

TIPC RESEARCH REPORT
APRIL 2019

TOWARDS A FRAMEWORK FOR TRANSFORMATIVE INNOVATION POLICY EVALUATION

Alejandra Boni (1), Sandro Giachi (2)
and Jordi Molas-Gallart (1)

INGENIO (CSIC-Universitat Politècnica de València)
SPRU, University of Sussex

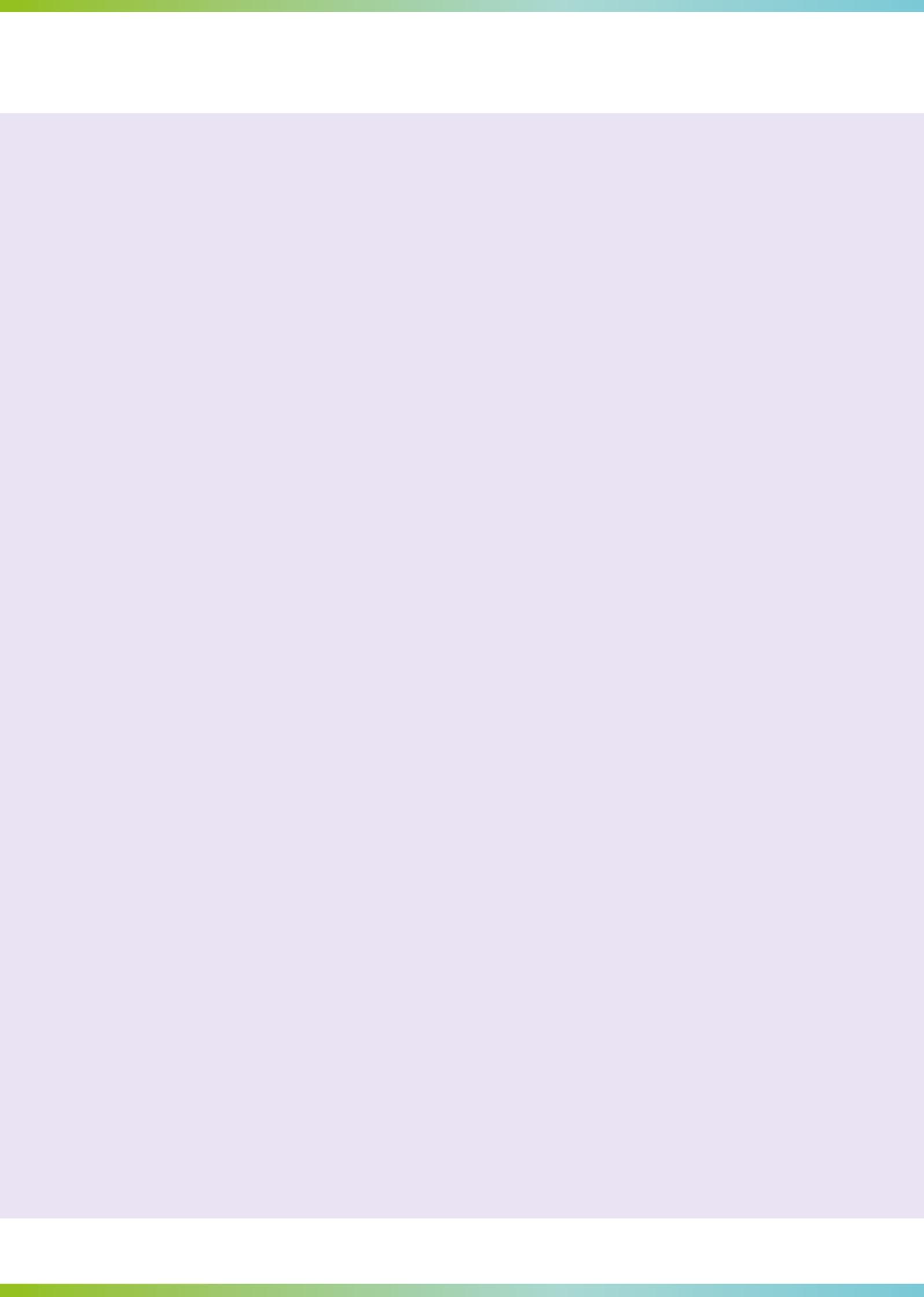


BUSINESS
SCHOOL



TRANSFORMATIVE
INNOVATION
POLICY
CONSORTIUM





CONTENTS

1. Introduction	2	5. A compendium of relevant evaluation techniques	29
1.1 Context of this report	2	5.1 Multi-criteria mapping (MCM): defining (and changing) the problem and strategies of a policy issue	29
1.2 An underpinning for the evaluation of transformative innovation	2	5.2 Portfolio analysis (PA): assessing gaps and potentialities of existing policy through big data sets	30
1.3 Objectives, contents and approach	3	5.3 Outcome mapping (OM): planning and monitoring improvements combining top-down and bottom-up strategies	31
2. Literature review	4	5.4 Participatory impact pathways analysis (PIPA): planning and monitoring impacts using theories of change	33
2.1 Evaluating sustainability transitions and climate change	4	5.5 Diversity approach to research evaluation (DARE): mapping diversity and modifications in the network structure	34
2.2 Insights from systemic approaches	10	5.6 Identifying relevant techniques: some concluding considerations	35
2.3 Delivering lessons from evaluation literature and experience	15	6. References	36
3. Key principles for TIP evaluation	17		
3.1 Principles of TI policy and their implications for evaluation practice	17		
3.2 The principles of 'formative' evaluation and its application to TI policy	20		
4. A formative approach to TI policy evaluation using a theory of change (ToC)	22		
4.1 A systemic approach in programme theory	22		
4.2 A generic ToC for TI policy	24		
4.3 A three-step sequence to build the ToC	25		

1. INTRODUCTION

1.1 CONTEXT OF THIS REPORT

The Transformative Innovation Policy Consortium (TIPC) brings together global actors to explore the future of innovation policy – its foundation, formulation and governance. The Consortium’s key objective is to examine and expand on current policy approaches to assist in solving urgent social, environmental and economic issues. This is a challenge for both the Global North and the Global South. TIPC aims to shape and deliver a new transformative innovation policy framework alongside other associated ‘policy mixes’. During the process, all participants are positioned as active co-researchers and policy co-designers.

In both TIPC’s pilot year (2017) and TIPC’s 1st year (2018), evaluation was one of the key areas highlighted along with research, experimentation and implementation. A group of researchers from SPRU and INGENIO have developed this guiding framework for the evaluation of Transformative Innovation Policies (TIP). We expect that this framework can guide the evaluation of the specific policy experiments to be developed by TIPC.

1.2 AN UNDERPINNING FOR THE EVALUATION OF TRANSFORMATIVE INNOVATION POLICIES

Transformative Innovation Policies (TIP) start off from a critical consideration of the shortcomings of current approaches to Science, Technology and Innovation (STI) policies. Schot and Steinmueller distinguish three “frames” of innovation policy (Schot and Steinmueller, 2018). Many current policies are still, implicitly or explicitly, informed by the “first frame”, which sees innovation mainly as a linear process emerging from the generation of new knowledge through basic and applied research, it is then further developed into new technologies through technological development, that is then applied through innovations that subsequently generate welfare and growth. The innovation systems literature has provided a second frame for innovation policy, stressing that the progression from new knowledge, to new technologies, innovation and growth is far from automatic and does not necessarily move in a single direction: new technological development can, for instance, spur basic and fundamental research, and the extent to which new knowledge and technologies

will lead to innovation and growth is contingent upon a variety of institutional factors and the linkages among different participants in an innovation system. The focus on systemic failures has provided a different rationale for innovation policies, moving beyond R&D investment levels to the institutional conditions and inter-organisational links that can promote innovation. Yet, the extent to which such processes address the challenges our societies face is left to decisions about the sectors and areas where investments are made. The alternative “third frame” postulates that to address key societal problems, like those laid out in the UN Sustainable Development Goals (SDGs), profound changes in current socio-technical systems are required, which in turn then calls for a different, broader mix of STI policies.

Socio-technical transitions can be triggered by policy experiments in protected niches. These experiments have explicit directionality; that is, they are expected to stimulate socio-technical regime change in specific directions aimed at addressing specific societal challenges. To achieve their transformational goals, policy interventions pursue changes in the structure and culture of governance emphasising inclusive participatory processes. These changes have to be extended to policy evaluation methods and practices, and imply crucial modifications in the roles of the actors involved in evaluation activities, as well as changes in the organisational routines within which these evaluation practices are inserted.

Although several evaluation frameworks and approaches have already been developed for sustainable innovation policies (Popper et al. 2017), sustainable transition policies (Taanman, 2014, Luederitz et al. 2017), and responsible research to leverage sustainable transformations (Daedlow et al., 2016) among others, our approach is directed at very specific types of research and innovation policies: those that follow the TIP approach as laid out by Schot and Steinmueller (2018). The framework we develop aims to be coherent with the key assumptions of this approach, while being operational and adaptable to the different contexts in which such policies are being implemented. The framework we propose aims to be applicable to the specific policy experiments that are developed by the Transformative Innovation Policy Consortium (TIPC).

We postulate that, to be consistent with the assumptions of TIPs, evaluation should:

- **Adopt a formative approach;** that is, it should aim at improving policy definition and implementation through the involvement in the evaluation process of all relevant policy stakeholders. Formative evaluation requires the development of specific capabilities in the organisations that implements it, and it calls for a close collaboration between Evaluation and other TIPC activities: capacity building, experimentation and research.
- **Integrate evaluation within policy design and implementation processes.**
- Be able to address the different levels of policy intervention, from specific projects at niche and local level, through programmes incorporating sets of projects, to complex policy interventions involving different programmes. Evaluation should be conducted at all levels in a coherent and consistent way. To this end, we propose a **nested approach to evaluation** defining the evaluation goals and approaches, and how they feed and contribute to each other.
- Facilitate **participation** and open debate, channelling power conflicts, and differences in interest and perceptions. At the same time, evaluation should be open and reflexive enough to capture emerging trends and problems, and incorporate new goals and actors.
- Be flexible from a methodological perspective using a mix of **methods and techniques** as needed. Rather than observing strict research protocols and using specific techniques, evaluation practices need to be adaptable and flexible according to the context, and the participatory and transformative nature of the policy.
- Adopt a **Theory of Change** informed by socio-technical transitions theoretical approaches to help distinguish the key dimensions of an intervention: the specific *context* in which the experiment operates, the specific *problems* it needs to tackle to achieve its objectives, the *mechanisms* (processes) the experiment aims to deploy with the *resources* available, the expected outcomes it aims to achieve and the changes they involve, and how these outcomes are important for the unfolding of a socio-technical transformation.

Drawing from the above features, we propose the following understanding of how evaluation can be framed for TIPC:



The diverse ways in which policymakers engage with processes of societal experimentation for sustainable transformation: initiating, supporting or mobilising, and evaluating such initiatives for informing decision-making, enabling processes of social learning, developing alternative pathways and enacting desirable futures.

1.3 OBJECTIVES, CONTENTS AND APPROACH

The main aim of this report is to develop a framework for the evaluation of TIP. The approach aims to be applicable to the specific policy experiments that will be developed by the TIPC. The report includes:

- A literature review describing how evaluation has been approached in two key areas related to transformative innovation (TI): sustainability transitions and climate policy, and systemic approaches for the evaluation of complex policy areas
- A set of principles to guide TIP evaluation
- A description of a generic theory of change on which to build the specific ToC to be used in the evaluation of TIP experiments
- A set of methods and techniques than can be used in different stages of the evaluation process

This framework has been developed through the involvement of the three co-authors of this report in TIPC activities, including: (1) in the bilateral contract with Colciencias (the Colombian Agency for Innovation) in 2017-2018 in which the authors trained and mentored local actors, interacted with policy makers and prepared a contribution for the “Green Book 2030: National Science and Innovation Policy for Sustainable Development” (Colciencias, 2018); and (2) their participation in the TIPC evaluation team in 2018, including a half-day workshop with all members of TIPC and a set of interviews and workshops with TIPC members in Colombia, Mexico and Sweden.

2. LITERATURE REVIEW

The objective of this section is to present the results of a literature review on two areas close to TIP evaluation: sustainability transitions and climate policy, and systemic approaches for the evaluation of complex policy areas. TIP evaluation is a new area whose boundaries are – partially – formed by other research fields, which are also emerging (e.g. evaluation of policy mixes, evaluation of sustainability transitions, etc.). TIP evaluation involves assessing the transformations associated to socio-technical systems, in itself a very challenging task. In the words of Kivimaa et al. (2017: 118): “Extending policy evaluations to transitions is complicated by the difficulties to set boundaries to what is the evaluand, i.e. the focus of evaluation, and how to evaluate ‘transition’ when it is still under way”.

Given this situation, we have selected for review, policy evaluation studies and approaches in areas facing complex evaluation challenges that are very close to TI policy aims, such as Sustainability Transition and systemic approaches to the evaluation of complex policy areas.

2.1 EVALUATING SUSTAINABILITY TRANSITIONS AND CLIMATE POLICY

The multidisciplinary set of studies on sustainability transitions and climate policy has not dedicated much attention to evaluative issues. In our review, we identified a small set of studies in this area, and only a few of them showed some link with the development of a formative approach for evaluating TI policy. It is important to mention that we are including in our review of sustainability transitions literature, environmental (e.g. climate) policy in order to find commonalities and derive useful implications for TI policy, instead of setting the state-of-the-art for each policy area separately.

Our review focused on the following studies:

- Hildén et al. (2014), who developed an evaluation perspective for climate policy innovation. They found a lack of consolidated approaches and methods in this field, and concluded that the policy tools currently used by the European Union to fight against climate change lack innovativeness and focus on a small set of policy instruments.
- Taanman (2014), who built an evaluative framework to monitor sustainable transitions policies and processes. Through a thorough literature review and an extended application to real cases and pilots, he identified two evaluation strategies: transformative and visionary.
- Luederitz et al. (2017), who developed a generic evaluative scheme for sustainability transition experiments. To check their previous assumptions and develop their proposal, they performed an extensive review of 61 case studies from peer review, Scopus and Google Scholar databases.
- Heiskanen and Matschoss (2018), who focused on the problem of evaluating local experimentation for sustainability transitions, specifically on climate governance and low-carbon experiments. They drew on two perspectives on learning from experiments: strategic niche management and participants’ programme evaluation.
- Holzer et al. (2018), who performed a literature review of the evaluation of socio-ecological research and its mechanisms. They identified five methods currently used in transdisciplinary research, including both qualitative and quantitative methods, and derived a six-stage framework for the assessment of socio-ecological research.

Roles and types of evaluation

In first instance, there is a general recognition of the need for more evaluation knowledge and capacities in this field, enabling us to identify “how judgments of success are shaped, and the extent to which evaluation plays a role by constructing knowledge in certain ways” (Hildén et al, 2014: 898). For instance, Holzer et al. (2018) mention that “[...] it has been explicitly noted that regarding the transformation in the production of research toward being more collaborative, more interdisciplinary, more transdisciplinary, and more oriented toward societal needs, assessment “does not happen without difficulties because there is no broad consensus about how to evaluate research in a more comprehensive way” (Spaapen 2015, pg. 36).” (pg. 817).

In this sense, Hildén et al. (2014) focus on innovation in policy practices as the essential way for addressing the challenges related to climate change, but they do not explicitly address the problem of reflexivity or the need for

continuous learning throughout the whole implementation of the policy. Yet, they do highlight the need for an integral approach to climate policy evaluation, identifying that “in many cases [...] because ex ante and ex post policy evaluation systems (and any coordination between them) are lacking in the environmental field (Mickwitz and Birnbaum 2009), policymakers seem to know relatively little about what the instruments they might introduce will achieve or, once adopted, what they are achieving or have achieved (Mickwitz 2012)” (pg. 885). They also note that the politics of innovation in the evaluation and monitoring process of climate policy has often been left aside, mentioning that: “We have confirmed that current evaluation practices are centrally concerned with the relatively narrow issue of compliance with aggregate targets. Some bigger – and often much more political – topics and questions have been pushed aside” (pg. 900) and that “the development of monitoring and evaluation of the policy measures that are expected to contribute to required emission reductions has progressed slowly compared with the more technical development of the monitoring of GHG emissions” (pg. 887).

By contrast, Taanman (2014) focuses specifically on monitoring, although he recognises the need for an evaluation following the whole intervention process and considering the transformation of the current regimes including their political dimensions, and it also delivers some ideas for improving ex-post evaluation of transitions programmes. Starting from the assumption that programmes are not needed in situations where the appropriate goals and means are known, the main task would be creating appropriate goals and means, as we will show later. In Taanman’s (2014) approach, learning is not something disconnected or abstracted from action, but situated in action. Reflexive processes are needed in which professionals practically learn about situations and become able to question what they had previously understood. At the same time, monitoring requires co-production by researchers and policy makers (programme managers, project leaders) because of the uncertainties and stakes involved.

Luederitz et al. (2017) consider different roles and types of evaluation: ex-ante, formative, summative ex-post, etc. In every case, they argue, evaluation should contribute to personal and organisational learning, promoting ongoing

change and up-scaling impact. This points to a prominent role for formative and process-oriented evaluation. The inclusion of ex-ante evaluation means that assessments to improve the design and operation of sustainability experiments can be made. Their proposal focuses only on experiments (short or medium term), even if they aim to a larger goal (transition) which requires cumulative evaluation. This creates some problems, because not every type of evaluation is able to capture time delays, and so ex-post evaluation should be planned from the start of an experiment. Evaluation requires financial and human resources and, ideally, is planned when designing the experiment proposal. To complete their proposal, Luederitz et al. (2017: 71) annexed a long list of evaluation criteria for appraising sustainability transition experiment outputs and outcomes, drawn from the literature review. This list of criteria is very diverse, mixing accountability and formative dimensions, scalability issues or scientific reliability.

Heiskanen and Matschoss (2018) explicitly address learning issues and attribute a formative role to evaluation. They use two approaches to promote learning: Strategic Niche Management, and realistic and programme evaluation perspectives, showing that ‘solutions’ and their ‘context’ are deeply interrelated. This implies that evaluations must address not just solutions but the overall governance framework and, then, follow a theory-of-change-based evaluation approach “in order to facilitate the embedding of the lessons learned from experiments beyond their local origins” (pg. 184). However, they also identified some serious challenges for evaluation, like for instance, the lack of appropriate evaluation methods to address local sustainability experiments; the difficulties of achieving a truly participatory evaluation; and the difficulties in increasing the scope of evaluations from what, where and when questions to evaluation that is “distributed learning, inspiration and (the) development of commitment and confidence” (pg. 197).

Theory-driven evaluation vs ‘open’ approaches

The reviewed studies present contrasting views about the ‘openness’ and the use of a theory-driven approach to evaluation. For instance, Holzer et al. (2018) or Hildén et al. (2014) do not follow a specific theory-based approach. Holzer et al. favour a transdisciplinary

approach as a way to foster sustainability through socio-ecological research, but do not consider the development or use of a programme theory. Hildén et al. simply focus on the effects of a policy on emission reduction as a guiding criterion for their evaluation approach. By contrast, Heiskanen and Matschoss (2018), follow a “realistic evaluation” approach to integrate Strategic Niche Management theories into the evaluation design, thus recognising the relevance of building theories of change as a starting point, to reflect the different interpretations that the experiment participants have about innovation and change mechanisms. This is a theory-driven approach but rooted in diversity and requires a participatory approach to build the underlying theory of change for the experiment. These scholars point out that “[...] evaluation procedures need to not only incorporate but also go beyond realistic evaluation designs to capture emergent aims and the socially constructed outcomes of local climate experiments” (pg. 197).

The position of Luederitz et al. (2017) is also closer to theory-based evaluation, although in a different way. These scholars recognise that there are different kinds of experiments: transitions policy programmes, transition management projects, technical innovation projects, community initiatives, social innovation processes, and including a new type of transition experiment characterised by cross-organisational collaboration between actors from academia and society. Despite this diversity, all of them have in common that they want to be transformational interventions towards greater sustainability. Although the experiments often focus on defined small-scale settings, specific to a place and socio-cultural context, the intention is to create positive outcomes that are replicable, transferable and scalable to society at large. Their preliminary assumption is that such experiments can produce evidence regarding both the persistent unsustainability of dominant regimes and the possible solutions to given sustainability problems. This idea is deeply rooted in sustainability transitions theory.

Taanman (2014) takes a different cue from transitions theory, by assuming that transitions dynamics are difficult to predict and, therefore, monitor. The same applies to the effects of interventions to influence them (experiments). However, both transition dynamics and

their effects can be influenced by experiments. Good monitoring is a way to influence experiments and, so, transitions and their effects, but this would require a better scientific understanding of the transition processes and, specifically, looking at more specific patterns and regularities, referred to as ‘middle range theories’ (Geels, 2007). Rather than monitoring everything, transition monitoring should be an instrument that takes advantage of ‘middle-range transition theories of change’ to make complexity manageable and understandable for specific programmes. A good example of this is posed by Van den Bosch (2010) who describes how transitions experiments contribute to a transition through mechanisms of deepening, broadening and scaling up:

1. Deepening is a learning process through which actors learn as much as possible about a transition experiment within a specific context.
2. Broadening is repeating a transition experiment in different contexts and linking it to other innovations.
3. Scaling up is embedding a transition experiment in dominant ways of thinking (culture), doing (practices) and organising (structure), at the level of a societal system.

Applying these elements to a monitoring framework generates the following set of expectations: in a process of deepening, transition scenarios are reflected upon and possibly modified; broadening can concern aligning and marketing the scenario within the same organisation, peer groups or similar projects; and scaling up the transition scenario occurs by anchoring the scenario within the sector at large.

Following Taanman (2014), we recognise that transition theories are useful for evaluation if they are already used in the programme design, because theories of change have normative implications. For example, using transition management as a basis for monitoring implies that long-term transition visions are important elements of a programme. There are also some instrumental implications: for instance, monitoring requires going beyond general discussions; considerable shared understanding in the monitoring team is needed to define operational indicators that correctly reflect relevant variables and result in relevant recommendations.

Another important distinction pointed by Taanman (2014) refers to transformative and visionary strategies. A visionary strategy starts from a vision of what a valuable future would be and then tries to shape the environment in order to realise it. The vision is actively used to influence other actors in the transition field. The programme develops new standards and different alternatives are tested in practice to learn quickly about ways to realise the vision. Therefore, it incorporates from the start a theory of change. By contrast, a transformative programme starts with experiments to explore persistent barriers and, potentially, promising long-term changes. Based on these initial projects, the programme builds the identity, knowledge and networks to take a next step, which requires commitment of new stakeholders. Through every step more knowledge is gained, networks are created and tightened, and a stronger sense of identity is constructed, while the theory of change must be built through the process.

Dimensions and levels of evaluation

Holzer et al. (2018) employ a traditional differentiation between inputs, processes, outputs and outcomes in evaluation, mixing different evaluation methods, through a flexible approach for soliciting “insightful comments” that can be applied either at the programme or project level. Luederitz et al. (2017) also follow a ‘traditional’ approach to evaluative dimensions, conceptualising inputs, processes, outputs and outcomes as basic category of the evaluation scheme and providing a comprehensive collection of critical features for these categories. Departing from an evaluative scheme developed in a study on urban sustainability experiments (Wiek et al, 2015) they consider the inputs that are invested in the experiment, the processes that are performed, the outputs (outcomes?) that are generated and the sustainability outcomes (impacts?) that are accomplished by the experiment (Figure 1).

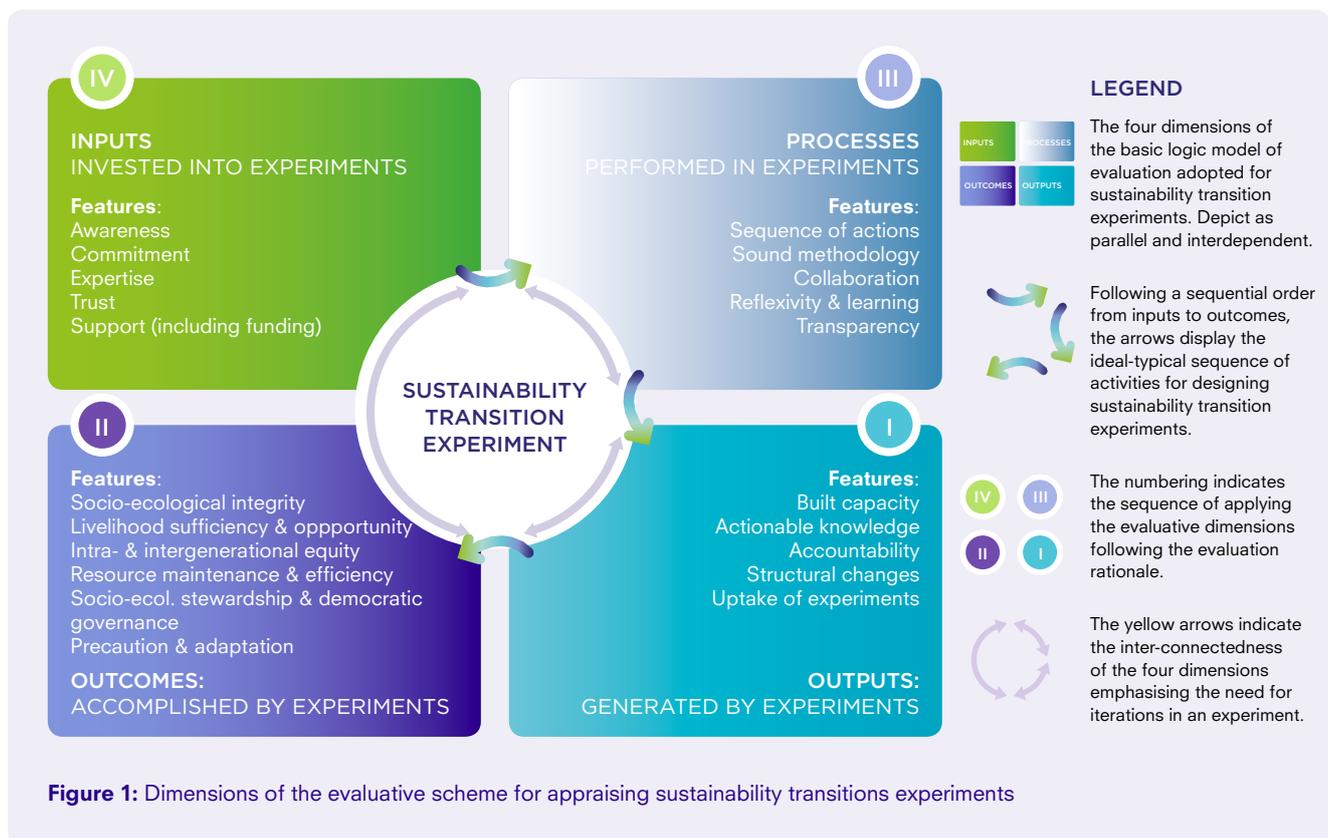


Figure 1: Dimensions of the evaluative scheme for appraising sustainability transitions experiments

There is an ideal sequence order for designing experiments (following the thick arrows) but also an ideal sequence of applying evaluative dimensions (numbers) that changes the sequence of the basic logic model of evaluation (Inputs-Processes-Outputs-Outcomes). The sequence of actions in experimentation needs to include:

1. Defining a baseline and a goal for the interventions
2. Creating a specific set-up to administer interventions
3. Measuring the effects of the interventions against the baseline and the goal
4. Evaluating the effects against sustainability criteria
5. Offering evidence-supported recommendations on how to implement the results

About the inputs, the following criteria are key:

- Awareness: refers to the ability and awareness of participants to acknowledge the need for radical real-world changes prior to and during their engagement in the experiment
- Trust: refers to the mutual willingness to collaborate on equal footing, reconcile divergent worldviews, as well as acknowledge different interests

However, these components are parallel and interdependent, i.e. an output could be related with an input and a process, but also, a new output can generate a new process. Applications to multiple experiments will allow to identify causal mechanism, influencing factors, weights, etc. Luederitz et al. use their list of criteria, drawn from the literature review, for appraising outputs and outcomes. For instance, regarding outputs, they seek evidence of building capacities and roles, of generating of different types of actionable knowledge (analytical, normative, transformational), and of “accountability” issues like participants’ commitment. Among the outputs identified they included the generation of intra and inter-generational equity.

In short, Luederitz et al. (2017) proposed a generic scheme, designed to be applicable to a broad range of sustainability transition experiments types. At the same time, the application of such a scheme to a specific experiment may require the modification of the scheme: criteria could be merged, subdivided or revised. The scheme aims to be comprehensive, considering different dimensions (outputs, outcomes, inputs, processes) and a collection of critical features from a broad range of experiments types, i.e. definitions, typical indicators, illustrative examples, and evaluative questions (Figure 1)

The approach followed by Taanman (2014) is slightly different. He identified three dimensions of transitions monitoring:

- ‘Transition scenarios’: long term ambitions of actors. A transition scenario is a description of the current situation (as is), a desired future situation (to be) and a ‘script’, which explains how the current situation can be changed into the desired future situation.
- Concrete and current changes in culture, structure and practices. ‘Structure’ includes physical structure (artefacts, land use, built environment, tangible resources) and social structure (institutions, organisations and networks, regulations, norms, responsibilities, budgets and contracts). Culture refers to paradigms, discourse, values and the knowledge base. Dominant culture and structure are constantly recreated through routine practices.
- Sustainability criteria: should be chosen by programme stakeholders. Transition monitoring therefore has two functions: not only measuring the ‘sustainability performance’ of the transition field, programmes and projects, but also stimulating the discussion about the sustainability norms at stake.

These three dimensions of transitions monitoring must be consistent between three evaluation levels: transition field, programmes and projects, as depicted in Figure 2.

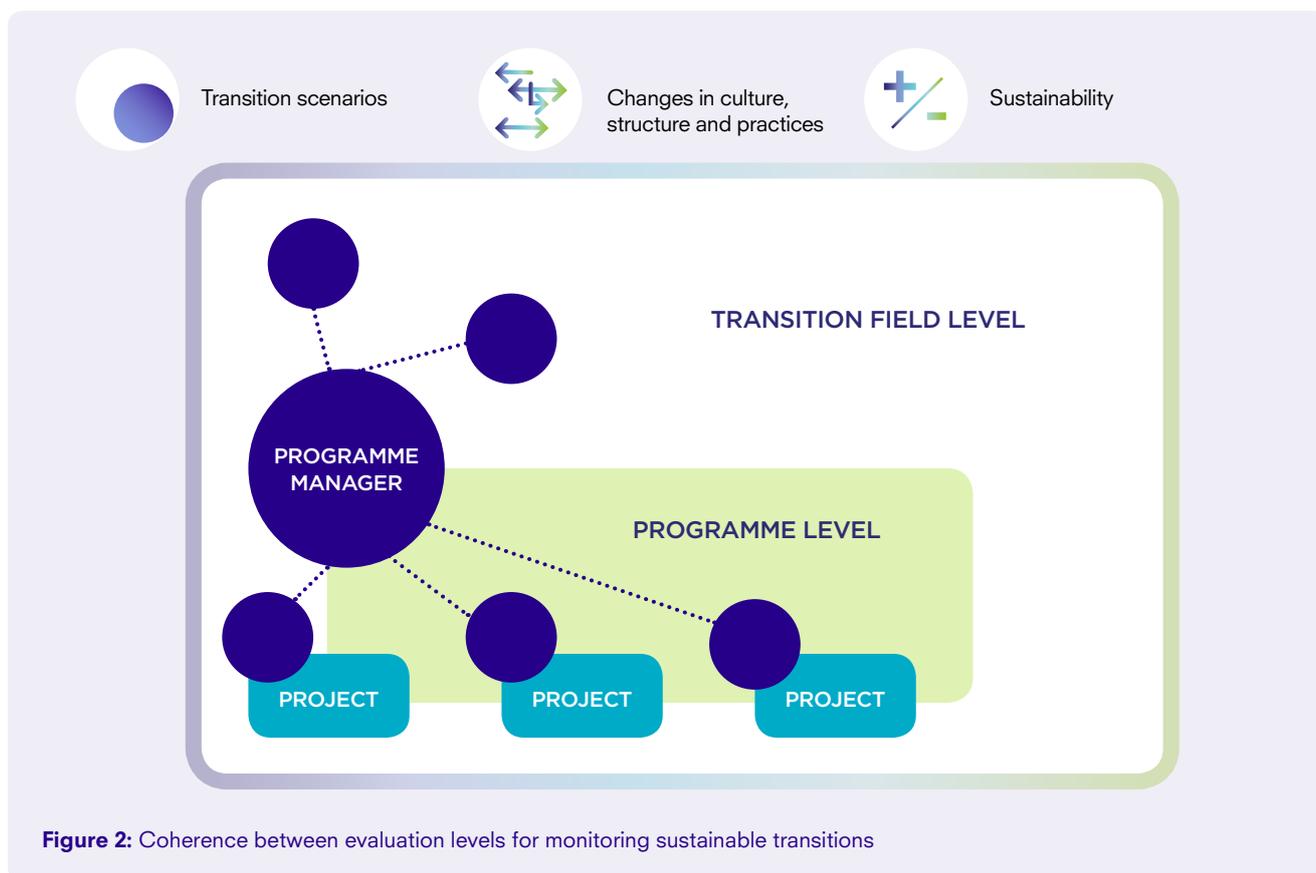


Figure 2: Coherence between evaluation levels for monitoring sustainable transitions

As previously mentioned, Taanman (2014) argues that defining appropriate goals and means can act as a trigger for change. While goals establish what type of transition is desirable and how it can be pursued (outcomes), means including several elements:

1. innovations that have increasingly better sustainability effects (outputs);
2. naming key structural and cultural barriers as well as effective ways to deal with these (outputs);
3. defining a real transition scenario for the long term (output);
4. creating an advocacy network of well-connected actors that develop this transition (output);
5. considering the growing amount of time, financial, organisational and other resources to do all of that (inputs).

In short, the inputs-resources (time, financial, organisational, etc.) feed into different forms of means-outputs (Innovations to sustainability, identification of structural and cultural barriers, transitions scenario, advocacy network of actors to develop this transition, etc.). In turn, these feed into the goals, related to the desirable type of transition.

Inclusiveness, participation and conflict

All the reviewed studies recognised the need for stakeholder and societal participation, although for different purposes. While Holzer et al. (2018) generically proposed to involve the research network (including the stakeholders) into the evaluation process of socio-ecological research programmes and projects, Hildén et al. (2014) highlighted the political dimension of policy innovation, and underlined the role of political debate in

evaluating climate policy: evaluation thus emerges as a political activity. When highlighting the political character of evaluation practice, it is important to consider the existence of institutional differences among countries, the differences between formal and informal evaluations and the multilevel nature of the governance system. Hildén et al. also mentioned the need for independent evaluations, to prevent the spread of ineffectual policies.

Luederitz et al. (2017) claimed that experiments go beyond inclusion and participation of a diverse array of social groups into creating opportunities in actively empowering them to be part of on-going and future sustainability transitions. This leads them to include intra and inter-generational equity among their criteria for assessing outcomes, understanding equity to emerge when sufficient and effective choices are generated to reduce disparity between rich and poor and to enhance future generations' opportunities to pursue sustainable lives. Despite recognising the existence of problems related to power and vested interests, they do not articulate a detailed proposal for dealing with the political dimension of evaluation. While they argue that the core members of an experiment should carry out the evaluation or at least support it externally (an inclusive model), they do not address how evaluators can give beneficiaries among least privileged groups a proper say in the evaluation process.

A similar view can be found in Heiskanen and Matschoss (2018), who emphasise the central role of the participant's perspective in the evaluation of local interventions and the limitations of current practices and knowledge, using Theories of Change, as a means to articulate different views and interests. They argue that the inclusion of different perspectives should not be restricted to the initial definition of the rationale of an experiment but should extend to the processes by which such rationale is redefined. They point out (197) that experiments transform the reality of the stakeholders beyond the boundaries of the experiments, affecting commitment, understandings, and social identities. The evaluation of success or failure is deeply interrelated with the participants' context, because experiment's processes are strongly influenced by local factors and issues of power relationships between the participants (Heiskanen and Matschoss, 2018: 194).

This consideration has led to the recognition of the facilitator role that evaluators can play: evaluators can help articulate different perspectives, anticipate potential conflicts (Blamey and McKenzie, 2007), and identify emergent trends and common goals. This role is especially relevant when interventions take place within network governance modes, based on bottom-up forms of horizontal coordination, crossing of administrative boundaries (e.g. public-private), and informal control (Hertting and Vedung, 2012), as it is often the case in the sustainability transitions field.

The articulation of different perspectives in evaluation raises an additional difficulty: participants' views on the experiment may influence its very performance. There are potential 'unintended consequences', 'placebo effects' of 'magic programmes' effects (Dahler-Larsen, 2001) that should be considered and that may be generated by the way in which participants' views are articulated within the experiment and the evaluation.

2.2 INSIGHTS FROM SYSTEMIC APPROACHES

The systemic approach to the evaluation of innovation policy gained increasing attention during the last two decades: it came along with the recognition of the systemic nature of innovation policy and represented a shift in interest from the project or programme levels as an evaluation subject to the level of the whole system that generic policy objectives are trying to affect. The interest of the evaluators thus changes from the analysis of a specific intervention (a project or a programme) to how different programmes interact; i.e. policy mixes or policy portfolios (Arnold, 2004; Edler et al., 2008). These perspectives can be aligned with what Schot and Steinmueller (2018) called 'Frame 2' policies and a focus on innovation systems.

Systemic approaches to evaluation are relevant for the evaluation of TIPs because:

- TI policy often implies complexity and a complex system of relationships across policy levels, governance levels, and actors. Systemic approaches aim to deal with complexity and multi-level governance.

- Systemic approaches in general – and in particular the consideration of policy mixes as a research subject – are a relevant way to address climate and sustainability transition policies, and could be applied in evaluation projects.

We identified the following examples of systemic approaches to evaluation:

- Magro and Wilson (2013) address the complexity of innovation policy mixes, due to their systemic nature and the lack of adequate evaluation methods. Drawing on the experience of a regional innovation policy system, they present a six-step evaluation protocol.
- Caffrey and Munro (2017) describe the use of a systemic approach to policy analysis, using systems methodology for evaluating complex social policy.
- Kivimaa et al. (2017) use a client-oriented evaluation approach to examine energy policies from the perspective of low-carbon buildings transition, connecting the literatures of policy evaluation, policy mixes and sustainability transitions.
- Janssen (2019) aims to assess the ‘value for money’ of policy mixes for sustainability transformations. He studies a policy mix within a national research and innovation pilot strategy. The study is based on 107 face-to-face interviews, which are used to assess the economic feasibility of the policy, its impact, and the challenges to measuring socio-technical transformations.

Evaluation approaches and roles

One of the highlights from Caffrey and Munro (2017) is the focus on analysing policies taking into consideration their interactions with other policies. Their systemic approach focuses on processes and not on structures, and it is useful for contributing to new evaluative knowledge and understanding complex policy problems. Their systems framework, focusing on empirically investigating how policy is interpreted as it interacts with other systems, provides a means to evaluate critically the policy in its context and to pinpoint any unexpected (and unwanted) effects. Another implication is that systems thinking acknowledges that in complex systems, unpredictability is inherent and so policy makers cannot assume to have the ‘right’ answer for every context.

Caffrey and Munro (2017) also mention that the systems approach can be a form of theory-based evaluation and it is useful to understand the dynamic properties of the context into which a policy is introduced and the influence of it on policy outcomes. They also emphasise the relevance of their systems approach as a learning system (Caffrey and Munro, 2017: 465-466).

From a different viewpoint, Kivimaa et al. (2017) also follow a theory-driven approach to evaluation under the perspective of stakeholder theory. They focus on a sole target group of energy intermediaries (i.e. energy service companies) as a means to reduce the complexity of policy mixes for sustainability transitions. They reconstruct the intervention theories of the policies according to the stakeholders’ views, focusing on the causal ‘theory of the intervention’, which they define as “Notions concerning how a given intervention/ programme directly or indirectly through its process of implementation and delivery of outputs will (a) have an impact on the causes of the underlying problem at issue so that the problem will disappear, be reduced, or prevented from becoming aggravated and (b) have possible effects in other areas during the entire process” (Kivimaa et al., 2017: 118).

Regarding methods, Kivimaa et al. (2017: 118) stressed the role of real-time qualitative evaluations involving stakeholders, complemented by quantitative and summative evaluation analyses. The latter “can point out unexpected synergies or conflicts, or problems that have occurred in policy processes, i.e. the preparation or implementation phases. Quantitative methods struggle to address policy mixes involving non-economic instruments, which are crucial for sustainability transitions. For instance, they found limited the usability of client-oriented evaluation on its own for the analysis of the overall policy mix but useful in complementing top-down policy evaluations, in order to “identify issues not necessarily evident in top-down policy mix evaluations, including what actors regard as the most influential policy goals and instruments and how actors experience in practice the influence of a mix of policy goals, instruments and processes crossing administrative boundaries and levels” (Kivimaa et al., 2017: 124).

Despite some similarities in the underlying view on policy mixes and methodology, Magro and Wilson (2013) do not follow a theory-driven approach in their proposal. Their framework “enables individual evaluations to be integrated in an evaluation mix appropriate for their systemic context, and that facilitates policy learning at each step” (Magro and Wilson, 2013:1655). Their proposal is quite comprehensive and structured. They propose to use an ‘evaluation mix’ and to triangulate evaluation techniques for approaching the complexity of policy mixes for innovation policy systems. They further argue that there is a need for a “more precise articulation of the practical steps” and to develop new forms of evaluation processes to capture the impacts of complex policy systems. By emphasising the complexity of evaluating innovation policies, they conclude that “the typically-adopted approach of employing specific

techniques to evaluate isolated policy interventions has strong limitations in systemic contexts, and its widespread adoption may, in fact, provide an obstacle to more sophisticated understanding of innovation policy mixes and their evolution” (Magro and Wilson, 2013: 1647). They call attention to the relevance of systemic evaluations to understand the impact of policy interactions and highlight the need for new evaluation approaches that can capture the “complex interactions that take place within policy systems” (Magro and Wilson, 2013: 1649). The reason is that quantitative evaluation tools are more adequate for linear rationales than for complex interactions such as those entailed by policy mixes. Table 1 summarises the different steps of this proposal. Applying this protocol, they show the wide overlapping between policy instruments.

STEP 1	Draw the policy system and establish its boundaries in terms of rationales, domains and instruments (policy mix dimension) and administrative scales (multi-level dimension)
STEP 2	Select a policy rationale
STEP 3	Analysis of the mix of domains and instruments at different administrative levels that fall under the selected rationale, looking for overlaps and complementarities
STEP 4	Identify current evaluation practices and the extent to which they take into account interactions between policy instruments
STEP 5	Design and conduct an integrated evaluation (including policy interactions) of policy instruments following the same rationale
(REPEAT STEPS 3-5 FOR EACH RATIONALE) STEP 6	Intergrate rationales’ evaluation into a holistic evaluation: evaluation mix

Table 1: Evaluation Mix Protocol. Source: Magro and Wilson (2013)

The work performed by Jansen (2019) presents some similarities with the approaches just mentioned but also some important differences. He highlights that the methods traditionally used to evaluate policy are inadequate for transformative policy. His definition of transformative policy is closer to economic policy than TIPC's: "any kind of policy approach that strives for diversifying an economy's industrial structure and underlying capabilities" (Janssen, 2019: 79). He also identified three key properties of transformative policy:

- Selective, i.e. related to specific socio-technical systems or techno-economic pathways
- Process-oriented, i.e. formative, involving continuous policy adaptation
- Multi-instrumental, i.e. based on policy mixes and requiring methodological heterogeneity

He claims that transformative policy poses severe challenges to evaluation for accountability purposes. Evaluation methods for accounting the returns from R&D are ill-suited when policy mixes are supposed to guide directed transformations. Due to the difficulty of estimating the individual impact of interventions that are part of larger sets of interventions, the traditional indicators do not capture the relevant characteristics of a policy. This difficulty is compounded by the difficulty in identifying good counterfactuals to estimate the impact of an intervention. He further explores the perspective of policy mixes and the challenges to assessing changes fostered by those policies, developing an assessment scheme to identify the organisation, orientation, and aggregate impact of transformative policy mixes.

Janssen (2019) points out the challenges that multi-instrumentality poses to the evaluation of each individual intervention. This challenge is important because "the sum of narrowly targeted interventions is likely to exceed their individual impacts (Magro and Wilson, 2013)" (Janssen, 2019: 84). Identifying specific impacts in transformative policy is complicated because interventions tend to mutually reinforce each other's contributions. Following Rodrik (2014), he also recognises the difficulty of applying rigorous evaluation techniques because of the differences existing in the programme components across recipients (Janssen, 2019: 82).

Elements of evaluation: levels, interactions and actors

The application of systemic approaches to the field of evaluation has focused on the generic concepts of interrelations, perspectives and boundaries. For instance, Reynolds (2014: 81), suggests that in the context of evaluation, these concepts relate to "understanding interrelationships associated with a situation; engaging with contrasting perspectives regarding a situation; and reflecting on boundaries of such representations and interactions". By contrast, Caffrey and Munro (2017) highlight that systems approaches focus on "functional abstraction rather than structural decomposition". They address the complexity of policy evaluation by using an engineering systems approach, in order to "illustrate how the systems concepts of 'emergence', 'local rationality', 'socio-technical systems' and 'feedback for learning' can contribute new knowledge and understanding to a complex policy evaluation problem" (Caffrey and Munro, 2017: 464). Scholars like Pawson (2013) or May (2006) have criticised the systems approach for being too abstract and inappropriate to carry out evaluations at the micro level. Caffrey and Munro (2017: 473) reply that "different systems models are appropriate for different research questions", i.e. the system approach used to carry out an evaluation should align with the level of analysis and final objectives of the evaluation.

Kivimaa et al. (2017) perform a policy analysis from the perspective of a boundary actor, considering the goals, expectations, concerns and needs of the policy recipients as rationale and criteria for evaluation. Their approach aims to integrate traditional evaluation models for energy efficiency with pluralistic methods. For instance, their findings showed sharp divergences among stakeholders, attributed to the incoherence in policy implementation processes. They also find that changes in organisational and institutional practices should be taken into account to provide coherence to policy mixes: "[...] we propose that new organisations or the replacement of key actors needs to be supported by the reorientation of incumbent policy actors through new organisational or institutional practices to support more coherent policy implementation" (Kivimaa et al., 2017: 124).

A similar view about inclusiveness and structural change is present in the work by Janssen (2019). Following new industrial policy (NIP) and transition thinking, Janssen claims that “achieving structural change requires governments, industry and research to collaborate in formulating targeted yet adaptive strategies”. He introduces “a framework that structures investigation of the extent to which a multi-instrumental policy approach is effective, decisive and targeted at the system functions most relevant for advancing the TIS” (pg. 79). This discussion relies on how the evaluation methods correspond to single-goal instruments instead of multi-instrumental policy approaches.

The assessment scheme that he proposed connects policy design review with dimensions of structural change. In his empirical application to a research and innovation pilot national strategy, he found that the policy design specificities explain larger impact of the experimental policy, mostly through engaging private parties to strengthen existing knowledge networks, rather than transforming them. Janssen’s framework builds on reform analytics (Hausmann et al, 2005) and Technological Innovation System (TIS) governance, as depicted in Table 2. The rows present TIS functions and the columns critical variables. This framework is complemented with specific questions to guide the analysis.

	λ_j	{n1...N}	P_j	∞_j	β_{ji}
F FUNCTION	How important is strengthening function j for the success of TIS x?	What did government do to strengthen function j?	how effective was policy in strengthening function j?	how decisive is policy for strengthening function j?	Did solving function j strengthen other functions?
F1 GUIDING DIRECTION OF SEARCH	Which goals were not shared or recognised?	(E.g. publish joint vision/agenda, studies)	In which way have interests been aligned?	Why did interests use to be divergent?	In the review period, which were the most pervasive dependencies between the function? Did strengthening a function have a positive or negative effect on other functions?
F2 KNOWLEDGE DEVELOPMENT	What knowledge was missing/insufficient?	(E.g. targeting public science)	What research results? (quality and topics)	Why is interest for certain new technologies low?	
F3 KNOWLEDGE EXCHANGE	Which particular connections were weak?	(E.g. vouchers, joint R&D, TTO, platforms)	What new interactions have been facilitated?	What mechanisms were blocking interactions?	
F4 ENTREPRENEURIAL EXPERIMENTATION	Which notable opportunities are not yet explored?	(E.g. attracting foreign firms, training centres)	What new firm capabilities/which firms settled?	Why no markets for training, or for technology itself?	
F5 RESOURCE MOBILISATION	Which activities hampered by lacking resources?	(E.g. innovation funding, R&D facilities, HR)	How much increase in R&D funding/R&D staff?	Public resources matched by private expenses?	
F6 LEGITIMATION/COUNTERACTING RESISTANCE	Which changes did lobby/laws used to block?	(E.g. adapt legislation, lobby, do assessments)	Did policy change discourse? Less legal barriers?	What hindered acceptance and compliance?	
F7 MARKET FORMATION	Which suppliers/demand are crucial?	(E.g. public procurement/public goods)	Which signals given to market parties	Why were up- or downstream markets missing?	

Table 2: Impact Assessment Framework for Transformative Policy. Source: Janssen (2019: 84)

2.3 DERIVING LESSONS FROM EVALUATION LITERATURE AND EXPERIENCE

Despite the diversity in approaches, a series of recommendations can be drawn from our literature review.

1. First, there is a general agreement about the need for a **comprehensive type of policy evaluation** due to the complexity of the policy processes and the structures they address. There is an overall recognition that ex-ante and ex-post evaluations are necessarily incomplete and that the analysis of implementation is crucial for disentangling complexity and learning from evaluation. However, how a 'comprehensive' evaluation might be is not clear and there is diversity in the proposals emerging from the literature review. For instance, while Hildén et al. (2014) speak about a 'comprehensive and reflexive' approach coordinating ex-ante and ex-post evaluations, Taanman (2014) focusses on following up the whole policy process from the very beginning. The implementation of this monitoring approach is open to experimentation with different evaluation strategies and methods.
2. The literature shows some **differences about the role of evaluation**. Taanman (2014) and Heiskanen and Matschoss (2018) explicitly advocate for formative evaluations, while Luederitz et al. (2017) and Holzer et al. (2018) mix formative and accountability criteria in a kind of hybrid approach, despite both attributing greater importance to learning and improvement as evaluation goals. Similarly, the systemic approach often sees no contradictions between learning and accountability purposes for evaluation and the works by Magro and Wilson (2013), and Caffrey and Munro (2017) seem to follow this trend, although, again, they explicitly recognise a prominent role for formative evaluations and policy learning. Only Janssen (2019) seems to attribute greater relevance to accountability issues (consistently with his research goals) but he also considers the importance of reflexivity and learning processes for transformative policy.
3. In short, although different authors are guided by different evaluation functions and roles, **learning** – in its different forms: reflexivity, policy learning, feedbacks, etc. – emerges as a prominent goal in the studies analysed.
4. The literature review shows a **wide variety in evaluation approaches and methods**. While some studies followed a theory-driven approach, others followed more 'open' strategies. The realistic approach followed by Taanman (2014), Caffrey and Munro (2017) and Heiskanen and Matschoss (2018) led to the use of Theories of Change in their evaluations. In turn, Magro and Wilson (2013), Luederitz et al. (2017) and Janssen (2019) proposed formal protocols for evaluations to be applied in sequential stages but without a Theory of Change. Finally, Hildén et al. (2014), Kivimaa et al. (2017) and Holzer et al. (2018) did not follow a theory-driven approach and did not deliver a specific strategy: they limited themselves to highlight some specific elements or criteria, like transdisciplinarity in socio-ecological research, or employing stakeholders' perspectives for the analysis. In short, it is difficult to derive some specific lessons about evaluation approaches and methods from such a variety of perspectives linked to similarly varied evaluation objectives and functions. In general, the use of ToCs is related to a realistic approach and, often, a formative role for evaluation, while systemic approaches mixing reflexivity, policy learning and accountability issues seem to prefer structured protocols. A common element in the literature seems to be the necessity of **mixing different evaluation methods, in order to deal with the complexity and process-based nature of innovation**. Therefore, an approach for TI policy evaluation should embrace this diversity, consider different methods, especially those that are participatory in nature, and experiment and perform more research and informal evaluation to improve and combine the current instruments.
5. Most of the reviewed literature recognised the **multidimensional and multilevel nature of innovation policy** systems and opted for 'traditional' descriptors of the policy process like inputs, outcomes and impacts. In this sense, the literature is consistent and not particularly innovative. The studies from

the systemic literature explicitly recognised the existence of different levels and their importance for policy implementation and evaluation. Taanman (2014) identified three specific levels for sustainability transitions: projects, programmes and transition field. Something similar can be observed in Heiskanen and Matschoss (2018) and Holzer et al. (2018), although not explicitly formalised.

6. Besides, Taanman (2014) complemented traditional evaluation elements (i.e. inputs, outcomes etc.) with others specific to sustainability transitions: transition scenarios, changes in organisation, practices and culture, and sustainability criteria. These dimensions could be used to integrate the traditional dimension of evaluation in order to describe how TI policy contributes to transitions. This is consistent with the findings of Kivimaa et al. (2017) about the relevance of changes in institutions and organisations, as well as some specific dimensions identified by Caffrey and Munro (2017), Hildén et al. (2014), Luederitz et al. (2017), Heiskanen and Matschoss (2018), and Janssen (2019), related to the transformation of socio-technical systems beyond the traditional indicators used in research and innovation policy.
7. Finally, from the literature reviewed we can derive some interesting implications about the **governance** of evaluation and the **participation** of stakeholders. Almost all the reviewed studies recognised the need for stakeholder and societal participation, although for different means and sometimes in different ways. For instance, Holzer et al. (2018) only make a generic mention about including the research network within the evaluation, Janssen (2019) highlights the opportunity to involve private firms to strengthen existing knowledge networks, while Kivimaa et al. (2017) present an evaluative research design where stakeholders (in this case, intermediary companies) have been used as the focal point for assessing different perspectives on policy implementation and impact.

Some authors have developed a theorisation of the relevance of **inclusion and participation** but without a specific proposal for governing the evaluation process. For instance, Caffrey and Munro (2017) tries to apply the systemic approach at the micro-level of social interactions and to consider the aggregate level of the agency of all the individuals involved into the policy process but they fall short of describing in detail how to implement the approach and address the problems it may encounter. Similarly, Luederitz et al. (2017) make a relevant point that inclusion does not only mean participation, but also creating opportunities, empowering and give a voice to the excluded into the policy process. However, they do not specify how this can be done, e.g. how the evaluators can avoid reproducing the existing structure of power and interests. In this regard, Heiskanen and Matschoss (2018) position is interesting: they argue that the performance of transition experiments can be influenced by participants' perspectives and beliefs, and that the success (or failure) of an experiment should be assessed considering the context, going beyond the initial purposes of the experiment. Evaluators should act as facilitators managing not only conflicts of interest and power, but also different cognitive and cultural perspectives, and foreseeing and addressing emerging changes and trends, as well as new perspectives going beyond the initial boundaries of the experiment.

3. KEY PRINCIPLES FOR TIP EVALUATION

3.1 PRINCIPLES OF TI POLICY AND THEIR IMPLICATIONS FOR EVALUATION PRACTICE

As we highlighted at the beginning of this report, following the model proposed by Schot and Steinmueller (2018) we recognise the existence of three innovation policy frames: 1) Linear modes, 2) Innovation Systems, and 3) Transformative Change. Frame 1 policy follows a linear model of innovation based on R&D, where innovation emerges from a process that starts with the generation of new knowledge through basic and applied research, the further development of such knowledge into new technologies, which are then applied through innovations that generate welfare and growth. Within Frame 1, the focus is on input, output and outcome indicators. Policy objectives can be narrowly defined and easily operationalise.

By contrast, Frame 2 policies recognise the systemic and complex nature of the innovation process, based on inter-organisational relationships and learning. Interactions among different types of actors and organisations are necessary to generate economic development and – ultimately– societal welfare. The innovation systems literature has provided a second frame for innovation policy, stressing that the progression from new knowledge, to new technologies, innovation and growth is far from automatic and does not necessarily move in a single direction: new technological development can, for instance, spur basic and fundamental research, and the extent to which new knowledge and technologies will lead to innovation and growth is contingent upon a variety of institutional factors and the linkages among different participants in an innovation system. The focus on systemic failures has provided a different rationale for innovation policies, moving beyond R&D investment levels to the institutional conditional and inter-organisational links that can promote innovation. Yet, the extent to which such processes will address the challenges our societies face is left to decisions about the sectors and areas where investments are made.

Therefore, Frame 2 introduces systemic complexity calling for more factors, variables and indicators to be included into the evaluation designs. However, the selection and definition of the proper indicators for Frame 2 innovation policy is not easy: here the focus is not on inputs, outputs

and outcomes, but on innovation processes. Unfolding the complexity of innovation processes for evaluative purposes requires different techniques, like those developed by the “Social Impact Assessment Methods for research and funding instruments” project through the study of ‘productive interactions’ between science and society (SIAMPI), or impact pathways techniques (e.g. ASIRPA). While in Frames 1 policy evaluation can be conducted summatively, evaluation approaches like SIAMPI and ASIRPAe focus on processes and learning and therefore, implicitly assume formative approaches to evaluation.

Frame 3 (Transformative Change) starts from very different premises with important implications for evaluation. It considers that previous innovation models have not been able to address societal needs and global challenges, including sustainability. Addressing these needs and challenges requires deep change in socio-technical systems toward a sustainable development model, i.e. sustainability transitions. Being Frame 3 still emerging in mainstream policy practice, TIPC defined a set of six criteria to provide a clearer definition of a TI policy and operationalise its dimensions: 1) directionality, 2) inclusiveness, 3) societal goal, 4) systemic impact, 5) deep learning and reflexivity, and 6) conflict vs consensus (Chataway et al., 2017; Schot et al., 2017).

The **Directionality** principle recognises that TI policy will foster some directions for innovation – that are desirable – and block others, while the **‘Inclusiveness’** principle refers to the involvement of civil society, users, local communities and those stakeholders usually marginalised by policy-making. Inclusiveness requires participation to promote inclusive processes and debates about the normative criteria of the policy-making. This is tightly linked with the Directionality principle: defining and deciding which directions are desirable is a political process.

The **Societal Goal** principle implies that environmental sustainability and societal challenges (like those encapsulated in the Sustainable Development Goals – SDGs) should be explicit goals of TI policy, while **systemic impact** refers to specifically addressing change and impact at the level of socio-technical systems. To achieve these systemic impacts, TI policy requires

learning and reflexivity in its practices and embracing conflict whenever this emerges. Reflexivity refers to ‘second-order’ (deep) learning, i.e. “problematization of operating routines of different actors and the creation of spaces for experimentation” (Chataway et al., 2017: 17). Second-order learning can only emerge if diversity of opinions and beliefs between the stakeholders are acknowledged and encouraged, because focussing on disruptive change will typically result in disagreements among the stakeholders. Therefore, conflict should be embraced as a natural element of TI policy.

Additionally, Frame 3 policies can be based on theoretical foundations that, implicitly, provide a generic theory of change: the socio-technical transitions framework focuses on the transformation of the key features of the production and consumption systems (food, transport, housing, finance or energy) and adopts a multi-level perspective considering that a socio-technical system consists of niches, regimes and landscapes (Geels, 2002; 2010, Geels and Schot, 2007).

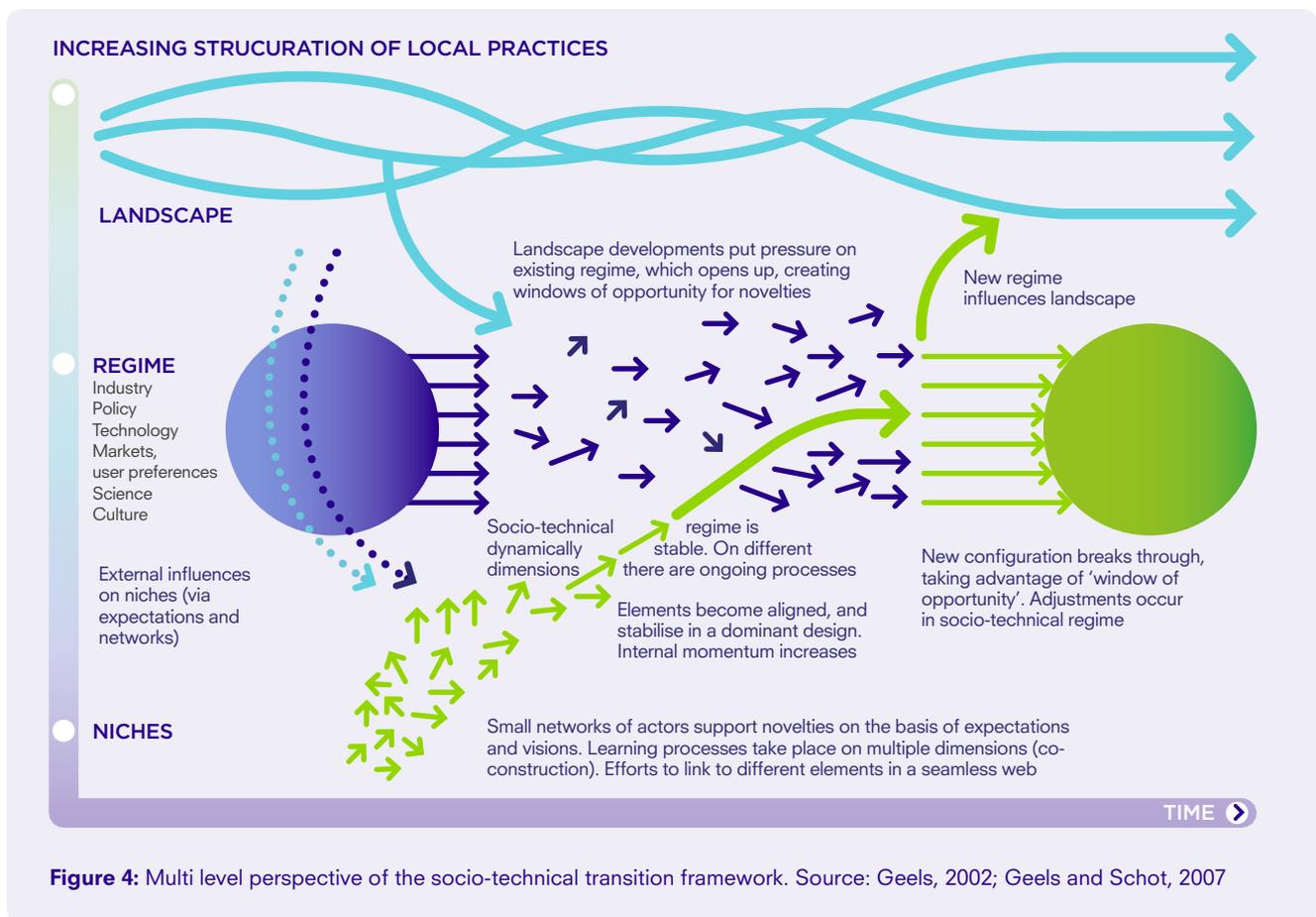


Figure 3: Principles of Transformative Innovation Policy. Source: TIPC, 2018

Socio-technical configurations in regimes are stable and dominant ways of realising a particular societal function, that is, they are the dominant configurations of practices, relations, discourses, etc. (Smith et al. 2010). However, a number of niches exist in a system: spaces in which alternative, less visible practices take place. These are protective spaces where different ideas, models, configurations and ways of doing try to survive and develop. Niches present configurations whose characteristics are different from those of the regime: they may work with different principles; may use different technologies; present different relations between stakeholders, channels and user practices; or may privilege different sources of knowledge and alternative cultures. Regimes are usually stable, whereas niches usually evolve quickly as they are spaces of

experimentation and change (Geels, 2002). Niches are the place of transformative ideas and practices, but their potential is constrained or enabled through the more powerful structures of the regime (Bos and Grin, 2008).

Following the ideas of the multi-level perspective (Geels, 2010, 2002), regimes try to survive and remain stable, but they are permanently exposed to pressures derived from external, powerful and long-term economic, social, cultural or environmental trends (Rotmans et al., 2001), which constitute the landscape. Transition in systems may take place when the regime is destabilised because of the heavy pressure of the landscape, so windows of opportunity may be open for niches – if they are mature enough – to influence or even completely replace the regime (Geels, 2002).



3.2 THE PRINCIPLES OF ‘FORMATIVE’ EVALUATION AND ITS APPLICATION TO TI POLICY

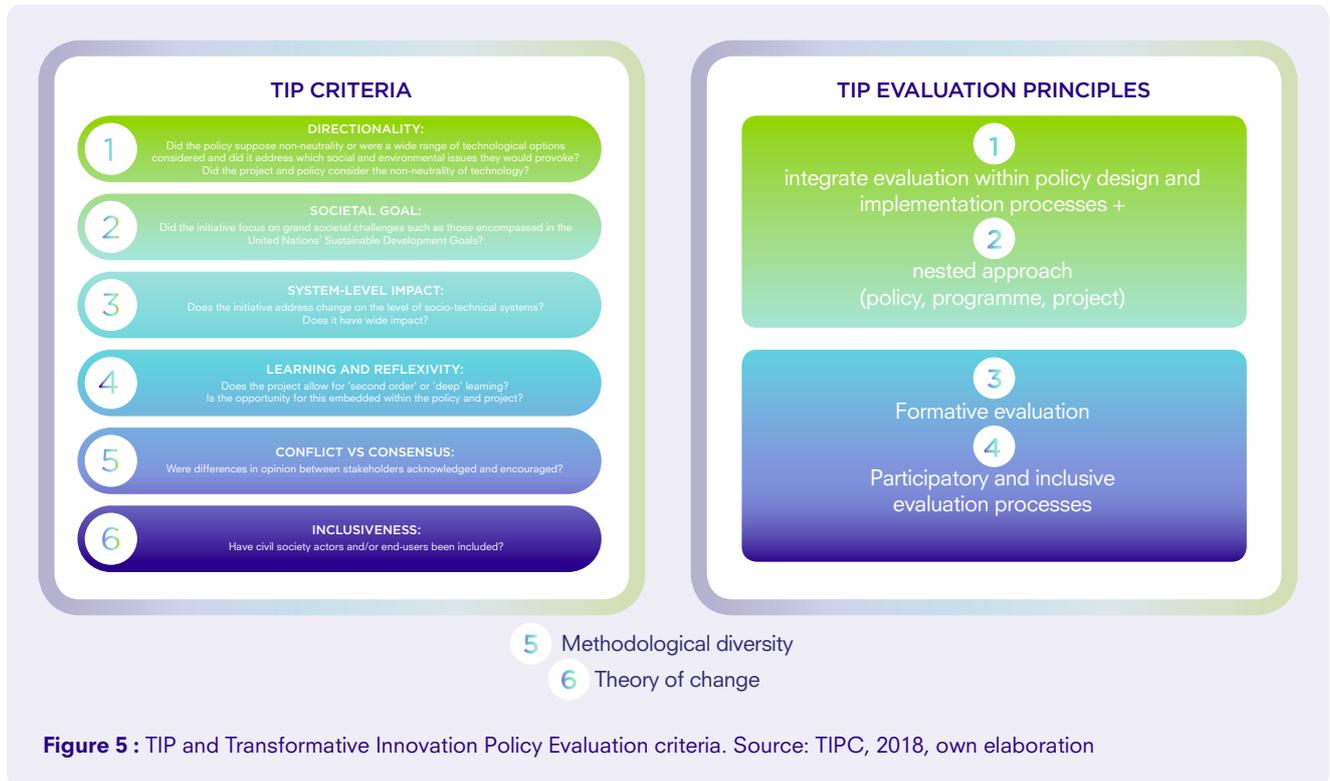
Inspired by TIP criteria and with insights provided by the literature review, we propose a set of 6 principles that can guide an approach to evaluation for TIPC:

- A **formative** approach to evaluation based on learning and reflexivity for the participants, consistent with the reflexivity and second-order learning criteria for TI policy. The formative goal aims to develop capacities especially inside TIPC member organisations. A formative approach also implies dealing with failure. This is very important, because failure can provide learning opportunities on the context, conditions and activities conducive to transformation processes. In this sense, evaluation can be understood as an experiment in itself and reinforces the necessary connections between evaluation and other core activities of the Consortium: capacity building first, but also experimentation and research. Evaluations can also help refining transformative innovation theory by providing information about different change processes triggered by different policies under different landscapes and dominant regime conditions.
- **Integrate evaluation within policy design and implementation processes.** We see evaluation as part of the policy process and, therefore, as a task that should share the overall characteristics we aim the process to have. Specific policies, their implementation and evaluation should be coherent with overall stated STI policy goals (directionality, societal goals and system impact criteria for TI policy). Evaluation is understood as a **strategic** part of the design and implementation process of a public policy.
- A **nested approach to evaluation** that will help us identify the policy level we aim to evaluate and its governance. The level of intervention could be: (1) a **policy** covering a broad area of public sector activity and aimed at generating an explicit set of results; (2) a **programme** of activities with an allocated budget and a pre-defined timeline for its implementation that involves several discrete activities; and (3) many programmes will fund a number of **projects**, specific activities implemented by an individual or team to address specific aspects and objectives of the programme to which they contribute.
- Evaluation should facilitate **participation** and open debate, and should not avoid conflicts of power, interest and perceptions (coherent with learning and reflexivity, inclusiveness conflict vs consensus criteria of TIP). The groups and communities with a role in the evaluation process will be varied and have different access to resources and, even, different interests. Managers and grassroots participants, for instance, may have different perspectives on the definition of the problems to be addressed, and be unequal in terms of the power they hold. An evaluation design should be attentive to such differences. At the same time, evaluation should be open enough open and reflexive, so as to the capture emerging trends and problems, and new goals, problems and actors.
- An **appropriate mix of methods and techniques** can be used in evaluation of TIP. Rather than formalised standard protocols, we need to be adaptable and flexible, selecting different methods and techniques according to the context, and the participatory and transformative nature of the policy. **Quantitative techniques** can provide synthetic assessments that allow for comparison across different units of assessment, and can provide, under specific conditions, robust assessments of the net impact of an intervention based on data compiled without the biases introduced by the personal recollection of individual participations. However, the measurement of social and environmental impact is difficult to be achieved with “standard” indicators. This difficulty is in part attributable to the nature of social values, which are often linked to incommensurable dimensions and perceived differently depending on cultural background and personal preferences. In these situations, **qualitative methods** can provide a better approximation to the study of social impacts by providing a fine-grained, contextualised description of change processes through detailed narratives. Finally, **participatory techniques** can help increase participation and inclusiveness of the evaluation process and allow a more horizontal involvement of the actors.

- The TIP criteria and theoretical foundations will inform the definition of a generic **theory of change** to support the evaluation of TIPs. Each evaluation of a specific TIP intervention will be based on specific theories of change consistent with the generic one. The specific theory of change will be flexible (it can be revisited as a result of the formative evaluation process) and will distinguish the key dimensions of an intervention: the specific context in which the experiment operates,

the specific problems it needs to tackle to achieve its objectives, the mechanisms (processes) the experiment aims to deploy with the resources available, the expected outcomes it aims to achieve and how these outcomes are important for the unfolding of a socio-technical transformation.

The following figure represents the TIP evaluation criteria and their links with TIP criteria as described previously.



Finally, based on the principles above, we propose a tentative definition of how evaluation can be framed for the TIPC:

Evaluation can be understood as a formative and strategic dimension of a project, programme or policy process. Evaluation should be informed by TIP principles, and deploy a participatory and inclusive process through which learning and capacities are developed among the participants.

4. A FORMATIVE APPROACH TO TI POLICY EVALUATION USING A THEORY OF CHANGE (TOC)

4.1 A SYSTEMIC APPROACH IN PROGRAMME THEORY

The idea of ToC draws on the programme theory approach to evaluation (Weiss, 1997). A ToC is an evaluative planning tool that enables stakeholders to work backwards from policy objectives, mapping expected outcomes and outputs, processes and inputs that will need to happen in order to achieve the project goals. Through this, an explicit pathway is laid out for the framework and the evaluators.

The ToC approach to evaluation is closely related with theory-based and 'realist' approaches (Pawson and Tilley, 1997). It seeks to provide a theory about how the changes a policy, programme or project intends to trigger are expected to come about. As Stame (2004, 2004:60) stated: "Theories should be made explicit, and the evaluation steps should be built around them: by elaborating on assumptions, revealing causal chains, and engaging all concerned parties in the exercise".

ToC has its roots in programme theory. The use of programme theories in evaluation has a long history. The basic idea is already present in Schman's work during the 60s, suggesting that evaluation should focus on the chain of goals of the intervention. The same idea is repeated by Bennet (1975), introducing the concept of outcomes hierarchy (Rogers et al., 2001). Progressively, other scholars and professionals started to support the idea that it is essential to develop a theoretical framework as a rationale underpinning the programme. Carol Weiss (1972) recommended to identify potential causal models of the programmes, which she defined as 'Theories of change': "the chain of assumptions explaining how activities lead step by step to the expected outcomes" (Weiss, 1998:2).

More recently, programme theory evolved following a systemic approach focusing on three key dimensions: outcomes, processes and structural elements (Ligero, 2011). The approach suggests focussing on outcomes previously defined by programme goals, and observing all the potential changes produced by an initiative even when they bear no connection with the intended goals.

Outcomes and goals are deeply intertwined. Goals are projected solutions previously defined, while outcomes are observed changes provoked by the intervention. From our point of view, we define outcomes as changes in people, organisations and practices, e.g. acquisition of capabilities, new relationships, developing new skills, etc. Some programme theories distinguish between **outputs** (the direct results generated by the intervention), **outcomes** (the immediate benefits for the beneficiaries of the intervention) and the wider **impacts** on society (Boekholt et al. 2014). In any case the way the labels attached to the flow of results traceable to an intervention have no effect on the intervention logic itself.

Outcomes are generated through **processes**. In a systemic approach to programme theory-based evaluation, a process is a sequence of activities generating a value for the beneficiaries. Through a chain of activities something new is produced, i.e. a product, a service, an orientation or a change in the observed population. By contrast, the **structure** of the programme is defined as "the (relatively stable) organisation of different types of resources in order to obtain project goals" (Cohen and Franco, 1992). Structural elements include several characteristics of a programme needed in order to make the processes work: budget, the nature of human resources, management, situation, plan of activities, methods to deliver a service, operational principles, inter-institutional partnerships, norms, etc. Finally, the context influences the structure. The **context** is not directly related to the intervention resources but has an influence on the intervention, e.g. other public policies. Table 3 synthesises the definitions of key elements of the systemic approach to programme theory while Figure 6 represents graphically the connections between the elements.

ELEMENTS	DEFINITION
CONTEXT	Not directly related to the intervention resources but has an influence on the intervention
STRUCTURE	The (relatively stable) organisation of different types of resources in order to obtain project goals
PROCESS	A sequence of activities generating a value for the beneficiaries
OUTPUTS	Immediate effects of the intervention
OUTCOMES	Intermediate effects produced by outputs
IMPACT	Long term effects produced by outcomes

Table 3: Definitions of key elements of programme theory. Source: Magro and Wilson (2013)

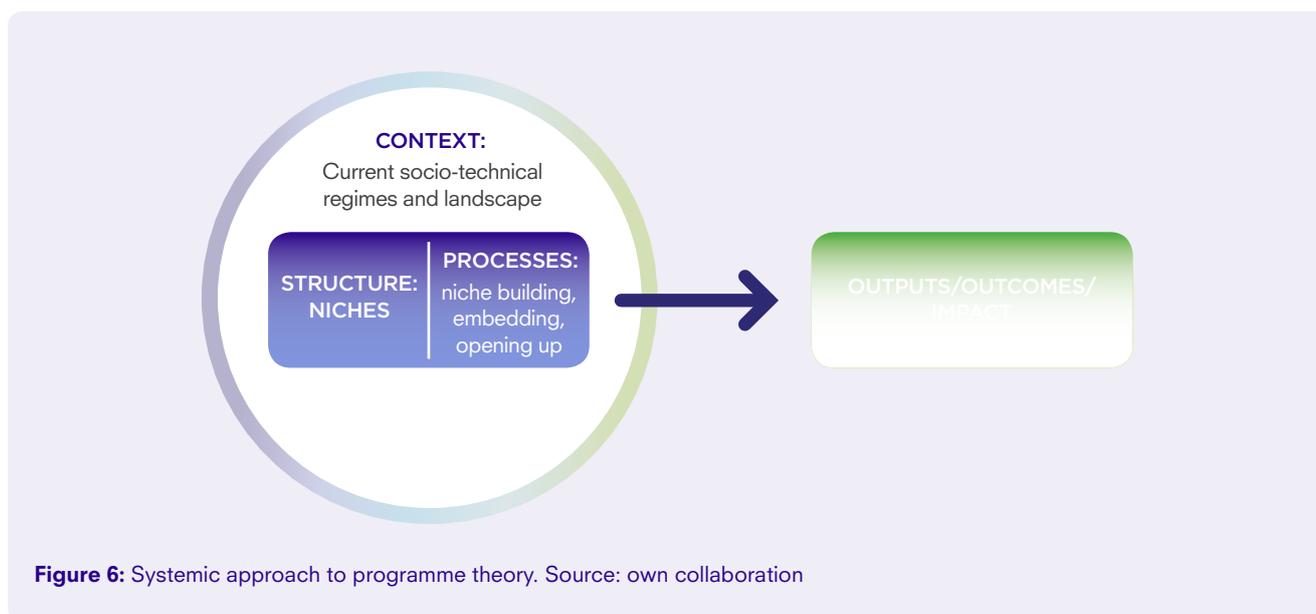


Figure 6: Systemic approach to programme theory. Source: own collaboration

Following this logic, the key evaluative questions are:

- Which outcomes have been obtained (both planned and not planned) and their relationships with the goals? Has there been negative, not beneficial, or unexpected outcomes?
- What are the key features of the processes that the interventions has generated?
- How can the outcomes be attributed to the intervention and through which processes have they emerged?
- What are the relevant “elements of the structure” with an influence on the results of the intervention?

4.2 A GENERIC TOC FOR TI POLICY

Starting from this systemic understanding of the ToC, we should specify the key characteristics of a ToC for TI policy. This has two components: 1) how the assumptions in TIP can be reflected in its generic ToC; and 2) how the implementation of a ToC approach to the evaluation of TIPs can be done in a formative manner.

The ToC of TIP

A generic ToC for TIP, should be informed by transitions theory. A key assumption (as for instance in the “third frame” of innovation policy) is that to address key societal problems like those laid out in the SDGs requires profound changes in the current socio-technical systems, which in turn call for a broader and different mix of STI policies. This policy mix starts with the developments of policy experiments in protected niches. It is expected that these experiments will stimulate socio-technical regime change in specific directions aimed at addressing specific societal challenges. Therefore, they have explicit directionality. The bottom-up processes triggered by such niche experiments need to be nurtured and empowered if they are to develop past the initial set of niches, deepen their scope and widen their reach. If successful, the experiments will turn into drivers of transformation, creating new expectations expanding beyond the niches and altering the socio-technical system into new configurations that respond to the societal challenges being addressed.

Besides, to achieve their transformational goals, all policy interventions pursue changes in the structure and culture of governance, emphasising inclusive participatory processes that involve a multiplicity of organisational forms. These changes must be extensive to evaluation methods and practices and imply crucial modifications in the roles and identities of the participating actors (including evaluators), shifts in organisational routines, and the emergence of new ways of framing of problems.

Consistently with the overall ToC approach, we suggest a specific understanding of the elements of a ToC for TIP. We define the **context** element as including elements of the landscape and regime that provide the overall frame within which the intervention takes place, but are not directly addressed by it. **Structure** refers to all the inputs necessary to carry out the experiments. These can be part of the regimes or the niches. **Outcomes** are changes in people and organisations and, according to TIP theory, can be related to changes in learning, expectations, networks and different elements of regimes. **Impact** is a long term effect important for the unfolding of a socio-technical transition to address societal challenges.

ELEMENTS	DEFINITION
CONTEXT	Elements of landscape and regimes
STRUCTURE	Inputs necessary to carry out the experiments (human, financial, organisational, legal, ethical, etc. could be part of niches, regimes, intermediaries, etc)
PROCESS	The experiments
OUTCOMES	Changes in people and organisations
IMPACT	Long term effects produced by outcomes (related with big societal challenges)

Table 4: Elements of a ToC for TIP

A formative approach

Because our main goal is not to provide a summative judgment for accountability purposes, but instead to foster (second order) learning and reflexivity among the participants through the evaluation process, the ToC should not be understood as a logical and fixed causal chain between resources, processes, outcomes and impact. Instead, a ToC can help decompose a policy, programme or project experiment to aid reflection on their interactions, and, especially, to assess if the experiment is contributing to generating outcomes that the TIP considers relevant to bring about change: first and second order learning, networks, changes in governance, etc.

Nevertheless, several evaluation problems originate from the ambitious, systemic objectives of TIPs (addressing key societal problems – SDGs – or fostering systemic changes with directionality). They amplify the ‘attribution’ problem, common in the field of research impact assessment: the results of an intervention can be no more than a contribution to (and not a determining cause of) the systemic changes being pursued. The ‘distance’ between an experiment and the change in a specific socio-technical system poses an even bigger attribution problem that is the norm in more “traditional” “Frame 1” policies. Besides, TIP policy relies on a set of interrelated experiments operating at different levels. This involves greater complexity, e.g. different projects defining local experiments operating in local niches, strategies to deepen and scale up, etc. The processes, timelines and immediate objectives of each initiative are different. TIP policy requires considering and mapping the interactions among this diverse set of policy initiatives.

To tackle these problems, we suggest, as we will describe with more detail further below, not to consider the impact dimension when evaluating projects or programmes. Only at a policy level, can system-level impacts be evaluated.

Finally, it is possible that different participants in TIP experiments will hold different ToCs in mind and pursue different objectives, driven by different interests and beliefs; then, and according to two TIP principles (inclusiveness and conflict vs consensus), it is possible that participants and involved stakeholders will not be able to agree on a single ToC. Therefore, we must be prepared to work with different ToCs, which, additionally can be altered during the evaluation process. The process of discussing these changes should be considered a means to reflect on TIP and revisit existing transformative innovation theories.

The rest of this section will develop in more detail the steps that need to be taken to implement this evaluation approach and illustrate them with examples of relevant methodologies and evaluation techniques that can be used to carry out specific evaluations within this framework. The approach proposed requires three initial steps:

1. Identify the level of the TI experiment (whether we are considering a single project, a programme of activities or an overall policy), and the main actors involved.
2. Identify key evaluative dimensions to be assessed. These will be different depending on the level of the TI policy or experiment: project, programme or policy.
3. Discuss the ToC and check their consistency with TIP criteria. Reformulate the ToC if necessary.

4.3 A THREE-STEP SEQUENCE TO BUILD THE TOC

Identify the level of the TI policy or experiment

Identifying the level of the TI policy experiment implies reflecting whether it is considering a single project, a programme of activities, or an overall policy including different programmes:

- Projects in protected niches set up as experiments to stimulate local socio-technical changes. They emphasise changes in governance to support broad participation.
- Programmes link several projects to enhance synergies and learning, and to protect and nurture niches. They seek to deepen the scope and widen the reach of the niche projects.
- A TI policy is constituted of several programmes targeting a specific socio-technical system or problem area.

The definition of ToCs and their implementation in evaluation need to deal the participation of stakeholders and the ensuing politics of evaluation. Cronbach argued that evaluations that only allow one group to define the evaluation questions tend not to be useful. Democratic societies are characterised by pluralism and evaluation should consider this feature (Chen: 61). Accepting this principle implies asking what actors are involved in and influenced by the policy/programme/project, which ones among them have legitimacy to participate in the evaluation process and how to make that evaluations be sensitive to their different needs in order to be useful.

The orientation towards actors has been called pluralism or ‘responsiveness’ (Stake, 2006), meaning that the evaluation is sensible to the needs of the different actors involved into the policy action.

Identify key evaluative dimensions to be assessed

Generally, evaluations focus on four areas: resources, processes, outcomes, and impacts. Each area poses specific questions and challenges.

1. Resources: what is invested? Often evaluators and policymakers focus on key inputs such as financial or human resources. While these need to be considered, analysing inputs to TIP experiments requires us to address also less tangible but equally important resources. If participatory evaluation is to be successfully achieved, trust must be present as well as support, commitment, and expertise. Is there a collective willingness to collaborate on a mutual footing? Are different and conflicting interests acknowledged?

2. Processes: what activities are being carried out? Charting and assessing the process of an experimentation can often shed light on how success (or lack thereof) is achieved, and how the process could be improved. It should be noted that, taking a formative perspective, an unsuccessful TI policy experiment will not be written off as just a failure; there will still be valuable learning lessons from that experiment’s process that will lead to learning. For TIPC, the policy experiments must provide opportunities for reflexivity and deep learning. How did power dynamics affect the decision-making and learning processes? Have different directions for technology and pathways of change been considered?

3. Outcomes: what are the relevant changes that the experiment has helped generate? In our definition, outcomes include the relevant organisational and behavioural changes enabled by the experiment. In TIPC, we are interested in outcomes that support socio-technical transitions; for instance, the project may lead to first, and particularly second, order learning by the participants, deep and broad networks, and changes in expectations. It is important to distinguish between learning from the participants, and the learning generated by the experiment activities. Have the actors

involved in the experiment learnt any new, relatable skills that could be used in furthering experimentation in transformative change? Have the participants changed their perception, or challenged their view about socio-technical routines? Have practices and routines changed? Can we observe transformations in narratives, behaviours and networks?

4. Impacts: how socio-technical systems have changed? Impact is often what focuses the attention of policymakers interested in policy evaluation. For TIPC, impacts of interest should relate to change in socio-technical systems that allow societies to address social and/or environmental challenges. Taking the SDGs as a reference point, we could consider whether specific SDGs have been achieved and the way in which socio-technical transitions have contributed to this achievement. Note here that the emphasis is in identifying the contribution of a policy, rather than “attributing” impact to a single policy source. How can we assess the contribution of a specific policy to a process of socio-technical transition towards sustainability?

The key dimensions to be assessed within these four areas will be different depending on the level of the experiment: project, programme or policy (Figure 6).

At the **project** level, we can distinguish:

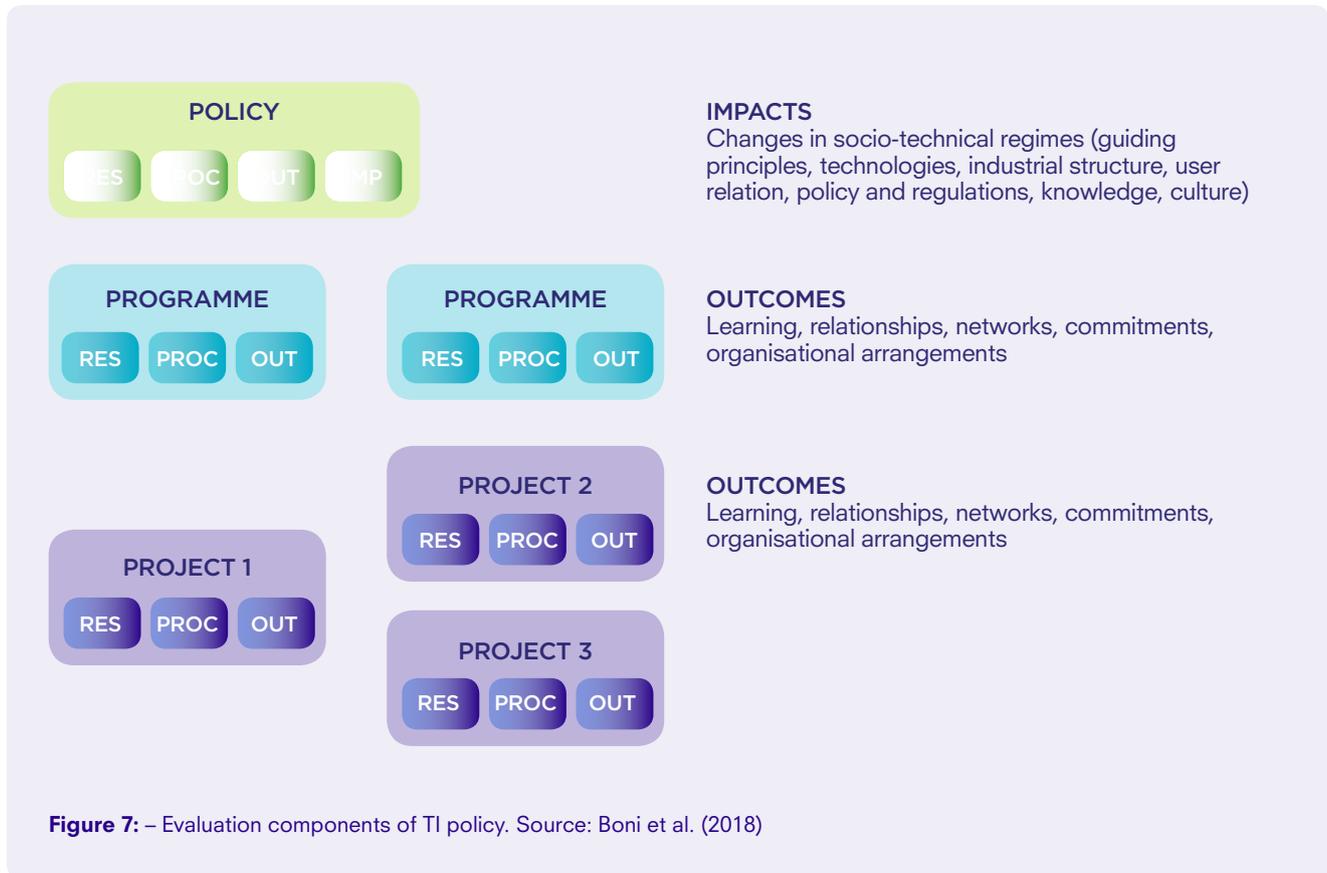
- Resources: the inputs needed to carry out the project (human, financial, material, organisational, etc.).
- Processes: all the activities the project aims to carry out and the processes they are expected to trigger
- Outcomes: the main changes generated directly by the experiment including first and second order learning, and those organisational and behavioural changes enabled by the direct changes, including new relationships, changes in the role of different actors, new commitments, new organisational arrangements, changes in the structure of governance, creation of networks, etc.
- Impacts: projects are understood as the first step in a long chain of events leading to systemic impacts in the form of socio-technical transitions, but a specific project in itself cannot cause such impact; therefore, impact assessments should not be used to evaluate single projects.

At the **programme** level, we can distinguish between the following:

- Resources: the inputs needed to carry out the programme (in terms of human, financial, material, legal, organisational, etc.) including the inputs invested in all the specific projects conforming the programme and those required for programme level activities (management, coordination, niche protection and empowering, etc.)
- Processes: how different activities conducted by the projects conforming the programme relate to each other, and how the niches associated with these projects are protected, nurtured and empowered.
- Outcomes: at the programme level, the outcomes observed will be the sum of those generated by the projects plus (1) additional changes in learning, expectations, commitments, organisational arrangements, and governance structures generated through the interactions among projects and activities conducted at the programme level, and with organisations and communities that are not participants in the programme activities, and (2) the upscaling, replication, circulation and institutionalisation of the niches associated with the projects as part of the process of accelerating and embedding niche innovations (Turnheim et al., 2019).
- Impacts: as at project level, impacts cannot be observed within the scope and timeframe of a programme.

Finally, at the **policy** level we can distinguish between the following:

- Resources: the accumulation of resources invested to achieve a socio-technical transition, including but not limited to those invested in the programmes.
- Processes: how the different programmes interact among each other and complement other initiatives oriented to protecting, nurturing and empowering the project experiments (including changes in regulations, institutional adaptations, etc.)
- Outcomes: changes in learning, expectations, commitments, organisational arrangements, and governance structures observed outside the groups involved in the initial projects but which can be traced back to such projects (i.e. outcomes attributable to scaling up processes)
- Impacts: observable changes in socio-technical regime that can be traced back to activities generated through the programmes. These changes will include shifts in the overall ideas and assumptions driving the system, the technologies used; the industrial structure (i.e., the relationships between stakeholders in production processes), user relations and channels to access goods and services (i.e., infrastructure, modalities, possibilities and interactions for accessing goods and services); the forms and sources of knowledge produced, legitimised and used by the system, and cultural changes (i.e., social and cultural patterns in the socio-technical configuration). These changes in the socio-technical system should be reflected in some of the aggregate indicators that conform the SDGs.



Make explicit the ToC of the experiment coherent with TIP criteria

As we pointed out before, each participant can have a different ToC in mind. From a formative evaluation perspective, one of the roles of the evaluator is to enable and manage a discussion on the different interpretations, and if possible, agree on an initial ToC that can be modified during the evaluation process. If such agreement is not possible, the different ToC interpretations should be incorporated and discussed along the evaluation.

The evaluator should also take care of the incorporation of TIP criteria in the evaluation design. TIP criteria will orient the evaluation toward the consideration of the following results:

- Deep learning: it happens when actors question their underlying assumptions, critically assess their own

preferences and experiment with alternatives.

- Directionality: articulation of new shared expectations and visions aligned with social and environmental goals (social needs, sustainable and inclusive societies).
- New networks of participants sharing this directionality: opening-up and broadening networks; ensuring incumbents are not dominant.
- Deepening networks: bringing in actors with more mobilising power in terms of resources and underlying networks and strengthening intermediation.
- New knowledge: often multifaceted (sustainability, history, development, etc.), involving a dialogue between expert and non-expert knowledge.
- Changes in policies: coordination of policies from various domains (e.g. transport, energy, etc.) and from various levels (local, regional, national, international).
- Changes in governance towards more flexible and open models.

5. A COMPENDIUM OF RELEVANT EVALUATION TECHNIQUES

The aim of this review is to sketch a comparison of some techniques and tools currently used in the evaluation of research, and innovation policies. Specifically, we focus on those techniques that we consider well-suited to address TI policy evaluation aims, and are easy to access by SPRU researchers because they have been developed or implemented within SPRU.

We provide an outline of the key points of each technique, the reasons for its use, its assumptions, advantages and limitations, and the kind of problems that can generate when applied to TI policy evaluation. We distinguish between those techniques originating from the field of research evaluation or the assessment of science policy (MCM, DARE, research portfolios) and those developed in the context of the evaluation of innovation and development policy, programmes and projects (OM, PIPA, policy mixes analysis). It is also worth distinguishing between those techniques that have been developed within SPRU or in collaboration with SPRU researchers (MCM, DARE) from others that are known in SPRU (OM, Portfolio Analysis, Policy mixes) or received significant contributions by SPRU researchers (PIPA).

In the final section, we summarise the main overlaps and differences between techniques, and we present a brief assessment for the purposes of building a toolkit for TI policy evaluation.

5.1 MULTI-CRITERIA MAPPING (MCM): DEFINING (AND CHANGING) THE PROBLEM AND THE STRATEGIES OF A POLICY ISSUE

MCM is a technique developed by Andy Stirling and Josie Coburn (SPRU), drawing on previous theoretical and empirical studies (Burgess et al., 2007; Stirling, 2007; Stirling, 2008). It has been applied in the context of research and science policy deliberation, and to a lesser extent in innovation policy and development.

MCM is a technique for analysing complex issues and appraising different options to address a specific aim against a set of criteria set up by the participants. These are usually policymakers, stakeholders and experts. Based on the value of plurality, MCM assumes that there is no 'best strategy' for policy options. Following Arrow's

impossibility theorem about democracy, MCM assumes that aggregating individual preferences cannot lead to an 'optimum' choice.

MCM can be practiced as follows (Coburn and Stirling, 2016): stakeholders are gathered in a workshop organised by the facilitators. A set of potential options or strategies for the policy problem to be addressed is shortlisted. Then, participants set up their own appraisal criteria and their weights, and they are also free to imagine and set up additional options. With the aid of the facilitators and a software assisting visualisation, participants can assess the alternative options according to their criteria. All the participants' results are gathered by the facilitators using the software and the results are used to open a discussion. This discussion is the core of the technique: thanks to the software support, it is possible to see the potential benefits and pitfalls of each strategy/option according to the evaluation criteria, and which criteria are more important. There is no protocol nor software support to the discussion dynamics but this can be organised following usual qualitative techniques (e.g. for workshops, focus groups, etc.).

MCM has two main advantages. First, it embraces uncertainty, and provides a big picture of alternative pathways and their relative assessment. It is particularly suited for addressing complex and unpredictable issues such as innovation and socio-technical transformations, because it provides detailed reasons why participants prefer certain options, aims to open-up discussions and (in the words of Andy Stirling) "policymakers cannot blame anything nor anyone if the policy eventually does not work". Second, it is an open approach based on participation and diversity: participants set up their own appraisal criteria and their weights (sometimes additional options too), they are constantly aware of data production and can re-define their judgements during the process. Besides, the facilitators can group the preferences according to types of participants, making them able to address issues of inequality or power relations. All these characteristics are consistent with the type of evaluation required for TI policy (focus on learning, reflexivity, diversity, participation, etc.). An extra advantage is the availability of software assistance, and its ease of use.

However, MCM requires one or more time-consuming workshops. In addition, it requires well-trained facilitators and a detailed previous work for defining the set of options and the goals of the assessment. Another problem is directly related to its core advantage: openness. As there is no specific mechanism for ‘closing down’ the decision-making process, sometimes the exercise does not reach a clear end, or can derive in a situation of potential conflict. Commitment, collaboration, and acceptance of diversity by the participants (and the facilitators) are key for the feasibility of MCM.

MCM can be a useful tool for TIP formative evaluation designs as it can be used to support decision-making processes during policy definition and implementation. For instance, it can provide a useful mechanism of ex-ante evaluation for deciding which solution(s) should be funded/presented to a call by a funding agency/research community. Thanks to its participatory nature, it can be useful for defining a specific strategy for policy and community experimentation.

Extending the technique beyond its original application area, an interesting possibility for TIP is using MCM as a tool for defining a specific ToC for an experiment drawing on the generic ToC presented in Section 4. In this sense, the generic ToC serves as a base to structure and inform the workshop, facilitating the work of the evaluators, i.e. providing (or helping to find) an initial set of options and criteria. In this way, MCM can provide a ‘standard’, open and participatory tool for deciding how the experiment might (contribute to) transform a specific socio-technical system in a specific context. Another new application of MCM can be as a “Real Time” evaluation technique, e.g. to modify or adapt the initial specific ToC of an experiment after some period of implementation. In this sense, the previous experience could deliver the inputs of the workshop, structured more as a ‘reflexive exercise’ instead of a decision-making process. Current options could be reviewed, confirmed or withdrawn, and new options could be tested. For instance, participants may assess the contribution of each strategy to the goal and assess its (temporary) effectiveness. New criteria could emerge, and new participants may be engaged.

5.2 PORTFOLIO ANALYSIS (PA): ASSESSING GAPS AND POTENTIALITIES OF EXISTING POLICY THROUGH BIG DATA SETS

Portfolio Analysis is a technique for decision-making and evaluation in policy and management, coming from finance and – specifically – from financial portfolio theory. The original aim of PA is assessing the expected returns of an investment portfolio based on the expected returns of the constituent stocks. It has been applied also to scientific research, using as unit of analysis the research portfolio (Oros et al., 2007; Srivastava et al., 2007). Usually bibliometricians analyse research portfolios with software support to perform graphic maps or network analysis (Wallace and Rafols, 2015). A similar perspective can be extended to policy studies, e.g. including the analysis of innovation policy mixes (Magro and Wilson, 2013; Kivimaa et al., 2017).

Portfolio Analysis is largely descriptive and comparative in nature. Overall, PA is a technique in which alternative options are evaluated by how they contribute to different objectives and according to several ‘balancing of ambitions’ and risk-reduction criteria, including economic efficiency criteria.¹ “According to Srivastava et al. (2007), “A ‘portfolio’ is any collection of things – objects, ideas, assets, accomplishments, etc. – physically or metaphorically bound together by a common thread.” For instance, “research portfolios can be defined by virtually any theme of interest, including those related to administrative needs, organisational structure, funding streams, goals, and results” – like research programmes (Srivastava et al. 2007). In the field of research, PA tools are used as an additional source of information in support of research management. Some research evaluation efforts already incorporate portfolio analysis activities. Sometimes, PA simply aims to provide managers or other stakeholders with information about the size, composition, diversity and other essential features of an existing research portfolio as an end unto itself, often for mere administrative purposes.

¹ According to Davis et al. (2008), a portfolio analysis “assesses alternative investment options by diverse quantitative or qualitative objectives, including risk mitigation. The analysis aids in balancing investments within a portfolio, i.e., in a mix of instruments. The intent is to address all objectives and mitigate all risks, but to different degrees, depending on priorities, budgets, and feasibilities.” The same report defines a portfolio as “[...] a collection, as in investments in a number of categories. A strategic portfolio might have categories for different functional classes of capability or categories for different theatres of interest; a portfolio within a capability area would have categories for different contributors to that area’s overall capability”.

Usually, PA rests on several building blocks:

- A recognition of the networked, cumulative nature of knowledge generation and application, and of the systemic nature of policy-making, making necessary assessing not only individual projects, but also programmes or portfolios of projects.
- The increasing breadth, depth, quality, and sophistication of available datasets, upon which PA techniques rely.
- Growing consensus among science policy makers that theories, models, and tools borrowed from other fields can and should be used to make research management more “scientific” (e.g. Marburger, 2005).

PA can also be useful to identify areas of overlap in existing research portfolios, and such information can inform resource allocation. Alternatively, identifying areas of overlap can help suggest priority areas for coordination or collaboration across existing programmes, initiatives, or organisational boundaries. It can also help to identify gaps in research portfolios relative to existing goals or obligations. Comparing internal portfolios to external ones may allow research organisations to understand the role of internal research portfolios relative to a broader context.

However, PA techniques suffer also from some limitations. They do not typically extend beyond providing a summary of the attributes of a specific portfolio, or a straightforward comparison of the attributes of one portfolio to those of another or an external standard. Therefore, they are tailored for use in theory-driven evaluation approaches. Another problem concerns data availability. Although this is improving, PA typically requires detailed and robust data and can be affected by the diversity of data sources and the resulting variability in data quality, granularity, and structure. These pose challenges that can be partially addressed through the use of data mining and visualisation techniques for portfolio analysis, allowing for the analysis of unstructured data.

In terms of policy, evaluations of research portfolios must be designed carefully to address questions related to process, outcome, effectiveness, efficiency, progress relative to goals, etc. Evaluators may also be required to make prescriptive recommendations to portfolio managers about how best to achieve a given set of goals. In terms of performance, the quality, inclusiveness, and transparency of that organisation’s data system is critical for addressing the methodological challenges posed by PA (Srivastava et al., 2007).

5.3 OUTCOME MAPPING (OM): PLANNING AND MONITORING IMPROVEMENTS COMBINING TOP-DOWN AND BOTTOM-UP STRATEGIES

Outcome mapping (OM) is a technique developed in the field of international cooperation and sustainable development for planning, monitoring and evaluating development projects and programmes (Earl et al., 2001; Smutylo, 2005). However, it has also been applied to projects and programmes in the fields of research communication and uptake (Young and Mendizabal, 2009). It focus its attention on the middle of the chain linking upstream interventions with downstream impacts. OM has developed standard protocols for its application.

OM INCLUDES THREE STAGES OF APPLICATION

A. INTENTIONAL DESIGN: HELPS THE WORKING TEAM TO BE SPECIFIC ABOUT TARGETED ACTORS, EXPECTED CHANGES, AND STRATEGIES TO BE USED:

1. Vision
2. Mission
3. Identification of boundary partners
4. Outcomes challenges statements
5. Progress Markers
6. Strategy maps
7. Organisational practices

B. PERFORMANCE MONITORING: SYSTEMATIC SELF-ASSESSMENT OF PROGRAMME/ PROJECT'S ACTIONS AND PARTNERS' PROGRESS TOWARD THE ACHIEVEMENT OUTCOMES (THROUGH JOURNALS)

8. Monitoring priorities
9. Outcomes journals
10. Strategy journals
11. Performance journals

C. EVALUATION PLANNING: IDENTIFYING EVALUATION PRIORITIES AND DEVELOPING THE EVALUATION PLAN. OM UNPACKS THE POLICY'S TOC AND DATASETS TO BE USED:

12. Evaluation plan

Source: <https://www.outcomemapping.ca/>

The first seven steps (and, specifically, the progress markers) serve to outline the logic and strategy of the intervention, including vision, mission, boundary partners, outcome challenges, graduated progress markers, strategy maps, and organisational practices.

OM defines outcomes as changes in behaviour, relationships, activities, or actions of 'boundary partners' – the people or groups with whom a programme/project directly works – instead of the end users or beneficiaries. Such changes are linked (but not attributed) to programme/project activities. OM assesses the contributions to outcomes that, in turn, may enhance impacts. In addition, it provides a reflection on 'how' changes happen. OM assumes that change occurs as result of many actors and factors, many of them specific to the context in which an intervention takes place.

OM has several advantages. It is a robust methodology that can be adapted to a wide range of contexts, and

is supported by a variety of 'standard' procedures and protocols and a well-developed community of practice. It also addresses the problem of impact attribution in an original way, focusing on contributions to impact and relevant, but often overlooked, dimensions (behavioural change, organisational change, etc.). Besides, OM can be used in combination with other evaluation techniques such as Most Significant Change Analysis, Force Field Analysis, or Stakeholder Analysis.

However, OM also faces limitations. First, despite the engagement of 'boundary partners' in design, data collection, ownership and use of findings, they are not fully involved in the evaluation study, which is mainly carried out by specialist evaluators. This can raise concerns about the inclusiveness of the technique. The second main limitation is the need for skilled facilitators and a dedicated budget and time. OM requires a substantial amount of resources, and it has been usually applied for big programmes or by resourceful

organisations. Additionally, it often requires a radical ‘mind shift’ for the actors involved and this can generate misunderstandings, requiring proper strategies and resources to be addressed.

Although initially developed for application in development projects, OM seems applicable to research and innovation policies, where some modification could be required, like the development of a ‘rapid’ version of the technique including a theory-driven approach i.e. developing a ToC during the planning phase (Young and Mendizabal, 2009). This transformation can make OM very similar to PIPA, as we show below.

Overall, OM can be useful for TI policy evaluation because it introduces monitoring and evaluation considerations from the very planning stage of the experiment and engages the work team in continuous evaluation and self-assessment. It also raises awareness of the immediate experiment goals, builds consensus and produces empowerment for the work team. This is consistent with the focus on formation, reflexivity and inclusion of TI policy evaluation. Besides, it has been usually applied to the planning and implementation phases. Therefore, it can be useful for a formative approach aiming to assess the policy implementation from the very beginning of the experiment. Finally, OM focuses on some dimensions like organisational change, behavioural change, learning, interactions, and unintended outcomes, which have been identified as important by our literature review on sustainability transitions and innovation policy systems (see Section 2). Mapping these changes as they are produced can be useful for the development of the experiment.

5.4 PARTICIPATORY IMPACT PATHWAYS ANALYSIS (PIPA): PLANNING AND MONITORING IMPACTS USING THEORIES OF CHANGE

PIPA is a planning, monitoring and evaluation technique developed in the field of development programmes and projects, especially for the agricultural sector (Douthwaite et al., 2007; Álvarez et al., 2010) The STEPS centre of the University of Sussex used PIPA for evaluating action-research, researchers’ engagement and societal impacts of research in the field of sustainable transitions and socio-ecological research (Ely and Oxley, 2014).

PIPA is a practical tool for complex projects, to help participants collectively define a programme theory using participatory workshops. Participants make explicit how they see themselves achieving their goals (through and beyond the project) and, then, derive outcome targets, milestones measuring progress towards them and design a monitoring and evaluation plan (to make corrections in the implementation). It usually begins with a participatory workshop (usually 3-days long) where:

1. Stakeholders make explicit their assumptions about the intervention, using problem trees to clarify their expectations of the ‘impact pathways’.
2. Stakeholders draw and discuss actor and network maps, to find the actions and strategies to establish or strengthen partnerships and relationships thereby, ensuring that the impacts achieved will benefit the expected beneficiary groups.
3. Stakeholders write an outcomes logic model and its narrative, describing the project’s short, medium, and long-term objectives in the form of hypotheses (e.g. which actors need to change; what are these changes; which strategies are needed to realise these changes; how will the project affect people’s livelihoods, etc.). They deliver outcome targets and milestones which will be regularly revised during the M&E phases.

PIPA mixes the idea of ‘impact pathways’ (social, technical and environmental interactions for system change) with the consideration of power and power relations using narratives and causal hypotheses. A project impact pathway describes how it will develop its outputs and who outside the project needs to use them to achieve the desired impact. PIPA recognises the relevance of how actors actually see and frame the system, and includes such framings when drawing the pathways. In addition, PIPA recognises the relevance of learning processes for policy-making and policy evaluation. For instance, the initial programme theories developed during the kick-off workshop are usually amended later.

PIPA goes beyond the traditional use of logic models and log frames by engaging stakeholders in a structured participatory process, promoting learning and providing a framework for 'action research' on processes of change. It involves a broad set of stakeholders: project implementers, 'next users' (people who will use what the project will produce), 'end users' (people served by the 'next users') and politically important actors. Importantly for TIPs, although some grade of consensus should be achieved, differences are accommodated and acknowledged in the records of discussions. The flexibility and adaptability of the techniques make PIPA instruments suitable for evaluating TI policy, e.g. building a specific ToC for the experiment and assessing its improvements and contributions to socio-technical transformations. Besides, PIPA is compatible with the use of quantitative indicators, which are set beforehand in a participatory way as monitoring targets.

A limitation of PIPA lies in the considerable amount of time to be consumed for both building the specific programme theory and the ensuing monitoring efforts, requiring good training and facilitating skills. Also, once the programme theory and a set of goals and associated objectives is established, the monitoring process focuses on these goals, offering little scope for revisiting the initial assumptions, the agreed programme theory and for identifying unexpected outcomes falling outside the agreed and expected impact pathway. However, within TIPC there is limited knowledge and experience about the performance of M&E activities in the shorter and more flexible version of PIPA used by the STEPS centre, because researchers focussed only on the planning phase of the methods.

5.5 DIVERSITY APPROACH TO RESEARCH EVALUATION (DARE): MAPPING DIVERSITY AND MODIFICATIONS IN THE NETWORK STRUCTURE

DARE is a technique for research evaluation developed in collaboration by scholars from SPRU and INGENIO-CSIC/UPV (Bone et al., 2017), building on previous research the DARE team had conducted (Rafols and Meyer, 2009; Molas-Gallart and Tang, 2011; Rafols, 2014; Molas-Gallart et al., 2016). DARE is an approach for assessing research projects, programmes and organisations at the intersection between fundamental

knowledge and applied solutions (e.g. translational research), showing whether (and how) a specific research initiative (e.g. a centre, a project) fosters the creation of interactions across different stakeholders. The changes in such interactions are then used to assess the potential of the intervention for generating socially beneficial impacts. Therefore, DARE does not look at impacts, but at structures and the processes generating outputs and outcomes, in order to assess if such structures and processes are likely to lead to research impacts.

Specifically, DARE uses social network analysis to create maps and indicators of the research network's structure according to different dimensions: cognitive, social, organisational, institutional, and geographical. By analysing such networks DARE studies their diversity and how the intervention has influenced it. The underlying assumption is that, for research initiative to generate impact they need to engage and link diverse communities. Formal, quantitative network analysis is complemented by a qualitative analysis of interviews with project participants and project documentation, in order to build a narrative of the specific intervention has helped bridging the diverse communities involved in it.

DARE builds on several underlying assumptions:

1. The fundamental knowledge on which specific applications are built needs to be adapted to the specific conditions in which it will be used. Such adaptation requires detailed knowledge of both the fundamental knowledge and of the context of application. Further, such application will require the combination of diverse types of knowledge, originating from a variety of sources including the practical contextual knowledge of final users and beneficiaries.
2. The processes by which these different types of knowledge are combined involve the collaboration of many different stakeholders, from different, even distant, organisations. Therefore, the problem of connecting scientific knowledge for solutions to practical problems can be interpreted as the challenge of orchestrating diverse social groups, used to operating in different ways, to share and combine the information and knowledge required to develop new applications.

3. Research impacts or the translation of fundamental knowledge to applied solutions often takes a very long time. Therefore, it is difficult to use impact assessment for providing monitoring or formative evaluation in the short- and mid-term.

The main advantages of DARE are that it recognises the importance of diversity among the stakeholders within a research initiative, and of the ways in which a specific initiative helps bridge and connect such diversity. It defines and operationalises the notions of diversity and coherence across different relevant domains: in terms of their cognitive background, position in the organisation, social networks, institutional norms and geographical locations. Specifically, DARE differs from mainstream research evaluation and impact assessment practices in that it recognises that research outcomes and impacts cannot be attributed only to the research efforts being assessed. Therefore, from a policy perspective, DARE aims to assess the diversity of research structures and processes as a key factor for the generation of applicable knowledge. By contrast, one important limitation of DARE is that it is time and resource-consuming. It requires extensive data, some of which can only be collected through interviews.

DARE has been explicitly developed for research settings, although it aims to assess the potential for applicability of the research outcomes. There are not direct applications to innovation projects, although this should be feasible. If DARE was to be applied to the assessment of innovation activities, it would be necessary to consider the extent to which stakeholder diversity would be beneficial for innovation outcomes and impacts, in the specific context under analysis.

DARE has been designed as a 'formative' evaluation tool, and therefore its use is consistent with TI policy evaluation. One of DARE's strengths is its potential use in real time monitoring and, therefore to inform, the adaptation of policy interventions as they are being implemented. However, one of the main concerns about the applicability of DARE to the evaluation of TI experiments lies in the conceptualisation of dimensions: whether the dimensions of proximity identified by DARE are relevant for the objectives pursued by TIP.

5.6 IDENTIFYING RELEVANT TECHNIQUES: SOME CONCLUDING CONSIDERATIONS

Some of the techniques reviewed in this section come from the field of research evaluation, while others have their roots in the assessment of international development projects and programmes. Additionally, some of them emerged as techniques for planning, ex-ante assessment or supporting decision-making, while others have been created with the purpose of monitoring or assessing ex-post outcomes and impacts. Yet, our assessment is that they can all be adapted to be used as tools in TIP evaluation, although each would need to be further elaborated and their adequacy be ultimately tested empirically.

All the reviewed techniques can be useful for a formative evaluation, although some have already been designed as formative evaluation tools. For instance, PIPA aims to follow up the whole process of implementation in a participatory way, while DARE aims to track the transformation in the network structure in order to understand and learn from these transformation and related research processes. By contrast, PA initially emerged for quantitative assessment of risk and efficiency and needs complementary tools and modifications in order to be used in formative evaluation. MCM – despite aiming to stimulate discussion, reflexivity and participation – needs to be adapted to an ongoing and formative scope for assessment because it originated as a tool for ex-ante assessment with no follow-ups. In this sense, OM occupies an intermediate position: it is suitable for a formative evaluation but the mechanisms through which learning and improvements are fed back into the policy and into the evaluation process itself need to be further specified.



REFERENCES



- Alvarez, S., Douthwaite, B., Thiele, G., Mackay, R., Córdoba, D., Tehelen, K. (2010). Participatory impact pathways analysis: a practical method for project planning and evaluation. *Development in Practice*, 20(8), 946-958.
- Arnold, E. (2004). Evaluating research and innovation policy: a systems world needs systems evaluations. *Research Evaluation*, 13 (1), 3–17.
- Blamey, A., Mackenzie, M. (2007). Theories of change and realistic evaluation peas in a pod or apples and oranges? *Evaluation*, 13(4), 439–455.
- Boekholt, P. Arnold, E., Giarracca, F., Ploeg, M. (2014). Evaluation Reference Model. For TAFTIE's Taskforce Benchmarking Impact, Effectiveness and Efficiency of Innovation Instruments. Brighton: Technopolis Group
- Bone, F., Hopkins, M.M., Rafols, I., Tang, P., Molas-Gallart, J., Davey, G., Carr, A.M. (2017). User Guide to DARE. A Diversity Approach for Research Evaluation. Available at: http://www.sussex.ac.uk/spru/research/dare/data/DARE_User_Guide_2017.pdf
- Boni, A., Molas-Gallart, J., Belda, S. (2018). Towards a framework for the evaluation of Transformative Innovation Policies. Paper presented at the EU-SPRI Conference, Paris, 6-8 June.
- Bos, B., Grin, J. (2008). "Doing" reflexive modernisation in pig husbandry: the hard work of changing the course of a river. *Science, Technology, & Human Values*, 33(4), 480-507.
- Burgess, J., Stirling, A., Clark, J., Davies, G., Eames, M., Staley, K., Williamson, S. (2007). Deliberative mapping: a novel analytic-deliberative methodology to support contested science-policy decisions. *Public Understanding of Science*, 16(3), 299-322.
- Burt, R. S. (2009). *Structural holes: The social structure of competition*. Harvard university press.
- Caffrey, L., & Munro, E. (2017). A systems approach to policy evaluation. *Evaluation*, 23(4), 463-478.
- Chataway, J., Daniels, C., Kanger, L., Schot, J., Steinmueller, E. (2017). *Developing and Enacting Transformative Innovation Policy*. Presented at the 8th International Sustainability Transitions Conference, Gothenburg, Sweden, 18 – 21 June.
- Chen H-T. (1990). *Theory-Driven Evaluations*. California: Sage.
- Coburn, J., Stirling, A. (2016). *Multicriteria Mapping Manual*. Available at: <http://www.multicriteriamapping.com/>
- Cohen, E., Franco, R. (1992). *Rationalising social policy: evaluation and viability*. Santiago: Cepal.
- Colciencias (2018), *Green Book 2030: National Science and Innovation Policy for Sustainable Development*, available at http://libroverde2030.gov.co/wp-content/uploads/2018/10/GreenBook2030_8oct-web2.pdf
- Daedlow, K., A. Podhora, M. Winkelmann, J. Kopfmüller, R. Walz and K. Helming (2016). "Socially responsible research processes for sustainability transformation: an integrated assessment framework." *Current Opinion in Environmental Sustainability* 23: 1-11.
- Dahler-Larsen, P. (2001). From programme theory to constructivism on tragic, magic and competing programmes. *Evaluation*, 7(3), 331–349.
- Davis, P. K., Shaver, R. D., Beck, J. (2008). *Portfolio-analysis methods for assessing capability options*. Santa Monica: Rand Corporation.
- Douthwaite, B., Alvarez, S., Thiele, G., Mackay, R. (2008). *Participatory Impact Pathways Analysis: A practical method for project planning and evaluation*. ILAC Brief.
- Earl, S., Carden, F., Smutylo, T. (2001). *Outcome mapping: Building learning and reflection into development programmes*. Ottawa, ON, CA: IDRC.
- Edler, J., Ebersberger, B., Lo, V. (2008). Improving policy understanding by means of secondary analyses of policy evaluation. *Research evaluation*, 17(3), 175-186.

- Ely, A., Oxley, N. (2014). STEPS centre research: Our approach to impact. STEPS Working Paper 60, Brighton: STEPS Centre.
- Geels, F.W., 2002. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, 31, 1257–1274.
- Geels, F. (2007). Feelings of Discontent and the Promise of Middle Range Theory for STS: Examples from Technology Dynamics. *Science Technology Human Values* (32), 627-651. Geels, F.W., 2010. Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Res. Policy* 39, 495–510. doi:10.1016/j.respol.2010.01.022
- Geels, F. W., 2011. The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental innovation and societal transitions*, 1(1), 24-40. <https://doi.org/10.1016/j.eist.2011.02.002>
- Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research policy*, 36(3), 399-417.
- Hausmann, R., Rodrik, D., Velasco, A. (2005). Growth Diagnostics. Center for International Development, John F. Kennedy School of Government, Harvard University.
- Heiskanen E., Matschoss K. (2018). Evaluating Climate Governance Experiments. Participants' Perspectives on Low-Carbon Experiments in Finland. In B. Turnheim, P. Kivimaa, F. Berkhout (Eds.), *Innovating climate governance: Moving beyond experiments*, Chap. 10: 182-200. Cambridge: Cambridge University Press.
- Hertting, N., Vedung, E. (2012). Purposes and criteria in network governance evaluation: How far does standard evaluation vocabulary takes us? *Evaluation*, 18(1), 27–46.
- Hildén, M., Jordan, A., Rayner, T. (2014). Climate policy innovation: Developing an evaluation perspective. *Environmental Politics*, 23(5), 1-22.
- Holzer, J., Carmon, N., & Orenstein, D. (2018). A methodology for evaluating transdisciplinary research on coupled socio-ecological systems. *Ecological Indicators*, 85, 808-819.
- Janssen, J. M. (2019). What bangs for your buck? Assessing the design and impact of Dutch transformative policy. *Technological Forecasting & Social Change*, 138, 78–94.
- Kivimaa, P., Kangas, H. L., Lazarevic, D. (2017). Client-oriented evaluation of 'creative destruction' in policy mixes: Finnish policies on building energy efficiency transition. *Energy Research & Social Science*, 33, 115-127.
- Ligero Lasa, J.A. (2011). Dos métodos de evaluación: criterios y teoría del programmea. Documento de Trabajo, Serie CECOD, 15/2011. Madrid: CEU Ediciones.
- Luederitz, C., N. Schöpke, A. Wiek, D. J. Lang, M. Bergmann, J. J. Bos, S. Burch, A. Davies, J. Evans, A. König, M. A. Farrelly, N. Forrest, N. Frantzeskaki, R. B. Gibson, B. Kay, D. Loorbach, K. McCormick, O. Parodi, F. Rauschmayer, U. Schneidewind, M. Stauffacher, F. Stelzer, G. Trencher, J. Venjakob, P. J. Vergragt, H. von Wehrden and F. R. Westley (2017). "Learning through evaluation – A tentative evaluative scheme for sustainability transition experiments." *Journal of Cleaner Production* 169: 61-76.
- Magro, E., Wilson, J. (2013). Complex innovation policy systems: Towards an evaluation mix. *Research Policy*, 42(9), 1647-1656.
- Marburger, J.H. (2005). Wanted: better benchmarks. *Science*, 308(5725), 1087.
- May, C. (2006). A rational model for assessing and evaluating complex interventions in health care. *BMC Health Services Research*, 6, 86.
- Molas-Gallart, J. (2015). "Research evaluation and the assessment of public value." *Arts & Humanities in Higher Education* 14(1): 116-126.

- Molas-Gallart, J., D'Este, P., Llopis, O., Rafols, I. (2015). Towards an alternative framework for the evaluation of translational research initiatives. *Research Evaluation*, 25(3), 235-243.
- Molas-Gallart, J., Tang, P. (2011). Tracing “productive interactions” to identify social impacts: an example from the social sciences. *Research Evaluation*, 20(3), 219–226.
- Narberhaus, M., L. Hopkins and J. Backhaus (2016). Theories of change and the vision of the Great Transition. Results and conclusions from an initial exploration at and around the Smart CSOx Workshop in October 2016, SMART CSOs Lab.
- Oros, C. J., Doan, H. M., Adoum, D. D., MacDonald, R. C. (2007). Portfolio review expert panel (PREP) process: a tool for accountability and strategic planning to meet research goals. *Research Evaluation*, 16(3), 157-167.
- Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. London: SAGE.
- Pawson, R., Tilley, N. (1997). *Realistic Evaluation*. London, SAGE Publications.
- Popper, R., Velasco, G., Popper, M. (2017). CASI-F: Common Framework for the Assessment and Management of Sustainable Innovation, CASI project report. Deliverable 6.2.
- Rafols, I. (2014) Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence, in Ding, Y., Rousseau, R., and Dietmar, W. (Eds.), *Measuring Scholarly Impact: Methods and Practice*, Springer, 169-190.
- Rafols, I., Meyer, M. (2009). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.
- Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. *Evaluation*, 20(1), 75–95.
- Rodrik, D. (2014). Green industrial policy. *Oxford Review of Economic Policy*, 30 (3), 469–491.
- Rogers, P.J., Hacsí, T.A., Petrosino, A., Huebner, T.A (2000). Programme Theory Evaluation: Practice, Promise, and Problems, in Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (Eds.) *Programme Theory in Evaluation: Challenges and Opportunities. New Directions for Evaluations*, 87.
- Rotmans, J., Kemp, R., Van Asselt, M. (2001). More evolution than revolution: transition management in public policy. *Foresight*, 3(1), 15-31.
- Russo, M., & Rossi, F. (2009). Cooperation Networks and Innovation: A Complex Systems Perspective to the Analysis and Evaluation of a Regional Innovation Policy Programme. *Evaluation*, 15(1), 75-99.
- Schot, J., Daniels, C., Torrens, J., Bloomfield, G. (2017). *Developing a Shared Understanding of Transformative Innovation Policy*. Brighton: TIPC Research Brief, 2017-01.
- 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.
- Shapira, P., Kuhlmann, S. (2003). *Learning from science and technology policy evaluation : Experiences from the United States and Europe*. Cheltenham: Edward Elgar.
- Smith, A., Voß, J. P., Grin, J. (2010). Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges. *Research policy*, 39(4), 435-448.
- Smutylo, T. (2005). Outcome mapping: A method for tracking behavioural changes in development programmes. ILAC Brief 7.
- Srivastava, C. V., Towery, N. D., Zuckerman, B. (2007). Challenges and opportunities for research portfolio analysis, management, and evaluation. *Research Evaluation*, 16(3), 152-156.
- Stake, R. E. (2006). *Evaluación comprensiva y evaluación basada en estándares (Vol. 10)*. Barcelona: Graó.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707-719.

Stirling, A. (2008). "Opening up" and "closing down" power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values*, 33(2), 262-294.

Taanman, M. (2014). Looking for Transitions. Monitoring approach for sustainable transition programmes. Doctor, Erasmus Universiteit Rotterdam.

Turnheim, B., Kivimaa, P., Berkhout, F. (2018). Experiments and beyond: An emerging agenda for climate governance innovation. In B. Turnheim, P. Kivimaa, F. Berkhout (Eds.), *Innovating climate governance: Moving beyond experiments*, Chap. 12: 216-241. Cambridge: Cambridge University Press.

Van den Bosch, S. (2010). Transition Experiments, experiments that can contribute to sustainability transitions. FSW. Rotterdam, Erasmus University. PhD Dissertation. Wallace, M. L., Rafols, I. (2015). Research portfolio analysis in science policy: moving from financial returns to societal benefits. *Minerva*, 53(2), 89-115.

Weiss, C. H. (1998). *Evaluation. Methods for studying programmes and policies*. London: Prentice Hall.

Weiss, C. H. (1997). How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review* 21(4): 501-524.

Wiek, A., Kay, B., Forrest, N., 2015. Worth the Trouble?! An evaluative scheme for urban sustainability transition labs (USTL) and an application to the USTL in Phoenix, Arizona. In: Frantzeskaki, N., Coenen, L., Castan Broto, V., Loorbach, D. (Eds.), *Urban Sustainability Transitions*. Routledge. Young, J., Mendizabal, E. (2009). Helping researchers become policy entrepreneurs: how to develop engagement strategies for evidence-based policy-making. Briefing Paper, 53. London: Overseas Development Institute.

TO UPDATE



AUTHORS

Alejandra Boni (1)

Sandro Giachi (2)

Jordi Molas-Gallart (1)

INGENIO (CSIC-Universitat Politècnica de València)

SPRU

University of Sussex



CONTACT

Science Policy Research Unit,
University of Sussex Business School
University of Sussex, Brighton, BN1 9SL United Kingdom



EMAIL

TIPC@sussex.ac.uk



WEBSITE

www.tipconsortium.net



TWITTER

@TIPConsortium

