European Commission

JRC TECHNICAL REPORT

Suggestions for Monitoring and Evaluation of Transformative Innovation Policy

JRC Working Papers on
Territorial Modelling and Analysis
No 03/2023

Authors:
Marques Santos, A.
Coad, A.

2023

Joint Research Centre

Contact information
Name: Anabela M. Santos
Address: Edificio Expo, C/Inca Garcilaso 3, 41092 Sevilla (Spain)
Email: anabela.MARQUES-SANTOS@ec.europa.eu
Tel.: +34 95 448 71 61

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC132492

Seville: European Commission, 2023

How to cite this report: Marques Santos, A., and Coad, A. (2023). Suggestions for Monitoring and Evaluation of Transformative Innovation Policy. JRC Working Papers on Territorial Modelling and Analysis No. 03/2023, European Commission, Seville, JRC132492.

The JRC Working Papers on Territorial Modelling and Analysis are published under the supervision of Simone Salotti, Andrea Conte, and Anabela M. Santos of JRC Seville, European Commission. This series mainly addresses the economic analysis related to the regional and territorial policies carried out in the European Union. The Working Papers of the series are mainly targeted to policy analysts and to the academic community and are to be considered as early-stage scientific papers containing relevant policy implications. They are meant to communicate to a broad audience preliminary research findings and to generate a debate and attract feedback for further improvements.

Suggestions for Monitoring and Evaluation of Transformative Innovation Policy

Anabela M. Santos and Alex Coad***

*Joint Research Center, European Commission, Seville, Spain
**Waseda Business School, Tokyo, Japan

JRC TECHNICAL REPORT

JRC Working Papers on Territorial Modelling and Analysis No 03/2023

INDEX

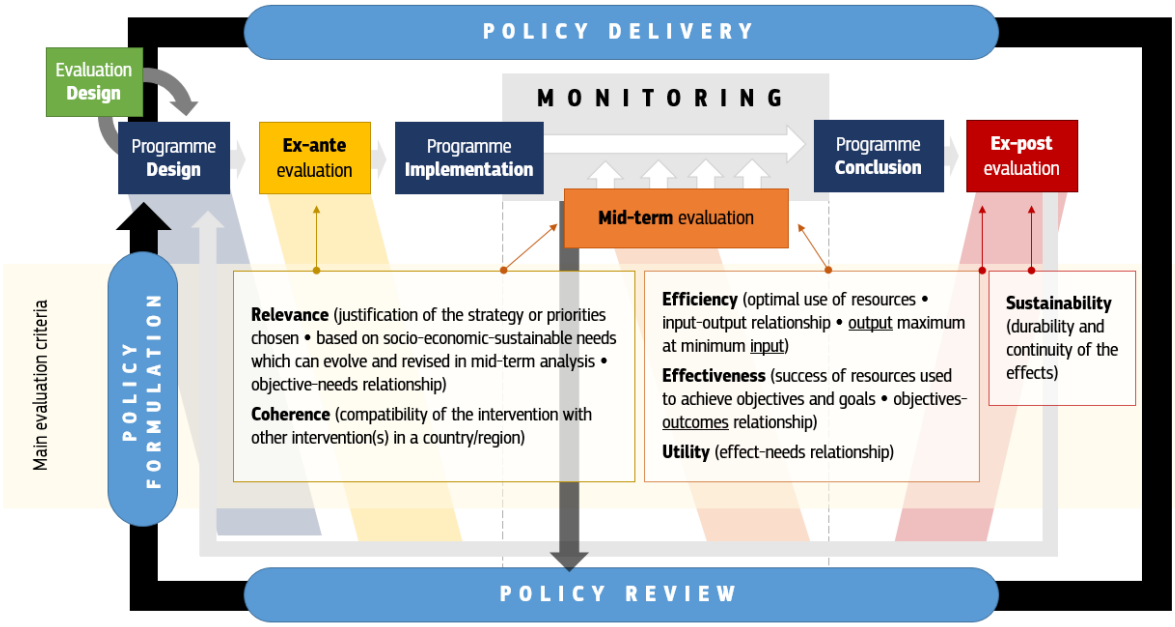
Executive Summary

Monitoring and evaluation (M&E) are essential components of the policy cycle, especially in the light with of the European Commission's Better Regulation. M&E aim to ensure that EU policies and legislation are evidence-based, well-designed, and achieve their intended outcomes efficiently and effectively.

Facing up to today's grand challenges (climate change, sustainable development goals) requires considerable efforts for Transformative Innovation Policy (TIP). TIP includes various policy instruments, such as R&D tax credits, direct R&D grants, intellectual property (IP) protection, public procurement, guaranteed loans, and so on. TIP differs from standard Innovation Policy (IP) because it puts more emphasis on longer term effects of innovation policy, spreading out over broad regions, taking into account a broad range of indicators of societal and environmental well-being, paying attention to coherence across various interventions that make up a "policy mix", and paying attention to externalities (spillovers) across regions and sectors. TIP holds the potential to steer economic activity towards adopting new products, cleaner processes, and better business practices that are compatible with a sustainable economy. To ensure that TIP resources are used efficiently and effectively, TIP requires rigorous Monitoring and Evaluation (M&E).

M&E is crucially important, because there is a world of difference between the grand visions of high-level policymakers and the reality of how actual policy initiatives affect incentives and behaviour. This paper focuses on TIP's M&E, touching upon various themes and discussing the requirements and making suggestions for how to improve TIP's M&E in future. Figure 1 shows our proposed Monitoring and Evaluation framework.

Figure 1. Proposed Monitoring and Evaluation framework



We discuss the various degrees of sophistication that can be found in M&E exercises, by referring to Storey (2000)'s "6 steps to heaven" scale. This scale ranges from basic monitoring activities at step 1 (such as interviews asking recipients whether they are happy to receive funding), to rigorous econometric evaluation analyses at step 6, which are set up to obtain quantitative estimates of causal effects. We then provide a survey of causal inference techniques that reach the 6th step on this scale (Randomized Field Experiments, Regression Discontinuity Design, Instrumental Variables, Difference-in-Differences, etc), and analyse the degree of sophistication of recent EU Cohesion project evaluations.

We conclude that evaluation completed by EU Member States using causal inference techniques only represents 8% of the total evaluations conducted for period 2014-2020, and this percentage is even lower when we look at innovation or environment-related programmes. We emphasize, therefore, that M&E activities should strive to use rigorous evaluation methods in order to have a more accurate picture of the efficiency and effectiveness of TIP interventions.

We identify some gaps in the observed M&E of EU Member States and we provide some recommendations for how to set up M&E, contrasting traditional M&E with modern M&E. M&E activities should not occur exclusively after a policy intervention has finished, but should occur at all stages (before, during, after), and contribute to the design and steering of the policy initiative. Of crucial importance is real-time data, that allows for monitoring of the policy intervention during the time that it is running, as well as a prompt evaluation after the end of the intervention's duration. Policy goals should be clearly stated before the start of the policy intervention (in practice, this is not always done). In sum, we state that M&E needs improving, and we suggest how this might be done.

Suggestions for Monitoring and Evaluation of Transformative Innovation Policy

Anabela M. Santos (corresponding author)
Joint Research Center, European Commission, Seville, Spain
anablea.marques-santos@ec.europa.eu

Alex Coad
Waseda Business School, Tokyo, Japan
alex.coad@waseda.jp

Abstract

Transformative Innovation Policy (TIP) has an important role in the sudden transition that our economies require to face up to today's grand challenges (climate change, sustainable development goals). In the European Union (EU), Cohesion policy funds are one of the main financing instruments to support innovation and a fair transition. This paper focuses on TIP's monitoring and evaluation (M&E). To begin with, we discuss the various degrees of sophistication that can be found in monitoring and evaluation exercises (i.e. Storey's "6 steps to heaven" scale), ranging from interviews asking recipients whether they are happy to receive funding, to full-blown causal econometric analyses. We then provide a survey of causal inference techniques that reach the 6th step on this scale, and analyse the degree of sophistication of recent EU Cohesion project evaluations. We conclude that evaluation completed by EU Member States using causal inference techniques only represents 8% of the total evaluations conducted for period 2014-2020, and this percentage is even lower when we look at innovation or environmental-related programmes. We identify some gaps in the observed M&E of EU Member States and we provide some recommendations for how to set up M&E, contrasting traditional M&E with modern M&E, and highlighting the need for real-time data. In sum, we state that M&E needs improving, and we suggest how this might be done.

Keywords: Innovation; Sustainability; Policy Monitoring and Evaluation.
JEL Classification: Q54, Q55, O38

Acknowledgements: The authors would like to thanks Dimitrios Pontikakis and Daria Ciriaci from Joint Research Centre, European Commission, as well as the participants in the Workshop "EvEUCoP 2022: Evaluating challenges in the implementation of EU Cohesion Policy" for the valuable comments and suggestions received in early version of the paper.

1. Introduction

Innovation policy should be designed to face up to the challenges of the exceptional times in which we live. Climate change is one of the biggest challenges of the century, with effects on the environment (Mowery, Nelson and Martin, 2010), population health (Patz et al., 2005; Haines et al. 2006) and social (Fankhauser et al., 2008) and economic development (Fankhauser and Tol, 2005). Climate change is translated into extreme weather, an increase in temperature, changes in precipitations, and sea-level rise (IPCC, 2021), as well as the ecological emergency of mass extinctions (Ceballos et al., 2015). Its direct effects on the environment are observed in water quality and quantity, soil fertility, erosion, loss of biodiversity, and the rise of natural disasters. Such changes have adverse impacts on health (e.g. allergic and infectious diseases, malnutrition, and deaths) and economic outcomes (e.g. lower output and income, costs with physical infrastructure). Even though all regions have been affected by climate change, their vulnerability to climate change issues is also associated with their sectorial patterns. Poorer regions may struggle to adapt to the challenges of climate change, especially considering that the industry composition of poorer regions often disproportionately features polluting industries (McCann and Soete, 2020). Economies strongly dependent on agriculture, forestry, and tourism are being the most affected since these activities are linked to weather conditions. However, sectoral and economic transitions struggle against vested interests, inertia, and obstacles related to creative destruction.[i] As a result, public policy plays a key role in accelerating the transitions towards a better economy and society. Furthermore, to be more effective and efficient policy actions should be local and place-based instead of being global and non-tailor-made (Barca et al., 2012; Neumark and Simpson, 2015), due to the heterogeneous effect of the climate change in the different territories and the specific needs of each territory along the transition pathway. Additionally, the concept of place-based policy also lies in the involvement and interaction with stakeholders to support its design and implementation (Barca et al., 2012; Neumark and Simpson, 2015).

Innovation is seen as an important driver for the development of new technologies, product and process to support climate change mitigation and green transition (Chiou et al, 2011; Conding and Habidin, 2012; Küçükoğlu and Pınar, 2015). Public policy is crucial to provide effective incentives for this pathway (Pontikakis et al., 2022) and to reduce the market gap (Cowling and Liu, 2021; Xiang et al., 2022), especially when studies have already pointed out to an existing green financing gap (e.g. Polzin and Sanders, 2020; Cowling and Liu, 2021).

In a context of an ecological emergency, where trial-and-error is too expensive, monitoring and evaluation (M&E) of a public policy should play an important role in steering the use of public money to be more efficient and effective. However, a traditional approach of M&E based on a linear model and only socio-economic indicators to measure the success of a policy instrument may only provide a partial overview of its effect. Furthermore, without real-time data for monitoring, as the backbone of the evaluation process,

[i] Despite the rhetoric, subsidies for fossil fuels are still often higher than subsidies for renewable energy, even today: according to analysis from the International Monetary Fund (IMF), fossil fuels received subsidies of $5.9 trillion in 2020 (Parry, Black and Vernon, 2021).

useful information to support the policy cycle may arrive too late to policymakers.

The present paper aims to propose a new conceptual framework for M&E of Transformative Innovation Policy, which is inspired on the concept of place-based innovation policy for sustainability. As a methodological approach, we begin with a critical literature review and the identification of a mismatch between policy targets and existing M&E systems. To this purpose, we draw on both scientific and grey literature, and we complement our analysis using data of European Union (EU) Cohesion policy programme evaluations completed by the EU Member States for the programming period 2014-2020, and extracted from Cohesion Open Data Platform. This database allows classifying the sophistication of the Member States' M&E using information from the evaluation techniques used by evaluators. For the present study, we use data from Cohesion Policy because the funds allocated to this policy represent one the main sources to finance place-based innovation policy, as well as to support a fair transition. For instance, for the programming period 2014-2020, the EU budget associated to Research & Innovation (R&I) – Thematic Objective 1 – was more than €43.5 billion (11% of the total). The European Commission (EC) imposed as ex-ante conditionality to EU Member States to receive these funds to have in place a regional and/or national Smart Specialisation Strategy (S3). The concept of S3 is anchored to a governance model of R&I funds associated with the place-based approach of innovation policy (Foray et al., 2009).

In particular, the paper contributes to knowledge regarding M&E for Transformative Innovation Policy.

- We begin with a description of the traditional M&E approach along the policy cycle, focusing in particular on aspects of monitoring and evaluation.
- We suggest how the traditional M&E approach can be further strengthened to address some emerging themes in a framework that we call the non-traditional M&E approach of Transformational Innovation Policy (TIP).
- We describe how elements of our new non-traditional M&E approach are observed in the design of emerging innovation policy (in particular, the EU Partnerships for Regional Innovation[ii] (PRI) – for more details see Pontikakis et al., 2022). Indeed, the present study may represent the backbone for the design of a M&E system for PRI.
- While we draw on a rich literature in the area of sustainability transitions and transformative innovation policy (e.g. Haddad et al., 2022; Meissner and Kergroach, 2021; Magro and Wilson, 2019; Schot and Steinmueller, 2018; Diercks et al., 2019; Molas-Gallart et al., 2021; Haddad and Bergek, 2023) and place-based policy for sustainability (McCann and Soete, 2020), nevertheless our contribution relates to the less-commonly-discussed topic of aspects of monitoring and

[ii] The Partnerships for Regional Innovation (PRI) are a new approach to place-based innovation policy which builds on positive experiences with smart specialisation strategies. The concept was launched in 2022, as the result of a joint initiative between the Committee of Regions (CoR) and the European Commission's Joint Research Centre (JRC) to explore the development of innovation partnerships fit for the green and digital transition.

evaluation. Our paper differs from other contributions regarding M&E for TIP (e.g. Kroll, 2019; Janssen et al., 2022; Haddad and Bergek, 2023), in that we emphasize that rigorous evaluations require quantitative analysis of outcomes with respect to a counterfactual (Storey, 2000; OECD, 2008, 2023; Heckman, 2010), and we highlight that a priority for M&E for TIP is obtaining real-time data. We provide some suggestions for (close to) real-time data sources for regional-level M&E.

The paper is divided into 5 sections. After the introduction, section 2 presents the background concepts of transformative innovation policy for sustainability. Section 3 surveys the methodologies applied for monitoring and evaluation, drawing on Storey's (2000) "six steps to heaven" scale, and focusing on the more sophisticated evaluation techniques at the sixth step. Section 4 analyses the setup of monitoring and evaluation exercises in the context of policy interventions, discussing what to monitor and evaluate, and how to design M&E. Section 5 concludes with some recommendations for policy.

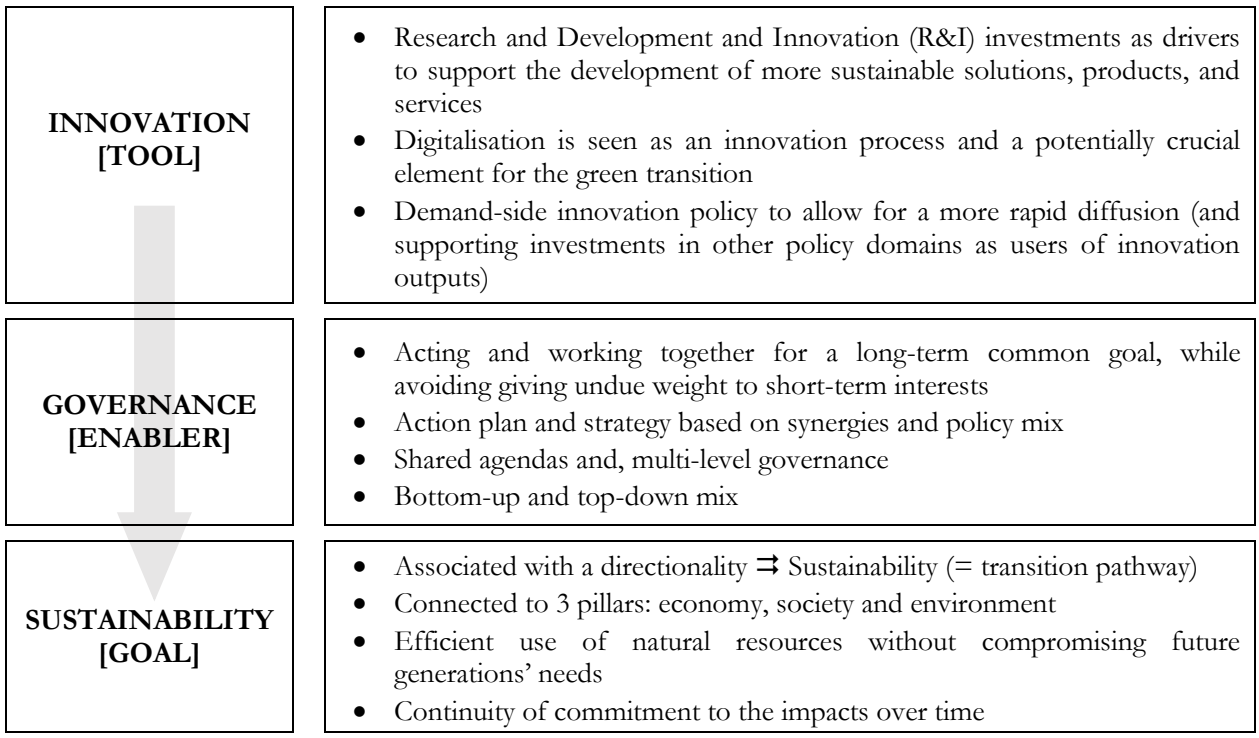
2. Transformative Innovation Policy for sustainability: Setting the Scene

The starting point to effectively design a monitoring and evaluation system lies in understanding the logic of intervention of a policy, programme or instrument. It implies decomposing each of its elements to be able to translate them into dimensions, metrics and indicators to monitor and evaluate.

Innovation Policy (IP) can be defined as a public intervention to have an impact or effect on innovation, including elements of R&D policy, technology policy, infrastructure policy and education policy (Borrás and Edquist, 2019), as well as less obviously-connected areas such as tax policy and immigration policy. It also seen as a way to influence the speed and direction of the innovation process (Borrás and Edquist, 2019).

In the framework of the present study, the direction is sustainability, and innovation is the tool to achieve it, and the governance model is the enabler to achieve the goal (Figure 2). Sustainability refers to lifestyles and socio-economic practices that are compatible with staying within environmental limits, to ensure the safe co-existence of humans with other species on our fragile planet. Sustainable development refers to the socio-economic trajectories to move us from unsustainable present practices towards a sustainable future. Innovation refers to the changes in technologies, processes, and products used by firms to produce goods and services to satisfy consumer demand. In the context of a transition to a sustainable economy, governance lies in a policy-mix between place-based innovation approach (to identify the pathway and specific territorial needs and solutions) and a top-down to complement and support bottom-up approach (Haddad et al., 2022).

Figure 2. Key concepts: definitions and descriptions



Source: Own elaboration.

Various innovation policy instruments are used, such as R&D tax credits, direct R&D grants, regulations regarding intellectual property rights (patents, trademarks, copyright), public procurement, research council funding for scientific research, guaranteed loans, vouchers, state-backed schemes for financing high-tech startups, export assistance schemes, sponsored networking events and matchmaking initiatives, tariffs and customs duties, information provision schemes; as well as standards and regulations (for the protection of consumers, employees and the environment), fines and penalties for undesirable behaviour, outright bans on toxic products, windfall taxes, forced divestiture and break-up of monopoly power guided by competition policy, and so on. A rigorous quantitative evaluation of each of these policy instruments can be attempted at the level of each policy intervention taken individually. Moreover, a rigorous quantitative evaluation can also be applied at a more aggregate level which takes into account possible interactions between policy interventions. An example of a policy interaction would be if firms draw on multiple sources of funding for different aspects of a planned initiative from multiple providers, without this being intended or coordinated at a higher political level (Kroll, 2019:642).

Innovation policy may focus on narrowly-defined areas (e.g. keeping polluting emissions below a certain concentration; keeping energy consumption of specific products below certain fixed levels to comply with environmental standards) or broadly-defined areas, such as regional-level productivity growth. In the latter case of innovation policy with broadly-defined outcomes, we use the term impact-based policy. While traditional innovation policy (IP) is often evaluated in terms of benefits to narrow groups (e.g. whether grant recipients have higher employment growth than non-recipients), impact-based policy focuses on

broader impacts on regions and society. While traditional innovation policy often focuses on narrowly-defined goals (e.g. stimulating patenting activity, encouraging firms to invest more in R&D), transformative innovation policy (TIP) has received growing interest in recent years (Haddad et al., 2022; Meissner and Kergroach, 2021; Magro and Wilson, 2019; Kuhlmann and Rip, 2018; Schot and Steinmueller, 2018; Molas-Gallart et al., 2021) because it seeks to comprehensively address the rapid and many-sided grand challenges required by our societies as we transition towards sustainability. TIP is not put forward here as a sudden and radically new approach to innovation policy, instead its emergence is a gradual and incremental development (Grillitsch et al., 2021). We do not seek to overstate the differences between traditional IP and TIP, although we distinguish between the two here for pedagogical reasons – to have a simpler but clearer storyline.

In this article, we discuss the shift from traditional M&E of IP towards an M&E anchored in the concept of TIP (Schot and Steinmueller, 2018; Molas-Gallart et al., 2021). Table 1 contrasts some characteristics of the standard M&E of IP with M&E based on the TIP approach. While traditional IP focuses on maximizing GDP growth, TIP focuses on a broader range of targets, including environmental sustainability, economic equality, etc, and aligned with the UN SDGs (Sustainable Development Goals). While traditional IP often focuses on firm-level outcomes, TIP is more sensitive to regional-level and societal-level outcomes, where more emphasis is placed on indirect effects and spillovers. (See Appendix 2 for a discussion of the various types of relevant spillovers.)

Table 1. Contrasting some characteristics of standard M&E for Innovation Policy with M&E for Transformative Innovation Policy (TIP)

Dimensions	Traditional Innovation Policy (IP)	Transformative Innovation Policy (TIP)
Main goal	GDP growth	Sustainable Development Goals (SDGs)
Main themes	Internal validity	Spillovers and policy interactions
Number of criteria for evaluation	One output	Multiple outputs
Academic paradigm	Causal econometrics for increasingly well-defined identification of treatment effects in narrow areas	Getting the best available evidence on the most important socio-economic and environmental questions
Research context	Focusing on areas amenable to precise causal identification (e.g. areas where data is rich)	Focusing on areas of economic, social, and environmental importance
Unit of analysis	Individual firms (usually)	Regions and countries, society as a whole
Data availability	Rich data on well-defined (often firm-level) contexts, allowing causal identification of past interventions	Need for real-time data on regionally-aggregated outcomes (which include region-specific spillovers), that is not available from traditional sources such as national statistical offices

Source: Own elaboration.

Similar in spirit to Table 1, Mazzucato (2018, her Table 1) contrasts old and new mission-oriented innovation policy. Old mission-oriented projects focused on technical aspects, decided by a small group of

experts, including a small group of participating firms, and diffusion of the results was either unimportant or discouraged. New mission-oriented projects, however, have "opened up" to focus on addressing societal problems, involving a wide range of stakeholders and actors, with the goal of widely diffusing the results throughout society.

3. Monitoring and Evaluation of Policy: methodological considerations

This section begins with Storey's (2000) "six steps to heaven" scale for monitoring and evaluation of policy interventions, before providing a survey of advanced techniques for policy evaluation, and looking at the current state of policy evaluation in practice.

3.1 Storey's 6 steps to heaven

Innovation policy has a potentially crucial role in the necessary transition from a fossil fuel-based economy to a sustainable economy. However, even the best policies need to be rigorously monitored and evaluated to ensure that they are effectively reaching the intended recipients to have the desired effects. Monitoring can help to steer, calibrate, and fine-tune policy interventions in real time, without waiting for the usual trial-and-error cycle to finish. Without evaluation, there is no way of knowing whether policy interventions are an effective use of tax-payer funds. Tax payers deserve to know whether their funds are being properly administrated.

Storey (2000) wrote that Monitoring and Evaluation of public policy interventions are often not given the attention they need, and this still remains a major challenge for TIP. Evaluations may be undertaken at the end of a programme, when the budget is spent and when time is running out. Ideally, however, the requirements of evaluations would be taken into account at the very start, when collecting data on recipients and non-recipients before the programme has even started (Storey, 2000). Furthermore, the objectives of the policy should be clearly defined (in quantitative terms) before the start of the policy, in the same way that the goalposts should be immobile before a penalty kick is taken.

In the minds of econometricians, it is clear how a program evaluation needs to be set up to obtain rigorous estimates of the treatment effect. However, what ultimately matters is not how program evaluation is beheld in the minds of econometricians, but how program evaluation is beheld in the minds of policymakers and regional authorities, who may not always have formal training in causal econometrics. Somewhere along the way, actual implementation of a policy, as well as implementation of its evaluation, may lead to compromises and modifications such that the final evaluation falls short of the setup that would be ideal for econometricians. Ultimately, it may simply be impossible to set up an evaluation if the initial requirements were not correctly set up (e.g. precise definition of the policy objectives, collecting data on the counterfactual group of non-recipients at the time of policy receipt).

Storey's "6 steps to heaven" provides a useful scale for measuring the appropriateness of monitoring and evaluation exercises, that is shown in Table 2 below. The first three steps correspond to monitoring, which merely either takes a record of activity under the program, or reports participants' perception of the value of the scheme. The second three steps refer to evaluation, i.e. more rigorous attempts to determine the causal impact of the policy initiatives. Storey's scale is a useful reminder that we should not talk about "evaluation" when what is actually being done is merely monitoring. While monitoring (the first three steps) refers to the observation of the numbers of recipients (e.g for accounting purposes, to check that money is being spent correctly), as well as a description of the characteristics and viewpoints of recipients, evaluation (the last three steps) refers to a comparison of recipients to a carefully-selected counterfactual group to obtain quantitative estimates of the treatment effect. Step 6 is the most advanced step that gives the most reliable estimates of the causal effect of the policy intervention. Step 6 is the standard to which policy evaluations should strive to achieve. At a minimum, evaluations should reach stage 5:

> "Governments are failing in their responsibilities to their taxpayers if they continue to finance evaluations less rigorous than those of stage 5" (Storey, 2000:190).

However, the authorities that administered the policy programmes may be biased toward running unsophisticated evaluations, not only because these are cheaper and faster, but also because these are more likely to be biased towards favourably evaluating the overall performance of the policy programme (the "cheap and cheerful" evaluations discussed in Storey, 2000).

Table 2. Storey's "6 steps to heaven" scale for the monitoring and evaluation of policy interventions.

Monitoring	STEP 1	Take up of schemes (numbers collected for simple accounting purposes)
	STEP 2	Recipients opinions on the scheme ("the so-called happy sheets")
	STEP 3	Recipients' views of the difference made by the Assistance
Evaluation	STEP 4	Comparison of the Performance of "Assisted" with "Typical" firms
	STEP 5	Comparison with "matched" firms
	STEP 6	Taking account of selection bias and unobservables

Source: based on Storey (2000); see also OECD (2008, Appendix B).

These days, the rise of computational power, the increasing refinement of econometric techniques for causal inference (Cunningham, 2021), and the richer possibilities for obtaining real-time data in our current age of "big data", all mean that it should be easier to achieve a step 6 evaluation now than when Storey (2000) was published.

Haddad and Bergek (2023, published in Research Policy) propose a framework for the evaluation of TIP. However, the techniques that they propose are primarily qualitative (not quantitative) and do not involve a counterfactual. At best, such schemes would correspond to Step 3, and would be referred to here as

monitoring and not evaluation. We emphasize that innovation policy urgently needs recent (ideally, real-time) data on key outcomes, as well as the application of rigorous techniques for causal inference (See Table 3) and perhaps also involving counterfactuals constructed for programme evaluation from structural models (Heckman, 2010) in order to reach step 6 of Storey's scale. We also emphasize that learning about the causal mechanisms of a policy is a valuable but insufficient endeavour: evaluation requires not only knowledge of the causal mechanisms or 'direction' of a policy initiative, but also a quantification of estimated costs and benefits. Even if the presumed causal mechanisms were functioning as intended, nevertheless if the costs exceed the benefits, then the policy should be evaluated as ineffective.

3.2 Econometric techniques designed for the heavenly 6th step

Table 3 below provides an overview of econometric techniques designed to reach Step 6 (to wit: step 6 is the standard to which policy evaluations should strive to achieve). A variety of econometric techniques are used for causal identification in the mainstream economic literature on policy evaluations (Nichols, 2007; Pless et al., 2020). These techniques are based on the understanding that, for impact assessment, correlations are not enough, and we need to think of causal effects that refer to additionality brought about by innovation policy interventions with reference to a suitable counterfactual.

Table 3. Overview of econometric techniques used for causal identification in the literature on evaluations

METHODS	ADVANTAGES	DRAWBACKS/LIMITATIONS
Randomized Controlled Trial (RCT)	Clear identification of the causal effect of a treatment Often called the "gold standard" (Imbens, 2010)	The researcher must be in control of the experiment: i.e. allocate participants to treatment vs control. This is very expensive in firm-level and region-level economic contexts. Statistical drawbacks include unknown external validity, and RCTs can only identify the mean (not the median, variance, nor percentiles, Deaton and Cartwright, 2018)
Natural experiments	Clear causal identification without the costs of setting up the experiment oneself	Assumptions may not always be valid The research can only focus on contexts where (by chance) a natural experiment has occurred. Limited choice of research topic/context.
Regression Discontinuity Design (RDD)	Clever approach to obtaining causal estimates	Precise requirements regarding the data (e.g. judges' scores of applicants) that are not applicable to most datasets Data collection must often be designed before the actual event
Instrumental Variables	Potentially a way to get causal estimates	The IV must behave like the randomization assignment in an RCT IV assumptions do not get causal estimates from nothing, but require prior causal knowledge (Coad, 2021) IV assumptions cannot be tested. In other words, there is no way of knowing (apart from theoretical assumption) whether or not the IVs are valid (Coad, 2021)

Continues on the next page…

Table 3. Overview of econometric techniques used for causal identification in the literature on evaluations (continuation)

METHODS	ADVANTAGES	DRAWBACKS/LIMITATIONS
Difference-in-difference regressions	Potentially a way to get causal estimates in panel data contexts	Data must be in panel data format Identification is more difficult in staggered treatment settings (Roth et al., 2022) Causal inference is unreliable if assumptions are not met (e.g. the parallel trends assumption)
Matching estimators (Propensity Score matching (PSM), nearest neighbour matching, Coarsened Exact matching (CEM), etc)	Econometric technique to estimate the causal effect of a treatment More precise estimates from a smaller data sample (because poor matches in the treatment and control groups are discarded; Caliendo and Kopeinig, 2008)	Assumption of covariate balance is usually (but not always) satisfied Assumption of unconfoundedness is unlikely to be satisfied, because it is unlikely that all causally relevant confounders are controlled for PSM is generally seen to be less reliable than CEM
OLS regression	Well-known and transparent technique to obtain an estimate of the conditional association between variables	Unlikely to deliver results that allow a causal interpretation

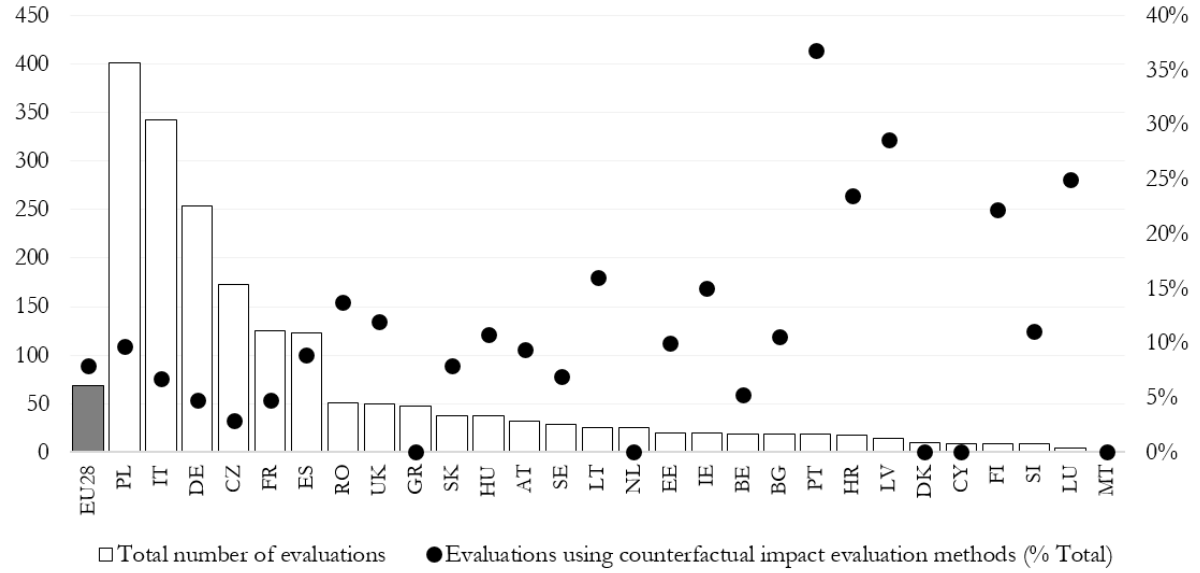
Notes: techniques are arranged in an order inspired by the hierarchy of internal validity in Nichols (2007), with more reliable techniques appearing nearer the top.

3.3 The quality of programme evaluations in practice

In practice, when policymakers, governments or their subcontractors are performing programme evaluation, the use of most sophisticated techniques as counterfactual impact evaluation (CIE) methods (e.g. matching estimators and difference-in-difference regressions) have only been used in about 8% of the Cohesion Policy's evaluations performed by EU members since 2015 (Figure 3). These methods presumably correspond at least to step 5 in Storey's scale. However, this degree of sophistication is strongly heterogeneous across EU countries, as well as the quantity of evaluations performed by the different EU member states. For instance, in Poland more than 400 evaluations to Cohesion Policy 2014-2020 were performed and 10% of them conducted with CIE methods, whereas in Portugal, despite a number of 19 evaluations only, 37% were conducted with CIE methods (Figure 3).

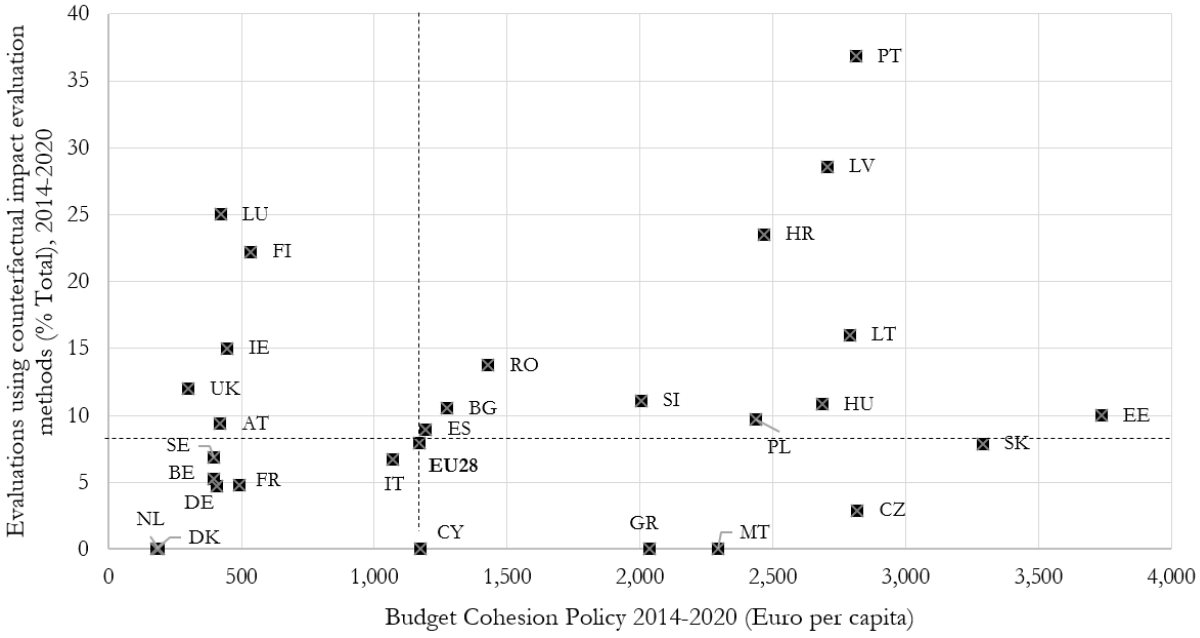
The degree of sophistication of the evaluation techniques seems also not be strongly associated with the Cohesion Policy budget allocated to the countries in 2014-2020, even if, on average, we observe that higher the budget per capita, the higher the intensity of sophisticated techniques used for Cohesion Policy evaluation, with some exceptions, namely Slovakia, Czech Republic, Malta and Greece (Figure 4).

Figure 3. Relationship between the number of Cohesion Policy evaluations (2014-2020) and the use of counterfactual impact evaluation techniques (% Total), by EU28 country



Source: Own elaboration based on data from Cohesion Open Data Platform (extracted on 05/12/2022).
Note: Figure refer to evaluation for the programming period 2014-2020 conducted since 2015. The original dataset does not include ex-ante evaluations for 2014-2020 (Art. 55 Common Provisions Regulation – Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013).

Figure 4. Relationship between the number of Cohesion Policy evaluations (2014-2020) and the use of counterfactual impact evaluation techniques (% Total), by EU28 country

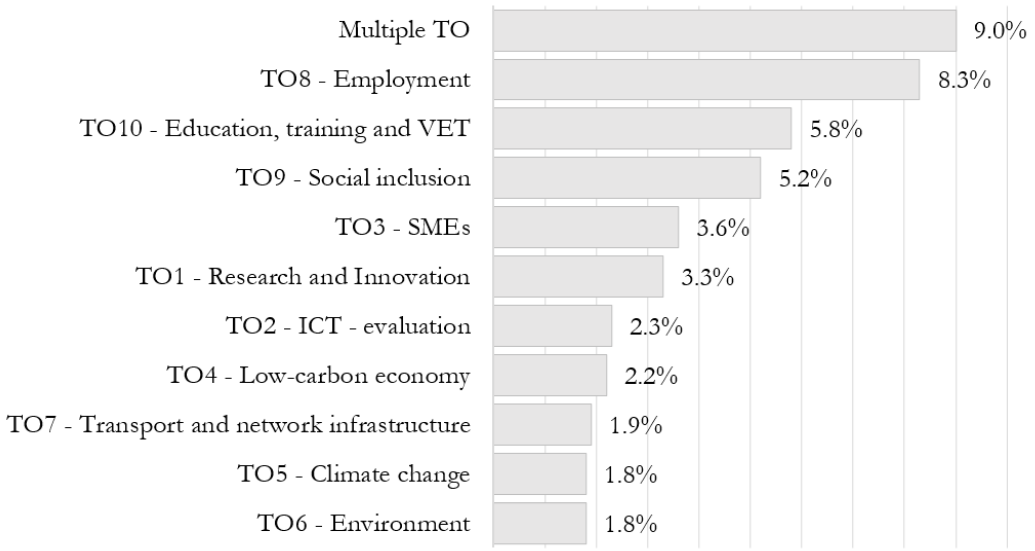


Source: Own elaboration based on data from Cohesion Open Data Platform (extracted on 05/12/2022).

Note: Figure refer to evaluation for the programming period 2014-2020 conducted since 2015. The original dataset does not include ex-ante evaluations for 2014-2020 (Art. 55 Common Provisions Regulation – Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013).

When looking at the counterfactual impact evaluation techniques by thematic objectives (Figure 5) of Cohesion Policy funds, we notice a higher concentration of the use of most sophisticated methods in programmes associated with social goals (e.g. employment and education), whereas the ones targeting environmental and climate change issues are among those with the less use of such more advanced evaluation methods.

Figure 5. The percentage of evaluations using counterfactual impact techniques by thematic objectives (TO), (% Total) Cohesion Policy 2014-2020, EU28



Source: Own elaboration based on data from Cohesion Open Data Platform (extracted on 05/12/2022).
Note: Figure refer to evaluation for the programming period 2014-2020 conducted since 2015. The original dataset does not include ex-ante evaluations for 2014-2020 (Art. 55 Common Provisions Regulation – Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013).

The analysis in this sub-section suggests that the evaluations completed by the EU Member States are still far from step 6 (i.e. "heaven") in Storey's scale, especially when we zoom in on areas related to environment or innovation dimensions of the EU Cohesion Policy. With the present paper, we propose a framework to make M&E more accurate and supportive of the policy cycle, especially considering that the EC aims to become a climate neutral economy by 2050, with innovation playing an important role in the transition.

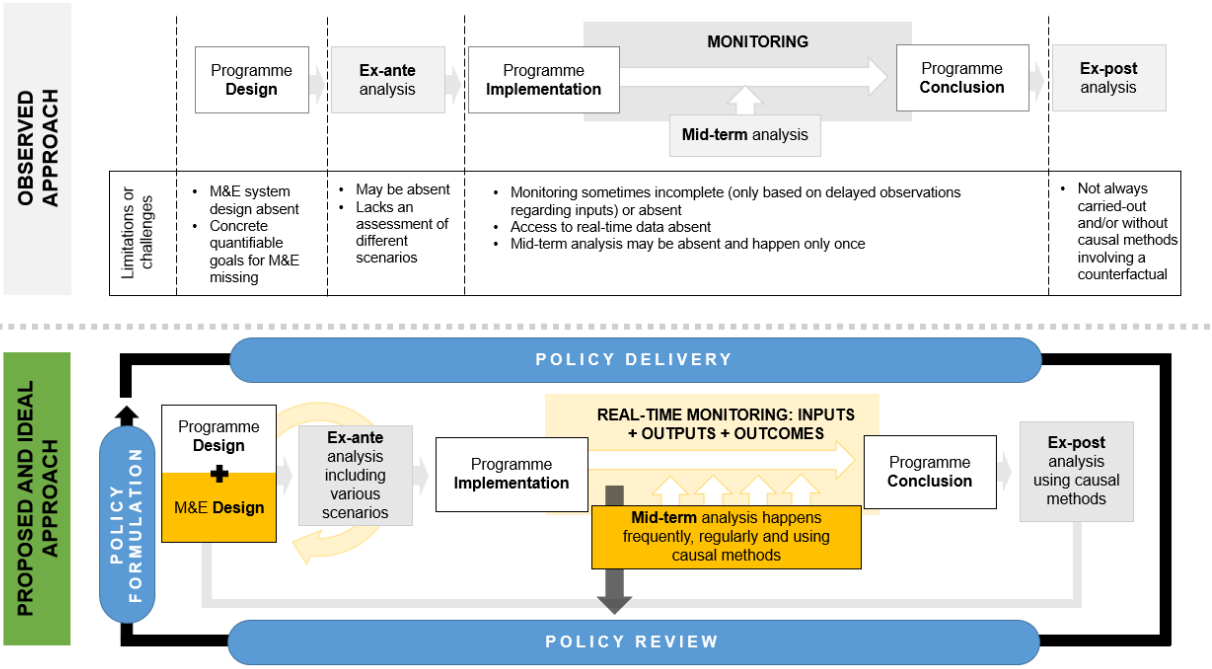
4. Proposed framework

4.1. How to design the M&E system?

In our proposed framework, M&E are key elements in the different phases of the policy cycle, and they are more than a legal obligation. They are part of the policy intelligence, and lessons learned from previous

evaluations should support programme design and/or policy formulation, instead of having a M&E system appearing after the programme implementation stage only (see Figure 6, below). Consequently, and following Molas-Gallart et al. (2021), we highlight the need of designing and implementing a M&E system in parallel with programme design, to support the definition of quantifiable goals and policy actions to achieve them. Designing a M&E implies to plan the different steps and players of the M&E system, namely, to define the indicators, the source for data collection and the methods for evaluation.

Figure 6. Top: Unsophisticated M&E. Bottom: Proposed and ideal M&E system



Source: Own elaboration based on data analysis about Cohesion Policy evaluation – Cohesion Open Data Platform (Top figure), European Commission (2013), Shahab et al. (2019) and previous considerations (Bottom figure).

Our proposed framework also implies to involve stakeholders to support M&E, as preconized by Molas-Gallart et al. (2021), as well as the beneficiaries of the R&I funds. This feature may allow to facilitate data collection for M&E (e.g. through a better stakeholders and beneficiaries interactions with managing authorities), as well as to ensure citizens ownership, engagement and acceptability of the policy actions. The engagement of stakeholders in M&E may help to better define the objectives of the policy intervention, collecting data behind the existing ones and to better interpret the results achieved (UNESCO, 2011). Citizens' involvement in the monitoring of government expenditures is crucial because they may help ensuring that public resources are used effectively and efficiently, and can lead to improved government accountability and transparency (Ramkumar, 2008). Citizens are also considered an important element in the transition pathway, as users and drivers of innovation, for instance, by influencing market needs, e.g. pushing firms to innovate in a given direction (Trischler et al, 2022). Related to this, we also see the need to include additional evaluation criteria (e.g. equity and acceptability) alongside the traditional

ones (relevance, coherence, efficiency, effectiveness, utility and sustainability). For more details about the list criteria, see Table 4. Our approach also follows Adler's (2012) suggestion that policy evaluation should include criteria such as well-being and equitable distribution in addition to standard cost-benefit analysis.

Table 4. Definition of the most relevant evaluation criteria

Criteria	Definition
Relevance	• Justification of the strategy or priorities chosen • Based on socio-economic-sustainable needs which can evolve and be revised in mid-term analysis • Focus: objective-needs relationships
Coherence	• Compatibility of the intervention with other intervention(s) in a country/region
Acceptability	• Support of policy design and implementation by society, decision-makers and decision-takers
Equity	• Intragenerational and intergenerational effects, and associated with fair distribution
Efficiency	• Optimal use of resources • Focus: input-output relationship ➡ output maximum and minimum input
Effectiveness	• Success of resources used to achieve objectives and goals • Focus: objectives-outcomes relationship
Utility	• Focus: effect-needs relationship
Sustainability	• Durability and continuity of the effects

Source*:* Own elaboration based on European Commission (2013) and Shahab *et al.* (2019).

The key differences between a so-called traditional M&E (and observed by EU Member States) and an improved M&E are summarized below:

- M&E does not take place uniquely at the end of the policy implementation, but throughout, as highlighted by the references to ex ante evaluation, mid-term evaluation, and ex post evaluation
- Objectives of the policy, and the M&E strategy, are determined and stated before the start of the policy's implementation
- At the start, relevant variables are collected, relating to the stated objectives (intended effects) as well as other indicators of interest (to check whether there may have been some unintended effects, e.g. using placebo tests)
- At the start, data is collected on the recipients (treatment group) as well as a suitable group of non-recipients (control group, for analysis of counterfactuals)

- The design of the policy may be refined through various iterations at the start, but the design and objectives become fixed before the start of the policy, to ensure accurate evaluation
- Monitoring takes place continuously throughout the policy implementation, ideally using high-frequency data from low-cost data sources shown on publicly-available dashboards
- The final evaluation links back to the previously-stated objectives, satisfying the requirements of Storey's step 6
- The final evaluation feeds back into a policy review and subsequent policy formulation

We also recognize that a rigorous evaluation requires committing to a data collection plan before the start of the policy implementation, that includes variables (unintended effects) and actors (as counterfactuals) and application-stage information on the ranking scores of applicants[iii] that are not even intended to be affected by the policy. This deliberate collection of seemingly irrelevant data may not be immediately obvious to regional authorities, although it is a requirement of rigorous evaluation. Relevant here are the recent developments in causal inference regarding "placebo tests" (i.e. checking whether the expected null effects are actually found when looking in places where effects should be absent (Athey and Imbens, 2017). Placebo testing could be a useful way of checking whether the intended causal mechanisms are in place, as well as checking for indirect effects of policies. Placebo testing could benefit from data that is not directly related to the stated policy goals, but which is nevertheless of interest to the economy and society in general.

Figure 6 (bottom) presents M&E activities referred to as *ex-ante* evaluation, mid-term evaluation, and *ex-post* evaluation. Only the latter (i.e. *ex-post* evaluation) corresponds to an "evaluation" in the sense of Storey's (2000) framework, however.

- *Ex-ante* evaluation refers to a plan or analysis set up to ensure that the intervention is relevant and coherent, to provide a prior assessment regarding the design and expected impacts of the assessment (European Commission, 2013). An *ex-ante* evaluation may include simulation analysis, with various scenarios, to explore relevant tradeoffs and to compare a simulated treatment group outcome with a simulated counterfactual. An ex ante evaluation is not an "evaluation" in the sense of Table 2, because at the time of an *ex-ante* evaluation, the data is not available for a rigorous comparison of performance outcomes with respect to a counterfactual.
- Mid-term evaluation refers to M&E activity that occurs mid-way through the intervention's implementation, to provide feedback and improve the intervention's management. To the extent that mid-term evaluation does not make use of the final data on outcomes (for neither the treated

[iii] Information on the application-stage ranking of applicants (i.e. the scores given by the committee of judges who decide who gets funding) is an important requirement of Regression Discontinuity design, an econometric technique discussed in Table 3 that obtains causal estimates by comparing only-just-qualified with near-misses, under the assumption that the only difference between these two is the receipt of the policy support.

nor control groups) and prioritizes speed over counterfactual analysis, a mid-term evaluation would be categorized by Table 2 as being more of a monitoring exercise than an evaluation.

4.2. What to monitor?

Monitoring is about describing and observing, whereas evaluation is about reaching a judgment on the overall value of an activity. Monitoring can be defined as:

> "The continuous process of examining the context of the programme and the delivery of programme outputs to intended beneficiaries, which is carried out during the execution of a programme with the intention of immediately correcting any deviation from operational objectives." (European Commission, 2013).

According to Storey (2000:180), monitoring: "merely either documents activity under the program or reports participants' perception of the value of the scheme" (Storey, 2000) and therefore is no substitute for an evaluation. Monitoring focuses more on checking up on the allocation of inputs, whereas evaluation focuses on decisions regarding the overall satisfaction with the outputs.

For example, regional authorities implementing innovation policies may discuss these policies in terms of total amount of resources invested in the policy (e.g. "our region has invested EUR100m in this…"). This is an input indicator though, not an output. Just because a lot of money is being spent, this is in itself no guarantee that the money is being spent wisely and effectively.[iv]

Monitoring takes place (ideally) *during* the policy programme and not just at the end of the programme. Ideally, at the end of the process there would be a rigorous evaluation (rather than just monitoring). Clearly, monitoring requires real-time (or recent) data. The usual lags (of perhaps 3 years – e.g. as happens with access to regional data at EU level) affecting the release of data (such as representative firm-level or regionally-aggregated census data) from national statistical offices mean that such data sources are not compatible with the requirements of project monitoring. A priority for TIP is to enrich monitoring with real-time information from big-data-era analytics (e.g. analysis of media or social media). Appendix 1 provides an overview of some potentially-interesting real-time data sources for regional-level socio-economic monitoring.

Under a so-called "traditional approach", monitoring refers to a periodic process of collecting data on inputs used to implement the programme (e.g. R&I subsidies) and the outputs generated by them (e.g. private R&D investment carried out by subsidized firms). In the context of TIP, it must go beyond and also focus on collecting evidence of the outcomes or impacts (e.g. the effect of the policy intervention). In the context of the present study, and based on European Commission (2013), we define inputs as all the financial, human, material, organisational and regulatory resources mobilised to implement the policy

[iv] A more useful statistic, if it could be calculated via an appropriate evaluation, would be something like "our region has invested EUR100m in this and generated EUR120m in benefits, corresponding to a rate of return on investment of 20%".

actions, and outputs as the indicator describing the "product" generated by the use of the inputs. Outcomes are associated with the effect quantified using counterfactual analysis, and measured before and after the policy intervention, and between treated and non-treated groups, and related to the concept of policy impact (European Commission, 2013).

Monitoring should be carried out during the execution phase of a programme or policy intervention, with the aim of correcting any deviation from desired objectives/goals. Therefore, our proposed framework differs in terms of scope, dimensions and focus of analysis in the basis of its singular characteristics in comparison with a traditional (or observed) approach, as summarized below:

- Instead of monitoring achievements, measured by indicators associated to subsidized beneficiaries, it should focus on the monitoring of outcomes and impacts
- It should also screen the spillovers at territorial-level to assess not only the desired effects but also the non-desired effects
- To include more than socio-economic indicators, namely environmental and governance indicators (e.g. policy-mix – synergies and complementarities – and the multi-level perspective)

4.3. What and how to evaluate?

Evaluation refers to the process of determining the success or failure of a policy or programme (European Commission, 2013). In the so-called traditional approach there are "in theory" three main different types of evaluation:

i) *ex-ante evaluation* conducted before programme implementation to ensure that an intervention is relevant and coherent;

ii) *mid-term evaluation* carried out once during the period of implementation;

iii) *ex-post evaluation* aiming to account for the achievement of expected goals (effectiveness), the efficiency of interventions and the sustainability of impacts.

The proposed evaluation framework follows a traditional approach but including some extra steps and a more continuous process. It aims to go *hand by hand* with the evaluation of investment projects to be implemented in the territory and the monitoring process. The starting point lies in defining the expected impact(s) and then to design the programme/policy intervention and to identity the inputs to achieve it (or them). As a methodological approach and to satisfy the requirements of Storey's step 6, the evaluation process should allow to estimate the effects of the policy using counterfactual methods and to quantify not only the direct but also the indirect effects. In line with Adler (2012) who argues that policies should be evaluated based on their impact on the well-being of individuals and society as a whole, therefore, should include more than economic factors.

We also believe that a single and unique methodological approach is not enough to cover the complexity of the TIP, therefore, we recommend using a mix of techniques and methods for evaluation, combining counterfactual analysis with other qualitative and quantitative approaches. Qualitative methods refer to surveys, interviews, focus groups and case studies. They can help identify important variables and mechanisms to measure in quantitative analysis. Furthermore, the counterfactual situation (without policy intervention) can potentially be designed or estimated thanks to specific questions included in the interviews or focus group. Quantitative methods include combining counterfactual analysis with other techniques such as Multi-Criteria Analysis (MCA) and Cost-Benefit Analysis (CBA). For more details about when to use MCA and CBA, see Table 5.

Table 5. Multi-Criteria Analysis and Cost-Benefit Analysis: When to use them?

Method	Definition	When to use it?
Multi-Criteria Analysis (MCA)	• Tool used to compare several interventions in relation to several criteria • It may involve weighting, reflecting the relative importance attributed to each of the criteria	• *Ex-ante* evaluation for comparing policy options or projects proposals for clarification purpose • *Ex-post* evaluation to compare the relative success of the different components of the intervention
Cost-Benefit Analysis (CBA)	• Tool for judging the advantages of the intervention from the point of view of all the groups concerned, and on the basis of a monetary value attributed to all the positive and negative consequences of the intervention • It estimates a fictive price or the willingness of beneficiaries to pay to obtain positive impacts or avoid negative ones. It can also be estimated by the loss of earnings in the absence of the intervention	• Usually for ex-ante evaluation • When it is not possible to use market price to estimate a gain or a loss of an intervention • To reveal missing information of the effect of the intervention • It allows (even if challenging) to express environmental impacts in monetary terms • Should be used with multi-criteria analysis

Source*:* Own elaboration based on European Commission (2013), Gamper and Turcanu (2007), Hanley and Barbier (2009).

5. Conclusion and policy challenges

Transformative Innovation Policy (TIP) has an important role in the sudden transition that our economies require to face up to today's grand challenges (climate change, sustainable development goals). This paper discussed various themes relating to the monitoring and evaluation (M&E) of TIP. There is a world of difference between high-level strategic intentions, and how these grand visions are translated and implemented by policymakers into actual policy initiatives, that may or may not have effects that are consistent and effective (Kroll, 2019). Monitoring and Evaluation, therefore, is vital to check the functioning of the long chain of steps from grand strategic visions to changes in real-world behaviour that are genuinely caused by policy interventions. To begin with, the various degrees of sophistication that can

be found in monitoring and evaluation exercises (i.e. Storey's "6 steps to heaven" scale) were discussed. These M&E activities range from interviews asking recipients whether they are happy to receive funding, to full-blown causal econometric analyses such as Randomized Controlled Trials (RCTs). We then provide a survey of causal inference techniques that reach the 6th step on this scale, and analyse the degree of sophistication of recent EU Cohesion project evaluations. Our analysis shows that the use of most sophisticated techniques as counterfactual impact evaluation methods have only been used for few Cohesion Policy's evaluations performed by EU members since 2015, and the likelihood to use these techniques is even lower when the evaluation targets thematic objectives associated with green dimensions of the policy intervention.

Our discussion led to a variety of recommendations for improving M&E activities in the context of TIP in the modern era. Considerations of evaluating a policy should not occur exclusively at the end of a policy programme, but should occur at all stages, and feed into the design of the policy programme before it even starts. Furthermore, the policy goals should be clearly stated before the start of the policy programme (in reality, this is not always done). Monitoring should ideally take place throughout the policy programme, using publicly-available dashboards that are updated with real-time data. Data should be collected throughout on seemingly-irrelevant variables, such as data on non-recipients of policy support (to create a counterfactual control group for comparisons) and also for outcome variables that are not officially presented as policy objectives (as a check for possible unintended indirect effects of the policy).

Like all research, our paper is not without limitations. We acknowledge that the present paper does not aspire to fully cover all the literature and dimensions of the TIP, indeed, it only aims to cover the M&E dimension of the TIP. Nevertheless, we hope that this paper contributes to improving discussions of monitoring and evaluation in the area of TIP among innovation scholars, and that it can be useful to policymakers who have the crucially important task of implementing monitoring and evaluation of real-world innovation policies during these times requiring urgent change.

References

Adler, M. (2012), Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis, Oxford: Oxford University Press

AON (2022). 2021 Weather, Climate and Catastrophe Insight. Available here: https://www.aon.com/weather-climate-catastrophe/index.aspx [Accessed on 29/01/2022]

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives, 31(2), 3-32.

Barca, F.; McCann, P. and Rodríguez-Pose, A. (2012). "The case for regional development intervention: place-based versus place-neutral approaches", *Journal of Regional Science* 52(1): 134-152.

Bjerke, M. B., & Renger, R. (2017). Being smart about writing SMART objectives. Evaluation and Program Planning, 61, 125-127.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. Annual Review of Economics, 10, 615-643.

Borrás, S. and Edquist, C. (2019). Holistic Innovation Policy: Theoretical Foundations, Policy Problems, and Instrument Choices, Oxford University Press. DOI:10.1093/oso/9780198809807.001.0001

Bortoli, C., Combes, S., & Renault, T. (2018). Nowcasting GDP growth by reading newspapers. Economie et Statistique, 505(1), 17-33.

Caliendo M., Kopeinig S., (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys 22, 1, 31-72

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science advances, 1(5), e1400253.

Ceron, A., & Negri, F. (2016). The "social side" of public policy: Monitoring online public opinion and its mobilization during the policy cycle. Policy & Internet, 8(2), 131-147.

Chiou, T., Chan, H.K., Lettice, F & Chung, S.H. (2011). The influence of greening the suppliers and green innovation on environmental performance and competitive advantage in Taiwan, Transportation Research Part E: Logistics and Transportation Review, 47:822–836.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88, 2-9.

Coad, A. (2021). Econometrics and the growth of firms: perspectives from evolutionary economics. Strategy Science, 6 (4), 338–352. https://doi.org/10.1287/stsc.2021.0132

Conding, J. & Habidin, N.F. (2012). The structural analysis of green innovation and green performance in Malaysian automotive industry, Research Journal of Finance and Accounting, 3:172–178.

Cowling, M. and Liu, W. (2021). "Access to Finance for Cleantech Innovation and Investment: Evidence from U.K. Small- and Medium-Sized Enterprises," in IEEE Transactions on Engineering Management, doi: 10.1109/TEM.2021.3066685.

de Rassenfosse, G., Kozak, J. & Seliger, F. Geocoding of worldwide patent data. Sci Data 6, 260 (2019). https://doi.org/10.1038/s41597-019-0264-6

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, 2-21.

Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an

emerging policy paradigm. Research Policy, 48(4), 880-894.

Edler J., Fagerberg J., (2017). Innovation policy: what, why, and how, Oxford Review of Economic Policy, 33, (1), 2–23. https://doi.org/10.1093/oxrep/grx001

European Commission (2013). EVALSED: The resource for the evaluation of Socio-Economic Development. Available here: https://ec.europa.eu/regional_policy/en/information/publications/evaluations-guidance-documents/2013/evalsed-the-resource-for-the-evaluation-of-socio-economic-development-evaluation-guide

European Parliament (2021). "Innovation Policy", Fact Sheets on the European Union.

Fagerberg, J. (2016). "Innovation Policy: Rationales, Lessons and Challenges", Journal of Economic Surveys 31(2): 497-512. https://doi.org/10.1111/joes.12164

Fankhauser, S. and Tol, R.S.J. (2005). "On climate change and economic growth", Resource and Energy Economics, 27(1):1-17, https://doi.org/10.1016/j.reseneeco.2004.03.003.

Fankhauser, S.; Sehlleier, F. and Stern, N. (2008). "Climate change, innovation and jobs", Climate Policy, 8(4):421-429. DOI: 10.3763/cpol.2008.0513

Foray, D., P. David, and B. Hall. 2009. "Smart Specialisation – The concept", Knowledge Economists Policy Brief n° 9.

Forsythe, E., Kahn, L. B., Lange, F., & Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. Journal of Public Economics, 189, 104238.

Grillitsch, M., Hansen, T., & Madsen, S. (2021). Transformative innovation policy: a novel approach? In: Handbook on Alternative Theories of Innovation (pp. 276-291). Edward Elgar Publishing.

Haddad, C. R., & Bergek, A. (2023). Towards an integrated framework for evaluating transformative innovation policy. Research Policy, 52(2), 104676.

Haddad, C. R., Nakić, V., Bergek, A., & Hellsmark, H. (2022). Transformative innovation policy: A systematic review. Environmental Innovation and Societal Transitions, 43, 14-40.

Haines, A.R.S.; Kovats, D.; Campbell-Lendrum and C. Corvalan (2006). "Climate change and human health: Impacts, vulnerability and public health", Public Health, 120(7): 585-596. https://doi.org/10.1016/j.puhe.2006.01.002.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. Journal of Economic Literature, 48(2), 356-98.

Imbens G.W., (2010). Better LATE than nothing: some comments on Deaton 2009 and Heckman and Urzua 2009. Journal of Economic Literature 48, 399-423.

IPCC (2021). Climate Change 2021 The Physical Science Basis, Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Janssen, M. J., Bergek, A., & Wesseling, J. H. (2022). Evaluating systemic innovation and transition programmes: Towards a culture of learning. PLOS Sustainability and Transformation, 1(3), e0000008.

Kroll H. (2019). How to evaluate innovation strategies with a transformative ambition? A proposal for a structured, process-based approach. Science and Public Policy, 46:635–47. https://doi.org/10.1093/scipol/scz016

Küçükoğlu, M.T., Pınar, R.İ. (2015). "Positive Influences of Green Innovation on Company Performance", Procedia-Social and Behavioral Sciences, 195: 1232-1237.

https://doi.org/10.1016/j.sbspro.2015.06.261.

Kuhlmann, S., & Rip, A. (2018). Next-generation innovation policy and grand challenges. Science and Public Policy, 45(4), 448-454.

Magro, E., & Wilson, J. R. (2019). Policy-mix evaluation: Governance challenges from new place-based innovation policies. Research Policy, 48(10), 103612.

Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. Industrial and Corporate Change, 27(5), 803-815.

Mccann, P. and Soete, L., (2020). Place-based innovation for sustainability. Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-20392-6, doi:10.2760/250023, JRC121271.

Meissner, D., & Kergroach, S. (2021). Innovation policy mix: mapping and measurement. Journal of Technology Transfer, 46(1), 197-222.

Molas-Gallart, J., Boni, A., Giachi, S., & Schot, J. (2021). A formative approach to the evaluation of Transformative Innovation Policies. Research Evaluation, 30(4), 431-442.

Mowery, D. C., Nelson, R. R., & Martin, B. R. (2010). Technology policy and global warming: Why new policy models are needed (or why putting new wine in old bottles won't work). Research Policy, 39(8), 1011-1023.

Neumark, D. and Simpson, H. (2015). "Place-Based Policies". In (Eds) Duranton, G.; Henderson, J.V. and Strange, W.C, Handbook of Regional and Urban Economics, Vol 5, pp. 1197-1287,

Nichols A., (2007). Causal inference with observational data. Stata Journal 7 (4), 507-541.

OECD (2008). OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes, OECD Publishing, Paris, https://doi.org/10.1787/9789264040090-en.

OECD (2023), Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes 2023, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://doi.org/10.1787/a4c818d1-en

Parry I., Black S., Vernon N., (2021). Still not getting energy prices right: A global and country update of fossil fuel subsidies. IMF working paper 2021/236. https://www.imf.org/en/Publications/WP/Issues/2021/09/23/Still-Not-Getting-Energy-Prices-Right-A-Global-and-Country-Update-of-Fossil-Fuel-Subsidies-466004

Patz, J., Campbell-Lendrum, D., Holloway, T. et al. (2005). "Impact of regional climate change on human health", Nature, 438: 310–317. https://doi.org/10.1038/nature04188

Pless, J., Hepburn, C., & Farrell, N. (2020). Bringing rigour to energy innovation policy evaluation. Nature Energy, 5(4), 284-290.

Polzin, F. and Sanders, M. (2020). "How to finance the transition to low-carbon energy in Europe?", Energy Policy, 147. https://doi.org/10.1016/j.enpol.2020.111863

Pontikakis, D., Gonzalez Vazquez, I., Bianchi, G., Ranga, L., Marques Santos, A., Reimeris, R., Mifsud, S., Morgan, K., Madrid Gonzalez, C. and Stierna, K., (2022). Partnerships for Regional Innovation Playbook, EUR 31064 EN, Publications Office of the European Union, Luxembourg. ISBN 978-92-76-52326-0, doi:10.2760/292307, JRC129327. Available online: https://s3platform.jrc.ec.europa.eu/pri-playbook

Ramkumar, V. (2008). Our Money, Our Responsibility – A citizens' guide to monitoring government expenditures, International Budget Project, Washington, DC (https://internationalbudget.org/wp-

content/uploads/Our-Money-Our-Responsibility-A-Citizens-Guide-to-Monitoring-Government-Expenditures-English.pdf).

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. arXiv preprint arXiv:2201.01194.

Schot, J., & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. Research Policy, 47(9), 1554-1567.

Shahab, S., Clinch, J. P., & O'Neill, E. (2019). Impact-based planning evaluation: Advancing normative criteria for policy analysis. Environment and Planning B: Urban Analytics and City Science, 46(3), 534-550.

Statista (2021). Greenhouse gases emissions.

Storey, D. J. (2000). Six steps to heaven: evaluating the impact of public policies to support small businesses in developed economies. Chapter 9 (pp176-197) in: Sexton DL, Landström H., (Editors), The Blackwell Handbook of Entrepreneurship. Blackwell publishers: Oxford, UK.

Trischler, J.; Svensson, P.O; Williams, H. and Wikström, F. (2022). "Citizens as an innovation source in sustainability transitions – linking the directionality of innovations with the locus of the problem in transformative innovation policy", *Public Management Review*, DOI: 10.1080/14719037.2022.2062041

UNESCO (2011), 'Results-based programming, management, monitoring and reporting (RBM) approach as applied at UNESCO: guiding principles' (https://focusintl.com/data/documents/RBM012-177568e.pdf).

Van der Wielen, W., & Barrios, S. (2021). Economic sentiment during the COVID pandemic: Evidence from search behaviour in the EU. Journal of Economics and Business, 115, 105970.

Xiang, X.; Liu, C.; Yang, M. (2022). "Who is financing corporate green innovation?", International Review of Economics & Finance, 78:321-337.https://doi.org/10.1016/j.iref.2021.12.011

Zhao, Q.; Guo, Y.; Ye, T.; Gasparrini, A. et al. (2021). "Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study", Lancet Planet Health, 5:e415–25. DOI: https://doi.org/10.1016/S2542-5196(21)00081-4

Appendix

Appendix 1. An overview of real-time data sources for regional-level socio-economic monitoring

Indicators should be SMART (Specific, Measurable, Achievable, Relevant and Time-bound) (Bjerke and Renger, 2017). Ideally, monitoring and evaluation of transformative innovation policy (TIP) would be facilitated by a user-friendly dashboard that offers real-time updates on the circumstances and performance of regions. For this, TIP requires real-time data. Such data would need to be updated regularly to remain current. However, manually updating such a TIP dashboard would have high maintenance costs (in terms of working hours) and would be infeasible. Instead, TIP would ideally update the user interface dashboard with data that is passively collected from online sources.

Conventional economic data (e.g. from usual data sources such as National Statistical Offices) usually becomes available after a long time lag of perhaps 2-3 years. Such time lags are problematic if the task is to provide an up-to-date user dashboard that features real-time data. Therefore, a different approach is needed, putting together data from various real-time sources.

- Real-time air quality data from the EEA (European Environmental Agency): https://www.eea.europa.eu/themes/air/air-quality-index . EEA provides recent data for a variety of environmental indicators here: https://www.eea.europa.eu/ims . In addition to data on air quality, future research might potentially collect data on water quality, water consumption, total quantity of waste, and percentage of energy from renewables.

- Google search data can be analyzed for the purposes of forecasting and nowcasting, as described in Choi and Varian (2012). This data has been analyzed by researchers at the EC-JRC (van der Wielen and Barrios, 2021) to explore economic sentiment at the country-level in Europe during the COVID pandemic. Google Trends data allows for data to be downloaded at the level of NUTS-3 regions, which is particularly interesting for TIP M&E. Perhaps Google search data could be used to analyze consumer preferences and behaviour regarding eco-friendly products and solutions.

- Nowcasting regarding labour market conditions using information on job vacancies. Forsythe et al (2020) use data from vacancy postings and unemployment insurance claims to get close to real-time indicators on labour demand at the time of the COVID shock. See also this NESTA work using firm website data.[v] Such work could focus on the number of jobs being advertised, as well as the type of jobs and skills being demanded. Data on job vacancies posted online (e.g. on LinkedIn) that give details regarding the skills required for specific jobs (e.g. Python, R software) could potentially be analysed, which could be of interest given the skill requirements of the twin

[v]

transition. Future research could potentially investigate data from universities (graduation rates, numbers of students, grades and subjects taken, courses on offer) to check alignment between the supply of education and the requirements for facing the twin transition.

- Social media can be a useful source of real-time data for policymakers (Ceron and Negri, 2016). This could show real-time information on overall sentiment, for example; Hedonometer provides sentiment analysis from Twitter in various languages, here is an example for tweets in French: https://hedonometer.org/timeseries/fr_all/?from=2021-04-17&to=2022-10-16 . Social media could potentially be used to analyse attitudes towards sustainability and environmental issues, although we might also expect lots of greenwashing and misinformation here.

- European Media Monitor (EMM) monitors thousands of online news sources across the world. EMM uses advanced information extraction techniques to automatically determine what is being reported in the news, where things are happening, who is involved and what they said. https://emm.newsbrief.eu/overview.html

- The interest of investors in the green economy could be monitored by tracking the performance of sustainable investment funds in financial markets. For example, Santander launched its "Santander Sostenible" range targeted at investors wishing to invest in sustainable and green areas: https://www.bancosantander.es/en/particulares/ahorro-inversion/fondos-inversion/sostenibles-responsables Such funds can be tracked in real time on e.g. Yahoo finance: https://finance.yahoo.com/

- Nowcasting on public health data,[vi] such as real-time data on daily COVID cases https://ourworldindata.org/covid-cases). Perhaps data on other aspects of public health might be available, such as death rates, suicide rates, sickness rates from hospital, regional incidence of specific types of cancer death.

- GDP nowcasting: Bok et al (2018) provide a comprehensive overview of the nowcasting of GDP done by the New York Fed Staff Nowcast. GDP growth can be nowcasted using text analysis of newspapers (Bortoli et al., 2018).

- Regional-level outputs in terms of academic research in scientific areas relating to sustainability and green transitions could potentially be tracked using university-level information on the latest research outputs. For example, here is a list of researchers at the University of Sevilla: https://scholar.google.co.jp/citations?view_op=view_org&hl=en&org=3333805543372288621. However, it seems that while google scholar offers thematic alerts, nevertheless google scholar does not offer alerts specifically relating to outputs from specific universities. It might be possible to write a data-collection routine that collects information on research outputs by individuals at the research institutes in a region, to have a real-time region-level aggregated indicator of research

[vi] Public health may also be affected by climate change, of course.

contributions by a region's universities/research institutes in a given area (i.e. sustainability and green tech).

- Patent data contains geographical information, and has been geocoded (de Rassenfosse et al., 2019), potentially allowing for regional-level analysis of patenting and technological capabilities.

Appendix 2. Types of spillovers

Table A2. Types of spillovers

Type	Description
Firm-to-firm knowledge spillovers	Poaching employees, employee mobility, imitation, reverse engineering, benefits of co-location
Firm-to-firm competitive interactions	Benchmarking, catch-up, strategic imitation (e.g. copycat products, similar R&D investment rules, similar employee compensation schemes) RCTs (Randomized Controlled Trials) in the development literature and the SUTVA assumption (Stable Unit Treatment Values Assumption), which might be violated if treated firms win at the expense of the non-treated in the context of a zero-sum regional market.
Regional spillovers	Investments in one region may benefit another For example, regional knowledge spillovers if regions benefit from the investments in education and R&I from neighbouring regions
"Race to the bottom" strategic interactions	Negative spillovers from undercutting competitors and "beggar-thy-neighbour" policies Patent box as a harmful "race-to-the-bottom" style of international tax competition with negligible effects on innovation (Griffith et al., 2010)[vii] Risk of attrition of working standards for employees
Spillovers across domains	For example, productivity growth leading to a low-carbon transition Women's rights leading to improved education outcomes, lower inequality, and greater environmental protection Workers' rights leading to societal well-being and less inequality
Environmental spillovers and externalities	The environment may be negatively affected by economic activity in ways that are not captured in traditional indicators such as GDP

[vii] Griffith R., Miller H., O'Connell M (2010). Corporate Taxes and Intellectual Property: Simulating the Effect of Patent Boxes. IFS Briefing Note 112, Institute for Fiscal Studies.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science