Public Sector Innovation Indicators: Towards a New Evaluative Framework

Research Report no. 6 of the LIPSE Project (Work Package 6)

Authors:
Rainer Kattel* (Tallinn University of Technology), Aleksandrs Cepilovs (Tallinn University of Technology), Tarmo Kalvet (Tallinn University of Technology), Veiko Lember (Tallinn University of Technology) and Piret Tõnurist (Tallinn University of Technology)

* Corresponding author
Ragnar Nurkse School of Innovation and Governance
Tallinn University of Technology
Akadeemia tee 3, 12618, Tallinn (Estonia)
rainer.kattel@ttu.ee

The research leading to these results has received funding from the European Union’s Seventh Framework Programme under grant agreement No. 320090 (Project Learning from Innovation in Public Sector Environments, LIPSE), Socioeconomic Sciences and Humanities. LIPSE is a research programme under the European Commission’s 7th Framework Programme as a Small or Medium-Scale Focused Research Project (2011-2014). The project focuses on studying social innovations in the public sector (www.lipse.org).
# Table of contents

- Executive summary .................................................................................................................. 4
- 1. Introduction .......................................................................................................................... 8
  - 2.1 Defining public sector innovation ...................................................................................... 10
  - 2.2. Indicators on private sector innovation .............................................................................. 20
  - 2.3. Measuring change and productivity in the public sector .................................................. 25
  - 2.4 Big data, social media, and randomised trials: Moneyball for indicators? ...................... 33
    - Big data and social media ...................................................................................................... 34
    - Randomised trials ................................................................................................................ 35
  - 2.5 Summary: key issues from the literature ........................................................................... 37
- 3. Current attempts to develop indicators for public sector innovations .................................. 39
- 4. Work package 1-5 input for indicators .................................................................................. 48
- 5. Evaluative framework .......................................................................................................... 52
- 6. Empirical investigation: cases from the city of Tallinn ......................................................... 57
  - 6.1 Technology and public procurement .................................................................................. 57
  - 6.2 Methodology ..................................................................................................................... 60
  - 6.3 The case of the city of Tallinn ............................................................................................ 61
    - Spatial planning registry ....................................................................................................... 61
    - Property registry .................................................................................................................. 62
    - Operative information database .......................................................................................... 63
    - Indicators used ...................................................................................................................... 65
    - Indicators not used – changes in power, control, and accountability .................................... 67
  - 6.4 Discussion: ‘Future ain’t what it used to be’ ..................................................................... 74
- 7. Experimental spaces for public sector innovation: new avenues of measurement ............. 80
  - 7.1 Introduction: surge of experimental spaces in the public sector ........................................ 80
  - 7.2 Methodology ..................................................................................................................... 82
  - 7.3 Innovation labs in the public sector .................................................................................... 83
General characteristics ........................................................................................................................................... 83
Evaluation as a gateway to autonomy ................................................................................................................... 86
Role of i-labs in the public sector ........................................................................................................................... 88
7.3 Discussion ..................................................................................................................................................... 89
8. Summary and key policy recommendations ......................................................................................................... 90
References ............................................................................................................................................................ 97
Appendix ............................................................................................................................................................... 110
Executive summary

In this report we summarise the research results of work package six of the LIPSE project. The main focus of the work package was on how to evaluate and measure public sector (social) innovations.

Policy evaluation and, in particular, measurement of policy impacts and representing the results via indicators has become one of the key domains of public policy research. As Osborne and Gaebler argued in 1992, ‘what gets measured, gets done’. Simply put, our research shows that what gets measured is not necessarily what needs to get measured and that evaluating innovations should be conceived in a much more complex framework than is currently done.

In this research report, we analyse the foundations of public sector innovation measurement and develop a new analytical framework to evaluate public sector innovation sets. A main finding reported in this paper is that rather than using individual indicators to understand and measure public sector innovation, it is instead advisable to use complex evaluative frameworks. Such frameworks are not riddled with conceptual and measurement difficulties endemic to many activities in the public arena; evaluative frameworks allow for both internal and external evaluations to take place over longer time periods and, equally important, such frameworks make it easier to analyse failed or partially failed public sector innovation attempts and to locate reasons for failure and key lessons to be drawn. The latter aspect, in turn, enables the connection of public sector innovation evaluations with existing performance management, auditing and risk governance practices – in other words, public sector innovation evaluation could become part of larger evaluation, feedback and policy learning tools and practices. Eventually, evaluative frameworks could ensure that feedback that emerges from evaluation processes reaches policymakers in an understandable and useful form.

Our research took place in four phases:

The first phase of the research project consisted of a state of the art review of public sector innovation measurement. We found that next to productivity and performance dynamics, public sector innovation is, on the most abstract level, related to public sector authority and legitimacy. New technological developments and data sources (incl. social media, big data) create novel opportunities to capture the latter; as of now, however, these possibilities are essentially not utilized in public sector settings. Thus, a wide range of information about the public sector innovation process is missed in measurement practices. Furthermore, public sector innovation indicators do not capture the role of the public sector in driving technological change.
In the second phase of the research, we studied prior innovation measurement efforts (MEPIN, EPSIS, APSII, NESTA, GII) and interviewed the experts involved with these initiatives. Our study showed that political demand for public sector innovation indicators is crucial in initiating, facilitating and funding measurement efforts. Prior attempts can be directly tied to justifying and legitimizing public sector activities; however, none of the reviewed measurement efforts questioned the influence of innovation on public sector legitimacy and trust. Surprisingly, productivity and performance measurement also did not play a large role in prior measurement efforts. As surprisingly, none of the attempts to create public sector innovation indicators used performance measurement and quality management tools (e.g. self-evaluations, etc.) widely used in many public organisations. Thus, existing public sector innovation indicators fail to take on board key lessons learned in performance measurement and quality management over the last decades. More complex relationships (e.g. capturing the effects of networks, or co-creation) within the public sector innovation process were also not included.

Following the aforementioned work, in the third phase of our research we developed a new evaluative framework for the study of public sector innovation indicators. We proposed an evaluative framework for public sector innovations based on three basic logics: first, the logic of changes; second, the logic of feedback, and third, the logic of technology. This creates a new, three-dimensional evaluation framework for public sector innovation evaluation. This was discussed with various stakeholders, which led to further elaboration of the 3D evaluative framework.

In the fourth (final) phase, the framework was actualised in two empirical studies: e-procurement case studies from the city of Tallinn and a global survey and analysis of new experimental spaces for innovation in the public sector. The cases from Tallinn illustrated that currently used evaluation frameworks are relatively narrow and often determined by limited public procurement frameworks. The study of innovation labs (i-labs) focused on new forms of public sector organisations that attempt to fundamentally challenge how innovations are conceptualised, designed and evaluated in the public sector. We show that while i-labs are set up to disrupt current design, evaluation and implementation practices, most such labs currently operate on the relative fringes of the public sector; the key challenge is how to bring lessons learned in i-labs to the core of public policymaking.

In total, to write this report over 40 semi-structured interviews, additional participatory research findings, document analysis, and two seminars with practitioners were undertaken.

Based on the research we can outline the following policy suggestions:

- Policymakers and civil servants should refrain from using single figure indicators (such as organisational or country rankings) in designing and evaluating new solutions, both technological and others. Public sector innovations should be designed and evaluated in a
wider comprehensive set of evaluative frameworks that allow more complex issues to come to the fore and also allow evaluative exercises to take place over longer periods of time.

- Evaluative frameworks should pay attention, first, to efficiency gains and wider policy goals (such as trust in the public sector, legitimacy of public sector activities), second, to where and how proposed innovations will have the most impact, and third, to how proposed changes relate to the technological (mostly ICT) capabilities of existing civil servants and users.

- It might be advisable to establish specific organisations (such as i-labs, publicly owned innovation companies) to experiment on a small scale with new ideas, solutions and partners. In this way, the involved stakeholders and the public sector can learn from mistakes and failed innovations, and also use diverse evaluation methodologies and frameworks. However, it is important to devise mechanisms for key lessons from such organisations to reach core areas of the public sector.

- Evaluation and measurement efforts need to reflect innovation capacities and, in particular, their dynamic nature: thus measurement should reflect factors influencing innovation capacities (such as those needed to be involved with networks, for co-creation practices, and to carry out risk governance regimes, auditing and performance measurement practices).

- It is, accordingly, important to connect public sector innovation measurement efforts with wider performance measurement attempts, accountability procedures, and risk governance measures. Public sector innovation indicators will invariably feed into accountability systems and risk governance practices.

- Rapidly evolving ICT will have a sizable impact on the nature of data available to the public sector, especially on the city level (smart city solutions). Without developing data-driven and dynamic user interfaces, key public values (such as trust) will remain only vaguely captured in the design and evaluation of new services. However, these new interfaces require quite new and different capacities both from the individual bureaucrats and the institutions involved (in terms of giving larger access to users in design and evaluation). Otherwise, public sector organisations may end up using data metrics that leave out the most important characteristics of public sector innovation dynamics.

- New organisational forms and experimentation (with co-creation, auditing, procurement) are important elements to take into account in measurement practices. More interaction-enabling public procurement frameworks are essential for removing some of the main barriers of innovation and technological developments. This can be achieved, for example, by using commonly negotiated procedures or communicating technology needs early. Importantly, it is not just what the law is, but also how the law gets interpreted in certain contexts. Therefore, investments in procurement capabilities – including into the legitimisation of risk-taking – constitute an important avenue for changing public sector innovation feedback mechanisms.
One of the key recommendations from our research is that organisations should vocalize and formalise their innovation and procurement strategies in evaluative frameworks before they set out to procure new technological solutions. The research results indicate that the evaluative framework offered in this report serves as a useful starting point for that.

And last, but not least, prior to putting together new public sector innovation evaluative frameworks, political support needs to be in place to develop, fund, and facilitate such efforts. This means that public sector innovation indicator frameworks should also be subject to a wider political debate, so that they do not fall victim to political criticism when they feed into broader accountability systems and encounter public scrutiny.
1. Introduction

In the Description of Work, the framework for our research was presented through the following research goals for Work Package Six:

- To adjust and refine the existing overall framework based on an updated literature review examining the factors that influence social innovation in the public sector
- To compare the results of WP 1-5 based on a cross country and cross sectoral meta-analysis
- To identify social and public sector innovation types, indicators and indices
- To identify major theoretical, methodological and practical strengths and weaknesses of existing sets of indicators and indices
- To develop a theoretical and methodological framework for the development of social innovation indicators in the public sector as well as formulate specific indicators
- To test the developed framework and set of indicators as well as the guidelines on how to use both
- To disseminate research results and policy recommendations

This report covers all of the objectives listed above except for the first objective, which is left out as it is covered elsewhere (see Voorberg et al. 2015; de Vries et al. 2015). In addition, while undertaking the work package, we added an additional task. Namely, it became clear that the emergence of innovation labs (i-labs) in the public sector plays a crucial role in how innovations are perceived, evaluated and measured in many countries. Thus, we decided to add the following objective:

- To develop a database of existing i-labs, understand the morphology of these new organisations, and, in particular, their role in redefining innovation in the public sector and in social contexts

Consequently, the report consists of three empirical studies: the state of the art in public sector innovation efforts, an investigation of the use of public sector innovation indicators in the city of Tallinn, and a study of public sector innovation labs. For the first, a significant desk research was undertaken combined with 13 in-depth, semi-structured interviews with project members from various public sector innovation initiatives around the world. For the case studies in Tallinn, in addition to document analysis, over 25 interviews with public and private sector stakeholders were carried out. The research team also followed the activities of the city of Tallinn between December 2013 and June 2015: participated in their development meetings (among them the e-service working group) and followed the management
meetings of the ICT developments selected for review in this study. For the study of innovation labs, an in-depth survey was carried out in addition to 11 interviews with the heads of innovation labs from around the world.

The findings of this three-part analysis were presented at the LIPSE mid-term conference (February 2015) for researcher and practitioner input and in two seminars organised by the research team. The first seminar, to test the findings of working package 6 (scientific validity and policy relevance of the new analytical framework), was held for the social innovation network of Estonia (April 28, 2015). The second seminar was part of the quarterly meeting of the e-service development group in the city of Tallinn (May 5, 2015) for an audience of policy practitioners and heads of municipal departments (more than 20 policymakers participated). While the first seminar was more geared towards an academic audience, the second was more practical and included a presentation from the city of Tallinn on the system of public sector indicators the city itself was developing. Feedback from these seminars was used to develop the new analytical framework further (e.g. the 3D model was born only after the aforementioned discussions). Additionally, the findings of the study were presented at the annual conferences of EGPA (2013-2015), IRSPM (2015), and ICPP (2015), in addition to other smaller conferences.

This report is structured as follows:

- We first discuss measurement dilemmas in the public sector context (section 2)
- Second, we discuss current attempts to develop indicators for public sector innovations (section 3)
- Section 4 summarises work packages 1-5 of this project and their main input for the evaluation framework to be developed in the next section
- Section 5 details the new evaluative framework for public sector innovations and utilizes this framework in the context of current public sector innovation indicators (section 2)
- Section 6 tests our evaluative framework with three case studies from the city of Tallinn
- In section 7, we discuss the emergence of i-labs and their impact on public sector innovation debates and evaluations
- The final section summarises our work and offers key policy recommendations
2. Theoretical framework: measuring public sector innovations – overview of key issues

Henry Mintzberg (1996) has succinctly summarised problems with measuring public sector performance in general: ‘Many activities are in the public sector precisely because of measurement problem: If everything was crystal clear and every benefit so easily attributable, those activities would have been in the private sector long ago’. On the other hand, William Baumol states in his 1967 essay that ‘the very progress of the technologically progressive sectors inevitably adds to the costs of the technologically unchanging sectors of the economy, unless somehow the labour markets in these areas can be sealed off and wages held absolutely constant, a most unlikely possibility’. In his view, many public services fall under this ‘unprogressive’ category, meaning that the costs of public services are bound to increase as overall productivity of a society increases. These two statements give an overall layout of the difficulties faced when measuring public sector innovations: first, many activities in the public sector are inherently difficult to measure and, second, the very nature of these services in relation to the rest of the economy indicates that we might need a relatively different approach to measuring improvements and failures resulting from public sector innovations. Discussions of the past 150 years over the nature of public sector innovations have indeed centred around the specifics of public sector innovations. In the next subsection, we give a brief overview of these discussions.

2.1 Defining public sector innovation

By and large we can divide scholarly efforts to delineate and conceptualise public sector innovation into three periods: 1) Schumpeterian period: innovations and the public sector are related to a larger theory of how evolutionary change takes place in societies, mainly associated with Schumpeter (1912 and 1939); 2) organisational theory period: innovations in the public sector are seen similar to innovations in private companies, mostly associated with early organisational theory and with Wilson (1989); 3) autochthonous theory period: the most recent trend to disassociate public and private sector innovations.

The Schumpeterian period is characterised by Schumpeter’s theory of innovation, which is an application in economics and business of his wider theory of how evolutionary change takes place in societies. Alas, Schumpeter never really developed his wider theory of social change (see also Andersen 2009). In his Business Cycles (1939, 97), Schumpeter states, in a footnote, that he ‘believes,

---

1 Here and beyond public sector is defined as follows: ‘those parts of the economy that are either in state ownership or under contract to the state, plus those parts that are regulated and/or subsidized in the public interest’ (Flynn 2007, 2).

2 We use public sector innovation to also include social innovation; that is, we do not delineate between these concepts and discuss social innovation only tentatively in later parts of this paper.
although he cannot stay to show, that theory [of innovation] here expounded is but a special case, adapted to the economic sphere, of a much larger theory which applies to change in all spheres of social life, science and art included’. His 1912 Theorie der wirtschaftlichen Entwicklung (The Theory of Economic Development)\(^3\) apparently assumes a similar theory, again without going into greater detail. We can deduce that what Schumpeter meant by this larger theory of change in social life is that change is driven by entrepreneurial, creative persons, or ‘new men’ as he called them in 1939, that look for ‘new combinations’ (innovative solutions) and thus bring forth evolutionary changes, entirely new ways of doing things (in business, politics, art, science, etc.) that will spread, in some cases more than others, throughout the given sphere of life.\(^4\) Some of these changes will change value systems and disrupt incumbent hierarchies.\(^5\)

In the economic sphere, such individuals drive innovations and, thus, economic growth. The role of the public sector in entrepreneurial innovation is twofold: first, the public sector can take on the role of the entrepreneur (in fact, Schumpeter argues that in socialism, as there is no private ownership, the state will be the sole innovator; 1912, 173); second, innovations in businesses can also be ‘called forth’ by governments (1939, 84).

In summary, what we can take from Schumpeter is that since early theories of innovation, the public sector has had a dual character vis-à-vis innovation: the state itself can be changed by innovators and the state can play a crucial role for business innovations as well (either by directly leading or indirectly supporting entrepreneurial activity). As we will see below, this foreshadows rather closely the currently emerging conceptual dichotomy between innovations in the public sector and innovations through the public sector (European Commission 2013a; EU Expert Group on Public Sector Innovation 2013).

**Organisational theory period.** Research explicitly dealing with innovation in the public sector goes back to at least the 1960s; however, its inception seems somewhat accidental in nature. Researchers in organisational theory dealing with innovation and how organisational structure supports creative work and novel ideas often did not differentiate between public and private sector organisations (this non-differentiation goes, in fact, back to Taylor’s Principles of Scientific Management as well as to Weber’s bureaucracy as an ideal type for both public and private organisations). For instance, Thompson talks

---

\(^3\) We use here the original German first edition because in later editions (which served as the basis for English translations), these discussions were cut by Schumpeter; for example, the second chapter of the original edition (in which Schumpeter discusses his theory of innovation) runs almost 100 pages, the English translation carries only half as many pages.

\(^4\) ‘Das erste Moment, die Freude am Neugestalten, am Schaffen neuer Formen der wirtschaftlichen Dinge ruht auf ganz denselben Grundlagen wie das schöpferische Tun des Künstlers, des Denkers oder des Staatsmannes’ (1912, 142).

explicitly about business and government organisations and their ‘capacity to innovate’ (1965, 1), and defines innovation as the ‘generation, acceptance, and implementation of new ideas, processes, products or services. Innovation therefore implies the capacity to change or adapt’ (1965, 2; see also, for example, Mohr 1969). Much of the subsequent management and organisation theory literature dealing with innovation moves effortlessly from the private to public sector and back, and deals mostly with the paradox of managers calling for innovative ideas that end up meeting resistance in implementation often from the same managers or organisational structures (Lynn 1997). This strand of research deals mostly with diversity of tasks and incentives in an organisation (Becker and Whisler 1967 is a good overview). One of the key figures in this tradition is James Q. Wilson, who’s definition of (public sector) organisational innovation remained largely the same from the 1960s to 1980s: ‘real innovations are those that alter core tasks; most changes add to or alter peripheral tasks’ (1989, 225). Wilson, without referring to Schumpeter, understood these alternations in core tasks to be evolutionary in nature and in impact: ‘Government agencies change all the time, but the most common changes are add-ons; new program is added on to existing tasks without changing the core tasks or altering the organisational culture’ (Ibid.).

Thus, there is a rather extensive literature that emerged from organisational theory that incidentally or purposefully deals with public sector innovation and where the latter is defined more or less similarly from the 1960s to 1990s. This literature uses more or less varied Schumpeterian notions of innovation, but it almost does not differentiate at all between private and public sectors, thus innovations in any organisation can be defined as significant and enduring changes in core tasks. In this way innovation should be different from incremental changes in organisations (public or private) and are similar to (technological) breakthroughs familiar to the private sector evolutionary literature (see, for example, Lynn 1997 who explicitly uses the concept of breakthrough).6

Autochthonous theory period. In the 2000s, literature dealing with public sector innovation tries to move away both from private sector Schumpeterian approaches emphasizing novelty in action and from organisational level changes towards innovation genuinely attributable to the public sector and towards discussing innovations in public services and governance (see, for example, Hartley 2005; Moore and Hartley 2008; also Verhoest et al. 2006; Pollitt 2011; De Vries et al. 2015 offers a thorough overview). However, while there is a distinct attempt to discuss public sector phenomena (i.e. decentralization of

6 Ironically, while this is indeed important for the early Schumpeterian literature, from the 1970s and 1980s onwards, evolutionary economics develops complex theoretical frameworks that show how routine-based individual skills and company level behaviour leads towards a higher level of complexity and helps to explain how Schumpeterian creative destruction shapes economies and competitive environments (see Dosi 1984; Nelson and Winter 1982). This leads to learning economies and national systems of innovation approaches that seek to explain innovations not only as breakthroughs but indeed as incremental everyday changes in company routines, learning, and various levels of interactions (e.g. user-producer; see Freeman 1982 and 1987; Lundvall 1992). Thus, evolutionary economics dealing with private sector innovations moves during the 1980s almost exactly in the opposite direction as the emerging public sector literature.
agencies or regions) and move away from private sector categorization and concepts (such as product, service and other types of innovations, concepts of life cycles and trajectories), there is hardly any substantial change in terms of conceptually differentiating public sector innovations from private sector ones. The main tenets are still changes that are new to the organisation and that are large and durable enough (e.g. Hartley 2005, 27; Moore and Hartley 2008, 5). Hartley, for instance, delivers a useful discussion of the difference between public sector innovations in traditional, new public management, and network-based paradigms of public administration (2005, 28-30). Yet, her conceptual framework is hardly different from Wilson. Similarly to organisational theory literature, the most recent literature on public sector innovation also sees innovations in the public sector as something different from incremental improvements and that can also fail and not lead to better public service. Thus, e-voting would constitute for most public sector researchers a real innovation and yet some would argue that this innovation did not really bring any improvement or at least that the jury is still out. However, in most cases, the line between innovation or not, improvement or not, is not only tenuous at best, often it seems plainly arbitrary. Moore and Hartley 2008, for instance, use as examples contracting out and private public partnerships, in other words, public sector innovation is another term for NPM-style reform practices. Thus, what is and what is not an innovation, seems rather arbitrary or subjective and this is further complicated by the fact that most attempts at measuring public sector innovations use surveys (as we will show below) – in essence further cementing subjectivity into the discussion.

In Table 1 we sample this most recent period in defining public sector innovations.

**Table 1. Public sector innovation definitions from the past two decades**

<table>
<thead>
<tr>
<th>Source</th>
<th>Definition</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynn (1997)</td>
<td>'Innovation [in government] is properly defined as an original, disruptive and fundamental transformation of an organisation's core tasks. Innovation changes deep structures and changes them permanently'.</td>
<td>The definition echoes Wilson's 1989 use of the concept and is aimed at differentiating any change in the public sector from deeper transformations that can be called innovations.</td>
</tr>
<tr>
<td>Moore et al. (1997)</td>
<td>'Changes worth recognizing as innovation should be…new to the organisation, be large enough and durable enough to appreciably affect the operations or character of the organization'.</td>
<td>General definition involving novelty and change in relation to the organisation and its overall operations. This means widespread improvements in both governance and service performance to increase public value (Moore, 1995).</td>
</tr>
</tbody>
</table>

See Drechsler 2005 on the role academic and policy discussion plays in such relabelling practices.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Definition</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newman et al. (2001)</td>
<td>Public sector innovation can be constituted as a ‘discontinuous or step change, as something which was completely new to a particular local authority (though which may have previously been applied elsewhere), and a change which had already been implemented rather than just an aspiration or planned initiative’.</td>
<td>Change-based definition that accentuates also incremental innovation and puts the focus on implementation.</td>
</tr>
<tr>
<td>Green et al. (2001)</td>
<td>[...] doing something new i.e. introducing a new practice or process, creating a new product (good or service), or adopting a new pattern of intra- or inter-organisational relationships (including the delivery of goods and services).</td>
<td>Emphasizes that simple organisational change does not equate innovation. Definition originally concentrates on service innovation.</td>
</tr>
<tr>
<td>Mulgan and Albury (2003)</td>
<td>‘New ideas that work […] successful innovation is the creation and implementation of new processes, products, services and methods of delivery which result in significant improvements in outcomes efficiency, effectiveness or quality’.</td>
<td>Emphasis put on implementation and successful innovations that have a significant impact in the public sector (implying radical change).</td>
</tr>
<tr>
<td>Hartley (2005)</td>
<td>The definition starts from the fact that innovation is not always ‘a physical artefact at all, but a change in the relationships between service providers and their users [...]Therefore, we should consider innovations, particular radical or complex ones, to be multidimensional, specifying the dimensions (and the size of the innovation in those dimensions) in the interests of systematic comparison.’</td>
<td>Takes note of the definition from Moore (see 1995; 1997) and extends the public value oriented approach to include different forms of innovation: product, service, process, position, strategic, governance and rhetorical innovations. Also diffusion and dissemination – spreading good practice and adapting/adapting existing innovations – is included as a significant part of public innovation.</td>
</tr>
<tr>
<td>Osborne and Brown (2005; 2013)</td>
<td>‘The introduction of newness into a system usually, but not always, in relative terms and by the application (and occasionally invention) of a new idea. This produces a process of transformation that brings about a discontinuity in terms of the subject itself (such as a product or service) and/or its environment (such as an organisation, market or a community)’.</td>
<td>Emphasizes ‘newness’ and discontinuity of change in the public sector.</td>
</tr>
<tr>
<td>Albury (2005)</td>
<td>Public sector innovation is ‘the creation and implementation of new processes, products, services and methods of delivery which result in significant improvements in outcomes efficiency, effectiveness or quality’.</td>
<td>Emphasizes implementation, significant improvement and creativity.</td>
</tr>
<tr>
<td>Koch and Hauknes (2005)</td>
<td>‘Innovation is a social entity’s implementation and performance of a new specific form or repertoire’</td>
<td>Functional distinction of public sector innovation that is shaped within the context</td>
</tr>
</tbody>
</table>
Halvorsen et al. (2005) define public sector innovation as ‘change in behaviour’. Very broad definition concentrated on change.

Mulgan (2007) offers a definition: ‘The simplest definition is that public sector innovation is about new ideas that work at creating public value. The ideas have to be at least in part new (rather than improvements); they have to be taken up (rather than just being good ideas); and they have to be useful’. The definition makes an additional requirement of implementation – of being ‘taken up’ – meaning that, in terms of measurement, some time lapse before change and impact is required.

European Commission 2013a; EU Expert Group on Public Sector Innovation 2013 structure their inventory along two lines: initiatives that would be readily considered innovations inside the public administrations, such as the shift to ICT tools and HR management (innovation IN); and initiatives that foster innovation elsewhere in society, such as the public procurement of innovation, the unitary patent or support to social entrepreneurship (innovation THROUGH). This definition tries to separate different modalities in public sector innovations rather than define in detail what innovations are.

Source: compiled by the authors.

However, in contrast to earlier periods of public sector innovation concepts, and with the exception of Lynn (1997; see also Lynn 2013), the current period of scholarship pays much less attention to the evolutionary character of changes described as innovations. This is not to say that there is not an acute awareness that one has to differentiate ordinary change from innovation. For instance, Osborne and Brown 2013 argue ‘the management of innovation is an entirely different task from the management of developmental change’ (2013, 3); Lynn similarly concurs that all non-transformative change is ‘“innovation lite”, which is indistinguishable from ordinary change’ (2013, 32). Yet, how this transformative change in fact works in the public sector – and differs from typical private sector dynamics – remains almost always unclear. Even the most advanced concepts of public sector innovation do not address in detail how selection mechanisms and other processes take place that would enable us to distinguish innovations from ordinary changes. What makes one reform or new service an innovation and the other not? Often there seem to be normative connotations involved in distinguishing innovation from change: as innovation is good, a successful reform must be innovative.

On the other hand, evolutionary dynamics dominate private sector innovation literature, evident in such concepts as backward and forward linkages, increasing returns to scale, first mover advantage, winner-takes-all markets, imperfect competition, externalities (many of these concept are present already in
Schumpeter, especially in 1939 *Business Cycles*). In fact, innovation research in the private sector is all about evolutionary change: how and why certain products, services, technologies, technology systems, but also organisational forms and institutional frameworks become dominant over others that, in turn, become obsolete or vanish altogether (Nelson and Winter 1982; Perez 2002). The role of technology, particularly large-scale shifts following technological revolutions that lead to whole new paradigms, is difficult to underestimate here.

However, such evolutionary practices and processes are simply much less evident or even lacking in the public sector. Moreover, many of these processes would not be desirable in the context of public organisations, such as monopoly rents garnered by first movers or undercutting the same first movers by imitation. There is hardly any competition within the public sector for such evolutionary processes to take place. The way innovations diffuse in the market environment, via imperfect competition and imitation, is hardly a way for public sector innovations to emerge and to diffuse. Furthermore, in business innovations, there are lots of failures at innovation and lots of losses through innovation or imitations by competitors. Again, these phenomena seem to not be present in the public sector or present themselves in a different form.

That is not to say there is no evolutionary change in the public sector. Almost all literature on public sector innovation assumes there is evolutionary change, but conceptualising the evolutionary changes in the public sector seems to have been lost in private sector terminology. The key lesson from previous literature, accordingly, seems to be that we should not attempt to look for similar processes to take place within the public sector; rather we should try to focus on evolutionary processes within the public sector that originate from logics of the public sector and pertain to such phenomena as power, legitimacy, and trust. This is arguable exactly the topic of perhaps the earliest ‘discussion’ on public sector innovation, namely between Tocqueville and Weber on the state level public administrations in the United States.⁸ Tocqueville’s analysis, and admiration, of state level administration is famous, Weber’s counterarguments are much more scattered and less well-known (Tocqueville’s was published in 1835 and 1840;⁹ Weber’s remarks can be found in *Wirtschaft und Gesellschaft* from 1922 and elsewhere).¹⁰

Tocqueville’s main question in looking at U.S. state and especially township level administration was how can diverse townships in New England, without central administration, still provide relatively uniform public services, especially under an administrative system where most public functions are fulfilled by elected officials (1876, 92). He explained this with judicial oversight of administrations and

---

⁸ For a longer historical discussion of the concept of innovation, see Godin 2008 and 2012.
⁹ We refer here to the 1848 French edition, available via Project Gutenberg, and to the 1876 English translation.
¹⁰ In Weber’s case, we use the 2009 German edition. For a comparative discussion of Tocqueville’s and Weber’s discussions on the United States, see Kalberg 1997.
called both – decentralized administration and judicial oversight – innovations (ibid.).\textsuperscript{11} In Tocqueville’s view, decentralized administration with elected officials and judicial oversight work better than centralized administrations (which, he argued, was an innovation of the French revolution; 121): centralized administrations have more resources, are good at regulating business, maintaining social order and security but also keep society equally from improvement and decline (113); centralized administrations are good at mastering resources to combat problems but they are poor at rejuvenating what might be called sociopolitical resources for change (109).

When we jump two thirds of a century later, we can see that all the ills of centralized administration described by Tocqueville become positives in Weber’s view:\textsuperscript{12} in order to keep social order, that is to retain authority and society functioning, centralized bureaucracy is the ‘technically’ better instrument over elected officials (2009, 156; further also 545-550 and 561). Elected officials and other ”schöpferische” Betätigung der Beamten’ leads rather to unpredictability and to bureaucracy that seeks to retain its own power, in other words to rent seeking behaviour (565).\textsuperscript{13}

While Tocqueville and Weber had different normative goals – the former describing the benefits of active civic life, the latter describing the benefits of a well-functioning and predictable state apparatus (also Freund 1974) – both eventually discuss how authority, to use Weber’s term, is maintained in society with competitive interests via institutional and administrative innovations (although Weber does not use the term). We can paraphrase Weber: the modern state is defined by its authority to use violence to uphold the very same authority. Above all, Tocqueville and Weber show how such innovations lead to differing sociopolitical relationships and networks, and institutional and organisational structures and cultures; in other words: how these innovations drive different evolutionary change. However, both also show why evolutionary processes in the public sector are punctured by political, legal, institutional and administrative constraints (see also Karo and Kattel 2013).

In fact, these very constraints are part of these evolutionary processes, forming simultaneously internal factors that are changed and external factors limiting changes. Constraints are intrinsic to the public

\textsuperscript{11} ‘C’est ce qui ne se découvre pas au premier coup d’œil. Les gouvernants regardent comme une première concession de rendre les fonctions électives, et comme une seconde concession de soumettre le magistrat élu aux arrêts des juges. Ils redoutent également ces deux innovations’ (‘The communities therefore in which the secondary functionaries of the government are elected are perforce obliged to make great use of judicial penalties as a means of administration. This is not evident at first sight; for those in power are apt to look upon the institution of elective functionaries as one concession, and the subjection of the elected magistrate to the judges of the land as another. They are equally averse to both these innovations’).

\textsuperscript{12} Weber’s Wirtschaft und Gesellschaft does not mention Tocqueville. However, as Weber’s wife, Marianne Weber has argued, ‘A mon avis, il est hors de doute que Max Weber connaissait les ouvrages de Tocqueville, bien que je n’ en aie aucune preuve... La parenté spirituelle entre les vues historiques et sociologiques des deux penseurs me semble une diose très plausible’ (quoted in Freund 1974, 457). Weber travelled the United States for 3 months in 1904, see extensive discussion in Kaesler 2014, 563-637.

\textsuperscript{13} Weber summarises his discussion of bureaucratic creativity as follows: ‘Entscheidend ist für uns nur: dass prinzipiell hinter jeder Tat echt bürokratischer Verwaltung ein system rational diskutabler “Gründe”, d.h. entweder: Subsumtion unter Normen, oder: Abwägung von Zwecken und Mitteln steht’ (565).
sector. Thus, to use Tocqueville’s example, judicial oversight in small townships acted as a constraint on elected officials, yet this same constraint led to better services for the citizens. Weber, on the other hand, writing two thirds of a century later, argued that modern societies have become increasingly more complex and thus require centralized administrations that can act simultaneously as constraints and enablers.

Consequently, following Tocqueville and Weber, we can argue that instead of competition as a driver and diffuser of evolutionary processes, as is the case in the private sector, intrinsic public sector features act simultaneously as constraints and enablers and engender punctured evolutionary processes as a consequence of public sector innovations. Notice that in both cases the innovations influence organisational level capacities, institutional interactions and, eventually, political authority of a state. Their recommendation, as it were, would be to look at changes in the public sector that lead to 1) changes in constraints and enablers, that relate directly to how authority is obtained/retained and 2) engender clearly discernible evolutionary trajectories in their respective ecosystem – such changes could be termed public sector innovations. Essentially, public sector innovations are such changes in the public sector, according to our interpretation of the Tocqueville-Weber debate, that realign enablers and constraints and one way or another influence the authority and legitimacy of the given public sector actor. This dimension of authority and legitimacy is almost completely missing in all historical and contemporary debates.

It can be argued that a recently emerging literature on social innovation (see Bekkers et al. 2013 for an overview) tries to fill the gap in public sector innovation literature by looking at values and social relevance and thus moves the discussions towards issues of authority, trust, etc. The concept of social innovation can be construed wholly as value-based: broadly put as democratic commitment to social change (Andersen and Bilfeldt, 2013). This can be described as change towards meeting social needs. However, the concept is also widely applied to the private sector: the effect of corporate social responsibility, business ethics, and social businesses, not to mention the development of civil society (Osburg, 2013; Schöning, 2013). Consequently, it is sometimes used as an opposite to what we perceive as the traditional public sector – the inclusion of market-reliant or philanthropic solutions that are found to be more inclusive and ‘better’ (Moulaert et al., 2013). First and foremost, the engagement of citizens is emphasized in social innovation. Nevertheless, when the positive change in responding to social needs is taken as primary, social innovation can be both bottom-up and top-down, not to mention output or process related, legislative or cultural. The important distinction is that social innovation reveals and responds (better) to social needs by creating new services or expanding equality of access to them (Martinelli, 2013). This can be done in unison with empowering users or specific social groups, thus modifying power-relations between service users and providers. Hence, it is hard to distinguish
and demarcate the extent of public sector innovation through the concept of social innovation; however, this can be accompanied into the public value based understanding of transformative change within the public sector. In other words, while conceptually social innovations are situated between private and public sector innovations (their origins can be in both sectors), it is their perceived impact of empowering, of changing power and authority relations, that makes this strand of literature also conceptually interesting.

Furthermore, social innovation as a concept can legitimize more socially oriented and solidarity-based political economy which also creates momentum behind public sector services (see Fraisse 2013).

Summarizing 150 years of discussion on conceptualising public sector innovations and innovations generally, we can, first, conceptualise the relationships within the whole innovation arena in one figure as a taxonomy (Figure 1) and, second, draw the following conclusions:

A) From the oldest literature discussing public sector innovations (Tocqueville, Weber):
   1) Public sector innovations are in the most abstract sense related to public authority;
   2) Innovations lead to evolutionary changes in constraints and enablers that are intrinsic to the public sector (rules, relationships, institutions);

B) From recent public sector innovation literature:
   3) These evolutionary processes use different modalities (innovations within and through the public sector), agency (the public sector proactively initiates changes or reacts to technological, environmental, etc., changes), and morphology (from incremental to discontinuous changes);
   4) Literature on public sector innovations rarely deals with authority (and related phenomena such as legitimacy and trust) but rather with relatively specific features of these changes, for example, with specific modalities (within public sector organisations), agency (reactions to external stimuli such as technology, politics, and social challenges) and morphology (incremental changes); most of these changes are in fact not evolutionary or their impact remains difficult to discern;
   5) Innovation is too often defined from a normative viewpoint (as something leading to significant improvement in public service delivery) rather than a process that explains how profound changes take place in the public sector;
   6) In defining innovation, the literature has focused mostly on the organisational and policy levels, but in doing so it has neglected the wider, public sector level constraints and enablers. What is argued here is that there is a need for a systemic perspective that goes beyond single instruments or decisions and that this would offer a framework against which changes in core routines on the organisational or policy level could be measured against;
Accordingly, disproportionally large areas of public sector activity in relations to innovations are under-researched and, we will argue in what follows, this leads to relatively simplistic attempts to measure public sector innovations.

Figure 1. Taxonomy of innovations, 150 years of discussion

Figure 1 is based on the assumption, derived from our literature review, that innovations in a society take place in two different contexts – technology and authority, to put it very simply – that lead to different and often parallel evolutionary changes in the society. While in capitalist economies, technology influences private sector innovations more directly than public sector ones, authority (and all that follows from this) influences more dramatically public sector innovations than private sector ones. Social innovations can be described in this context as a case of innovations where this parallelism of changing technology and authority comes most clearly to the fore, at least in our current conceptual state of the art.

2.2. Indicators on private sector innovation

The first well known effort to measure private sector innovation took place in the United States: the National Research Council started to collect statistics that we today commonly relate with innovation
Public Sector Innovation Indicators

(e.g. on R&D activities and their impact on changes in the volume of sales) in 1933 (see Holland and Spraragen 1933). Such measurement activities became more common in the 1960s when several other organisations (e.g. the Federation of British Industries) carried out related industrial R&D surveys (for an overview, see Godin 2002). The linear approach – overemphasizing R&D based creative activities to develop technical inventions – dominated until the early 1980s when the re-discovery of the Schumpeterian theories took place (Arundel & Hollanders 2008). As of today, there are basically two strands of literature that deal with indicators and the measurement of innovation in the private sector: first, competitiveness indices and indicators; second innovation and technological dynamics indices and indicators.

The first set, competitiveness measurements, do not have a clear-cut theoretical basis, but come from management research and are methodologically mostly survey based. Examples of such approaches include various applications of Porter’s approach to cluster and competitiveness (Porter 1990, 1998), but also the closely related work of the World Economic Forum (e.g. Porter et al. 2008; Schwab et al. 1999) and the Institute for Management Development (IMD 2012). At the micro (company) level, competitiveness is generally understood to refer to the ability of a firm to increase in size, market share and profitability; studies have increasingly paid more attention to non-price factors being important contributors to competitiveness, including human resource endowments, technical factors as well as managerial and organisational factors (Clark and Guy 1998).

The second approach – science, technology and innovation measurements – has as clearer basis in Schumpeterian theory and is usually based on empirical measurements and increasingly also on surveys. The following are indicated as the main areas of indicators: (1) R&D data, (2) data on patent applications, grants and citations; and (3) bibliometric data (i.e. data on scientific publication and citation; Smith 2005). While statistical analysis along those indicators is common in innovation research, specific databases and surveys have generated more insights into the matter; for example, the SPRU database on major technical innovation in the British industry (see Pavitt 1984 for the results) or DISKO surveys on technological collaboration of the University of Aalborg. Most notably the OECD, by synthesizing various innovation measurements, has developed the so-called Oslo Manual (OECD 1992; OECD 2005) that largely guides the measuring of private sector innovation activities today. The most common application is the Community Innovation Survey (CIS), carried out with biannual frequency by EU member states and a number of other countries, has given a basis for extensive econometric and statistical analysis of innovation (for an overview of such studies, see Arundel and Smith 2013).
What is common to both approaches is that the impact of the innovation on a firm’s performance is expected. The firm specific indicators – the additional turnover generated, impacts on costs, employment, and productivity – are generally applied (see OECD 2005), although several limitations were already proposed in the initial OECD manual (2005, 109-112), including the multiplicity of factors that influence innovation impacts (can be exogenous to the company), time lag between an innovation and its impact and the difficulty of asking ‘for quantitative measures of the effects of innovations in surveys, even for very rough estimates, as the calculations often require substantial analysis on the part of the enterprise’ (109).

The use of such overall proxies to grasp the true nature of innovation impacts is revealed, for example, on research on the relationship between innovation and employment; it has been concluded that ‘On the whole, economists cannot propose a clear-cut diagnosis about the employment impact of innovation, either theoretically or empirically’ (Vivarelli 2007, 729). The locus of such research has moved from the question ‘does technology create or destroy jobs?’ to questions like ‘what type of jobs are created or destroyed by innovation?’ and ‘how does the composition of skills and wage structure change’ on firm, industry and macro levels (Pianta 2004).

Most of the work on the effects of innovation concern productivity – defined as the ratio of output over input – and more specifically on labour productivity rather than total factor productivity, due to lack of data on capital and other inputs (Mairesse and Mohnen 2010). The research is largely based on the framework from Griliches (1979; 1995) linking investment in knowledge to productivity growth; its elaboration is explained in Figure 2 (by Hall and Mairesse 2006), where the square boxes denote generally measurable quantities, while key elements – knowledge capital and innovation output – are marked as oval boxes for which only coarse proxies are used. This has led to the conclusion that, although measuring the effects of innovative activities on firms’ productivity has been an active area of research for several decades, the literature still does not provide a unique answer in terms of the magnitude of this impact due to the variability and uncertainty that is inherent in innovation (Hall et al. 2009; Mohnen and Hall 2013).
While the measurement and improvement of productivity are better established in the manufacturing sector, measuring productivity of a service is not yet as well-developed or well-established (Grönroos and Ojasalo 2004), though increasing rapidly in the case of so-called information intensive (Table 2) industries. Still, services are generally related to intangibility, simultaneous consumption and production, and less R&D content, leaving several aspects for which well-developed metrics have been developed (Figure 2) uncovered. In productivity measurements, a wider view on productivity is called for (i.e. not only measuring efficiency, but a deeper inclusion of the concept of effectiveness is needed; Grönroos and Ojasalo 2004; Johnston and Jones 2004; Gallouj and Savona 2010, Gallouj and Djellal 2010) and remains challenging for measuring public sector services as well.

Both approaches share more common problems as well. First, a key challenge is related to what are the most appropriate indicators to include, both generally as well as specifically. Research on private sector innovation has made it quite clear that companies and the environments they operate in – sectors, national context, etc. – are different, posing difficulties to measure innovation across companies. Table 2 presents a commonly accepted view indicating that the main sources of technology and innovation are different across companies (the OECD’s 2010 innovation strategy also proposes categories such as ‘cost-driven’, ‘demand-driven’, ‘user-driven’ and ‘employee-driven’ innovations): for companies in some sectors it is the R&D that mostly matters, while for others innovation becomes
available once new products (e.g. new seeds) or new production processes (e.g. new machinery) are available. The nature of innovation can also be different, varying from radical product innovation and process innovations (which are mostly researched) to (less researched) incremental marketing and organisational innovation. This also leads to difficulties in measuring innovation in various companies in a consistent and statistically comparable way.

Table 2. Types of technology firm

<table>
<thead>
<tr>
<th>Supplier-dominated</th>
<th>Scale-intensive</th>
<th>Information-intensive</th>
<th>Science-based</th>
<th>Specialized suppliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical core sectors</td>
<td>Agriculture Services Traditional manufacturing</td>
<td>Bulk materials Automobiles Civil engineering</td>
<td>Finance Retailing Publishing Travel</td>
<td>Electronics Chemicals Drugs</td>
</tr>
<tr>
<td>Main sources of technology</td>
<td>Suppliers Learning from production</td>
<td>Production, engineering Learning from Design offices Specialized suppliers</td>
<td>Software and systems departments Specialized suppliers</td>
<td>R &amp; D Basic research</td>
</tr>
</tbody>
</table>


To measure sectors and countries in a consistent and statistically comparable way poses even more complications related to the aggregation problem – firm level management tools and indicators are difficult to aggregate on the national level as sectors and their dynamics differ widely. Research on national systems of innovation that emerged in the late 1980s (see Freeman 1987; Lundvall 1985; Nelson 1993) gave impetus to advance our understanding on those national contexts, promoting the idea of the economy as ‘an ensemble of connected elements not an aggregate entity’ (Metcalfe 2002), and ample research has become available since then. Still, it has remained rather weak theory-wise, also posing measurement problems.

CIS, the major innovation survey, specifically, has been criticized on the basis of definitional restrictions with respect to innovation inputs and outputs and on the limitations of applying the survey – developed on the basis of manufacturing companies – to the service sector as well as the weak inclusion of the human capital development aspect (Smith 2005). In order to integrate emerging topics in more recent innovation research, the Oslo Manual has undergone revisions. First, by extending innovation surveys to the service sector and then, by acknowledging the importance of non-technological innovation, by adding organisational (the introduction of new systems and management methods and new types of work organisation and business models) and marketing innovation (the introduction of new commercial methods, including changes in product design and promotional strategies). Still, both strands of
research face difficulties measuring and relating to innovation intangible assets, such as skills, that have a key place in evolutionary work. The OECD (2010), for example, emphasizes that more attention should be put on measures of education, entrepreneurship, economic, environmental and social outcomes, and the framework conditions that support or inhibit innovation.

To measure evolutionary dynamics on a sectoral level and even more so on a country level are complicated tasks. In recent years, we have seen how productivity has become increasingly used as a simple one data point proxy for evolutionary dynamics. The problems with this are obvious: while productivity changes indicate change within a company, sector or country, it does not actually indicate the sources of the change (massive lay-offs due to a recession can also lead to productivity increases without any innovation or skill upgrading). However, it is politically highly poignant and easy to use such figures as they are seemingly understandable to non-specialists. As we show below, this logic is also increasingly important in the case of public sector performance measurement efforts.

In summary, innovation measurement in the private sector has developed towards more multidisciplinary and interdisciplinary approaches, while the impact on productivity has been the key aspect studied. On the policymaking level, we see a tendency to emphasize single high importance figures such as productivity, R&D expenditure in GDP, or rankings in various competitiveness indices.

2.3. Measuring change and productivity in the public sector

Productivity has become one of the main proxy-concepts used in measuring changes in private sector innovation performance and there exists a rather strong consensus on what private sector productivity measurements are on the company level, less so on how useful it is on the country level. This is, however, more complicated in the case of public sector organisations and the public sector in general.

In the last two decades, the public sector has increasingly imported values from the private sector. It is expected that the former focus more on customers, on outputs rather than inputs and on more effective and efficient performance (Hoque, 2008). Since the early 1990s, this process has gathered steam with the introduction of various performance instruments (incl. many performance indicators) to the public sector (see the seminal work by Osborne and Gaebler, 1992). The goals of this process are usually tied to ideas of advancing transparency, learning, apprising, sanctioning, and showing accountability in the public sector. Simply put, ‘what gets measured, gets done’ (Osborne and Gaebler, 1992). While decentralization and responsiveness were deemed central to this reform trajectory, the idea that accountability equals performance does not inherently concur with the former ideas (Kelly, 2005). Consequently, many problems resulting from the use of indicators as performance management tools
have been exposed for decades; while the critiques are cyclically repeated and sometimes advance (e.g. Ridgeway 1956; Smith, 1995; van Thiel and Leeuw, 2002; Miller, 2003; Pidd, 2008), very few solutions have been presented to tackle these problems. In line with this, the quality of data from various sources is a perennial problem for measurement systems and, further, various rationalities can result in different interpretations of the same data (Townley, 2008).

Many different aspects can be measured in terms of services provided by government, ranging from inputs/resources, throughputs/processes, outputs, and outcomes/impacts (see e.g. Packard, 2010; Kuhlmann, 2010; Sillanpää, 2013). This has led to the proliferation of performance indicators, which does not necessarily mean that the quality of indicators themselves has improved (Lonti and Gregory, 2007). This can be a source of government overload (Lewis and Triantafillou, 2012). With the growth of indicators, a high degree of confusion regarding priorities has also been noted (Micheli and Neely, 2010). Thus, a large number of indicators in public organisations nowadays may be irrelevant, but very hard to remove (Fryer et al., 2009). Even when trying to avoid it, usually due to credibility and comparability more technocratic performance measurement is preferred (for public administrators, department workloads are usually more important than long-term outcomes or positive changes; Ammons, 2004). The simpler the representation of results, the easier it is to recall them later (see Hibbard et al., 2002). However, made simple, the measurement system can severely misrepresent result and thus, become useless.\(^{14}\)

Effectiveness in the public sector is seen in theory as value creation to the citizen, which has no real maximum and thus, is very hard to quantify (Tangen, 2005). Various measurement systems have been suggested to the public sector, from balanced scorecards to quality management models, to deal with measurement (Hasan and Kerr, 2003; Sahay, 2005), but they have also been severely critiqued for simplistic, output-centred approaches or unintended effects. It is very easy to develop ‘tunnel vision’, myopia, misrepresentation and misinterpretation, gaming and ossification, etc. (as outlined by Smith, 1995). While the general goal of indicators is to enable comparability (to select the good out of the bad), they can also add to the impetus to ignore local circumstances and tacit knowledge (already brought out by Carnegie and Wolnizer, 1996).

Consequently, the main problems of measurement in the public sector, not only technical, but also conceptual, can be summarised as follows:

\(^{14}\) This is the usual problem with single point league table ranking systems and indices (e.g. comparing schools by average exam results; Goldstein and Spiegelhalter, 1996) that is also present in the private sector (see previous section). When reasonable estimates of statistical variation would be introduced to these comparisons, most differences would disappear (Pidd, 2008). This means that the extremes – top and bottom results – can considerably vary, but for most others there can be very little difference.
1. The diverse nature of public sector services, the wide range of users and the difficulties in defining targets (Arnaboldi and Azzone, 2010). Targets do not adhere to singular profit imperatives in the public sector (Van Thiel and Leeuw, 2002). When multidimensional impacts – tangible and intangible, financial and value based, individual and system level – are concerned, prospects on how to measure these effects are largely missing in academic literature.

2. Many economic impact evaluation methods are almost impossible 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. As such, measurement in the public sector is usually limited to the ‘product’ rather than a process, ‘throughput’. Consequently, measures are faced with the problem of ‘product’ definition: for example, 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).

3. At the same time, measurement in general is static, while the processes are clearly dynamic. There are time lags connected to the effects of many policies and also public sector performance. When a long-term view is taken into consideration, present actions can be questioned (Brax, 2007). Welfare services, with traditionally very high numbers of performance indicators, are found to lack measures to demonstrate the various long-term effects (Sillanpää, 2013). These are problematic to measure, not only because of the unknown, but also most strategic planning cycles are maximally 4-5 years long. However, in many cases, success can only be shown through long-term effectiveness that is usually only possible to describe in qualitative effects (e.g. Porter, 2010).

4. As such, the application of quantitative performance measurement usually rests on proper measurement scales that decrease ambiguity and uncertainty (see Pidd, 2008). However, there are, in this regard, clearly situations where quantitative indicators are not the best measures. This occurs especially when practices in the public sector are in transition, complex issues are faced and when standards in place are contested (see Noordegraaf and Abma, 2003). This usually leads to a value-debate for which quantitative numerical indicators are not useful (they simply represent one limited dimension of value) and expert opinion based on professional agreement is more advanced for evaluation.

5. Furthermore, in situations that have been described as ‘wicked problems’ in the public sector (Rittel and Webber, 1973) – i.e. innovative and sometimes conflicting processes that are not routine – conventional control systems (incl. static measurement) are found to make no sense as a certain level of ambiguity and uncertainty is necessary in the processes (this should not be

---

15 Moreover, in innovation related long-term goals, the static outtakes of the current situation may not be that informative and they may even stand in the way of innovation. One of the major lines of critique most pertinent to innovation indicators is the fact that measurement is found to simply block innovation or reproduce the existent (Behn and Kant, 1999).
confused with statistical uncertainty that comes from the measurement system (data derived from indicators being neither reliable nor valid)).

The quest for measuring public sector productivity only amplifies the conceptual and technical problems related to general public sector performance (change) indicators. Public sector productivity has many different meanings and its significance has changed over time (Andrews and Entwistle 2013; Rutgers and van der Meer 2010; Schachter 2004). Today it is mostly regarded as a technical term, which refers to the ratio of outputs to inputs in producing public services (Pollitt and Bouckaert 2011; Andrews and Entwistle 2013; Rutgers and van der Meer 2010; Dunleavy and Carrera 2013). The definition of productivity is often used as a synonym for public sector technical efficiency and in fact the terms of public sector productivity and efficiency tend to be used interchangeably in the literature (see various definitions outlined in Rutgers and van der Meer 2010; but cf. Dunleavy and Carrera 2013). At the same time, a clear distinction is usually made between productivity and related concepts. For example, while productivity (efficiency) refers to the ‘amount of resources used to produce a unit of output’, effectiveness is mostly understood as the ‘degree to which an organization realizes its goals’ (Etzioni 1964 in Lane 1993, 191). This, in turn, implies the need to distinguish productivity from other related concepts (in addition to effectiveness, e.g. economy, cost-effectiveness, and value-for-money) when dealing with public sector performance (see also Rutgers and van der Meer 2010; Dunleavy and Carrera 2013).

This technical definition of public sector productivity has become firmly rooted in today’s public administration and management rhetoric (see only Pollitt and Bouckaert 2011). In the 2000s, the discussion moved collectively onward to also include public sector productivity and efficiency frontiers (benchmarking similar organisations in their input/outcome ratio). Challenges that both the private and public sector face – diminishing manpower in an ageing society and less available capital – are confronted with a solution from the private sector: increasing productivity growth (e.g. Waller, 2006). As is the case with the private sector – mostly assumed to be more productive than the public sector (Roessner 1977) – productivity measurements hold the allure of expressing complex dynamic processes in one number that is equally understandable to experts and lay persons.

At the same time, this technical approach to public sector productivity has been repeatedly challenged (see Rutgers and van der Meer 2010 for an overview); in spite of the simple and eloquent definition of

---

16 Eradicating the former from the process by strict, narrowly defined indicators and control system may also make action impossible. Reducing uncertainty and ambiguity is, however, the usual reaction when dealing with public sector control systems. The more autonomy is assigned with more complicated tasks – which innovation related activities surely are – the more accountability is of course expected. This is used to evaluate actions of single organisations, but also to ‘benchmark’ organisations against each other. Thus, the increase in reporting can come at the expense of taking risks and being innovative (de Bruijn, 2002). Consequently, the whole process can be subjected to ‘pigeon-holing’, due to the measurement system thinking and acting in pre-established categories.
public sector productivity, the concept has proven to be all but unambiguous and uncontested (see Table 3 for an overview of different concepts). The central normative debate has been about the proper place of productivity among other public sector values and whether the technical approach is at all useful for the public sector. For example, Gulick (in Rutgers and van der Meer 2010) has claimed that administrative productivity should be seen as a fundamental value in itself, a value which may conflict with other public values. Others (e.g. Waldo) have rejected this idea and insist that productivity cannot be in conflict with other public purposes; productivity has a meaning only if measured in terms of other values; it is part of wider value system and can have only subordinate value for some other (higher) public purposes (ibid.).

Table 3. Different meanings of productivity and efficiency

<table>
<thead>
<tr>
<th>Different meanings of productivity</th>
<th>Definition</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical efficiency</td>
<td>Ratio of organisation’s outputs to inputs</td>
<td>Dunleavy and Carrera 2013; Atkinson 2005</td>
</tr>
<tr>
<td>Allocative efficiency</td>
<td>Match between the demand for services and their supply (responsiveness)</td>
<td>Andrews and Entwistle 2013</td>
</tr>
<tr>
<td>Distributive efficiency</td>
<td>Distribution of resources or services between citizens and the relative cost to government (and society) of that distribution</td>
<td>Andrews and Entwistle 2013</td>
</tr>
<tr>
<td>Dynamic efficiency</td>
<td>Allocation of resources between current and future consumption</td>
<td>Andrews and Entwistle 2013</td>
</tr>
<tr>
<td>Substantial efficiency</td>
<td>Capacity to produce an effect</td>
<td>Rutgers and van der Meer 2010</td>
</tr>
</tbody>
</table>

For the strongest opponents, the general pursuit of productive efficiency should be rejected altogether since public managers need to foremost realize multiple democratic values rather than base their decisions on productive rationality (Andrews and Entwistle 2013; Dunleavy and Carrera 2013). According to this view, public administration should be guided by wider public purposes and related democratic values (such as equity, probity, accountability) that are necessarily contested and cannot be reduced to the ratio of inputs and outputs. Moore’s (1994) public value concept, which aims at a shift away from productivity to broader public value creation, is an example of a more recent approach reflecting the general discontent with regard to the market-loaded productivity thesis. According to the public value framework, it is the citizens who should decide and express through the democratic
process what kind of values are to be created by the public sector and how (O’Flynn 2007). Productivity may not even be among the parameters by which the processes and outcomes of the delivery of public values is evaluated against. In short, the critics insist that it is responsiveness to citizens that matters the most rather than how to structure for efficiency.\(^{17}\)

With the rise of the debate around public values, the focus in measurement has moved even more towards outcomes and processes – what value is added to the public sphere, by whom and how (see Benington and Moore, 2011). This cannot be singularly captured in individual satisfaction surveys, as many of the effects may not be directly visible to individual service users. For example, Moore, who popularised the concept in the mid-1990s, has used the balanced scorecard model for the private sector (Kaplan, Norton) for PV – ‘public value scorecard’ (2003). For broader evaluation, the ‘competing values framework’ has been proposed (Talbot, 2008). However, most studies on public value concern themselves with measuring public sector employees’ individual values and motivations, mostly in the form of surveys (e.g. O’Toole et al. 2005; Lyons et al., 2006; Stackman et al., 2006; Buelens and Van den Broeck, 2007; Meynhardt and Metelmann 2009). The new public value framework brings out competing priorities in the public sphere based on ‘values’ that also include trade-offs (Benington and Moore, 2011). Measurement systems in the public sector rarely answer how trade-offs between different values are mitigated and emerging problems are dealt with (a good overview of current literature, also on measurement, or lack thereof, is Williams and Shearer 2011).

In a rather similar vein, Rutgers and van der Meer (2010) argue that the technical concept of public sector efficiency is too narrow to embrace the true meaning of the term and that it is substantial efficiency that should be taken as the core value of public administration. For them, substantial efficiency refers to the Aristotelian meaning of efficiency as ‘the ability to get things done’ (ibid., 772), or to put it in other words, the ‘capacity to produce an effect’ (ibid., 772-773). The substantial meaning of efficiency is not related to outputs (which is part of the technical realm of the concept), but rather on force and ability; that is, ‘efficiency is the potential to pursue desired ends’ (ibid., 773).

In order to overcome the normative contradictions, Andrews and Entwistle (2013) argue that instead of abandoning the efficiency concept altogether, one should attempt to incorporate fundamental democratic values into the concept of efficiency. They stem from the position that ‘public service efficiency is essentially a product of the proper functioning of democratic institutions and the responsiveness of the state bureaucracy’ (2013, 258-259). Building upon mainstream economic theory, they delineate four dimensions of public service efficiency: productive efficiency, allocative efficiency, distributive efficiency and dynamic efficiency. Productive efficiency (or productivity) corresponds to the

\(^{17}\) Schachter (2004) provides an interesting historical overview on early debates on public sector productivity and where this dilemma is of special interest.
technical approach to public sector productivity; that is, the ratio of outputs to inputs. Both strengthening the bureaucracy (via centralization and consolidation) as well as breaking up bureaucracies (via agencification and contracting) may be outlined as central strategies for enhancing public sector productivity. Allocative efficiency refers to the match between the demand for services and their supply. This is a complex dimension of efficiency since citizens’ demand is always mediated (e.g. through politicians). The main policy options to enhance allocative efficiency may include electoral reform, fostering more participative forms of democracy, devolution and decentralization, and injecting choice and contestability into public service delivery. Distributive efficiency ‘refers to the distribution of resources or services between citizens and the relative cost to government (and society) of that distribution’ (255). Policy choices here include direct transfers via taxes and welfare payments as well as the provision of merit goods (e.g. education). Dynamic efficiency refers to the allocation of resources between current and future consumption (256). In other words, it refers to efficiency over time. It is a question of maintaining a balance between current and capital investments, and borrowing and debt repayment. It relates the issue of public sector efficiency to the rate of consumption of non-renewable resources as well as finding a proper balance between investments in social capital and ex post service delivery (e.g. health promotion vs. providing acute care). Dynamic efficiency is allegedly directly dependent on institutional design as it needs slack resources for R&D, innovation and improvement.

The normative claims that call for a much broader efficiency definition in the context of the public sector makes it obviously very difficult to actually measure public sector productivity and its dynamics. This would require one to considerably widen the scope of measurement and assume that outcomes and processes were incorporated into the efficiency calculations. Thus, further conceptualisation is needed before one can really apply this wider framework to study public sector productivity, and its relations with innovation, on an operational level. As there are many boundary crossing activities (incl. synergies, shared objectives, and measures) in the public sector, recently also system level performance measurement systems have been called for (Callender, 2011; Lönnqvist and Laihonen, 2012). As of yet, however, no such tools or measures for this kind of evaluation have been advanced. Thus, measurement should also include information on ability and support to system coordination (see Lönnqvist and Laihonen (2012) on welfare service system productivity).

Even if restricted to the narrow/technical definition of productivity, problems related to the measurability of public sector productivity are still abundant (Pollitt and Bouckaert 2011; van de Walle 2006; Atkinson 2005; Lane 1993; Dunleavy and Carrera 2013). This is mostly because of a lack of adequate data (static as well as over-time) and problems related to output valuation. In the private sector, price-weighted outputs are used to overcome complexities related to measuring and aggregating different output volumes and quality. In contrast, public services are delivered either for free or are subsidized;
thus there are no price-tags attached to public services that would provide easy and comparable indicators to value different outputs (Dunleavy and Carrera 2013). Consequently, even if the technical approach is seemingly widely used in policy and management rhetoric, it has too often been over-inflated by including a wide variety of performance criteria, it has had little to do with systematically measuring productivity change over time (Dunleavy and Carrera 2013), and it has too often been reduced to simple cost reduction campaigns with no apparent success (Hood and Dixon 2012).

All this has direct systemic level consequences. Due to measurability problems, the national accounts usually treat public sector outputs as equal with inputs, meaning that productivity is assumed to be always constant (i.e. productivity growth is assumed to be constantly zero). In addition, cross-country comparative analyses that use proxy measures have not been able to remedy the measurability problems as most of these studies have so far failed to capture public sector efficiency (productivity) dynamics in meaningful ways (van de Walle 2006).

What becomes evident is that although the concept of productivity seems to have occupied a highly prominent place in contemporary public policy and administration rhetoric, it has not, for various political, conceptual and analytical reasons, really been applied into the study and practice of public administration. Still, there have been some recent attempts at refining the narrow approach of public sector productivity (Atkinson 2005; Dunleavy and Carrera 2013). The central idea in Dunleavy and Carrera (2013) is that public productivity as a concept of output-input ratio makes only sense if applied to the organisational level and clearly de-linked from other public sector performance indicators (e.g. effectiveness and outcomes) as well as from macro-level innovations and policy-level or political factors. As they insist, ‘public sector productivity is (and must remain) a single, deliberately limited measure, focusing solely on how many outputs are produced for a given level of inputs’ (ibid., 12).

They dismiss the idea that policy-level changes could be incorporated into the study of productivity because there is simply no methodology that would differentiate between policy change and genuine innovation. Inspired by the methodologies developed for private sector total factor productivity and labour productivity measurement and earlier works on public sector productivity (e.g. Atkinson 2005),

---

18 Interestingly, this is often a politically rewarding position to have as it inflates GDP and enables artificially increasing GDP figures especially during economic downturns (Dunleavy and Carrera 2013).
19 Van de Walle (2006) has reviewed four international public sector performance studies (World Bank Governance Indicators, the European Central Bank’s public sector efficiency study, the Global Competitiveness Report and the World Competitiveness Yearbook) that one way or another attempt to measure public sector efficiency. As he demonstrates, the aggregate level rankings provide rather an unreliable basis for international comparisons as many of the datasets have inherent quality problems, it is difficult to determine what the used indicators actually measure, and the datasets often give a partial and biased view (e.g. inputs are often used as outputs, many important aspects are neglected, and efficiency criteria are often assessed only against free trade and government intervention).
20 This has also been suggested by other authors, e.g. Putnam 2003.
they suggest that unit costs can be used to derive a value of a public sector organisation’s output, which then can be divided by the total volume of inputs of that organisation.\textsuperscript{21}

This is, however, far from being an easy and automatic task. Serious efforts are needed to strike a proper balance in choosing over activities that represent the organisation best and this assumes collecting good-quality and stable data on government services’ inputs, outputs and productivity growth. It also assumes an organisation’s analytical capacity to calculate unit costs or proxies, such as share of administrative costs, and evaluate the true impact of part-time staff, outsourcing, and capital consumption on the organisation’s productivity. Moreover, quality adjustments are often needed, especially if dealing with complex services and services with high variation in quality across the public sector (ibid.).

Recent developments in public productivity measurement may offer a step forward in terms of organisational productivity, but the systemic perspective is still largely missing. If we look back at our discussion of public sector innovations and compare it to the productivity debate, we note a latent conflict: while productivity is best measured on an organisational basis, innovations are systemic, at least on the level of a sector; that is, their impact on a concrete organisation may be vague and indirect and likely also with serious time lags.

\textbf{2.4 Big data, social media, and randomised trials: Moneyball for indicators?}

Increasing collection and digitization of public and private data has led to the rise of big data guided policy efforts (e.g. predictive policing) and political campaigns, but also to efforts by governments to use social media in day-to-day work.\textsuperscript{22} Indeed, it can be argued that when it comes to measurement, big data and related trends (such as the internet of things) might offer new gold standards in what is measured and how this information is used in policymaking. Even now, e-services are measured through digital footprints (number of users, clicks, positive/negative posts, time of response), but these are rather simple and output oriented and thus, do not tell us much about how to reimagine service and better respond to citizens’ needs. Dissatisfaction with policy evaluations and one-measure indicators

\textsuperscript{21} More specifically, Dunleavy and Carrera (2013, 34) use cost-weighed outputs as a proxy value measure for private sector price-weighted outputs. They stress the need to consider the total number of the activities performed by an organisation and weigh these activities against each other according to their unit costs (or shares of administrative costs). On some occasions, the outputs should also be adjusted for quality factors, which is, however, difficult to do. To calculate total factor productivity for a given period of time, one should then aggregate the cost measures and divide these with the total volume of inputs (labour costs, intermediate administration and capital consumption) of that period. If the volume of output is divided by the total number or costs of full time equivalent staff, then a labour productivity measure is derived. All costs should be deflated if over-time analysis is carried out. Dunleavy and Carrera also provide an overview on some of the most common techniques to be used for measuring productivity change: index-based, parametric, and non-parametric techniques (ibid., 47-55).

\textsuperscript{22} See Drechsler 2015 for a wider discussion of the influence of ICT on public administration research.
Public Sector Innovation Indicators

has led to the increasing use of experimental approaches – such as the use of randomised control trials (RCTs) – in policy evaluations. What can these two emerging trends teach us about indicators for public sector innovations? Is this the dawn of Moneyball for measurement and indicators in the public sector?23 This certainly seems to be the case in the private sector, where both big data and RCTs are increasingly used in performance measurement (see, for example, Porter and Heppelmann 2015; Thomke and Manzi 2014).

Big data and social media

The potential big data and social media analyses offer to public sector performance measurement is to move from substituting real life phenomena with indicators (and underlying models that are by definition limited and partial) to trying to replicate reality as closely as possible. We can also use here the concept of digital twins used in the private sector. This is essentially an avatar of an actual product that changes as new data streams in (from production, customers, etc.; see Porter and Heppelmann 2015, 100-101). Big data and social media analysis could make it possible to create avatars or digital twins of public services and thus offer possibilities to change/better them according to feedback (on usage, satisfaction, etc.) streaming. We can imagine that in such services as transportation and utilities, such digital twins would appear rather soon as there is a growing trend to outfit especially cities with sensors, etc. (so-called smart city solutions). This could open ways for emergent and predictive governance – services are delivered on-demand (i.e. public transportation becomes a mobility service fusing public and private offerings) and/or where solutions will be needed in the near future (environmental damage, natural disasters, or policing; see Lee et al. 2014 on predictive production and self-learning machines underlying so-called Industry 4.0). This will probably also signal the dawn of predictive public sector innovation indicators: first an avatar’s performance will be measured and then, the avatar’s predictive power’s in real life.

In addition, through designing new social technologies and new ways of interaction, citizens could be better positioned to use their existing knowledge for articulating specific needs and novel ideas, and quickly provide effective solutions through either individual initiatives or collective ones (hackatons, app contests, etc.; see Townsend 2013; Morabito 2015, esp. 23-45). All this can increase the micro-control of citizens over the physical and social world. Thus, social media barometers on different policy and public service issues can become the norm instead of satisfaction surveys and other time-lagged

23 Moneyball: The Art of Winning an Unfair Game is a 2003 book by Michael Lewis that deals with the use of enhanced statistics in sports, specifically in baseball, to form competitive teams with relatively small budgets.
indicators. At the same time, IT solutions can, to a considerable extent, enable more interactive and inclusive participation in public affairs and consequently increase the legitimacy of the public sector.

However, existing efforts to use big data and social media analysis in public service delivery and measurement are rather far away from the future envisioned in this section. There are overviews on how government offices use social media (rather sporadically; see Mergel 2012; also Brainard and Ednlis 2015), attempt to use crowdsourcing apps/sites (e.g. seeclickfix.com; see Morabito 2015), and emerging analysis using twitter to understand, for instance, riots (Procter et al. 2014). Furthermore, third sector think thanks are also developing social media tools to evaluate how governments are performing. Perhaps the best knows cases are electoral software used in political campaigns (mostly in the United States; see Pasick and Fernholz 2015; but also Brazil: Lenhard et al., 2015) and predictive policing efforts (again mostly in the United States; see Ellwanger 2014; Hunt et al. 2014). To our knowledge, there are no efforts to use big data and/or social media analysis in the context of public sector innovations.

**Randomised trials**

Where standard evolutionary selection through the market is not an option, experimental or quasi-experimental designs offer an opportunity to test a new service in a particular environment, which can shed light on the feasibility of the project as well as on the effects of this service on an organisation providing the service and on service recipients. Experimental designs, such as randomised controlled trials, originating from medicinal drug trials, have recently gained significant traction in development economics (e.g. Banerjee and Duflo, 2008; Banerjee et al., 2013) as well as in education and social science research (Torgerson and Torgerson, 2008). About a decade ago the U.S. Department of Education went as far as proposing that funding should be provided to those

‘[…] program projects proposing an evaluation plan that is based on rigorous scientifically based research methods to assess the effectiveness of a particular intervention. […] this priority will allow program participants and the Department to determine whether the project produces meaningful effects on student achievement or teacher performance’ (Department of Education, 2003 in Barahona, 2010, 5).

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). This makes us question how legitimate can public sector
innovation indicators be if the RCT is the gold standard? Some authors argue that RCTs are particularly suitable for the evaluation of socially-complex interventions as they are able to filter the relevant data from background noise (Sheldon and Oakley, 2002). Other claims in defence of RCTs and other experimental and quasi-experimental designs as reliable methodologies include: claims regarding external validity; internal validity ensured by random allocation; control and reference groups.

Proponents of using RCTs in the evaluation of policy interventions also argue that experimental designs with randomisation provide a means to establish causal relations between the initial conditions, intervention and final outcome, thus making it possible to make generalisations beyond the initial sample (Torgerson and Torgerson, 2008). A very common issue, however, when one applies RCTs beyond their initial domain of drug efficacy testing is exactly the lack of ecological validity and generalisability beyond the initial sample. In order to ensure the ecological validity of an experiment the conditions under which the experiment is run must be of maximum resemblance to those in which the policy intervention (or in our case a social innovation) will be applied on a larger scale, which can potentially cause difficulties. If a trial is run under conditions that do not resemble the normal conditions under which a studied intervention will be implemented, the results of the trial will not be generalisable and, therefore, the value of RCT will be lost. A less mechanistic approach to doing RCTs based on a mixed method approach can potentially, at least to a degree, address the issue of generalisability beyond the initial sample, thus allowing for judgments regarding further diffusion of successful innovations.

If compared to the evaluation of drug efficacy, new medical treatments or other comparatively simple interventions, social policy interventions, including social innovations are much more complex and, therefore, require a different approach to designing a trial. First of all, as Kieron Kirkland argues, conducting an RCT makes sense only when a pilot of the policy intervention has been carried out and it is already clear what the possible pitfalls are and how the trial can be replicated. Designing and implementing an RCT prematurely, without going through the iterative process of adjustment in a policy pilot can result in unreliable results or costs and time required far exceeding the projected costs. In fact, in case of complex interventions, where all potential variables that can affect the impact of an intervention cannot be acknowledged, an RCT alone is not the most suitable tool for evaluation. In order to produce policy-relevant results, an RCT has to be complemented by qualitative methods, such as ethnographic studies, action research, and interviews. Thus, qualitative methods (in-depth interviews, action research) could help to account for context specific determinants which can become essential for the success of public sector innovation indicators (consider the illustrative failure of the PlayPump case in Africa; see Hobbes 2014).
In recent guidelines to designing and implementing complex policy interventions, the authors suggest the following list of dimensions that make an intervention complex (Craig et al. 2013, 7):

- Number of components and interactions between those within the experimental and control interventions;
- Number and difficulty of behaviours required from those delivering or receiving the intervention;
- Number of groups or organisational levels targeted by the intervention;
- Number and variability of outcomes;
- Degree of flexibility or tailoring of the intervention permitted.

All of these dimensions apply to the field of social policy and particularly to the field of social innovation as a policy intervention, due to the number of actors involved and interactions among those, complexity of the environment in which the innovation is implemented as well as the high degree of flexibility that is necessary in case previously untested innovative interventions are implemented. As already mentioned, running an RCT helps us to explain the effects of the treatment on a group of people. It does not, however, provide us a possibility to understand why the intervention worked in some instances and not in others, and how the intervention worked. Therefore, in order to both explain the efficacy of an intervention as well as to understand the process of implementation, a mixed-methods approach is necessary. This is an important finding for the aim of creating new public sector innovation indicators, whether they use input from RCTs or not.

In the context of public sector innovation measurements, NESTA is perhaps the leading organisation in pushing for RCTs to also be used in this context (see, for example, its Alliance for Useful Evidence\textsuperscript{24} and work done under Innovation Policy\textsuperscript{25}).

### 2.5 Summary: key issues from the literature

Summarizing our discussion, we can formulate the following key issues regarding public sector innovation indicators:

1) We can expect a relatively strong influence of private sector innovation measurement literature on public sector innovation indicator development and thus expect that
   a. efficiency and productivity are key drivers why and how indicators for public sector innovations are created and used;

\textsuperscript{24} See \url{http://www.nesta.org.uk/project/alliance-useful-evidence}.
\textsuperscript{25} See \url{http://www.nesta.org.uk/areas-work/innovation-policy}.
b. indicators mostly capture incremental changes rather than discontinuous change.

2) As there are large gaps in the public sector innovation literature on trust, legitimacy and similar issues, we can expect that
   a. these play almost no role in why and how indicators are created,
   b. but they play some role in how they are used.

3) Given the widespread use of performance measurement systems in the public sector, we expect attempts at connecting these with innovation indicators.

4) As legitimacy, trust and similar issues are difficult to measure, we expect that
   a. they are not captured at all by indicators and
   b. more direct feedback systems (big data and social media) are not used.

5) Public sector indicators capture either innovations within public organisations or through the public sector leading/supporting businesses, but not both.

6) Public sector innovation indicators capture public sector reactions to technological change and not public sector initiatives to drive technological change.
3. Current attempts to develop indicators for public sector innovations

Quantitative large-scale studies aimed at measuring innovation have previously mostly targeted innovation in the private sector (e.g. Community Innovation Survey). Although there have been previous attempts at measuring public sector innovativeness (e.g. Roessner 1977), systemic approaches to public sector innovation surveys only started to appear during the last decade. Arundel and Huber (2013) identified, through a literature search, 18 studies aimed at evaluating innovation in the public sector using 15 large scale data sets. These studies focused on developed economies, and also excluded health and education service providers. Most of the studies did not cover all five categories of the public sector, focusing mostly on public administration. We have identified 5 such experiments that resulted in actual indicators developed and used at least once:

- **MEPIN (Measuring Public Innovation in the Nordic Countries).** The MEPIN project aimed at devising a conceptual framework and a survey questionnaire. The conceptual framework is built upon insights from the Community Innovation Survey (CIS), which is intended for measuring innovation in the private sector, adjusted to suit the needs of the public sector, as well as on existing work on public sector innovation. The preliminary conceptual framework was piloted in all five Nordic countries.

- **European Public Sector Innovation Scoreboard (EPSIS).** As part of a preliminary work in developing public sector innovation indicators, an Innobarometer survey was conducted in 2010, which consisted of 24 questions. The survey involved 500 public organisations coming from across public administration, higher education, local authorities and hospitals from all 27 member states. EPSIS was developed along the lines of the Innovation Union Scoreboard that targets innovation in business enterprises, distinguishing between three factors – enablers, activities and outcomes. However, in contrast to other studies mentioned here, EPSIS also includes measurements of the impact of public sector innovation on the performance of businesses.

- **Australian Public Sector Innovation Indicators (APSII).** The Australian Public Sector Innovation Indicators project is the latest among efforts around the world to measure innovation in the public sector and it incorporates all the lessons learned from earlier efforts of other countries discussed here. In contrast to the methodology applied in the European surveys (mirroring the CIS methodology), the APSII project proposed a methodology for a pilot based on a survey conducted in two modules – an agency and an employee survey. The conceptual framework used in the APSII project for measuring innovation in the public sector incorporates...
five main themes: inputs to innovation, innovation processes, outputs of innovation, outcomes of innovation, and environmental conditions that affect innovation in the public sector.

- **United Kingdom (NESTA).** In 2008-2009 the National Endowment for Science Technology and Arts (NESTA) commissioned six exploratory studies on public sector innovation with an aim to develop a new Innovation Index. The London School of Economics Public Policy Group (LSEPPG) developed its public sector innovation index on the basis of a survey of studies performed previously, thus taking into account the possible shortcomings and benefits of the previous efforts.

- **Korean Government Innovation Index.** Government Innovation Index (GII) is an online innovation measurement tool devised in 2005 by the Headquarters for Government Innovation under the Ministry of Government Administration and Home Affairs of Korea. The GII is one of the early efforts aimed at measuring innovation in the public sector, thus providing insights for the subsequent innovation indices developed elsewhere. The GII is designed to measure innovation in government agencies, focusing on a set of innovation management components, including: innovation leadership; vision and strategy; personnel capacity; implementation of innovation; improvement of performance; and barriers to innovation. The data collection process required three randomly selected representatives of respondent agencies to fill in an online questionnaire.

In Table 4 we summarise all 5 experiments and their composition according to what they attempted to measure: inputs, processes, outputs, outcomes, impacts, and other issues.
## Table 4. Public sector innovation experiments

<table>
<thead>
<tr>
<th>Inputs/Enablers</th>
<th>Activities /Processes</th>
<th>Outputs</th>
<th>Outcomes</th>
<th>Impacts/Effects/Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EPSIS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of 'creative occupations'; Share of employees with a university degree; <strong>Quality of public services</strong> Government effectiveness; Regulatory quality; Increased efficiency of government services due to ICT; Online availability of public services; e-government development index.</td>
<td>Capacities Share of service innovations in-house; Share of process innovations in-house; <strong>Drivers and barriers</strong> Internal barriers to innovation; External barriers to innovation; Management involvement in innovation; Importance of external knowledge; Share of employees involved in innovation.</td>
<td>Innovators Share of organisations in public administration with different types of innovation; Share of new services out of all services innovations; Public sector productivity; <strong>Effects on business performance</strong> Improvements in public services for business; Impact of innovative public services on business; <strong>Government procurement</strong> Gov procurement as a driver of business innovation; Gov proc. of advanced technology products; Importance of innovation in procurement.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UK NESTA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovation capability: leadership and culture; management of innovation; organisational enablers of innovation; <strong>Wider sector conditions for innovation</strong> incentives; autonomy; leadership and culture; enablers.</td>
<td>Accessing new ideas; Selecting and developing ideas; implementing ideas; diffusing what works.</td>
<td></td>
<td></td>
<td>Improvement in organisational key performance indicators; improvement in service evaluation; improvement in efficiency; improvement context.</td>
</tr>
<tr>
<td><strong>APSII</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investment in intangible assets; Innovation expenditures; Human resources devoted to innovation; Staff skills for innovation; Staff attitudes and attributes of innovation; Sources of information; Technological infrastructure for innovation.</td>
<td>Explicit innovation strategy and targets; Systematic, internal measurement and evaluation of innovation; Management practices for innovation; Incentive and reward structures; Practices for learning and diffusing knowledge and innovations; Innovation collaboration and alliances; Perception of enablers and barriers to innovation.</td>
<td>On-going innovation projects; Types of innovations; Degree of novelty and scope of innovations (incremental vs. radical innovation); Innovation intensity; Related, intangible outputs (patents, trademarks).</td>
<td>Quality and efficiency; Productivity; User satisfaction; Employee satisfaction; Societal and environmental impacts; Other intangible effects (increased trust, legitimacy); Effects of innovations;</td>
<td><strong>Environmental conditions</strong> User innovation; Supplier innovation; Wider public sector culture and leadership identified as drivers or barriers; External political and legislative factors identified as drivers or barriers; Leadership and culture; Public tolerance of risk.</td>
</tr>
</tbody>
</table>
### LSEPPG

| R&D activities (e.g. dedicated innovation or R&D unit; dedicated strategy unit; expenditure on consumer or market research; expenditure on development and implementation of innovations); |
| Consultancy and strategic alliances (e.g. Number of joint ventures; consulting expenditures; collaboration programmes with universities); |
| Intangible assets (e.g. patents; IPR development activities; unit responsible for IPR; trademarks); |
| ICT infrastructure; |
| Human resources (e.g. staff with graduate education; job satisfaction; performance-based promotion system; churn of personnel; etc.) |

| Institutional performance (e.g. percentage of targets met; average time to deliver outputs; change programme in place; and prizes and awards for innovations) |
| e-government, online services (percentage of services that can be requested online; % of service that can be delivered online); |
| origins of innovation (e.g. how many: innovations as a result of EU regulations; innovations as a result of ministerial/political suggestions; customer suggestions; management suggestions) |

| Number of innovations developed for delivery of new outputs; number of innovations improving existing outputs; number of innovations altogether; New outcomes; |

| Number of innovations joining-up across other government organisations; number of innovations improving performance; Number of people that have been affected by innovations introduced in the government organisation for provision of new or existing outputs. |

### MEPIN

| Information sources: channels of knowledge transfer; interactions between public organisations and other actors (i.e. enterprises, citizens); |
| Driving forces: people, organisations and other factors that push organisations to innovate (e.g. new policy priorities, regulations, citizen feedback, staff, management) |
| The role of ICT; |
| Barriers: political factors, bureaucracy, other internal conditions such as lack of incentives for staff to innovate; external conditions such as resistance of users to change. |

| In-house activities: in-house R&D; internal or external training and education of staff for innovation activities; other in-house innovation activities (e.g. planning and design; market research; feasibility studies, testing and other preparatory work for implementation of innovations); |
| External activities: external R&D; other consultancy services; acquisition of external know-how (patents, licenses, etc.); acquisition of equipment/software. |
| Procurement practices - acquisition of services, components or software from ICT suppliers, contracting for management services. |
| Organizing innovation: innovation strategy; the role of management; organizing innovation activities; and organizing competences. |

### Measuring effects or objectives: efficiency, quality, ICT, organisations and staff, other factors – health and safety. |

---

Source: authors' elaboration on the basis of LSEPPG 2008; Bloch 2011; Bugge et al. 2011; DIISR 2011; Arundel and Huber 2013; Allman et al. 2011; European Commission 2013b; Eun Kim et al., 2007.
We interviewed members from all five projects. In total, 13 interviews with the team members of the aforementioned initiatives were carried out. In what follows, we summarise key results from the interviews.

1) **Motivations behind creating indicators.** In most cases, some form of political demand played an important role: as public sector innovations are an emerging issue for many governments, there is also a need to understand its impact better. Also, supranational organisations such as the OECD and the EU play a crucial role in initiating, funding and facilitating such attempts. In all cases, organisations taking up the task of creating public sector innovation indicators had significant prior experience with measuring business innovations. However, the overarching idea behind creating new indicators for public sector innovations is strongly associated with justifying and legitimizing public sector activities.

Within the wider context of political demand, current attempts to measure public sector innovation tend to also reflect the individual preferences of sponsoring organisations. For example, EPSIS was designed in cooperation with various EC directorates and thus ended up measuring both innovation in the public sector as well as innovation through the public sector, whereas other attempts focused more on innovation within the public sector. The nature of political demand also had other important consequences on how the measurement efforts evolved. The demand in general has been short-term oriented leading to the application of existing, but partially questionable, private sector innovation methodologies. The ability to compare private and public sector innovation performance was one of the driving forces behind the application of the Oslo Manual to the public sector context. In effect, the goal was to show that the public sector was as or at least comparably innovative. At the same time, the political demand has been rather inconsistent; in most cases it is not known if and how the measurement attempts will be prolonged.

2) **Theoretical discussions and ideas behind indicators.** In all cases, key people involved with developing indicators were economists by background and readily admitted that business innovations were an important theoretical vantage point. However, all also admitted that the public sector has unique logics that are often difficult to separate from each other and, consequently, to measure. Significantly, all projects realized only after the empirical exercises that they would have needed substantial input from public administration and governance scholarship. In addition, to the need to add public sector specific logics, important difficulties were discovered in applying the existing private sector innovation measurement methodologies in the public sector. Issues like innovation output, significance of innovation, and innovation failures that occupy a central place in private sector innovation measurement methodologies
were almost impossible to measure in the context of the public sector. One of the lessons was that innovation perception is different (more problematic to discern between process and product innovation or to explain what is and what is not innovation) in the public sector and thus indeed assumes a revised theoretical framework for indicators and the definition of innovation itself for the public sector. At the same time, all projects seem to have been in discussion with each other and trying to learn from each other’s successes and failures.

3) **Relevance of productivity measurement.** While productivity in the public sector was deemed an important driver of attempts to measure public sector performance, most experts argued that it is much too difficult to measure in the public sector and thus their indicators do not reflect this or do so only indirectly by studying favourable conditions and inputs assumingly capable of leading to higher productivity. Outcome measurement as a whole was very weak in all the measurement exercises.

4) **Relevance of performance measurement.** Performance measurement practices have played almost no role at all in developing public sector innovation indicators. At the same time, performance measurement has been one of the most studied issues in recent public management debates and could provide ample ideas for linking performance measurement with innovation. This seems to be strongly related to missing public administration and management competencies in all projects.

5) **Relevance of legitimacy, trust.** Legitimacy and trust played no role at all in developing indicators. In a rare occasion, it was seen as an important input to innovation, but it was not included in the measurement effort. As put by one of the respondents: ‘Trust is important, but we don’t capture it in the measurement’. Again, we can guess that it has to do with the fact that it is methodologically difficult to capture these issues and that experts with such skills were not involved.

6) **Who uses indicators and how.** Predicted use of indicators varies quite significantly from in-house management tools to country-to-country comparison and benchmarking. Also, as indicated previously, the ability to compare private and public sector innovation performance was one of the motivations behind launching these attempts. The variety of use of the indicators reflects both the mixed motivations of the individual sponsors of the measurement attempts as well as the limited validity and reliability of the results. However, it is important to note here that all the interviewed stakeholders were very aware of the limitations of the measurement attempts, which is why we can also see a rather different mix of actual use of the measurement results.

7) **Future and evaluations of indicators.** No projects have been formally evaluated; it is also not certain that any of the projects will be repeated in the future (EPSIS seems to be most certain to be repeated) or developed into a working tool for other organisations (the OECD work on
Public Sector Innovation Indicators

updating the Oslo Manual will utilize some of the work done in MEPIN and EPSIS). This seems to stem from difficulties involved both on the technical side (availability, comparability of data) but also practicality (in most cases the results do not tell much to managers, politicians or citizens and it is expensive to gather new forms of indicators). Still, all the stakeholders – in spite of the existing limitations – are positive on the need to continue with the measurement attempts and develop the indicators further. There are, however, different ideas around how to best proceed with the public sector innovation measurement attempts. By and large, these ideas fall into two groups: a) more sophisticated surveys that would better address the existing limitations, and b) more focused and narrower studies on specific public sectors and phases of innovation (e.g. adoption/diffusion).

We can also now look at the expectations we formulated based on the literature review, summarised in Table 5.

Table 5. Expectations on indicator development and actual attempts to create indicators

<table>
<thead>
<tr>
<th>Expectations based on literature</th>
<th>Actual attempts to create indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Because of the influence of private sector innovation measurement literature, efficiency and productivity are key drivers why and how indicators for public sector innovations are created and used</em></td>
<td>This is only partially the case because private sector measurement practices play an important role as far as technical expertise of developers is concerned; at the same time, the developers were acutely aware that private sector instruments and indicators cannot be used in the public sector. Also the ideas of efficiency and productivity play an important role as guiding motivations for measuring public sector performance (they were used in the conceptualisation process), but at the same time, due to data and other technical difficulties, these ideas are not strongly reflected in actual indicators.</td>
</tr>
<tr>
<td><em>Because of the influence of private sector innovation measurement literature, indicators mostly capture incremental changes rather than discontinuous change</em></td>
<td>As all indicators have strong survey elements, this seems to be the case since incremental changes are easy to report while discontinuous change is technically more difficult to define and hence also to report. In most cases, the level of change – incremental or transformative – is not captured at all or only indirectly captured (by proxies, e.g. deployed person-months) in measurement effort. However, what was also indicated by the interviewees is that in most cases public sector innovation is perceived as something leading to positive impacts. Public servants find it very difficult to relate negative changes with innovation, meaning that many important incremental changes.</td>
</tr>
</tbody>
</table>
As there are large gaps in public sector innovation literature on trust, legitimacy and similar issues, these play almost no role in why and how indicators are created.

<table>
<thead>
<tr>
<th>Public Sector Innovation Indicators</th>
<th>stay out of the focus of public sector innovation measurement.</th>
</tr>
</thead>
<tbody>
<tr>
<td>As there are large gaps in public sector innovation literature on trust, legitimacy and similar issues, these play almost no role in why and how indicators are created.</td>
<td>This was indeed so; also because no experts with public administration and management skills were involved in any of the projects.</td>
</tr>
<tr>
<td>but they play some role in how they are used.</td>
<td>All people interviewed admitted that this indeed would be an important issue for future indicators. At the same time, they brought out problems of conceptualising and measuring legitimacy.</td>
</tr>
<tr>
<td>Given the widespread use of performance measurement systems in the public sector, we expect attempts at connecting these with innovation indicators.</td>
<td>This was not at all the case, also because no experts with public administration and management skills were involved in any of the projects. At the same time, it can be argued that most projects wanted to measure and capture public sector performance and why certain organisations perform better than others. This was, however, often clothed in different, private sector innovation language (e.g. in terms of barriers, drivers, funding).</td>
</tr>
<tr>
<td>As legitimacy, trust and similar issues are difficult to measure, they are not captured at all by indicators and direct feedback systems (big data and social media) are not used.</td>
<td>This was indeed the case as admitted by all respondents.</td>
</tr>
<tr>
<td>Public sector indicators capture either innovations within public organisations or through the public sector leading/supporting businesses, but not both.</td>
<td>This was indeed confirmed; there seem to be no attempts under way to use social media or relevant feedback mechanisms for public sector innovation measurements.</td>
</tr>
<tr>
<td>Public sector innovation indicators capture public sector reactions to technological change and not public sector initiatives to drive technological change.</td>
<td>The different indicators indeed tried to capture either one or the other. There were no attempts, for instance, to capture various PPP or co-creation practices, although PPP and outsourcing were studied as indicators of innovation through the public sector.</td>
</tr>
</tbody>
</table>

In summary, while the existing attempts to create public sector indicators are rather heterogeneous, they suffer from relatively similar weaknesses. However, there are some obvious strengths as well that are common to the existing attempts: first, all attempts are aimed at creating numerical values that offer quick overviews or snapshots for managers and policymakers; and second, enable comparisons and rankings (between organisations, countries)(see some examples of the aforementioned in figures 9-12 in the Appendix). This makes public sector innovation a political tool for ‘blaming and shaming’ usually
on the country level and motivates public servants to increase innovation efforts (however, this can become more about rising up the indicator and country ranking rather than actual changes). On the positive side, the high level of aggregation on these indicators shields specific public sector organisations from shaming (also in the case of APSII, where the organisational level indicators were only available to the organisation itself and comparisons are made with the average levels) and allows them to remain open to risks and failure (part of normal innovative activity), but at the same time it allows organisations to remain sluggish in innovation if there is no direct feedback to their activities.
4. Work package 1-5 input for indicators

As part of work package 6, we also use the main findings of prior, concluded work packages as input to discussing public sector innovation indicator needs. This is outlined in detail in a separate paper (see Tõnurist et al. 2015a).

The main contribution of LIPSE was to pull together the main drivers and barriers of social innovation in the public sector (Bekkers et al. 2013; see also Voorberg et al. 2015; de Vries et al. 2015). These included environmental triggers from the political and administration system, legal culture, governance traditions (corporatist, legal, market tradition), centralization tendencies and resource dependency, but also factors from the process itself – leadership, end-user involvement, risk management, etc. – and the adoption of innovations. Following the developed integrative framework, different work packages approached innovation in the public sector from different angles. This helped to describe conceptually the relationship between internal (learning and sense making) and external (innovation environment) processes that support social innovation in the public sector. In other work packages (see figure 3 below) these drivers and barriers were analysed in more detail (also adding the element of technology adoption to the mix (see Nasi et al. 2015)).

Figure 3. Overview of LIPSE work packages and their coverage

Throughout different LIPSE work packages (see Lewis et al. 2014; Voorberg et al. 2015; van Acker 2015; Flemig et al. 2015; Nasi et al. 2015) the importance of innovation capacity and its ties to external and internal factors became clear. Furthermore, the importance of leadership skills in motivation, collaboration, and risk taking as part of innovation capacity were highlighted. The studies found that administrative culture tends to be, in general, risk averse, thus risk-taking capabilities inside
the public sector become important. The studies also showed that innovative capacities outside the public sector are an important factor to study. Thus, citizens can partake in co-creating social innovation as co-implementers, co-designers, and co-initiators, and depending on these roles, requirements to their capacities can also differ. Furthermore, the research showed that depending on prior experience, level of development, and already existing capacities to innovate, different internal and external factors (also in accountability, feedback, and learning mechanisms) can become essential for the adoption and diffusion of innovation. Thus, public sector organisations can be differentiated by their status as innovators, followers, late adopters, and non-adopters (see Nasi et al. 2015). Depending on the type of adopter, main drivers and barriers to innovation can differ.

The main findings of LIPSE, thus, assert that public sector innovation is contextually determined, so different public sector organisations need to analyse their internal processes and external environment to encourage innovation with the highest possible impact. Some of the key takeaways from LIPSE research are the following:

- LIPSE research indicates that supporting best practice approaches are rife with concerns (due to different state traditions, administrative cultures) and additional measures may be needed for smooth adoption of social innovation initiatives.
- Formal structures proceed informal innovative processes, thus, differences on the state or organisational level can play a major role in changing how innovation processes look like and which drivers and barriers are more important in directing the process. This means that practitioners should carefully analyse how their formal structures influence public sector innovation and change those structures if they stand between boundary crossing activities needed for innovation.
- Leadership is key in public sector innovation: practitioners need to promote new types of skill sets and characteristics in public service - ‘motivator-risk taker’ – to take the lead in innovation projects. Thus, the change in public sector management is especially important for public sector innovation.
- Flexible risk governance and working accountability systems need to be created at the beginning of the innovation process to assure both the legitimacy and sustainability of the innovation process. While public scrutiny enforces blaming-and-shaming, working risk governance systems should outline the acceptable borders of both risk and uncertainty, so public servants would not feel the need to abstain from innovation projects.
- Involving citizens in achieving better outcomes through social innovation is important, but this can become meritocratic – involving only citizens with high social capital and high levels of innovative capacity. In the process of social innovation, practitioners should be wary about not ignoring more disadvantaged target groups with less innovative capacity.
It is important to evaluate innovation processes in the public sector to ensure that the desired outcomes are achieved. This means that not only the expected outcomes should be examined, but also the unexpected and negative effects of public sector innovation.

Consequently, based on LIPSE research, we can summarise the following discussion points for public sector innovation measurement:

1) Public sector innovations are inherently complex issues to measure as they take place in complex and multiple feedback settings.
2) Key elements in feedback are described by networks and multiple actors involved in public sector innovations.
3) Measurement efforts need to reflect innovation capacities and in particular their dynamic nature: thus measurement should reflect factors influencing innovation capacities (such as those needed to be involved in networks, co-creation practices, and carry out risk governance regimes, auditing, and performance measurement practices).
4) It is, accordingly, important to connect public sector innovation measurement efforts with wider performance measurement attempts, with accountability procedures, and risk governance measures. Public sector innovation indicators will invariably feed into accountability systems and risk governance practices.
5) Rapidly evolving ICT will have a sizable impact on the nature of data available to the public sector, especially on city level (smart city solutions).
6) New organisational forms and experimentation (with co-creation, auditing, procurement) are important elements to take into account in measurement practices.

In summary, research done in LIPSE and other research summarised in this report shows that it is essentially impossible and also useless to develop single figure indicators or sets of indicators to measure public sector innovations (as is the case with private sector indicators, such as scoreboards). Rather, in accordance with feedback received through work package 6 and in particular from Christopher Pollitt (2015), we should seek to develop evaluative frameworks to evaluate various public sector innovation undertakings. Furthermore, work package 3 on the recommendations of ombudsmen and audit offices showed how difficult this can be and highlighted the need for continuous evaluations, because when it comes to failures, public sector organisational memory could not be relied upon.

26 The feedback session took place during the LIPSE midterm conference, February 3-4, 2015 in Brussels. We also discussed our ideas for the evaluative framework during some of the interviews we undertook with the members involved in developing existing public sector innovation indicators. Also two seminars were held to test the indicators (see introduction of this report).
One of the interesting issues that various LIPSE studies pointed towards was the co-existence of two parallel, if not conflicting, goals of the public sector: to maintain the continuous functioning of public services while at the same time innovating. The latter assumes risk-taking, which is impeded by the need not to fail. This is of course similar to what private sector organisations face as well. However, the competitive pressure forces private sector organisations to overcome this dilemma and to take risks, otherwise they will not survive. This leads to failures and indeed the death of some private sector organisations. In short, this is what drives private sector innovation and economic development. In the public sector context, however, survival is mostly related to the ability to continuously provide public services without interruptions, i.e. it is stability that allows an organisation to survive and not the ability to outcompete other public sector organisations. The studies provided good insights into risk awareness and other factors; however, what future studies should explore more in-depth is how exactly government organisations take on risky projects and experiment, how do various governments maintain and develop organisational variety, and what kind of organisational processes evolve out these strategies. Here also the role of evaluation and innovation measurement can play a significant role as these efforts can be used for blaming and shaming and thus, can penalise experimentation in the public sector. Thus, evaluative frameworks need to enable learning from failed innovations as well.
5. Evaluative framework

When we summarise the preceding discussion, we see there are three key dimensions or logics to measure and evaluate public sector innovations: first, the logic of changes; second, the logic of feedback, and third, the logic of technology (see also Kattel et al. 2014). The last is especially illustrated through the discussion relating to the influence of ICT (big data, social media, etc.). We propose to base a new evaluative framework on these three logics.

1) Logic of proposed or implemented changes – what is the intended goal of innovation? Based on our preceding discussions, this dimension has, at its one extreme, efficiency (or productivity) as the main goal where changes deemed as innovations bring more for the same or less money; at the other extreme, legitimacy as the main goals where changes deemed as innovations bring more legitimacy and trust for public sector activities. Ideally, both extremes should to be captured analytically (how to conceptualise efficiency, legitimacy) and thus also measured differently. We can describe a rough taxonomy based on this dimension (Table 6) and what kind of measurement methods could be considered (these methods are not meant as an exhaustive list, but rather as indicative approximations). Between the extremes there are multiple options for innovation goals and respective measurements: for instance, performance management tools such as audits, diverse self-evaluation tools, and satisfactions/trust surveys.

Table 6. Logic of proposed changes, from efficiency to legitimacy

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>Legitimacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input/output productivity</td>
<td>Performance measurement, audits (outcomes, etc.)</td>
</tr>
</tbody>
</table>

Source: Authors.

2) Impacts and feedback loops of proposed or implemented changes – what kind of organisations are mostly influenced by innovations, how does this take place and how (by which organisations, through what interaction channels) is the impact articulated (feedback loop)? This dimension has at its one extreme impact/feedback within (a single) public organisation and at its other extreme impact/feedback within the private sector alone (in terms of productivity increases, new market creation, etc.). Again, we
can argue that both extremes need to be analysed and measured differently and again we can draw up a taxonomy of forms of impact/feedback that needs to be differentiated (Table 7). As with the previous logic, there are options between the two extremes on how the impact and feedback loops take place, for instance between different public sector organisations, through public-private partnerships, or co-creative practices.

Table 7. Feedback loops of public sector innovations

<table>
<thead>
<tr>
<th>In public sector</th>
<th>Through public sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within a public sector organisation</td>
<td>Between public sector organisations</td>
</tr>
</tbody>
</table>

These two dimensions should describe both the entirety of possible changes in the public sector that could be conceivably seen as innovations in the public sector and also describe ways to measure or at least analytically capture such changes in evaluation processes. Table 8 summarises the first two logics combined.

Table 8. Evaluative matrix for public sector innovation

<table>
<thead>
<tr>
<th>Efficiency IN</th>
<th>Productivity</th>
<th>Performance, audits</th>
<th>Self-evaluation</th>
<th>Surveys</th>
<th>Big data</th>
<th>Legitimacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within public organisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between public organisations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public-private partnerships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THROUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors.

The idea of this evaluative framework or matrix is to enable continuous and complex evaluation processes of innovations to be undertaken in the public sector. Such framework also enable the inclusion of more typical public sector performance measurement and auditing practices, but also emerging practices such as big data and social media analysis; it also showcases the role of networks
and co-creation practices. Such evaluative frameworks can also be used together with risk governance activities and capture failed innovations as well as locate more clearly reasons for failures. Such a framework can be used both for internal and external communication and also benchmarking. And, above all, the evaluative framework can be used to better understand innovative capacity. However, this two-dimensional framework does not fully capture the logic of public sector innovation, especially in terms of the role of the underlying technology and its adoption.

3) The role of underlying technology and adoption. Technology and, in particular, ICT seems to have a significant role in how the public sector innovates, from new data gathering methods to co-creative practices. There are large differences in terms of the technology used and how radical or incremental the changes are in reaction to public sector innovations. The public sector can try to lead ICT innovations, adapt existing private sector solutions (or other public sector solutions), or copy (or follow) private sector solutions. Within this continuum (following-adapting-leading ICT innovations), public sector organisations would exhibit highly different capacities and interactions with private and social actors. Thus, public sector organisations can be at the forefront of technological breakthroughs or following others’ examples and implementing incremental changes. For example, work package 5 looked at public sector innovation organisations on a continuum of innovation leaders and followers. Thus, technology (ICT) should be counted as a separate dimension of the public sector innovation evaluative framework: what are the ICT/technological complexities involved in a given innovation?

Consequently, the conceptual evaluation framework can be broadened to a three-dimensional evaluation framework encompassing the logic of impact (efficiency-legitimacy), logic of feedback (in-through), and logic of technology/ICT (leader-follower). See figure 4 below for a visual illustration.
As such, the ideal public sector innovation measurement and evaluation would be based on a three dimensional system encapsulating the feedback mechanism of the system (x), impact measurement (y) and the role of technology (z). This is a holistic approach to public sector innovation evaluation, helping both researchers and practitioners understand public sector innovation in a dynamic manner (both in a temporal and substantive sense), linking the goals of public sector changes to feedback mechanisms (and the administrative capacities this requires) to the effects of the underlying technology. This approach differs considerably from existing public sector innovation indicator efforts (see section 3) and should not be operationalised into single indicators or into sets of indicators as the value of such frameworks is not only to show what an innovation does but also what it does not (e.g. a new database that decreases the time used in applying for building permits can examine all 3 dimensions or logics). As was argued previously, the possible continuums of measurement on these different dimensions – for example, impact measurement (from efficiency oriented goals to social media and real-time surveillance of public opinions) – is so wide that detailing single measures within the framework (with specific x, y, z coordinates) would narrow the relevance of the framework too much. In fact, one can hypothesise that public sector innovation is driven by all three logics, but empirical research has not attempted to capture these dynamics in a comprehensive manner. Thus, at this stage, it is better to actualise the framework in a real-life setting to evaluate the merits of such an approach. With the increasing availability of information and data, the 3D evaluative framework could be developed into a
digital twin of proposed innovation and consequently used as continuous tool to understand what happens with innovations.

In the following section, we will try to actualise the framework we have developed through three case studies of public sector innovation projects relying on e-procurement in the city of Tallinn. First, we will observe which methods are in use to evaluate public sector innovation in the city of Tallinn, before moving onto discussing the developments in our proposed dimensions.
6. Empirical investigation: cases from the city of Tallinn

We will use our evaluative framework developed above in analysing three case studies in the city of Tallinn, where e-procurement was used for new innovative ICT-based solutions. Originally the Description of Work of LIPSE called for a single case study related to e-procurement to test the theoretical and methodological framework for the development of social innovation indicators, however, to show the difference in the application of the evaluation framework we chose to pursue three different cases from the city of Tallinn. As technological development in public sector organisations largely depends on procurement procedures, we first discuss the relevance of procurement procedures to the current study, before moving onto outlining the methodology of the case studies and the descriptions of the technological developments themselves. The empirical investigation is then analysed in light of the proposed public sector innovation measurement framework.

6.1 Technology and public procurement

Being a process that frames both formally as well as informally the ways public and private actors interact, public procurement has a direct bearing on what kind of technology is developed, how it is applied, what interests and aims get involved, and what consequences follow from new technologies. Importantly, public technology procurement can spur innovation both in the public as well as private sector (Lember et al. 2015). There are several categories that influence innovation in cases of public technology procurement.

*The first category* is related to the innovation strategy (aims) of public organisations. More specifically, public organisations can associate public procurement to technology and innovation in three ways (Edquist et al. 2000; Edler and Georghiou 2007; Hommen and Rolfstam 2009). First, the public sector can procure so-called off-the-shelf solutions (i.e. ordinary solutions that neither require nor lead to any innovation). Second, governments can facilitate radical innovation where, as a result of public procurement, private actors deliver new-to-the-world products or solutions. Here the public sector explicitly contracts for non-existing solutions, thus creating incentives for the private sector to engage in not just exploitative, but truly exploratory innovation processes. In doing so, governments need to possess specific market and technological know-how, resources, and competencies to pull off as well as use newly created innovations. Crucially, in addition to just developing innovative services, these services may also change the relationship, accountability and legitimacy structures between
government, market, and citizens (think only about the potential effects of emerging data-driven services on these relationships) (see Jayasuria 2005 for a more general point). Third, governments opt for incremental innovations, where the procured services are only new to the particular service area or user organisation but not to the ‘world’ or broader society. Here the innovation is adaptive or exploitative in its nature. Importantly, the above-mentioned three types of public procurement involve, to varying degrees, inter-organisational collaboration and learning, which is a key success factor of innovation processes (Edquist et al., 2015).

The second category is about technology contracting traditions that influence how ICT projects are developed (small and packaged projects vs. system-level). Here one can distinguish between three state traditions (see e.g. Dunleavy et al. 2006). First, there is the marketization strategy where ICT solutions are, as a rule, always outsourced. The emphasis is on spot contracts and maximum competition rather than long-term partnerships with proven contractors. Second, there are countries that prefer to balance the marketization strategy with a strong in-house ICT capacity, where a considerable amount of ICT services are produced as well as delivered by public sector units. Third, public organisations balance spot contracts with long-term partnerships. The idea is to have trusted private partners as preferred providers that possess in-depth knowledge of public sector needs and specificities. There is also an additional factor, which is the composition of specialised IT market sectors. Markets dominated by a handful of big players may lead to different dynamics compared to markets with many small and medium sized firms actively present.

The third category is related to in-house capacity to procure technological solutions. Foremost, it is about organisational ability to discover, obtain, understand and use new knowledge and/or technologies (i.e. absorptive capacity, see Cohen and Levinthal 1990). This capacity issue also includes ‘The extent to which government agencies retain the capacity to maintain or re-establish their own in-house IT service, and to design, coordinate, and implement substantial IT projects. A key question here is whether an in-house unit can at a minimum still take the systems integrator role or act as the procurement manager in an orthodox contracting set-up for meeting the agency’s IT needs’ (Dunleavy et al 2006). Yet, the technological capacity needs to be accompanied by sufficient contracting capacity in writing, tendering, and monitoring procurement contracts (Brown et al 2006).

The fourth category covers the contracting practices relevant for innovation. The key issue here is how to facilitate interaction and learning with providers before, during, and after contracting. There are several procurement-related activities that can facilitate and support innovation: the use of life-cost assessment (vs. lowest cost), acceptance of variants and usage of functional specification rather than input specifications, risk sharing between public and private partners, allocation of intellectual property rights, use of incentive contracts such as profit-sharing arrangements that emerge from fulfilling the project, advanced communication of future needs that would enable private firms to specialise, early
interaction with contractors to enhance learning, emphasis put on sustainability criteria, and use of innovation requirements in tenders (Uyarra et al 2014; Dunleavy et al 2006).

The fifth, and final, category is closely related to the previous point, that is, the specific procurement procedures conducive to innovation. The usage of competitive dialogue or other procedures enabling dialogue and joint learning is found to be especially useful here (ibid.).

Table 9 summarises the public procurement analytical categories.

Table 9. Factors influencing innovation in public ICT procurement

<table>
<thead>
<tr>
<th>Category</th>
<th>Specific factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Innovation strategy</td>
<td>- Off-the-shelf</td>
</tr>
<tr>
<td></td>
<td>- Incremental</td>
</tr>
<tr>
<td></td>
<td>- Radical</td>
</tr>
<tr>
<td>Contracting tradition/strategy</td>
<td>- Marketization (spot contracts)</td>
</tr>
<tr>
<td></td>
<td>- Corporate (long-term relationships)</td>
</tr>
<tr>
<td></td>
<td>- In-house centred</td>
</tr>
<tr>
<td></td>
<td>- Dominance of SME versus large firms</td>
</tr>
<tr>
<td>In-house capacity</td>
<td>- Technology (absorption) capacity</td>
</tr>
<tr>
<td></td>
<td>- Contracting capacity</td>
</tr>
<tr>
<td>Contracting practices relevant for innovation</td>
<td>- Life-cost assessment (vs. lowest costs)</td>
</tr>
<tr>
<td></td>
<td>- Acceptance of variants and usage of functional specification</td>
</tr>
<tr>
<td></td>
<td>- Risk sharing</td>
</tr>
<tr>
<td></td>
<td>- Provisions related to intellectual property</td>
</tr>
<tr>
<td></td>
<td>- Incentive contracts such as profit-sharing arrangements</td>
</tr>
<tr>
<td></td>
<td>- Advanced communication of future needs</td>
</tr>
<tr>
<td></td>
<td>- Early interaction with contractors</td>
</tr>
<tr>
<td></td>
<td>- Emphasis on sustainability criteria</td>
</tr>
<tr>
<td></td>
<td>- Innovation requirements in tenders</td>
</tr>
<tr>
<td>Procurement procedures</td>
<td>- From open competitive tendering to negotiated tendering and competitive dialogue</td>
</tr>
</tbody>
</table>
In essence, the factors influencing public sector innovative procurement practices also influence the way a given public sector organisation conceptualises technology, innovation, and its potential impact, and how this impact could and should be measured.

6.2 Methodology

In order to analyse the processes connected to the use of innovation indicators in the public sector we used the city of Tallinn as a test case. More specifically, we analysed the use and influencing factors of innovation indicators through the dynamics of the three largest e-service development projects of the city; all cases used public e-procurement as a way to acquire new solutions. The municipal level was chosen for the analysis due to the citizen-centric focus of recent web 2.0 technologies and the fact that prior research has indicated that governments are more responsive to citizens’ needs on the local level (see argumentation in Gil-Garcia and Martinez-Moyano 2007).

The cases were analysed through a participatory action research design to identify how and why the public sector uses innovation indicators, how they are related to public procurement institutions, and how they influence the evolution of administrative capacities. Consequently, as part of the research (in addition to document analysis and over 25 semi-structured interviews with public and private sector stakeholders), we followed the activities of the city of Tallinn between December 2013 and June 2015: participated in their development meetings (among them the e-service working group) and followed the management meetings of the aforementioned and ongoing ICT developments. Most of the interviews were recorded (this depended on the preference of interviewee); for the internal meetings the author relied on written notes.

While the municipality was very open to allowing the research team access to internal meetings and planning processes, the research did meet some set-backs during the investigation. For example, the research team tried to find internal support both within the high administrative and political levels of the city government to develop algorithms to tests social media based real-time surveillance of public opinion on different large scale innovative initiatives. However, while this was greeted with strong interest by public servants, the plan was halted due to political concerns. Hence, we have to note that new methods to study public sector innovation from the perspective of legitimacy are not easy to fulfil.
6.3 The case of the city of Tallinn

Estonia is globally seen as one of the leaders in e-government. The city of Tallinn is at the forefront of implementing electronic services in Estonia. In the last 5-7 years, the city has taken a service-specific focus in developing its ICT capabilities. The city has categorized 581 different services in 20 different policy fields. Close to 200 of the former exist electronically only in a descriptive form (1st level e-services) while 21 are semi-automated and 58 are fully automated e-services. For the 1st level e-services, electronic forms can be downloaded or requested for a service. Tallinn has also created a self-service portal for one-stop access to the offered e-services.

As a rule, the city of Tallinn procures software solutions and tries to license rather than buy to ensure that the IT developer has an incentive to continue to develop the software. Thus, procurement processes play a central role in the ICT systems development in the city. The developments we look at more specifically in what follows – the spatial planning registry, the city’s internal property registry, and the operative information database for closing streets and planning road work – were the city’s biggest development projects of 2014. All of the above also have a geographic component and can be described as geospatial web – GeoWeb – solutions (Cinnamon and Schuurman 2012). All cases can also be seen as key infrastructures for smart city services and data collection. The document trail showed that all of the developments can be traced back to the recommendations of internal audits to increase transparency, user-friendliness, and accountability in the specific fields they were initiated in. The initiatives are briefly described below.

Spatial planning registry

The new spatial planning registry is by far the biggest development of the three cases. It is built on the pre-existing electronic system for planning, building projects, and architectural conditions created in 2005. The prior registry allowed for a semi-automated planning process; alphanumeric and spatial data were not integrated, and it was difficult to change and configure the software to match internal processes and to use the map interface. Nonetheless, civil servants were used to working with the system. It was also very well-known that the system was not user-friendly and very confusing for the average user; nevertheless, as frequent external users of the system – architects, developers, etc. – had learnt to use the system, no concrete strategy to change the registry were planned. In 2011, an internal audit reviewed the system and severely critiqued the lack of speed, control, and transparency of spatial planning processes in the city and recommended that the process should be fully automated and a new registry for it developed. These goals can be connected to both efficiency – speed – and also legitimacy – need for additional control and transparency – in our previously described evaluative framework. In effect, this gave the Urban Planning Department (UPD) the justification to ask for
additional funding to start planning the development. Thus, following the audit, the city changed its building decree in November 2012 and started the procurement process for a 4-step development process of the new registry which included the analysis of the process, legal framework, composition of the initial assignment, and the software development process. Compared to the other two cases, the role of the central IT department of the city was more consultative and the development process was led by the Urban Planning Department, who was, however, not the direct user of the future registry. The registry was supposed to be ready on April 30, 2014, but the delivery of the registry was postponed for more than a year to March 2015. The goal of the development was to make the planning process fully electronic and shorten the processing time of spatial plans – both detailed and general plans – and make information about and access to the process more simple and intuitive by also increasing the user-friendliness of the new interface.

**Property registry**

The development of the property registry started in 2009 and at the time of writing had not yet been finished. Following audit procedures, the central City Property Department was created in 2009 which generated a need to centralize city property information. The city owns more than 17 thousands different objects; data on Tallinn city property has been stored in various datasets in city departments and district offices. These datasets did not follow a uniform structure (the most common form was to collect the data in Excel spreadsheets) nor was it possible to link the data to other registries. The system was, thus, not very transparent and arguably could enable corruptive practices. The property registry was meant to increase internal efficiency and create an overview of the land, real-estate, and other city property (incl. care, renting, and other business processes connected to said property). While the motivation was downplayed during our interviews by the civil servants involved in the projects (due to possible backlash from the involved city departments), the newly formed City Property Department aimed to gain control over the city’s property management through a new technological solution. One can link this to a need to increase internal control as the registry creates the possibility for statistical analysis of the data and a digital audit trail for all the changes connected with city property management. The possibility to interface the registry with other data systems will decrease mistakes via centralization of all data. On the whole, this is an internal tool for central control and management of city property. It is also important to note that such a digital central control and management system should, ideally, also diminish opportunities for corruption and nepotism. As this development touched most of the city’s organisational units, there was a lot of internal uncertainty and resistance to the creation of the program. The initial assignment for the procurement process was set only at the end of 2012 and the development process started in 2013. Compared to the other two cases, the functionality of the future system is much lower: first, due to the long-term process (which has partially made the technological solution subservient to new technological solutions), but also due to the city’s inability to describe new
functionalities and the private developer’s lack of capabilities (e.g. the GeoWeb solution is not connected to the main city maps held by UPD as the IT developer did not have experience with the aforementioned software solutions). Finally, at the end of 2014 the pilot testing program started, which uncovered a multitude of mistakes in the functioning of the software. Currently the development is in its finalisation phase.

Operative information database
There were three main city departments that were involved with the development: Municipal Engineering Department, Transport Department and the Urban Planning Department. The operative information database for closing down streets, excavation permits, and operative information was finished at the end of 2013 and is a fully automated e-service. This has been described by the city government as ‘one of the fastest ICT development projects’ in the local municipality’s history. An evaluation of the previous semi-automated system was held in mid-2012, when the prior contract with the software provider was concluding. In the beginning of 2013, additional sources outside the city government were found and project funding was applied for. The relative speed of the project was explained by the project team as a ‘co-alignment of circumstances’: the IT service of the city found a source for external funding in EU project funding and ‘sold’ the project to the city departments involved; furthermore, a motivated leader for the project was found in the Municipal Engineering Department who carried the project to its finish with close help from the IT services of the city. The initial assignment was compiled in January 2013, the work started in May and by December the database was ready to be tested. The goal was to cut down the time it took to process applications for permits and make the overall process more transparent, simple, and accessible to involved stakeholders. This also meant that information on the closing of streets and on municipal works was to become available online to all citizens, with the possibility for citizens to follow the processes online on a map interface in real-time. The new database has been functional since the beginning of 2014 and considerably cut down the time needed to apply for permits in the connected policy area (from two weeks to two days). As it is mostly used by field specialists as well as various water and electricity works companies, the database was quickly adopted by its users.

Table 10 gives an overview of the outlined cases; we use the evaluative 3D framework developed in the previous section to describe the cases, in particular what kind of evaluative measures could be used or expected to be used.
### Table 10. Overview of cases

<table>
<thead>
<tr>
<th>Logic of impact</th>
<th>Spatial planning registry (not finished)</th>
<th>Property registry (not finished)</th>
<th>Operative information database (finished)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Possible input-output gains (productivity increases measured in time saved in and through the public sector)</td>
<td>No clear productivity gains (possible work load increase on the lower administrative levels)</td>
<td>Clear input-output gains (productivity increases measured in time outside the public sector)</td>
</tr>
<tr>
<td></td>
<td>Possible external legitimacy gains</td>
<td>Self-evaluation/ performance evaluation, audits</td>
<td>External legitimacy gains (both in terms of involved private partners and the public scrutiny of operative works in the city) not evaluated</td>
</tr>
<tr>
<td></td>
<td>Productivity, performance audits, surveys, and social media possible to adopt</td>
<td></td>
<td>Productivity, performance audits, surveys, and social media possible to adopt</td>
</tr>
<tr>
<td>Logic of feedback</td>
<td>Between public sector organisations</td>
<td>Within a public sector organisation</td>
<td>Co-creation practices (tasks externalised)</td>
</tr>
<tr>
<td></td>
<td>Co-creation practices (tasks externalised)</td>
<td></td>
<td>Private sector activities</td>
</tr>
<tr>
<td></td>
<td>Private sector activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logic of technology</td>
<td>Technology allowed to change planning processes thoroughly (not implemented, regulative barriers + internal resistance)</td>
<td>Follower, incremental change</td>
<td>Technology takes a control function and allows for externalising tasks</td>
</tr>
<tr>
<td></td>
<td>Technology takes a control function and allows for externalising tasks</td>
<td></td>
<td>Technological solution simplifies information overview for the general public (interactive GeoWeb solution)</td>
</tr>
<tr>
<td></td>
<td>Technological solution simplifies information overview for the general public (interactive GeoWeb solution)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Possible two-way interaction between the government and citizens</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors.
Indicators used

Although we suggested that there exist ample ways to measure innovation in the public sector context (see Table 8 in section 5), we didn't find evidence of the systemic use of innovation indicators in the city of Tallinn. In fact, most of the potential public sector innovation indicators are never used and the applied ones are used in an unsystematic manner. The following quote describes the former:

‘Normally there are no general indicators connected to procurement tenders. Usually they say that the system or the service has to become “better”. That is not measurable to an engineer. This means that there is no indicator. There should be one dominating indicator for ICT developments that allows for flexibly: the process becomes quicker, more effective, or transparent to the service user…. If that is achieved, then we shouldn’t argue about the details. What in reality happens in scope disputes is that there is no main goal, public servants take the lower level process indicators and start to nit-pick, although the main goal may have been fulfilled long before’ (private IT contractor).

On a general level, the city of Tallinn has not developed indicators for e-services centrally nor are the effects of new policy and administrative initiatives systematically measured. On a departmental level, web-based user satisfaction surveys are frequently used as the main feedback vehicle. Recommendation indices, where citizens are given the option to rank services on a 10-point scale, are also used to measure satisfaction with e-services. The city has mapped and described all of its 581 services and counts the proportion of fully automated (level 5) services as the baseline indicator for further developments. Mostly, the success of various services is measured by the number of visitors on their websites. This can easily be used for a productive measure (input/output), but is not. Additionally, the city uses the number of visits to different self-service environments and the number of free text message service users to report the process in e-service development (this also includes the number of bills paid through the self-service environment). Also, the number of questions, complaints, and proposals connected to different services are counted. The IT-service has shown initiative to evaluate the level of interoperability of different services, however, this is, at most, connected to the possibility of paying for the service through the self-service environment. Hence, indicators currently in use in the city of Tallinn are highly output centred. Yet, these indicators are either too abstract or narrow to inform about the nature and rate of innovation in the public sector and are thus seldom systematically used as an inherent part of management systems. Moreover, as we will demonstrate, much of what drives innovation in the public sector – especially authority and control related issues – fall outside these

---

27 In general, objective, output centred user engagement with public sector information systems is well-covered by technical solutions that can generate statistics on how many unique sessions occurred, how long request proceedings took to clicks on different web pages and menu items (see e.g. Trinidad Consulting 2014).
Public Sector Innovation Indicators

indicators. Also, in the discussions we participated in, the indicators used by the city played almost no role in how new ICT solutions were designed and procured.

When it comes to the specific three cases, the city did aim at increased productivity (mostly in terms of time saved for internal as well as external stakeholders) and general performance (from better functionality to paperless communication) prior to launching the developments of the new service platforms. Yet, achievement of these aims was never formally measured during or after the implementation of the developments. Even so, in all cases at least some productivity and performance increase was achieved.

Similarly, the issues of citizens’ reactions (satisfaction, trust), ease of service use, as well as transparency (internal and external control of processes) figured prominently as goals of the three developments, but were not actually measured. As was put by two of the IT experts involved with the city’s developments:

‘From the get-go they wanted to make the process of approving spatial plans more clear and understandable. However, I am being honest, they only changed the process visually, how it looks to the outside world – they were not willing to change their internal business processes’.

‘There is a fundamental contradiction in the public sector procurement process: [the] public sector is supposed to guarantee the quality of the public service to its citizens, while at the same time being the one that is supposed to provide the service. If the service quality for the citizen comes in conflict with providing the service in a convenient way in the development process, of course the public sector is going to prioritize the latter’.

Importantly, the citizens’ reaction is still taken into account indirectly by making sure that a system was fully functional before going live. Yet, for many interviewees this hinders innovation:

‘The private sector uses minimum product functionality: you finish the most crucial part of the development as soon as possible to go live and then develop the system around it. Of course the private sector gets its money back right away. They really tend to follow the process of going live very precisely and most innovation to the system actually comes from that phase. The public sector has a lot to learn from that’. (IT systems architect)

Discussion on the productivity and efficiency of e-services – return of investment - usually rises during budget discussions when investments on ICT developments have to be approved. While the city is horizontally managed (with different departments and offices having relatively high autonomy from the central Tallinn City Office), financial services are centralized, which gives the financial department the most power in the city to question and direct developments. The central IT department, in comparison, is in a much weaker position because IT investment is part of different departments’ and districts’ own
budgets, making the IT department a consultative rather than development unit. Hence, prior to starting procurement procedures, the city department has to justify spending money on e-services, in doing so the efficiency/productivity gains should also be monetarily evaluated. In reality, most departments fail to provide substantial information as they cannot foretell possible savings or performance gains from ICT developments. In other words, innovation indicators are not effectively used and make almost no impact in guiding or influencing city development processes.

Importantly, for city officials, innovation was mostly associated with internal improvements: how innovation inside the public sector (internal work processes) would lead to higher productivity and enhanced citizen’s satisfaction and trust (although the latter remains vague). To some extent, the city aimed at also making use of co-creation practices. This tendency manifested itself though forcing external stakeholders to participate in service provision. The most significant change we found in the externally oriented operative information database and the urban planning registry was connected to the ‘responsibilization’ of citizens that was enacted through the development process of these new web-systems. First and foremost, in both cases the external users become explicitly responsible for the spatial information they add to the database and the registry. Thus, mistakes made in entry are the fault of users alone. This is a powerful shift in responsibility and, accordingly, in accountability; we can argue that this represents a case of contracting out accountability via technological solutions (if files, data, etc., do not fit predetermined requirements, applicant cannot move on to the next phase within the electronic environment - there is no civil servant, but rather a machine saying no).

Importantly, private providers find it difficult to suggest meaningful indicators to measure innovation in the public sector. On the one hand it was suggested that:

‘I think that the main indicator for e-services should be that there are less people knocking on the door about non-important things; that both the citizens can save time and public servants can actually do their jobs’ (IT systems architect).

Yet, on the other hand, interviewees acknowledged the difficulties related to the public sector:

‘Good service availability is transparent – you cannot measure that’.

Indicators not used – changes in power, control, and accountability

At their core, all three developments are managerial in nature and mostly meant for specialist use; though, the operational information database and the spatial planning registry also introduce functionalities to the general public (e.g. the possibility to follow in real-time road-blockages, road maintenance work in the city, and be warned of maintenance work beforehand; or in the case of the urban planning registry, follow planning procedures in your neighbourhood or the city at large and also
give online feedback). Both the operational information database and the spatial planning registry are working tools, information channels, and archives of processes, and both are important channels for government-private sector interactions. The property registry is the most administrative and internal control oriented in nature. The core task of the new registry is to provide a transparent overview of the municipality’s management of real-estate to the central city office. As mentioned previously, the ability to evaluate renting, sales contracts, and other property oriented information uniformly also deters corruption. As different city departments and district offices have been historically rather independent of the central city government, it is not surprising that the development has been difficult and different offices have tried to postpone the system. The control function of the database is clear and the new system does not offer a lot of value added to specific city departments. Even though the system is meant for internal use, there are almost no feedback functions built into the system. One of the IT personnel of the city described it as a ‘more elaborate Excel table’. The potential to also use the database to increase transparency around the city’s property use was left undeveloped. Because the development has taken a lot of time, it is also partially the case that the functionalities of technologies have grown beyond the initial assessment and need, while it is very difficult to change the official procurement process after it had already started. Thus, the GeoWeb solutions of the development were not interfaced with the official interactive city maps that the Urban Planning Department uses. Consequently, making the city’s property management more transparent has mainly increased the power of the central city office. Although it would have been possible to make most of the registry information accessible to the general public, the city government and the City Property Department see it as an internal tool meant for increasing administrative efficiency and accountability.

The operative information database and the spatial planning registry introduce more complex patterns of relationships. It is important to note here that compared to the spatial planning registry, the operative information database is on a much smaller scale and the processes are, in general, much simpler than processing detailed urban plans. However, in both cases the idea was to control the work-flow electronically and make it possible for different city departments to approve permits or plans parallel to each other. With these kinds of case-management software solutions the capacity for government to delay processes decreases as citizens are able to follow government decision-making and ask for justification for delays (Garcia-Murillo 2013). This makes both areas more transparent and up for public scrutiny, which has also been the case for the implemented operative information system: timelines of roadworks have become much easier for the general public and the media to follow in the GeoWeb application. In addition, in the case of the operative information system, the time saved on evaluating permit applications was rather drastic. It is also important to note that with the new solution responsibilities of various city offices and private companies applying for permits became clearer and thus also easier to control.
In the case of the spatial planning registry, it is difficult to know how the processes will work. In the new system it is also possible for citizens to follow the processes online and see which city department is holding up the process. At the same time, there are opportunities to develop apps for public use that increase transparency on applying for permits, etc., but these options have not been used nor have they played an important role in developing the new registry. As debates on and review of general urban plans can stretch on for years, in the initial assessment phase of the registry development, the specialists involved advised a radical solution: if city departments are unable to approve urban plans in the set timeframe, they will be automatically approved. This did not reach the final phases of the development. As the development started with the analysis of the process itself, there was also a possibility to redesign the urban planning process more thoroughly. While changes were in the end made (e.g. initial planning procedures were simplified for the users), the civil servants in the city opposed more drastic changes to the process itself. There is still a high level of uncertainty inside the city connected with the new registry as various city departments have to use it during planning processes. Thus, as the spatial planning registry is by its nature a much more complex system than the operative information system, here we can see how administrative power issues intertwined strongly with technological developments.

During the process of development, public servants in different city departments were well aware of the control function these new web-based solutions created and those involved with the development process tried to minimize the pressure landing on specific public officials. For example, in the more complex urban planning registry, specific information of who specifically is looking over spatial plans from a specific department is not open to public view. This information is, of course, available in the system itself for administrative personnel as tasks are assigned and completed within the registry itself. Consequently, the statistical information that is given to the general user is less specific than is available for the municipal government itself.

The most significant change we found in the externally oriented operative information database and the urban planning registry was connected to the ‘responsiblezation’ of citizens that was enacted through the development process of these new web-systems. First and foremost, in both cases the external users become explicitly responsible for the spatial information they add to the database and the registry. Thus, mistakes made in entry are the faults of users alone and these mistakes are machine-controlled. This is a powerful shift in responsibility and, accordingly, in accountability; we can argue that this represents a case of contracting out accountability via technological solutions (if files, data, etc., does not fit, applicant cannot move on to the next phase). In the case of the operative information database, the exactness of data entry (e.g. drawing on the map the extent of the road blockage needed for specific works) will also determine the fees imposed for the service. Thus, the service becomes, to a degree, also dependent on the skill level of the user. This also applies to the urban planning registry
where personalised accounts and digital signatures are imposed to increase personal responsibility. Thus, each user gets a digital work table in the registry and, depending on the role (UPD’s worker, City Office specialist, external stakeholder, and general user), also access to various information and tasks. Furthermore, in the more complex urban planning registry, the goal of the Urban Planning Department with the digitalization process was also to make the developers more responsible for getting agreements from different city departments and citizens from the specific neighbourhoods prior to different steps in the registry workflow. The registry also provides the opportunity to give direct tasks to developers connected with mistakes made in incorrect data import.

A case can also be made here for the digital/democratic divide. In the operative information database, the move to an electronic service was very smooth and big companies were able to almost instantaneously start applying for permits online and use the GeoWeb interface to map out works. While the main users of both the operative information database and the urban planning registry are specialists in nature, the urban planning process is much more conflictual and also political in nature, thus, public interest of these processes is much higher. However, the urban planning process is considered complicated and already overly technical by average users and neighbourhood associations who were interviewed as part of this study. In the new registry the process is online and while there are public debates held in specific urban plans, the opinions and specific data are only accessible online. In the urban planning registry case it is also clear that the local government prefers to primarily use electronic channels for the process. With some notification tasks compulsory by law in the urban planning process, the municipality has built an interface with the official state government e-service portal (eesti.ee) that maintains official e-mail addresses for citizens. If this cannot be used, then paper-based notifications are seen as a last resort.

While in none of the cases can we see effective two way interactions, the urban planning registry creates opportunities for identified citizens to give opinions and express views on different detailed urban plans. The city municipality can also answer citizens through the system. As the new system has not been in use for a long time, it is difficult to foresee how much these channels are actually going to be used and if this will speed up communication between government and citizens.

What is perhaps most noticeable is almost a complete lack of discussion around how to use data that is created in the new solutions for evaluation purposes or how to create some social features (feedback, discussion forums) to these databases.

It is also noticeable that in none of the cases city officials differentiated between evaluating impact within the public sector and through the public sector.
Public procurement and innovation indicators

We expected public procurement to play a significant role in shaping public sector innovation processes as well the use of indicators. Several observations can be made in that regard.

First, the city of Tallinn has no explicit strategy for procuring innovations per se. Although ICT platforms are to an extent always innovative – that is, these are usually tailor-made solutions – the ways the city carries out public procurements assumes, according to contractors, routine work from private developers. This was echoed by a city official:

‘How to procure innovation with public procurement rules? Even after 25 years of experience, I don’t know how to do that’ (public sector IT manager).

Second, the city’s public procurement is heavily based on a spot contracting strategy, meaning that the city contracts out single, packaged tasks rather than relying on internal development capabilities or long-term partnerships. Also, usually no system-level contracting takes place.

‘Long-term partnerships are more effective. When you need to think about the whole life cycle costs of the ICT system – for example 10 years – then you start to think about what you initially invest in the development. Also these things wouldn’t happen (authors: as in spatial planning registry) that you have some analytics who have done the previous system engaged with another project, because it is more profitable to the firm’ (IT developer).

‘We try to act as partners to the public sector and finish the spatial planning registry. But will it help us in the next procurement? No’ (IT developer).

Still, many of the contracts are won by companies with proven track records.

‘ICT companies do a better job if you value their work and also hype their developments – you did a good thing, we go to conferences, present it, give you some free publicity’ (public sector IT manager).

Third, although the city has a dedicated IT department whose responsibility is to assist city structures with IT projects, these technological capabilities play no central role in city administration decision-making structures. The IT department has no direct power over technology development in the city nor does the IT department or any other technology unit figure highly in the administrative structure.

‘It was a bit frustrating to deal with the city. I don’t know if it was the project team or it is how they do things in the public sector, but it seemed that the operations manager didn’t make any decisions. Nobody wanted to take responsibility. So, everything had to be taken to the higher-
ups, so, while we had already move on with the development, it wasn't uncommon that the project team came back to us and said: no, actually we cannot do it this way' (IT developer).

'I regret that we didn't hold our ground and draw the new process as it should have been and stuck with the reality. This probably cements the processes even further in the organisation. I have learnt from that for the future' (IT systems architect).

'It is very difficult to automatize processes. While technically you can close a process before you start another process, but in practise it is not so easy. The city is centrally managed and the City Council says what you can or cannot do’ (public IT engineer).

The city of Tallinn's capacity to understand the technological trends and emerging possibilities as well as its capacity to absorb new emerging technological solutions is probably among the best in Estonian local government, but as the city itself does not invest in long-term exploratory development projects, its absorptive capabilities can be regarded as somewhat limited.

'The honest answer is that neither we nor the city of Tallinn understood how complicated the development was. There were professionals working on both sides (we have a lot of technical capability and they know the process), but we couldn't foresee all the interlinkages' (private sector IT developer).

At the same time, the overall legal as well as procurement context seems not to have made this particular task easier:

'There have been occasions that in a friendly collaboration with the IT-developer we find out that we could do things differently or we cannot do something at all (for example if we are dependent on another public sector organisation and they don’t fulfil their part). We should do things differently, but we cannot. The public procurement unit tells us that we don’t have grounds to change the procurement contract mid-process, we would be breaking the law’ (public IT engineer).

'It is difficult to draw new IT systems and their different outlooks if you don’t know for sure what the legal system is going to look like. The public sector is still in the process of changing laws while we have to prototype new solutions and fulfil our procurement contract on time. In the case of the Spatial Planning Registry, we didn't know if the state was going to take over part of the building planning process or when they were planning to do that. In the end, we had to go with the solution and the model that described the then-current system' (IT systems architect).

The city itself regards its contracting capabilities to be on a very good level. This assessment is further supported by the fact that, on average, the city has a very small number of procurements with
challenges. Yet, this does not resonate directly in public opinions about the city. Also, private providers
do not necessarily share this perspective.

‘The public sector is not a good procurer from the perspective of Auntie Maali (authors: ordinary
citizen)’ (city official).

‘In these procurement documents almost everything is described, as if you are solving all the
world’s problems. The client should know what is important, what is the main functionality.
When resources are limited then you have to know what to let go. However, in the public sector
the tendency is to do everything at least somehow – that is the worst. It is pointless’ (IT
developer).

‘It seems that in the public sector they want to keep the deadlines to the last second. The time
frames in the procurement process were absolutely unreal. It seemed that if we analysed
something and came up with new ideas that seemed to be better, we were so busy that we
couldn’t develop them and had to move on. Then everything was left as it was already in the
initial project documents’ (IT systems architect).

Fourth, the city of Tallinn makes limited use of innovation enabling contracting practices, such as the
use of life-cost assessment, acceptance of variants and usage of functional specification, risk sharing
between public and private partners, effective allocation of intellectual property rights, use of incentive
contracts such as profit-sharing arrangements, or advanced communication of future needs. In a similar
vein, the city seldom makes use of competitive dialogue and similar procedures that would enable more
interaction and learning prior as well as during public procurement. On the one hand, the city
encourages functional rather than input-based thinking:

‘My ideology is to propose tasks, not solutions, in procurement tenders. It gives some room to
think for the developer. There is nothing I dislike more than an IT-developer who comes to me
and says that this thing wasn’t in the procurement document’ (city official).

Yet, this is limited due to time and cost constraints:

‘The patterns in the public sector are very similar: contract conditions are concrete, funds for
additional activities are low and this does not bode well for managing projects’ scope flexibility.
Not in terms of time, money or tasks. The only thing that the public sector is slightly flexible on is
time, but for a developer this means working hours – that is money’ (IT developer).

‘In a system where the cheapest offer rules, it is difficult to develop IT systems. The reality is
that IT systems are so interlinked and should be interoperable that it is difficult to do just the one
single part that was ordered. Another city department sees the development and finds the
results, the created data interesting for them and asks to link it to their databases. This wasn’t in the official offer, but then we are told that we “promised to make the system whole” (private sector IT developer).

For city officials as well as for some private contractors this is a direct consequence of the public procurement law.

‘People are generally nice and hard-working in the public sector, thus, it is not public servants personally that don’t allow for innovation in public procurement, it is the structure in which public procurements are organized’ (private sector IT developer).

‘I don’t think that agile development is possible in the public sector in the near future, but the meantime solution might be to divide the process into different parts. First to procure the pre-analysis that ends with a system prototype and then go into the main development process. Then it is more clear for the developer as well… if they want a castle or a small hut… and the procurement offers will be more realistic as well’ (IT systems architect).

In terms of indicators and evaluation frameworks, it is noticeable how fundamental the impact of procurement capacities is on what kind of indicators are used in new technological solutions. In essence, existing procurement rules significantly narrow the choice of indicators used.

6.4 Discussion: ‘Future ain’t what it used to be’

From the previous discussion, we draw the following preliminary conclusions:

- New technologies in the public sector can increase productivity and performance, but also affect organisational change and legitimacy and power relationships with the public.

- Importantly, all cases indicated that productivity increases were mostly associated with saved time (internally as well as for service users), whereas all stakeholders stressed the importance of transparency and citizens’ re-defined roles in service provision.

- Using single innovation indicators provides only a limited opportunity to understand and measure the extent and depth of public sector change. The Tallinn cases demonstrate that evaluative frameworks that go beyond simple indicators can reveal much wider change in the dynamics at play compared to what single indicators can disclose. Importantly, the problematic nature of innovation measurement – and thus the need for better metrics – was acknowledged by all stakeholders.
The Tallinn cases exposed the presence of the classic innovation measurement problem: ICT brings about numerous ways to track the changes taking place in public sector innovation projects, yet these tend to be of limited use when one is to find out the wider effects of innovation and change in the public sector. The usual Web 2.0 related indicators, such as usage statistics, tell us very little about organisational productivity dynamics or change in power and control relationships.

Analysing changes in control, power, and legitimacy relationships in the case studies reveals that more complex evaluative framework for public sector innovation measurement provides valuable insights into public sector change. Here the internal change (or resistance to change) of work processes and administrative power dynamics play crucial roles in how technologies get developed and adopted.

The public sector tends to take into account various logics of change, but this is done implicitly through internal communication and interactions, rather than explicitly through clearly identified indicators or frameworks. There is an inherent problem for public sector stakeholders to describe or quantify expected productivity improvements and, even more so, expected changes in authority and legitimacy. If in conflict, internal productivity and control of information prevail against external legitimacy concerns (e.g. ease of use and transparency). This is reinforced by the fact that linking legitimacy metrics to innovations is very difficult and hardly ever done (i.e. no real-time measurement of citizens’ reactions, etc.), although technologically possible. Real-time interaction with citizens and real-time evaluations can also mean that policies and actions within the public sector are sped up.

There are different kinds of feedback loops in operation, yet the innovation processes are most strongly influenced by the expectations associated with processes within the public sector. Even if new technologies are created and procured in cooperation with the private sector, the potential positive effects of innovation through the public sector are not directly taken into account. That means that the potential positive effects arising from public procurement of technology in terms of new private sector capabilities does not play a significant role as innovation strategies mostly aim at off-the-shelf rather than radical innovations.

Yet, the very context of public-private partnership (here manifested through public procurement of technology) heavily influenced the innovation feedback processes and thus the extent to which new technologies changed public service provision. Public procurement strategy and in-house capacity to engage with private providers as well as contracting practices and procurement procedures all significantly influenced the decision-making process and ultimately the effectiveness of technology development. Most importantly, it is challenging for the public
sector to institutionalise innovation-enabling interaction and a learning environment within the existing procurement institutions.

- More interaction-enabling public procurement frameworks are essential to removing some of the main barriers in innovation and technology developments. This can be achieved, for example, through more often using negotiated procedures or communicating technology needs early. Importantly, it is not just what the law is, but also how the law gets interpreted in certain contexts. Therefore, investments into procurement capabilities – including into legitimisation of risk-taking – constitute an important avenue for changing the public sector innovation feedback mechanisms.

To summarise: perhaps the most startling conclusion we can draw from the Tallinn case studies is that new technological solutions change the perception of time, as baseball legend Yogi Berra put it, the ‘Future ain’t what it used to be’. Technological advances and innovations lead to parallel temporalities in evaluating the public sector:

First, the shortening of time horizons in public sector innovation where the logic of change (efficiency gains) is strong and easy to measure (both within and through the public sector; as in the case of the operative information system in which the average time to obtain permits fell from 2 weeks to 2 days), where user skills match new technological solutions (as in the case of the spatial planning database in which case users have the sole responsibility for data input and completeness is judged by algorithms), and where existing technological opportunities are easy to absorb (low level of uncertainties, as was the case with the operative information system).

Second, the prolongation of time horizons in areas determined predominantly by trust and legitimacy concerns, where user feedback is driven by surveys and similar highly roundabout tools, which virtually secures that the input is not taken into account in further developing the technological tools (none of our cases considered using social media tools to track the use of, problems, and satisfactions with new tools, either internally or externally; the most complex tools considered were simple feedback forms to report problems or ‘thumbs-up’ signs). The main driving factor in these cases remains the issue of administrative power – whether technology undermines or enables it. In other words, the highly abstract logic of change, indirect feedback loops and technological uncertainties related to its potential impact on service delivery as well as administrative power can create contested public sector innovation environments that assume the use of interactive evaluation tools or the application of digital twins in order to make sense of innovation impacts.

Such impacts of increasingly digitized service design and evaluation are in fact similar to what private companies experience when using big data in marketing their products: because of big data analytics, some companies know almost too well what their customers want and end up undercutting their own
long-term brand-building efforts (as customers start to associate them with quick, cheap offers) (see Horst and Duboff 2015). In the public sector case, we can argue that something similar happens when new ICT infrastructure is being built: quick efficiency gains and easy to use control mechanisms are set up (although in Tallinn’s case poorly measured and evaluated), but long-term ‘brand-building’ (what should the city stand for in the eyes of citizens, companies) is almost not considered at all in terms of how to build new technological tools and, hence, there are no advances in how to measure and evaluate their wider impact. As a result, on the axis of impact – from efficiency to legitimacy – public sector organisations are ignoring most of the scale of possible evaluations. Whether new technological solutions increase trust in city services and the legitimacy of a city’s activities, is essentially ‘measured’ through local elections. However, it is very difficult to attribute these effects to technological projects or even larger municipal policies, due to compounding issues and a possible ‘negativity’ bias of the local electorate (see e.g. Boyne et al. 2010). Especially in the highly polarized environment of Tallinn local politics (where the political divide tends to be starkly along ethnic lines: the ruling Centre Party is centre-left party with close to universal support among the local Russian-speaking minority), this type of feedback has almost no bearing on city services and no bearing on technological change within the city hall. Thus, the direction of innovation is mostly determined by ICT capacities of city officials, administrative power concerns and financial constraints in procuring new solutions. Even productivity gains are not systematically evaluated or measured.

We can thus argue that in procuring new technological solutions, the Tallinn city government relied strongly on its existing capacities to administer such procurements and, as a result, almost no new capacities and capabilities were developed. In this sense, we can argue that ICT solutions re-enforced existing capacities and problems with these capacities. At the same time, we can see trends towards the automatization of service provision in which control and responsibilities are being re-balanced (towards central city departments and highly skilled private users) with enhancing efficiency for some of the partners. Without developing data-driven and dynamic user interfaces, key public values (such as trust) will remain only vaguely captured in the design and evaluation of new services. However, these new interfaces require quite new and different capacities both from the individual bureaucrats and institutions involved (in terms of giving larger access to users in design and evaluation).

In summary, we can argue that in the case of Tallinn, the evaluation frameworks used are relatively narrow and often determined by limited public procurement frameworks. One of the key recommendations from our research is that organisations should vocalize and formalise their innovation and procurement strategies in evaluative frameworks before they set out to procure new technological solutions. The framework we have developed showed in the case of Tallinn were the weaknesses of current attempts at procuring new technological solutions are and how these could be fixed with new
organisational routines, better feedback mechanisms, the sophisticated use of data, and the creation of new data sources (such as social media based databases).

In terms of the 3D evaluative framework developed within our project, we draw the following conclusions:

Strengths of the 3D framework:

- The use of evaluative frameworks enables wider ranging discussions throughout innovative processes (from preparation to ex-post evaluations) and also allows for more precisely defining innovations and their importance (across different dynamic dimensions).
- These discussions can be used as useful tools for both internal and external communication.
- Evaluative frameworks offer quick ways to bring out alternative measurement and evaluation tactics and techniques.
- The framework enables the involvement of performance and risk management tools in innovative processes within public sector organisations.
- The framework puts a lot of emphasis on processes outside public organisations (such as technology and trust) and allows for describing the public sector as an enabler of innovation.

Weaknesses of the 3D framework:

- The use of such complex evaluative frameworks assumes strong buy-in from key people involved in projects/processes.
- Equally important, the use of the 3D framework assumes that most of the stakeholders have a relatively similar understanding of the key issues at stake when embarking on new innovative processes or products in the public sector; or at least a version of results that can be defined through the three dynamic dimensions we identified.
- The framework relies on data generated in other contexts (such as trust surveys or performance measurement tools) that often do not (yet) take into account innovations in the public sector.
- Real-time data analysis presents public organisations with the challenge to develop entirely new administrative and policy capacities, which may be out of reach for many organisations. However, one can hypothesise that in the longer term the development of these capabilities becomes inevitable.
- The 3D evaluative framework does not yield results that can be easily compared across organisations or countries. Thus, it cannot be used as a country ranking tool nor for blaming and shaming purposes.
Future development of the 3D evaluative framework should involve further testing in diverse settings and also further attempts at integrating various measurement tools into more comprehensive frameworks.
7. Experimental spaces for public sector innovation: new avenues of measurement

One of the key developments in public sector innovation is the emergence of organisations specifically designed to experiment with various public sector innovation elements. In our view, such organisations are profoundly changing the way we understand public sector innovations. Innovation labs (and design labs and other similar organisations in the public sector on both the central and local level) use user-centric design approaches to rapidly create new prototypes of potential solutions, test and evaluate them, and eventually help to scale them up into full services or products. Thus, they use profoundly different design and evaluation methods than typical public organisations. As such, we undertook to create an overview of these new emerging organisations and the way they evaluate public sector innovation. In theory, these innovation labs should also be technologically at the forefront and leading public sector innovation, and thus allow us to look more specifically into what impact and feedback mechanisms these units use. Here we concentrate on the nature, working techniques, and evaluation methods of such units – a more comprehensive review of the findings can be found in Tõnurist, Kattel and Lember (2015b).

7.1 Introduction: surge of experimental spaces in the public sector

In 2013, Parsons DESIS lab (the New School for Design) published the ‘Government Innovation Labs Constellation 1.0’ covering 16 such innovation outfits. Subsequently, Nesta and Bloomberg Philanthropies published a report on public sector innovation labs that covered 20 such units around the world (Puttick et al. 2014). While these reports have been informative in nature, there is very little research on public sector innovation labs beyond descriptive – and at times normative – overviews. Mostly i-labs are described as versions of various existing organisations: as hybrids of think tanks, digital R&D labs, social enterprises, and charitable organisations (Williamson 2014). The nature, organisational structure, and need for such units within the public sector is largely unexamined. In this section, we will try to take a first systematic step to fill this gap in academic literature.

There are some relatively well-known and obvious wider social and technological processes that have led to the surge of i-labs in the public sector:

First, recent public sector reforms, and perhaps even more so academic discussions of such reforms (new public governance), increasingly focus on service processes and outcomes (Osborne and Brown 2005). This has opened the public sector to an influx of various process-oriented methods, such as design thinking, that focus on the user experience, frequent experimentation, and multiple analytical perspectives (Bason 2010). This is epitomized by such organisations as MindLab in Denmark.
Second, many governments operate under some form of fiscal austerity constraints that reinforce focus on public service processes in order to increase productivity and citizen trust simultaneously (e.g. through bringing in user experience via co-production methods of participation). On the flip side of fiscal constraints arguments, are discussions about the entrepreneurial state that focus on the public sector’s role in engendering new generations of technological revolutions (Mazzucato 2013). Here the focus shifts towards different kinds of policy experimentations, for example, innovative public procurements and other so-called demand-side innovation policy measures that seek to create new markets (Lember et al. 2014).

Third, perhaps most importantly, governments are under siege from the adoption and diffusion of ICT and the possibilities this has created, from participatory feedback mechanisms to utilising web analytics and big data.

Consequently, the emergence of i-labs in the public sector can be tied to larger theoretical debates about organisational and technological (radical) innovations. First, Weber argued that new organisational forms (or change from one form to another) emerge through conflicts between old and new leaders and staff (Weber 2009, 154-155). Particularly interesting for our context is Weber’s charismatic organisation, as succinctly summarised by Samier (2005, 71): ‘A charismatic organization is consistent with its own principles, that is a new organization with its own language, mores, myths, and roles derived from the personality and belief system of the charismatic founder, affecting staffing, working patterns, social behaviour, and the material environment’. Second, such new charismatic organisations can be linked to the effects of disruptive or revolutionary innovation (see Yu and Hang 2010). The disruptive innovation theory popularised by Clayton Christensen (1997) in *The Innovator’s Dilemma* is worth mentioning in this context. Simply put, disruptive technologies are technologies that are significantly different from traditional technologies, and while they can be initially inferior in the performance criteria mostly valued by mainstream customers, disruptive technologies are able to create new markets and provide new functionalities. These technologies disrupt the existing market structure, linkages and, on the organisational level, destroy competences (ibid.). Established companies tend to excel in incremental innovation, but wane when confronted with breakthrough innovations (Tushman and Anderson 1986). Smaller companies are presumed to be better suited for technological breakthrough (Baumol 2004). While organisational death can be a legitimate part of the evolutionary perspective of the economy, it has lead management researchers to ask how incumbent firms can deal with these organisational challenges and which solutions there could be for the ‘innovator’s dilemma’.

As such, i-labs as an attempt to structure (radical) change processes within public organisations is not entirely a new phenomenon (see e.g. Thompson and Riccucci (1998) on the U.S. reinvention labs in the 1990s). However, what is different in the current wave of i-labs is the context and logic why these structures have emerged, that is, the combination of user-driven service production logic, ever-
increasing computing power, and fiscal austerity. The existing literature on i-labs shows that the surge of innovation labs also in the public sector can be tied to (lead) user-centred approaches (e.g. co-creation, co-design, and co-production)(Bason 2013; Mulgan 2014). Thus, the spread of user driven innovations and lead users (von Hippel 1986; 2005) has been a major contributor to the spread of innovation labs. Furthermore, these models have been powered by the popularity of ‘open’ innovation models during the previous decade (Chesbrough 2003). Thus, the key has been to tap into ‘collaborative’ innovation. What is different in the public sector context is that direct feedback mechanisms from the market – which forces companies to change and acquire new capabilities to keep up or perish – do not exist. Thus, the public sector has the luxury to be more ‘self-absorbed’, meaning that the discussion surrounding public sector innovation has jolted between internal performance and efficiency gains and external legitimacy, not so much public value based concerns (Kattel et al. 2014; Tõnurist 2015).

7.2 Methodology

On the whole, innovation labs both in the private and public sector are very heterogeneous – in terms of their activities, scale, and organisational structures – making them difficult to map and analyse. Thus, we decided to base this research on a two-step approach: first, a comprehensive survey was carried out directed at the management, followed by an extensive in-depth interview with the same managing figures of i-lab. The survey is based on long-term and large scale research into public sector organisations in Europe – the COBRA research project. Based on a proven structure and logic, the COBRA questionnaire addresses the autonomy of agencies towards their political and administrative principals on different dimensions. The questionnaire also helps to shed light on the way agencies are controlled by their principals and what kind of internal management tools the agencies use. This helps us to compare i-labs to other (semi-)autonomous public sector organisations. However, due to the specific nature of i-labs, the questionnaire had to be significantly updated to fit the purposes of our research. The survey was followed by an in-depth interview in which we more specifically covered the reasons behind the creation of the lab, team characteristics and main tools, network partners, activities and goals, outcomes, and steering and control. The research design was tested prior to use with a representative from MindLab, Denmark.

Based on prior reports by Nesta, IBM (Puttick et al. 2014; Burstein and Black 2014), Parsons ‘Gov Innovation Labs Constellation 1.0’, and web-based searches, we identified 35 i-labs in or directly funded by the public sector. Most of them could be found in Europe and North America, although, Asia is also showing a growing number of such labs. In developing countries, these labs (primarily social innovation labs) are usually found in the third sector and thus outside of the scope of our research.

29 We are grateful to Prof. Koen Verhoest for his help in updating the questionnaire.
Furthermore, i-labs established under the United Nations (incl. the UNDP Public Service Innovation Lab) were not considered for this research. Prior to the survey and interviews, we made a profile for all i-labs in our sample based on document analysis. From the aforementioned 35 i-labs, we were able to find a direct contact for 25 labs, from which 16 answered our initial interview request. In the end, 11 i-labs joined the full study (3 of these had closed down by the time of our in-depth study; see the list of interviews in the Appendix). Our study includes six i-labs from Europe, four from North America and one from Australia.

### 7.3 Innovation labs in the public sector

#### General characteristics

Innovation labs in our sample were established between 1999 and 2013, seven of the i-labs were established after 2010. From the total sample (35) of i-labs, around one third were established on the municipal level (if we would have considered innovation officers in U.S. cities then this proportion would have been much higher), while others were created on the state or federal level. Approximately half of i-labs in our sample had their own legal personality separate from their parent organisation (both vested in public and private law). While others were identified as independent parts of a ministry or municipal department (e.g. DesignGov, Laboratorio para la Ciudad) or did not exist in the formal organisation at all, which was the case for one of the most well-known i-labs – Helsinki Design Lab – in SITRA.

I-labs in our sample employed from 2 to 17 people, with an average team size of 6-7 persons. While it was difficult for some i-labs to differentiate their own budget from the overall budget of the parent organisation, the maximum budget in our sample was €1.5 million in their previous financial year. For over 60% of the teams the primary source of income was self-generated, that is project-based funding (for more than half of the labs under review this constituted more than half of their budget – see figure 5), though closely followed by direct budgetary transfers from the government. In many cases, internal funding (in addition to operating costs) comes from specific project or program partners within the public sector. This structure of finances was seen as important by i-lab executives: while internal funding encourages ownership of projects inside the public sector, external funding gives i-labs the flexibility to try new things.
Figure 5. Source of income providing more than half of the total budget (% of i-labs)

Source: Authors.

This also gives an indication of the power and control relations that separate these teams from the rest of the public sector – self-generated income and low operating budgets mean that most i-labs do not illicit strenuous performance evaluations nor the need to collect quantitative metrics to make the output of the labs measurable. This makes most i-labs small and agile, and is seen by most interviewed managers as necessary: the lean, start-up type structure enables much quicker communication, forces lab to do things ‘quick and dirty’ (e.g. some work with the time-frame of 6-weeks-discovery with the new design for a new service solution ready in 12 weeks), because there are not enough people nor budget to draw out the process. When projects become internally too big, then invariably i-labs run against existing structures (e.g. IT departments and ICT architecture) and procurement rules. This was seen as cause for loss of momentum as ‘existing standards override everything’. Thus, innovation labs primarily engage in rapid prototyping and were less interested in long-term engagement, although scalability is one of the most stressed aspects in new social innovation solutions (see also Kieboom 2014). Thus, in this sense simple solutionism (rapid prototyping, quick and dirty approaches) takes hold while complex system dynamics can be underestimated; especially in terms of social innovation some consider more long-term engagement important to have a real impact (Mulgan 2009).

Usually i-labs are built around a particular user-design led method such as human-centred design (MindLab), the ‘Friendly Hacker’ method (La 27e Région), or four-step Innovation Delivery model (New Orleans Innovation Delivery Team)(for more see Puttick et al. 2014). However, when it comes to specific analysis techniques and skills, i-labs use a variety of approaches: randomised control trials,
ethnography, or action research to work directly with the people impacted by social problems and use a variety of techniques (story-telling and story-boarding; character profiles; service journeys; experience maps; actors map; interactive tables and whiteboards; headlines and postcards from the future; and foresight) to increase empathy with the target group and social issue (see also Bellefontaine 2012; Puttick et al. 2014). This also means that labs have to incorporate a variety of skills to accommodate these approaches. Innovation labs usually bring together heterogeneous teams of researchers, designers, and stakeholders to discover and analyse problems from different angles and develop, test, and improve prototypes for their practical application. Our interviews showed that i-labs employed both people from backgrounds generally new to the public sector – design, anthropology, ethnography, social geography – and more traditional skill sets – political science, sociology, communication, etc. What was striking was that while these labs are often associated with new ICT solutions and hackathons, there were not many IT engineers present in the labs that were in our sample – these skills were acquired from outside partners. In some sense, this can be seen as an attempt to make technology subservient to social change rather than letting the technology be the catalyst of the latter (see also Townsend 2013 on this point). At the same time, a cause for this was also the fact that during the period of austerity i-labs were not allowed to hire outside of the public sector.

Overall, i-labs were created to enable cross-disciplinary and citizen-driven approaches. At the same time they produce most of their work for or with the ministerial departments and other government agencies. The level of collaboration is high due to the user-centred approaches that i-labs employ both in and outside of the public sector. Thus, collaboration – both inside and outside the public sector – and the ability to coordinate interdisciplinary user-needs across different partners is key for i-labs. As such, it is not surprising that the self-reported characteristics of i-labs are a concentration of activities on building trust, individual, relational aspects, cooperation, and empathy – see figure 6.
One of the most important aspects of i-labs is their level of autonomy, which should allow the units to pursue discontinuous and disruptive innovations without direct interference from traditional organisational structures. The survey outlined that most i-labs in our sample were indeed characterised by high levels of autonomy, with most units taking most of the decisions themselves, with the minister/parent department only slightly involved. Half of the surveyed i-labs considered their autonomy sufficient to a degree, others found it totally sufficient or hard to evaluate. When we look specifically at different factors of autonomy, we can see that two thirds of the labs had control over setting salaries as well as appointing and evaluating most of their staff. Furthermore, most organisations set their goals themselves, with only a third of i-labs having to consult their parent organisation. The same holds true for negotiations with external actors, also those from the international level.

Source: Authors.

Evaluation as a gateway to autonomy
Usually formal indicators are used to communicate results monthly or quarterly between the i-lab and parent organisation, while more ethnographic methods (description of activities, video diaries, blogging, etc.) are used internally and to communicate results to the wider network of lab partners. The latter is more to legitimate lab activities in the eyes of the general public. Goal attainment is usually evaluated inside the organisation itself and there are no direct performance rewards for results (apart from the possible increase in budget in few of the reviewed cases).

However, in terms of finances, outside control over i-labs is more rigorous: meaning that investment and annual budgets have to be coordinated with the parent organisation or the financial department. Most are subject to external audits concentrating on financial issues, legality, and rule-compliance. One third of i-labs found that their use of resources is evaluated to a very great extent and another third to some extent. Nevertheless, measurable targets are usually not tied to budget allocation and targets are in most cases set only for internal use. Used indicators usually describe activities and task performance, measuring quality of services, and are both qualitative and quantitative in nature. However, most labs did not use any traditional quality measurement systems and concentrate on public reporting of outcomes - see figure 7.

**Figure 7. Traditional evaluation modes**

```
<table>
<thead>
<tr>
<th>Evaluation Mode</th>
<th>Not at all</th>
<th>To little extent</th>
<th>Hard to evaluate</th>
<th>To some extent</th>
<th>To large extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>User or customer-panels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal units monitoring quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality management systems (i.e. ISO, BSC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer surveys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality standards for production/service delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public reporting on the performance of the organization in yearly reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Source: Authors.

However, when the budgets get higher and the activities become more visible, calls for more precise control and ‘meaningful’ performance (including efficiency) measures emerge (see e.g. the case of the OPM lab in GAO (2014)). This is also the reason why, as mentioned previously, most executives saw the small size of i-labs as key to the success of their activities, otherwise the centre of control would go elsewhere and the steering of the i-lab would become more standardized, invariably influencing the core activities of the lab itself.
Role of i-labs in the public sector

As argued previously, the primary tasks of the organisations were service centred (developing prototypes, helping to scale new solutions, and building capacity and networks outside of the public service (see also figure 8), especially for those innovation labs on the municipal level dealing with social innovation. Thus, the role of i-labs can differ in the extent to which they are called to experiment and redesign existing services and processes, relishing the ‘skunk works’ mentality, or primarily empowering citizens and enterprises to bring forth change – innovation through the public sector – in an open innovation environment. The latter are exemplified by very lean budgets, crowdsourcing and lightweight structures (see also the case of New Urban Mechanics in Townsend 2013, 214-215).

Figure 8. Selected activities characteristic of i-labs

I-labs try to capitalize on the growing trend of open data based civic apps, as more complex political changes are outside of their control. While most i-labs did not measure the long-term effects of their activities, evaluating their results and impact 3-6 months after the projects, several executives acknowledged that the prototype and accompanying change may only manifest itself after some years. Hence, in many cases there are high lead times between the project and the implementation. Thus, i-labs try to document and publicise their involvement in blogs on social media as much as possible.

While the impact of labs can be measured on different levels – the lab itself, the spin-offs it creates, innovations and innovators it supports, and innovation discourse it helps to establish (Tiesinga and Berkhout 2014, 106) – soft outcomes (networks, discourse change, etc.) – are by the account of i-labs themselves easier to achieve. At the same time, the question remains how apt are i-labs in facilitating
system level change or is their role in the public sector more connected to specific project as examples and legitimisers of further change in the other parts of the public sector organisation.

7.3 Discussion

To summarise our results in detail, we can describe the i-labs as follows. I-labs as they are created today are rather unique in their mission, expected to act as change agents within the public sector and enjoy large autonomy in setting their targets and working methods. I-labs are typically structurally separated from the rest of the public sector and expected to be able to attract external funding as well as ‘sell’ their ideas and solutions to the public sector. They also operate outside of typical performance measurement practices and bring into the public sector new ways to measure performance (use of RCTs, but also big data analytics). However, depending on context, their organisational build-up can considerably differ. As a rule, i-labs have no authority over other public sector structures, thus their effectiveness heavily depends on their ability to communicate and persuade other public sector units through informal networking and using evaluations to gain legitimacy both in and outside the public sector. This provides the i-labs the autonomy as well as incentive to experiment with new solutions and processes. Furthermore, i-labs typically have relatively low budgets and are generally small fluid organisations and are thus, dependent on the resources (funds, human resources) they are able to co-opt to their activities externally.

Yet, this kind of set-up also limits the ability of i-labs to catalyse and push through public sector-wide changes. I-labs tend be small structures, specialising in quick experimentations that usually lack the capabilities and authority to significantly influence upscaling of the new solutions or processes. As their budget does not allow for large demonstration projects, solutions prototyped in these units are technologically more incremental, reliant on co-creation and the involvement of different partners. Interestingly, IT capabilities are not that prominently present in the studied i-labs. However, they could have profound effects on how the public sector conceptualises, designs, and evaluates new innovations and technological solutions: instead of rolling out service changes or new software for an entire customer base, i-labs focus on quick prototyping, diverse evaluation methods, and consequently can have a rapid turn-over of new ideas, discover problems relatively quickly, and have multiple enhancement rounds before any customer tries out the service or product in a real life setting.

However, i-labs, though prominent in many modern public management strategies, are yet far from becoming an organic part of the public sector (there are some notable exceptions, e.g. Helsinki’s innovation company Forum Virium, which was not included in our sample). The main source of autonomy as well as survival is high level political and/or administrative support, meaning that once an i-lab loses its sponsors, the survival chances diminish radically. Here, evaluations new to the public sector are used to gain wider support.
8. Summary and key policy recommendations

This report summarised the LIPSE project work done in Work Package 6 on public sector innovation measurement and indicators. By drawing on a thorough public and private sector innovation literature review, research results from LIPSE WP 1-5, and three original empirical studies the report provided several new insights into public sector innovation measurement and evaluation.

Next to the usual public sector productivity and performance issues, public sector innovations are, in the most abstract sense, related to public authority, where innovations lead to evolutionary changes in constraints and enablers that are intrinsic to the public sector (rules, relationships, institutions). These evolutionary processes use different modalities (innovations within and through the public sector), agency (the public sector proactively initiates changes or reacts to technological, environmental, etc. changes), and morphology (from incremental to discontinuous changes). As such, in the ideal case public sector innovation indicators would be based in a three dimensional system encapsulating impact measurement, the feedback mechanism of the system, and the role of technology.

Yet, these aspects rarely affect public sector innovation or its evaluation directly: public sector impact is difficult to measure, the feedback processes tend to be indirect and slow, and technological change tends to have uneven impact on different public sector activities. All this means that public sector innovation measurement is far more complicated than can be captured by single indicators. Using single innovation indicators provides only a limited opportunity to understand and measure the extent and depth of public sector change. For example, our empirical studies showed that simple metrics are employed to justify investment into ICT projects in the public sector (in the case of Tallinn), but these are rarely used to guide the process of public sector innovation. Even more, it is difficult to only see efficiency oriented indicators as input into public value based debates. The Tallinn cases demonstrate that evaluative frameworks that go beyond simple indicators can reveal much wider change dynamics at play compared to what single indicators can disclose.

We also showed that there are two countervailing dynamics in how to measure public sector innovation: first, 5 large scale attempts (or projects) to create public sector indicators (expressed in concrete numbers and rankings); and second, the emergence of i-labs that base policy design and evaluation on user-centric design approaches that prefer quick prototyping of new services or products, and quick, yet diverse methods in experimentation, feedback, and evaluation. In the case of large-scale indicators, it is surprising that none of the existing attempts relate in a detailed way with widespread public sector performance and auditing activities; and there is no specific attempt to capture a wider set of values.
such as trust and legitimacy. In the case of i-labs, we can argue that these organisations are attempting to transform how public sectors conceptualise, design, and evaluate public sector innovations, however, most of these organisations currently operate on the fringes of the public sector, often with weak impact on core activities and processes. Nevertheless, their evaluative practices signal wider branding and legitimacy goals which are used to gain autonomy to continue their core activities.

As demonstrated by the LIPSE studies, public sector innovations are inherently complex issues to measure as they take place in complex and multiple feedback settings. Efficiency and productivity are key drivers of why and how public sector innovation is made sense of and measured by public sector organisations as well as ‘outsiders’, yet the actual use of specific indicators is problematic due to data and technical difficulties. Therefore, in most cases the level of change – incremental or transformative – is not captured at all or only indirectly captured. Legitimacy, trust, and similar issues are even more difficult to measure. As a result, all these issues are captured and unpacked by public organisations mostly informally through interaction and internal as well as external communication, and thus these issues feed back into public sector innovation processes indirectly rather than directly. In other words, key elements in feedback are described by networks and multiple actors involved in public sector innovations. If in conflict, internal productivity and control of information prevail against external legitimacy concerns (e.g. ease of use and transparency). This is reinforced by the fact that linking legitimacy metrics to innovations is very difficult and hardly ever done (i.e. no real-time measurement of citizens’ reactions, etc.). Although the current ICT era could provide answers to these problems, more direct feedback systems (big data and social media) are also not used.

Perhaps the most startling conclusion is that new technological solutions change the perception of time in the public sector: technological advances and innovations lead to parallel temporalities in evaluating the public sector. First, one can observe the shortening of time horizons in public sector innovation where the logic of change (e.g. efficiency gains) are strong and easy to measure, where user skills match new technological solutions, and where existing technological opportunities are easy to absorb.

Second, the prolongation of time horizons in areas driven by trust and legitimacy concerns, where user feedback is driven by surveys and similar highly roundabout tools, which virtually secures that the input is not taken into account in further developing the technological tools.

The institutional context of public sector innovation processes (in ICT mostly manifested through public procurement of technology) heavily influences innovation feedback processes and thus the extent to which new technologies changed public service provision. Public procurement strategy and in-house capacity to engage with private providers as well as contracting practices and procurement procedures all significantly influenced the decision-making process and ultimately the effectiveness of technology development. Most importantly, it is challenging for the public sector to institutionalise innovation-
enabling interaction and a learning environment within the existing procurement institutions. Existing public sector indicators capture either innovations within public organisations or through the public sector leading/supporting businesses, but not both. Also, public sector innovation indicators capture public sector reactions to technological change and not public sector initiatives to drive technological change.

Based on the research we can outline the following policy suggestions:

- Policymakers and civil servants should refrain from using single figure indicators (such as organisational or country rankings) in designing and evaluating new solutions, both technological and others. Public sector innovations should be designed and evaluated in a wider comprehensive set of evaluative frameworks that allow more complex issues to come to the fore and also allow evaluative exercises to take place over longer periods of time.

- Evaluative frameworks should pay attention, first, to efficiency gains and wider policy goals (such as trust in the public sector, legitimacy of public sector activities), second, to where and how proposed innovations will have the most impact, and third, to how proposed changes relate to the technological (mostly ICT) capabilities of existing civil servants and users.

- It might be advisable to establish specific organisations (such as i-labs, publicly owned innovation companies) to experiment on a small scale with new ideas, solutions and partners. In this way, the involved stakeholders and the public sector can learn from mistakes and failed innovations, and also use diverse evaluation methodologies and frameworks. However, it is important to devise mechanisms for key lessons from such organisations to reach core areas of the public sector.

- Evaluation and measurement efforts need to reflect innovation capacities and, in particular, their dynamic nature: thus measurement should reflect factors influencing innovation capacities (such as those needed to be involved with networks, for co-creation practices, and to carry out risk governance regimes, auditing and performance measurement practices).

- It is, accordingly, important to connect public sector innovation measurement efforts with wider performance measurement attempts, accountability procedures, and risk governance measures. Public sector innovation indicators will invariably feed into accountability systems and risk governance practices.

- Rapidly evolving ICT will have a sizable impact on the nature of data available to the public sector, especially on the city level (smart city solutions). Without developing data-driven and dynamic user interfaces, key public values (such as trust) will remain only vaguely captured in the design and evaluation of new services. However, these new interfaces require quite new and different capacities both from the individual bureaucrats and the institutions involved (in terms of giving larger access to users in design and evaluation). Otherwise, public sector
organisations may end up using data metrics that leave out the most important characteristics of public sector innovation dynamics.

- New organisational forms and experimentation (with co-creation, auditing, procurement) are important elements to take into account in measurement practices. More interaction-enabling public procurement frameworks are essential for removing some of the main barriers of innovation and technological developments. This can be achieved, for example, by using commonly negotiated procedures (competitive dialogues) or communicating technology needs early. Importantly, it is not just what the law is, but also how the law gets interpreted in certain contexts. Therefore, investments in procurement capabilities – including into the legitimisation of risk-taking – constitute an important avenue for changing public sector innovation feedback mechanisms.

- One of the key recommendations from our research is that organisations should vocalize and formalise their innovation and procurement strategies in evaluative frameworks before they set out to procure new technological solutions. The research results indicate that the evaluative framework offered in this report serves as a useful starting point for that.

- And last, but not least, prior to putting together new public sector innovation evaluative frameworks, political support needs to be in place to develop, fund, and facilitate such efforts. This means that public sector innovation indicator frameworks should also be subject to a wider political debate, so that they do not fall victim to political criticism when they feed into broader accountability systems and encounter public scrutiny.

We finish this report with a practical 5-step guide on how to use the new 3D evaluative framework developed in the course of our research. We propose the following steps:

Step 1: Familiarise, test, modify

It is important to understand from the outset what the 3D model is for, what the key elements of the model are and what its dimensions actually want to measure. In chapter 5 we provided a thorough explanation of what the 3D model is. The dimensions of the model are summarised in the following figure.
As such, the ideal public sector innovation measurement and evaluation would be based on a three dimensional system encapsulating the feedback mechanism of the system (x), impact measurement (y), and the role of technology (z). We suggest testing the framework using a previous successful case from within the organisation as it is possible to map the previous case from its initiation to its impact. If possible, comparing two distinct innovations within the public sector (e.g. technology heavy project geared towards internal efficiency vs. innovations with a large consumer base) using the analytical framework may be a good approach to familiarise public servants with the 3D model. It is likely that you will find some aspects of the model that perhaps do not play such an important role in your case or in your organisation. We suggest modifying the model as a key element in familiarising yourself with the model.

Step 2: Set goals, identify capacities

As the next step, we suggest mapping the goals of a proposed change (new technological solution, etc.): what is it and what do internal and external stakeholders understand the main goal of the proposed innovation to be? Is it productivity (e.g. for public organisations or citizens), legitimacy (e.g. higher citizen participation or more responsive service provision), or even technological development through the public sector (e.g. in frameworks of local entrepreneurship policies). Facilitating interaction and communication among stakeholders in defining innovation goals takes time. For this mapping, the three dimensions of the evaluative framework offer initial roadmaps. As
a next step, match the goals of the proposed innovation and the existing capacities of all partners involved. Again, we suggest a quick mapping exercise to locate the strengths and weaknesses of stakeholders. This will likely lead to additional capacity-building exercises, re-configuration of stakeholder networks, or a new assessment of goals. In this stage, it is also very important to gain political support for the model, so, there will not be a political backlash after carrying out the exercise.

Step 3: Select methods, collect data, and compare results

The position of an innovation in the 3D model will also suggest some methodological solutions for measurement. Not all measurement approaches should be used in parallel, but rather, the most appropriate to single cases should be selected (e.g. in the case of public sector internal efficiency, it is probably not useful to study social media reactions). Once there is an understanding of what you want to achieve and whether there are capacities to achieve the goals, it is important to plan data gathering throughout the innovation process: who gathers what kind of data, how are they analysed, and also how will they be communicated – and to whom and by whom. It makes sense to consult all stakeholders as well as external experts on the validity and reliability of selected methods in order to avoid misunderstandings and erosion of interest during later stages. Comparing the results should have both formal as well as interactive dimensions. If the former provides the stakeholders with ‘hard’ data, the latter enables mutually interpreting the facts. This is important if innovation measurement was expected to make substantial impact for all stakeholders.

Step 4: Communicate results

Try to find the key figures in public sector organisations – innovation leaders – who can initiate changes. Try to avoid blaming and shaming, but provide organisation/team specific information to the managerial level. Concentrate on the process: how the measurement process enabled or disabled organisational learning and what kind of impact it had on organisational change. If possible, also bring out positive results to encourage positive competition among public sector organisation managers. Make sure that the information the model provides is actually useful for the target audience and that it is communicated in an understandable way. Change the model if necessary and also give feedback to the target audience in that regard. Be aware that public sector organisations may use measurement results to their own gain (in budget negotiations, interactions with the public, etc.).

Step 5: Ensure continuity

Finally, we suggest you take steps to ensure that evaluation of your innovation is continuous within the 3D model: make sure that information from stakeholders is fed into the evaluation process and
compared against initial goals and capacities. If appropriate, formalise successfully adapted measurement methods in order to further facilitate public sector innovation. Furthermore, previous steps and feedback are also very important to keep measurement efforts financed in the future. Hence, the usefulness of these indicators is key; however, one should not fall into the trap of producing positive results, becoming a PR tool for the government. The perceived and actual objectivity of measurement efforts is key to remain reliable and useful inside and outside of the public sector.
Public Sector Innovation Indicators

References


Public Sector Innovation Indicators


Schumpeter, J. A. (1912), Theorie der wirtschaftlichen Entwicklung, Berlin: Duncker&Humblot


Figure 9. MEPIN various types of innovations by country, 2008-2009 (%)


Figure 10. Novel and any product innovation by country, all government levels, 2008-2009 (%)

Source: Bloch 2011, 33

Figure 11. EPSIS scorecard

Source: EPSIS 2013, 54.
Figure 12. APSII outtake from the State of the Service report