming periods, a lot of space was devoted to the debate about the best method to evaluate impact, leading sometimes to complex machineries. We agree that all those approaches should be part of the evaluator's toolbox, but that the question of 'what approach is best?' should be replaced by "what approach could answer what question towards what kind of use?" (Delahais & Quadrant Conseil 2017, Delahais et al. 2022)—taking a turn that has already been taken by many international organisations. Proposing simple and robust methodologies to answer routine questions, such as that of relevance, coherence or added value, is also a domain in which the Commission could thrive. As C. Hutchings rightfully said at the closing conference of the European Evaluation Society in 2016, we need *Fiats working in real-world contexts, more than methodological Ferraris*.

Audit organisations such as the ECA can help too. Performance audits usually aim at verifying whether the set of necessary conditions related to an expected result can be observed. There is a work to be done today to develop an auditing framework focused on the conditions for use of evidence within European institutions and member states; one that would not merely stop at an evaluation function, but that would consider administrative, organisational, governance, policy-making processes; whether there is room in these processes for the collaboration between evidence producers and users; for reconsidering prejudice on what works or not; for improving implementation based on empirical feedback; and so on. Such a framework would need to go farther than institutional provisions, to look rather at actual practices; and allow for whole-institution recommendations towards evidence-informed policy making.

Ultimately, a culture of evidence takes time to build and is a fragile thing. Nor is it a political utopia: major 'modernisation' initiatives of member states are often accompanied by evaluation efforts, including in the area of structural funds (e.g. recently in Portugal). The recognition of the uncertainty in which policy actors must operate; the intractable difficulties associated with super-wicked problems (Levin et al. 2012) such as climate change, biodiversity collapse, rising inequalities and so on, also call for a new, 'post-normal' role for evaluation, to which only robust evaluation systems may be able to contribute (Schwandt 2019). It is still time to prepare.

[9] For a full run-down and discussion of these evolutions, see Naldini 2018.