Observatory of
Public Sector Innovation

Evaluating Public Sector Innovation
Support or hindrance to innovation?

ALPHA VERSION: FOR DISCUSSION AND COMMENT

The Observatory of Public Sector Innovation collects and analyses examples and shared experiences of public sector innovation to provide practical advice to countries on how to make innovation work.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671526.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Co-funded by the Horizon 2020
Framework Programme of the
European Union

Executive summary

Governments around the world are facing unprecedented change and public sector needs innovation more than ever. Yet, how can we be sure that innovation in the public sector is actually delivering positive change? How can we be certain that innovations are moving in the right directions, that there is innovation happening in the public sector at all? Innovation is not inherently good or bad, but by definition, it is something new to the context it is applied in; and that brings a value shift. What this value shift is or who wins or loses, is not specified. With major changes with increased digitalisation on the way and the need to respond to complex policy problems, there is need for critical evaluation of the effects innovations are delivering to the public sector and if the sector itself is conducive to innovation in the first place.

At the same time, there is plenty of evidence that evaluations can themselves negatively affect innovation processes by ramping up risk aversion and fear of failure, sometimes locking in action to established practices. Evaluations may rationalise something that is often unintended, unplanned, creative activity, messy and uncertain by its nature. The novelty of innovation makes learning from the past and, thus, traditional ex post evaluations difficult if not useless, but evaluation may contribute directly to sustainability of innovation in the public sector itself.

This report outlines some of these core contradictions and conflicts between innovation and evaluation, which are probably the cause why no comprehensive public sector innovation evaluation framework currently exists. And maybe it should not exist. It might be more important to skill established evaluators on innovation principles and contradictions, rather than creating approaches for a specific community; and challenge the assumptions of evaluation methodologies themselves. As said, innovation is needed in all field in government and invariably if will be evaluated by various parties. Furthermore, innovation evaluation should be a continuous process (it cannot be otherwise as its aims and approaches are subject to change) and more reflexive practises may make more sense for innovators for learning then pre-defined criteria. Hence, the report argues that it is important to distinguish why evaluations are undertaken in the first place – for accountability, legitimacy, learning etc. Different aims may mean different approaches.

Furthermore, different types of innovation – in line with the OPSI innovation facets model – may necessitate diverging approaches as well. Thus, the report gives a broad overview of different clusters of evaluation (performance and impact evaluation, collaborative and user-centred approaches, in project reflection, automatized evaluations, future oriented evaluations, systems based approaches, systematic reviews and meta-analysis and triangulation and mixed methods) that have relevance to innovation due to their wide application or direct usefulness. The interactions between different methods and innovation require further analysis and empirical testing, and the report hopes to raise pertinent questions to accompany the former.

Table of contents

Executive summary **2**

1. Introduction **6**

1.1. Evaluation surge in the public sector 6
1.2. Innovation lifecycle 7
1.3. The core questions connected to innovation evaluation 10

2. The role of evaluation in the policymaking process **14**

2.1. The role of evaluation 14
2.2. Evidence-informed policymaking and the institutionalisation of evaluation 22
2.3. How evaluation functions as part of learning and innovation processes? 27
2.4. The dark side of measurement and evaluations 32
2.4.1. Replicability over complexity and causality 32
2.4.2. Forgetting about time and place 33
2.4.3. Conflicting expectations 33
2.4.4. Formalism and goal displacement 33
2.4.5. Disincentives to responsibility and innovation 34
2.4.6. Unintended use and unanticipated effects 34
2.4.7. Evaluation complexity and opacity 35

3. Evaluating innovation in a changing context **36**

3.1. Can evaluations keep up with change? 36
3.2. Pushing past old dogmas 38
3.3. The rise of public value 39

4. Better evaluation to fit different types of innovation: multi-faceted nature of innovation **42**

4.1. Enhancement-oriented innovation 43
4.2. Mission-oriented innovation 43
4.3. Adaptive innovation 44
4.4. Anticipatory innovation 44

5. Tools, methods and approaches to support innovation evaluation **45**

5.1. Evaluation of performance and outcomes 45
5.1.1. Impact assessment 45
5.1.2. Economic evaluation 53
5.1.3. Experimental design 54
5.1.4. Audits 57
5.1.5. Logic modelling 57
5.1.6. Theory-based approaches 58
5.2. Collaborative and user-centric evaluation 60
5.2.1. Joint and participatory evaluation 62
5.2.2. Most significant change 62
5.3. In project reflection 63
5.3.1. Monitoring 63
5.3.2. Learning agendas 64
5.3.3. Rapid experimentation 65

5.3.4. Adaptive pathways 67
5.4. Automitised/Digitalised evaluation tools 67
5.4.1. Big data and text mining 67
5.4.2. Data dashboards and interactive data visualization 69
5.4.3. Citizen experts and P2P production of evidence 69
5.5. Future-oriented evaluation tools 70
5.5.1. Risk analysis 70
5.5.2. Strategic evaluation 70
5.5.3. Foresight evaluation 70
5.5.4. Horizon scanning 71
5.6. Systems-based approaches to evaluation 71
5.7. Systematic reviews and meta-evaluations 72
5.8. Triangulation and mixed methods approaches 74

6. Conclusions 76

7. Remaining issues 77

Bibliography 78

Annex A. Tools to support evaluation 95

Tables

Table 2.1. Different perspectives on evaluation 18
Table 2.2. Five forums of accountability 19
Table 2.3. Barriers connected to evidence-informed policymaking 26
Table 3.1. Value sets and categories 40
Table 5.1. Practical information for the use of the matrix 50
Table 5.2. Phases and key actions for the expert review 51
Table 5.3. Instrument for measuring the public innovation capacity 64
Table 5.4. Two Basic Approaches to Evaluate Public Sector Innovation 72
Table 5.5. Summary of challenges for policy mix evaluations 74

Figures

Figure 1.1. Innovation lifecycle 8
Figure 2.1. What Works Centres areas of activity 23
Figure 2.2. Number of institutions conducting evaluations in OECD Member countries 24
Figure 2.3. Embedding evaluation diagnostic tool 24
Figure 2.4. Evidence planning worksheet 26
Figure 2.5. Heuristic model of the FAL process 29
Figure 2.6. Structure and methods 31
Figure 3.1. Investment Dimensions of UNDP's Innovation Facility Fund Portfolio 37
Figure 4.1. Public sector innovation model 43
Figure 5.1. Ten steps of an outcome evidence process 47
Figure 5.2. Best application of QCA approaches 49
Figure 5.3. Basic design of randomised control trials 55
Figure 5.4. The realist evaluation cycle 60
Figure 5.5. Open innovation model of smart city service development. 61
Figure 5.6. Evaluation continuum. 62

Figure 5.7. NESTA's continuum of experimentation 66
Figure 5.8. Key concepts in adaptation pathways approaches 67
Figure 5.9. Determinants of public sector innovation. 72
Figure 5.10. What Works Centres systemic review methodology 73
Figure 5.11. Three governance modes for policy mix evaluation 75

Boxes

Box 2.1. OECD's definition of Evaluation 14
Box 2.2. Core questions of traditional evaluations 15
Box 2.3. Formative vs summative evaluations 16
Box 2.4. The 17 Step Utilization-Focused Evaluation Framework 21
Box 2.5. The rise of evidence-based policymaking in the UK 22
Box 2.6. NESTA's resources for evidence informed policymaking 25
Box 2.7. Tacit and explicit knowledge 28
Box 2.8. Types of learning 30
Box 2.9. Evaluation Guide to Public Sector Innovation in Denmark 31
Box 5.1. Evaluation matrix for innovative solutions in the active and healthy ageing (AHA) field 49
Box 5.2. Core assumptions of public value mapping (PVM) 52
Box 5.3. Evaluating to improve efficiency in New Zeeland 53
Box 5.4. The UK's Cross-Government Trial Advice Panel 56
Box 5.5. The role of National Audit Offices in public sector innovation 57
Box 5.6. Undertaking a Theories of Change Evaluation 58
Box 5.7. Capturing casual moments 61
Box 5.8. Urban Data Centres (the Netherlands) 69
Box 5.9. Future reviews in Finland 70
Box 5.10. The three orientations of systemic evaluation 71
Box 5.11. Examples of evidence clearinghouses 73

1. Introduction

1.1. Evaluation surge in the public sector

In the last four decades, in conjunction with the spread of New Public Management (NPM), the issue of performance became the central topic of public administration (Pollitt 2008; 2010; Van Dooren, Bouckaert and Halligan, 2010). The main claim of NPM based on public choice theory and other economics-inspired inputs was that the civil service as a whole exerted unbalanced influence on the size, structure and outputs of government and thus, had to be checked (see the overview of the influence of economic theories to public administration in Tõnurist and Bækgaard, 2018). This brought about an increase in auditing, monitoring and a variety of evaluation methods. Thus, the goal to undergo evaluation was usually connected to advancing transparency, apprising, sanctioning, and showing accountability – managing for results –, less so to learning (Ebrahim, 2005). The predominance of economic values such as the '3Es' – economy, efficiency and effectiveness (Glynn and Murphy, 1996) – and also the need to demonstrate value for money reinforced these trends. Based on these assumptions and narratives, complex patterns of influence, learning and effects of evaluation emerged and resulted in the rise of performance management and the specialisation and professionalization of evaluation in the public sector (Boyle and Lemaire, 1999). This was not necessarily bad, but it also divorced evaluation from the continuous practice of policymaking usually giving it an ex post or ex ante role in the policy cycle and handing the role and responsibility over to specialised auditors and evaluators; thus, cutting feedback loops between policy design and implementation.

Furthermore, variety of measurement system from balanced scorecards to different quality management models were introduced to the public sector (e.g., Hasan and Kerr, 2003; Sahay 2005). However, most of these have been severely criticised for being output-centred, overly simplistic and ignorant of unintended effects. It is very easy to fall prey to many negative side-effects of measurement – 'tunnel vision', myopia, misrepresentation and misinterpretation, gaming and ossification, etc. (Smith, 1995; 2005) – not to mention the ease of ignoring local circumstances and tacit knowledge connected to projects. Thus, with the rise of evaluation in the public sector, the creation of evidence started to institutionalise, but it also produced many negative side effects.

Public sector innovation, which as a topic came to the forefront invariably later, did not escape the influence of the NPM; one could even argue that it with values such as user centricity and entrepreneurship is the product of the latter paradigm (Bartlett and Dibben, 2002). Yet, decade later, the academic debate around innovation performance in the public sector is still fairly minor (Nasi, Cucciniello and Degara, 2018). To be sure, empirical research into public sector innovation has increased rapidly in the last decade (Bekkers, Tummers and Voorberg, 2013; Kattel et al. 2013; De Vries, Bekkers and Tummers, 2016). Some measurement efforts applying econometric methods and survey-based analysis have been utilized (e.g., Bugge, Mortensen and Bloch 2011; Hughes, Moore, and Kataria, 2011; Arundel and Huber, 2013; COI, 2015; Arundel, Bloch, and Ferguson, 2018). Most of these, however, do not evaluate the outcomes of 'innovation' projects, but put the focus on the inputs and enablers, processes and at most, outputs of innovation (Kattel et al. 2013). On the innovation project level, qualitative analysis and case studies have been the dominant form of evaluation; although, single innovation or 'object' for evaluation can be also used

in a survey-based format for evaluation (Bornis 2001; Bianchi, Marin and Zanfei, 2018; Pärna and von Tunzelmann 2007). Nevertheless, most research to date is still cross-sectional (Damanpour, Walker, and Avellaneda 2009; Walker 2014) rather than longitudinal and thus, the long-term effects of public sector innovation are yet to be robustly evaluated. Thus, issues of representativeness, prevalence and frequency of evaluation connected to innovation raise its head. Furthermore, most published results concentrate predominantly on success stories rather than a fair account of innovations whatever happened to them (Edgerton 2007; Kelman 2008).

Outside the quality of evaluation itself, the rise of the evaluation agenda in the public sector has had both negative and positive effects on innovation. On the one hand, it has the potential to legitimise change, create windows of opportunity for innovation if evaluation shows that current forms of intervention are not working. On the other, it can stifle and direct change by pre-defining what the solutions could be (by over specific recommendations) and thus, create a false sense of certainty about what works. Alternatively, fear of sanctions produce an environment of paralysis where risks are not taken or minimized at all cost (Brown and Osborne 2013; Flemig, Osborne and Kinder 2016), and uncertainty is ignored; thus, enforcing existing path-dependencies. Public sector innovation has a variety of barriers (Bekkers, Tummers and Voorberg, 2013; Cinar, Trott and Simms, 2018) that evaluation can (unintentionally) exasperate.

This report looks at all of the above from the specific angle of the innovation life cycle and puts the attention on innovation projects. The aim is not to solve all the issues outlined above, but to create a more holistic, multi-method approach to evaluation that takes into account the various goals of innovative activity. In doing so, Chapter 1 introduces the topic by briefly describing the process of innovation and the core questions evaluation brings to the table. Chapter 2 outlines the 'basics of evaluation' – why evaluation is useful, how to plan it, what capabilities and capacity public sector needs, how it can become institutionalised in organisations etc. Chapter 3 starts the discussion about what evaluation in the context of change means and what kind of issues and myths connected to innovation and evaluation have to be tackled to get to better outcomes. Chapter 4 introduces a fit for purpose model for innovation evaluation based on the OPSI innovation facets model. Chapter 5 gives an overview of the different tools, methods and approaches one can use to evaluate innovation, before putting them into the context of different innovation facets. Chapter 6 concludes the report and introduces additional topics to be aware of when evaluating innovation.

1.2. Innovation lifecycle

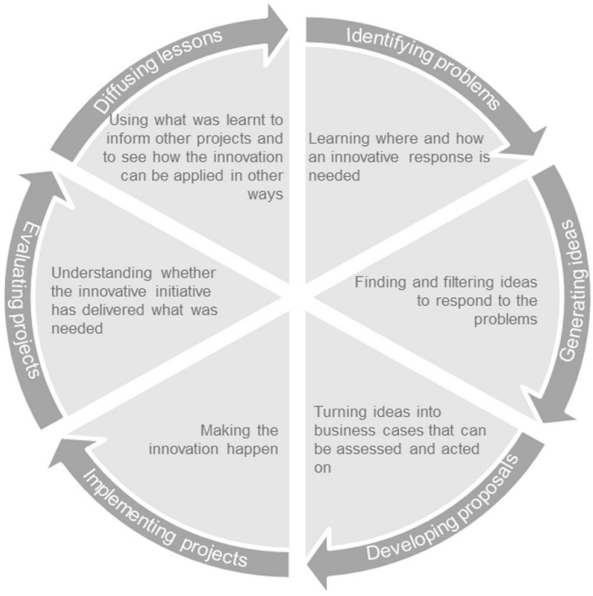
This report is part of a series of studies the OECD is conducting to better understand the innovation process. This series of studies, funded under the European Commission's Horizon 2020 program, aims to take stock and review what is known about public sector innovation, identify possible gaps in that knowledge, and to provide guidance about:

- The issues faced by innovators and organisations when trying to introduce novel initiatives or ways of thinking.
- What tools and methods are most appropriate at different stages of the innovation process and under what conditions?
- How a stronger innovation capability may fit with existing processes and initiatives?

As mentioned above, the number of empirical studies around public sector innovation is increasing. Nevertheless, there are there are fairly few systemic accounts about public sector innovation on the project level which look into the how new ideas get generated and developed into projects and how innovations get taken up by other units and organisations. To be sure, there is plenty of research on all these topics if not in the public sector literature then in the private sector context, but it compendium of evidence is fairly fragmented or specialized into specific streams (e.g., learning and innovation; accountability and risk) and often ignores the whole continuous process and political economy, where these innovations develop. Thus, this series of studies will contribute to a better understanding of how public sector organisations can effectively use the innovation process to get better outcomes, including by:

- Identifying problems and learning where and how an innovative response is needed.
- Generating and sourcing ideas to respond to those problems.
- Developing proposals that turn those ideas into business cases that can be assessed and acted upon.
- Implementing the innovation projects that proceed.
- Evaluating (and integrating) the outcomes of those innovation projects and whether the innovative initiative has delivered what was needed.
- Diffusing the lessons from those evaluations, and using those lessons to inform other projects and how other problems might be responded to.

Figure 1.1. Innovation lifecycle



Note: The lifecycle is not meant as a linear progression of different processes, but it depicts core elements of any innovation project. These processes can overlap each other in practise.
Source: OECD

The cyclical, but interconnected innovation process is shown in Figure 1.1. As with the general public policy cycle (see Howlett, Ramesh and Perl, 2009), this is a simplification of the subject matter to structure the analysis and it comes at a cost. The underlying connection to linear policymaking and Easton's input-output model even in stages perspective cannot be ignored (Wegrich and Jann, 2006). Innovation as a process in many cases drives on complexity – sometimes even chaos (Peters 1989; Bakhshi, Ireland and Gorod, 2016). Thus, while all of the studies in the series described below focus on one aspect of the innovation lifecycle, they invariably cover topics that are relevant to other parts of the lifecycle, among others evaluation. For example, generating ideas or developing proposals in the real world does not emerge from a blank slate. Even when innovation deals with discontinues change, the processes themselves are path-dependent and established paradigms of thought, legacy systems influence new ideas, not to mention the effect of rule-driven processes ingrained in the public sector (Bernier and Hafsi, 2007). New ideas are grounded in (or at least influenced by) previous experiences, their perceived success, learning – all potentially, results of evaluation – and the capacity to use this knowledge. Implementing innovation projects is often built on quick iteration, rapid testing and agile development (Mergel 2016; Tõnurist, Kattel and Lember, 2017) – all of the former require continuous evaluation (stop, go, rethink). Moreover, while most public sector innovation programmes nowadays emphasise diffusion, it is only useful if the ideas, processes and practices in question are actually effective, which means that the effects of innovation have to be evaluated, tested and verified to some degree beforehand (Word et al. 2011; Askim, Hjelmar and Pedersen 2018).

Nevertheless, the lifecycle approach helps to delve into different aspects of the innovation process and put on a particular lens of examination. Thus, the specific studies in this series are the following:

- The first study in this series, *What's the Problem? Learning to Identify and Understand the Need for Innovation*, outlined the importance of understanding the problems that require an innovative response, and how organisations can learn about them.
- The second study, *What's Possible? Generating Innovative Ideas*, discusses how to go about identifying options and considering which are likely to be the most promising, the ones that may have the best chance of becoming a reality.
- The third study, *What's Good? Developing, Testing and Assessing Proposals* discusses how to convert these ideas into business case justifications and proposals for action.
- The forth study, *How do we Make it Happen? Implementing Public Sector Innovation Projects* covers the "Implementing projects" stage of the innovation process.
- The fifth study, *Evaluating Public Sector Innovation. Support or hindrance to innovation?* discusses the contradictory role of evaluation in the innovation process.
- The series concludes with the sixth study, *Transferring and Adapting: Diffusion of Innovation Knowledge and Lessons*, which covers the diffusion and learning processes of public sector innovation.

1.3. The core questions connected to innovation evaluation

"*The Americans have need of the telephone, but we do not. We have plenty of messenger boys*." William Preece, British Post Office, 1876

Applied definitions of innovation in the public sector tend to be narrow and it is easy to underestimate the connected processes in terms of their complexity. Codifying and measuring innovation in many instances is very difficult, as innovation is inherently a tacit phenomenon (Word, Stream and Lukasiak, 2011). Venturing into the unknown and untried cannot be described beforehand and thus, ex post evaluations of implemented innovations are not necessarily predictors of innovations to come. If they are used in such a manner, organisations tend to lock themselves in and limit their potential for radical change. Paraphrasing Soren Kierkegaard's famous quote: innovations can only be understood backwards; but they have to be developed forward. This raises questions about the evaluability of innovation (the extent to which it can be evaluated in a reliable and credible fashion) and brings forth one of the most fundamental questions when talking about evaluation and innovation:

- *How to make evaluation useful for the innovation practise that it is not only about taking account of the past, but it allows for learning for the future?*

Usually evaluation is understood in the ex post setting – how public policy has performed in action. This is, however, not true in practise: the linear policy cycle leads us to presume that there is a notable 'end' to the policy process which kick-starts evaluation, yet, in real life new policy proposals are tested and developed while current solutions are still in use. Furthermore, new technologies allow for much faster policy evaluation – almost real time input into policymaking (e.g., big data analytics – Höchtl, Parycek, and Schöllhammer, 2016). Thus, evaluation is (and through the help of technology is even more so) a *continuous process*. With fast-based changes outside of the public sector governments have to be open to permanent reiteration, reassessment, and consideration. Thus, the question arises:

- *How to integrate evaluation throughout the innovation lifecycle?*

Innovation is not a uniform activity. There are different types of innovation (e.g., product, service, process, organisational, communication and possibly policy (Arundel, Bloch and Ferguson, 2018)) and they are at different maturity levels. Service innovations have – for many years – remained invisible (Fuglsang, 2010; Djellal, Gallouj and Miles, 2013) because they rely on a bricolage of bottom-up changes. Depended on the innovation maturity, its scale and scope, the connected risk and uncertainty levels can be very different, also time-horizons to show results. Some argue that the role of government is to invest when uncertainty is the highest (e.g., Mazzucato 2015). Innovations can add to existing solutions in an incremental way or radically challenge them entirely. Innovations can also be directed or undirected in nature (Edler and Boon, 2018). Thus, there is a need for different metrics and methods to evaluate innovation.

- *How to account for different types and forms of innovation in evaluation efforts?*

Innovation is not always a planned and intentional activity in relation to an understood problem (Koch et al. 2005; Fuglsang, 2010). It can be ad hoc activity based on posteriori knowledge (i.e. knowledge derived from experience) and tacit in nature requiring 'lived experience' rather than codified processes. This has been previously referred to as a "*posteriori recognition of innovation*" (Toivonen, Touminen and Brax, 2007). Moreover, the process itself can be gradual, where the more transformative change is cumulative and

happens over time. In another quite common situation, the need for innovation might be clear, but what this might look like is not – thus, a lot of innovation activity requires 'fuzzy-front ends' (Tate et al. 2018).

- *How to evaluate unplanned activity without clear goals or timeframes?*

Kattel and Mazzucato (2018) describe the effects of the reform waves of the 1980s and 1990s created a particularly strongly unified and specific vision of public sector performance. These focused on 'visible performance' of lower-level activities – the 'frog view' – and not on higher-level (e.g., cross-organizational) policy fields. The result of this view is a precision-target driven government support on the project level and ever more complex and mostly market failure-driven policy evaluation tools. The approach assumes that there are correct answers to specific projects (*Ibid*.). There, however, is not always one right answer for policy problem or one best way of organizing. Governance issues are not clearly bounded, they are intrinsically complex. The same applies to their solutions. Innovation may be about testing different solutions that fit in some situations and not in others.

- *How evaluations can account for different options for innovation?*

Innovation builds on heterogeneous resources and interactive processes. No organization alone has all the resources and solutions and the ability to connect and collaborate across sub-systems and disciplines through open innovation is important (Chesbrough, 2006; Von Hippel, 2006). While internal relations quality is connected to incremental innovation, then external relations quality has been linked to radical innovation (Obal, Kannan-Narasimhan and Ko, 2016). This is imperative now as citizens' expectations and technological shifts are driving radical change inside and outside of the public sector. Thus, recently there has been a wave of literature discussing collaborative innovation (Crosby, Hart, and Torfing, 2017; Bekkers and Tummers, 2018; Torfing, 2018). It is a way to not only get to know user preference better, but also crowd in resources. This is very much in line with the broader ideas of network governance and New Public Governance (Osborne, 2010) – governments do not function in isolation, in command and control manner anymore and nor does public sector innovation. However, that also means that governments alone do not produce the results of public sector innovation: openness, relationships and network structures will influence what kinds of innovations emerge. Furthermore, in a collaborative setting it is difficult to assign accountability to one actor and draw out actor-specific causal relationships between outcomes and inputs. This all complicates evaluation, but as it is a core feature of innovation, it should not be ignored. Thus, evaluation needs to take also a cross-disciplinary and cross-boundary perspective.

- How to evaluate cross-boundary effects of innovation?

Another issue connected to innovation evaluation is the fact that current feedback systems may produce the wrong kind of information and may kill off promising innovations. Managers and leaders, even when confronted with fast-paced change can become committed to 'strategic status quo' (Geletkanycz and Black, 2001). This does not happen because they are unqualified for their jobs, but because feedback from the system undermines the potential of new developments. This is the quintessential innovators dilemma (Christensen, 2013): value to innovation is often shaped as an S-curve and it takes many iterations to improve and perfect an innovative solution; meanwhile, incumbents and established organisations have large user bases (with high quality expectation) and large prospective revenues, with which innovations in early phases of development cannot compete. This means that feedback from the system is to stop pursuing projects outside of

the scope of interest of the current user base with small (initial) revenue potential. As such, organisational size has a direct positive effect on incremental innovation performance and the internal knowledge creation capability (Forés and Camisón, 2016). Before radical innovations are applied in practise and widely adopted, their effects are not clear in many cases (in terms of technology this has been described by the Collingridge's dilemma (Liebert and Schmidt, 2010)). Thus, the relationship between the established legacy system and innovations that challenges the former is difficult. Especially in the public sector, as there may not be any incentives to critically analyse incumbent instruments and policies (as the direct feedback from the market is lacking).

- *How can evaluation help counteract lock-in in organisations and not kill of transformative change?*

Not all innovation projects become the source of transformative change. Many fail. There is by definition a lot of uncertainty connected to the innovation process and that is normal. This is the nature of innovating and this uncertainty cannot be boiled down to calculable risks. What tends to happen in the public sector characterised by high accountability culture, is that failure is not reported and innovation projects either become 'too big to fail' or they are killed off early due to the fear of failure. Risk aversion and the fear of failure are one of the most often cited barriers to innovation in the public sector (Bloch and Bugge, 2013; De Vries, Bekkers and Tummers, 2016). There is no question that measurement influences human behaviour and governance structures. Excessive control can stifle innovation. Even under best circumstances, it is very difficult for government officials to communicate to the public that it is acceptable to spend public money on things that turn out to be failures (Pollitt, 2011). Policymakers and politicians alike are often harshly penalized both by accountability mechanisms and the media (Pollitt, 2011; Gilson et al., 2009). Governments have started to experiment with measure to give 'licence to fail' in connection to innovation (Tõnurist, 2018), but the efforts are slightly diminished due to the misconception that uncertainty is the same as risk and it can be either minimised or stabilised. Evaluation can here add pressure by enforcing accountability, but it can also facilitate learning from failure that otherwise does not seem to happen.

- *How can evaluation create room for experimentation and facilitate learning from failure?*

Finally yet importantly, one of the most difficult issues with public sector innovation is the lack of comparison between before and after or the adequate measure to account for change. In the private sector this is defined as competitive advantage in the true Schumpeterian (1934) meaning and can be measured in monetary terms. In public sector, it is almost impossible (Kattel et al. 2013). Yet, how can we evaluate innovation when we do not have a constant unit of measurement? In many cases, when ex post evaluations of innovation are attempted the baseline is simply missing and as processes, services, products etc. tend to change fundamentally through innovation, it becomes difficult to evaluate even costs connected to the process. Thus, evaluation tends to become activity-based (looking at most at inputs, through-puts or on occasion outputs), but not outcomes or effectiveness in a comparative sense. Yet, there is a difference between measuring impact of individual innovations and measuring organizations ability to generate innovations. For example, the attempts to measure public sector productivity or to link that to innovation have so far remained unsuccessful (Lau, Lonti and and Schultz, 2017; Dunleavy, 2017). Furthermore, innovation can also bring forth 'creative destruction' or 'competence destroying' innovation (Tushman and Anderson, 1986). As such, the impact of innovation on employment is not simple – no clear-cut diagnosis exists either theoretically or empirically

(Vivarelli 2007, 729). Innovation and technology can both create and destroy jobs; consequently, the question is more what type of jobs are created/destroyed and how does it affect the skill- and wage-structure of organisations. Thus, new technologies offer opportunities to improve economic efficiency and quality of life, but they also bring many uncertainties, unintended consequences and risks. Whose (competitive) advantage should prevail in evaluating innovation? Here innovation does not take on a normative, positive or negative role; it may benefit some, but may also disadvantage others.

- *How to describe the value-added of innovation without an adequate baseline comparison or a uniformly understood measure of value?*

Above described are some of the core issues linked to innovation and evaluation that the following report will try to address.

2. The role of evaluation in the policymaking process

"What gets measured, gets done." (Osborne and Gaebler, 1992)

2.1. The role of evaluation

In broadest terms, valuation is the assessments of the implementation, output, and outcome of government measures in order to effect deeper understanding and well-grounded policy decisions (Vedung, 2017). Compared to monitoring, appraisal and other forms of generating information, evaluation usually looks for causal relationships and is thus, capable of answering how and why outcomes came about (see OECD's definition of evaluation in Box 2.1).

Box 2.1. OECD's definition of Evaluation

Evaluation is the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.

Evaluation also refers to the process of determining the worth or significance of an activity, policy or program. An assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention.

Evaluation in some instances involves the definition of appropriate standards, the examination of performance against those standards, an assessment of actual and expected results and the identification of relevant lessons.

Source: OECD, 2011: 21-22.

The aim is to assist people and organisations to improve their plans, policies and practises on behalf of citizens (Weiss, 1999). Evaluation in the public sector serves many purposes: it can help to distinguish the worthwhile from the worthless, create confidence in the programs selected, legitimacy in public policy, but also hold government account. Thus, evaluation is useful to policymakers when they want to (*Ibid.*; Patton, 1975):

- Make better *sense of policy issues* (their scope, scale, frames of reference etc.) and *facilitate improvements* (e.g., choose between instruments).
- *Render judgements* by making sure they can trust policy outcomes (e.g., their effectiveness or identify unintended effects) or actions of actors (e.g., their ethics, morality) or information connected to the policy field itself.

- *Provide legitimacy* to the policy field (e.g., input, output and throughput legitimacy (Schmidt, 2013)) and demonstrate public value.
- Find evidence or theory that *supports their position* or be considered up to date.

Box 2.2. Core questions of traditional evaluations

Should it work? (Theory of change)

What is the underlying 'theory of change' which explains how the policy will make an impact? An understanding of the theory of change that underpins the project will ensure that we measure the things that really matter during the evaluation.

Can it work? (Process/Implementation evaluation)

How was the policy implemented? Has the policy been properly implemented? What were the challenges to implementation and how were they overcome?

Does it work? (Impact evaluation)

Many of our evaluations investigate the impact of the intervention.

Is it worth it? (Economic evaluation)

It is anticipated that, if successful, policys/interventions might receive a wider roll-out. It will therefore be important to consider whether such an approach is cost effective and cost-beneficial.

Source: Acquah, 2018.

Based on the usage of evaluation, it can be categorised also by (Weiss, 1999; see also the account in Van Acker, 2017):

- *direct usage* of the information in the decision-making process;
- *indirect usage*, where the evaluation does not immediately results in changes, but does contribute to the understanding of the problem, and;
- *symbolic us*e, which refers to the usage of results to comply with regulations, or for internal or external political motives.

Only symbolic use of evaluation is unlikely to contribute to the improvement and sustainability of the interventions, incl. innovations.

In its most basic categorisation evaluation can be either (1) formative or summative and; (2) ex ante or ex post (e.g., Klecun and Cornford, 2005). Box 2.2 outlines the difference between formative and summative evaluation. In reality, the difference between the two in practice are disappearing and more creative use cases of evaluation are emerging. The other core categorisation of evaluation (ex ante, ex post) is purely temporal. Ex ante evaluation analyses a chosen process, service, product, developed system or technology before it has been acquired, designed, constructed, or implemented – ex post evaluation looks at it after (Klecun and Cornford, 2005). Ex ante evaluations are predictive in nature: they evaluate and estimate the potential effect in the future. While one might think that all ex post

evaluations are summative, it is not always the case. A summative evaluation may be required on an ex ante basis (e.g., for continuation approval) and ex post evaluations may also have formative purposes (Venable, Pries-Heje and Baskerville, 2016). Ex ante, ex post or in real time, different aspects of government performance can be measured including inputs/resources, throughputs/processes, outputs, and outcomes/impacts (Kuhlmann, 2010).

Box 2.3. Formative vs summative evaluations

- *Formative evaluations* produce empirically based interpretations that provide a basis for successful action in improving the characteristics or performance of the evaluated. It is an ongoing evaluation that is not fixed but is still in the process of change. Formative evaluations focus on consequences – what is and what is not working well – and support the kinds of decisions that intend to improve the evaluated policy, program or interventions. In theory, formative evaluations do not make a final judgment on the relative merits of the program.
- *Summative evaluations* produce empirically based interpretations that provide a basis for creating shared meanings about the evaluated in the face of different contexts. It supports the selection of the evaluated for an application. Often, it tries to encapsulate all the evidence up to a given point and examine both the intended and unintended outcomes to judge the merits of a fixed, unchanging program as a finished product, relative to potential alternative programs. This for summative evaluation is final point of judgement.

Source: Scriven 1967; Wiliam and Black, 1996; Teras 2005; Venable, Pries-Heje and Baskerville, 2016; Figueredo et al. 2014.

Other classification of evaluation can be based on the approaches and techniques applied – e.g., qualitative or quantitative methods, subjective or objective techniques. These distinctions rely on the reasons *why*, *when* or *how* evaluation is undertaken or *what* is evaluated in the first place (Stufflebeam and Coryn, 2014). However, any examination evaluation starts with the question who is evaluating and why, the process (how) and the content (what) is actually being considered. Here Stufflebeam (2001) has proposed a comprehensive classification consisting 4 main categories and 22 alternative approaches. The four categories are:

1. pseudo-evaluations (motivated by political reasons; e.g., public relations studies and politically motivated studies);
2. questions and methods evaluation approaches or quasi-evaluation studies (geared towards answering a particular question of applying a specific method);
3. improvement/accountability-oriented approaches (concentrating on outcome variables and expansive in approach and methods used) and;
4. social agenda and advocacy (geared toward directly benefitting the community in which they are implemented).

As the above shows there are very different approaches to evaluation and its role. Table 2.1. outlines different perspectives on evaluation – critical, socio-technical, social constructivist and hermeneutic – and the different questions these pose for the evaluation process. Put together these highlight some critical recommendations that evaluation should strive towards (Klecun and Cornford, 2005):

- Evaluation should try to *critically expose historically created conditions* (including technological conditions) that limit people's lives, and consider how a system under study may alter them (in a positive or negative way) and bear in mind the accumulative power of technology and innovation.
- Evaluation should attempt to give *voice to all persons and groups that have interests in* or are potentially affected by the phenomenon being evaluated, or by the evaluation itself. Evaluation should be inclusive and if possible, participatory in nature.
- Evaluation process and content – *logics-in-use* (e.g. the construction of criteria and methods of evaluation) – cannot be separated from the critical principles and the situation under study. Approaches and methods used will invariably influence the types of outcomes the evaluation will reach. Furthermore, building on the hermeneutic tradition, one needs to keep in mind that critical interpretation is invariably based on the evaluators' experience and knowledge.
- Evaluation should not only be a snapshot in time, but cover implementation over time. To make sure to what extent to which they become embedded (or not) in work practices and institutionalised in organisations will all change over time. Evaluation should be *fundamentally reflexive*.

Table 2.1. Different perspectives on evaluation

Different perspectives	*Critical*	*Socio-technical*	*Social constructivism*	*Hermeneutic*
Main aspects	Evaluation as a political, nonobjective process. Different stakeholders may have conflicting (vested) interests and exercise unequal power. Technology is not neutral but is socially constructed, presenting opportunities for furthering the case of emancipation or detracting from it. Evaluation must be normative (guided by norms and values), and represent interests of all groups affected by the technology. It should be based on learning and dialogue (striving for ideal speech situation).	Evaluation as a political process. Different interested parties should have a chance to voice their opinions about the system and its potential effect. An 'optimal' solution can be arrived at (designed) through a rational process of negotiations. IS are a mixture of social and technical, and thus evaluation should be a socio-technical process itself. The distinction between social and technical is largely preserved.	Evaluation as a political, nonobjective process. Different relevant social groups may have different interests. For the technology (or evaluation) to succeed these groups must be 'enrolled' in the process, i.e. their interests must be engaged. Technology is socially constructed and evaluation is a part of this (contingent and messy) process. The social and technological are closely interwoven and cannot be artificially separated.	Evaluation as an informal, subjective and situated process. Meaning is situated, i.e. technology/IS are understood in a particular context (e.g. organisational) and within a canvass of our past experiences. Understanding (and thus evaluation) is an incremental, circular process, involving re-interpretations.
Questions about the evaluation process	What is the macro context of the evaluation? In what way macro trends (and their perceptions by different actors) influence the evaluation? Who sponsors the evaluation? What are its underlying (hidden) aims? Are the interests of different groups represented in the evaluation, and how? What are the power relations between different groups? How do the process of evaluation and the choice of measurements effect the evaluation outcome?	Who are the stakeholders? What are the goals and measures relevant to the business, employee and customer perspectives? Does the system help to achieve those goals? Are working lives reflected in the designing of technology?	What are the relevant social groups? What are their views about the system? How are they enrolled in the process of stabilising (and evaluating) technology? What organisational vision does the information system support/appear to support/or suppose to support in the eyes of different groups? In what organisational context is the evaluation being conducted? How is the information system constructed through evaluation?	What are people's daily experiences of the system? Is the information system embedded in their work practices? How do they perceive the system? How such perceptions are constructed? (E.g. What theories-in-use influence their perceptions; What are their past experiences of similar information systems?)
Primary focus	Society/organisation	Group/organisation	Group/organisation	Individual

Source: Klecun and Cornford, 2005.

After evaluation, the policy/proposal is either put in action, maintained or reconceptualised (Howlett, Ramesh and Perl, 2009) based on the questions the evaluation is asking. However, people and organisations can also be held account for the results of the programme or project. In general, accountability can be either formal mechanism or relationship (an obligation of an actor to explain and justify its conduct to a significant other) or it can be also a norm or a virtue (a norm to act in an accountable way similar to responsiveness, transparency, fairness or responsible behaviour) (Schillemans and Bovens, 2011). Accountability can serve two goals: it can either be the reason why an organisation will start to look for feedback information in the first place or it can be a valuable source of feedback information itself (Van Acker 2017). Accountability can bediscussed in different forums, where the actor has an obligation to explain and justify its conduct to the

forum, in which the forum can pass judgment, after which the actor may face consequences (Bovens, Schillemans and Hart, 2008) (Table 2.2). Accountability arrangements, if organised properly, should feed public managers on a regular basis with feedback information about their own organization that may merit further evaluations (Bovens, Schillemans and Hart, 2008). The mare existence of accountability forums can lead public organisations to seek evaluation to prevent critical accountability reports (Pollitt et al., 1999). At the same time, accountability can have missing, unexpected, contradictory or even perverse effects (e.g., accountability overload – Halachmi 2014). Thus, the connected processes are characterised by complexity and multiplicity (Bovens et al 2014; Byrkjeflot et al 2014).

Table 2.2. Five forums of accountability

	Forums
Political	**Citizens as voters** **Members of Parliament** • Plenary assembly • Parliamentary committees • Special committees (e.g. policy disasters) **Direct democratic tools** • Referenda, citizen surveys/juries, etc. **Political parties**
Judicial	Administrative tribunals Judicial courts
Administrative	**Internal audit** • Court of audit • Auditors General/audit office • **Inspections** • Visitations • **Advisory councils** • Ombudsmen • Commissioners • Professional peer review • Whistleblowers **External audit** • Regulatory and monitory bodies • Private auditing firms
Public	Mass media Organized civil society (formal) 'Ad hoc' action groups or individuals (informal)
Private	Shareholders/owners Consumers

Source: Willems and Van Dooren 2012.

Utilisation of results is one of the primary goals of evaluations. When it comes to learning from evaluations (not simply accountability relations) the stakeholders' willingness to consider the implications of the findings of evaluations is key. This is connected to many factors of how and why the evaluation was conducted in the first place (Posavac 2015). Based on the above discussion, when planning evaluations the following should be thought trough:

- What is the aim of the evaluation? The philosophy, focus and future uses of an evaluation should be agreed upon beforehand.

- Does the aim of the research require to go deep into the intricacies of the project and cover all contextual variables connected to the project/programme (contingency enhancing internal validity) or the aim is to generalise the results (external validity)?
- What is the unit of analysis? Will the evaluation be conducted on the project, programme or systemic level?
- What is the frequency of evaluation? Is it a one-time assessment or a continuous process (summative or formative evaluation)?
- Is there capacity to carry out the evaluation or how should the evaluator be selected and what kind of criteria should be used to insure privacy, capacity and transparency (no conflict of interest) within the process.
- What kind of data is available for the evaluation? Is the data collected in a standardized, internationally accepted format so that it can be replicated? Is it disseminated in a standardized format to improve consistency and interoperability? How often is the data updated? How far back does the data go? Are there forecasts of the data? Is the data disaggregated by gender, income levels, race, rural/urban, nationality and age or other variable of interest?
- How are the evaluation criteria elected? Has the evaluated agreed to the criteria?
- How will the analysis be conducted to assure methodological rigor in analysis, its credibility, fairness and transparency? Key concepts here are internal validity, external validity, statistical conclusion validity, and construct validity.[1] To assure internal validity the selection bias should be minimised. For external validity randomized experiments on representative samples are considered the best strategy (Figueredo et al. 2014).
- How will feedback from evaluation be organised? Feedback here denotes the transmission of evaluation findings to parties for whom it is relevant and useful so as to facilitate learning. This may involve the collection and dissemination of findings, conclusions, recommendations and lessons from experience.
- How will the evaluation results be communicated effectively? Communication is a multidimensional, interactive process and this should be reflected in the evaluation reports, because if the information in them is not understood then it cannot be utilised (Posavac 2015).
- What should be take into account on the uptake of evaluation findings and support of decision making? Lack of institutionalization of evaluations, a lack of priority given to evaluations, a lack of evaluation capacity, and low quality of the evaluations and dissemination are the most common factors why evaluations are not used (Jüngen 2013). Factors that are positively correlated with evaluation impact are the relevancy of the content, its quality and credibility of the evaluation, the type of users, the user orientation and, most notably, the interaction with

[1] Internal validity denotes the establishment of a causal relationship between two variables such as treatment and outcome. External validity enables to support the generalization of results beyond a specific study. Statistical conclusion validity refers to applying statistical techniques appropriately to a given problem. Construct validity but specifically consists of assessing and understanding program components and outcomes accurately. (See further Cook, Campbell and Shadish, 2002)

stakeholders in the evaluation process (ibid.). Other factors that have been found to influence the usage of evaluation are communication quality, timing of the evaluation (aligning reporting with stakeholder decision-making cycles), personal characteristic of the (intended) user, the receptiveness of the users to the evaluation, and finally the involvement of evaluation stakeholders in the process (Johnson et al. 2009).

Some of these questions are also described also in the utilisation-focused evaluation framework (Box 2.4).

Box 2.4. The 17 Step Utilization-Focused Evaluation Framework

Utilization-Focused Evaluation (UFE), developed by Michael Quinn Patton, is an approach based on the principle that an evaluation should be useful to its intended users and should be judged based on the former. Consequently, evaluations should be planned and conducted in ways that enhance the likely utilisation of both the findings and of the process itself to inform decisions and improve performance. The two essential elements of the framework are that first, primary intended users of the evaluation must be clearly identified and personally engaged at the beginning of the evaluation process to ensure that their primary intended uses can be identified. Second, evaluators must ensure that these intended uses of the evaluation by the primary intended users guide all other decisions that are made about the evaluation process. The idea is that intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings (Patton, 2008). The frameworks checklist goes as follows:

1. Assess and build program and organizational readiness for utilization-focused evaluation
2. Assess and enhance evaluator readiness and competence to undertake a utilization-focused evaluation
3. Identify, organize, and engage primary intended users: the personal factor
4. Situation analysis conducted jointly with primary intended users
5. Identify and prioritize primary intended uses by determining priority purposes
6. Consider and build in process uses if and as appropriate
7. Focus priority evaluation questions
8. Check that fundamental areas for evaluation inquiry are being adequately addressed: implementation, outcomes, and attribution questions
9. Determine what intervention model or theory of change is being evaluated
10. Negotiate appropriate methods to generate credible findings that support intended use by intended users
11. Make sure intended users understand potential methods controversies and their implications
12. Simulate use of findings: evaluation's equivalent of a dress rehearsal
13. Gather data with ongoing attention to use

14. Organize and present the data for interpretation and use by primary intended users: analysis, interpretation, judgment, and recommendations
15. Prepare an evaluation report to facilitate use and disseminate significant findings to expand influence
16. Follow up with primary intended users to facilitate and enhance use
17. Meta-evaluation of use: be accountable, learn, and improve

Source: Patton and Horton, 2008 and https://www.betterevaluation.org/en/plan/approach/utilization_focused_evaluation

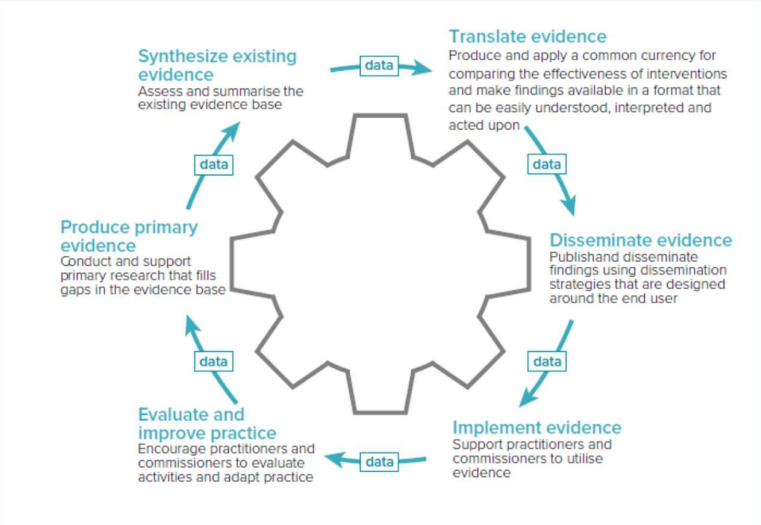
2.2. Evidence-informed policymaking and the institutionalisation of evaluation

The last chapter ended with suggestions of key questions connected to planning effective evaluations, but this is just one side of the coin – there also has to be demand for evaluation and evidence in the first place. While there is a large body of literature on program evaluation, relatively little attention has been given to how evidence is used within public bureaucracies in the policy development work (Head 2016). Yet, the topic of "evidence-based" or "evidence-informed" decision making gets increasingly more attention (see the UK experience in Box 2.6), especially in the era of 'post truth' and fake information. Yet, decision-making cannot be fully rationalised and room for constructive and imaginative judgements of practitioners who are implementing the decisions is needed. Thus, currently the consensus is that decisions and processes can be 'evidence informed' (Nevo and Slonim-Nevo, 2011). For the latter, policymakers need to access and use evidence to inform policy, have the capacity and incentives to do so.

Box 2.5. The rise of evidence-based policymaking in the UK

The UK has also been a leader in the development of evidence-based policymaking. Nearly 20 years ago, the Cabinet Office produced "The Modernising Government White Paper," which promised changes to policy making to ensure that policies are strategic, outcome focused, collaborative, inclusive, flexible, innovative and robust. This resulted in the creation of a network of "What Works Centres" in the UK. The centres reflects a belief that the provision of high-quality evidence can improve public policy decisions. Currently, the network now consists of 10 independent What Works Centres. These centres have pioneered new ways of increasing the supply of evidence in areas such as policing, education, local economic growth, and health and social care. By beginning of 2018, the centres had collectively produced more than 280 evidence reviews and commissioned or supported over 160 trials. The centres assess the existing evidence base, try to address evidence gaps and try to increase the reach of evidence by putting it into more user-friendly format.

Figure 2.1. What Works Centres areas of activity

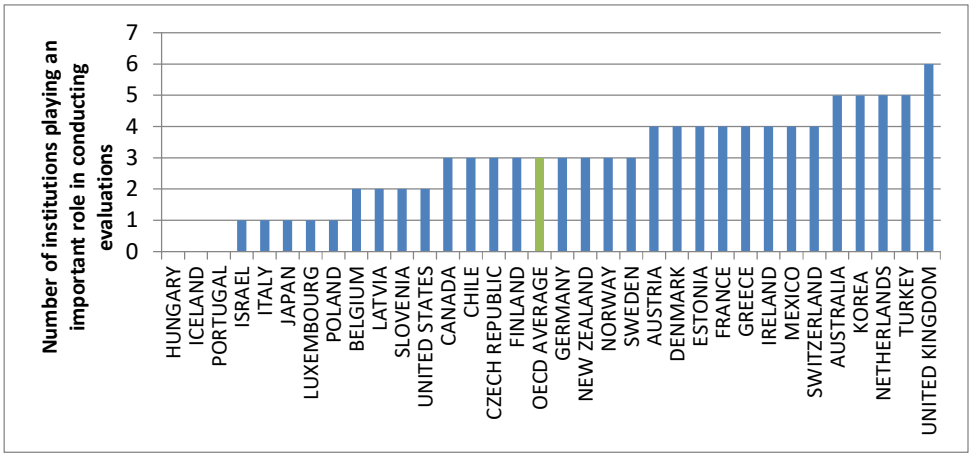


Note: What Works Centres have adopted the approach from the Digital and Trustworthy Evidence Ecosystem produced by MAGIC, 2016.
Source: The What Works Network 5 years on, 2018.

Evidence-informed policymaking is increasingly institutionalising in governments and more and more public organisations carry out evaluations (Figure 2.3). More and more governments apply knowledge management systems, a set of processes and practices in organizations by which knowledge is recognized, acquired, captured, codified, recorded, stored, aggregated, communicated, shared, transferred, converted, retrieved and reassessed Gilson, Dunleavy and Tinkler, 2009; Hislop, Bosua, and Helms, 2018). However, knowledge management systems in the public sector are fairly nascent (Al Ahbabi 2018). Furthermore, institutionalisation of evaluation does not only support capacity development in public sector, but also outside of government agencies as many evaluations are outsourced to various evaluation professionals, consultancies, think thanks and research centres. Furthermore, the push to provide evidence of impact and outcomes is also transferred to public sector partners, whose work the sector funds. Thus, also, the third sector's evaluation capacity has had to increase and many resources to support the former have sprung up (see for example Inspiring Impact[2] repository of resources).

[2] https://www.inspiringimpact.org/resource-hub/

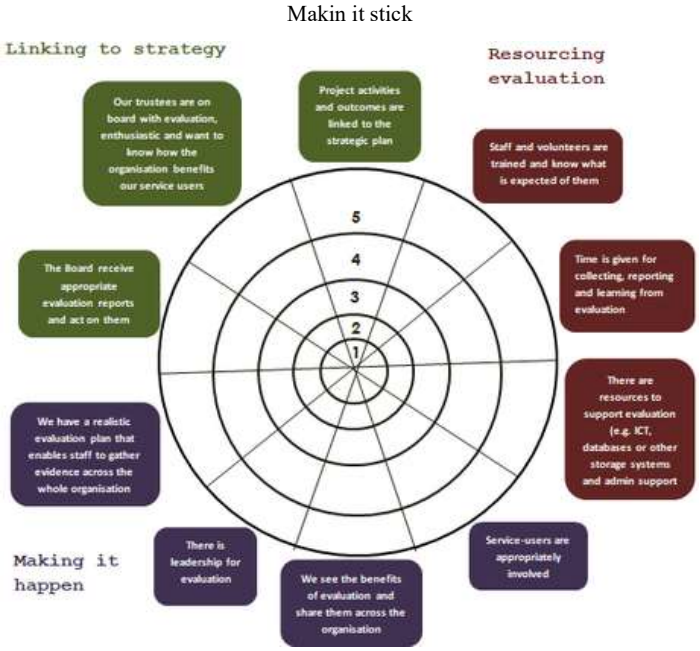
Figure 2.2. Number of institutions conducting evaluations in OECD Member countries



Source: OECD Performance Budgeting Practices and Procedures database.

While countries are progressively pursuing the institutionalising of evaluation (see Jacob et al., 2015; Pattyn, 2017), few normative claims exist on how to embed evaluation in countries' governance architecture (Jacob et al. 2015). Factors such as the political system, public administration cultures, and the rationale for evaluation all shape the development and characteristics of evaluation cultures. Nevertheless, there are still tools to help embed evaluations into the broader knowledge management system, but also in the broader institutional setup of policymaking and service delivery (see figure 2.4).

Figure 2.3. Embedding evaluation diagnostic tool



Note: The tool is designed to help organisations look at all areas of their organisation and to see how far evaluation is embedded into their systems.

Source: Evaluation Support Scotland (2018) Making it stick - A guide to embedding evaluation. http://evaluationsupportscotland.org.uk/resources/411/

The professionalization of evaluation has been supported by two agendas: the accountability agenda concerned with efficient and effective use of public money; and the policy effectiveness and innovation agenda (Head 2016). The first is usually liked to a system of key performance indicators and variety of accountability forums from independent auditing to public scrutiny are used to keep governments at check (Van Dooren, Bouckaert and Halligan, 2015). The second, however, surpasses the efficiency, fine-tuning and reliability concerns and is ties to finding the most effective methods to achieve positive outcomes taking into account the diverse context (Osborne and Brown 2013). The latter is of course more difficult. Nevertheless, many resources are available help to policymakers carry out evaluations and use evidence (e.g., International Atlas of Evaluation – Jacobs, Speer, Furubo, 2015; Box 2.6; see also Annex A for the selection of tools).

Box 2.6. NESTA's resources for evidence informed policymaking

NESTA published a range of reports on how governments, politicians, charities, practitioners and others are using evidence to inform decision-making, and uncover the best approaches, and create practical guides to help understand and implement the ideas that will work best. These include the Using Research Evidence practice guide; the "Using Evidence: What Works" report which is an introduction to the findings of The Science of Using Science, which studied how decision-makers were using research evidence (and how to enable this, including appointing 'evidence champions' to promote the use of evidence in their own organisations). NESTA has also produced a number of in-depth reviews with case studies, examples and ideas, such as a report exploring the Scottish Approach to Evidence, and a paper that looks at how to develop Better Public Services Through Experimental Government. Below is the example of the evidence planning toolkit.

Evidence planning toolkit

How can it help me?

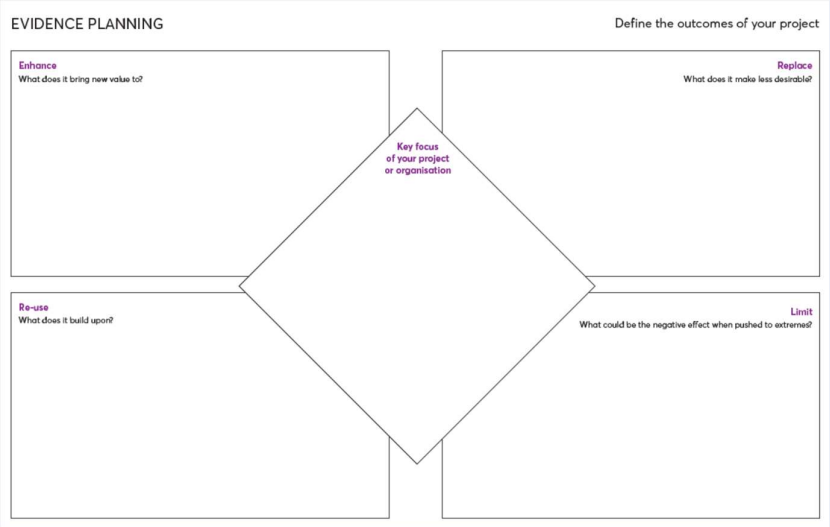
- Evidence planning helps you to communicate and refine your goal and build an evidence-based case for your planned impact.
- This worksheet provides prompts to help you consider how your work will affect other people and organisations. It also helps to highlight any potential issues early on.

How do I use it?

- Work from the centre outwards (Figure 2.1) by starting with the focus of your work. Then respond to the questions to reflect on the impact your work could have.

Consider the following groups when asking the questions: yourself (the impact on you personally); your stakeholders; your sector and the broader world.

Figure 2.4. Evidence planning worksheet



Note: The tool is inspired by Nesta (2009) Worksheet 2b, Evidence Modelling, Creative Enterprise Toolkit.
Source: https://www.nesta.org.uk/toolkit/evidence-planning/

Yet there are still barriers to use evidence of policymakers (Oliver et al. 2014) connected to collaboration, organisation and resources, research characteristics, characteristics of both policymakers and policy itself (see also table 2.3). Little is known about the skills and capacities and practises of policymakers themselves (how they design policies and review evidence, what sources they trust, how they process feedback etc.) (Hall and Jennings 2010; Head 2013; Head et al. 2014; Jennings and Hall 2012; Head 2016). Hence, governance of evaluation is often neglected in the evaluation literature: academic debates have traditionally been focused on the external/internal evaluator issue with little emphasis on the process of governance itself (Margo and Wilson 2018).

Table 2.3. Barriers connected to evidence-informed policymaking

Barriers concerning the policy-making process	**Facilitators concerning the policy-making process**
Lack of a culture of dialogue	Political commitment to evidence-informed policymaking
Primacy of political priorities	Dialogue between stakeholders
Ambiguity over mandate for evidence-informed policymaking	Strengthening demand for evidence
Weak long term policy planning	Supporting from the international community
Inflexible and non – transparent policy processes	Ensuring quality of evidence
Lack of trust between science and policymaking communities	
Barriers concerning the institutional set up	**Facilitators concerning the institutional set up**
Limited resources	Sufficient Resources
Weak incentives for evidence-informed policy making	Strong leadership and institutional memory
	Maximising positive disruptive power.

Lack of capability to engage in evidence informed policymaking	Flexibility in the job description and performance

Source: OECD, forthcoming.

Weiss (1999) outlines four reasons – the "Four I's" – why the uptake of evidence from evaluations is arduous: competing interest; diverging ideologies of different stakeholder groups; the information in the report itself can be a barrier (as other competing information sources say something different); and institutional characteristics (the structure and history of behaviour). Evaluator who may not have the necessary skills to properly perform the evaluation. Poorly implemented evaluations may result in decisions that unfairly reward or punish organizations and the individuals that they serve based upon biased or incomplete data. Policy makers can just select out information and interpretations that support existing policies and narratives (e.g., Stevens 2011) or exert pressure to shape the findings and recommendations themselves (Morris and Clark, 2013). Thus, the quality of an evaluation report does not guarantee that the information will be used in the subsequent deliberation or that is timely for the policymaking process. Rather the engagement and interaction between the client and the evaluator is sometimes more important to maximising the use of evidence (Johnson et al., 2009: 389). Furthermore, as policymakers are continuously under time pressures they tend to use the 'best available' evidence – relying on a broader conception of usable knowledge and recognizing the value of relevant professional expertise – rather than wait for information from scientifically rigorous examinations, e.g., RCTs or experimental assessment designs (Shillabeer, Buss, and Rousseau 2011; Head 2016).

2.3. How evaluation functions as part of learning and innovation processes?

"*We constantly have to justify our existence and effort by making it possible to see what value we've created.*"[3] (Thomas Prehn, Transformation Lead, Africa – BØRNEfonden*)*

Innovation typically builds on smart information gathering, with a good combination of both thin and thick accounts (Van Dooren and Willems, 2018) – the first scratch the surface and instigate debate; and the latter are better suited to generating explanations of differences in performance, designing future strategies and raising critical issues. Many argue that transparent measurement systems for assessing innovation success or otherwise in the public sector are vital to robust analysis and for creating cultures of learning (Mulgan and Albury, 2003; Bloch and Bugge, 2013). Yet, as mentioned above, knowledge management systems in the public sector are fairly nascent (Al Ahbabi 2018) and when it comes to innovation, some aspects are fairly tacit and cannot be easily captured. Box 2.7 outlines the distinction between tacit and explicit knowledge.

[3] https://apolitical.co/solution_article/innovation-labs-measure-success-to-justify-themselves-but-it-cant-be-done/

Box 2.7. Tacit and explicit knowledge

Explicit knowledge can be readily articulated, codified, transmitted and accessed using formal systems of language, numbers, recording etc. It is the knowledge of 'know-what' that can be stored on various forms of media.

Tacit knowledge is knowledge that is difficult to transfer to another person by means of writing or verbalising. It is personal, contextual, and often embedded in practice (learning by doing, knowhow, action-oriented knowledge, crafts and skills that apply to specific contexts).

The transfer of knowledge – from tacit to explicit and vice versa– is reliant on close social interaction and cannot be done easily. Hartley and Allison (2002: 105-6) provide four modes of knowledge conversion through which tacit and explicit knowledge can be created and transferred between individuals and groups:

- *Socialization*: a process of sharing experiences and thereby sharing tacit knowledge such as shared mental models and technical skills. It includes the processes of observation and imitation.
- *Externalization*: the process of articulating tacit knowledge into explicit concept and ideas.
- *Combination*: the process of systematizing concepts into a knowledge system and it occurs through combining and converting different forms of explicit knowledge.
- *Internalization*: the process of converting explicit to tacit knowledge and this tends to be achieved through practice, by 'having a go'. Manuals and other documentation can help to embed tacit knowledge but the 'embodiment' of knowledge through action is critical.

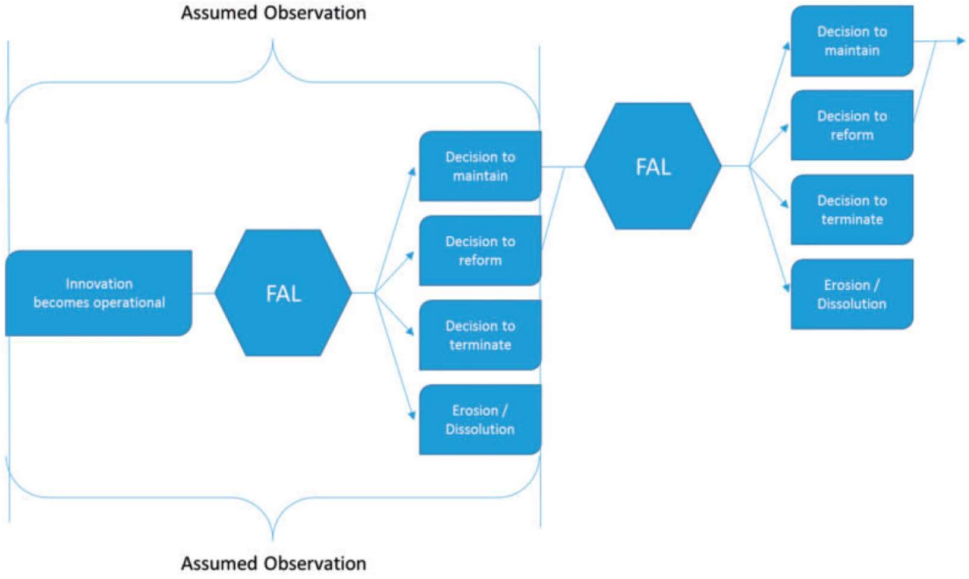
Source: Hartley and Allison, 2002.

When it comes to innovation, there are some innovation-related responsibilities organisations could be held accountable for: for example, creating an open environment that fosters the intake and circulation of ideas; providing for a safe space for change or sack resources to engage with new ideas; giving attention for diversity in the organization; scanning of the environment; or prototyping, experimenting and piloting (Albury, 2005; Hartley, 2005; Van Dooren and Willems 2018). Yet, most of these are activities and inputs to innovation rather than its impact. The problem is that the occurrence of innovation is *de facto* unpredictable and thus, too diffuse base to connect direct accountability measures to (Osborne and Brown, 2013). Furthermore, chance events (from political crises, natural disasters to unpredicted meetings) are an important, albeit unpredictable, part of the innovation process. Chance events also activate different accountability forums (Rixen 2013) that can put pressure that leads to innovation (Dooren and Willems 2018). Scrutiny of accountability forums can also be activated by innovation failures and accountability regime which focuses too harshly on mistakes and sanctions may discourage change (Van Loocke and Put, 2011).

The rapid pace of change – economic, technological, demographic, geo-political – makes it difficult even for expert analysts to understand trends and provide evidence-informed advice (Head, 2017). Thus, fast-paced changes and chance events can create windows of opportunity for innovation, but it also means that there may not be time for informed decision-making or consideration of all factors connected to innovations. Thus, there is also a need to question if existing evaluation frameworks are sufficient to describe and explain these new phenomena and provide timely input when needed. While private sector has many innovation evaluation guidelines and practises, including in the field of innovation policy (e.g., Technopolis Group and Mioir, 2012), the field of public sector innovation is not as developed.

Still, Van Acker and Bouckaert (2017; building on Frees and Bouckaert, 2015) propose a framework focusing on three concepts: Feedback, Accountability and Learning (FAL for short) specifically for the public sector. The FAL model assumes that the extent to which organizations (a) gather feedback information on the performance of their innovations, (b) are subject to accountability mechanisms, and (c) use this information through learning processes to improve the innovations, will influence the survival chances of innovations.

Figure 2.5. Heuristic model of the FAL process



Source: van Acker and Bouckaert, 2017.

The first component, feedback gives "*information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way*," (Ramaprasad, 1983: 4). Effective feedback allows organisations to correct errors, adjust its goals, to restore its performance levels (if performance gaps occur), and to align itself with its environment (Morgan, 2006; Walker, 2013). Internal feedback can be preoccupied with efficiency (how to do more cheaper, in more productive ways, how to use alternative methods for the same objectives), while other forms of feedback may be more concerned with the functioning of the overall system in connection to the changing environment (van Acker and Bouckaert, 2017). The second component, accountability, which was also discussed in previous two paragraphs, confronts public managers on a regular basis with

feedback information about the functioning of the organisation. If functioning correctly, it can stimulate and debate about the successes and failures of past policies, both separately and in dialogue with one another (van Acker, 2017). Accountability can produce information internally to correct for errors and find innovative solutions, but it can also raise broader debate in the broader accountability forums described in table 2.2 in chapter 2.1. These debates can give rise to also dissonant voices and thus, break conformist attitudes in organisations and more room for transformative change. Indeed, organisations can be subject to group think and conformist tendencies as the members over times start to share the same beliefs and values – this can become a serious barrier for innovation (see about workplace determinants of innovation in Nanda and Singh, 2009). Hence, sources from the outside – including critical outside evaluations and audits – are sometimes in a better position to question long-held behaviours and assumptions (Salge and Vera, 2012).

When it comes to learning, a receptive attitude towards different opinions and alternative ways of doing things, in conducive to innovation. This also means a tolerance for errors and risk-taking. At the organizational level, this should be supplemented with structural arrangements which allow organizations to process relevant information as a basis for change and innovation (Greiling and Halachmi, 2013). Also personal factors matter a lot for learning, for example, the top leadership's stance to supporting and investing in support systems for learning (Garvin, Edmondson and Gino, 2008). To understand the role of evaluation in innovation also the different types of learning have to be understood. These include single loop learning, double loop learning and deutero learning (Box 2.8). The types of learning applied in organisations reflect on its learning process itself and this can be also seen in evaluations these organisations conduct or commission (be they on the current functioning level; do they challenge underlying norms and assumptions or they look at the systems elements themselves).

Box 2.8. Types of learning

Single loop learning happens when an organization investigates its effectiveness, and adjusts its functioning if it finds a discrepancy. If the organization looks deeper, they may find need for changes in the program's underlying norms, assumptions and policies, not just the program's functioning. This is described by Argyris and Schön (1978) as *double-loop learning* (1978). The organization can also reflect on what prevented them from seeing that the system needed changing in the first place. This third level is called *deutero learning* (Schön, 1975).

Source: Argyris and Schön 1978; van Acker, 2017.

With its different components, FAL model (figure 2.5) proposes that once the interlinkages between the feedback, accountability and learning are in place they actually contribute to sustainability of innovations (van Acker and Bouckaert, 2017). This, however depends on the feedback and accountability measures to be in place and receptive learning systems present in organisations to capture the information. Furthermore, it is also crucial what the evaluators actually report about. If evaluations are only conducted to assure legal compliance or they concentrate on efficiency rather than effectiveness or the robustness of the system in its entirety, they will not contribute to innovation in the broader sense.

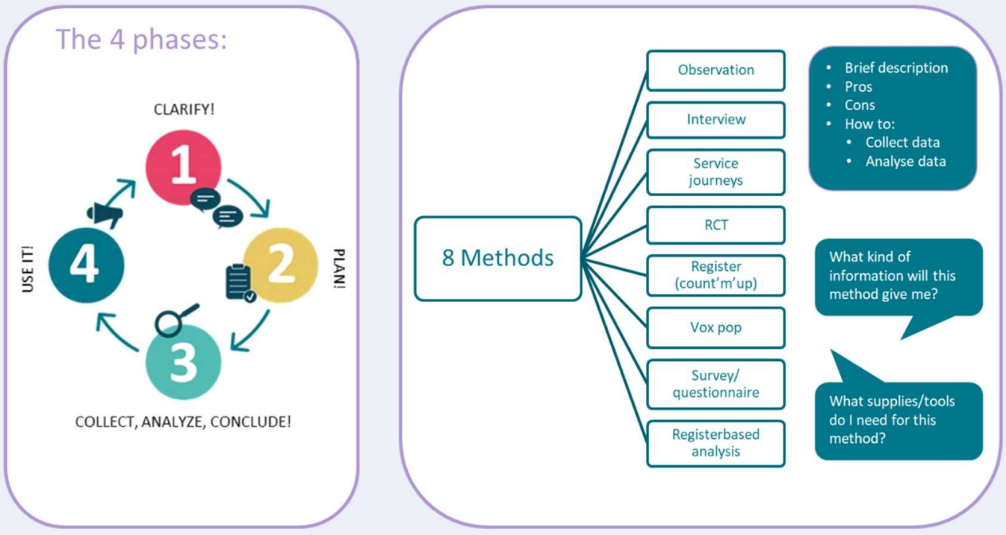
Moreover, through evaluations organizations can learn also from each other about the effectiveness of adopted innovations and the existence of alternatives – thus, it can both lead to the adoption and abandonment of innovations (Volden, 2010).

Thus, there is a great potential in the evaluation of innovation. Evaluations can help to qualify the innovation process to retain knowledge and provide reasons for implementing change. Evaluations can also identify the value of the innovative interventions and thus, make it easier for others to reuse the new solution. But it can also be a way to create legitimacy into the innovation process itself (as described in 2.1). At the same time, there are not too many public sector innovation specific evaluation approaches developed – one of outliers is the Danish government who has invested in developing an evaluation guidebook for public sector innovation (Box 2.8).

Box 2.9. Evaluation Guide to Public Sector Innovation in Denmark

The National Centre for Public Innovation in Denmark aims to increase awareness that the two disciplines are interdependent enable evaluation capacity building amongst innovation practitioners in the public sector. The centre in with collaboration with experts and practitioners has composed an evaluation guidebook, which gives an introduction to evaluating innovation, providing a map of different theoretical approaches and overall guiding practical advice for evaluation. The guidebook aids the practitioners through every step of the evaluation. The guidebook includes 10 separate tools that guides through specific actions from beginning to end when evaluating an innovation/innovative process. The guidebook is structured around four phases – (1) clarify, (2) plan, (3) collect, analyse and (4) conclude and uses eight main methods to evaluate innovations (Figure 2.6).

Figure 2.6. Structure and methods



Note: Learn more at: coi.dk/evaluering
Source: Lene Krogh Jeppesen and Maja Brita Hauan. Evaluating public sector innovation in Denmark. Presentation to the OECD, 2017.

2.4. The dark side of measurement and evaluations

> "*I cannot help fearing that men may reach a point where they look on every new theory as a danger, every innovation as a toilsome trouble, every social advance as a first step toward revolution, and that they may absolutely refuse to move at all*." (Alexis de Tocqueville, 1840)

While measurement can bring greater accountability, efficiency and learning, there is a considerable body of literature highlighting numerous negative features of performance measurement and evaluations (e.g., Hood 2007; Jackson 2011; Le Galès 2011; Lehtonen, 2015). These can also affect innovation evaluations and evaluation innovations. Some of these are outlined below:

2.4.1. Replicability over complexity and causality

The application of economic evaluation models "depends on reductive abstraction, sacrifices verisimilitude and nuance but gains in transparency and replicability" (Hill and Lynn, 2005, 175). When applied, these models are highly data-driven and thus, reliant on available data, meaning that many analyses are based on cross-sectional samples and not longitudinal datasets (ibid. 189). This does not however guarantee causality; what is mostly studied are associations. Someone is always selecting causal factors and then attributing causality and there is always some partiality connected to it (Reynolds et al. 2016).Claims about causality are therefore often made based on theory rather than demonstrated empirically (Yang, Hsieh and Li, 2009). Thus, in many cases, data from evaluations shows positive change, but it is difficult to confirm that the specific intervention caused the change. Furthermore, causal mechanisms through which policies achieve their intended impacts may not be those intended by their designers but may be due to the hidden influence of those implementing the policies. In an evolutionary perspective, all actors, not just policy-makers and innovators, learn from past experiences and the changes in behaviour that are induced over time through learning may complicate attempts to evaluate cause-effect relationships (Wegner 2003; Witt 2003). Richer empirical understanding of actual 'policy histories' is needed than is generally seen in innovation research or evaluation research. Understanding agency requires a 'narrative approach' that follows the actors and studies processes in real time, without the treatment of a sequence of events as inevitable (Garud, Kumaraswamy, and Karnøe, 2010).

In complex situations causal attribution can be very problematic and misleading (Forss, Marra and Schwartz, 2011). Furthermore, complexity science shows that situations under evaluation are always non-linear and comprising of feedback loops (De Haan, 2006). Yet, an evaluator's concern is with more directly observable social and public impacts and with monitoring the direct effects of research on such impacts (Reynolds et al., 2016). Thus, often evaluation scope remains too limited and may lead to sub-optimization: e.g., the improvement of a sub-system at the expense of the organization as a whole (Van Loocke and Put, 2010). Governments increasingly acknowledge the wicked nature of some of the most pressing problems they face, so they also need to realised that existing policy

toolboxes (of design, coordination, and evaluation) are not enough to tackle these challenges (Kattel and Mazzucato 2018). On the positive note, some systems and complexity researchers (e.g. Funnell and Rogers, 2011; Patton, 2011; Reynolds et al. 2016) are in-cooperating complexity into evaluation approaches.

2.4.2. Forgetting about time and place

The interventions evaluated as the evaluation process itself are subject to continuous change and uncertain effects (Reynolds et al., 2016), thus, static evaluations in a very dynamic and uncertain situation do not work. Ongoing change, especially when dealing with innovation, needs to be taken into account in evaluations in an adaptive way. Instruments may vary so much across time and space that attempts at the rational meta-evaluation of 'what works' are rendered meaningless for all but the simplest of interventions (Flanagan and Uyarra 2016). When it comes to private sector oriented innovation policies and innovations, there is more evidence of policy outcomes. Manchester 'Compendium' review of evidence on innovation policies (Edler et al. 2016) found no clear evidence that any class of innovation policy instrument studied works consistently from place to place or time to time. Uyarra and Ramlogan (2016) found wide differences in policy outcomes resulting from variation not just in objectives and implementation, but also due to context-specific institutional configurations and policy path dependencies.

2.4.3. Conflicting expectations

Vagueness of outcomes, multiple criteria and not defining key concepts such as quality and productivity are the source of conflict, confusion and stress when it comes to evaluation (Tan et al. 2014; Togni et al., 2015; Cabaj et al., 2015). One consistent source of conflict is reported in social innovation evaluations is between macro-level information needs (e.g. of funders for summative decisions or for use across contexts) and programme/project level needs (e.g., the innovators responsible for the project) (Milley et al., 2018). These can be both implicit and unspoken, only coming out in the later stages of the evaluation process. Sufficient time and flexibility to describe reporting methods tailored to various needs are required (McKegg et al., 2015), but there may not be always scope in the evaluations to do so and different sides will be left unsatisfied with the result (e.g., Mathie and Peters, 2014).

Schillemans and Bovens (2011) describe how actors may also be confronted with different accountability forums, each with its own set of evaluation criteria, themselves differing from expectations formulated by for example parliament or powerful interest groups. These sets might be partially overlapping, but also partially diverging, and perhaps even mutually contradictory. It may be arduous to combine these different expectations or to prioritize between them. Therefore, organizations trying to meet conflicting expectations are likely to end up in a state of dysfunctional paralysis. They tend to oscillate between behaviours which are consistent with conflicting notions of accountability (Ibid.; Koppell, 2005; Bovens, et al., 2008). If not carefully managed, role confusion can lead to unsustainable 'scope creep' (e.g. Poth, Pinto and Howery, 2012), and can contribute to oscillation between evaluation approaches (Ibid., Cabaj et al., 2015; see also – Milley et al., 2018).

2.4.4. Formalism and goal displacement

Evaluation is often associated with being a compliance exercise. As such, there is a always a risk of blame or fault for example through political capture and sensationalisation in media. An accountability regime which is too rigorous, may turn public institutions into

formalistic bureaucracies, which are obsessed with conformity with rules and procedures (Schillemans and Boven, 2011). Instead of a means to provide insight in, and reflection about performances and processes, the account giving may become a goal in itself (Bovens and Hart, 2005). Thus, evaluation and the actions leafing to it can become subject to goal-shifting and gaming. As such, rigorous sanctioning mechanisms might also have perverse effects as organizations will find ways to circumvent or manipulate the rules and invent measures to "play the game" to their own benefit (Kuhlmann and Bogumil, 2018).

2.4.5. Disincentives to responsibility and innovation

The expectation of negative or positive consequences, as a result of a certain behaviour is likely to put administrations under pressure because institutional actors seek to avoid negative consequences and to maximize benefits (Askim, Hjelmar and Pedersen, 2018). Public managers may get very good at fulfilling the requirements imposed by their accountability forums, rather than delivering effective outcomes; thus, developing tunnel vision towards the measurement system rather than having a broad view of the problems.

Too rigid evaluations – focussing too harshly on finding mistakes and punitive action – may discourage entrepreneurship, risk-taking, initiative and creativity (Hartley, 2008; Van Loocke and Put, 2010; Bovens, 2005; Bekkers et al., 2013). As stated above, mistakes and failure are part of the innovation and overall learning process and incentivizing public managers to shield from the former can lead to organisational paralysis. Thus, either intentionally or unintentionally, evaluations can legitimize and reinforcement prevailing power structures and institutional setups.

So, for example, an accountability forum that expects innovation activity, but punishes risk taking my induce civil servants to engage in 'innovation theatre' rather than actual performance. Consequently, civil servants may create a façade of plans, procedures and goals – even projects to look like innovation – to satisfy the forum, while behind the façade, everything continues as before (Bovens, Schillemans and Hart, 2008; Van Loocke and Put, 2010). Furthermore, accountability forums discussing the results of evaluations may systematically focus on certain aspects, while ignoring others. For example, focusing on performance, but ignoring broader systemic effects such as transparency, privacy etc.

2.4.6. Unintended use and unanticipated effects

On the one hand, evaluations often ignore positive or negative side- or ripple effects of innovations and other inventions, as these fall out of their evaluation frame. Yet, unintended effects may create new problems that displace the original policy problem (Bardach 2006). On the other hand, evaluations themselves can have 'unintended' or 'unanticipated' effects (cf. Morell, 2010).

Furthermore, the fixation on evaluations and measurement itself also contributed to government overload (Lewis and Triantafillou, 2012) – meaning that only evaluation gets done and there is no bandwidth for substantive change or innovation for that matter. Thus, evaluations not generating meaningful learning can draw resources away from serving the mission. Furthermore, the evaluations themselves can be interpreted in unforeseen ways and used against the main messages of the research. Public sector innovation is in many cases about political choices, change and is contentious by nature, thus, evaluations can be captured by political debate in the public forum.

2.4.7. Evaluation complexity and opacity

Before the complexity of the subject matter was covered, but also evaluations can be complex. Dialogue and deliberation can be considerably reduced due to the difficulty in interpreting results (Grimsley and Meehan, 2007). This can lead to misuse and misunderstanding of results of evaluations or simply, disinterest in them.

3. Evaluating innovation in a changing context

"*He that will not apply new remedies must expect new evils, for time is the greatest innovator*." (Bacon, 1625)

3.1. Can evaluations keep up with change?

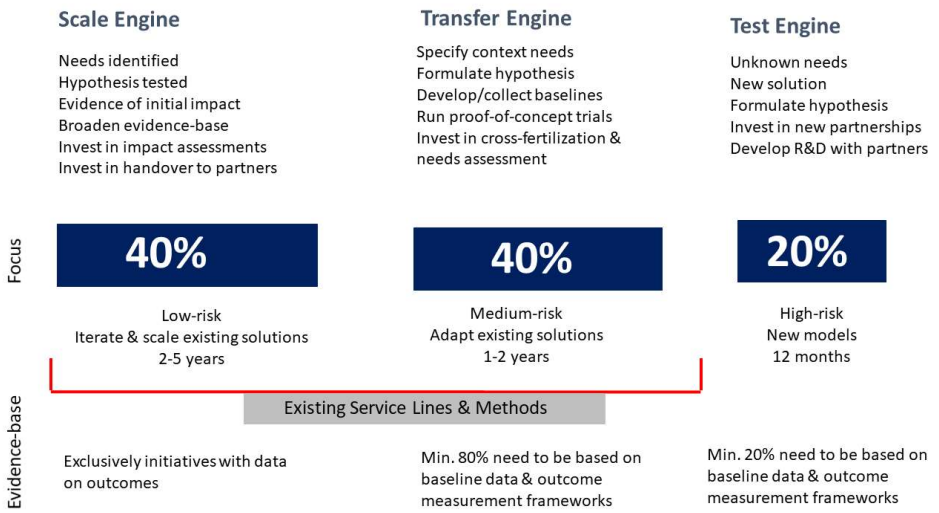
Literature on evidence-based policy suggests that technocratic models downplay the complex and contested nature of the policy process and many barriers to the actual uptake of information produced in evaluations exist (see Bristow, Carter and Martin, 2015). There is still a lack of connection between innovation and evaluation (Word, Stream and Lukasiak, 2011; Askim, Hjelmar and Pedersen, 2018). Furthermore, as the pace of change is continuously increasing, innovation is entering all fields and all projects. Thus, evaluating innovation cannot be discussed separately from the broader public policy system or its evaluations – it should be ingrained within all evaluation frameworks and practices. Hence, also established evaluation methods and approaches need to be reformed to take into account the amplified need for innovation across all government activity. Chapter 1 outlined some of the key challenges dealing with innovation brings to evaluations.

Primarily, these are specifically connected to the uncertainty and risk connected to the innovation process and the possibility of failure. Here it is fundamental to understand that risk does not equal uncertainty – the latter can be mitigated, but it cannot be assessed and calculated in the same manner as risk. Consequently, effects become apparent only when new methods and approaches are applied. Thus, continuous evaluations are needed, so, that projects in the public sector do not become 'too big to fail' as to avoid wasting time and resources with dead-end projects. However, these judgements are difficult, because it may take time for an innovative practice to perform and have the impact required. This means that evaluation needs to happen at all stages of the innovation lifecycle for feedback and learning to occur in a timely manner. One cannot wait for ex post evaluations for this feedback to come – evaluation has to become ingrained in the innovation projects themselves, but not become a barrier to taking risks.

Innovation projects will fail – it part of the uncertainty connected to the innovation process. Yet, more information is needed about why they fail –failure ca occur due to technological risks (Pärna and von Tunzelmann, 2007), rejection by potential users, or a lack of resources and capabilities for developing and implementing an innovation (Kay and Goldspink, 2012). All of these give different signals to the way forward. Moreover, public sector and its evaluation systems need to also agree upon – or at least debate about – an acceptable degree of failure when it comes to innovation. See the UNDP's Innovation Facility Fund overview in Figure 3.1 and the accepted levels of failure. Evaluations should not become too punitive, accountability forums too over-zealous and start to kill off innovations.

Figure 3.1. Investment Dimensions of UNDP's Innovation Facility Fund Portfolio

Type the subtitle here. If you do not need a subtitle, please delete this line.



Source: https://www.unhcr.org/innovation/innovation-metrics-for-human-development-what-have-we-learned/

Yet evaluations are necessary, because innovation is not inherently good nor bad and its impacts should be critically evaluated before scaling initiatives. Innovation changes the *status quo* and therefore creates winners and losers, who may have divergent views about the value and effectiveness of an innovative projects. To deal with adheres effects policymakers and innovators have to be aware of them. Furthermore, innovation projects may often be part of a portfolio of projects: investing in one innovation, may cut the potential for another. Assessments and evaluations of innovative projects should consider the innovation portfolio and the wider system; and how individual innovation projects may interrelate with one another within one organisation and within a broader eco-system.

As outlined above, creating legitimacy and trust is one of the core features of evaluation. This may be one of the core goals of evaluation of innovation already (e.g., public sector innovation labs and units produce evaluation information in different forms to demonstrate their usefulness as often their actions do not align with current organisational goals (Tõnurist, Kattel and Lember, 2015)), but it is often ignored as a purpose when planning innovations. Evaluation can assist by giving confidence that the project or programme has considered the right things, but the practice of evaluation might also contribute to doubt about innovation, as innovations will rarely run smoothly or better than existing initiatives. The context of innovative projects matters. It may be very difficult to isolate the precise contributions that shape an innovative project and its impact.

All of these issues mean that an understanding of evaluation, evaluation skills, and familiarity with evaluation practices will need to be much more widespread among innovators themselves. Moreover, professional evaluators need to start taking into account that most programmes and projects will have an innovation component to them that need to be addressed differently than business as usual.

3.2. Pushing past old dogmas

With the rise of the entrepreneurial state (Mazzucato, 2015) and more discussion about government as a 'technology maker' (Karo and Kattel, 2016), new dynamic capabilities towards innovation and evaluation are required. Doing things differently also necessitates looking and old assumptions and established evaluation canons with a critical perspective. There are various assumptions connected to both scientific theories and evaluation – the power of statistics, the 'gold standards' of large RCTs, cognitive agency, statistical independence, bias as a de facto negative, cause and effect, absolute/objective truth, and nowadays the overreliance on big data (Brockman, 2015)[4] – that need debunking.

Primarily, evaluations and evidence cannot be divorced from intrinsic roles of values, ideologies and economic interest that shape the democratic political system and the policymaking within. When addressing complex policy and program areas, collaborative approaches to knowledge sharing and adaptive management in light of experience will be necessary and not everything can be solved by 'objective' evidence alone (Schorr and Farrow, 2014; Head and Alford 2015). Thus, rigor and relevance of information is sometimes more important than the specific method-led approaches (Goodman, Epstein, and Sullivan, 2018). For one, the RCT lobby has affected the types of evaluations countries carry out (Jacob, Speer and Furubo 2015). This is a critical insight during the upsurge of randomised control trials (RCTs) as the golden standard of evaluation. RCTs in some situations may be very useful, but they are also complex, time consuming, and expensive (Luce et al. 2009). The idolisation of a specific, positivist, type of knowledge (Raadschelders and Lee, 2011) needs to stop and public servants need to think more carefully what information and evidence in their specific context is timely and useful to them. Thus, for example the juxtaposition of qualitative or quantitative methods as an either or is slowly disappearing in practice.

It is often assumed that evaluation would bring rationality to policymaking (Weiss 1999) that all bias is bad. Yet, policymaking and innovations for that matter are normative, not ideologically neutral or apolitical, and ignoring those factors make evaluations less relevant to the political economy they land in. Different interest, ideologies, information and institutions shape policy and innovations and thus, competing interests and power, knowledge, dominant rationalities and ideologies, as well as alienating and constraining practices should be part of evaluation. The desire to comply with 'professional' and rational standards – as a good in itself – may actually turn into a threat of moral inversion as evaluators often fail to question seemingly rational practices and assume that instrumental practices are neutral and legitimate (Dillard and Ruchala, 2005). For example, in the field of innovation policy there has been a underlying faith in the existence of 'rational design' and coherence of policies, which may be fundamentally flawed (Flanagan and Uyarra 2016). These assumptions also mean that legacy systems and policies are not questioned as critically as the ideas that challenge them creating great barriers to entry to new, innovative solutions. On the flip side, under the hermeneutic evaluation tradition (outlined in table 2.1 in chapter 2.1), Jones and Hughes (2001) advocate the use of situated evaluations, where

[4] See also a good account here: Ofir, Z. (2018) Updating the DAC Evaluation Criteria, Part 7. Let these ideas die! http://zendaofir.com/dac-evaluation-criteria-2-0-part-7/

people are seen as 'active makers of physical and social reality' and their informal opinions and gossip – important to them and peer groups – and should be part of evaluations.

Seeing policies as atemporal is another source of concern in evaluation or that there is one best way of doing things or 'what works.' There is no policy that is perfect – policies succeed and fail in the context they are in. Innovations as well as policies are adopted into pre-existing institutional frameworks and roles of actors that have evolved over long periods of time. Path-dependency implies bounded change and when dealing with innovation, positive feedback is not always your friend as evaluations may reinforce existing structures and institutions.

The old maxim of 'under promise, over deliver' is actually counterintuitive to the new rise of mission-oriented innovations. Evaluation may exert pressure on managers to 'cherry pick' only those goals, which are achievable or easily measured, but this will not get governments towards bold new solutions. Furthermore, the silo-isation in evaluation – taking a programme or an individual project level as the primary interest – is problematic especially within the context of mission-oriented innovations and among them SDGs which should deliver across national, regional and global level. Here approaches such as policy and innovation lifecycles are actually not helping, because they as well rationalise and specialise an inherently messy processes. By appointing different sets of professionals to work on their respective lifecycle phases – including evaluation – the feedback loops from practice are broken (Howlett and Wellstead, 2011) and only a small part of the senior staff may be in the position to grasp the 'big picture' of the various elements of the process (Head, 2016).

Policy problems are wicked, uncertain and complex and so is most of innovation activity. The reaction to that is usually to ignore the issue – as it is too complex to tackle – or assume that approaches such as experimentation or adaptive learning will solve the issue. The reality is that they do not and simplified theories of change and linear evaluation methods (e.g., logframes) may actually create advice that is not useful in practice. These are issues that cannot be 'solved'; they can only be addressed in action, in being continuously reflective of assumptions, mental models and values during evaluations.

3.3. The rise of public value

> "*Innovation is a dangerous concept for social scientists who study government and governance. That is because it is both conceptually unstable and normatively loaded*."
> (Christopher Pollitt, 2015)

One of the defining characteristic of evaluation is the judgement of value and when it comes to transformative change it is difficult – at least in the public sector – to talk about little else than value shifts. Thus, value should also be put at the centre of evaluating public sector innovation.

Value can often be defined in terms of merit (an intrinsic context-free value), worth (a contextually determined, place-bound value), and significance (a value that is related to some norm or state of affairs) (Reynolds et al. 2016), but there are a variety of different public values (table 3.1). In general, public value represents a normative consensus of prerogatives, principles, benefits and rights that can be attributed to both governments and citizens (Jorgensen and Bozeman, 2007) and can be linked to more values of good

governance in general like transparency, participation, integrity and lawfulness. Thus, public value can pertain to both the content of the service itself and how it is delivered. The heterogeneity of what public value can mean is at the heart to public sector work. Yet, policy systems are notoriously ill-equipped with dealing with complex problems not to mention the public values connected to them (OECD 2017).

Table 3.1. Value sets and categories

CATEGORY	PUBLIC SECTOR CONTRIBUTION TO SOCIETY	TRANSFORMATION OF INTERESTS TO DECISIONS	RELATIONS BETWEEN PA AND POLITICIANS	RELATIONS BETWEEN PA AND ITS ENVIRONMENT	INTER-ORGANISATIONAL ASPECTS OF PA	BEHAVIOUR OF PUBLIC SECTOR EMPLOYEES	RELATIONSHIP BETWEEN PA AND CITIZENS
VALUE SET	*Common good* *- public interest* *- social cohesion* *Altruism* *- human dignity* *Sustainability* *- voice of the future* *Regime dignity* *- regime stability*	*Majority rule* *- democracy* *- will of the people* *- collective choice* *User democracy* *- local governance* *- citizen involvement* *Protection of minorities* *- protection of individuals rights*	*Political loyalty* *- account-ability* *- responsive-ness*	*Openness-secrecy* *- responsiveness* *- listening to the public opinion* *Advocacy-neutrality* *- compromise* *- balance of interests* *Competitiveness-cooperativeness* *- stakeholder or shareholder value*	*Robustness* *- adaptability* *- stability* *- reliability* *- timeliness* *Innovation* *- enthusiasm* *- risk readiness* *Productivity* *- effectiveness* *- parsimony* *- business-like approach* *Self-development of employees* *- good working environment*	*Accountability* *- professionalism* *- honesty* *- moral standards* *- ethical consciousness* *- integrity*	*Legality* *- protection of rights of the individual* *- equal treatment* *- rule of law* *- justice* *Equity* *- reasonableness* *- fairness* *- professionalism* *Dialogue* *- responsiveness* *- user democracy* *- citizen involvement* *- citizen's self-development* *User orientation* *- timeliness* *- fairness*

Source: Based on Jorgensen and Bozeman 2007, 360-361.

In the overall context of 'value pluralism', it is well established that civil servants often face value conflicts and controversial value-laden issues daily (de Graaf and Paanakker, 2015; Page et al. 2015; Schott, Van Kleef and Steen 2015). For example, a corrections officer may be aware that the rehabilitative justice will be more effective in the long term, but must deal with politically supported punitive approaches or a solution supported by cost-benefit analysis puts marginalised groups in worse situation (see Alford et al. 2017). The Right-versus-Right dilemma is one of the most common ethical dilemmas faced by policymakers and these dilemmas should be also evaluated in the context of winners and losers of innovations. These dilemmas are often the result of two intersecting values or issues that meet in such a way that policies or regulations provide little to no guidance: for example, choosing beween ideals such as truth versus loyalty, individual versus community, short-term versus long-term and justice versus mercy (Word, Stream and Lukasiak, 2011; Head 2017). As these dilemmas are at the basis of innovations as well, evaluative judgements cannot be made across different context and cultures without acknowledging the values behind such arguments.

Value does not only have a cost dimension, but also can be described through social outcomes and trust/democracy (see Kelly et al. 2002; also Bason 2010 in the context of

innovation). Consequently, not all innovation projects are developed with 'efficiency' or "productivity" as a particular end goal. When effectiveness becomes something beyond 'value for money,' then the latter does not have a real maximum nor is it easy to quantify (Tangen, 2005). Projects aimed at collaboration, citizen engagement, etc. are more about improving relationships, customer satisfaction, ownership, and trust. For example, positive added values – e.g., shortened service delivery time, increased legitimacy, life-expectancy etc. – are difficult to link to technological performance of a particular public organization. There are multiplicities of factors that influence innovation impacts also in the public sector (within the innovation ecosystem) that are exogenous to the specific organization.

4. Better evaluation to fit different types of innovation: multi-faceted nature of innovation

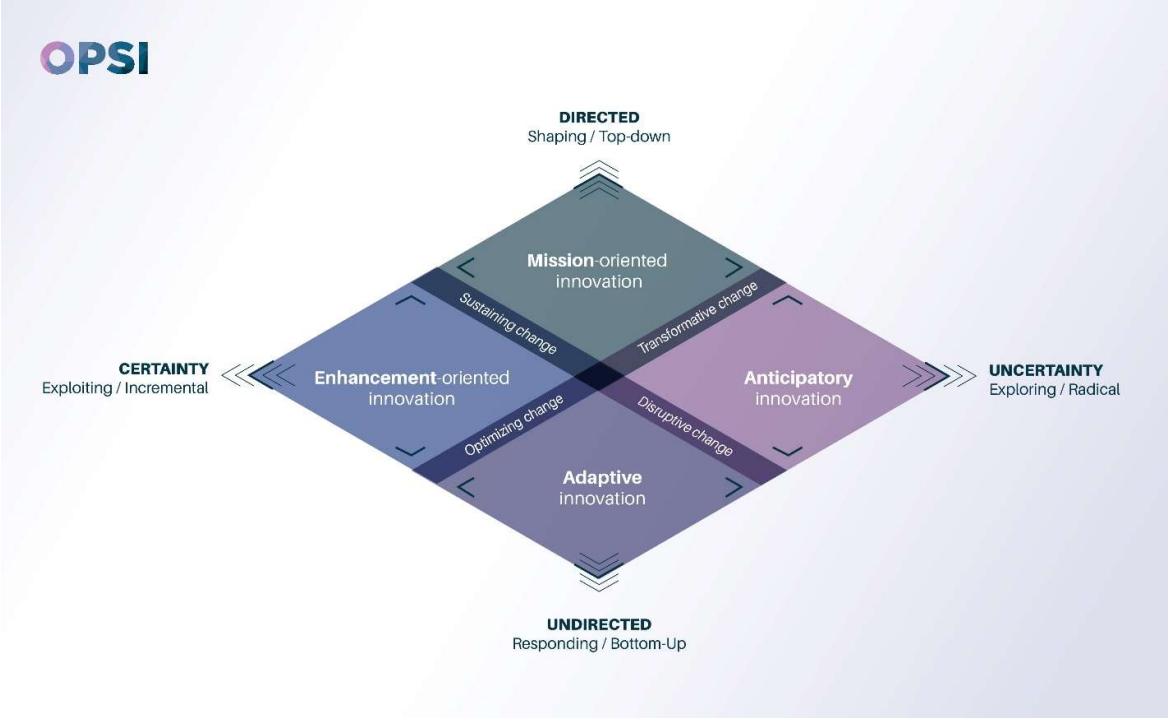
Despite an increasing number of studies on innovation, studies have generally treated innovative activity as a homogeneous phenomenon. Nevertheless, past research has argued that distinguishing different types or dimensions of innovation is necessary for understanding organizations' innovative behaviour, because they have different characteristics and organisational responses (De Vries, Bekkers and Tummers, 2016; Torgase and Arundel, 2016). For example, there is a sub-stream of 'complex innovations' (Damanpour, Walker, and Avellaneda, 2009; Goffin and Mitchell, 2010; Demircioglu and Audretsch, 2018). Others have applied the radical or incremental innovation dichotomy (Albury, 2005; Osborne and Brown, 2013).

Assuming that differentiating innovations brings value to their stewardship, the Observatory of Public Sector Innovation has proposed a public sector innovation facet model based on two core characteristics – directionality and uncertainty. The current report hypothesises that also evaluation efforts for these different facets will differ. As such, public sector innovation will occur in contexts with different levels of uncertainty, and those different contexts will require different strategies, working methods, and types of dissemination and diffusion. An innovation portfolio can be understood in terms of facets, depending on two factors:

- Is the innovation directed? Does it have a clear intent/objective that it is trying to achieve, or is more about discovery and responding (proactively or reactively) to externally generated change?
- Is the innovation dealing with high uncertainty? For example, is the context one of exploring completely new ground, or is it one where the challenge and context is relatively understood?

Based on these two factors, four facets emerge (Figure 4.1). The specific features of each facet are outlined below.

Figure 4.1. Public sector innovation model



Note: The model is developed as part of the work of the Observatory of Public Sector Innovation, OECD.
Source: OECD.

4.1. Enhancement-oriented innovation

This facet focuses on upgrading practices, achieving efficiencies and better results, and building on existing structures, rather than challenging the status quo.

It will generally exploit existing knowledge and seeks to exploit previous innovations. This type of innovation often builds efficiency, effectiveness and impact via existing processes and programmes.

This is traditionally where most governments have focused their innovation efforts.

Innovation evaluation here will be concentrated on efficiency and effectiveness and more traditional measurement approaches can be applied.

4.2. Mission-oriented innovation

This facet involves a clear outcome or overarching objective for which innovation is leveraged. There is a clear direction, even if the specifics of how it will be achieved may be uncertain.

This type of innovation can range from the incremental to the more radical, but will often fit within, rather than subverting, existing paradigms.

Such innovation can be very important for achieving societal goals, though it also works at an organisational or individual level to align activities. Public sector bureaucracies are naturally attuned to this sort of innovation, provided there is sufficient political will.

Missions can generate motivation and inspiration, a sense of what is trying to be achieved beyond the day-to-day process work, as well as guidance and reassurance when specific plans fall off track. A clear goal makes the value of diffusion and learning apparent.

Evaluation of mission-oriented innovation has to take a cross-boundary and systemic perspective and cannot rely on linear input-output tools.

4.3. Adaptive innovation

In this facet, the purpose to innovate may be the discovery process itself, driven by new knowledge or the changing external environment. When the environment changes, perhaps because of the introduction of innovation by others (e.g. a new technology, business model, or new practices), it can be necessary to respond in kind with innovation that helps adapt to the change or put forward something just because it has become possible.

This type of innovation can also range from the incremental to the more radical. However the more radical adaptive innovation is, the more likely that a public sector organisation will either endorse it from a leadership level or seek to suppress it or force it outside of the organisation.

Adaptive innovation can be extremely valuable in matching external change to internal practices and usually it cannot be directed top down, because people's developing needs cannot be prescribed. Adaptive innovation will generally be driven from the bottom-up, as those closest to citizens and services will often be the ones who see the need for change and react accordingly.

Evaluation of adaptive innovation is very difficult to conduct, as economic discovery is an uncharted process. This type of innovation is better evaluated through its enabling factors and activities rather than innovations themselves.

4.4. Anticipatory innovation

This facet involves exploration and engagement with emergent issues that might shape future priorities and future commitments. It has the potential to subvert existing paradigms. Very new ideas generally do not cohabit well with existing reporting structures, processes, and workflows. Anticipatory innovation therefore generally requires being sheltered from core business and having its own autonomy. Otherwise the pressures of very tangible existing priorities (such as existing missions) are likely to cannibalise any resources that are dedicated to something preliminary, uncertain, and with no guarantee of success.

Anticipatory innovation is important because big changes are often easiest (and cheapest) to engage with and shape when they are still emergent and not locked-in.

This type of innovative activity is the most uncertain and future oriented (option theory) base evaluation methods and approaches should be applied. This may also mean that feedback from the current system has to be to a degree ignored to assure the ambidextrous stance of evaluations.

5. Tools, methods and approaches to support innovation evaluation

Evaluating innovation requires flexible evaluation methods, which can accommodate the complexity and changeability innovations so often contain. Many existing evaluation methods are likely to still be of use for innovative projects, however there may be needed changes in focus or perspective (e.g. the boundary around which innovation failures are assessed). There are also different levels – project, programme and systems – of evaluation than need to be taken into account. The key ingredients of evaluation are to select the right mix of method for the innovation at hand (based on the underlying problem, timing and aim of evaluation) and collect and analyse data systematically.

Evaluation in the public sector is not by any means a new concept and there are multiple methods and practices that have been developed. In some ways this diversity is a challenge – when there are so many, it can be difficult to choose which might be most appropriate when. The tools ranging from qualitative methods such as interviews and observation, to quantitative methods based on numbers, statistics and economic modelling. Both approaches are important in evaluating innovation and as argued before it is not conducive to argue which is produces 'better' evidence, rather the argument should be: when is an approach useful.

5.1. Evaluation of performance and outcomes

5.1.1. Impact assessment

Impact assessment is a means to measure the effectiveness of organisational activities and changes brought on by those activities. It relies on establishing the cause of observed changes (impacts) referred to as causal attribution (also referred to as causal inference). Thus, impact assessment methodologies are not universally applicable, but depend on the objective of the impact assessment exercise, its timing (ex ante and/or ex post); and the scope and nature of the problem. Impact assessment can cover both economic and non-economic impacts (Godin and Doré, 2006). The interventions can be at the project, programme or on the broader policy level. Impact assessments have usually been applied ex ante as part of needs analysis and planning activity; or ex post to assess effects of an intervention.

Impact assessment is usually part of evaluations, not a separate product of the latter. Evaluations in general are usually broader analysing the appropriateness of the intervention design, its cost and efficiency, its unintended effects and possible learning for the design of future interventions. For this reason using impact assessment for innovation evaluation, the approach should be strengthened by other evaluation approaches and methods to capture core features of innovation.

Read more about impact assessment in OECD (2015).[5] Impact assessment methods include variety approaches and a selection of these is outlined below.

[5] See also https://www.betterevaluation.org/en/themes/impact_evaluation#ImpactEvaluation_1

Outcome mapping

Outcome mapping[6] is a methodology for planning and assessing projects that aim to bring about tangible change. Outcomes here are defined as changes in the behavior, relationships, activities, or actions of the people, groups, and organizations with whom a program works directly (Earl, Carden and Smutylo, 2001: 1). Outcome mapping can be applied both ex ante and ex post, but also for ongoing monitoring. It originates from international development, and can also be applied to projects (or programme) relating to research communication, policy influence and research uptake.

At the planning stage, the process it helps a project team or program be specific about the actors it intends to target, the changes it hopes to see and the strategies appropriate to achieve these. In ongoing monitoring, outcome monitoring helps to measure the most significant changes of actor behaviour towards specific outcomes (Smutylo, 2005).

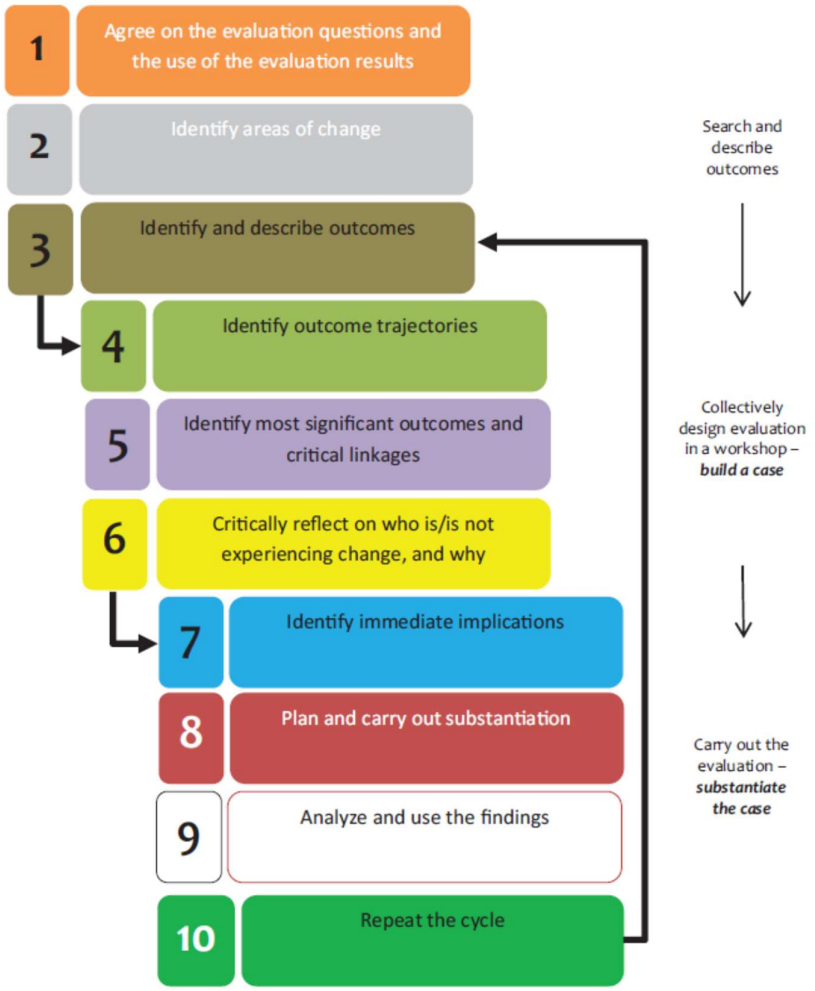
As an ex post evaluation approach, outcome mapping unpacks an initiative's theory of change, provides a framework to collect data on immediate, basic changes that lead to longer, more transformative change, and allows for the plausible assessment of the initiative's contribution to results.

Outcome harvesting

Outcome Harvesting is a method that enables evaluators to identify, formulate, verify, and make sense of outcomes (Wilson-Grau and Brit, 2012). Using the method the evaluator gathers information from reports, personal interviews, workshops and other sources to document how a given program or initiative has contributed to outcomes. This requires considerable skill and time making the approach resource intensive. These outcomes can be positive or negative, intended or unintended, but the connection between the initiative and the outcomes should be verifiable (Ibid.). The approach is quite close to outcome mapping and is described in Figure 5.1. What makes the approach interesting for innovation evaluation is the fact that it discover results without reference to predetermined objectives (in many cases during the innovation process these change considerably) (Britt and Patsalides, 2013).

[6] Read further: https://www.betterevaluation.org/en/plan/approach/outcome_mapping

Figure 5.1. Ten steps of an outcome evidence process



Source: Paz-Ybarnegaray and Douthwaite, 2017.

Compared to other impact-oriented evaluation methods outcome harvesting does not measure progress towards predetermined outcomes, but works backward to determine whether and how the project or intervention contributed to the change. Thus, the method is good at establishing the links how innovative change happened, but does not give a lot of insight into what type of change is required in the future.

Contribution analysis

Contribution Analysis is an approach for assessing causal questions and inferring causality in real-life program evaluations (Mayne, 2011). It offers a step-by-step approach to analyse what contributions project/programme has made towards specific outcomes. The essential value of the approach comes from the possibility to reduce uncertainty about the contribution the intervention is making to the observed results through an increased understanding of why the observed results have occurred or not. Contribution analysis also looks at the roles played by the intervention and other internal and external factors.

Contribution analysis follows generally six steps: (1) circumscribing the causal issue, (2) developing the theory of change, (3) gathering evidence, (4) defining the contribution story, (5) strengthening the draft contribution story by review and quality assessment, and (6) finalising the contribution story (Delahais and Toulemonde, 2012).

Contribution analysis is particularly useful when the innovation is not experimental, i.e. when it is based on a clearly articulated theory of change and there is little or no scope for varying how the program is implemented. The analysis either verifies the postulated theory of change or suggests revisions in the theory where the reality appears otherwise, but does not uncover new positions. This means that the method may be useful in enhancement-oriented innovation evaluation, but may fall short in other cases. Still, when there are several theories of change at play contribution analysis can be used to explore this collection of theories of change, probably in an incremental and iterative fashion to make the task practical (Mayne, 2011).

Case studies

Case studies can be used for different types of evaluation purposes including the establishment of causal relationships and impact. They usually focus on a particular unit (e.g., innovation project) and can often combine both qualitative and quantitative data. Case studies are a good approach to understand the relationships between different elements within a broader innovation eco-system. They can be used to capture program effects, illustrate evaluations (adding realism and in-depth examples), as exploratory components to the study (to generate hypotheses), as ways to capture critical instances (e.g., positive deviance from the norm), as implementation studies of different operations or in a cumulative manner (bringing together information from different case studies) (Morra and Friedlander, n.d.)

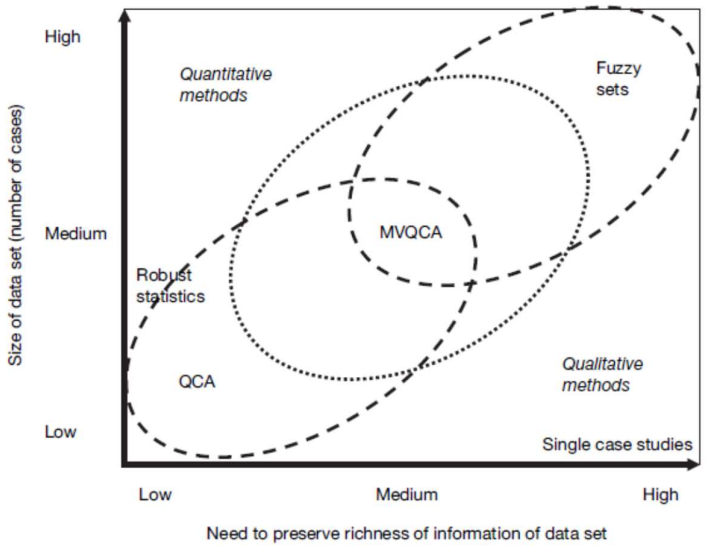
Case studies provide a rich descriptions of contextual variables (contingency enhancing internal validity), but they have little methodological capacity for external validity (generalizability) (Granato and Scioli, 2004, 314). They are very good to explore the effects of innovations when they have not broadly diffused. Case studies are also good ways to capture critical instances or innovation occurring or failing to understand the possible reasons behind the former. However, usually case studies analyses do not have the sufficient power to filter through many plausible rival explanations. There is a higher probability of sampling errors and less potential to notice large, nominal trends in data (Banfield and Cayago-Gicain, 2006). As case study analysis are usually very intensive then they tend to be also expensive forms of evaluation.

Qualitative comparative analysis

The qualitative comparative analysis (QCA) is a hybrid method developed by combining the advantages of both qualitative analysis (case-oriented research) and quantitative analysis (variable-oriented research) (Ragin, 2000). The approach is based on choosing multiple case studies for analysis. The methods tries to merge on the one, insights from different cases and capture the complexity of the cases; and on the other hand, produce some level of generalisation (Rihoux 2006). QCA can be used to either summarise data, check coherence within the data, test existing theories or assumptions, test some new ideas formulated by the evaluator and finally QCA also allos to elaborate new assumptions (ibid.). Thus, QCA method analyses the causal complexity of how the various causal conditions are combined and how it affects the outcome, and has strengths from small to medium-level case studies (Yong and Park, 2017). It does not decomposing cases into

variable units, emphasizing 'equifinality' arguing that there are various causes and paths that can produce the same result. The approach has inspired contributory techniques such as fuzzy set, multi-value QCA and other techniques. The method relies on finding comparable cases, which may be an issue when dealing with innovation. However, in sector specific innovation projects it may be an interesting method to apply.

Figure 5.2. Best application of QCA approaches



Source: Rihoux, 2006.

Scorecards

Innovation scorecards provide a rapid overall assessment of the practices adopted or the result achieved. Scorecards do this by assessing achievement or progress towards a particular goal, usually deploying predefined categories of assessment. They are useful in situations where there is little time for trials or other means of assessment and decisions have to be taken fast (see example in Box 5.1). The most known of the approach are Balanced Scorecards that were widely applied in the private sector also in the field of innovation and then on heavily critiqued for being excessively rigid, mechanistic in mind-set, static in nature and unable to account for the external environment and systemic linkages (Voelpel, Leibold and Eckhoff, 2006). This limits the perspective on innovation rather than broadening it. As such potential unintended side effects may be missed.

Box 5.1. Evaluation matrix for innovative solutions in the active and healthy ageing (AHA) field

This matrix was developed by the PROEIPAHA project to enable public purchasers in the health and medical-social sector to evaluate the relevance of an innovative solution

according to a previously identified need, in order to facilitate any procurement decision. The matrix covers the definition of need, enterprise characteristics, solution characteristics, benefit and financing assessment.

- The purpose of the "Enterprise" area is to ensure that the enterprise offering the innovative solution identified can successfully engage in a possible collaboration.
- The objective of the "Solution" area is to ensure that the characteristics of the innovative solution identified are appropriate to address the identified need.
- The objective of the "Benefit" area is to ensure that the potential benefits/gains of the innovative solution identified are appropriate for the gains/benefits sought.
- The objective of the "Finance" area is to ensure that the financial aspects connected to the purchase of the innovative solution identified are appropriate for the need.

Table 5.1. Practical information for the use of the matrix

Objectives of the matrix	Prerequisites for use	Potential users
- Analyse the relevance of an innovative solution according to an identified need. - Facilitate the procurement decision.	1) Identify innovative solutions. 2) Collect information concerning the identified solution(s) and the enterprise(s) 3) Know the specificities of the sector/market concerned.	- Purchaser or any other person responsible for the purchase - Management - End users - Expert on the sector concerned - Innovation expert

Source: Add the source here. If you do not need a source, please delete this line.

Source: https://ec.europa.eu/eip/ageing/public-procurement-platform/aha-innovative-solutions/3-matrix-evaluation-innovation_en

Benchmarking

Benchmarking is a reference point or standard against which performance or achievements can be assessed (OECD, 2011: 18). Benchmarking originates from the quality movement of the 1980 and 1990s. Benchmarking models can be classified either according to their content (either process, functional, performance or strategic) or their purpose (either competitive or collaborative) (Anand and Kodali, 2008). Specifically, performance benchmarking is concerned with quantifiable outcome characteristics. Usually benchmarking tries to identify best practises and compare against them, which may be problematic in the innovation setting. Furthermore, benchmarking has a high danger of comparing apples to oranges. Different countries and organisations have different legacy systems and pursue different reform trajectories. The unit of meaning – entity in charge – can vary from country to country, to organisation to organisation (Wollman 2003).

Expert reviews

Expert reviews. They can be relatively quick, low-cost, well-known, widely accepted and versatile tool which can be used to answer a variety of questions throughout the project performance cycle as well as in other applications (OECD 2011). They also provides an opportunity for mutual learning. However, it is difficult to ensure the accuracy and quality

of the resulting evaluations. Expert review has limited usefulness as a method to guarantee reliability and consistency (or repeatability).

Table 5.2. Phases and key actions for the expert review

Phases	Actions
Pre-review	**Establishing the foundations of the review** • Initiating the review: Assigning responsibilities (K) • Identifying the purpose and scope of the review • Identifying information needed and data collection/analysis processes • Identifying the evaluation criteria and review questions to be used • Identifying the types of review group and the audience (K) • Establishing timeline and determining logistics for the review **Selecting and inviting the reviewers** • Identifying criteria for selecting reviewers • Developing a list of possible reviewers and nominating • Gathering background information and developing initial selection list • Selecting the chairperson and reviewers from list of nominees **Preparing tools and materials** • Developing guidelines and tools for the review • Developing the presentations • Providing evaluation materials • Creating the expert review record)
Conducting a review	• Providing final instructions to the reviewers • Presenting the programme and Q&A • Discussing and judgement • Synthesising evaluation results from reviewers • Developing review documents and report
Post review	• Integrating additional comments • Writing a final report • Making the report available to the public • Assigning action items and evaluating response to action items • Evaluating the expert review process itself, including lessons learned

Note: The OECD adaption is based on EERE (2004), EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review, August; Kostoff, Ronald N. (2003), Science and Technology Peer Review: GPRA, Office of Naval Research; Kostoff, Ronald N. (2004), Research Program Peer Review: Purposes, Principles, Practices, Protocols, Office of Naval Research; Rigby, John (2002), "Expert Panels and Peer Review", in Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexande Tubke, and Klaus Zinocker (eds.), RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies. IPTS Technical Report Series, EUR 20382 EN.
Source: OECD 2011.

Public value mapping

Public value mapping is a practical approach rooted in public value theory. It is not a methodology per se, but a loose set of heuristics for developing analyses of public values. Public value mapping often begins with a set of core assumptions. Usually PVM starts with case study analysis where public values are analysed. Then the criteria of analysis are applied and value analysis chains are developed (Bozeman and Sarewitz, 2011). These approaches may benefit the analysis of transformative change when public values shift through innovations.

Box 5.2. Core assumptions of public value mapping (PVM)

1. PVM is either prospective (analysing planned or projected research activities), "formative" (analysing such activities as they occur), or "summative" (evaluating activities and their impacts after they have occurred).

2. It seeks to take into account the highest order impacts of activities (i.e. broad social aggregates) and thus focuses on social indices and social indicators.

3. It is multilevel in its analysis, seeking to show linkages among particular programme activities of an agency or institution, activities of other agencies or institutions, relationships – intended or not – among various institutional actors and their activities.

4. PVM is concerned with understanding the environmental context of research and related programmatic activities, with locating the activities and their institutional actors in terms of other actors in the environment, and with the constraints, opportunities and resources present in the environment.

5. Research in any field by any method is embedded in a social context; in PVM analysis of the social context of research (i.e. characteristics of research performers, their attributes and social relations) is part of the analysis.

6. PVM is guided by a "public value model of outcomes", rather than a market-based or market failure model. PVM explicitly rejects evaluation and assessment based on commodification of values and outcomes. Market prices are viewed as weak partial indicators of the social value and outcomes. Even as a partial indicator, market value is considered in terms not only of magnitude but also of distribution and equity.

7. Since market value is eschewed in PVM and since social values are not interpersonally transmissible, PVM anchors its outcomes values in a wide range of criteria derived from diverse sources including: official, legitimated statements of policy goals; goals implicit in poorly articulated policy statements; government agencies' goal statements in strategic plans; aggregated statements of value represented in opinion polls; official policy statements by government actors; and official policy statements by relevant non-governmental organisations (NGOs).

8. PVM analyses (maps) the causal logic relating goals (any of the above) to measured and hypothesised impacts and outcomes of science and research activities. When possible, the analysis begins from the causal logic articulated by responsible officials. The causal logic, explicit or implicit, is then considered in relation to various plausible alternative hypotheses and alternative causal logics invented by the analyst.

9. PVM is not an analytical technique or even a set of analytical techniques, but a model that includes a guiding theoretical framework (public value theory) and a set of assumptions and procedures. Research techniques employed in PVM depend upon the needs and possibilities afforded by the context of application. The only technical approach used in all applications of PVM is the case study method.

10. After gathering data to test hypotheses about causal logics and outcomes, the hypotheses are tested using appropriate analytical techniques and the impacts and outcomes are measured. The results of the analysis focus on relations among the causal logic, the environmental context, and the measured impacts and outcomes.

11. PVM links impact and outcome measures back to aggregate social indicators or other broad-based, trans-institutional, or trans-research programme measures of social well-being.

12. PVM concludes the analysis with recommendations for possible changes that seem likely to lead to better social outcomes.

Source: OECD, 2011; Bozeman, 2003.

5.1.2. Economic evaluation

Many economic impact evaluation methods are very difficult to use in the public sector simply because they require that effects (also intangible, e.g. improved health and quality of life) should be monetized (Kattel et al. 2013). Consequently, measures are faced with the problem of definition, which is usually based on values: e.g., how to measure research (by scientific publication), successful treatment (reoccurrence) or even deterred crime. For this, tolerance of multiple definitions has been suggested as a possible solution that could capture multiple values (de Bruijn, 2002).

Box 5.3. Evaluating to improve efficiency in New Zeeland

New Zealand's government has been using evaluation to improve efficiency of social invest by setting measurable outcomes (e.g. Welfare actuarial approach). They generate information to better understand peoples' needs (e.g. Treasury's Analytics and Insights team) and analyse what works for whom by applying cost-benefit analysis in budget. This information is used to purchase outcomes (e.g. Productivity Commission social services inquiry) and progress is to improve services and inform future investment decisions (e.g. Social Investment Agency).

Cost-benefit analysis

Cost-benefit analysis (CBA) is an established methodology in welfare economics that applies a systematic approach to estimating the strengths and weaknesses of alternatives (for example in transactions, activities, functional business requirements). At the core of CBA is an evaluation (ex-ante or ex-post) of the project socio-economic benefits and costs, all expressed in units of a welfare numeraire (usually money in present value terms) (Florio et al., 2016). A quantitative performance indicator (the net present value, or the internal rate of return, or a benefit/cost ratio) computes the net effect on society. Another approach based on CBA is the cost-effectiveness analysis, which compares the relative costs and outcomes (effects) of different courses of action, but it specifically assigns monetary value to the measure of effects. CBA and its derivative approaches are notoriously difficult in the field of innovation as (especially in the public sector) uncertain impacts of innovation and difficulty in measuring them are the rule rather than an exception.

Break even analysis

Break even analysis is usually applied in the private sector to determine the point at which revenue received equals the costs associated with receiving the revenue. Break-even analysis calculates a margin of safety – the amount that revenues exceed the break-even point. As in revenues are largely missing in the public sector it is difficult to apply. However, it has been in some cases been applied for evaluating technology development.

Social Return on Investment (SROI)

Social return on investment (SROI) is a performance measurement tool trying to capture social and economic value of projects. SROI can be evaluative, conducted retrospectively and based on actual outcomes that have already taken place; or a forecast, which predicts how much social value will be created if the activities meet their intended outcomes (Millar and Hall, 2013). It can also be used internally can also be useful internally as an instrument for organizational learning by enabling staff to analyse and improve their services (Ibid.). Yet, there is a high risk of error and poor quality information. The method by design assumes that organisations have good evidence base and financial proxies, but the reality might be very different. For example, it is very difficult to monetize soft outcomes. SROI also needs some assumptions and data about the counterfactual (what would have happened anyway), but this in cases of transformative innovations is very difficult to obtain.

Actuarial valuations

Actuarial valuations are another critical part of the social investment approach. This involves projecting future cash flows and relating them to the assets held to finance them and is used to project the liability for social benefits and inform priority areas for investment and intervention development.

5.1.3. Experimental design

In essence, experimental approaches to evaluation have 'secessionist' view – meaning that the approaches assume that causality cannot be seen, but only inferred from repeated succession of one such event by another (Pawson and Tilley, 1997: 5). Experimental designs try to exclude other conceivable rival cause from the experiment to be left with just one secure causal link. Due to high claims regarding external validity; internal validity ensured by random allocation; control and reference groups experimental and quasi-experimental are seen as highly reliable evaluation methods.

As such, experimental designs can be very useful for public sector innovation as they allow to test and trial new causal linkages between the intervention and results. However, practitioners need a better understanding when experiments are useful and what happens after an experiment is run. More discussion about the contextual effects and scale of experiments is needed as ill-defined directives and miss-conception about experiments may hurt the agenda. For example, experimental designs assume that the evaluator knows the goal (or purpose) of the innovation process, that it known before assessment, so, that it can be measured (Askim, Hjelmar and Pedersen, 2018). With innovation that may not always be the case.

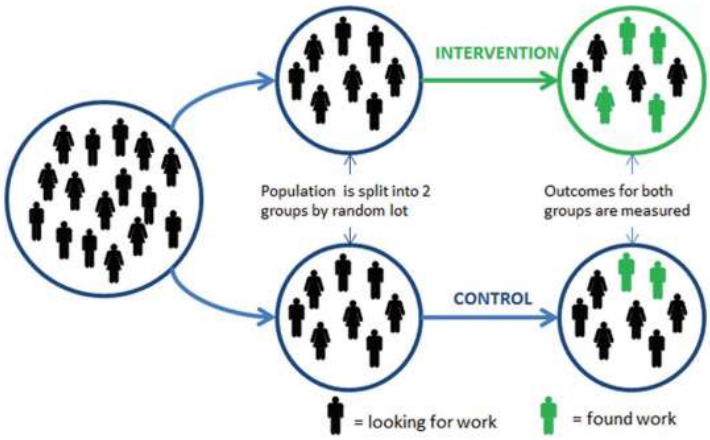
Many organisations in the field have come out with their frameworks for public sector experimentation. This is further confusing the field, between experimental study designs vs experimental mindset. For example, NESTA's has come forward with a draft framework

"Continuum of experimentation" which includes RCT's, quasi-experimental design, pilots, prototyping as well as direction shaping tools like horizon scanning and foresight. Experiments and experimental designs, however, are not prototypes and pilots and classifying them as such makes the concepts meaningless.

Randomised control trials (RCTs)

RCTs help to identify causal impacts of a programme by randomly assign participants to treatment and control groups (Figure 5.2). Threats to internal validity and selection bias are minimised by random assignment and the assurance that each group is approximately equivalent in their characteristics. In those conditions administering the intervention to the treatment group and comparing their performance to those in the control group differences can be safely attributed to receiving or not receiving the treatment. When it comes to experimental evaluations of effectiveness of policy interventions, randomised controlled trials (RCT) are considered to be the 'gold standard'. In this way they have become stronger legitimisers of changes within the public sector compared to standard statistical measures (incl. traditional public sector innovation indicators).

Figure 5.3. Basic design of randomised control trials



Source: Haynes et al., 2012.

Practise, however, has shown that RCTs are difficult to repeat with the same results in different context. There are also other issues with the approach (Cook et al., 2010). For sample, some characteristics – e.g., gender or race cannot be randomly assigned – and thus, questions relating to these variables cannot be methodologically answered. Moreover, RCT do not by themselves provide information about why intervention in some instances worked and why in others they did not, and how they worked. Therefore, to explain the efficacy of an intervention as well as to its process of implementation, a mixed-methods approach is necessary. Furthermore, in the public sector context it may be difficult (or sometimes even unethical) to form control groups especially in an evaluator setting.

Putting together robust RCTs also requires quite a lot of expertise. In the UK, where trails are most used also in the field of innovation, a cross Governmental Trial Advice Panel has been set up (Box 5.3).

Box 5.4. The UK's Cross-Government Trial Advice Panel

In 2015 the 'What Works' team in the Cabinet Office brought together a group of the top trialling experts from across Government together with twenty-five external academics (supported by ESRC) with expertise in experimental and quasi-experimental methods to provide a free to use service for all civil servants. The UKs Cross-Government Trial Advice Panel is a free to use service for civil servants bringing together top trialling and experimentation experts from across government and academia to provide to advise on all aspects of experimental and quasi-experimental design.

Source:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/599093/Cross_Government_Trial_Advice_Panel.pdf

Quasi-experimental methods

Where randomised control trials are not an option (e.g., due to issues with random assignment), 'quasi-experimental' approaches of randomisation can help. Quasi-experiments are different in RCTs in one element – participants are not assigned to groups randomly and thus groups cannot be not equivalent. Statistical techniques such as Ordinary Least Squares (OLS) and matching can be used to address in these circumstances. However, higher quality impact evaluations use identification strategies to construct a control group and then try to control for remaining differences in observable characteristics. The combination makes OLS or matching more powerful, as alone the concerns about the extent to which unobservable characteristics determine both treatment and outcomes and thus bias the evaluation remain.

Interrupted time series designs

Interrupted time series designs are another format of quasi experimental design (see Penfold and Zhang, 2013). The method involves a 'time series' of repeated observations of a particular event collected over time divided into 2 segments in the simplest case. The first segment comprises rates of the event before the intervention or policy, and the second segment is the rates after the intervention. Segmented regression is applied to measure the changes in level and slope in the post- intervention period compared to the pre-intervention period. In other words, segmented regression is used to measure immediate (level) changes in the rate of the outcome as well as changes in the trend (slope).

Regression discontinuity design

Regression discontinuity designs (RDD) were first introduced in the 1960 to the evaluation community The RDD contains both a treatment and a control group, but unlike quasi-experimental designs, the determination of group membership is known as the assignment follows a particular cut-off point. That is, in the RDD, participants are assigned to either a treatment or control group based on a particular cut-off that determines group membership (e.g., assigning students to control and treatment groups based on their reading comprehension scores). The threat to internal validity are historical factors. Furthermore, RDD is an extremely complex methodology and functional forms of analyses (linear,

quadratic etc.) can considerably influence the potential to determining true impacts (Figueredo et al., 2014).

5.1.4. Audits

Similar to other sources of information, ombudsmen reports and performance audits can provide public sector organizations with feedback information about important performance dimensions and the success of innovations. What makes audits different from general evaluations that they do not usually function in a client-contractor relationship and require a higher degree of evaluator's independence (even when they are part of the internal audit team) (see further in Van Acker, 2017). Audit teams conduct their work specifically with public accountability in mind and in many cases have access to high level accountability forums (parliament etc.) to present the results in. Thus, they have a higher chance in legitimising change when change is needed, but they can also illicit more fair of failure of innovations. Furthermore, as audits usually apply their own frames of reference – economic or otherwise – there is limited possibility to influence the questions, norms and standards connected to the work (also to take into account innovation needs).

Box 5.5. The role of National Audit Offices in public sector innovation

Various National Audit Offices around the world have been starting to input into the domain of public sector innovation. For example:

- In Australia the National Audit Office gave out a guidebook "Innovation in the Public Sector: Enabling Better Performance, Driving New Directions" in 2009.
- In Canada the Auditor General has been critiquing the public sector for its bureaucratic culture in light of big innovation failures.
- The National Audit Office in the UK released a report in 2009 on "Innovation Across Government"

Source: OECD compilation.

Audits can both concentrate on value for money, but also on performance. Evidence shows that internal auditing at least is steadily moving from traditional accounting and financial control towards operational control, risk management, and corporate governance issues (e.g., Arena and Azzone, 2009) and even project management (Arena and Jeppesen, 2016). Thus, it is predictable that internal audits at least in the public sector will start looking into the issue of innovation soon (see also Liston-Heyes and Juillet, 2018).

5.1.5. Logic modelling

Logic models (programme theory) outline inputs and activities and their anticipated outputs and outcomes. They help to identify evaluation measures and indicators. This entails identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumptions or risks that may influence success and failure.. Findings from this process subsequently can be fed back to the innovation process

in order to adapt and refine the intervention itself (Keller et al. 2009; Askim, Hjelmar and Pedersen, 2018). The problem, however, with logic modelling is that the intended ways in which outputs lead to outcomes are blurred in innovative processes. This does not mean that evaluatons based on logic modelling cannot be applied in an innovation context, but it imples a strong collaboration with the participants in the innovative project to strengthen the programme theory, make elements of the programme measurable, and conduct research during the innovative process to test the linkages between outputs and outcomes (Ibid.).

One can also draw theoretical assumptions from a practice itself. For example, Grounded Theory is one such analytic approach. The job of the evaluator/analyst is to systematically consider all of the data and to extract theory from the data (Strauss and Corbin, 1990). This in truly radical innovations may be a preferable method of analysis.

5.1.6. Theory-based approaches

Theory-based evaluation is a mode of evaluation that uncovers the underlying, assumptions about why a program, project, innovation will work (Weiss, 2000). Theory-based evaluations examine conditions of program implementation and mechanisms that mediate between processes and outcomes as a means to understand when and how programs work (Weiss 1997). Theory of change and realist evaluation are considered as specific schools under theory-based approaches.

Theories of change

Articulating theories of change (Weiss, 1995) is not simple – it is a collective and collaborative process (see Box 5.6; Mason and Barnes, 2007). It takes several iterations, political negotiations and conflict resolutions inherent in the process, a set of steps is outlined that, if followed sensitively, will result in a project's theory of change (Mackenzie and Blamey, 2005). Usually theories of change process is undertaken in the planning phase of the initiative. Thus, it is most useful in circumstances when objectives and activities can be identified and planned beforehand; or when there is a need to adapt in response to emerging issues and to decisions made by partners and other stakeholders.

Box 5.6. Undertaking a Theories of Change Evaluation

To elicit the theory of change underlying a planned programme, the evaluator works with a wide range of stakeholders in a collaborative manner. Part of the evaluator's role is to facilitate the articulation of the relevant theories and to highlight conflicting and discrepant theories. To help capture expectations of change, stakeholders are asked to focus explicitly on the following steps and to reflect on the contextual factors that influence their decision-making. These will involve the following steps:

Step 1: The focus here is on the long-term vision of an initiative and is likely to relate to a timescale that lies beyond the timeframe of the initiative. Its aim should be closely linked to the existence of a local or national problem.

Step 2: Having agreed the ultimate aim of the programme, stakeholders are encouraged to consider the necessary outcomes that will be required by the end of the programme if such an aim is to be met in the longer term.

Steps 3 and 4: Stakeholders are then asked to articulate the types of outputs and short-term outcomes that will help them to achieve the specified targets.

Step 5: At this stage those involved with the programme consider the most appropriate activities or interventions required to bring about the required change.

Step 6: Finally, stakeholders are required to consider the resources that can realistically be brought to bear on the planned interventions. These will include staff and organizational capacity, the existence of supportive networks and facilities as well as financial capability.

Following a collective and iterative process the resulting programme theory must fulfil a set of pre-specified criteria: that it must be *plausible*, *doable* and *testable*.

First then, the theory of change that is elicited should be interrogated to ensure that the under lying logic is one that is acceptable to stakeholders either because of its existing evidence base or because it seems likely to be true in a normative sense.

Second, the implementation theory itself should be questioned to ensure that timescales, fi nancial resources and capacities add up to the aspirations of the programme.

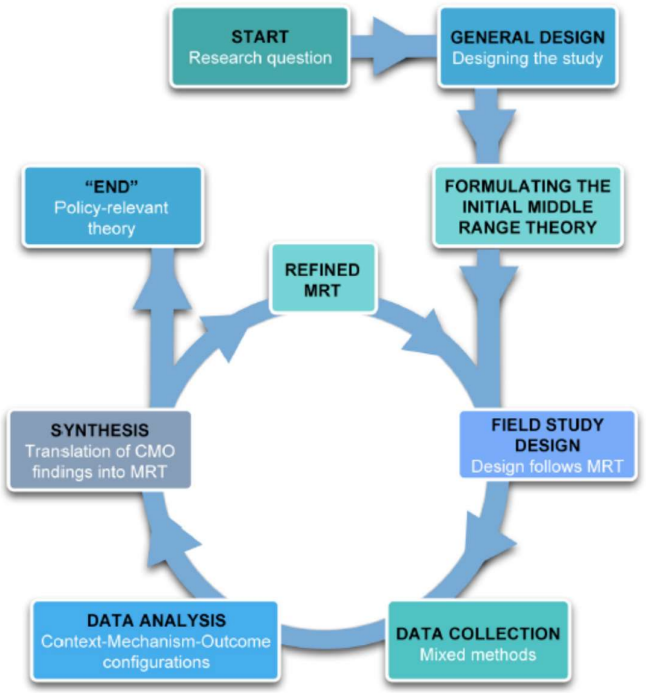
Finally, the Theory of Change needs to be articulated in such a way that it can be open to evaluation; this is only possible where there is a high degree of specificity concerning the outcomes of the programme. The proponents of the approach provide examples of overall programme theories. A Theories of Change evaluator then takes the programme map generated through this process and, using standard multi-method approaches as relevant, monitors the unfolding of the programme in practice.

Source: Blamey and Mackenzie, 2007.

Realistic Evaluation

Realistic approach is another form of theory-based evaluation which is based on the foundations of critical realism (Pawson and Tilley, 1997). The account of the processes that explain how an intervention leads to a particular outcome is formulated as a middle-range theory (Marchal et al., 2012). The process is described in figure 5.4. The main assumption of the approach is that it is not good enough to establish what works, but what works in which circumstances and for whom are questions that the evaluators need to answer. Thus, the approach assumes that social phenomena have multiple and conjunctural causes. Realist evaluators consider causality to be generative in nature, i.e., believe that actors have a potential for change by their very nature. This is in line with the contextual nature of innovation.

Figure 5.4. The realist evaluation cycle



Note: MRT in the figure denotes 'middle range theory,' CMO denotes the main realist evaluation imagining tool context-mechanisms-outcomes.
Source: Marchal et al., 2012: 196.

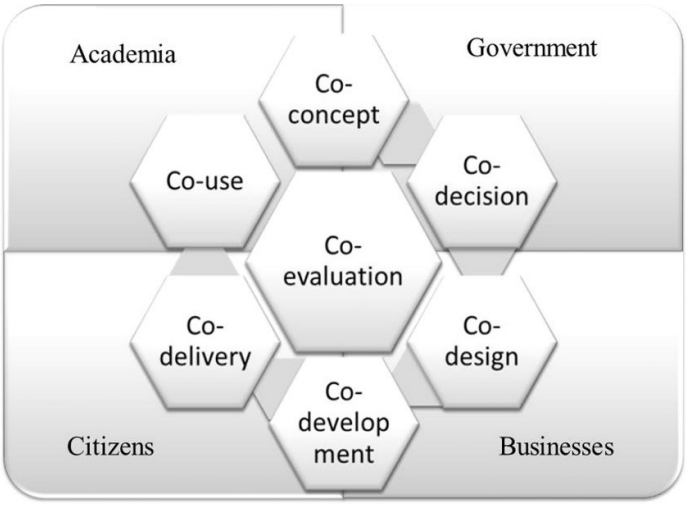
5.2. Collaborative and user-centric evaluation

Collaborative innovation is on the rise and it is reliant on interdependencies across actors and boarders. Collaboration is also a key in transfer of tacit knowledge and building collective know-how (e.g., Scarbrough, 2003)Yet, involving citizens in co-production – as a collaborative process of service development – and in co-evaluation – as an open innovation assessment method – is far from easy (Paskaleva and Cooper, 2018). In practice, public managers may not have all of the skills and capabilities to do it they need to do this effectively. Mandated participation has severe limitations (Cabaj et al., 2015; Wilson-Grau et al., 2015). There are not many tested methods or tools for effectively co-evaluation as an inclusive assessment method. Yet the field is emerging and there is more talk around also empowerment evaluations (Fetterman et al., 1996). The first step to applying participatory evaluation is to start using user-centric evaluation (with the help of design thinking, ethnographies etc.). It ensures interventions respond to user needs, and that evaluation focuses on meaningful questions to users (Gripper et al. 2017). At the same time, it requires deep culture change to properly implement, so susceptible to lip service or tokenism (Ibid.)

Especially, as more and more citizens are engaged with government decision making especially on the city level. Part of the focus of co-evaluation here is also self-assessment In its concerns for issues as ownership, relevance, involvement, improvement, co-evaluation corresponds to other evaluation approaches – participatory, collaborative,

stakeholder-involving, and utilization-focused evaluation (Patton, 2005). As such, co-evaluation is seen as part of open innovation models (Figure 5.3).

Figure 5.5. Open innovation model of smart city service development.



Source: Paskaleva and Cooper, 2018.

Citizens perspectives within a collaborative setting can be captured in a variety of ways (e.g., Box 5.6). These include also customer journey mapping: journey maps are visual depictions that describe either an individual's unique life path or experience within a specific service context. Journey maps are rooted in Human-Centered Design, a practice that emphasizes empathy and understanding; they can lead to deep connections between individuals.

Box 5.7. Capturing casual moments

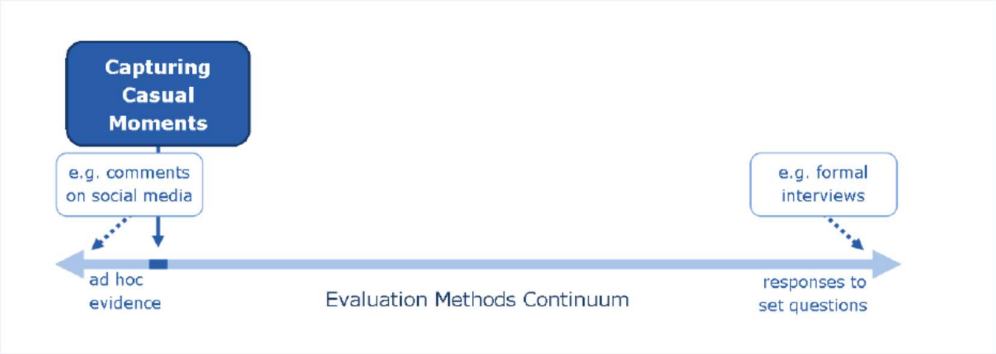
When staff or volunteers want to record comments in non-formal circumstances. This method is a simple means for writing down service users' actual words.

Useful for:

- Gaining honest responses when the pressure is off
- Following up in a more formal setting e.g. one to one support session
- Contributing to a good mix of evidence
- Getting the attention of your audience

This can be followed in the continuum (Figure 5.3) of capturing emotional touchpoints, which help to identify key points in people's experiences of services and is then used to prompt feedback or stories from service users. It can also act as a framework for collecting stories.

Figure 5.6. Evaluation continuum.



Source: Evaluation Method: Capturing Casual Moments. Available at: http://evaluationsupportscotland.org.uk/media/uploads/resources/method_-_capture_casual_april_17.pdf

Source: http://evaluationsupportscotland.org.uk/resources/355/

5.2.1. Joint and participatory evaluation

Participatory evaluation is an evaluation method in which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation. They enable more inclusive monitoring and evaluation processes. Participatory evaluations allow comparison and faster accumulation of knowledge. Saves resources and supports sector level change (Gripper et al. 2017). These approaches, however, need flexible shared tools and time to establish interventions across boarders. These approaches may facilitate inter-organizational learning, which sometime can be highly innovative and explorative in nature (Holmqvist, 2003). However, uncertainty is also present in these evaluation, especially when a degree of control rests with participants (Mathie and Peters 2014).

5.2.2. Most significant change

'Most Significant Change' (MSC) technique was created by the international development industry. In essence, MSC is a participatory monitoring technique that involves the collection of 'significant change' stories from the field that are systematically gathered and selected by groups of stakeholders and staff (Davies and Dart 2005). The approach is useful when the focus of evaluation and monitoring is on learning and not Marely accountability and also when the outcomes themselves are difficult to quantify (Willetts and Crawford, 2007). This might occur in complex situations where outcomes are diverse. One could argue that storytelling techniques used by innovation labs in the public sector (Tõnurist, Kattel and Lember, 2015) also follow the MSC logic, albeit to a less rigorous level.

MSC technique captures broad range results (intended/unintended, positive/negative) and makes diverse perspectives explicit. In addition, it allows to showcase the perspectives of different stakeholder groups as they are part of determining the criteria for a significant

5.3. In project reflection

Regular, reflective practice sessions build trust and understanding (McKegg et al., 2015; Zamir and Abu Jaber, 2015), provide a 'safe place' for critical thinking (Gamble et al., 2015) and address sources of stress and frustration among team members (Togni et al., 2015). In project reflections can also address tacit learning needs in organisations as tacit knowledge can be acquired by practical experience, including observation, imitation and learning by doing (Hartley and Rashman, 2018). Lifecycle report on diffusion (OECD 2018) in the current series covered ritualised in-project reflection. For some, institutional reflexivity is a key precondition for organizational learning and sustainable modernization (Kuhlmann and Bogumil, 2018). There are many ways to facilitate in project reflection for evaluation purposes (incl. active self-evaluation).

5.3.1. Monitoring

Monitoring is a continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds. (OECD 2011).

Table 5.3. Instrument for measuring the public innovation capacity

Function	Statements for self-assessment
Mobilizing	M1. Employees in *City X* with ideas about *data-driven innovation* easily find the right persons in the city to jointly realize these ideas.*
	M2. The people in charge of *data-driven innovation* in *City X* succeed in engaging companies, researchers and citizens in the development of new ideas.
	M3. *City X* has a strong structural network of companies, researchers and citizens connected to *data-driven innovation*.
	M4. The people in charge of *data-driven innovation* in *City X* succeed in stimulating the development of new ideas among colleagues in *City X*.
	M5. *City X* has a strong network of employees with an interest in *data-driven innovation*.
	M6. A company, researcher or citizen with good ideas for *data-driven innovation* easily finds the right persons within *City X* to develop these ideas further.
Experimenting	I1. *City X* is successful in setting up experiments.
	I2. *City X* has societal support (from citizens, NGOs, companies, etc.) for experiments on *data-driven innovation*.**
	I3. Political institutions in *City X* – representatives, aldermen – support experiments with *data-driven innovation*.
	I4. The administrative executives of *City X* support experiments with *data-driven innovation*.
	I5. *City X* makes sufficient funds available for experimenting.
	I6. If necessary, *City X* engages other governments, companies and societal organizations in experiments around *data-driven innovation*.
Institutionalizing	R1. *City X* is successful in scaling up experiments.
	R2. *City X* adopts *data-driven innovation* that have proven to be successful on a small scale in the organizational routines.
	R3. *City X* evaluates experiments with *data-driven innovation* well.
	R4. *City X* succeeds in turning experimental collaboration with governments, companies and societal organizations into structural forms of collaboration.
Balancing	B1. *City X* succeeds in identifying risks, disadvantages and tensions around *data-driven innovation*.
	B2. *City X* initiates the public debate about the risks, disadvantages and tensions around *data-driven innovation* and how to deal with these.
	B3. If there are conflicts, *City X* is good at mediating conflicts around *data-driven innovation*.
	B4. In *City X*, ethical aspects of *data-driven innovation* are discussed well.
Coordinating	C1. *City X* makes financial means available for *data-driven innovation* on a structural basis.
	C2. There is a good exchange of information on *data-driven innovation* between all actors in *City X*.
	C3. *City X* has a culture that stimulates *data-driven innovation*.
	C4. *City X* creates the right conditions for *data-driven innovation* (training, information exchange, instruments, etc.).
	C5. *City X* has a clear vision on *data-driven innovation*.
	C6. Political institutions in *City X* – representatives, aldermen – are prepared to allocate financial means in the budget for *data-driven innovation*.

Source: Meijer 2018.

5.3.2. Learning agendas

A learning agenda is a set of questions, that identify what needs to be learned before a project can be planned and implemented. Once the questions are identified, a learning agenda also prioritizes and establishes a plan to answer short- and long-term questions of the highest value across relevant program and policy areas. Learning agendas have been actively applied in the Office of Management and Budget in the US. Learning agendas can:

- systematically identify gaps in their knowledge and conduct research and evaluation to fill them;
- organize research and evaluation efforts within budget and program timeframes;

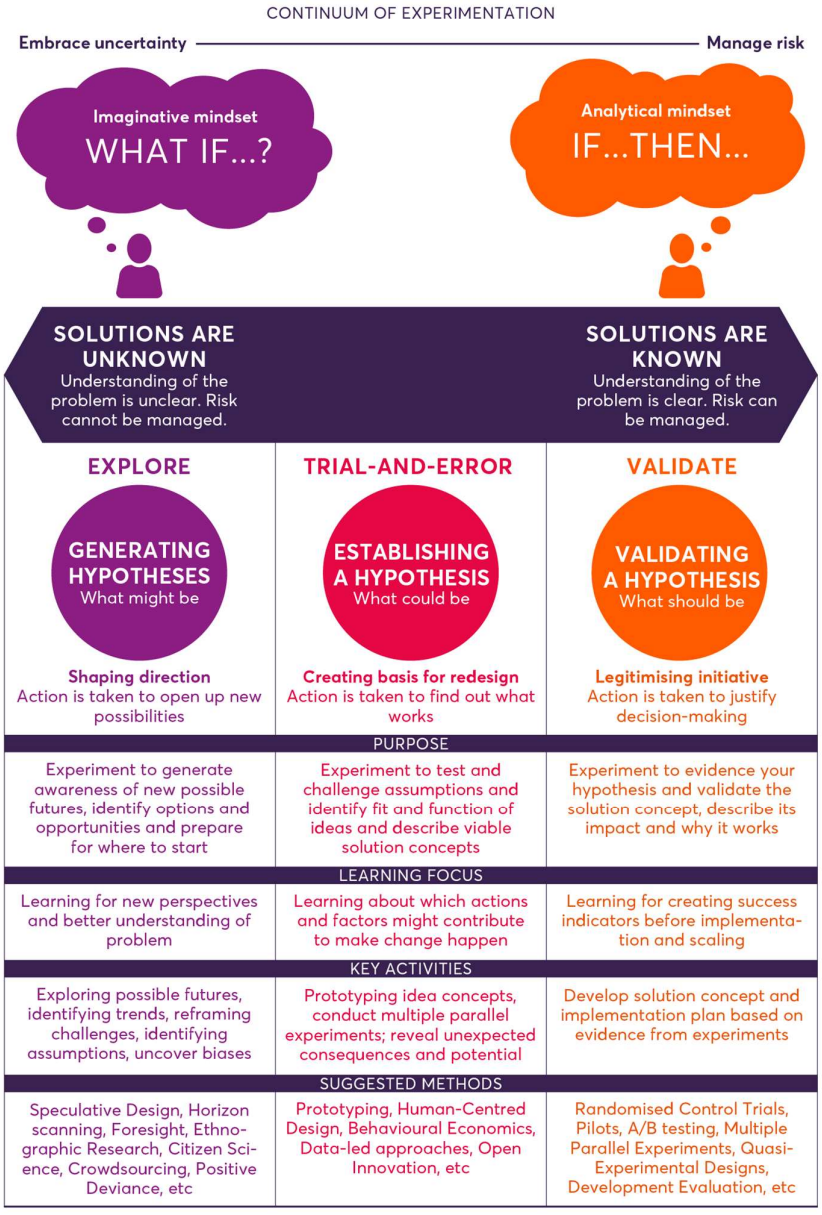
- align research and evaluations with their missions and legal responsibilities;
- coordinate the different types of evaluations and research conducted by various offices;
- create an environment that encourages individuals and offices to learn from their evidence and
- from others, which can lead to adaptation and innovation;
- foster collaboration and evidence-sharing internally and with other agencies and external
- stakeholders; and
- reinforce organizational change to more effectively build evidence and conduct high-quality,
- rigorous research and evaluations.[7]

5.3.3. Rapid experimentation

Most reports or audits of public sector innovation contain rapid experimentation or agile processes as one of the many recommended actions, particularly in regard to evidence-based policy (Staley, 2008, Banks, 2009; Potts and Kastelle, 2010). These are working approaches that are common to lean start-ups. Rapid experimentation is not about robust research designs, but shorter feedback looks to learn faster. As such, the approach fits the more pragmatic, relevant evidence approach which also is described by the NESTA's experimentation continuum (Figure 5.7).

[7] See the Evidence toolkit. Available at: https://www.urban.org/sites/default/files/publication/97406/evidence_toolkit_learning_agendas_2.pdf

Figure 5.7. NESTA's continuum of experimentation

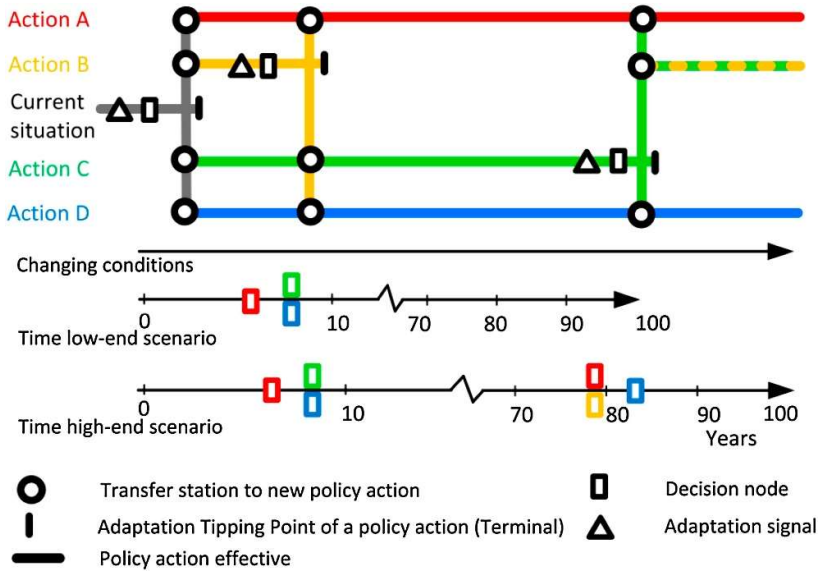


Note: NESTA's contribution has been inspired by 'Continuum of experimentation' (inspired by multiple resources including Danish Design Centre's 'Designing policy experimentation' and Donald Schön's 'The Reflective Practitioner'.
Source: NESTA (2018)
https://media.nesta.org.uk/documents/experimentalculture_continuumofexperimentation.pdfSpecialisation indices

5.3.4. Adaptive pathways

Adaptation pathways stress the notion that, at various points in time, different decisions are possible, whereby path-dependence links different decisions in a longer-term pathway (Hermans et al., 2017). These policy decisions are visualised in a dynamic sequence (Figure 5.5). Pathways approaches differ from more conventional planning approaches, as they include different alternative adaptation options over time, which are triggered by pre-specified conditions (Ibid.). This implies a forward-looking flexible character of adaptation pathways plans. Adaptive analytical approaches can reduce the sample size, time, and cost required to obtain decision-relevant information (Luce et al. 2009).

Figure 5.8. Key concepts in adaptation pathways approaches



Source: Hermans et al. 2017.

5.4. Automitised/Digitalised evaluation tools

New technological solutions are continuously improving the collection, availability and use of evidence. Public opinion can be gathered from social media using web-scraping and sentiment analysis (Serrano-Guerrero et al., 2015). Open data movements in government are making also more data available and this facilitates more collaborative practices in developing new services, but also around evaluation of the latter (Engin and Treleaven, 2018). Thus, new forms automated and digitalised evaluations are emerging.

5.4.1. Big data and text mining

Big data analytics are based on examining very large data sets to uncover hidden patterns, unknown correlations, etc. (Engin and Treleaven, 2018). Usually, these data sets are so complex that traditional data processing application software is not enough to utilise them properly and thus, various other technologies (incl. big data lakes, cloud computing, NoSQL database, Hadoop programming, in-memory analytics) are deployed. The vast amount of administrative data governments collect combined with (real time) data from social media, cameras and sensors make big data based evaluations extremely important

for the public sector in the future (Giest 2017). Big data allows individual data-points to be considered collectively and also in comparison to each other, either geographically, demographically or behaviourally (Tate et al., 2018). This enables the expansion of behavioural/predictive analytics, which rely on the analysis of large and varied data sets to uncover hidden patterns, unknown correlations, user preferences, etc. to help make informed decisions (Engin and Treleaven, 2018). This can provide new insight into the events, life experiences, and patterns and trends in society as digital signals (Kowalkiewicz, Safrudin and Schulze, 2017). The more the capacity of predictive analytics approves, the more it will start to challenge traditional forms of evaluation. In essence, predictive analytics 'forecast' what might happen in the future with an acceptable level of reliability, and includes what-if scenarios and risk assessment (Tate et al., 2018). For example, learning analytics platforms capture data from children's educational activities to track and algorithmically optimize their educational experience; predicting the future performance of the system and the student (Williamson 2016).

Furthermore, already now, different solutions and approaches connected specifically to evaluation are applied in this space in practice. For example, Ceron and Negri (2016) have used Supervised Aggregated Sentiment Analysis for ex-post evaluation of policies (school and job market reform) and measured the reaction of online public opinion on policy alternatives and monitoring mobilization of opposition groups. Law enforcement and security services are at the moment at the forefront of utilising big data analytics (e.g., the Los Angeles Police Department (LAPD) has invested heavily in its data collection, analysis, and deployment capacities and uses the platform Palantir to use big data better (Dunleavy, 2016, 15)). U.S. Army has developed its Automated Continuous Evaluation System which uses big data analytics and context aware security to investigate government, commercial, and social media data to uncover patterns of applicants (Höchtl, Parycek, and Schöllhammer,. 2016).

These are all steps towards algorithmic evaluation (with real-time analytics; behavioural and predictive analytics), which potentially comes at cost: this, first, to privacy as citizens may not know when they come under the prevue of government nor can government explain why algorithms have flag an interest in them or made a particular call on their specific case.[8] These issues increase with the move towards more artificial intelligence based evaluations and knowledge based systems (these are wither rule-based or case-based automated reasoning approaches). Machine learning, natural language understanding and sentiment analysis are already important for public opinion monitoring, policy and fraud detection.

Remote sensing and geospatial analysis

Also, Internet of Things (IoT) is the inter-networking of 'smart' physical devices, vehicles, buildings, etc. that enable these objects to collect and exchange data. This can also be used for new types of continuous evaluations. For example, law enforcement now uses Automatic License Plate Readers (ALPRs) to record cars and their GPS coordinates and immediately compare them to outstanding warrants and organise geo-located responses (Brayne, 2017). On a wider scale, participatory evaluations of new developments are already happening in cities under smart city and living lab initiatives (Schuurman and Tõnurist, 2016).

[8] Read here about the recent controversy in Denmark: https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/

5.4.2. Data dashboards and interactive data visualization

Another influence of new technologies are on new digital and interactive data visualizations (Giest 2017), thus, also evaluation data and indicators can be communicated in a different, often interactive way. US cities have been vanguards of these approaches and have created various machine-readable administratively collected data platforms (Merge, Kleibrink and Sörvik, 2018). Nevertheless, these new data platforms around core government outcomes are also emerging in Europe (Box 5.9).

Box 5.8. Urban Data Centres (the Netherlands)

In 2016, CBS (Statistics Netherlands) started to develop Urban Data Centres (UDC) by combining national data and data expertise with smart, data-driven city needs. The Dutch city of Eindhoven jointly developed an UDC with CBS. The centres are built around the city's interests and needs – smaller towns and big metropolises variably have different interests – by combining national survey, administrative and big data with city data. After launching the first CBS Urban Data Centre, seven additional centres were established in just one year. The concept can also be adapted to and implemented in developing countries, and can contribute to the realisation of the SDGs.

As such, CBS combines its considerable data expertise with real-life urban problems and city policy knowledge. As a federal body, CBS works to support cities through the provision of expertise that cities often lack. The resulting jointly developed Urban Data Centres help to better understanding the current situation and problem dynamics in a city. The centres create location and problem-specific data-driven input for local policy making that can lead to transformative change. Thus, this is not only about getting data onto the digital platform, but defining issues and problems together and finding ways to make them visible and actionable for local governments.

Source: read more in OPSI, *forthcoming*. Transformation of Public Value Cities as the Playground for the Future.

5.4.3. Citizen experts and P2P production of evidence

The last, but not least, trend in digitalisation and evaluation is the increased peer-to-peer production of evidence and also evaluation. Peer production describes a special kind of production system where individuals act in response to their own needs and interests in a decentralized way (Navarro, 2016). These advancements are facilitated through the spread of distributed ledger technologies (e.g., blockchain) and the plethora of citizen engagement platforms. Various tools and platforms such as CitySourced, FixMyStreet, Monithon, Where's my Villo? etc. were designed to allow citizens directly report their issues to the government, beyond generic satisfactions surveys. Yet also more passive, crowdsourced data is being used for evaluation purposes (e.g., mobile positioning data) (see a broader overview of digital co-production in Lember, Brandsen, Tõnurist, forthcoming).

5.5. Future-oriented evaluation tools

Bottom-up and knowledge scanning agencies out-perform policy-dependent agencies on innovation outcomes (Arundel, Casali and Hollanders, 2015) meaning that future-oriented evaluations and monitoring are key to long-term innovation success. This especially when connected to anticipatory innovation. Most of the future-oriented methods and approaches are used as forms of ex ante evaluation. They most often are based on qualitative measures (but also quantitative approaches exist, e.g., simulation methodologies, agent-based modelling are used to evaluate policy alternatives before implementing them), which aim to determine which programmes provide the greatest benefit to stakeholders. Expert opinion and review can also be used in ex ante evaluation.

5.5.1. Risk analysis

Risk analysis is an assessment of factors which affect or are likely to affect the successful achievement of an intervention's objectives. A detailed examination of the potential unwanted and negative consequences to human life, health, property, or the environment posed by development interventions; a systematic process to provide information regarding such undesirable consequences; the process of quantification of the probabilities and expected impacts for identified risks.

5.5.2. Strategic evaluation

Strategic evaluation is concerned with determining the effectiveness of a strategy in achieving the organisational objectives and taking corrective actions wherever required. Strategic evaluation draws on techniques such as road-mapping and strategic policy intelligence. Road-mapping is a strategic planning technique that is integral to creating and delivering strategy and innovation by matching and long-term goals with specific or potential solutions. Strategic policy intelligence is a set of actions to search, process, diffuse and protect information in order to make it available to the right person at the right time in order to make the right decision maker (Strategic Policy Intelligence 2001)

5.5.3. Foresight evaluation

Foresight and analysis of past experiences are useful for developing policy alternatives. It has been examined by a variety of scholars (e.g., Rijkens-Klomp and van der Duin, 2011) and various evaluation frameworks have been developed through several Foresight evaluation engages with factors of foresight success, areas of foresight impact, and different aspects of the foresight process (Makarova and Sokolova, 2014). Part of foresight evaluations are the future reviews implemented in the Finnish government (Box 5.11).

Box 5.9. Future reviews in Finland

The futures reviews of the ministries describe Finland's key questions in the years ahead. Their purpose is to assess situations and developments in society and examine issues for political decision making connected to the future. The aim is to generate public debate and provide information for the forthcoming elections and government formation talks. There have been five iterations of futures reviews: the earlier ones were published in 2003, 2006,

2010 and 2014 and the last in 2018. The ministries' foresight working group coordinates the drafting process and has members from each ministry.

Source: Hallinnonalojen tulevaisuuskatsaukset. Available at: https://vnk.fi/tulevaisuuskatsaukset; https://valtioneuvosto.fi/en/artikkeli/-/asset_publisher/10616/tulevaisuuskatsaukset-suomi-sopeutuu-omin-vahvuuksin-globaaliin-muutokseen

5.5.4. Horizon scanning

Horizon scanning is used to identify emerging issues and challenges and to contemplate how they might evolve and interact with other characteristics of the system. If used well, it can help organisations engage with uncertainty and recognise that their assumptions are just that – assumptions. Such a perspective can help engender more flexibility in thinking and in outlook. Horizon scanning systems also seek to identify, monitor, and evaluate new and emerging technologies and trends (together with technology assessments) which are essential to anticipatory innovations.

5.6. Systems-based approaches to evaluation

The systems-based approach to evaluation takes its departure in the challenges faced when dealing with the open-ended nature of problems and issues including innovation and the goal complexity of the connected processes (Williams and Iraj, 2007; Askim, Hjelmar and Pedersen, 2018). Innovation evaluation from a systems perspective has to evolve with the intervention – be reflective in action. Also insights from complexity theory are often drawn in with the concept of 'adaptation' which helps evaluators to reconstruct a programme theory of change from understanding that resources are not the final constraint nor is optimal allocation known in advance (see Reynolds et al., 2016 for a very good overview of systems approaches to innovation evaluation). OECD has been also moving towards a systems approach to public sector innovation and has developed a model to look at innovation activities from an individual, organisational and systemic lens (Figure 5.9 below).

Box 5.10. The three orientations of systemic evaluation

Engaging in a systemic evaluation of a situation requires three orientations (Williams and Hummelbrunner, 2011; Reynolds 2016):

- *Reflectiveness*: being aware (and checking the validity) of assumptions, mental models and values – and how they affect what we see or hear. Evaluations should be conceived as reflective practice processes consisting of successive reflective loops.
- *Respect and trust self-organization*: paying due attention to emergence and value differences from plans, as these can provide useful clues for improvements. Evaluations should look at the entire range of processes triggered and beyond original intentions.
- *Think-act-think circularity*: enacting non – linear praxis rather than linearity: Because many interventions are still predominantly conceived in a 'linear' fashion,

evaluations should transform them into 'circular' ones, i.e. structured as a system and linked in a recursive logic to the relevant operational context.

Source: Reynolds 2016.

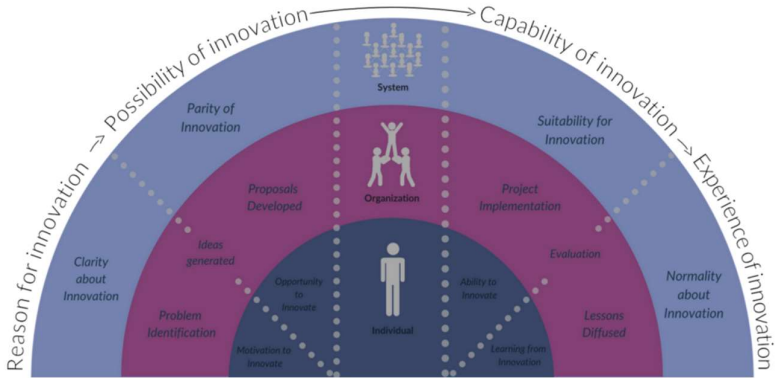
Thus, it is not experimental, but heuristic – it is geared towards practical problem solving, discovery and learning. In a systems perspective knowledge generated from evaluations is valid as long as it useful and used by the stakeholders in the system. These are fairly significant differences compared to traditional impact assessment (these are further outlined in Table 5.4).

Table 5.4. Two Basic Approaches to Evaluate Public Sector Innovation

	Impact-focused approach	Systems approach
Purpose	Solid knowledge about effects and causality	Generate new practices; nurture innovation and learning
Participants	Controlled selection, few in number, randomisation, strategic sampling, control group	Self-selection, many participants, no control group
Contents	Narrow scope and accommodation	Wide scope; emergent, liberal accommodation
Evaluation	Comprehensive, external party, effect-oriented	Inspiring depictions, context-sensitive narratives, process-oriented

Source: Askim, Hjelmar and Pedersen, 2018.

Figure 5.9. Determinants of public sector innovation.



Note: This model was developed as part of the review "Innovation System of the Public Service of Canada" (OECD, 2018)
Source: OECD.

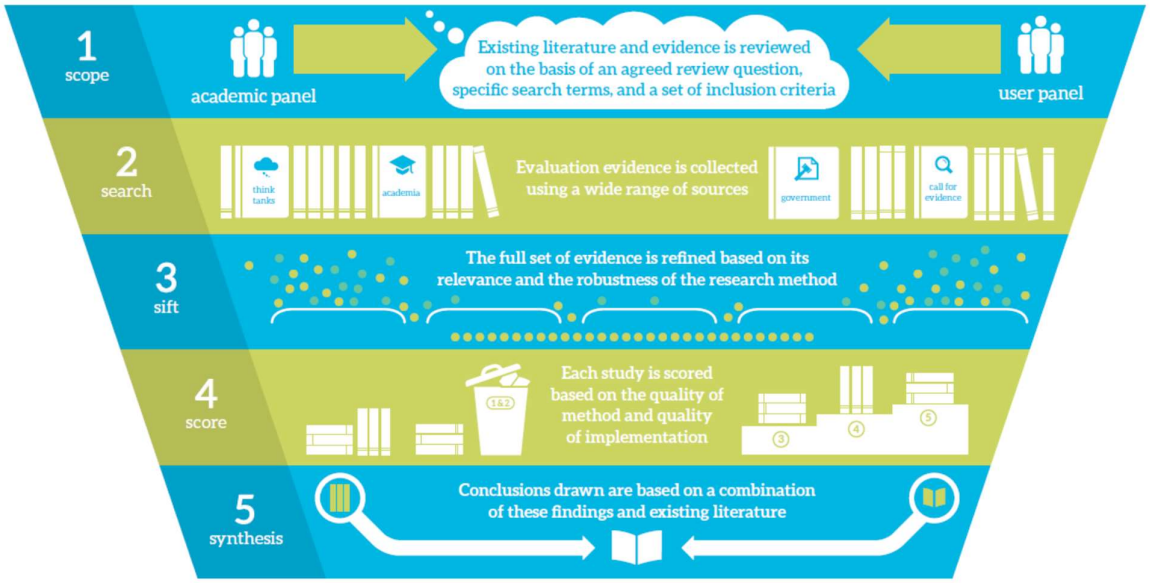
5.7. Systematic reviews and meta-evaluations

One should not confuse systems approaches with systematic reviews. The latter is a summary of existing evidence that answers a specific question and contains a thorough, unbiased search of the relevant literature, explicit criteria for assessing studies and

structured presentation of the results (see figure 5.10). Meta-evaluation are evaluations designed to aggregate findings from a series of evaluations. Meta-evaluations differ from systematic reviews as they usually incorporates quantitative pooling of similar studies to produce an overall summary of treatment effects. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators (OECD, 2011). Systematic reviews and meta-analyses provide a useful and convenient summary of existing knowledge in the field of study, but they are also of varying quality. They are predominantly used in medical sciences, but their use has risen recently also in social sciences and some systematic reviews have been applied to public sector innovation as well (e.g., De Vries, Bekkers and Tummers, 2016).

Figure 5.10. What Works Centres systemic review methodology

The five stage process.



Note: The What Works Centres have developed their own methodologies for conducting systematic reviews.
Source: A What Works Centre for Local Economic Growth, 2014.

Another form of systematically collecting and gathering and validating innovations are in the format of evidence clearinghouses (Box 5.11). Usually clearing houses are used topically to review existing research on different programs and products connected to specific fields or policy problems. Usually clearinghouses create searchable datasets for research evidence and give guidance on critical use of evidence in using critical evidence for decision making to choose, implement and sustain interventions. Not all evaluations are equal: some are designed and conducted better and thus, more meaningful for decision-making – not all evaluations are equal. As such, most importantly, clearinghouses should validate the trustworthiness of evidence itself – be curators of knowledge – which also means that the rigor of their work should be also critically reviewed.

Box 5.11. Examples of evidence clearinghouses

Evidence/research clearinghouses help users in the public and non-profit sectors identify effective programs and interventions by conducting systematic literature reviews and assessing programs based on the evidence that exists of their effectiveness. The findings are usually made available through online web portals.

Some examples include:

- The California evidence-base clearinghouse for child welfare
- What Works Clearinghouse for education programs
- Ohio's evidence-base clearinghouse for education
- The National Registry of Evidence-based Programs and Practices on mental health
- and substance use interventions
- The Pew-MacArthur Results First Initiative
- Clearinghouse for military family readiness

Source: Davies and Silloway, 2016.

5.8. Triangulation and mixed methods approaches

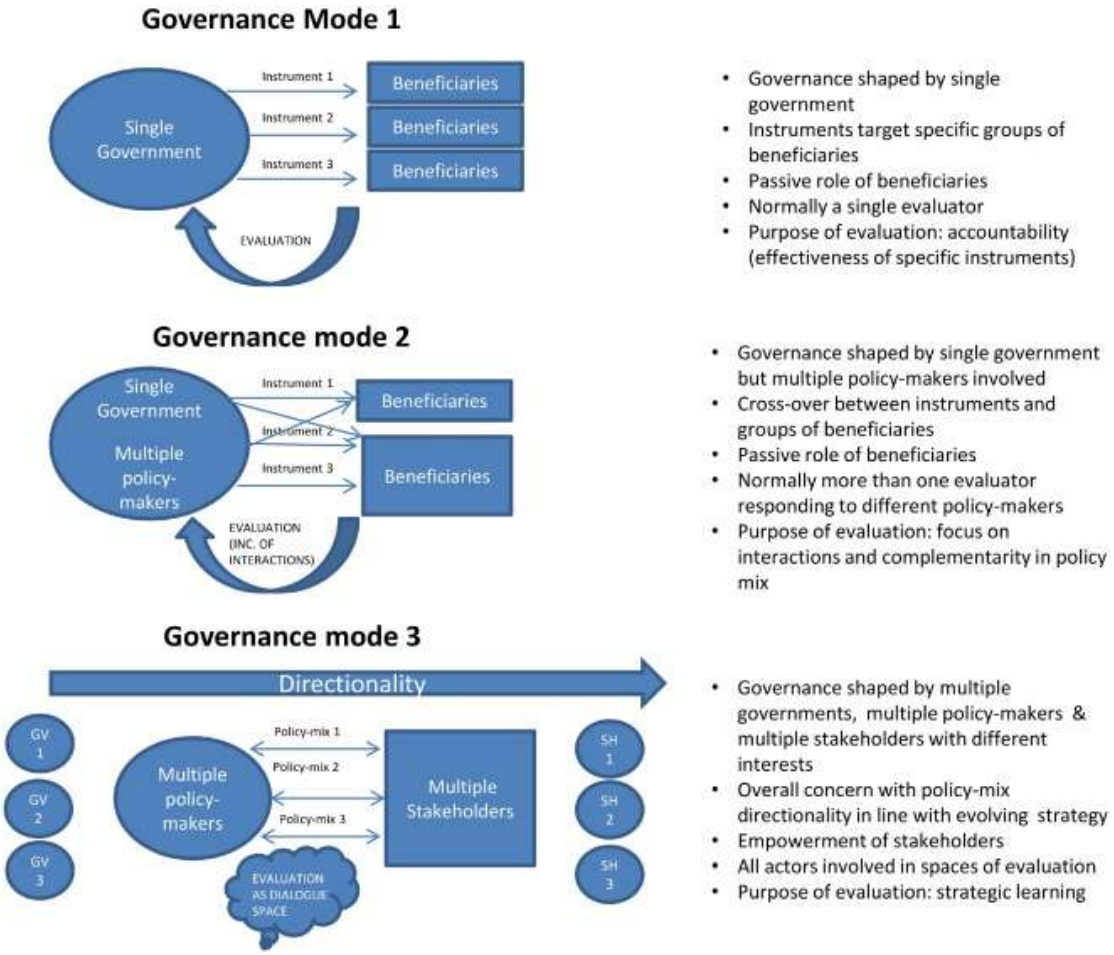
A mixed methods approaches to evaluation are useful because they allow for cross-validation (triangulation) of results and complementary effects, where the relative weakness of one method becomes the relative strength of the other (Little, 2013). Here the process of triangulation can come in handy. Triangulation is the use of multiple theories, sources or types of information, or types of analysis to verify and substantiate an assessment (OECD, 2011). There is a need for multi-method approaches and triangulation of data, as innovation are influenced by policy mixes rather than linear policies. However, moving from individual policy evaluation to policy-mix evaluation can be arduous (table 5.6.) and also may require different governance responses (figure 5.8; Margo and Wilson, 2013; Margo et al. 2014).

Table 5.5. Summary of challenges for policy mix evaluations

Characteristics of policy-mixes	Challenges related to evaluation governance
Combination of directional and neutral instruments	Deal with plurality of vested interests (and existence of policy silos)
Multi-level or vertical dimension	Deal with coordination of evaluation information across levels of government
Promoting experimentation	Able to detect and integrate into evaluation practices impacts on actors behaviour and structural systemic changes in the context of a dynamic EDP
Multiple actors as both policy-makers and beneficiaries	Able to support government and other actors in developing the capabilities required to adapt the policy-mix to the evolving strategy

Source: Margo and Wilson, 2018.

Figure 5.11. Three governance modes for policy mix evaluation



Source: Margo and Wilson 2018.

6. Conclusions

"*Life is pretty simple: You do some stuff. Most fails. Some works. You do more of what works. If it works big, others quickly copy it. Then you do something else. The trick is the doing something else*." (Peters, 1994, Chicago Tribune)

Evaluation and public sector innovation have not merged their agendas. Apart from a few forerunners, there are no public sector innovation specific evaluation tools and methods. Yet, the evaluation literature itself is fast, comprehensive and complex. Should there be public sector innovation specific approaches? Maybe not. Innovation by now is pervasive, it has entered all government activities and programmes; thus, rather than developing public sector innovation specific tools the evaluation methods themselves or their application as a whole should change and innovation as a concept should be applied to the broader practise of evaluation.

It is important to note and acknowledge the conflicting expectations, goals, paradigms and values connected to evaluation processes themselves, esp. as digitalisation is going to fix or exacerbate many of them. What this report throughout has tried to do, is to highlight the inherent biases and adverse effects of evaluation approaches and combine the latter with the complexity and non-rationality of the innovation process. The results of the mix are sometimes very difficult to predict, yet, it should be one of the guiding questions when choosing evaluation methods not only for innovation, but for broader policy problems. Of course, not all established methods of evaluation are not going to be reformed nor is there need for it. All approaches to project and programme evaluation have advantages and disadvantages when it comes to innovation that in some cases are specific to different innovation facets. Thus, a reflexive mixed methods approach and triangulation of data would suit innovation the best. As innovation itself cannot be meticulously planned nor can the evaluations of the former – they need to evolve as the aims and application of innovation evolves.

Primarily, however, practitioners need to understand why innovation evaluation is undertaken in the first place. As outlined in Chapter 2 evaluation has a variety of aims: it can help develop and improve innovations as it facilitates learning processes, it can create room for innovation by showing the inadequacy of the current situation, but it can also inform people of the worth and significance of innovation itself. Evaluations, thus, support oversight and accountability which public sector innovation cannot escape. Different types of evaluation users may have simultaneously different needs and requirements for the evaluation, so, it is important to know, what the reason to evaluate innovation project is. Otherwise, evaluations will face continuous scope creep by diverging demands and interest. Thus, the aggregate needs of accountability forums and 'local' learning needs of innovation developers and users need to be balanced or put at the forefront in applying different approaches.

What public sector innovation community does need to move beyond are success stories and positive message case studies. Evaluation in this form was essentially a communication tool to justify and legitimise public sector innovation in the government context, but the agenda has moved beyond that and more robust and generalizable (if possible) evaluations are required. Most importantly, they are impeding learning in the public sector.

7. Remaining issues

This is an alpha version of a study – i.e. it has been developed to seek input and test various ideas and features. In that light, feedback is sought about the report and where it may need to be improved, where there may be assumptions or arguments that should be challenged, and whether the report provides a sufficient basis for providing guidance to public sector organisations.

Some possible questions for consideration include:

- What might be missing?
- Is there anything that does not fit with the lived experience of innovation in the public sector?
- Does the report adequately provide an overview of the relevant factors for evaluating innovation?
- Are there additional (or better) examples or case studies that could be used to illustrate the process of evaluating innovation?

Feedback can be provided to the Observatory of Public Sector Innovation team at opsi@oecd.org. This will contribute to a beta version of the report, which will then be tested with representatives from OECD member countries and interested public servants.

Bibliography

Acquah, D. 2018. Presentation to the World Bank. Brasilia.

Al Ahbabi, S.A., Singh, S.K., Balasubramanian, S. and Gaur, S.S., 2018. Employee perception of impact of knowledge management processes on public sector performance. Journal of Knowledge Management.

Albury, D. 2005. 'Fostering innovation in public services', Public Money & Management 25(1): 51–6. http://doi.org/10.1111/j.1467–9302.2005.00450.x

Alford, J., Douglas, S., Geuijen, K. and 't Hart, P., 2017. Ventures in public value management: introduction to the symposium.

Anand, G. and Kodali, R., 2008. Benchmarking the benchmarking models. Benchmarking: An international journal, 15(3), pp.257-291.

Andersen, S.C. and Jakobsen, M.L., 2018. Political Pressure, Conformity Pressure, and Performance Information as Drivers of Public Sector Innovation Adoption. International Public Management Journal, 21(2), pp.213-242.

Andrews, R., & Entwistle, T. 2013. Four Faces of Public Service Efficiency: What, how, when and for whom to produce. Public Management Review, 15(2), 246-264.

Arena, M., and G. Azzone. 2009. "Identifying Organizational Drivers of Internal Audit Effectiveness." International Journal of Auditing 13 (1): 43–60. doi: 10.1111/j.1099-1123.2008.00392.x

Arena, M., and K. K. Jeppesen. 2016. "Practice Variation in Public Sector Internal Auditing: An Institutional Analysis." European Accounting Review 25 (2): 319–345. doi:10.1080/09638180.2015.1018917

Argyris, C. and Schön, D., 1978. Organizational learning: A theory of action approach. Reading, MA: Addision Wesley.

Arundel, A., & Huber, D. (2013). From too little to too much innovation? Issues in measuring innovation in the public sector. Structural Change and Economic Dynamics, 27, 146–159. doi:10.1016/j.strueco.2013.06.009

Arundel, A., Bloch, C. and Ferguson, B., 2018. Advancing innovation in the public sector: Aligning innovation measurement with policy goals. Research Policy.

Arundel, A., Casali, L. and Hollanders, H., 2015. How European public sector agencies innovate: The use of bottom-up, policy-dependent and knowledge-scanning innovation methods. Research Policy, 44(7), pp.1271-1282.

Askim, J., Hjelmar, U. and Pedersen, L.H., 2018. Turning Innovation into Evidence-based Policies: Lessons Learned from Free Commune Experiments. Scandinavian Political Studies, 41(4), pp.288-308.

Bacon, F. (1625) Of innovations (see Oxford Dictionary of Quotations , 2009, Oxford, Oxford University Press, pp45-49).

Bakhshi, J., Ireland, V. and Gorod, A., 2016. Clarifying the project complexity construct: Past, present and future. International Journal of Project Management, 34(7), pp.1199-1213.

Banfield, G. and Cayago-Gicain, M., 2006. Qualitative Approaches to Educational Evaluation: A Regional Conference-Workshop. International Education Journal, 7(4), pp.510-513.

Banks G (2009) Challenges of evidence-based policymaking, Contemporary Government Challenges series, Australian Public Service Commission and Productivity Commission.

Bardach, E. 2006. "Policy Dynamics." In The Oxford Handbook of Public Policy, edited by M. Moran, M. Rein, and R. E. Goodin. Oxford: Oxford University Press.

Bartlett, D. and Dibben, P., 2002. Public sector innovation and entrepreneurship: Case studies from local government. Local government studies, 28(4), pp.107-121.

Bason, C., 2010. Leading public sector innovation: Co-creating for a better society. Policy Press.

Baumol, W. J. and W.G. Bowen. 1966. Performing Arts, the Economic Dilemma: A Study of Problems Common to Theater, Opera, Music, and Dance. New York: Twentieth Century Fund.

Baumol, W.J., 1967. Macroeconomics of unbalanced growth: the anatomy of urban crisis. The American Economic Review, 415-426.

Bekkers, V. and Tummers, L., 2018. Innovation in the public sector: Towards an open and collaborative approach. International Review of Administrative Sciences, p.0020852318761797.

Bekkers, V. J. J. M., Tummers, L. G., & Voorberg, W. H. (2013). From public innovation to social innovation in the public sector: A literature review of relevant drivers and barriers. Rotterdam, NL: Erasmus University Rotterdam.

Bernier, L. & T. Hafsi (2007), The changing nature of public entrepreneurship, in: Public Administration Review, 67(3); 488-503.

Bianchi, A.; G. Marin, A. Zanfei. New perspectives in public service innovation. In: F. Gallouj, F. Djellal (Eds.), A Research Agenda for Service Innovation., Edward Elgar, Cheltenham (2018), pp. 166-186.

Blamey, A. and Mackenzie, M., 2007. Theories of change and realistic evaluation: peas in a pod or apples and oranges?. Evaluation, 13(4), pp.439-455.

Bloch, C., and M. M. Bugge. 2013. "Public Sector innovation – From Theory to Measurement." Structural Change and Economic Dynamics 27: 133–145. doi:10.1016/j.strueco.2013.06.008.

Borins, S. 2001. The Challenge of Innovating in Government. Arlington, VA: PricewaterhouseCoopers Endowment for the Business of Government.

Bovens, M., Schillemans, T. and Hart, P.T., 2008. Does public accountability work? An assessment tool. Public administration, 86(1), pp.225-242.

Bovens, Mark, Robert E. Goodin, and Thomas Schillemans. 2014. The Oxford Handbook of Public Accountability. Oxford: Oxford University Press.

Boyle, R. and Lemaire, D. eds., 1999. Building effective evaluation capacity: Lessons from practice (Vol. 8). Transaction Publishers.

Bozeman, B. (2003), "Public Value Mapping of Science Outcomes: Theory and Method", in D. Sarewitz et al., Knowledge Flows & Knowledge Collectives: Understanding the Role of Science & Technology Policies in Development, 2(1).

Bozeman, B. and Sarewitz, D., 2011. Public value mapping and science policy evaluation. Minerva, 49(1), pp.1-23.

Brayne, S. (2017). Big Data Surveillance: The Case of Policing. American Sociological Review, 82(5), 977-1008. doi: 10.1177/0003122417725865

Bristow, D., Carter, L. and Martin, S., (2015). Using evidence to improve policy and practice: the UK What Works Centres. Contemporary social science, 10(2), pp.126-137.

Britt, H. and Patsalides, M., 2013. Complexity-aware monitoring. Discussion Note, Monitoring and Evaluation Series. Washington, DC: USAID, December.

Brockman, J., 2015. This Idea Must Die: Scientific Theories That Are Blocking Progress (Edge Question Series). Harper Perennial: New York.

Brown, L. and Osborne, S.P., 2013. Risk and innovation: Towards a framework for risk governance in public services. Public Management Review, 15(2), pp.186-208.

Bugge, M., Mortensen, P.S. and Bloch, C., 2011. Measuring Public Innovation in Nordic Countries. Report on the Nordic Pilot studies-Analyses of methodology and results.

Byrkjeflot, H., Christensen, T., & Lægreid, P. (2014). The many faces of accountability: Comparing reforms in welfare, hospitals and migration. Scandinavian Political Studies, 37(2), 171-195.

Cabaj M, Leviten-Reid E, Vocisano D, et al. (2015) An example of patch evaluation: Vibrant Communities Canada. In: Patton M, McKegg K and Wehipeihana N (eds) Developmental Evaluation Exemplars: Principles in Practice. New York: The Guilford Press, 216–33.

Center for Offentlig Innovation (COI) (2015) Innovationsbarometeret - Foreløbig Baggrundsrapport. http://innovationsbarometer.coi.dk/media/1284/pdfgrundlag-2242015.pdf

Ceron, A., & Negri, F. (2016). The "Social Side" of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle. Policy & Internet, 8(2), 131-147. doi: 10.1002/poi3.117

Chesbrough, H.W., 2006. Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

Christensen, C., 2013. The innovator's dilemma: when new technologies cause great firms to fail. Harvard Business Review Press.

Cinar, E., Trott, P. and Simms, C., 2018. A systematic review of barriers to public sector innovation process. Public Management Review, pp.1-27.

Cook, T.D., Campbell, D.T. and Shadish, W., 2002. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Cook, T.D., Scriven, M., Coryn, C.L. and Evergreen, S.D., 2010. Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven. American Journal of Evaluation, 31(1), pp.105-117.

Crosby, B.C., 't Hart, P. and Torfing, J., 2017. Public value creation through collaborative innovation. Public Management Review, 19(5), pp.655-669.

Damanpour, F., R. M. Walker, and C. N. Avellaneda. 2009. "Combinative Effects of Innovation Types and Organizational Performance: A Longitudinal Study of Service Organizations." Journal of Management Studies 46 (4): 650–675. doi:10.1111/j.1467-6486.2008.00814.x.

Davies, R. and J. Dart (2005) 'The "Most Significant Change" (MSC) Technique; A Guide to its Use', www.mande.co.uk/docs/MSCGuide.pdf (retrieved 25 July 2006).

Davies, E. and Silloway, T. (2016) Research Clearinghouses. https://www.evidencecollaborative.org/toolkits/research-clearinghouses (retrieved 7 December 2018)

De Graaf, G. and Paanakker, H., 2015. Good governance: Performance values and procedural values in conflict. The American Review of Public Administration, 45(6), pp.635-652.

de Haan, J., 2006. How emergence arises. Ecological Complexity, 3(4), pp.293-301.

De Vries, H., Bekkers, V., & Tummers, L. (2016). Innovation in the public sector: A systematic review and future research agenda. Public Administration, 94(1), 146–166. doi:10.1111/padm.2016.94.issue-1

de Vries, H., Tummers, L. and Bekkers, V., 2018. A stakeholder perspective on public sector innovation: why position matters. International Review of Administrative Sciences, 84(2), pp.269-287.

Dean, E.D.W.I.N., 2009. International comparisons of non-marketed output: the role of productivity data. Available at: Worldbank, http://siteresources. worldbank. org/ICPINT/Resources/270056-1255977254560/InternationalComparisonofNonMarketedOutput_E. Dean. pdf.

Delahais, T. and Toulemonde, J., 2012. Applying contribution analysis: Lessons from five years of practice. Evaluation, 18(3), pp.281-293.

Demircioglu, M.A. and Audretsch, D.B., 2018. Conditions for complex innovations: evidence from public organizations. The Journal of Technology Transfer, pp.1-24.

Demircioglu, M.A. and Audretsch, D.B., 2018. Conditions for complex innovations: evidence from public organizations. The Journal of Technology Transfer, pp.1-24.

Dillard, J.F. and Ruchala, L., 2005. The rules are no game: From instrumental rationality to administrative evil. Accounting, auditing & accountability journal, 18(5), pp.608-630.

Djellal, F., Gallouj, F. and Miles, I., 2013. Two decades of research on innovation in services: Which place for public services?. Structural change and economic dynamics, 27, pp.98-117.

Dunleavy, P. (2016). 'Big data' and policy learning. In G. Stoker, & M. Evans (Eds.), Evidence-based Policy Making in the Social Sciences: Methods That Matter (pp. 143-151). Bristol: Policy Press.

Dunleavy, P. and Carrera, L., 2013. Growing the productivity of government services. Edward Elgar Publishing.

Dunleavy, P., 2015. Public sector productivity: puzzles, conundrums, dilemmas and their solutions. Managing Under Austerity, Delivering Under Pressure, p. 25.

Dunleavy, P., 2017. Public sector productivity. OECD Journal on Budgeting, 17(1), pp.1-28.

Earl, S., Carden, F. and Smutylo, T. 2001. Outcome mapping: building learning and reflection into development program [online], Ottawa: International Development Research Centre. Available from: http://www.outcomemapping.ca/resource/resource.php?id=83

Ebrahim, A., 2005. Accountability myopia: Losing sight of organizational learning. Nonprofit and voluntary sector quarterly, 34(1), pp.56-87.

Edgerton, D. (2007) The shock of the old: technology and global history since 1900, Oxford, Oxford University Press.

Edler et al. (2016) and: http://www.innovation-policy.org.uk/

Edler, J. and Boon, W.P., 2018. 'The next generation of innovation policy: Directionality and the role of demand-oriented instruments'—Introduction to the special section. Science and Public Policy.

Engin, Z. and Treleaven, P., 2018. Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies. The Computer Journal.

Evaluation Support Scotland (2018) Making it stick - A guide to embedding evaluation. http://evaluationsupportscotland.org.uk/resources/411/

Figueredo AJ, Olderbak SG, Schlomer GL, Garcia RA, Wolf PS. 17 Program Evaluation: Principles, Procedures, and Practices. The Oxford handbook of quantitative methods. 2014:332.

Flanagan, K. & Elvira Uyarra (2016) Four dangers in innovation policy studies – and how to avoid them, Industry and Innovation, 23:2, 177-188, DOI:10.1080/13662716.2016.1146126

Flemig, S., Osborne, S. and Kinder, T., 2016. Risky business—reconceptualizing risk and innovation in public services. Public Money & Management, 36(6), pp.425-432.

Florio, M., Forte, S., Pancotti, C., Sirtori, E. and Vignetti, S., 2016. Exploring cost-benefit analysis of research, development and innovation infrastructures: an evaluation framework. arXiv preprint arXiv:1603.03654.

Forés, B. and Camisón, C., 2016. Does incremental and radical innovation performance depend on different types of knowledge accumulation capabilities and organizational size?. Journal of Business Research, 69(2), pp.831-848.

Forss, K., Marra, M. & Schwartz, R. (2011). Evaluating the complex, New Brunswick and London, Transaction Publishers.

Frees, W. and Bouckaert, G., 2015. The role of Feedback, Accountability and Learning in Organizational Change and Innovation: A theoretical framework.

Fuglsang, L., 2010. Bricolage and invisible innovation in public service innovation. Journal of Innovation Economics & Management, (1), pp.67-87.

Funnell, S.C. and Rogers, P.J., 2011. Purposeful program theory: Effective use of theories of change and logic models(Vol. 31). John Wiley & Sons.

Garicano, L. and Heaton, P., 2010. Information technology, organization, and productivity in the public sector: Evidence from police departments. Journal of Labor Economics, 28(1), pp.167-201.

Garud, R., A. Kumaraswamy, and P. Karnøe. 2010. "Path Dependence or Path Creation?" Journal of Management Studies 47: 760–774.

Garvin, D., Edmondson, A. & Gino, F. (2008). Is yours a learning organization? Harvard Business Review, March, pp. 109-116.

Geletkanycz, M.A. and Black, S.S., 2001. Bound by the past? Experience-based effects on commitment to the strategic status quo. Journal of Management, 27(1), pp.3-21.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? Policy Sciences, 50(3), 367-382. doi: 10.1007/s11077-017- 9293-1

Gilson, C., Dunleavy, P. & Tinkler, J. (2009). Organizational learning in government sector organizations: literature review. LSE Public Policy Group, London, UK. Available online at: http://eprints.lse.ac.uk/59228/1/Gilson_Dunleavy_Tinkler_Organizational-learning-ingovernment-sector-organizations_2009.pdf

Glouberman S, Zimmerman B. 2004. Complicated and complex systems: what would successful reform of Medicare look like? In Health Care Services and the Process of Change .21-53, Forest GP, McIntosh T, Marchildon G (eds.), vol. 2, Originally published as Discussion Paper No 8 (2002). Commission on the Future of Health Care in Canada Romanow Papers: Ottawa; 21–53.

Glynn, J.J. and Murphy, M.P., 1996. Public management: Failing accountabilities and failing performance review. International Journal of Public Sector Management, 9(5/6), pp.125-137.

Godenhjelm, S. and Johanson, J.E., 2018. The effect of stakeholder inclusion on public sector project innovation. International Review of Administrative Sciences, 84(1), pp.42-62.

Godin, B. and C. Doré (2006), "Measuring the Impacts of Science: beyond the Economic Dimension", Working Paper, mimeo.

Goffin, K., and R. Mitchell. 2010. Innovation Management: Strategy and Implementation Using the Pentathlon Framework. Basingstoke: Palgrave Macmillan.

Goodman, L.A., Epstein, D. and Sullivan, C.M., 2018. Beyond the RCT: Integrating Rigor and Relevance to Evaluate the Outcomes of Domestic Violence Programs. American Journal of Evaluation, 39(1), pp.58-70.

Granato, J. and Scioli, F., 2004. Puzzles, proverbs, and omega matrices: The scientific and social significance of empirical implications of theoretical models (EITM). Perspectives on Politics, 2(2), pp.313-323.

Greiling, D. and Halachmi, A., 2013. Accountability and organizational learning in the public sector. Public Performance & Management Review, 36(3), pp.380-406.

Grimsley, M. and Meehan, A., 2007. e-Government information systems: Evaluation-led design for public value and client trust. European Journal of Information Systems, 16(2), pp.134-148.

Gripper, R., Kazimirski, A., Kenley, A., McLeod, R. and Weston, A., 2017. Global Innovations in Measurement and Evaluation. NPC, online, available at: http://www.

thinknpc. org/wp-content/uploads/2017/06/NPC-Innovations-in-Measurement-Final-Interactive-20170614. pdf [accessed 1 December 2017].

Halachmi, A., 2014. Accountability overloads. In The Oxford handbook of public accountability.

Hall, Jeremy L., and Edward T. Jennings. 2010. Assessing the Use and Weight of Information and Evidence in U.S. State Policy Decisions. Policy & Society 29(2): 137–47.

Hartley, J. 2005. 'Innovation in governance and public services: Past and present', Public Money and Management 25(1): 27–34. http://doi.org/10.1111/j.1467– 9302.2005.00447.x.

Hartley, J. and Allison, M., 2002. Good, better, best? Inter-organizational learning in a network of local authorities. Public Management Review, 4(1), pp.101-118.

Hartley, J. and Rashman, L., 2018. Innovation and inter-organizational learning in the context of public service reform. International Review of Administrative Sciences, p.0020852318762309.

Hasan, M., & Kerr, R. M. (2003). The relationship between total quality management practices and organisational performance in service organisations.The TQM Magazine, 15(4), 286-291.

Head, B. 2013. How Do Government Agencies Use Evidence? Research report for the Swedish National Board of Health and Welfare. http://www.socialstyrelsen.se/ Lists/Artikelkatalog/Attachments/19163/2013-6-38.pdf [accessed October 7, 2015].

Head, B.W., 2016. Toward More "Evidence-Informed" Policy Making?. Public Administration Review, 76(3), pp.472-484.

Head, B.W., 2017. Assessing the value of political science in the era of relevance and impact. Australian Journal of Political Science, 52(4), pp.599-614.

Head, Brian W., Michele Ferguson, Adrian Cherney, and Paul Boreham. 2014. Are Policy-Makers Interested in Social Research? Exploring the Sources and Uses of Valued Information among Public Servants in Australia. Policy & Society 33(2): 89–101.

Hermans, L.M., Haasnoot, M., ter Maat, J. and Kwakkel, J.H., 2017. Designing monitoring arrangements for collaborative learning about adaptation pathways. Environmental Science & Policy, 69, pp.29-38.

Hill, C.J. and Lynn, L.E., 2005. Is hierarchical governance in decline? Evidence from empirical research. Journal of public administration research and theory, 15(2), pp.173-195.

Hislop, D., Bosua, R. and Helms, R., 2018. Knowledge management in organizations: A critical introduction. Oxford University Press.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. Journal of Organizational Computing and Electronic Commerce, 26(1-2), 147-169. doi: 10.1080/10919392.2015.1125187

Hollanders, H., A. Arundel, B. Buligescu, V. Peter, L. Roman, P. Simmonds, and N. Es-Sadki. 2013. Enterprise and Industry, European Commission, European Public Sector Innovation Scoreboard 2013 – A Pilot Exercise. Belgium: European Commission.

Howlett, M., Ramesh, M. and Perl, A., 2009. Studying public policy: Policy cycles and policy subsystems (Vol. 3). Oxford: Oxford University Press.

Hughes, A., Moore, K. and Kataria, N., 2011. Innovation in Public Sector Organisations: A pilot survey for measuring innovation across the public sector. London: Nesta.

Ison, R., 2018. Governing the human–environment relationship: systemic practice. Current Opinion in Environmental Sustainability, 33, pp.114-123.

Jacob, S., Speer, S. and Furubo, J.E., 2015. The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. Evaluation, 21(1), pp.6-31.

Jarvie, W. and Stewart, J., 2018. Conceptualizing learning in the public sector: the importance of context. International Journal of Public Sector Management, 31(1), pp.14-30.

Jennings, Edward T., and Jeremy L. Hall. 2012. Evidence-Based Practice and the Use of Information in State Agency Decision Making. Journal of Public Administration Research and Th eory 22(2): 245–66

Johnson, K., Greenseid, L.O., Toal, S.A., King, J.A., Lawrence, F. et al. (2009). Research on Evaluation Use. A Review of the Empirical Literature From 1986 to 2005. American Journal of Evaluation, 30:3, pp. 377-410.

Jones L, Watson B, Hobman E, et al. (2008) Employee perceptions of organizational change: Impact of hierarchical level. Leadership & Organization Development Journal 29(4): 294–316

Jones, S. and Hughes, J., 2001. Understanding IS evaluation as a complex social process: a case study of a UK local authority. European Journal of Information Systems, 10(4), pp.189-203.

Jones, S., Hackney, R. and Irani, Z., 2007. Towards e-government transformation: conceptualising "citizen engagement" A research note. Transforming Government: People, Process and Policy, 1(2), pp.145-152.

Jorgensen, T., and B. Bozeman. 2007. "Public Values: An Inventory." Administration & Society 39 (3): 354–381. doi:10.1177/0095399707300703.

Jüngen, A. (2013). Understanding the role of evaluation findings in the development aid policy process. Presented at the International Conference on Public Policy, Grenoble.

Kankanhalli, A., Zuiderwijk, A. and Tayi, G.K., 2017. Open innovation in the public sector: A research agenda.

Karo, E. and Kattel, R., 2016. Innovation and the State: Thinking of Government as "Technology Maker" and Implications for Public Administration Research. Administrative Culture, 17(1), pp.5-17.

Kattel, R. and Mazzucato, M., 2018. Mission-oriented innovation policy and dynamic capabilities in the public sector. Industrial and Corporate Change, 2018, Vol. 27, No. 5, 787–801 doi: 10.1093/icc/dty03

Kattel, R., Cepilovs, A., Drechsler, W., Kalvet, T., Lember, V. and Tõnurist, P., 2013. Can we measure public sector innovation? A literature review. LIPSE Project Paper.

Kay, R. and Goldspink, C., 2012. What CEOs mean when they say they want to innovate.

Kelly, G., G. Mulgan and S. Muers. 2002. Creating Public Value: An Analytical Framework for Public Service Reform. Discussion paper prepared by the Cabinet Office Strategy Unit, United Kingdom.

Kelman, S. (2008) 'The "Kennedy School School" research on innovation in government', pp28-51 in S.Borins (ed.) Innovations in government: research, recognition and replication, Washington DC, Brookings Institution.

Klecun, E. and Cornford, T., 2005. A critical approach to evaluation. European Journal of Information Systems, 14(3), pp.229-243.

Kowalkiewicz, M., Safrudin, N. and Schulze, B., 2017. The business consequences of a digitally transformed economy. In Shaping the Digital Enterprise (pp. 29-67). Springer, Cham.

Krugman, P. 1990. The Age of Diminished Expectations, MIT Press, Cambridge, Mass.

Kuhlmann, S. and Bogumil, J., 2018. Performance measurement and benchmarking as "reflexive institutions" for local governments: Germany, Sweden and England compared. International Journal of Public Sector Management, 31(4), pp.543-562.

Kuhlmann, S., 2010. Performance Measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany. International Review of Administrative Sciences, 76(2), pp.331-345.

Lau, E., Lonti, Z. and Schultz, R., 2017. Challenges in the measurement of public sector productivity in OECD countries. International Productivity Monitor, 32, pp.180-195.

Lehtonen, M., 2015. Indicators: tools for informing, monitoring or controlling?.

Lember, V., Kattel, R. and Tonurist, P., 2016. Public Administration, Technology and Administrative Capacity. Working Papers in Technology Governance and Economic Dynamics no. 71.

Lewis, J.M. and Triantafillou, P., 2012. From performance measurement to learning: a new source of government overload?. International Review of Administrative Sciences, 78(4), pp.597-614.

Liebert, W. and Schmidt, J.C., 2010. Collingridge's dilemma and technoscience. Poiesis & Praxis, 7(1-2), pp.55-71.

Lindsay, C., Findlay, P., McQuarrie, J., Bennie, M., Corcoran, E.D. and Van Der Meer, R., 2018. Collaborative Innovation, New Technologies, and Work Redesign. Public Administration Review, 78(2), pp.251-260.

Liston-Heyes, C. and Juillet, L., 2018. Employee isolation and support for change in the public sector: a study of the internal audit profession. Public Management Review, pp.1-23.

Little, T.D. ed., 2013. The Oxford handbook of quantitative methods, volume 1: Foundations. Oxford University Press.

Luce, B.R., Kramer, J.M., Goodman, S.N., Connor, J.T., Tunis, S., Whicher, D. and Schwartz, J.S., 2009. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. Annals of internal medicine, 151(3), pp.206-209.

Mackenzie, M. and Blamey, A., 2005. The practice and the theory: lessons from the application of a theories of change approach. Evaluation, 11(2), pp.151-168.

Magro, E. and Wilson, J.R., 2013. Complex innovation policy systems: Towards an evaluation mix. Research Policy, 42(9), pp.1647-1656.

Magro, E. and Wilson, J.R., 2018. Policy-mix evaluation: Governance challenges from new place-based innovation policies. Research Policy.

Magro, E., Navarro, M. and Zabala-Iturriagagoitia, J.M., 2014. Coordination-mix: The hidden face of STI policy. Review of Policy Research, 31(5), pp.367-389.

Mairesse, J. and Mohnen, P., 2010. Using innovation surveys for econometric analysis. Handbook of the Economics of Innovation, 2, 1129-1155.

Makarova, E. and Sokolova, A., 2014. Foresight evaluation: lessons from project management. Foresight, 16(1), pp.75-91.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. and Kegels, G., 2012. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. Evaluation, 18(2), pp.192-212.

Mason, P. and Barnes, M., 2007. Constructing theories of change: Methods and sources. Evaluation, 13(2), pp.151-170.

Mathie A and Peters B (2014) Joint (ad)ventures and (in)credible journeys evaluating innovation assetbased community development in Ethiopia. Development in Practice 24(3): 405–19.

Mayne, J., 2011. Contribution analysis: Addressing cause and effect. Evaluating the complex: Attribution, contribution, and beyond, 18, pp.53-96.

Mazzucato, M., 2015. The entrepreneurial state: Debunking public vs. private sector myths (Vol. 1). Anthem Press.

McKegg K, Wehipeihana N, Becroft M, et al. (2015) Developmental evaluation's role in supporting community-led solutions for Maori and pacific young people's educational success: The Foundation North Maori and Pacific Education Initiative. In: Patton M, McKegg K and Wehipeihana N (eds) Developmental Evaluation Exemplars: Principles in Practice. New York: The Guilford Press, 125–42.

Meijer, A., 2018. Public Innovation Capacity: Developing and Testing a Self-Assessment Survey Instrument. International Journal of Public Administration, pp.1-11.

Mergel, I., 2016. Agile innovation management in government: A research agenda. Government Information Quarterly, 33(3), pp.516-523.

Mergel, I., Kleibrink, A. and Sörvik, J., 2018. Open data outcomes: US cities between product and process innovation. Government Information Quarterly, 35(4), pp.622-632.

Millar, R. and Hall, K., 2013. Social return on investment (SROI) and performance measurement: The opportunities and barriers for social enterprises in health and social care. Public Management Review, 15(6), pp.923-941.

Milley, P., Szijarto, B., Svensson, K. and Cousins, J.B., 2018. The evaluation of social innovation: A review and integration of the current empirical knowledge base. Evaluation, 24(2), pp.237-258.

Misuraca, G., Codagnone, C., & Rossel, P. 2013. From practice to theory and back to practice: Reflexivity in measurement and evaluation for evidence-based policy making in the information society. Government Information Quarterly, 30, S68–S82.

Mohnen, P. and Hall, B.H., 2013. Innovation and productivity: An update. Eurasian Business Review, 3(1), pp.47-65.

Morgan, G. (2006). Unfolding logics of change: organization as flux and transformation. In Morgan, G. (Ed.): Images of Organization, pp. 241-290. Sage Publications, London, UK.

Morra, L.G. and Friedlander, A.C., (n.d.). Case Study Evaluations, The World Bank Group. Retrieved from: https://ieg.worldbankgroup.org/Data/reports/oed_wp1.pdf

Morris, Michael, and Brittany Clark. 2013. You Want Me to Do What? Evaluators and the Pressure to Misrepresent Findings. American Journal of Evaluation 34(1): 57–70.

Mulgan, G. and Albury, D., 2003. Innovation in the public sector. Strategy Unit, Cabinet Office, 1, p.40.

Nanda, T. and Singh, T.P., 2009. Determinants of creativity and innovation in the workplace: a comprehensive review. International Journal of Technology, Policy and Management, 9(1), pp.84-106.

Nasi, G., Cucciniello, M. and Degara, V., 2018. Evaluation of Innovation Performance in the Public Sector: A Systematic Review of Studies. In Outcome-Based Performance Management in the Public Sector (pp. 203-224). Springer, Cham.

Navarro, F.M., (2016) From Government Innovation to Public Innovation. The ICT as Key Tools. International Journal of New Technology and Research (IJNTR) ISSN:2454-4116, Volume-2, Issue-6, June 2016 Pages 04-07.

Nevo, I. and Slonim-Nevo, V., 2011. The myth of evidence-based practice: Towards evidence-informed practice. British Journal of Social Work, 41(6), pp.1176-1197.

Obal, M., Kannan-Narasimhan, R. and Ko, G., 2016. Whom should we talk to? Investigating the varying roles of internal and external relationship quality on radical and incremental innovation performance. Journal of Product Innovation Management, 33, pp.136-147.

OECD (2010). Glossary of key terms in evaluation and results based management. OECD Publishing, Paris, France.

OECD (2011) Enhancing Research Performance through Evaluation, Impact Assessment and Priority Setting.

OECD (2015) Impact assessment. Available at: https://www.innovationpolicyplatform.org/sites/default/files/general/What%20is%20impact%20assessment%20OECD%20Impact%20Assessment%20KT%202015_0.pdf

OECD (2017), Systems Approaches to Public Sector Challenges: Working with Change, OECD Publishing, Paris, https://doi.org/10.1787/9789264279865-en.

OECD. 2010. Measuring innovation. A new perspective. Paris: OECD Publishing.

Oliver, K., Innvar, S., Lorenc, T., Woodman, J. and Thomas, J., 2014. A systematic review of barriers to and facilitators of the use of evidence by policymakers. BMC health services research, 14(1), p.2.

Osborne, D. and Gaebler, T., 1992. Reinventing government: How the entrepreneurial spirit is transforming government. Reading Mass. Adison Wesley Public Comp.

Osborne, S.P. ed., 2010. The new public governance: Emerging perspectives on the theory and practice of public governance. Routledge.

Osimo, D., 2018. The Opportunities and Limitations of Crowdsourcing Policy Indicators. Governance Indicators: Approaches, Progress, Promise, p.186.

Page, S.B., Stone, M.M., Bryson, J.M. and Crosby, B.C., 2015. Public value creation by cross-sector collaborations: A framework and challenges of assessment. Public Administration, 93(3), pp.715-732.

Pärna, O. and von Tunzelmann, N., 2007. Innovation in the public sector: Key features influencing the development and implementation of technologically innovative public sector services in the UK, Denmark, Finland and Estonia. Information Polity, 12(3), pp.109-125.

Paskaleva, K. and Cooper, I., 2018. Open innovation and the evaluation of internet-enabled public services in smart cities. Technovation.

Patton, M. (1975). In Search of Impact: An Analysis of the Utilisation of Federal Health Evaluation Research. Minneapolis: Centre of Social Research, Minnesota University.

Patton, M.Q. and Horton, D. 2009. Utilization-Focused Evaluation for Agricultural Innovation. International Labor Accreditation Cooperation (ILAC) Brief No. 22. ILAC, Bioversity, Rome.

Patton, M.Q., 2005. Toward distinguishing empowerment evaluation and placing it in a larger context: Take two. American Journal of Evaluation, 26(3), pp.408-414.

Patton, M.Q., 2010. Developmental evaluation: Applying complexity concepts to enhance innovation and use. Guilford Press.

Patton, M.Q., 2011. Essentials of utilization-focused evaluation. Sage.

Pawson, R., Tilley, N. and Tilley, N., 1997. Realistic evaluation. Sage.

Paz-Ybarnegaray, R. and Douthwaite, B., 2017. Outcome evidencing: A method for enabling and evaluating program intervention in complex systems. American Journal of Evaluation, 38(2), pp.275-293.

Penfold, R.B. and Zhang, F., 2013. Use of interrupted time series analysis in evaluating health care quality improvements. Academic pediatrics, 13(6), pp.S38-S44.

Pestoff, V., 2012. Co-production and third sector social services in Europe: Some concepts and evidence. Voluntas: International Journal of Voluntary and Nonprofit Organizations, 23(4), pp.1102-1118.

Peters, P. (1994) "The Best Corporate Strategy? None, Of Course." Chicago Tribune, July 11, 1994

Peters, T. J. (1989), Thriving on Chaos: Handbook for a Management Revolution (Pan).

Pollitt C, Girre X, Lonsdale J, Mul R, Summa H and Waerness M (1999) Performance or Compliance? Performance Audit and Public Management in Five Countries. Oxford: Oxford University Press.

Pollitt, C. (2011). Innovation in the public sector: an introductory overview. In Bekkers, V.J.J.M., Edelenbos, J., & Steijn, B. (Eds.): Innovation in the public sector – Linking capacity and leadership, pp. 35-43. Palgrave MacMillan, New York, USA.

Pollitt, C. (2015) Innovation and Lipse: An Interim Assessment of an Innovative. Research Project Speech to Lipse Conference: Royal Flemish Academy of Belgium, Brussels, 3 February 2015. http://lipse.org/userfiles/uploads/INNOVATION-LIPSE-Keynote.pdf

Pollitt, C. 2008. The Essential Public Manager, Maidenhead: Open University Press.

Posavac, E.J., 2015. Program evaluation: Methods and case studies. Routledge.

Poth C, Pinto D and Howery K (2012) Addressing the challenges encountered during a developmental evaluation: Implications for evaluation practice. The Canadian Journal of Program Evaluation 26(1): 39–48.

Potts, J. and Kastelle, T., 2010. Public sector innovation research: What's next?. Innovation, 12(2), pp.122-137.

Raadschelders, J.C. and Lee, K.H., 2011. Trends in the study of public administration: Empirical and qualitative observations from Public Administration Review, 2000–2009. Public Administration Review, 71(1), pp.19-33.

Ragin C. Fuzzy-set social science. Chicago: University of Chicago; 2000.

Ramaprasad, A., 1983. On the definition of feedback. Behavioral Science, 28(1), pp.4-13.

Reynolds, M., Gates, E., Hummelbrunner, R., Marra, M. and Williams, B., 2016. Towards systemic evaluation. Systems Research and Behavioral Science, 33(5), pp.662-673.

Rihoux, B., 2006. Qualitative comparative analysis (QCA) and related systematic comparative methods: Recent advances and remaining challenges for social science research. International Sociology, 21(5), pp.679-706.

Rijkens-Klomp, N. and Duin van der, P. (2011), ―Evaluating Local Public Foresight Studies From A User Perspective‖, paper presented at the 4th International Seville Conference on Future-Oriented Technology Analysis (FTA), 12-13 May, Seville, Spain, available at: http://ftp.jrc.es/EURdoc/JRC64647.pdf.

Rixen, T. 2013. 'Why reregulation after the crisis is feeble: Shadow banking, offshore financial centers, and jurisdictional competition', Regulation & Governance 7(4): 435–59. http://doi.org/10.1111/rego.12024.

Rogers, E. M. 1983. Diffusion of Innovation. New York: Free Press.

Sahay, B. S. (2005). Multi-factor productivity measurement model for service organisation. International Journal of Productivity and Performance Management, 54(1), 7-22.

Salge, T.O. and Vera, A., 2012. Benefiting from public sector innovation: The moderating role of customer and learning orientation. Public Administration Review, 72(4), pp.550-559.

Sanderson, I., 2002. Evaluation, policy learning and evidence-based policy making. Public administration, 80(1), pp.1-22.

Savoldelli, A., Codagnone, C., & Misuraca, G. 2014. Understanding the e-government paradox: Learning from literature and practice on barriers to adoption. Government Information Quarterly, 31, S63–S71.

Schillemans, T. and Bovens, M., 2011. The Challenge of Multiple Accountability. Accountable governance: Problems and promises.

Schmidt, V.A., 2013. Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'. Political Studies, 61(1), pp.2-22.

Schön, D.A., 1975. Deutero-learning in organizations: Learning for. Organizational dynamics, 4(1), pp.2-16.

Schorr, L. and Farrow, F., 2014, November. An evidence framework to improve results. In Harold Richman Public Policy Symposium, Center for the Study of Social Policy.

Schott, C., van Kleef, D.D. and Steen, T., 2015. What does it mean and imply to be public service motivated?. The American Review of Public Administration, 45(6), pp.689-707.

Schumpeter, J.A., 1934. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle. Harvard University Press, Cambridge, MA.

Schuurman, D. and Tõnurist, P., 2016. Innovation in the public sector: Exploring the characteristics and potential of living labs and innovation labs. In OpenLivingLab Days 2016.

Semanjski, I., Bellens, R., Gautama, S., & Witlox, F. (2016). Integrating Big Data into a Sustainable Mobility Policy 2.0 Planning Support System. Sustainability, 8(11), 1-19. doi: http://10.3390/su8111142

Serrano-Guerrero, J., Olivas, J.A., Romero, F.P. and Herrera-Viedma, E., 2015. Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311, pp.18-38.

Shillabeer, A., Buss, T.F. and Rousseau, D.M. eds., 2011. Evidence-based public management: Practices, issues, and prospects. ME Sharpe.

Smith, K. (2005) Measuring innovation, in Fagerberg, J., Mowery, D.C. & Nelson, R.R. (Eds). The Oxford Handbook of Innovation. Oxford, UK: Oxford University Press, pp. 148-177.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. International Journal of Public Administration, 18, 277-310.

Smutylo, T., 2005. Outcome mapping: A method for tracking behavioural changes in development programs.

Staley L (2008) Evidence-based policy and public sector innovation, IPA Report. http://www.ipa. org.au/library/publication/1226382181_document_staley_vic_gov_innovation.pdf

Stevens, Alex. 2011. Telling Policy Stories: An Ethnographic Study of the Use of Evidence in Policy-Making in the U.K. Journal of Social Policy 40(2): 237–55.

Strategic Policy Intelligence (2001). Current Trends, the State of Play and Perspectives. IPTS Technical Report Series, Seville, December 2001.

Studinka, J. and Guenduez, A.A., 2018. The Use of Big Data in the Public Policy Process-Paving the Way for Evidence-Based Governance.

Stufflebeam, D.L. and Coryn, C., 2014. Daniel Stufflebeam's CIPP Model for evaluation: An improvement and accountabilityoriented approach. Evaluation theory, models, and applications, pp.309-340.

Stufflebeam,D. L. (2001). Evaluation models. New Directions for Evaluation, 89, 7–98.

Tan E, McGill S, Tanner E, et al. (2014) The evolution of an academic-community partnership in the design, implementation, and evaluation of experience corps Baltimore city a courtship model. The Gerontologist 54(2): 314–21.

Tangen, S. 2005. Demystifying productivity and performance. International Journal of Productivity and Performance Measurement, 54(1), 34-46.

Taras, M., 2005. Assessment–summative and formative–some theoretical reflections. British journal of educational studies, 53(4), pp.466-478.

Tate, M., Bongiovanni, I., Kowalkiewicz, M. and Townson, P., 2018. Managing the "Fuzzy front end" of open digital service innovation in the public sector: A methodology. International Journal of Information Management, 39, pp.186-198.

Technopolis Group and Mioir (2012): Evaluation of Innovation Activities. Guidance on methods and practices. Study funded by the European Commission, Directorate for Regional Policy.

Tocqueville, Alexis de (1840). Democracy in America, Volume II. OCLC 48653286

Togni S, Askew D, Rogers L, et al. (2015) Creating safety to explore: Strengthening innovation in an Australian indigenous primary health care setting through developmental evaluation. In: Patton M, McKegg K and Wehipeihana N (eds) Develop.

Toivonen, M., Touminen, T., & Brax, S. (2007). Innovation process interlinked with the process of service delivery: A management challenge in KIBS. Économies et Sociétés, 8(3), 355–384.

Tõnurist, P., (2018). New approaches in policy design and experimentation. OECD STI Outlook.

Tonurist, P., Kattel, R. and Lember, V., 2015. Discovering innovation labs in the public sector. The Other Canon, Foundation and Tallinn University of Technology Working Papers in Technology Governance and Economic Dynamics, TUT Ragnar Nurkse School of Innovation and Governance.

Tõnurist, P., Kattel, R. and Lember, V., 2017. Innovation labs in the public sector: what they are and what they do?. Public Management Review, 19(10), pp.1455-1479.

Tõnurist, Piret, and Martin Bækgaard. "Economics and PA: Public Choice Theory, Transaction Costs Theory, Theory of Expectations, and the Enduring Influence of Economics Modeling on PA—Comparing the Debate in the US and Europe." In The Palgrave Handbook of Public Administration and Management in Europe, pp. 1105-1120. Palgrave Macmillan, London, 2018.

Torfing, J., 2018. Collaborative innovation in the public sector: the argument. Public Management Review, pp.1-11.

Torugsa, N. and Arundel, A., 2016. Complexity of Innovation in the public sector: a workgroup-level analysis of related factors and outcomes. Public Management Review, 18(3), pp.392-416.

Tushman, M.L. and Anderson, P., 1986. Technological discontinuities and organizational environments. Administrative science quarterly, pp.439-465.

Uyarra, E., and R. Ramlogan. 2016. "The Effects of Cluster Policy on Innovation." In Handbook of Innovation Policy Impact, edited by J. Edler, P. Cunningham, A. Gok, and P. Shapira. London: Edward Elgar.

van Acker, W. and Bouckaert, G., 2017. What makes public sector innovations survive? An exploratory study of the influence of feedback, accountability and learning. International Review of Administrative Sciences, p.0020852317700481.

Van Acker, W., 2017. Sustainable Public Sector Innovations: How do feedback, accountability and learning matter?.

Van Dooren et al., 2010Van Dooren, W., Bouckaert, G.and Halligan, J. 2010. Performance Management in the Public Sector, London: Routledge.

Van Dooren, W. and Willems, T., 2018. Thinking Allowed: Reforming Indicator-Based Accountability to Enhance Innovation. Available at: https://www.researchgate.net/publication/328106343_Thinking_Allowed_Reforming_Indicator-Based_Accountability_to_Enhance_Innovation

Van Dooren, W., Bouckaert, G. and Halligan, J., 2015. Performance management in the public sector. Routledge.

Van Loocke E and Put V (2011) The impact of performance audits of SAIs: A review of the existing evidence. In: Lonsdale J, Wilkins P and Ling T (eds) Performance Auditing: Contributing to Accountability in Democratic Government. Cheltenham: Edward Elgar, pp. 175–208.

Van Loocke, E. and Put, V., 2010. De impact van performance auditing: slow and subtle?.

Vedung, E., 2017. Public policy and program evaluation. Routledge.

Venable, J., Pries-Heje, J. and Baskerville, R., 2016. FEDS: a framework for evaluation in design science research. European Journal of Information Systems, 25(1), pp.77-89.

Vivarelli, M. 2007. Innovation and Employment: A Survey. In: Hanusch, H. & A. Pyka (eds.) Elgar Companion to Neo-Schumpeterian Economics. Cheltenham: Elgar, 719-732.

Voelpel, S.C., Leibold, M. and Eckhoff, R.A., 2006. The tyranny of the Balanced Scorecard in the innovation economy. Journal of Intellectual Capital, 7(1), pp.43-60.

Volden, C. 2010. “Failures: Diffusion, Learning, and Policy Abandonment.” Paper presented at the American Political Science Association Annual Research Conference, August, Washington, DC.

Von Hippel, E., 2006. Democratizing Innovation (MIT Press). Cambridge, Massachusetts.

Walker RM (2013) Internal and external antecedents of process innovation: A review and extension. Public Management Review 16(1): 21–44.

Walker, R. M. 2014. “Internal and External Antecedents of Process Innovation: A Review and Extension.” Public Management Review 16(1):21–44. doi:10.1080/14719037.2013.771698.

Wegner, G. 2003. “Evolutionary Markets and the Design of Institutional Policy.” In The Evolutionary Analysis of Economic Policy, edited by P. Pelikan and G. Wegner, 46–66. Cheltenham: Edward Elgar.

Wegrich, K. and Jann, W., 2006. Theories of the policy cycle. In Handbook of public policy analysis (pp. 69-88). Routledge.

Weiss, C.H., 1997. Theory-based evaluation: Past, present, and future. New directions for evaluation, 1997(76), pp.41-55.

Weiss, C.H., 1999. The interface between evaluation and public policy. Evaluation, 5(4), pp.468-486.

Weiss, C.H., 2000. Theory-based evaluation: theories of change for poverty reduction programs. Evaluation and povery reduction.

What Works Centre for Local Economic Growth, 2014, http://www.whatworksgrowth.org/public/files/Methodology/14-03-20_About_our_Reviews.pdf

Wiliam, D. and Black, P., 1996. Meanings and consequences: a basis for distinguishing formative and summative functions of assessment?. British Educational Research Journal, 22(5), pp.537-548.

Willems, T. and Van Dooren, W., 2012. Coming to terms with accountability: Combining multiple forums and functions. Public Management Review, 14(7), pp.1011-1036.

Willetts, J. and Crawford, P., 2007. The most significant lessons about the Most Significant Change technique. Development in Practice, 17(3), pp.367-379.

Williams B, Hummelbrunner R. 2011. Systems Concepts in Action: A Practitioner's Toolkit. Stanford University Press: Stanford.

Williams, B. and Hummelbrunner, R., 2011. Systems dynamics. Systems Concepts in Action. Stanford University Press, Stanford, CA, pp.45-59.

Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. Journal of Education Policy, 31(2), 123-141. doi: 10.1080/02680939.2015.1035758

Wilson-Grau, R. and Britt, H., 2012. Outcome harvesting. Cairo: Ford Foundation. http://www. managingforimpact. org/sites/default/files/resource/outome_harvesting_brief_final_2012-05-2-1. pdf.

Witt, U. 2003. "Economic Policy Making in Evolutionary Perspective." Journal of Evolutionary Economics 13: 77–94.

Wollmann, H. ed., 2003. Evaluation in public-sector reform: concepts and practice in international perspective. Edward Elgar Publishing.

Word, J., Stream, C. and Lukasiak, K., 2011. What cannot be counted: Ethics, innovation, and evaluation in the delivery of public services. The Public Sector Innovation Journal, 16(2), pp.2-17.

Yang, K.; J.Y. Hsieh and T.S. Li (2009) 'Contracting capacity and perceived contracting performance: Nonlinear effects and the role of time', Public Administration Review, 69, pp.681-696.

Yong, T. and Park, C., 2017. A qualitative comparative analysis on factors affecting the deployment of electric vehicles. Energy Procedia, 128, pp.497-503.

Annex A. Tools to support evaluation

Inspiring Impact Resource Hub

Resource Hub is a one-stop shop for impact resources and tools relevant to improving impact practice, such as guidance, research, outcomes frameworks and surveys.

Koala Evaluation Tool

The Koala (KnOwledge And Learning evAluation) tool from ImaginationLancaster, University of Lancaster to evaluate facilitated workshops and events in a fun and informative way. The tool can be used to get feedback from participants of workshops or events and can be changed to suit your own context (i.e. the questions you want to ask) as long as the terms of the Creative Commons License are adhered to.

Mentimeter

Free online voting tool to gather baseline information and feedback about what participants wanted and then what they were taking away. The tool projects word clouds in real time giving instant feedback to the organisers and participants. Participants require a smart phone to vote and the venue has to have wifi.

miituu

miituu is an online video channel that enables organisations to capture, manage and share digital stories via your own secure private channel. You are able to respond using webcam, Android or Apple devices and miituu takes care of all the video transfer, hosting, encoding and bandwidth. The miituu App can be rebranded according to an organisation's brand guidelines. Once miituu App is purchased and you have collected people's stories you can share them privately, publicly, individually or collectively using the miituuTV videowall.

Wordle

Wordle creates "word clouds" from text that you provide. The clouds give greater weight to words that appear more frequently in your text. The images you create with Wordle are yours to use however you like. They can help you to think about the most important issues that come up in reports of your word.

Weaver's Triangle

This tool can help you to clarify the impact you want to make and separate your aims, outcomes and activities.

Theorymaker

This free and open-sourced web-based tool was made by Steve Powell as a quick and simple way of creating a theory of change.

The Rickter Scale®

The Rickter Scale® is a flexible, multi-sensory assessment process, designed specifically to measure soft outcomes and distance travelled. The individual uses a hand-held board that engages them to explore a baseline and build goals towards a desired state by scaling, producing a self assessed action plan. It can also be used to engage and measure Groupwork outcomes. The online Impact Management System is then recommended to record and evidence qualitative and quantative data for funders and stakeholders. There is a cost for the Rickter Scale®, training and the IMS.

The Evaluation Game

This board game is specifically designed by Leapfrogtools as a part of one of their short projects to gather stories (qualitative data) through creative evaluation.

The Blob Tree

The Blob Tree developed by Ian Long and Pip Wilson can be used as an evaluation tool. It has been developed particularly for schools, hospitals, youth workers, psychologists, carers, counsellors and anyone working with people. It can be used in many different settings. It is copyrighted and you need to purchase it before you use it. A downloadable copy and license for use can be purchased from the Blob Tree Shop.

Survey Monkey

Survey Monkey is an online survey tool. You can also pay to create more complicated surveys and to access the tools to analyse the results (otherwise you have to analyse by hand).

Social reporting: a primer

Social reporting is the use of social media to record and share these thoughts and discussions. In digital format, sharing becomes easy.

This guide tells you how to use social media to capture learning and discussion at event.

Self-evaluation health check

Using this simple assessment tool will help you discover how well you are evaluating your project, work or services. It uses the evaluation pathway to help you identify your areas of strength and the weaker spots.

Quirkos

Quirkos is a qualitative analysis software package which has been developed by social researchers to allow researchers to visually analyse qualitative text data and present them statistically. This tool can be used by the third sector organisations to use and present qualitative evidence such as feedback from volunteers, staff or service users. There is a one-off purchase cost. This is discounted for Third Sector organisations.

Big Picture Route Map

This cartoon map has different vehicles, routes and characters. This visual tool can help participants reflect on their current situation or experience of an event and to help start discussions on what the next step for them might be.

Evaluation Wheel

Creative and flexible tool to collect information on outcomes in a simple and accessible manner.

Useful for:

- Provides many different types of information.
- Showing journeys of change – results can be made into a spider diagram and overlaid giving a visual representation of progression.
- Results can be aggregated into wheels for annual reports.
- Simple analysis as it is essentially a scale questionnaire in circle form.

Sticky wall

A sticky wall is one way to collect qualitative feedback about outcomes from a large group. It can be particularly useful to find out whether you have achieved the planned outcomes at an event.

Useful for

- Fairly quickly collecting qualitative feedback from large numbers of people
- Allowing participants to see the evaluation feedback for themselves.