

Towards a Typology of Innovation System Evaluation Practices. Evidence from EU Member States

Mart Laatsit, Copenhagen Business School, Denmark. mla.dbp@cbs.dk (corresponding author)

Susana Borrás, Copenhagen Business School, Denmark.

Abstract:

An increasing number of European countries are introducing a novel form of evaluation which assesses innovation systems performance as a whole and the role of policies herein. This move towards ‘system evaluation’ comes from the need of an encompassing overview of how policy initiatives interact and affect innovation processes in the whole country. However, there is a conceptual inconclusiveness in the scholarly literature about what constitutes a ‘system evaluation’, which results in the inability to grasp country-level empirical observations meaningfully. Furthermore, there is a lack of evidence about the specific features that characterize actual system evaluation practices across countries. With the purpose of addressing both gaps this paper develops a clear conceptual framework about this phenomenon, which is then used to gather solid empirical evidence from the EU’s 28 member states. The preliminary findings show that a number of the 28 countries have developed ‘system evaluations’, and that these are unevenly distributed across a four-fold typology. The conclusions summarize the findings and point towards next steps in a research agenda that looks into the epistemological perspective of these evaluations (where innovation indicators play a central role in different national practices of ‘system evaluations’), as well as the policy learning that comes out from these ‘system evaluations’ in relation to the indicators used and the organizational features at national level.

Keywords: Evaluation, innovation policy, innovation indicators, European Union, innovation system

1. Introduction

There is a growing attention among policy-makers in Europe to move from individual research and innovation program evaluations towards system evaluations. That is, evaluations which are more encompassing, analysing innovation policy initiatives in their entirety and examining their role in innovation systems. Some countries have already started experimenting with system evaluations, carrying out the first steps into system-level analyses of their innovation policies (a few recent examples include Denmark, Finland, Estonia); while many more countries are contemplating it. These efforts are supported by international organisations, such as European Commission and OECD, who are trying to introduce new assessment tools and provide policy advice and assistance. ‘System evaluations’ are naturally rather ambitious exercises launched and owned by policy-makers in search of concrete clues to see the overall performance and evolution of the set of innovation policy instruments in the context of the innovation system. This new interest is possibly related to the growing focus of policy-makers on the issues of impact and in particular, on the socio-economic effects of complex combinations of policy instruments and public initiatives.

In contrast to the growing interest among policy-makers, there is still a limited scholarly attention to this matter. As we will review in the next section, the few scholarly publications about ‘system evaluation’ have addressed this phenomenon in rather broad terms. These scholars have provided a stimulating constellation of meanings and understandings associated to this new form of evaluations. However, they have not been able to overcome the inconclusiveness attached to this concept. This conceptual inconclusiveness has resulted in the inability to grasp country-level empirical observations meaningfully because too broad concepts are poorly equipped to operationalize the analysis and data collection. Hence, it is not surprising that the scarce literature in this topic has not provided empirical evidence about the current ‘state of affairs’ in terms of whether countries have engaged and developed ‘system evaluation’ practices or not; and if so, how have they gone about it. These are two important gaps.

With the purpose of addressing both gaps this paper develops a clear conceptual framework about this phenomenon, which is then used to gather solid empirical evidence from the EU’s 28 member states. It looks at an under-researched strand in the field of innovation studies, namely system evaluation. With more than 30 years of national system of innovation theoretical approaches, there is still a considerable lack of knowledge on how to assess the effectiveness of complex policy systems. Only a few examples of academic research on system evaluation (Arnold 2004) (Edler, Ebserberger et al. 2008) (Magro and Wilson 2013) have been published over the past decade, which is hardly a match to the increasing demands for novel approaches from policy-makers. In addition, we need to develop a better understanding about the innovation indicators that are being used by countries, considering whether these are systematically assessing systems of innovation (Jordan, Hage et al. 2008) (Hage, Jordan et al. 2007). All-in-all, pursuing new avenues for innovation policies must be matched with an adequate understanding of the efficiency and effectiveness of the current policy-

mixes, and about different ways of engaging in system-wide evaluation exercises that provide better basis for decision-making.

The paper proceeds as follows. Next section identifies the nature of the issue, by characterizing the gap between the needs of policy-making and the existing scholarly frameworks. In so doing, it reviews the scant literature on system evaluations examining in detail how this literature has defined the contours of this notion. Since the literature is rather inconclusive, section three of this paper engages in a process of providing a specific definition of the concept ‘system evaluation’, which is based in three concrete attributes. Thereafter, it defines two items that will help classifying different types of system evaluations according to two distinct dimensions, namely, whether the organizational responsibility of conducting system evaluations is mainly internal to the executive or external (using experts or international organizations), and whether the system evaluation is conducted routinely or in an ad-hoc manner. This conceptual framework will be used to collect empirical evidence in the 28 European Union countries, helping to establish how far system evaluations have spread across Europe, and what types of system evaluations are currently in place. Section four describes the methodological issues related to the choice of analytical methods and data collection, here including its reliability and validity. This paper reports preliminary findings from 30 interviews in EU countries. Section five reports the findings indicating which EU28 countries have currently developed system evaluations, and which ones not according to our conceptual definition. We present the data results, which are at the moment preliminary, showing that a number of the 28 countries have developed ‘system evaluations’. Next section, section six reports the findings in terms of the two dimensions mentioned above, which provide a four-fold typology of system evaluations. The preliminary findings show that there is an uneven distribution of our 28 cases across this typology. The final section summarizes the findings, answers the research question, and delves into a discussion about the existing frameworks for system evaluation across European countries. The latter is highly relevant for further considerations about the role of indicators and impact assessment methodologies in innovation policy making.

2. ‘System Evaluation’ in the Literature: Unclear Conceptual Boundaries

During the past two decades the innovation system approach has gained substantive endorsement among scholars and policy-makers alike. This approach sees innovation as a complex social process of cumulative nature, embedded in complex institutional and organizational national contexts (Lundvall 1992) (Nelson 1993). It brings forward the notion of innovation as the outcome of complex interactions and dynamics in the idiosyncratic socio-economic context of an economy (Edquist 1997). Yet, the more the innovation system approach has gained the upper hand, the more policy-makers have grown dissatisfied with the limitations of evaluating single R&D programs. Together with the appearance of more complex policy systems and mixes (Smits and Kuhlmann 2004), the innovation system approach has made apparent the need for advanced tools providing knowledge at the system

level. In particular, the holistic view of innovation policies (Edquist 2011) with the focus on interaction and interactive learning between organizations (Lundvall and Borrás 1998) has raised awareness about the need of more sophisticated tools to enable policy-makers better grasp the complexity of the impact of policy instruments.

In his seminar paper about the new frontiers of evaluation studies, Steven Feller (2007) reflected upon this growing dissatisfaction between the needs of policy-makers for more encompassing approaches stemming from the innovation system approach on the one hand; and the conventional praxis of research evaluation of individual R&D programs on the other. Single evaluations are increasingly perceived to be too limited to provide answers regarding the impacts of public initiatives in the wide framework of the economy. “Existing evaluations touch only lightly, however, on how the strategies, behavior, performance of the sectors or actors described in the national innovation taxonomy change as a result of the cumulative, long term impact of a cluster of programs” (Feller 2007).

In their review of the literature Molas-Gallart and Davis argue that “the practice of policy evaluation continues to lag behind advances in innovation theory. Innovation theory has produced successive generations of more sophisticated conceptual models that seek to explain how the relationship between scientific and technological research and the market opportunities for innovation occurs.” (Molas-Gallart and Davies 2006). However, much of the evaluation undertaken today is still performed at the program level, and is mainly based on simple models of impact assessment and accountability. Yet, the question of how to aggregate and integrate findings relating to specific policies and programs into an overarching evaluation framework that looks at the effect of policies into the national innovation systems is not an easy one. The innovation system approach and the theoretical framework it suggests “have proved difficult to use in the practice of evaluation, resulting in a gap between evaluation practice and Science Technology Innovation (STI) policy theory” (Molas-Gallart and Davies 2006). In spite of some very few attempts to link innovation system theory with evaluation frameworks, the literature of innovation system approach continues to have a largely ‘unfinished business’ of bringing together theory and the practice of policy-making and its evaluation (Borrás and Edquist 2016).

The few scholarly attempts to develop the theoretical framework are worth examining here in detail, as they provide different clues about how to define and shape innovation system-wide evaluation exercises. They all share the same starting point, namely, the understanding that R&D and innovation evaluation should move ahead and beyond from the myriad of isolated and individual program-focused evaluations. Governmental actors need encompassing evaluation exercises that provide them with insights for strategic policy action. However, in spite of the same starting point, these scholars have come to provide very different approaches. On the more ‘hands-on’ side, Jordan, Hage and others have engaged in a definition and re-definition of the indicators needed at the micro- meso- and macro-level of innovation system (Jordan, Hage et al. 2008) (Hage, Jordan et al. 2007). These authors are mainly concerned with providing a specific indicator-based overall methodology that will allow to develop innovation system-wide evaluations at the national level.

As a new concept, system evaluation tries to respond to the sophistication of innovation policy systems and provide knowledge on the interactive effects and policy performance of such systems (Magro and Wilson 2013, Edler et al 2008). Nevertheless, only three attempts have been made over the past decade to develop the concept of system analysis – Arnold (2004) suggesting combining different levels of evaluation into a system view, Edler et al (2008) proposing the concepts of *meta-analysis* and *evaluation-synthesis*, Magro and Wilson (2013) suggesting an “evaluation mix protocol” for a practical design of system evaluation.

In the earliest work of the three, Arnold (2004) does not offer an explicit definition for system evaluation, but describes it as a ‘system of evaluations’, comprising of different evaluations on several levels. For the broadest picture on innovation policy, he suggests using an ‘analysis of system health’. While unclear on the exact methods to be used in this analysis, he suggests it should cover the performance of the major institutional blocks in the system, connectivity within the system as well as knowledge and capabilities (Arnold 2004).

In contrast to that, Edler et al (2008) perceive system evaluation as “using existing evaluations to learn about policy performance and policy effects on the system level”. Inside this frame they separate two concepts, meta-analysis and evaluation synthesis, that serve as a systematic conceptual framework for utilizing and analyzing existing evaluation data. Meta-analysis allows for “an improved comparison and understanding of interventions and their effects by taking into account the results of a large number of evaluations” (Edler et al 2008). For its part, evaluation synthesis is understood as “an aggregated content analysis based on multiple evaluation reports on similar programs or projects” (Edler et al 2008).

Magro and Wilson (2013) define system evaluation as “meta-evaluations or secondary analyses that build on individual evaluations in trying to capture the systemic nature of policies; moving ahead from isolated, individual evaluations”. In that respect, they share a common viewpoint with Edler et al (2008) that a system evaluation should be based on individual evaluations, meaning that a system understanding results in the sum of knowledge derived from specific program or project evaluations.

There is a clear divide between the three approaches. Magro & Wilson (2013) and Edler et al (2008) share a common viewpoint that a system evaluation should be based on individual evaluations, suggesting that a system understanding is the sum of knowledge derived from specific program or project evaluations. On the other side of the spectrum, Arnold (2004), following the national system of innovation theory, sees that a system of innovation requires a ‘system of evaluations’, comprising of different evaluations on several levels.

These three approaches are relevant and stimulating in the sense of providing a first take in this matter. However, their concepts fall short of providing a clear-cut definition of what is and what is not ‘system evaluation’. More concretely, their conceptual inconclusiveness is due to the lack of specific formulation determining the constitutive attributes of the concept.

3. 'System Evaluation' Concept and Dimensions for a Typology

In the previous section we showed that there are several competing understandings on what a system evaluation of innovation policy would and should entail in terms of combining information from different levels of abstraction. We argue that we can see here the reflection of the classical problem in social sciences regarding the micro and macro-level of analysis. On one hand social phenomena are defined by the action of individual actors (such as individuals, firms, organizations, etc), on the other hand social phenomena is also defined by general trends at the collective/macro level (in our case, the innovation performance of a society and economy). While both levels are important already on their own, the key lies in linking them to each other at the meso-level.

In the context of the current research, it is important to keep in mind that we are looking at the innovation system from the perspective of innovation policy. We see innovation policy as the channel for resources to the economy and society, with the expectation to improve innovativeness at all its levels. For this reason, when studying evaluations of the system level, we look at three levels, namely, the evaluation of individual policy programs (micro-level of policy evaluation, which concerns the effects of individual policy instruments); policy-mix evaluations, which focuses on a specific subset of the policy instruments and particularly central instruments and institutions in the system, examining their combined effects on a wide context (meso-level of policy evaluation); and analyses of socio-economic innovative performance of the economy (macro-level effects of policy).

We need a clear definition of 'system evaluation' using the perspective of the national system of innovations theory in order to identify empirically which countries have and do not have evaluation systems, and what characterizes them. A clear definition, and its operationalization, will allow us to grasp the complexity of conducting innovation system evaluations at different levels of innovation policy, while accounting for the diversity and variation in which different countries have developed such evaluations. The classical understanding of the national system of innovation sees the two dimensions that make an innovation system as the policy-dimension (i.e. institutional set-up) and the innovation performance dimension (i.e. aggregate data on firm performance). Assessing the institutional set-up comprises both policy level evaluations as well as evaluating the overall policy-mix. We argue that it is the policy-mix level of assessment that provides the connection between program evaluations and overall economic indicators on innovation performance – it places information from both the individual evaluations as well as the macro-indicators into policy context, allowing for a 'system' view of the whole policy area. Therefore, we see the policy-mix dimension as crucial in understanding the link between policy and its economic effects, tying it together into one analytical framework.

Following this discussion, we define system evaluation as an evaluation framework that assesses the innovation policies and institutional frameworks that are part and parcel of the innovation system in a way that provide an encompassing evaluative view of that system. In other words, system evaluations are those with a wide coverage of the object under assessment (the innovation system) in a way that takes into consideration individual policy programs (micro-level), policy mixes (meso-

level) and innovative performance (macro-level). It is important to note that this definition of system evaluation is an 'ideal type', a notion defined in deductive terms. Reality, as we will see, is far more complex and muddled. Following recent methodological discussions in the social sciences about the analytical abilities and limits of concepts, we want to avoid problems of 'conceptual stretching' by setting a clear cut operationalization of what constitutes a systemic evaluation and what not (Sartori 1970).

It is important to underline that our concept above defines an ideal model, a 'perfect and abstract' case of system evaluation. Ideal models are formed deductively from theory and theorizing endeavors, providing a clear guidance for empirical analysis. However, because they are 'ideal' they are not to be found in their 'purity' or 'entirety' in the real world. They are abstractions, not replicated in the complexity of social phenomena. For that reason we follow the 'family resemblance' approach, which is a suitable method operationalizing ideal models (Goertz 2006). This approach states that only specific sub-sets of all the possible concept attributes are to be found empirically, and that it is the particular combination of those attributes that gives the analysis its ability to dissect complex empirical social phenomenon.

In our case, it means that we expect very rarely to find individual countries carrying out systemic evaluations at all three levels of assessment at once (the micro, meso and macro). Instead, in our empirical analysis we expect to find countries with specific combinations of these three levels. What is then a system evaluation? What is the 'cutting line' of the concept? When can we determine that a country has or not system evaluation? As mentioned above, our definition of system evaluations indicates that system evaluations are those providing an encompassing evaluative view of that innovation system. Here we refer to the paramount relevance of the 'meso-level' as the assessment of crucial mixes of policy instruments and/or institutions in the innovation policy. The 'meso-level' is a sort of 'glue' of the endeavors to provide a careful assessment of innovation policy and institutional frameworks, in a way that gives meaning to the users of the evaluations.

Therefore, we propose an evaluation to be called 'systemic' when it contains the meso-level as a necessary condition, and one or both of the other two levels as its sufficient condition. Hence, systemic evaluations are the following three possible combinations: when a country has evaluations of innovation policy instruments at the meso & micro levels; at the meso & macro levels; and at the micro-, meso- and macro-levels.

Towards a typology. The definition of system evaluation and its conceptual operationalization are useful tools to interpret and analyze the empirical research. However they say little about the possible variation of those forms of evaluations other than the coverage of what is being evaluated in terms of the three attributes. In order to understand the possible variation among those countries with system evaluations, we need to consider further analytical dimensions. The idea of 'evaluation machines' suggested by Dahler-Larsen (2011) comes at hand, because it provides those additional dimensions. His analytical interests are in the field of organizational sociology and public administration, and he is therefore interested in seeing how evaluations of any type of policy program have been organized and developed in different national administrative settings. He defines an evaluation machine as the

set of “mandatory procedures for automated and detailed surveillance that give an overview of organizational activities by means of documentation and intense data concentration” (Dahler-Larsen 2011). The ‘machines’ are these politico-administrative and legal procedures by which individual public administrations have defined, organized and carried on the evaluation of programs. In his view, there are five defining features characterizing every evaluation machine: coverage, permanence, organizational responsibility, prospective use, and epistemological perspective.

‘Coverage’ refers to the specific phenomenon that is being evaluated. Some evaluation machines in different countries will have broader or narrower scope of the evaluated phenomena. This notion of ‘coverage’ is directly related to our definition of system evaluation, and the way in which we operationalized it above. It constitutes our first step for conceptual clarification. More interesting for our purpose of identifying variations are the other four dimensions of Dahler-Larsen (2011); particularly two of them. ‘Permanence’ refers to the time dimension of the evaluation activities, asking whether these evaluations are carried out on a regular basis or on an ad-hoc basis. ‘Organizational responsibility’ has to do with the organizational-dimension of evaluations: whether they are embedded in organizational procedures, reflecting the ideas and values of the organization, not individual evaluators. Hence it asks what is the political and organizational placement of the unit/department/individuals conducting the evaluations?

Dahler-Larsen (2011) provides two further dimensions in his ‘evaluation machines’, which are, the epistemological perspective of evaluations, and the prospective use of evaluations. Both are very relevant, but will play a minor role in this paper. The first, the epistemological perspective, refers to the existence of a coherent worldview in an evaluation framework. It has to do with the set of indicators, measurements, and specific evaluation methodologies developed and applied in the course of the evaluation work. Given the sophistication of systemic evaluations, the type of indicators and methodologies used are crucial in order to understand the nature of the evaluation exercise and the type of results and outputs these evaluations generate. However interesting, this needs a very careful qualitative in-depth study that lies beyond the scope of the present paper. ‘Prospective use’ is the last dimension of that author, and has to do with the way in which the evaluation outputs are located in the politico-administrative context and are being used or not in the process of policy making. This is without any doubt a highly relevant issue. But we believe that the question about the extent to which an evaluation (systemic or not) has been used and has lead towards policy learning, is a matter that goes beyond a typology of evaluation dimensions, and belongs to questions about the effect of evaluation practices all in all. In other words, it has to do eventually with the transformative power of evaluations, rather than the way in which evaluations are actually conducted.

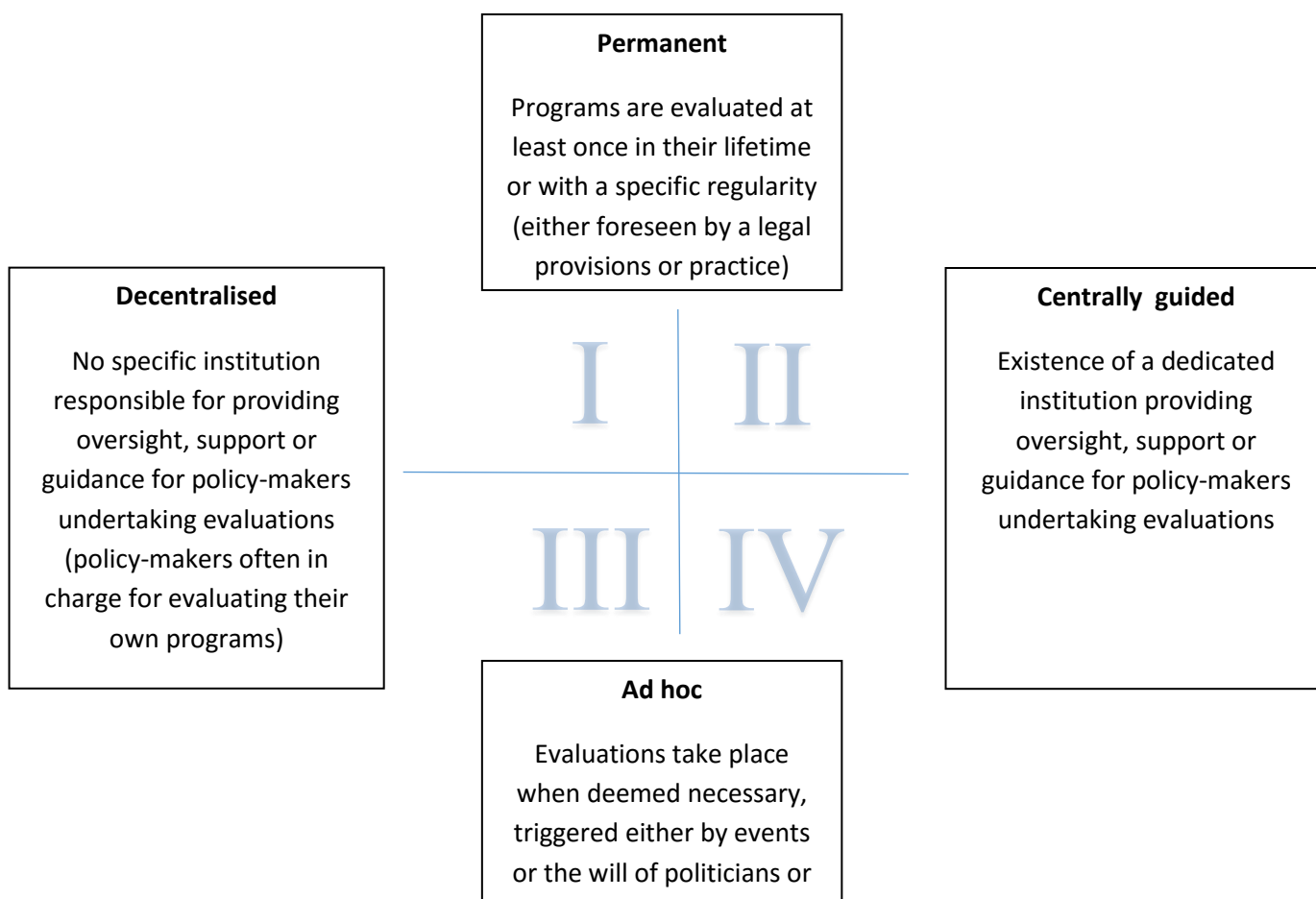
Based on the above, our concept defining system evaluation will be based on a specific interpretation of ‘coverage’ based on three sub-sets of levels, where the meso-level of evaluation must be present. Once we have identified which countries have system evaluations and which ones do not, we will generate a typology of the former according to their permanence and their organizational responsibility. Table 1 below summarizes these two further dimensions and Figure 1 provides a conceptual map of the different variations.

Table 1. Two dimensions defining different types of system evaluations

Dimension	Definition	Operationalisation
Permanence	The time-dimension of the evaluations: whether they take place on a regular basis, and are carried out in a continuous manner	- Are evaluations in innovation policy taking place regularly? - Is the evaluation process continuous, rather than ad hoc?
Organizational responsibility	The organizational-dimension of evaluations: whether they are embedded in organizational procedures, reflecting the ideas and values of the organization, not individual evaluators	What is the political and organizational placement of the units/departments/people conducting the evaluation?

Source: Our adaptation of Dahler-Larsen 2011 and Dahler-Larsen 2016

Figure 1. Defining features of the theoretical varieties of ‘evaluation machine’



4. Data and methods

The main source of empirical data is collected through semi-structured interviews with policy-makers from the 28 EU member states. A total of 31 interviews, with respondents in 26 countries were carried out in 2016. This gives a preliminary basis for the analysis in the coming two sections.

Interviews are carried out with the heads of innovation policy or other senior officials in the field of innovation policy, provided that they have a thorough understanding of the foundations of their evaluation framework. Via the interviews, data is collected on the evaluation practices in general and system evaluation practices in particular, with an emphasis both on current state as well as future plans. More specifically, we have gathered information on three aspects: how is the performance of a country's innovation policy evaluated, who are the actors involved, what are the main characteristics of the evaluations in terms of scope, regularity and methods.

The interviews were conducted following a specific questionnaire, which was initially tested and thereafter adjusted to make sure questions were clearly formulated and understood by the respondents.

This study is also based on secondary data sources, main of which are documents. These are main evaluation reports, evaluation protocols or other government documents that are relevant for the research question of this paper, and hence provide insights into the national evaluation practices. To that end, data from the OECD Innovation Policy Platform has also been consulted, and included in the analysis.

5. Analysis Part I

In this chapter we take a look at how much the existing formats for system-level reviews have been used by EU member states and how many of these countries have experimented with their own tailor-made reviews of innovation policy. The OECD reviews of innovation policy aim at providing a comprehensive overview of a country's innovation system. The European Research Area and Innovation Committee's (ERAC, formerly CREST) peer-reviews, facilitated by the European Commission, have the same aim of providing policy-makers with a comprehensive analysis of their innovation systems, but its format is more limited in scope and depth than the OECD reviews. However, the launch of the Policy Support Facility by the European Commission could also see changes the format of the peer-reviews of innovation policy.

From Table 1 we can see that most countries in the EU have over the past 10 years commissioned either an OECD or European Commission facilitated review. Some countries have carried out a policy-review on their own, using their own resources or independent experts. Pointing out these reviews is important since they are expected to provide the policy-mix level analysis that is crucial in

tying together the information from program as well as macro-indicator level. By doing that, these reviews create a necessary precondition for a system evaluation of innovation policy.

Table 2. EC, OECD or independent system-level reviews in EU member states

Country name	OECD country review	ERAC, CREST or PSF peer-review	<i>National review*</i>
Austria		CREST	2009
Belgium		ERAC, CREST	
Bulgaria		PSF, CREST	
Croatia	2014		
Cyprus		ERAC	
Czech Republic			
Denmark		ERAC	
Estonia		ERAC, CREST	
Finland	in process		2009
France	2014	CREST	
Germany			annual
Greece			
Hungary	2008	PSF (in process)	
Ireland			2015
Italy			
Latvia		ERAC, PSF (in process)	
Lithuania		CREST	
Luxembourg	2007, 2015		
Malta			
Netherlands	2014	CREST	
Poland			
Portugal			
Romania		CREST	
Slovakia			
Slovenia	2012	ERAC	2012
Spain		ERAC, CREST	
Sweden	2013, 2016	CREST	
United Kingdom		CREST	

Source: OECD, European Commission, our data

*data collection in process

While a majority of countries in the EU have participated in an ERAC/CREST/PSF peer-review exercise, many have also commissioned an OECD review. Some have also undertaken large-scale policy-evaluation exercises of their own – such as the Evaluation of the Finnish National Innovation System, carried out by a team of independent experts in 2009. This data shows a steady demand for

a system level understanding of the performance of national innovation policy. However, both OECD and ERAC/CREST/PSF are by their nature ‘one-off’ exercises (admittedly there are examples of countries having conducted already two rounds of reviews, but this is arguably not enough yet to establish a regularity) as they are conducted mainly by outside experts and the procedures are not internalized, in order to gain necessary traction for a more permanent setting. Therefore the ‘national’ reviews conducted by countries themselves might offer a basis for more permanent structures. Of course, one has to take into consideration that there are reasons for ordering a review from international organizations, often the most prominent one being the expected impartiality of outside experts.

6. Analysis Part II

This chapter looks at the empirical data collected on evaluation practices in EU member states and places it in our conceptual framework. First of all, we start by analyzing the data presented in the previous chapter in the context of our definition of system evaluation. With due regard to the advances in innovation policy reviews, we must take into account that following our definition an OECD or ERAC/CREST/PSF review does not constitute a system evaluation on its own, but rather acts as one building block of a system evaluation (see chapter 3 for a discussion on and definition of system evaluation). Therefore, in Table 3 we have added all dimensions, and consider whether the countries with system evaluations comply with our previous criteria of meso-level evaluation as the necessary condition, and the micro and/or the macro-level as the sufficient condition). We have checked the boxes with a capital ‘X’ where our data shows strong evidence of well-developed evaluation activities in a particular dimension. Where we have evidence of practices that are not as well-developed, we have marked the country with a simple ‘x’. Cells without an ‘x’ are those where we have evidence that there are no remarkable evaluation practices. The data lines are blank for two countries (Czech Republic and Italy) where we have not yet managed to acquire information on their evaluation practices. In this chapter we discuss mainly the findings on the program level of evaluations, since our data collection efforts have until this point of the research focused on the program and policy-mix levels. Data collection on macro level characteristics is still ongoing – since we are still researching data from a number of member states, we are not in a situation yet to present data on this level.

Our data (Table 3) shows that eleven countries of the 26 observed (we have not yet managed to acquire data on Czech Republic and Italy) demonstrate strong program evaluation practices. This means that the respondents have stated that there are steady rules or practices guiding program evaluations – every program is evaluated either once its lifetime or in a certain interval. Also eleven countries have reported weak practices or our current evidence suggests that there are probably weak evaluation practices in place for innovation policy programs. For example, many of the countries in this group have reported that their programs are evaluated “according to the rules of EU Structural Funds” – as the legislation governing the use of EU Structural Funds does not demand a

comprehensive and regular evaluation of each individual policy instrument but rather the output and outcome of patches of programs (often the Operational programs as a whole), evaluations “according to SF rules” cannot be automatically considered as strong evaluation practices. The same applies for policy-mix evaluations – merely referring to Structural Fund rules accounts for a modest practices, whereas we consider an OECD, ERAC/CREST/PSF or sophisticated national analyses as evidence of strong practice.

Table 3. Evidence of system evaluation practices in EU member states, along the three levels of system evaluation

Country name	Program level	Policy-mix level	Macro level*
Austria	X	X	
Belgium	X	X	
Bulgaria	x	X	
Croatia		X	
Cyprus		X	
Czech Republic	<i>no data</i>		
Denmark	X	X	
Estonia	x	X	
Finland	X	X	
France	X	X	
Germany	X	X	
Greece	x		
Hungary	X	X	
Ireland	X	X	
Italy	<i>no data</i>		
Latvia	x	X	
Lithuania	x	X	
Luxembourg		X	
Netherlands	X	X	
Poland	x	x	
Portugal	x	x	
Romania	x	x	
Malta	x		
Slovakia	x	x	
Slovenia	x	X	
Spain	x	X	
Sweden	X	X	
United Kingdom	X	x	

Source: our data

* data collection in process

Legend:

X: Strong evidence of well-developed evaluation practices;

x: Evidence of modest evaluation practices.

Surprisingly, four countries reported having no specific program evaluation practices at all. One possible reason for this could be that innovation policy in these countries has only recently started to be systematically developed and proper evaluation practices are still to take root. At the same time it could also partly be due to any gaps in our data - further investigation is necessary into those cases to verify the initial information.

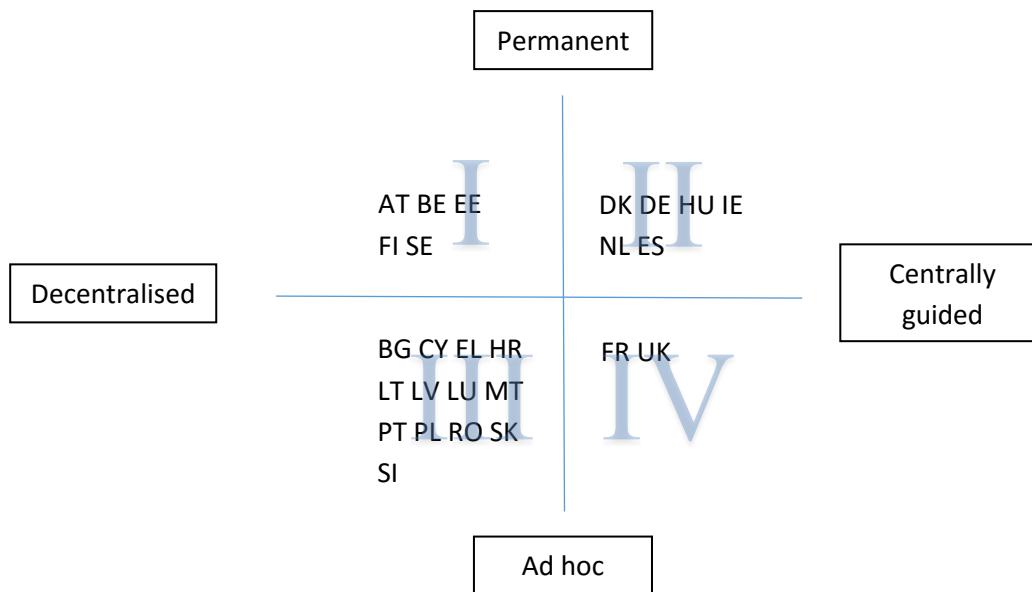
In order to assess the character of the national evaluation practices (stage 2 analysis, Figure 2), we plot two of the key features of 'evaluation machine' against each other. On the x-axis we look at 'organizational responsibility' – what is the organizational placement of the 'evaluation machine', on the y-axis we look at 'permanence' – do the evaluations run on a continuous basis or do they take place in an *ad hoc* manner (see chapter 3 for a thorough discussion of these features).

More specifically, on the 'permanence' axis we position countries according to the 'permanence' of their evaluation practices. Based on the information from the interviews and from secondary data, we have tried to determine whether evaluations of innovation policy take place with a certain regularity (for example being tied to the budgetary procedures) or are carried out on an *ad hoc* basis (for example the decisions to evaluate a program are being driven by certain events or political circumstances). It is important to note that of the three components constituting a system evaluation (program level, policy-mix level, macro-indicator level) evidence of regularity could for the time being only be determined on the program level and the macro-indicator level, while experimentation with evaluations on the policy-mix level has only started recently and any regularity is still to be established.

On the 'organizational responsibility' axis, we position countries according to the existence and placement of national coordination or support mechanisms for evaluating innovation policy. In many countries an organization or a structural unit within an organization is charged with overseeing the evaluation activities in the field of innovation policy. The oversight function can take different forms, ranging from strict coordination where all intended evaluation proposals must be approved by that institution, to a more supportive role, where the institution guides or advises the evaluation process but does not take responsibility. The existence of a central coordinating or supporting institution would suggest that the 'evaluation machine' is more centrally coordinated, while the lack of such institution would suggest a more decentralized approach.

The combination of the two axes create four conceptual sectors (see Figure 1). Firstly, in the upper left corner we would see countries where evaluation practices in innovation policy show high levels of permanence, but are based on a decentralized coordination/support structure. Secondly, in the upper right corner we can witness high level of permanence combined with more centralized coordination/support structures. Thirdly, in the lower left corner, there are countries where evaluations take place more on an *ad hoc* basis and where the coordination and support functions are relatively decentralized. Fourthly, in the lower right corner we can see a combination of *ad hoc* practices with centralized guidance structures for evaluations.

Figure 2. Types of ‘evaluation machine’ in EU member states



Source: our data

Our data shows a relatively uneven distribution of cases among the four-folded typology. In the first type we have five countries, in the second type six, the third type is the most populated with thirteen member states and the fourth type is in turn the least crowded with only two cases. As to the general patterns, we can see most of the ‘new’ member states concentrated in the lower left corner that is characterized by low degrees of centralization and permanence. This could be attributed to relative youth of their innovation policy systems, where steady practices still need to be developed. While neither decentralized organizational set-up nor the *ad hoc* nature of performing evaluations is in itself a sign of a policy system in its developing stage, the combination of the two seems to hint at it. Likewise interestingly, we see two of the larger EU member states in the lower right corner, with centralized guidance systems combined with a low degree of permanence. While the causes of this merit an analysis of its own, one might think on a possible hypothesis that a centralized guidance system provides a necessary degree of oversight and administrative stability of policy makers, making any explicit requirements on a permanent practice unnecessary. At the same time we see the top innovation performers on the upper level of permanence, but equally distributed between centralized and decentralized systems. This could suggest that advances in innovation policy require a steady evaluative effort to feed the policy-making cycle with information on policy performance and impact, thereby making a case in support of evidence-based policy-making.

7. Conclusions

This paper has provided a novel insight into an under-researched phenomenon in innovation studies – the practice of ‘system evaluations’. We have addressed the gaps in the previous literature by firstly improving the conceptual understanding of what constitutes ‘system evaluation’. Secondly, we have identified specific features that help to characterize system evaluation practices across countries. By developing this clear conceptual framework, we have been able to analyze the empirical evidence from 28 EU member states, using a four-fold typology. While our work has been helpful in understanding the current practices of system evaluation in innovation policy, it also opens up new avenues in the research agenda, such as the epistemological perspective of and policy learning through these kinds of system evaluations.

Firstly, in this paper we have opened up and improved the conceptual understanding of what a system evaluation is. We have demonstrated that in order to adequately grasp the whole of policies addressing system of innovation, three levels of assessment need to be present and complement each other. A system perspective requires information from the micro level (concerning policy programs), the meso-level (concerning the policy-mix) and macro level (concerning the socio-economic effects of innovation policy). We claim that while the micro and macro levels act as sufficient conditions to determine the existence of system evaluation practices, it is the meso-level that constitutes the necessary condition, providing a ‘glue’ for connecting the different levels into a coherent understanding of the performance of the system of innovation. Our data, gathered through 30 interviews with policy makers from EU member states, shows that most countries have at least some program evaluation practices in place and most countries have also had some sort of policy-mix analysis carried out during the last decade (often facilitated by OECD or the European Commission). Our data is still inconclusive on the macro-level, further data collection is currently in progress to provide better information there.

Secondly, in order to grasp the empirical complexity of system evaluation practices, we created a four-fold typology. We took as starting point the concept of ‘evaluation machine’ and used two of its features – permanence and organizational responsibility – to demonstrate how different countries have organized their evaluation practices. Our data shows that a large group of countries (mainly ‘new’ member states) are grouped in the corner that is characterized by low level of permanence and a decentralized coordination/support system for evaluations. Most of the countries that are perceived as top innovation performers tend to be placed in the upper-left and upper-right corners, showing a strong tendency towards more permanent evaluation practices. However intriguing these findings are, they still have to be perceived as preliminary and need to be strengthened by further data collection.

Thirdly, while this research has provided a new conceptual understanding of system evaluation and created a theory-based framework for analyzing its varieties, there is more work to be done on the subject and further avenues to be explored. While we have made use of three features of the concept of ‘evaluation machine’ – coverage, permanence and organizational placement – there are two additional characteristics that deserve further enquiry – epistemological perspective and prospective use of evaluations. Due to the large empirical field of our research (several levels of evaluation practices at all 28 EU member states), we have not been able to go into depth with each case. The

goal of this paper has been to provide conceptual clarification and an encompassing empirical data collection about features and types of system evaluation practices. Future research will focus on the particularities of individual countries, exploring in detail the epistemological foundations of those evaluations as well as the use of the evaluations in further policy-making. This latter point about prospective use is strongly related to issues of policy learning in innovation policy-making (Borrás and Højlund 2015), leading towards future research questions, such as how do different types of system evaluations affect policy learning across national governments, and what kinds of policy learning emerge from the theoretical and empirical varieties of evaluation systems.

To sum up, we have provided a novel insight into the concept of system evaluation. Firstly, we have developed a new concept for understanding the phenomenon, by proposing a definition combining the three levels of micro, meso and macro level evaluations. Secondly, we created a four-fold typology to account for the empirical varieties of 'system evaluation' and used it to analyze data collected from the 28 EU member states. Finally we have pointed out new directions for further research, suggesting to go deeper into the features of 'evaluation systems and to take into consideration the perspectives on policy learning.

References

- Arnold, E. (2004). "Evaluating research and innovation policy: a systems world needs systems evaluations." Research Evaluation **13**(1): 3-17.
- Borrás, S. and C. Edquist (2016). Conceptual Underpinnings for Innovation Policy Design: Instruments and Indicators in Context. OECD Blue Sky III: Informing Science and Innovation Policies. Ghent.
- Borrás, S. and S. Højlund (2015). "Evaluation and policy learning: The learners' perspective." European Journal of Political Research **54**(1): 99-120.
- Dahler-Larsen, P. (2011). The Evaluation Society. Stanford, Stanford University Press.
- Dahler-Larsen, P. (2016). Private correspondence, April 13, 2016.
- Edler, J., B. Ebserberger and V. Lo (2008). "Improving Policy Understanding by means of Secondary Evaluation." R&D Evaluation **17**(3): 175-186.
- Edquist, C., Ed. (1997). Systems of Innovation : Technologies, Institutions and Organizations. London, Pinter.
- Edquist, C. (2011). "Design of innovation policy through diagnostic analysis: identification of systemic problems (or failures)." Industrial and Corporate Change **20**(6): 1725-1753.
- Evaluation of the Finnish National Innovation System (2009). Helsinki, Helsinki University Press.

- Feller, I. (2007). "Mapping the frontiers of evaluation of public-sector R&D programs." Science and Public Policy **34**(10): 681-690.
- Goertz, G. (2006). Social Science Concepts. A User's Guide. Princeton, Princeton University Press.
- Hage, J., G. Jordan and J. Mote (2007). "A theory-based innovation systems framework for evaluating diverse portfolios of research, part two: Macro indicators and policy interventions." Science and Public Policy **34**(10): 731-741.
- Jordan, G. B., J. Hage and J. Mote (2008). "A theories-based systemic framework for evaluating diverse portfolios of scientific work, part 1: Micro and meso indicators." New Directions for Evaluation **2008**(118): 7-24.
- Lundvall, B.-Å., Ed. (1992). National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. London, Pinter.
- Lundvall, B.-Å. and S. Borrás (1998). The Globalising Learning Economy: Implications for Innovation Policy. Brussels, European Commission.
- Magro, E. and J. R. Wilson (2013). "Complex innovation policy systems: Towards an evaluation mix." Research Policy **42**(9): 1647-1656.
- Molas-Gallart, J. and A. Davies (2006). "Toward Theory-Led Evaluation: The Experience of European Science, Technology, and Innovation Policies." American Journal of Evaluation **27**(1): 64-82.
- Nelson, R. R., Ed. (1993). National Innovation Systems. A Comparative Analysis. Oxford, Oxford University Press.
- Sartori, G. (1970). "Concept Misformation in Comparative Politics." The American Political Science Review **64**(4): 1033-1053.
- Smits, R. and S. Kuhlmann (2004). "The rise of systemic instruments in innovation policy." International Journal of Foresight and Innovation Policy **1**(1-2): 4-32.

Annex1. List of interviews

- 1. Austria**
Head of Department
Austrian Ministry for Transport,
Innovation and Technology
29.04.16
- 2. Belgium**
Director
Scientific and Technical Information
Service
01.06.16
- 3. Bulgaria**
Director
Ministry of Economy
01.06.16
- 4. Croatia**
Head of Sector
Ministry of Science, Education and
Sports
06.05.16
- 5. Cyprus**
Director General
Research Promotion Foundation
23.05.16
- 6. Czech Republic**
missing
- 7. Denmark**
Head of Section
Danish Agency for Science,
Technology and Innovation
27.05.16
- 8. Estonia**
Head of Unit
Ministry of Economic Affairs and
Communications
27.01.16
- 9. Finland**
Head of Division
Ministry of Employment and the
Economy
20.01.16
- 10. France**
Head of Unit
Ministry for Economy, Industry and
Digital Affairs
09.12.15
- 11. Germany**
Head of Unit
Federal Ministry for Science and
Technology
28.01.16
- 12. Germany**
Professor/Director
Max Planck Institute for Innovation
and Competition
16.05.16
- 13. Greece**
Head of Directorate
Ministry of Education, Research and
Religious Affairs
04.05.16
- 14. Hungary**
Head of Department
National Research, Development and
Innovation Office
23.05.16

- 15. Ireland**
Principal Officer
Department of Jobs, Enterprise and
Innovation
15.06.16
- 16. Ireland**
Deputy Principal Official
Department of Jobs, Enterprise and
Innovation
21.06.16
- 17. Italy**
missing
- 18. Latvia**
Head of Division
Ministry of Economics of the Republic
of Latvia
28.01.16
- 19. Lithuania**
Head of Division
Ministry of Economics of the Republic
of Lithuania
17.03.16
- 20. Luxembourg**
Deputy Director
Ministry of Higher Education and
Research
02.06.16
- 21. Malta**
Executive officer
Malta Council for Science and
Technology
29.04.16
- 22. Netherlands**
Deputy Director
- Ministry of Economic Affairs
26.01.16
- 23. Poland**
Head of Unit
Ministry of Economic Development
19.05.16
- 24. Portugal**
Member of Board
National Innovation Agency
20.05.16
- 25. Romania**
Senior counsellor
National Authority for Scientific
Research and Innovation
02.06.16
- 26. Slovakia**
Director
Ministry of Economy
30.05.16
- 27. Slovenia**
Professor
University of Ljubljana
21.06.16
- 28. Slovenia**
Director General
Ministry of Economic Development
and Technology
01.07.16
- 29. Spain**
Deputy Director General
Ministry of Economy and
Competitiveness
02.06.16

30. Sweden

Head of Section

Ministry of Enterprise and Innovation

14.01.16

31. United Kingdom

Head of Unit

Department for Business, Innovation
& Skills

25.05.16