



HAL
open science

Measuring the performance of work integration social enterprises in an evolving policy environment : a comparative study between Denmark and France

Bryan Dufour

► **To cite this version:**

Bryan Dufour. Measuring the performance of work integration social enterprises in an evolving policy environment : a comparative study between Denmark and France. Economics and Finance. LEST; Aix Marseille Université, 2019. English. NNT : . tel-03560525

HAL Id: tel-03560525

<https://hal.science/tel-03560525>

Submitted on 9 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Aix-Marseille Université
Faculté d'Économie et Gestion (FEG)
École Doctorale de Sciences Économiques et de Gestion d'Aix-Marseille (ED 372)
Laboratoire d'Économie et de Sociologie du Travail-CNRS (UMR 7317)

Measuring the performance of work integration social enterprises in an evolving policy environment: a comparative study between Denmark and France

Bryan DUFOUR

Thèse présentée pour obtenir le grade universitaire de docteur en Sciences économiques

Soutenue le 25 avril 2019 devant le jury composé de :

Madame Francesca PETRELLA

Maître de conférences HDR, LEST-CNRS, Aix-Marseille Université, Directrice de thèse

Monsieur Lars HULGÅRD

Professeur, Université de Roskilde, Rapporteur

Madame Nathalie RAULET-CROSET

Professeure, Université Paris 1 Panthéon Sorbonne - IAE, Rapporteur

Monsieur Jérôme BLANC

Professeur, Sciences Po Lyon, Suffragant

Madame Nadine RICHEZ-BATTESTI

Maître de conférences, LEST-CNRS, Aix-Marseille Université, Suffragante

The university neither approves nor disapproves the opinions expressed in this thesis.

These opinions are to be considered the author's own.

Table of Contents

ACKNOWLEDGEMENTS	7
GENERAL INTRODUCTION	9
PART 1 CONTEXT, ANALYTICAL FRAMEWORK AND RESEARCH PROTOCOL	21
CHAPTER 1 THE PLACE OF WORK INTEGRATION SOCIAL ENTERPRISES IN PUBLIC ADMINISTRATION AND MANAGEMENT.....	23
1.1 THE PLACE OF SOCIAL ENTERPRISES IN THE CHANGING NATURE OF THE STATE	25
1.2 PERFORMANCE AND ITS MEASUREMENT	34
CHAPTER 2 THE SHIFTING SHAPES OF PUBLIC ADMINISTRATION AND MANAGEMENT.....	47
2.1 NEW PUBLIC MANAGEMENT.....	49
2.2 THE MANY HOUSES OF NEW PUBLIC MANAGEMENT	52
2.3 TWO MODELS OF INTEREST: THE NEO-WEBERIAN STATE AND THE NEW PUBLIC GOVERNANCE	57
2.4 PUBLIC VALUE THEORY: AN UNDERLYING LOGIC RATHER THAN A MODEL.....	66
CHAPTER 3 PERFORMANCE: ITS MEASUREMENT AND ITS REGIMES	83
3.1 PERFORMANCE AND PERFORMATIVITY.....	85
3.2 PERFORMANCE IN PUBLIC ADMINISTRATION AND MANAGEMENT	97
3.3 THE PERFORMANCE REGIMES	106
CHAPTER 4 THE PERFORMANCE OF WISES?	111
4.1 CAPTURING THE PERFORMANCE OF SOCIAL ENTERPRISES: ENTERS SOCIAL IMPACT ASSESSMENT	113
4.2 A SHORT HISTORY OF SOCIAL IMPACT ASSESSMENT	120
4.3 SOCIAL IMPACT ASSESSMENT: A STATE OF THE ART.....	127
CHAPTER 5 A PVT-INFORMED, CRITICAL REALIST ANALYTICAL FRAMEWORK.....	141
5.1 THE COMPLEMENTARITY OF OUR THREE THEORETICAL ANCHORS.....	142
5.2 OPERATIONALISING OUR THREE THEORETICAL ANCHORS IN OUR RESEARCH	145
5.3 DELINEATING OUR RESEARCH QUESTIONS	150
CHAPTER 6 METHODOLOGICAL FRAMING AND EMPIRICAL MATERIAL	155
6.1 PARADIGMATIC APPROACH	156
6.2 METHODOLOGICAL APPROACH	163
6.3 PRESENTATION OF OUR EMPIRICAL MATERIAL	182
PART 2 EMPIRICAL RESULTS	201
CHAPTER 7 MECHANICS OF PUBLIC VALUE CO-CREATION IN FRANCE AND IN DENMARK.....	205
7.1 FRANCE: WORK INTEGRATION AND SOCIAL ECONOMY IN A NEO-WEBERIAN STATE.....	207
7.2 DENMARK: WORK INTEGRATION AND SOCIAL ECONOMY IN THE NEW PUBLIC GOVERNANCE	222
CHAPTER 8 PERFORMANCE MEASUREMENT IN FRANCE: BETWEEN AUTOCRACY, MANAGERIALISM AND DELIBERATION	239
8.1 OBSERVED PRACTICES	241

8.2	EFFECTS OF NEO-WEBERIAN STATE MECHANISMS ON STRATEGIC INTERACTIONS	255
8.3	BETWEEN AUTOCRACY, MANAGERIALISM AND DELIBERATION: HOW PERFORMANCE MEASUREMENT EVOLVES UNDER THE NEO-WEBERIAN STATE	262
CHAPTER 9	PERFORMANCE MEASUREMENT IN DENMARK: DELIBERATION OR CHAOS?	267
9.1	OBSERVED PRACTICES	269
9.2	EFFECTS OF NEW PUBLIC GOVERNANCE MECHANISMS ON STRATEGIC INTERACTIONS.....	287
9.3	DELIBERATION OR CHAOS: HOW PERFORMANCE MEASUREMENT EVOLVES UNDER NPG	296
CHAPTER 10	HOW STRATEGIC INTERACTIONS FRAME THE MEASUREMENT OF PERFORMANCE IN DIFFERENT CONTEXTS	303
10.1	SUMMARISING THE OBSERVED PRACTICES	305
10.2	HOW PAM MODELS SHAPE THE PRACTICE OF PERFORMANCE MEASUREMENT?.....	311
10.3	STRATEGIC INTENTS AND WHAT THEY TELL US ON THE CONCEPTIONS OF PERFORMANCE AND ITS MEASUREMENT.....	318
10.4	THE ROLE OF SHARED PREFERENCES AND SOCIO-TECHNICAL ARRANGEMENTS	329
CHAPTER 11	GENERAL DISCUSSION	341
11.1	THEORETICAL CONTRIBUTIONS	342
11.2	METHODOLOGICAL CONTRIBUTIONS.....	353
11.3	EMPIRICAL CONTRIBUTIONS	356
11.4	LIMITS AND FUTURE RESEARCH ALLEYS	360
<u>CONCLUSION</u>		<u>365</u>
<u>BIBLIOGRAPHY.....</u>		<u>371</u>
<u>INDEXES OF TABLES, FIGURES AND APPENDICES</u>		<u>389</u>
<u>APPENDICES.....</u>		<u>393</u>
<u>SUMMARY</u>		<u>461</u>

Acknowledgements

This thesis represents the capstone of four tumultuous years, spent on spinning perhaps a little too many plates. A lot has happen in these four years, both professionally and personally, and during that whole time, this thesis has been a rather strange beacon: a goal to focus on while other things were in motion and sometimes uncertain, a source of delight to quench my curiosity, a pretext to travel and share experiences with a community of like-minded researchers, but also a cradle for doubts, frustrations and challenges that continue to shake my self-confidence. As daring as it has been, I know that this experience has made me grow and develop in ways I could not have foreseen. I owe this to the many people that are listed below and to their support.

My recognition first and foremost goes to my wife Majbrit, whose support never failed during these four years, and who leniently sacrificed many hours of our family life for this work to come to fruition. To my son Valder: you were not born when all this started, I was not married to your mother and the house we live in was not built. This work has taken too many hours away from the first years of your life, and for this I apologise.

My gratitude goes to Francesca Petrella and Nadine Richez Battesti, who have supported me and believed in me despite conditions that were far from optimal. Francesca, your guidance and your aptitude to ferreting out leads out of seeming dead ends have been instrumental in shaping my thesis, while your optimism and your breezy attitude have been invaluable in keeping the pressure to a manageable level. Nadine, your composure and your inquisitive mind are irreplaceable and are imprinted in many aspects of this work. My thankfulness extends to the whole team at the LEST, amongst which I have found a group of open and forthcoming people who supported me on several aspects of my thesis.

My gratitude also extends to Thomas Lagoarde Segot, who brought instrumental contributions to my research paradigm and who provided the decisive pointers that allowed me to publish some of my work.

I also want to thank Laurent from FR1, who dedicated a substantial amount of time and resources to my research. My thankfulness extends to Françoise and Mathieu from FR2, Maria from DK2 and Marianne from DK1, who have all provided time and resources for my case research to be carried out in the best possible conditions. As I write this, a special thought goes to Lars Hulgård, who not only supported me in identifying cases in Denmark, but who also encouraged me to take on the Ph.D. journey as early as 2012 when we first met in his office at Roskilde University.

I want to particularly thank Karen and Kurt, who crossed Denmark multiple times to give us a hand at home so that I could finalise this manuscript. And for providing moral support and a good laugh when I needed one, my thoughts go to my good friends: Alessandro, Antonino, Anne Louise, Antoine, Bjarke, Elina, Federica, Félix, Filippo Lasse and Tola. Special thanks go to Alexandre and Jean-Michel who welcomed me during four years for every stay I have had in Marseille and Aix-en-Provence.

I also found support amongst my colleagues: thank you Lotte for the cheery mood and the laughter. A particular gratitude goes to Edvard for being supportive and showing a high degree of flexibility in order for me to carry out research work while having a full-time job. My appreciation in this regard also extends to Joan. I also want to express my thankfulness towards Steen Thygesen, without whom I would not have been able to start my Ph.D. This did not go the way we anticipated, but it will remain a key milestone of my life.

Last, but certainly not least, I want to thank both my parents. They are the individuals who brought me up and made me the person I am, with the values and sensibilities that I carry into my work. Maman, you have always been supportive in my academic pursuits and you have done sacrifices I am too aware of so that I could succeed. You should know I will always be grateful. Papa, your support, your interest and your curiosity during these four years have meant a lot, and I look forward to having more occasions to visit you and Jeanne now that I will have some personal time back.

General Introduction

“What we measure shapes what we collectively strive to pursue” Stiglitz, Sen and Fitoussi, 2009

In January 2018, Jean-Marc Borello, a social entrepreneur supporting the current French government, delivered his report on work integration to the Ministry of Labour. When the results were presented to the Steering Committee of a national network of work integration services providers, we sat in the room. *“This is what we feared...”* the delegate muttered before presenting his summary of the report.

Considerations on performance are driving many of the leads Mr. Borello suggests to achieve a more inclusive labour market. As public budgets are getting tighter, and because work integration is typically a sector that relies, at least partially, on public subsidies, recommendations are made to tie funding to the performance of the organisations providing work integration services. The Committee in which we were sitting was not so much worried about being assessed, they were concerned about the form of this assessment. How does one measure such a performance? What indicators can be used to gauge the effect of an organisation on one’s employment track? Can these indicators be detrimental to the broader goal of inclusion? Are they encompassing the value that field actors perceive they bring to the citizenry? These are just a sample of the many questions that the performance of work integration services providers is raising, not only in France, but in an increasing number of places where such services are offered, and particularly when they are delivered by non-public providers.

Such providers are often labelled as work integration social enterprises (WISEs) in the literature (see e.g. Battilana, Sengul, Pache, & Model, 2015; Davister, Defourny, & Grégoire, 2004; Petrella & Richez-Battesti, 2016). The development of WISEs is often tied to the historical context of the late 1960s and early 1970s, a period that marked the end of the post-war growth era for Western economies. The last years of what the French call the *“Glorious Thirty”* came with swift social change and an accompanying shift in global policy making. One noticeable trend in the decade’s turmoil was the diffuse appearance of unemployment in

developed economies. Amid the collapse of the Fordist employment model, as family structures evolved and as governments started to be preoccupied with policy *performance* considerations, an expanding pool of individuals were left on the margin of society. It is in this context that WISEs came to play a growing role in welfare societies, as mission-driven organisations meant to cater to these individuals by helping them get (back) into employment.

Because of their efforts on work integration, WISEs are particularly praised in economies struggling with unemployment issues (OECD, 2013; Sibieude, 2010). Their appeal for policy maker is also often associated to the rise of the New Public Management (NPM, Hood, 1991). NPM is a Public Administration and Management (PAM) model that promotes the use of private-inspired management mechanisms in the public sector. NPM reforms started developing in the late 1970s, at a time where WISEs began to spread. This is of course not coincidental, as NPM paints the backdrop of the welfare shifts and focus on performance that we describe above. It is commonly admitted that in this context, and as third-party providers (i.e. external to the state apparatus), WISEs appeared (and still do so) as a good instrument to bear the burden of flexibility and efficiency asked by the NPM for the delivery of some welfare services (see e.g. Defourny & Nyssens, 2010; Hulgård, 2011; A. Nicholls, 2010; Nyssens, 2006; Pestoff, 2008).

Beyond this however, we make the observation in our thesis that public actors maintain collaborations with WISEs outside of pure NPM mechanisms. We for instance see these interactions foster in countries like France and Denmark (the two countries in which we ascribe our research), which are typically portrayed as reluctant to the full implementation of the NPM paradigm (Greve, Laegreid, & Rykkja, 2016; Hammerschmid, Van de Walle, Andrews, & Bezes, 2016). In these two countries, although we find some layers of NPM reforms (i.e. some components, but not the entire paradigm, Christensen and Laegreid, 2009), we also notice a rooting of social economy in traditions as well as cultural and economic developments anterior to the inception of the NPM. It is therefore through a variety of drivers that WISEs, together with the broader population of social enterprises, have gained momentum in the global political agenda in the past two decades. The multiple reports that

have emerged from organisations such as the European Commission (2013, 2014), the Organization for Economic Cooperation and Development (OECD, 2010, 2013, 2015) or the International Monetary Fund (IMF, 2012) in the support of their fostering are evidencing this trend.

As WISEs gain prominence as key partners in the provision of welfare services, and as they become a topic of interest on the political agenda, the questions pertaining to their performance that we raised in our first paragraph appear of a particular interest.

Our research object

We frame our research object around the three pillars that have formed the reflections leading us to our initial questioning: work integration social enterprises, their place in public administration and management (PAM) and considerations pertaining to performance measurement (in general and for WISEs specifically).

Work integration social enterprises

WISEs are typically considered a subset social enterprises (Davister et al., 2004; Nyssens, 2006). The definition of social enterprises is still in flux, which is a debate we briefly cover in our thesis. For now we will qualify them as organisations focused on the creation of social value, through entrepreneurial means likely to involve commercial revenue (see e.g. Dees, Ewing, Haas, & Haas, 1998; Mair & Martí, 2006; Yunus, 2006). On this basis, WISEs are social enterprises focusing on employment-related outcomes. This narrow locus grants them the benefit of boasting a relatively consensual definition (Davister et al., 2004).

The place of work integration social enterprises in public administration and management

By providing work integration services to marginalised individuals whose needs could not be catered to by public or market offerings, WISEs propose a remedy to a market failure (Bator, 1958; Zeckhauser & Stokey, 1978). We find that public value theory (PVT, Moore, 1995) associates such services to a form of public value (Alford & O'Flynn, 2009). PVT, from which we derive our central analytical framework (cf. below) considers that public value is, in essence, the value generated for the citizenry (Moore says that public value is to citizens what

dividends are to private companies' shareholders). This constitutes an interesting take on WISEs, which as social enterprises, have been identified by some PVT authors such as Bryson, Sancino, Benington, & Sørensen (2017) as co-creators of public value.

Such authors insist on the **co-creation** of public value that can occur when organisations like WISEs cooperate with public actors. However, with NPM in the background, even in countries like France and Denmark which do not strictly apply its philosophy, we quickly realise that performance measurement is part of the public value co-creation environment. Understanding the mechanisms framing this co-creation will therefore be a central point to clarify in our thesis: what are the co-creation models in France and in Denmark? How do they differ? How do they translate in terms of strategic interactions (cf. below) between WISEs and their public stakeholders? And how do they contribute to defining a baseline for performance?

Performance and its measurement

Performance measurement in general:

This last interrogation leads us to the third pillar of our initial questioning, which is the issues pertaining to performance measurement. Leaving aside our interest in WISEs for a moment, we see that performance measurement is a topic that have interested sociologists and economists alike. From the economic angle, we retain especially the questions pertaining to the creation of *new* performance indicators, often related to the debates on development indicators (see e.g. Gadrey & Jany-Catrice, 2016; Jackson & Stymne-Airey, 1996; Miringoff, 1995; Stiglitz, Sen, Fitoussi, & others, 2010). Gadrey and Jany-Catrice summarise these issues with the following quote: *“a new system of values and a new hierarchy of the ‘items that count’ is under way and will penetrate both political discourses and decisions”* (2016, 7).

Our thesis also abundantly delves into considerations introduced by sociologists such as Callon (2006, 2009) and Desrosières (2012, 2014), particularly around the notion of performativity. Performativity designates the tendency for *measurement* instruments to have an influence on the phenomena they are actually *measuring*. An example can be provided through the work of Salais (2010), who shows how a focus on lowering unemployment

statistics can actually be detrimental to employment outcomes. Other authors such as Jany-Catrice (2012), Gadrey and Jany-Catrice (2016) or Desrosières (2012) offer several examples of such occurrences (e.g. the negative externalities of some activities captured as positively contributing to economic growth). They also abound about the democratic implications of performativity, insisting that conventions (the arithmetic and assumptions going into the design of indicators) should be more transparent so as to allow better informed policy decisions.

Performance measurement for WISEs:

Generally speaking, performativity matters because it can impact the **outcomes** of the organisations or phenomena it measures (and we see very early in our research that outcomes are fundamental to assess WISEs' performance) and because of its **democratic** implications (and having clarity on WISEs' performance becomes a democratic issue since they are an active partner of labour policies). Furthermore, as mission-driven organisations pursuing a social purpose through market-oriented means, WISEs offer a compelling case study to research the emerging topic of new performance indicators at the micro scale.

With these issues in mind, **social impact assessment**¹ (SIA) is an object that sparked a particular interest in our research. SIA refers to a collection of approaches that have been developed by mission-driven organisations such as WISEs² in order to gauge their performance (for contributions presenting SIA as a means to gauge performance, see e.g. Austin, Stevenson, & Wei-Skillern, 2006; Grieco, Michelini, & Iasevoli, 2015). A common characteristic to the many SIA models available is that they strive to propose means to assess what Gianfaldoni (2018) or Lall (2017) call "**social performance**". These frameworks are extremely diverse (we identified 124 models in our research), but two particularities are observed in almost all of them. The first one is a focus on outcomes (as opposed to outputs, which *traditional* performance measurement approaches typically tend to capture). And the second is the deliberative process that is advocated for their design (i.e. stakeholders should

¹ By social impact assessment, we designate all the approaches developed by mission-driven organisations to gauge their performance, including what can be referred to as "*social impact measurement*"

² And including many other stakeholders of the social economy such as other social enterprises, foundations, associations, mutuals, cooperatives and even so-called *impact investors*

be involved in the design of processes and indicators). We found that these two structuring features of SIA interestingly echo the issue that performativity highlights.

Furthermore, it also appears that prompted by the increasing importance of WISEs as instrument of welfare, policy makers have started to consider SIA models for questions pertaining to the measurement of their performance (see e.g. European Commission / GECES, 2014; G8, 2014; OECD, 2015a, 2015b). This prods us to ask: are governments truly implementing SIA? Can this help alleviate issues pertaining to performativity? Knowing that popular SIA models such as SROI³ consist in a monetarisation of impacts through the use of financial proxies, can SIA convey relevant information for public actors? More broadly, we may ask whether these models (in part developed by impact investors) encompass the right information for policy makers to make informed democratic choices on topics pertaining to labour policy.

Initial questioning

As WISEs become increasingly relevant welfare actors, in a context where PAM models advocate for a quasi-systematic use of performance measurement, and in view of the trends we observed with SIA, the initial questioning that prompted us to write this thesis was: how is the development of SIA changing the way that public actors are measuring WISEs' performance?

Seeing as the development of WISEs and SIA were phenomena observed internationally, and that the question of performance measurement was bore by international organisations such as the OECD or the European Commission, we sought to parallel these developments in our research and ascribed our work to an international comparison. As two of the countries boasting some of the highest level of public welfare in the world, but through two very different models, a comparison between France and Denmark seemed particularly appealing.

³ Social Return On Investment

Part one: analytical framework and research protocol

The thesis that recounts our investigation is structured in two parts (cf. Figure 1). The first one presents and deepens the concepts that we mobilise in our research and provides the analytical framework we use to come to our conclusions. The second part builds upon this framework to present the results of our investigation and answer our research questions.

We start Part One with a chapter contextualising our research, detailing further the three pillars of our research object, which we just introduced (WISEs, their place in PAM and performance measurement). Having established the foundation of our research, Chapters 2 through 4 provide the analytical framework that we marshal in order to move from our initial questioning to a well delimited research question.

Chapter 2 introduces the scaffold that we mobilise to understand the interactions that WISEs have with public stakeholders. In our comparative analysis, these relationships are the complex product of different actors evolving in distinct cultural and political environments. Our analysis therefore distinguishes between the micro level, which is the direct interactions between WISEs and their public stakeholders, as informed by our case studies, and the macro level, which aims at understanding the environment in which the micro observations are formulated. For the macro level, the primary analytical framework we use is PAM models, which are different in the two countries we research: the **Neo-Weberian State** in France (NWS, a modernising administration concerned with preserving its traditional apparatus, Pollitt & Bouckaert, 2004) and the **New Public Governance** in Denmark (NPG, an open, network-driven paradigm focused on inclusive governance; Osborne, 2006). We also use Kuhlmann and Wollmann's administrative profiles (2014) as our auxiliary instrument in our analysis, reflecting the influence of local cultures and traditions: France is a member of the Napoleonic group and Denmark of the Nordic's.

For the micro level, we use PVT, which allows us to develop an original analysis of the interactions between WISEs and their public stakeholder around the public value they co-create. We use PVT's strategic triangle to this end (three vertices that have to be aligned in

order to produce public value: **legitimacy**, **operationalisation** and **public value** itself). We expand this initial framework into nine original entries (one of our thesis' contribution) that we mobilise to understand phenomena at the micro level, and particularly the **strategic interactions** between WISEs and their public stakeholders. What we call strategic interactions are the aspects of the relationship revolving around the co-creation of public value, of which performance measurement is a part. E.g. how do WISEs secure the necessary legitimacy to operate in a given PAM model? Can measures of performance help?

Performance measurement is precisely what Chapter 3 is about. An area of focus in this chapter is the complex notion of performativity, which captures many of the stakes associated to the measurement of performance such as the potential influence that indicators can have on outcomes (when indicators are misused) and on democratic processes (when the design of indicators is not well understood). As we unpack performativity, we introduce an original framework to apprehend some of performativity's causes and effects (another contribution of our thesis). The first element of our framework is **socio-technical arrangements** (STAs, Callon, 2006): the mutual interactions entertained between performance instruments, their users and the environment in which they are developed and operationalised. A SIA model such as SROI, because it provides a frame of reference to gauge performance, is typically a form of STA. According to Callon, STAs have to be enacted in order to become performative (i.e. all STAs are not performative). The second element is **retroaction** (Desrosières, 2012): the way actors may change their behaviour in reaction to indicators monitoring their performance. The last element is what we call **indicator-substitution**, or the tendency for indicators engineered through opaque conventions to replace the social realities they are meant to represent.

Our Chapter 3 also provides us with two additional frameworks that we use to parse the performance measurement models we come across. The first one is a general design to apprehend performance in a PAM setting, informed by seven performance criteria. The second is Jany-Catrice's four performance regimes (2012): a scaffold structuring performance models around two dimensions of (i) planned versus deliberated design and (ii) single versus multiple indicators. Jany-Catrice's contribution is particularly useful in our research in order

to characterise the performance measurement approaches we come across in our research, especially at the macro level, distinguishing for instance between the NWS' tendency of imposing top-down performance models and the NPG's incline towards deliberative designs.

These two schemata have a direct influence on the last analytical framework that we introduce in our Chapter 4 to scrutinise and parse the way that the performance measurement models we encounter capture and communicate elements pertaining to social performance. This last chapter of our literature review focuses on SIA, or the performance measurement models developed by mission-driven organisations such as WISEs. We present a state of the art of SIA identifying 18 models particularly fit for WISEs and we introduce our own scaffold to dissect such models. This framework is aimed at providing an understanding of all the performance models we encounter (i.e. not only SIA models). Since we are interested in WISEs' social performance (Gianfaldoni, 2018; Lall, 2017), SIA is an excellent place to start building our analytical tool because of its history and purpose as an instrument championed by mission-driven organisations. Our own research on performance measurement in Chapter 3 permits to complement our framework to make it more thorough. For instance, Jany-Catrice's regimes is what allows us to add an analytical dimension pertaining to the models' design (which can be either planned or deliberated), which is surprisingly absent from the 118 variables identified in the SIA literature.

Chapter 5 brings together all the elements of our literature review to propose a comprehensive analytical framework encompassing all aspects of our research object (the place of WISEs in PAM, the models through which they can co-create public value and performance measurement). We also capitalise on the knowledge brought forth to reframe our initial questioning around two aspects that are core to the issue at hand: the broader matter of performance measurement (rather than our initial focus on the restricted sphere of SIA) and the strategic interactions occurring between WISEs and their public actors, which is where PVT suggests we will find relevant elements to explore for performance measurement from one PAM model to another. We hence come to craft the following research question:

How strategic interactions between WISEs and public stakeholders are contributing to frame the measurement of performance in different PAM contexts?

We close the first part of our thesis by providing the research protocol that will allow us to answer this question. We articulate the philosophy underpinning our paradigmatic approach: the critical realism (CR, Bhaskar, 1975; Lawson, 2003). One of CR's distinctive attribute is the *stratified ontology* where three levels of reality are identified: the empirical (the facts as they are experienced: a leave is flying), the actual (the facts as they occur: the wind is carrying the leave) and the deep real (what causes the occurrence: air masses of different temperature are causing a depression). Because of its focus on social phenomena as open systems, the application of the critical realist paradigm in economics is particularly relevant to our research as we are interested in occurrences like performativity, which typically run the risk of being left out by the closed systems of more orthodox approaches (Lawson, 2003; 2006).

Chapter 6 also details the methodology through which we operationalise CR: a sequential dominant mixed research with multiple embedded case studies. The dominant stance of the design is qualitative (although more dominant than initially anticipated due to challengingly low participation rates in our quantitative research in France). Our data was treated in Nvivo 12 through an *ex-ante* coding grid complemented by emergent codes (Miles & Huberman, 2003), a method compliant with CR's retrodution (a form of abductive reasoning used by CR researchers). We surveyed 31 WISEs in Denmark (an estimated 10% of the country's total population) and 12 in France (0.3%), after which we carried two case researches in both countries, which we also present in Chapter 6.

Part 2: empirical results

The second part of our thesis delivers our results and the analysis we make of it based on the research question and the frameworks introduced above. Chapter 7 starts by providing a description of the public value co-creation environment observed in Denmark and in France, fleshing out the NWS and NPG models we introduced in our Chapter 2. The structuring difference between the two countries appears to be a **legalist**-based approach of work integration services in France, where WISEs must secure a formal institutional agreement (a

covenant) to be able to operate; and what we call a **relationship-based quasi-market** in Denmark, where WISEs offer work integration as a commercial service to their public stakeholders.

Chapters 8 and 9 describe the strategic interactions evidenced between WISEs and their public stakeholders as they co-create public value. In these chapters, our PVT analysis reveals that the French legalist environment tends to favour strategies focused on the generation of **legitimacy**, as WISEs strive to keep their covenant and therefore their *right* to operate. In Denmark we see that WISEs appear to deploy more **balanced** strategies as their competitive environment requires them to deliver public value based at market terms. The fundamentals of the relationships between WISEs and their public stakeholders that our analysis of strategic interactions capture appears essential to fully comprehend the different performance models we recount in these chapters and the way they are used.

This leads us to Chapter 10, where by assembling the results of the field research carried out in the two countries, we unearth and characterise the **strategic intents** tied to performance measurement. A key finding of our thesis, the strategic intents shed light on the motives tied to performance measurement and the use of one performance model over another. On this basis, we identify the strategic intents as one causal mechanism for the emergence of performativity through strategic interactions. In this chapter, we also evidence a second causal mechanism which is the public co-creation environments induced by PAM models.

Our discussion in Chapter 11 confronts these findings with the existing literature. As we do so, we develop further about STAs and we surface one of the key contributions of our thesis: **socio-technical arrangements' feedbacks**. We suggest that although STAs are indeed conveyors of performativity, their fostering under certain conditions can also be virtuous and limit performativity.

Before proceeding to the content of our thesis, Figure 1 on the next page provides a visual queue to the and structure of our work and its substance.

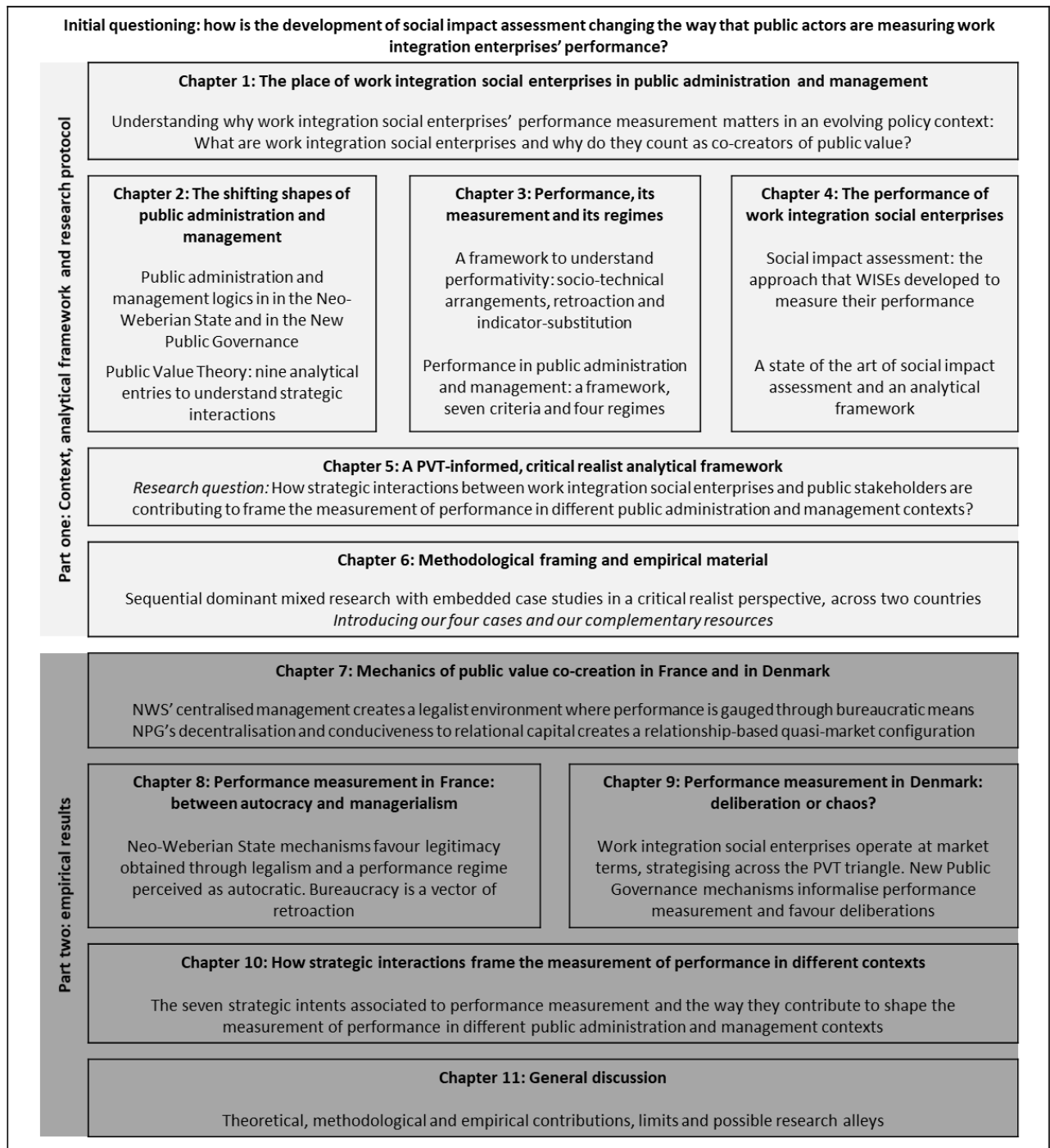


Figure 1: Summarised thesis content

Source: Author

Part 1 Context, analytical framework and research protocol

The first part of our thesis provides the analytical framework that we build to investigate our research topic. It concludes with the articulation of the research questions we aim to elucidate and the epistemological and methodological framework we propose to answer them. The choice of theories and analytical tools that we present in these six chapters is the outcome of a lengthy reflection process. The purpose of this introduction is to make this process more explicit by emphasising the practical considerations and the stakes behind our choices.

We start by introducing our principal research object, the **work integration social enterprises** (WISEs). By delivering work integration services to marginalised individuals whose needs have not been catered to by existing public or market offerings, WISEs propose **remedies to market failures** (Bator, 1958; Zeckhauser & Stokey, 1978). By doing so, WISEs contribute to the creation of **public value**, as framed in the public value theory (PVT, Moore, 1995). In a context of endemic unemployment and state withdrawal in the provision of welfare services, this positions WISEs as increasingly relevant partners for public actors.

As such, WISEs appear to be concerned with the phenomenon of **performance measurement**, which is spreading to a growing number of public administration and management (PAM) models. However, the mission-driven nature of WISEs poses the challenge of the measurement of **“social performance”** (Lall, 2017, Gianfaldoni, 2018). To address this issue, public authorities have manifested a growing interest for **social impact assessment** (SIA) models, or the approach that WISEs and other mission-driven actors have developed over the last decades to get a better grasp of their own performance. But do these models change the way public actors measure WISEs’ performance?

To bring an answer to this question, we must first understand what characterises performance for WISEs’ public stakeholders. Chapter 2 provides the foundations we use for this by introducing the two PAM models that we use (at the macro level) to frame the

mechanisms of public value co-creation in the two countries we research: the **Neo-Weberian State** in France and the **New Public Governance** in Denmark. In the same chapter, we also present the original analytical framework we build around PVT, which we use to understand occurrences at the micro level. A particular element of interest at this level is the **strategic interactions** between WISEs and public actors, which is where issues pertaining to performance measurement are addressed (*inter alia*).

How can strategic interactions affect performance measurement models? Chapter 3 introduces the fundamental concept we mobilise to capture an answer to this question: **performativity**, or the influence that performance instruments can have on what they *measure*. Performativity is an important notion that helps us identify the implications that performance measurement models and their indicators have on WISEs' outcomes. In order to frame these models and indicators with a systematic approach, this chapter also introduces two additional frameworks: the performance regimes and seven criteria for performance measurement.

What are the specificities of SIA models and what do they bring to performance measurement? Building upon the fundamental considerations on performance measurement highlighted above, Chapter 4 capitalises on the SIA literature to propose a last analytical framework dedicated to apprehend the specificities of performance measurement models used to capture social performance.

In Chapter 5, we present the overarching analytical framework that our literature review allows to build and that we use to articulate our research questions. Chapter 6, the last chapter of this section, introduces critical realism (Bhaskar, 1975; Lawson, 2003). as our research paradigm and the societal approach (Maurice, 1989) as the method we choose to carry out international comparison. Both of these scaffolds have a complementary ontology that allows us to research causal mechanisms as the (fallible) best explanation of reality.

Chapter 1 The place of work integration social enterprises in public administration and management

Ever since their emergence in the late 1960s after decades of growth prompted by post-war reconstruction, **work integration social enterprises** (WISEs) have played an increasingly important role in labour policies. Today, as economies struggle to eradicate unemployment in fringes of the population that seem more and more marginalised, WISEs' role keeps growing in relevance. Their cooperation with public actors is documented in several contributions (see e.g. Defourny & Nyssens 2010; Hulgård & Andersen, 2015, Petrella & Richez-Battesti, 2016). Yet, as these organisations can provide salutary services to individuals having exhausted options offered by public welfare services, their performance increasingly comes in question, as does an ever-expanding slab of public policies.

This first chapter offers to understand the backdrop of the picture we just painted, by introducing the elements essential to its comprehension. Defining WISEs and their focus on employment issues (Davister et al., 2004), we understand that the interest governments lend to them ties to their inherent mission of offering **employment services** to marginalised individuals. Doing so, WISEs provide a valued **market failure** remedy, i.e. a solution for these individuals that neither private actors nor public welfare service were able to successfully propose (Bator, 1958; Zeckhauser & Stokey, 1978).

We show that WISEs' appeal to policy makers verifies across public administration and management (PAM) models such as France's **Neo-Weberian State** (Pollitt & Bouckaert, 2004) or Denmark's **New Public Governance** (Osborne, 2006). But we highlight that **New Public Management** (NPM), a PAM model favouring mechanisms from the private sector (Hood, 1991), is particularly conducive to the development of WISEs, as a device to further market logics in welfare services. Doing so, we briefly introduce **public value theory** (PVT, Moore, 1995) as our analytical framework for issues cutting across different PAM models.

We then offer an explanation to the recent pre-eminence gained by the question of WISEs' performance, which we tie to two elements. The first is a trend favouring the development of performance measurement across most forms of public activities, which is driven mainly

by an ask for efficiency and accountability. Here again, NPM offers a particularly conducive framework but we show that this trend was already under way (albeit at a slower pace) before NPM came to prominence.

The second element is the inherent difficulty to capture the performance of hybrid organisations like WISEs (Battilana et al., 2015). Authors like André, Cho, & Laine, (2018), Gianfaldoni (2018) or Lall (2017) speak of **social performance**, but how do we *measure* the difference an organisation made for someone's employment record? This question leads us to the emergence of **social impact measurement** (SIA) models, which are solutions initially imagined by WISEs and other mission-driven actors to measure their performance, and that are today being studied by public actors. The question above also brings us to issues pertaining to **performativity**, or the influence that performance instruments can have on what they *measure* (Callon, 2006; Desrosières, 1993).

The latter crystallises the last fundamental issues we introduce in this chapter: as we ask the question of how WISE's performance should be measured, performativity reminds us that indicators design can have an influence on the actual **outcomes** of an organisation and can even carry issues pertaining to **democracy**. It is also the interest we take in performativity that allows us to tie our research to broader considerations in the field of economics, about the use indicators and the need for a renewal in the way they are conceptualised and operationalised (Gadrey & Jany-Catrice, 2016).

1.1 The place of social enterprises in the changing nature of the state

What exactly are work integration social enterprises (WISEs) and why do they matter for labour policies? These are two essential questions that must be elucidated before getting to our core issue of WISEs' performance measurement. The purpose of this section is to bring a detailed answer to the first question and to present the frameworks that help elucidating the second.

As social enterprises are not consensually defined in academia, we start by providing an overview of the structuring questions around the issue before positioning ourselves. Moving to our second question, we cover the linkage between WISEs and new public management reforms proposed in most of the existing contributions. We go further and also highlight WISEs' relevance in other public administration and management contexts, particularly as providers of **remedies to market failures**. We conclude by briefly introducing public value theory as the framework that will allow us to further analyse the way WISEs interact with public stakeholder to fulfil their mission (and therefore be subject to performance measurement).

1.1.1 What is a work integration social enterprise?

Before diving into the context of our research, we start by precisely delineating the object of our inquiry, namely work integration social enterprises (WISEs). As WISEs are in effect a subset of social enterprises (SEs), we start by defining the latter, before moving to the former.

1.1.1.1 A difficult notion to coin

There is no consensus in academia or among practitioners on the definition of the term "social enterprise". Looking at the literature, social enterprises are often outlined through the definitions of social entrepreneurship (Light, 2008; Mair & Martí, 2006; Moulaert, 2013), although the use of one or the other term is not neutral (Petrella & Richez-Battesti, 2014). Several schools of thoughts are confronting their views on what makes the essence of social entrepreneurship. Two prominent takes are one slant that focuses on social value creation through innovation (Dees et al. 1998; Roberts & Woods, 2005) and another perspective

putting the emphasis on social value creation through market mechanisms such as earned income strategies (Sagawa & Segal, 2000; Young & Salamon, 2002; Yunus, 2006). In these two perspectives, the creation of social value is the crystallisation point of what defines social entrepreneurship and a social enterprise (Hart et al., 2010), whether it is through commercial activities (not necessarily innovative) or novelty (not necessarily spread through business-like processes). Social value creation can essentially be seen as “the basic and long-standing needs of a society” (Certo & Miller, 2008, 267). We elaborate further on social value in Chapter 4.

Later approaches, and the one we choose to pursue here, have been to provide a set of criteria for what can be considered a social enterprise. That is the approach adopted by the EMES network in Europe, which proposes a set of indicators in three areas (social, economic and governance) rather than a concise definition (Defourny & Nyssens, 2012). EMES’ criteria are laid out as followed:

- a. Economic and entrepreneurial dimensions:
 - i. Continuous activity producing goods and / or services: social enterprises should have a focus (not necessarily exclusive) on productive activity.
 - ii. Significant level of economic risk: social enterprises’ economic and financial risk are borne by its owners, which is a key differentiation against public institutions.
 - iii. Minimum amount of paid work: social enterprises can combine monetary and non-monetary resources, as well as voluntary work, but should command a minimum threshold of paid labour.
- b. Social dimensions:
 - i. Explicit aim to benefit the community: social enterprise should have a clearly stated and acted upon social purpose.
 - ii. Initiative launched by a group of citizens or civil society organisations: social enterprises should maintain a sense of community-based endeavour that often characterises the goal they strive to achieve.
 - iii. Limited profit distribution: As a mean to affirm the primacy of their social goal, social enterprises should avoid profit maximisation although not excluding reasonable profit redistribution.
- c. Governance dimensions:

- i. High degree of autonomy: The independence expressed in this criterion is primarily towards public authorities.
- ii. Decision-making power not based on capital ownership: Social enterprises should aim at an inclusive governance where decision making is not only dictated by capital ownership.
- iii. Participatory governance involving various stakeholders affected by the activity: Social enterprises should strive to further democracy at the local level, particularly through the inclusion of a diversity of stakeholders in their governance.

This approach has been rather influential in Europe and has been followed by most legislation proposing a legal framework for defining what a social enterprise is (e.g. in France, Italy, Belgium, UK or Denmark), where different criteria are put forward in different countries (Borzaga & Galera, 2016; Triponel & Agapitova, 2016).

We acknowledge the lack of consensus around a definitive definition for what is a social enterprise. We note that views positing social enterprises as “hybrid organisations” (i.e. in between charity and business, see e.g. Battilana & Lee, 2014; Haveman & Rao, 2006) with business activities as a mean to sustain their operation (see e.g. Battilana & Dorado, 2010; Battilana, Sengul, Pache, & Model, 2015) tend to become more commonly accepted, without a necessary innovation criteria (see e.g. Austin, Stevenson, & Wei-Skillern, 2006; Mair & Martí, 2006). We subscribe to this position as all of our case studies have a substantial degree of commercial activity. Further, we choose to position our understanding of social enterprise based on the framework provided by EMES, which enjoys a political and legislative legitimacy in several European states. We also stress the variability inherent to this scaffold, with criteria that may or may not be accepted by different stakeholders or within different national contexts. For instance, criteria attached to governance, profit distribution or asset use are typically very differently observed across places and organisations.

1.1.1.2 Work integration social enterprises: a near-consensual type of social enterprises

Our focus on work integration social enterprises (WISEs) restricts the scope of our research and shields us from some of these definition issues. Davister et al. (2004, p. 3) propose one of the most detailed definition WISEs: “WISEs are autonomous economic entities whose main

objective is the professional integration – within the WISE itself or in mainstream enterprises – of people experiencing serious difficulties in the labour market. This integration is achieved through productive activity and tailored follow-up, or through training to qualify the workers.”

WISEs are therefore a type of social enterprise whose social mission primarily focuses on putting individuals with serious employment difficulties (back) into the working population. These individuals form a WISE’s target group, or **beneficiaries**. Often long-term unemployed, beneficiaries can vary from one WISE to the other, which can aim at specific target groups such as people suffering from a particular handicap (e.g. All Ears Marketing, focusing on deaf people in Denmark) or just cater to a maximum number of individuals in need of their support (e.g. Groupe La Varappe in France).

Focusing on the well-framed WISE concept offers several advantages:

1. It has a relatively consensual definition which allows to mitigate some of the difficulties arising from the lack of broadly agreed-upon definitions for the term “social enterprise”;
2. It is considered a representative subset of social enterprises in general and as a population coherent enough to build solid empirical analysis (Nyssens, 2006);
3. WISEs’ focus on employment, enablement and social integration allows to frame the social impact (defined as the sway an organisation’s activities manage to imprint in others’ life, cf. Chapter 4) they create around specific outcomes (e.g. job retention, skills development, social inclusion);
4. It is a widespread enough concept to be used across countries and by international organizations such as the United Nations (UN) or the Organisation for Economic Cooperation and Development (OECD).

While the definition of WISEs we retain from Davister et al. suggests a focus on “professional integration”, the authors identify four modes of integration in WISEs that we group in the three categories described below.

Transitional occupation: The WISE provides a temporary work experience to its beneficiaries under the form of direct employment (a transitional job) or on-the-job training (traineeship), with the aim to (re)habilitate individuals' capability to sustain a formal job in the labour market, out of the WISE. The essence of the transitional occupation integration is the limited term of the contract proposed to the beneficiaries. This type of integration is found in all of our case studies, which provide professional experiences to most of their beneficiaries in a (re)habilitation capacity, with no formal objective of having them as long-term employees

Creation of permanent jobs: The WISE creates sustainable jobs on the basis of commercial activities, either exclusively or with the help of subsidies. Positions are filled by beneficiaries with the objective to maintain them in the organisation, at market terms, as long as possible (employment contracts typically have an unlimited term). In this configuration, public subsidies may be used to compensate a lack of productivity of the hired workers, either throughout the employment process, or during early training or employment stages to then taper off. Davister et al. identify (2004) continuously subsidised job as a fourth type of integration, but we find it difficult to make this difference in practice. For instance, one of our Danish case studies follows this model of permanent jobs, while mixing subsidies profiles (with some staff benefiting from no subsidies at all).

Socialisation through a productive activity: Here the primary goal is not professional integration (although it is not excluded), but the (re)socialisation (i.e. the development of *soft skills*) of the beneficiaries through work activities. The WISEs seeks to bring a structured life-style, having them respecting ground rules (such as showing up on time, respecting their hierarchy or observing basic hygiene duties) in order to make them more capable of socialising and eventually sustain a job. In our research, we find such activities in the FR1's Integration workshops and construction sites⁴.

These modes of integration are all considered "**employment services**" by WISEs' public stakeholders. The above classification offers a charting map to better understand the nature

⁴Ateliers et chantiers d'insertion, or subsidies sheltered workplaces in France, cf. Chapter 7.

and the purpose of WISEs' activities. They also provide us with the fundamentals to comprehend and characterise the nature of their social impact, focused on providing their beneficiaries with the means to sustain a stable professional activity. Equipped with this understanding, we make a first observation that WISEs contribute to provide **remedies to market failure**. That is to say, they cater to individuals typically not well served by either the job market or established welfare services. We see later in our thesis (Chapter 2 and Chapter 4) that remedies to market failure can be considered a form of **public value**.

This observation brings us to our next section, where we show the role that SEs in general and WISEs in particular endorse in the political agenda.

1.1.2 The relevance of WISEs in the political agenda

Why do WISEs matter in today's labour policies? We provide two perspectives to answer this question. The first one builds upon the usual association that is made between the development of WISEs and the implementation of **New Public Management (NPM)** reforms in the 1980s and onward. From this perspective, WISEs are a tool that inject market mechanisms in the delivery of welfare services, with an implicit aim of delivering these services in a more efficient manner (i.e. presumably saving money and achieving similar or better results). The second perspective we provide is that beyond the considerations of efficiency (and whether they verify or not), WISEs are relevant to welfare systems because they propose remedies to **market failures**. We see that through Denmark and France's examples, which implement different PAM models and where WISEs enjoy a historic relevance. We end this sub-section by introducing **public value theory (PVT, Moore, 1995)** as our framework to analyse and understand how WISEs work with their public stakeholders to fulfil their mission across different PAM models.

1.1.2.1 A good tool to implement NPM-inspired reforms

WISEs have gained a considerable attention in the policy arena in the past two decades (Dees & IMF, 2012; European Commission, 2013), where they have been recommended as a key lever to focus on to alleviate particular forms of social and economic issues (European

Commission & OECD, 2013; OECD, 2010). While this is true for SEs in general, WISEs, whose mission of integration through employment translates into a potential for affecting macroeconomic indicators such as unemployment (OECD, 2013; Sibieude, 2010) have been the object of a prevalent interest.

Many academic contributions highlight how WISEs became particularly relevant in the context of the public reforms inspired by the NPM (see e.g. Defourny & Nyssens, 2010; Hulgård, 2011; A. Nicholls, 2010; Nyssens, 2006; Pestoff, 2008). NPM (Hood, 1991) is a form of PAM model that we describe at length in our Chapter 2. It is best outlined as the injection of private management-inspired mechanisms into the functioning of the public administration. Among other devices, NPM encourages the recourse to market mechanisms in order to generate presumed efficiencies in the delivery of public services. Such mechanisms can take several forms, among which are public-private partnerships (PPPs).

WISEs involved in PPPs become actors of the welfare system as they are contracted to provide services from which the state is either absent, not sufficiently present or withdrawing. Evers & Laville (2004) describe such mechanisms. Furthermore, within the NPM paradigm, WISEs can appear as a particularly appealing resource from the private sector to leverage in the public sphere for governments seeking to increase the *efficiency* of the services to deliver (Hulgård, 2011; A. Nicholls, 2010; Nyssens & Defourny, 2012; Pestoff, 2008). As private providers with which negotiations can be carried out at market terms, WISEs can be seen as an additional device to inject further market logics into public administration. Therefore, just like PPPs in general, the development of WISEs found a positive echo in the public sphere through the political reforms enthused by NPM.

1.1.2.2 A continued relevance that exceeds the usual association to NPM

We see that policy frameworks favouring the recourse to SEs, and particularly to WISEs for the delivery of employment services, has also occurred even in countries like France and Denmark, initially perceived as reluctant to implement NPM reforms (Greve et al., 2016; Hammerschmid et al., 2016). In our thesis, while conceding that PAM models are by no way clear-cut templates (we use Christensen and Laegreid's concept of public reforms layers,

2009), we associate France to the Neo-Weberian State (NWS, Pollitt and Bouckaert, 2004: a modernised administration seeking to retain the traditional state apparatus) and Denmark to the New Public Governance (NPG, Osborne, 2006: a model with an emphasis on inclusive governance). While both countries have indeed ended up implementing a degree (*layers*) of NPM-inspired policies, the development and the relevance of WISEs can also be found elsewhere, through historical and cultural considerations.

We provide the details of these considerations in Chapter 7. Here, we highlight that other drivers than PAM mechanisms can be found in these two countries for the development of WISEs. We see that the end of the of what is called “The Glorious Thirty”⁵ in France in the 1960s, together with its accompanying social transformations and a degree of welfare state withdrawal, has created a space for actors from the private sector to cater to growing social needs, among which were unemployment (Petrella & Richez-Battesti, 2016). A similar slow-down of economic expansion, combined to a recess of the traditional role of cooperatives and a knack for social experimentation has driven Denmark to a place where WISEs also became relevant providers of services, and a pertinent political object (Hulgård & Andersen, 2015; Hulgård & Bisballe, 2004).

It therefore appears that even in non (strictly speaking) NPM models like the NPG or the NWS, WISEs have been able to develop and provide **employment services** to those who needed it. Because these services are not offered by other market actors and are being phased out of public offerings, WISEs are still providing **remedies to market failure** under these circumstances and are therefore relevant welfare actors. Contributions such as Defourny & Nyssens (2013), Howlett, Kekez, & Poocharoen (2017), S. P. Osborne (2010) or Ridley-Duff & Bull (2015) all highlight a place for SEs as relevant welfare participants in PAM paradigms other than NPM.

However different PAM models call for different forms of cooperation between public stakeholders and their third-party providers (including WISEs). In the NPG, Osborne (2006)

⁵ Les trente glorieuses, the post-war period of thirty years characterized by the fast economic expansion of the Western world

highlights the importance of relational capital and thereby signals relationships of “*preferred providers*”. In the NWS, external providers are treated primarily through Weberian mechanisms: the legalism of contracts with extensive clauses and a hierarchical relationship (Gianfaldoni, 2018, speaks of “*subordinate rapports*” between French WISEs and their public stakeholders). These are two different ways of implementing collaboration models between public actors and private providers that differ from the pure market mechanisms advocated by the NPM.

1.1.2.3 Public value theory as an analytical framework to transcend PAM models

The development of WISEs is therefore anterior to the 1980s and the spread of NPM, and it endures as new PAM models emerge. One useful framework to comprehend these developments across different PAM models is **public value theory** (PVT, Moore, 1995). PVT focuses on public value (as the value created for the citizenry). Because WISEs remedy to market failures through the employment services they propose to vulnerable individuals, they can be considered creators of public value (Alford, 2009; Bryson et al., 2017). PVT is a cornerstone of the analysis we develop in our thesis (cf. Chapter 2) as it provides us with keys to answers questions such as how do French WISEs cope with the NWS’ legalist framework? Or do Danish WISEs capitalise on relational capital to deliver public value? And how WISEs **co-create** public value with their public stakeholders across PAM settings (Bryson et al., 2017; De Jong et al., 2017)?

Answering these questions is necessary to understand the underpinning of our initial questioning about performance measurement. Since whatever the PAM model and the co-production mechanisms it suggests for its external providers, none exclude the notion of performance when contracting third parties. Having clarified what WISEs are and why their role matter in different PAM contexts, it is on the notion of performance measurement that we develop our next section.

Textbox 1: A word on terminology

In this thesis, we develop about “social impact”, “social utility”⁶ “social value”, “social return” and “performance”, being “assessed”, “measured” or “evaluated”. Defining each of these terms could probably be the sole object of more than one thesis. We nevertheless try to provide as much clarity as possible around the use of these words. This textbox is meant as a shortcut to grasping an overview of our taxonomy.

We are interested in performance measurement (covered in Chapter 3) for work integration social enterprises (WISEs). We associate social impact assessment (SIA, covered in 4.1) to a form of performance measurement for WISEs (we cover the reasons motivating this parallel in 4.1). In 4.1, while admitting to lexical deviations, we endorse an interchangeable use of social impact (our customary term, cf. 4.1), social return and social value creation: social impact is the share of social value connected to a specific activity. We admit the notion of attribution or causality between impact and activities in SIA’s conceptualisation, but we reject it as a defining feature: cf. 4.2.

In 2.4, we define public value (as value generated for the citizenry), and we introduce the notion that WISEs produce public value primarily under the form of remedies to market failures. In 4.1, we pursue this reflection by highlighting a number of parallels between social value and public value, concluding that social value can be considered a subset of public value.

In Chapter 3, when developing about metrological conventions superseding quantification processes, and while acknowledging the differences between the terms, we rhetorically equate assessment (our customary term, cf. 3.1) to measurement and evaluation – we do so specifically and explicitly in the context of our research topic. In 8.1.1, as a result of field observations, we characterise the notion of social utility⁶ which encompasses social impact but is not preoccupied so much with attribution (i.e. proving causality, cf. 4.2).

As a general convention in our thesis, and in the absence of what can be characterised as SIA from a theoretical perspective among performance measurement methods used by public stakeholders, we reserve the phrasing SIA to the forms of performance measurement developed by WISEs and other mission-driven actors, while we use performance measurement as a generic term encompassing but not limited to SIA. In 3.2 we associate programme evaluation (PE) to a form of performance measurement for Public Administration and Management.

To summarise, we consider the following:

- Social impact (our customary term) = social return
- Social impact is a subset (attributed to activities) of (total) social value
- Social impact and social value are both subsets of social utility
- Social impact, social value and social utility are all subsets of public value
- Rhetorically, and for our topic: assessment (our customary term) = measurement = evaluation
- Social impact assessment (SIA) is a subset of performance measurement
- SIA is a shorthand for performance measurement initiated by WISEs
- Performance measurement is a generic term, which includes *inter alia* the measurement of WISEs’ social performance by public actors and practices of PE observed in public administration and management.

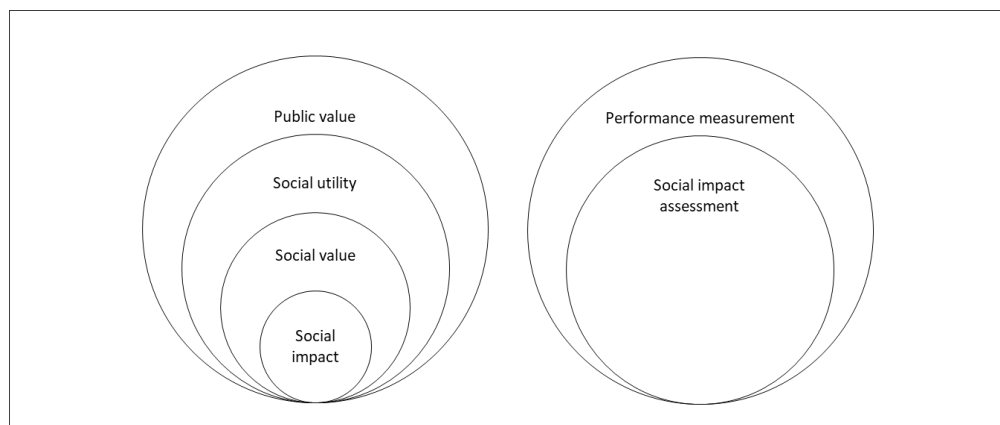


Figure 2: Key terminology

Source: Author

⁶ From the French “utilité sociale”

1.2 Performance and its measurement

In this last section of our first chapter we zoom in on our core issue: the measurement of WISEs' performance. In a first segment, we outline how vivid this question is when several reports and policy briefs were published on the topic by national governments and international organisations such as the European Commission or the OECD in the year of our writing (2018). We tie this urgency to a broader pregnancy of performance measurement practices in PAM as well as to the challenges inherent to the nature of WISEs' performance: how does one *measure* the difference made in a someone else's life?

This question brings us to our last segment, where we introduce the concept of **performativity**, or the implications that *measuring* involves on the organisations or people that are *measured*. We explain why such considerations are capital to avoid flaws in the design of performance measurement models that can lead to suboptimal outcomes (though misleading indicators) or even be detrimental to democracy (through opaque metrics framing policy development). As a conclusion we explain how these questions pertain to a broader research in economics on the topic of indicators development.

1.2.1 Performance measurement for WISEs in public administration and management

In order to understand why performance measurement is a topic of relevance for WISEs, we start by outlining why it matters in PAM in general. We show that here again, NPM-inspired reforms are the usual suspect, as they accelerated the adoption of performance measurement practices (present in European PAM since the early 1960s) in order to presumably increase efficiency and accountability. In a second part, we look at how the question of performance measurement is applied to WISEs. As we focus on their **social performance**, we introduce the practice of **social impact assessment**, a fundamental element of our thesis.

1.2.1.1 Performance measurement in public administration and management

When introducing NPM, Hood (1991) highlights performance measurement as the second doctrinal component of the paradigm, where targets for public managers are quantified. It is

on these premises that the development of NPM-inspired reforms in the last decades has driven a growing attention on programme evaluation (Dunn & Miller, 2007; Lapsley, 2009). In its quest for introducing management apparatuses inspired from the private sector into public administration, the NPM paradigm argues that evaluation (or programme evaluation, a term that in PAM, we equate to a form of performance measurement in Chapter 3) provides the information policy makers need to make better-informed choices, ultimately improving the efficiency of the public sector (Dunn & Miller, 2007; Lapsley, 2009).

The spurring of PPPs and other forms of contracting with external providers, (Broadbent and Laughlin 2003; Edwards and Shaoul 2003; Grimsey and Lewis 2007) have fostered new ways of delivering goods and services (including welfare-related services) traditionally channelled by the public sector. This trend also contributed to the development of a culture of evaluation among public authorities, facing the need to assess a growing number of third parties involved in the provision of public value.

Of course the introduction of new mechanisms for the delivery of public services did not start a trend for performance measurement. It however considerably accelerated a practice already under way: the first attempts to introduce formal evaluation mechanisms in Europe are often tied to the American Planning Programming Budgeting System (PPBS) introduced in 1965 (Greenhouse, 1966). It is the case of the French rationalisation of budgetary choices⁷, which was introduced in 1968 (Perret, 2006). Since then France has implemented many other frameworks of systematic evaluation. Chiefs among them are the creation of the Scientific Council for Evaluation in 1990, the passing of the Organic Law on Budget Legislation⁸ in 2001, the General Review of Public Policies⁹ (2007 – 2012) and the Modernisation of Public Action¹⁰ (2012 – 2017).

Denmark is not new to the game of evaluation either, even though it started later, in the late 1970s (Albæk & Rieper, 2002). Only after the general elections of 1998 the country had for

⁷ Rationalisation des Choix Budgétaires (RCB)

⁸ Loi organique relative aux lois de finance (LOLF)

⁹ Révision Générale des Politiques Publiques (RGPP)

¹⁰ Modernisation de l'Action Publique (MAP)

the first time a government that ascribed programme evaluation as part of its platform. This is not to say that evaluation did not take place in any form prior to that, quite the contrary, as Denmark has a history of forming investigative commissions on ad’hoc topics (Albæk & Rieper, 2002). It however happens through processes that, when compared to France and its approach anchored in legalism, are far less institutionalised (Furubo, Rist, & Sandahl, 2002). This kind of practice is quite characteristic of the NPG and its relationship-based approach to governance (cf. Chapter 2).

It therefore appears that performance evaluation, often carried out under the form of programme evaluation in PAM, is another feature that is rightly associated to the development of NPM, but that we observe in diverse forms in other countries applying their own “modernisation” approaches (Pollitt & Bouckaert, 2017).

1.2.1.2 WISEs do not escape the trend

In our previous section, we described the developing interest from policy makers for WISEs. In the PAM contexts we just exposed, this interest also sprung concerns for a better understanding of their performance. We show in Chapter 4 that what we call the hybrid (Battilana et al., 2015) performance nexus of WISEs is peculiar in the way that it combines the economic appraisal typically associated to the performance of a commercial organisation together with, what authors like André et al. (2018), Gianfaldoni (2018) or Lall (2017), call a form of “*social performance*”. While the former can be assessed through established metrics such as e.g. revenue, profit or return on assets, the latter poses challenges in the way it should be gauged.

Third sector¹¹ actors themselves, together with their funders, have started developing similar reflections several decades ago. These actors often tend to dub their social performance “*social impact*”, or “*social utility*” in France. In Chapter 4, we articulate more specifically what

¹¹ The third sector refers mostly to mission-driven non-profit organisations (often comprised of foundations, associations, cooperatives, mutuals and SEs). The term *third sector* is often used by trichotomy to define what belongs neither to the public sector nor to the private sector (understood here as a “for-profit” sector, since most members of the third sector are indeed private). We use the term third sector sporadically in our thesis to loosely refer to mission-driven actors in general (i.e. not limited to SEs or WISEs). A thorough and seminal definition of the third sector can be found in Pestoff (1992). Further refinement of the concept can be found in Salamon & Sokolowski (2016).

can be put behind these two notions (for social impact, we summarise it as the sway an organisation's activities manage to imprint in others' life) and how stakeholders intend to *measure* or *assess* their impact. The enthusiasm for SEs, paralleled with the development of *impact investing* practices (capital investment carried out in SE projects with the expectation to garner "*blended*" economic and social return), has led to an increasing focus on **social impact assessment** (SIA, Emerson, 2003).

This trend first emerged among private actors (philanthropists, social enterprises themselves and impact investors), eager to understand, prove and manage their impact. With the development of SIA accelerating in the past decade, a growing number of specialised organisations have emerged (e.g. B Lab, Social Asset Measurements or Sinzer), as well an ever-expanding list of methodologies for SIA (see e.g. Grieco, 2015; Olsen & Galimidi, 2008 for some of the most comprehensive overviews).

Prompted by their interest in SEs, but challenged by the measurement of their performance, public actors have also picked up this infatuation for SIA. In this context, we observe an increasing number of attempts to make SIA part of regulatory frameworks aiming to favour the third sector for the provision of certain welfare services. Initiatives attesting of this effort are, inter alia, the G8's Social Impact Investment Taskforce, the European Commission's SIA Sub-group -GECES, the 2015 policy brief from the European Commission and the OECD dedicated to the topic or the 2018 report on work integration from the French Ministry of Labour (where social impact *measurement* is a subject brought forth several times).

Just like the fostering of their development, the measurement of WISEs' performance is therefore an item that carries a certain weight in the political agenda. As public actors are tempted to take queues from the SIA model developed by third sector actors (cf. e.g. European Commission / GECES, 2014), this observation constitutes the basis of our initial questioning: **how is the development of SIA changing the way that public actors are measuring WISEs' performance?**

Before delving into the details of this question through the intricacies of our thesis, we want to conclude this chapter by enlarging our reflection to more general considerations on performance measurement in the field of economics.

1.2.2 Measuring performance: a recurrent problem in economics

This last segment of our first chapter aims at highlighting the core issues associated to performance measurement that we will delve into, and why they matter. We introduce the concept of **performativity**, which crystallises the challenges of performance measurement that we seek to contribute to. We use the notion of performativity to show that choices made around performance measurement have implications particularly in terms of outcomes (the way performance is measured contribute to shape organisations' behaviour and the outcomes they reach) and democracy (the way performance is measured drives the understanding and the tackling of political issues). In a last part, we elevate our considerations to broader considerations and explain how the reflections we introduce on performance measurement are part of a wider discussion on the design of economic indicators.

1.2.2.1 Ramifications of performance measurement

We learn from Desrosières (e.g. 1993, 2010, 2012, 2014) that to *measure* something, anything, presupposes the definition of **conventions**, i.e. assumptions about what we can summarise in two aspects (we develop further on the topic in Chapter 3):

1. What is worthy of measurement (the scope of the object or concept to be measured);
2. The metrological scale that should be applied (which system, and which value per unit).

On the basis of what we describe in the previous segment of this chapter regarding the performance nexus of social enterprises, we understand that neither of these aspects can be evidently grasped. This is why Desrosières insists on the necessity of "*public spaces*" (1992), where the conventions associated to a form of measurement can be debated, decided and made clear to anyone who is to use the information produced by the measurements stemming from these conventions.

The increased use of the phrasing performance “*measurement*” unfortunately does nothing to help this situation. This is a trend we observe very clearly in the third sector, where actors *measure* their social impact, borrowing metrical concepts from the natural sciences and providing the illusion that social phenomena can be objectively quantified. We can relate this to the notion of “*quantophrenia*”, borrowed from Sorokin (1961) in sociology by De Gaujelac (2005, 2011) to place it in a policy context closer to our area of interest. Quantophrenia designates a “*measurement pathology*”¹² (De Gaujelac, 2011, 188), where an increasing number of social phenomena *must* be *measured*. In our thesis, we do not develop further on quantophrenia. Rather, we use the similar but more developed thoughts of Desrosières on the way *measurement* takes precedence over *quantification* (cf. Chapter 3).

Our research distances itself from the use of this metrological approach to performance (cf. Chapter 3 and Chapter 4). One important downside associated to this “*replacement*” of evaluation by measurement (Beaud & Prévost, 2000) is the increased normative effect that the means applied to *measure* performance can have. With its positivist illusion of objective neutrality borrowed from the natural science, the measured performance becomes a norm to achieve. This is part of the notion of **performativity** (Jany-Catrice, 2012, Callon, 2006) that we develop at length in our thesis.

At this stage of our reflexion, we can illustrate one aspect of this concept through Desrosières’ **retroaction** (2012). Retroaction designates the phenomenon by which indicators (as a form of measurement) influence the behaviours of the organisations or people it measures, as they strive to create the conditions leading to an improvement of the measurement (rather than an improvement of the outcomes stemming from their activity). Retroaction hence illustrates how the indicators of performance measurement, as social constructs, end up shaping reality and **influence the outcomes** experienced by the actors of that reality.

Measuring performance is therefore a complex endeavour with numerous possible ramifications. This is particularly the case when talking about organisations like WISEs, whose

¹² “Pathologie de la mesure”

“social performance” is characterised by the life outcomes of their beneficiaries: intangible events that tend to occur on the long run and typically entangled with many other possible causal variables (particularly in a context of co-production). It is why “public spaces”, as proposed by Desrosières (1992) are so important: they can allow to avoid the pitfalls of misrepresentative performance metrics and / or of misinterpretation. Ogien (2010) illustrates this very accurately by quoting the so-called “Stiglitz report”¹³: “What we measure shapes what we collectively strive to pursue” (Stiglitz et al., 2010)

When concerned with dimensions of the public action (as it is with the performance of WISEs as co-creators of public value), these considerations have fundamental **democratic implications** highlighted in numerous contributions. One example can be found with Ogien (2010), who develops about the “moral value”¹⁴ that performance metrics can endorse (in terms of e.g. authority, power or legitimacy), which denatures the question of performance (replacing *how can we do better* by *can we measure this better*). Similarly, Jany-Catrice (2012) explains that quantities (through statistical indicators whose design is unclear to the masses) are used to shape political objects and visions. Another example is Salais (2010), showing that the pursuit of employment policies designed to decrease unemployment as measured through the metrics proposed by the International Labour Organisation (ILO) can be detrimental in terms of actual employment outcomes.

These thoughts lead us to our last consideration where we broaden our field of enquiry to put in perspective the issues attached to performance measurement in economics at large, where we see how the observations we just surfaced apply at a macro level.

1.2.2.2 Towards a new way of measuring performance?

More broadly speaking, the problem of performance indicators has gained prominence over the past two decades. This is particularly visible for macro indicators, which have been progressively decried to an extent never reached before. In France, Dominique Méda has contributed to reviving these questions with her 1999 book “What is wealth?”. Globally,

¹³ Report by the Commission on the Measurement of Economic Performance and Social Progress

¹⁴ « valeur sociale »

several developments like the 2007 the conference *Beyond GDP* (with *inter alia* the OECD, the European Commission and the Club of Rome), the 2007 Istanbul Declaration, the enshrinement of the Gross National Happiness (GNH) in the Bhutanese constitution in 2008 or the 2010 Commission on the Measurement of Economic Performance and Social Progress have all contributed to popularise the questions delving in the creation of **progress indicators**¹⁵. The OECD published the first iteration of its Better Life Index in 2011 (as part of an effort started in 2004). Meanwhile, questions pertaining to inequality have progressively joined this debate, gaining prominence on the basis of Thomas Piketty's work (2013). This state of affairs is what drives Gadrey and Jany-Catrice to write that "a new system of values and a new hierarchy of the 'items that count' is under way and will penetrate both political discourses and decisions" (2016, 7)¹⁶.

A topical epitome is the place of GDP, whose growth remains in most governments' platforms as a key proxy for performance. However GDP is a rather ambiguous indicator, starting with its inception: while Simon Kuznets, its "inventor", was clear it was not to be taken as a proxy for welfare ("*The welfare of a nation can, therefore, scarcely be inferred from a measurement of national income as defined above*", 1934, 7) he still presented it an approximation of it: "*Any claim to significance such a total would have would lie in its presumptive usefulness as an appraisal of the contribution of economic activity to the welfare of the country's inhabitants, present and future.*" (1934, 4).

This ambiguity is pursued in the way GDP is being implemented today: the System of National Accounts (SNA) adopted by the United Nations in 1993 explicitly states that illegal activities should be included in its calculations, citing "*the manufacture and distribution of narcotics, illegal transportation in the form of smuggling (...) and services such as prostitution*" (United Nations, 1993 p.152) as examples of activities to include in scope. The SNA is the result of a joint work from the European Commission, the IMF, the OECD and the World Bank in order to set global standards for national accounts.

¹⁵ Also called « alternative wealth indicators »

¹⁶ « Un nouveau système de valeur et une nouvelle hiérarchie des éléments 'qui comptent' vont gagner du terrain et investir le champ du discours et de la décision politique »

The latest version of the European Union's own national accounts framework (the European system of national and regional accounts 10 –ESA 10) implements and enforces the SNA's guidelines regarding illegal activities (already in scope in its previous version ESA 95), providing a standard method for estimates. In Europe, the inclusion of illegal activities in the scope of GDP following ESA 10 has for instance translated into GDP revisions of +9.5% for Cyprus and +7.6% for the Netherlands (Marianthi Dunn, Akritidis, & Biedma, 2014) or +0.7% for the United Kingdom (Abramsky & Drew, 2014) and Ireland (Brennan, 2014).

Such an implementation of the concept of National Income appears at odds with Kuznets' original vision, who was explicitly stating that *“if no criteria of social productivity are used, national income (...) would include the compensation of robbers, murderers, drug peddlers and smugglers (...). Such a judgementless estimate would be of little use.”* (1941 p.4).

This brings us back to the key issue of performativity and its implications on democracy, which Gadrey and Jany-Catrice (2016), illustrate with the large acceptance among the citizenry of GDP growth as a satisfactory horizon for public performance. The authors tie this fact to a lack of knowledge of the actual scope and arithmetic (the conventions) behind the monolithic growth rate communicated to the public. This ignorance is rendered acceptable by the positivist tradition of economics, often traced back to Léon Walras (see e.g. Walras, 1896), which presents economic instruments such as the GDP as providers of neutral and unbiased information.

This is however contradicted by Kuznets himself (as well as many others): *“The statistician who supposes that he can make a purely objective estimate of national income, not influenced by preconceptions concerning the 'facts', is deluding himself; for whenever he includes one item or excludes another he is implicitly accepting some standard of judgment, his own or that of the compiler of his data”* (1941 p.3). In our thesis, as we adopt a critical realist position (cf. Chapter 6), and through the considerations we bring forth on performativity, we reject the positivist stance.

We do not reject it on the basis that it ignores moral contemplations but on the premise that it stipulates a neutral and universal social truth expressed metrologically, epitomised by Walras as “*Only mathematics can teach us the condition of maximum utility*”¹⁷ (1926, 14). While mathematics does have a very relevant use in economics, we see the positivist claim of neutral truth (and therefore the near-impossibility of debating conventions) as pervasive for democratic values, on the basis of the arguments and examples developed above. To this extent, we emphasise the early phrasing of “*political arithmetic*” (Petty, 1691) as an interesting take on economics: at a time where economics, as a field of scientific inquiry, did not *exist yet*, this is the term that Petty (credited with being amongst the first contributors to the concept of national income) used to describe his statistical techniques, albeit his goal of rigor and soundness.

We ascribe our research to this trend of refocusing public actors on well-being and social progress rather than economic development (without necessarily antagonising the two) and exploring new ways for assessing the progress made towards these goals. The organisations at the core of our research, WISEs, encapsulate these views particularly well, leveraging economic activities to pursue a social mission.

Conclusion of Chapter 1:

In this chapter, we articulated the considerations that make our research relevant, and we have established the framework in which our work is carried out. As we explained how **work integration enterprises** (WISEs) increasingly became a partner in the implementation of labour policies, we showed the relevance of questioning the measurement of their performance from a public angle. A first question that can then come to mind is whether the use of the **social impact assessment** models we introduce in this chapter, as instruments developed by third sector actors and their funders can address the needs of public actors in terms of performance measurement. Another question can also be raised around the novelty that such models can bring to established performance measurement approaches in PAM.

¹⁷ La mathématique seule peut nous apprendre la condition du maximum d'utilité

We emphasised that although this increasing relevance of WISEs are observed across PAM models, including the **Neo-Weberian State** and the **New Public Governance**, these frameworks presuppose different approaches for the way WISEs co-create public value with their public stakeholders. Here we have an important variable for our work, as one can make the assumption that these different models can mandate different approaches to performance measurement. Under these premises, it is interesting to understand whether differences are actually observable, how can they be characterised, and whether they have an effect on WISEs' activity. For this last question, **performativity** is a key concept.

In the context of our research, performativity appears as particularly crucial because it serves as an exploratory device to understand two fundamental issues associated to performance measurement. The first one is how performance measurement influences the outcomes that organisations contribute to (and we see that outcomes are a key component of WISEs' social performance). The second issue is how through its democratic stakes, performance measurement influences the understanding and the tackling of political issues (and the first section of this chapter establishes why WISEs are relevant political issues). Together, these considerations ascribe our work to a broader research on the design of economic indicators.

Chapter 2 The shifting shapes of public administration and management

In Chapter 1, as we contextualised our research, we explained that public administration and management (PAM) models matter because they condition the way work integration social enterprises (WISEs) act with public actors to co-create public value. This chapter provides a detailed account of elements pertaining to this issue, helping to elucidate interrogations such as what are the characteristics of the French and Danish administrations? How do they differ? What do these differences mean for partners such as WISEs? Answering these questions is required before moving to considerations on WISEs' performance and understand how it can be fathomed by public actors. This chapter also serves as an introduction for the analytical framework we mobilise to answer these questions: **public value theory** (PVT, Moore, 1995).

In order to propose a holistic view of the researched phenomena, we differentiate between two main levels of analysis: macro and micro. At the macro level, we use **PAM models** as a primary analytical entry and **administrative profiles** as a secondary entry (Kuhlmann & Wollmann, 2014). This chapter provides a detailed overview of the three models we use for the former. The first one is the **New Public Management** (NPM, Hood, 1991), widely adopted and characterised essentially by the adoption, within the public sector, of management mechanisms coming from the private sector.

The second is the **Neo-Weberian State** (NWS, Pollitt & Bouckaert, 2004), which we associate to France. NWS preserves the traditional Weberian state apparatus (which we typify mostly through centralisation, legalism and bureaucracy) while modernising some of its aspects through an adoption of managerialism and a more decentralised form of democracy. The third is **New Public Governance** (NPG, Osborne, 2006), which we associate to Denmark. NPG focuses on inclusive governance, making civil society an active participant to the public life. It does so by drawing on network theory, building *“enduring inter-organisational relationships”*, which translates into an environment of high **relational capital**.

Administrative profiles provide an auxiliary framework that emphasises endogenous factors such as local culture and history, which can explain some of the variations in the adoption of global models like NPM. As we nuance the application of PAM models as monolithic analytical

devices, we also develop about reforms **layering** (Christensen & Laegreid, 2009, or the tendency of PAM reforms to superpose rather than linearly replace one another), and the interchangeable nature of policy tools (Pollitt & Bouckaert, 2017), or the fact that different policy elements, such as *“performance measurement”* can be observed in more than one PAM model).

Lastly, we introduce PVT for the analysis we carry out at the micro level. Starting with a literature review, we end with establishing an **original framework** that we use to chart the **strategic interactions** between WISEs and their public stakeholders (the aspects of their relationships pertaining to the creation of public value and where performance issues are discussed). This framework is one of our thesis' key theoretical contribution and lean on PVT's *“strategic triangle”* (the three components to align in order to create public value: legitimacy, operationalisation and public value itself), adding further analytical entries to each vertex. Doing so, we address some of PVT's criticism around its lack of clarity (see e.g. Rhodes & Wanna, 2007) while developing a charting tool to derive further insight from our empirical material than the three *“simple”* vertices would allow.

2.1 New public management

“*Public administration*” (PA) is the generic phrasing used to refer to the discipline that studies and inform the functioning of public institutions. It is still very much in use today but it contends with the notion of “*public management*” that started to be broadly used in the late 1970s. Stephen P. Osborne (2006) uses the term “*Public Administration and Management*” (PAM)¹⁸ to bridge the divide and talk indistinctively about the way public institutions are run.

PA has been formalised as a political form of organisation, governed by the rule of law where the focus is on developing and enforcing rules and guidelines (Iacovino, Barsanti, & Cinquini, 2017; S. P. Osborne, 2006; Riccucci, 2001). This bureaucratic way of functioning has attracted many criticisms, mainly focused around the lack of efficiency and an insufficient consideration for the tangible effects of policies (e.g. Chandler, 1991; Dunleavy, 1985; O’Connor, 2006; Ostrom & Ostrom, 1971). PA also has its advocates, praising its public sector ethos of working towards a greater good, and overall being more conducive to the sustainability of democracy (Lynn, 1996; Riccucci, 2001).

The notion of “public management” emerged in the 1980s as a response to PA’s critics. Epitomised by Ronald Regan’s statement during his inaugural address that “government is not the solution to our problem, government is the problem”, this zeitgeist consisted essentially in implementing oversight mechanisms typically found in the private sector (i.e. management) in order to curb PA’s bureaucracy. Public management in itself was never really theorised. Instead, “*new public management*” (NPM) was conceptualised at the end of the 1980s.

While the discipline of public administration was undergoing a deep transformation for over a decade, Christopher Hood is credited for conceptualising this phenomenon as NPM in 1991. The semantics of this name can be somewhat confusing as there is not really any “old public management”, but simply the anterior concept of public administration. In its seminal paper,

¹⁸ While Osborne used PAM throughout his seminal article about NPM, he seems to have abandoned that phrasing in subsequent publication (e.g. the introduction of Osborne 2010), referring explicitly to either PA or NPM.

Hood mentions the phrase of “old public management” only once (Hood, 1991, 4), referring to a 1989 contribution from Keating about... public administration.

Now that we have cleared this taxonomic ambiguity, let us proceed to define NPM. Put simply, NPM is the attempt to inject management mechanisms inspired by the private sector into the public sector. This usually translates into a focus on goals, efficiencies, or a market-driven provision of (some) public services, but not only. Hood encapsulates the precepts of NPM in seven doctrinal components (1991, 4):

1. *“Hands-on professional management”*, where leaders in the public administration are asked to act as managers and show accountability to avoid the diffusion of power.
2. *“Explicit standards and measures of performance”*, where goals, targets and indicators are clearly articulated and quantified. It allows to gauge efficiency and foster accountability.
3. *“Greater emphasis on output controls”*, where the focus is put on results rather than processes. It aims to avoid the formation of bureaucratic mechanisms and allows to measure performance.
4. *“Shift to disaggregation of units in the public sector”*, where decentralised management of smaller units are breaking down the traditional monolithic style of public administration. This allows for an increased flexibility and efficiency gains.
5. *“Shift to greater competition in the public sector”*, where rivalry is expected to lower costs and where the move to public tendering and term contracts is encouraged.
6. *“Stress on private-sector style of management”*, where management technics typically taught in business schools are encouraged to be deployed in order to move away from the public service ethic.
7. *“Stress on greater discipline and parsimony in resource use”*, or *“do more with less”*, where it is advocated to resist union demands and increase discipline across the strata of public institutions.

These overlapping doctrines paint a more precise picture of NPM. Furthermore, we see some clear intersections with the forces that have driven the adoption of social enterprises as providers of services typically offered by the public sector (cf. Chapter 1), such as a greater

competition or more discipline in the use of resources, that have ultimately led to the increased adoption of public private partnerships (PPPs).

Just as PA came with its critiques, so did NPM, most of them focusing on the subversion of the public service ethos, an excessive focus on efficiency and market mechanisms causing negative externalities in the provision of welfare services (e.g. Farnham & Horton, 1993; Flynn, 2007; Hood & Jackson, 1992; Metcalfe & Richards, 1990). Osborne articulates the differences between PA and NPM as followed (2006, 380):

“The strength of PA is in its exploration of the essentially political nature of PAM and of the complexities and nuances of the public policy making process. The extent to which (...) PA has been able to unpackage the differential influences upon public policy implementation has been disappointing, however. There is a tendency for implementation to be seen simply as a ‘black box’ with no apparent will to un-package the complex sub-processes of the management of the outputs of the policy process – public services themselves (...). By contrast, the strength of NPM has been in its ability to address precisely the complexities of this black box, though with an equally irritating tendency to see the public policy process as simply a ‘context’ within which the essential task of public management takes place. In its most extreme form, the NPM has even questioned the legitimacy of public policy as a context for public management, arguing that it imposes unreasonable democratic constraints onto the management and provision of public services.”

2.2 The many houses of new public management

How did NPM come to the international prominence it enjoys? And if NPM is a form of PAM hegemon, why are we interested in carrying out an international comparison? Now that we have introduced a more comprehensive definition of NPM, the two parts below shed light on how it has spread internationally through a first wave of reforms, before branching out in more country-specific models under the influence of **administrative profiles**. Doing so, we explain how PAM reforms tend to come in **layers** rather than having models linearly replacing one another and show that countries can retain a high level of PAM idiosyncrasies despite global reform waves.

2.2.1.1 The re-emergence of administrative models after decades of NPM hegemony

Hood (1991) associates the rise of NPM to four “megatrends”:

1. A will to stop government growth (in terms of spending and staffing);
2. The rise of privatization and subsidiarity in the provision of public services;
3. The development of automation;
4. The emergence of an international agenda for policy making and PAM.

Here again, we note that these drivers are particularly well aligned with the way social enterprises (and especially WISEs) have become privileged partners of public stakeholders (cf. Chapter 1).

The fourth megatrend is particularly relevant to our research, which investigates two countries. For a long time, PAM development had been idiosyncratic, budding within the boundaries of nations, ascribing itself to local legal frameworks, customs and cultural codes (Pollitt & Bouckaert, 2017). The post-war order and the creation of international fora such as the OECD or the UN and its Public Administration Network (UNPAN), as well as the emergence of multinational management consultancies, changed that paradigm and fostered the development of transnational standards for the development of reforms in PAM (Sahlin-Anderson & Engwall, 2002; Saint-Martin, 2005).

The doctrines promoted by NPM have taken full advantage of this new institutional framework to propagate globally. Building on this observation, Pollitt and Bouckaert stress

that NPM “is a term which has (rather confusingly) come to be used to cover a very wide range of reforms in an equally broad spread of countries” (2017, 6). During a first wave of reforms taking place between the late 1970s and the late 1990s, the adoption of NPM was therefore happening globally (at least in a majority of Western OECD countries) and with a rather uniform agenda (Pollitt, 1993).

As questions started to arise about governance, transparency or trust, the reforms’ tone commenced to change and different dynamics began to emerge in the late 1990s, leading to more variations in PAM reforms between countries and no real dominant model (Pollitt & Bouckaert, 2017). This last set of reforms was more impregnated with local customs and traditions, as “*administrative profiles*” started to re-emerge (cf. Table 1) but still infused with some NPM doctrines.

This led to a spurring of “*models*” identified in the literature, attached to demonstrate specificities of PAM reforms. Down at country level, governments came up with a number of models of their own, such as the “*Copernicus*” model in Belgium (De Visscher, 2005), the “*slim state*”¹⁹ in Germany (Sauer, 2000), the “*Relève*” in Canada (Bourgon, 1998) or the New Zealand Model (Boston, 1996). Academics participated in the flurry of proposed models by grouping some of the observed PAM dynamics together to design broader models such as the Digital-Era Governance (Dunleavy, Margetts, Bastow, & Tinkler, 2006), the Napoleonic model (Ongaro, 2009), the Neo-Weberian State (Pollitt & Bouckaert, 2004), the New Public Governance (S. P. Osborne, 2006), the New Public Service (Denhardt & Denhardt, 2015), the New Civic Politics (Boyte, 2011), or the Nordic model (Greve et al., 2016; Veggeland, 2007).

Parallel academic work to the development of PAM models is the identification of administrative “*traditions*” or “*administrative profiles*”, defining the basic culture, belief system and ideas of a country’s public institutions. Typical approaches consist in grouping countries by geography and PAM apparatus (PAM structures being the concrete translation of the underlying administrative tradition). Prominent work in this line of analysis include

¹⁹ Schlanker Staat

Kuhlmann & Wollmann (2014) and Painter & Peters, (2010). Table 1 draws on Kuhlmann and Wollmann to present a summary of European administrative profiles. These profiles aim to provide a more fundamental understanding of the inner workings of a country's administration and through that, they can help us understand the way new PAM paradigms can be understood and implemented differently in different countries.

Table 1: Administrative profiles in Europe

Administrative profile and group of countries²⁰	Administrative tradition	Administrative structures
Continental European Napoleonic (FR, IT, PT, GR, ES)	Rule-of-law (Rechtsstaat), legalism. Southern European sub-group: clientelism, party patronage, politicization.	Unitary-centralized, weak local government. Decentralization reforms since 1980s or 1990s.
Continental European Federal (DE, AU, CH; B)	Rule-of-law (Rechtsstaat), legalism. CH: weaker separation of state and society, weaker public service, legalism.	Federal-decentralized, strong local government.
Nordic (SE, NO, DK, FI)	Rule-of-law (Rechtsstaat) culture, transparency / contact culture, accessibility of administration for citizenship.	Unitary-decentralised, strong local government / civic self-determination
Anglo-Saxon (UK)	Public interest culture, pragmatism.	Unitary-centralised, strong local government, weakened since the 1980s.
Central Eastern European (HU, PL, CZ)	Socialist cadre administration "Stalinist" legacy, since the system change: re-establishment of	Unitary-decentralized, strong local government. Recentralisation since 2011 in HO.
South Eastern European (BG, RO)	pre-communist (rule-of-law) traditions	Unitary-centralised, weak local government

Source: adapted from Kuhlmann & Wollmann, 2014, 21

²⁰ B: Belgium, AU: Austria, BG: Bulgaria, HU: Hungary, CH: Switzerland, CZ: Czechia, DE: Germany, DK: Denmark, ES: Spain, FI: Finland, FR: France, GR: Greece, IT: Italy, NO: Norway, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, UK: United Kingdom

Both administrative traditions and PAM models are important drivers to understand and interpret field observations. Administrative profiles are more static than PAM models and are strongly focused on the public administration itself and its inner workings. For this reason, we choose to focus on PAM models such as NPM, NWS or NPG since they allow to better understand the role of stakeholders outside of the public administration (e.g. WISEs) and their focus on PAM reforms provide a better framework to comprehend the dynamics (the interactions between the different stakeholders and their evolution), that we observe in the field. While PAM models constitute our primary analytical framework, we sometime refer to administrative profiles as well in order to bring more nuance to our analysis and to avoid force-fitting reform paradigms to countries where the strict application of models may come at odds with the existing administrative culture.

2.2.1.2 Layers of PAM reforms rather than clear-cut PAM models

PAM models tend to coexist rather than being mutually exclusive (Iacovino et al., 2017). Besides, although reforms are usually associated to one PAM model or the other, the reality is that they are often implemented in strata, leading to a **layering** process of paradigms rather than a linear substitution of one by another (Christensen, 2012; Christensen & Laegreid, 2009; J. P. Olsen, 2010; S. P. Osborne, 2006). There is therefore no clear-cut implementation of for instance NWS versus NPG, but rather a set of tools that will be adopted and operationalised under the influence of the national administrative tradition or administrative profile.

To illustrate this, Pollitt and Bouckaert (2017) speak of menus and dishes (cf. Figure 3): broad models such as NPM, NWS or NPG are presented as menus, within which a set of dishes can be selected. Dishes are tools, mechanisms or ideas that can be implemented under a model. The intention of this approach is to show that dishes can be available in more than one menu, i.e. that ideas, tools or management techniques can be applied in several PAM paradigms, in similar or different ways. For instance, evaluation is rightly seen as a key feature of NPM. However it can also be implemented almost systematically in “*modernising*” NWS administrations. While blurring the lines between the different models, this approach helps refine our analysis by making it closer to the realities of PAM and PAM reforms. Besides, looking at how dishes are “ordered” in a specific PAM model (menu), has the interest of taking

into account a level of local detail such as the administrative profile that contributes to explain the observed phenomena beyond the PAM model.

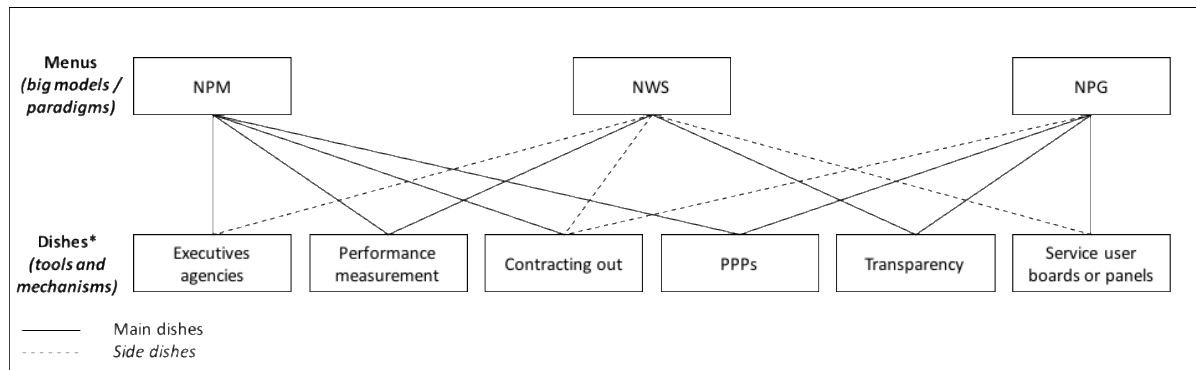


Figure 3: Menu and dishes in PAM

Source: Adapted from Pollitt and Bouckaert, 2017, 28

*Selected examples by the authors

These considerations have crucial ramifications for an international comparison like the one we are developing in our research. The first is that although the models we introduce here are intellectually compelling devices, fitting one country entirely in one of them can mislead our analysis, attributing observations in one country to the model's features rather than, for instance, some more relevant historical features of this country. On another hand, looking in details into the, social, political, economic and historical fabric of a country may make international comparison difficult: *"the problem with local detail, of course, is that, however illuminating by itself, it is just local detail. International comparison is not possible or meaningful unless some features can be identified which are sufficiently common to be compared across boundaries"* (Pollitt & Bouckaert, 2017, 26).

This is why the *"menus and dishes approach"* proposed by Pollitt and Bouckaert appears as particularly compelling to us. Although we do not resist associating the countries we are studying more with one model or another, we are wary not to take these paradigms as preceptual. The menus and dishes help us strike that balance: to derive useful analytical insight from models while providing enough awareness not to stumble into the pitfall of oversimplification derived by the systematic application of standard models. Furthermore, the *"societal approach"* (Maurice, 1989) we adopt as part of our methodological framework (cf. Chapter 6) is a useful scheme to enforce this approach.

2.3 Two models of interest: the Neo-Weberian State and the New Public Governance

We understand from the above that despite a certain hegemony enjoyed by NPM, countries still run their own PAM model. Our analysis therefore needs to capture the specificities of the French and the Danish PAMs if it is to comprehend how public actors work with WISEs. This understanding comes before even asking about performance, since we need to make sense of how these two public administrations work and how they cooperate with external providers such as WISEs in order to get to a representation of what performance can be to them.

This section therefore delves into the details of the New Public Governance (NPG) and the Neo-Weberian State (NWS), the PAM models that we attribute to Denmark and France respectively. We provide a first high level description in Table 2, alongside NPM, before diving into the details of each model in the following two sub-sections.

Table 2: Main PAM models used in our work

Model	Core claim	Key mechanisms	Selected sources
NPM	To make government more efficient and “client-responsive” by injecting business-like methods.	Market-type mechanisms, performance indicators, targets competitive contracts, quasi-market.	(Hood, 1991; Lane, 2002; D. Osborne & Gaebler, 1992; Pollitt, 2016)
NWS	To modernize the traditional state apparatus so that it becomes more professional, efficient and responsive to citizens. Business-like methods may have a subsidiary role, but the state remains a distinctive actor with its own rules, methods and culture.	Authority exercised through a disciplined hierarchy of impartial and professional officials.	(Drechsler, 2009; Lynn, 2009; Pollitt & Bouckaert, 2017)
NPG	To make government more effective and legitimate by including a wider range of social actors in both policy making and	Network of and partnerships between stakeholders. These	(Bellamy & Palumbo, 2010; Frederickson, 2006; Kaufmann, 2007; S. P. Osborne,

Model	Core claim	Key mechanisms	Selected sources
	implementation. Most governances emphasize horizontality over verticality	bring different skills and resources to address complex problems. Preferred suppliers with ongoing relationships.	2010; Pierre & Peters, 2000)

Source: adapted from Pollitt and Bouckaert, 2017

2.3.1 The New Public Governance: governance and network

NPG's genesis is attributed to Osborne (2006). The defining terms of this model, "public governance" and "governance" are extensively used in the PAM literature: e.g. as a central concept to frame institutional relationships within societies (Kooiman, 1999) or as the core component of 'self-organising inter-organisational networks' that partake in the delivery of public services (W. Kickert, 1993; Rhodes, 1997). Other contributions use this concept to explore different aspects of policy making and the inner workings of PAM (e.g. Kettl, 2006; W. J. Kickert, Klijn, & Koppenjan, 1997; Lester, 2002; Marsh & Rhodes, 1992).

When incepting the NPG, Osborne endorses Kickert and Rhodes' view (self-organising inter-organisational networks), while advocating that NPG has to push the reasoning further: *"the intention here is to carve a distinctive niche for the NPG that both has the capacity to be intellectually coherent and rigorous and has the capacity to capture the realities of PAM within the plural and pluralist complexities of PAM in the twenty-first century"* (Osborne, 2006, 381). Hence the *"new"* in new public governance: leading the concept of governance to a more central place in our understanding of PAM. By **plural**, the author means to stress the multiplicity of actors that partake in the delivery of public services (not all coming from the public sector). By pluralist, the focus is brought on the multiplicity of processes that comes into play in policy making mechanisms.

Pursuing this line of thought, Osborne writes that NPG *"focuses on inter-organisational relationships and the governance of processes, and it stresses service effectiveness and outcomes. Further, it lays emphasis on the design and evaluation of enduring inter-*

organisational relationships, where trust, relational capital and relational contracts act as the core governance mechanisms.” (2006, 384). Two essential components of NPG are highlighted here:

- First is the emphasis on **“service effectiveness and outcomes”**, as well on **“design and evaluation”**, which speaks directly to the performance of NPG and the ways to evaluate it.
- Second is the importance of **“enduring inter-organisational relationships”**, **“relational capital and relational contracts”**, which is very reminiscent of network theory.

A substantial part of our literature review is dedicated to performance and public evaluation, covering the concepts of effectiveness and outcomes extensively (cf. 3.2). At this stage though, we want to stress the double focus on **outcomes** (the long-term effects of a programme or policy – e.g. increased employability) promoted by the NPG rather than outputs (the immediate results of a programme – e.g. number of successful candidates) and on **effectiveness** (whether outputs actually turn into positive outcomes) rather than efficiency (the ratio of output to inputs). This brings the performance criteria of NPG to an entire different level from NPM, where the typical focus is on outputs and efficiency (Hood, 1991; Pollitt, 2016; Pollitt & Bouckaert, 2017).

As per network theory, Osborne himself claims NPG’s filiation to this theoretical stance. Among the gargantuan literature on network theory developed since the 1990s (see Powell, 2003 for a seminal and influential piece), a few contributions such as Agranoff, 2007; Klijn, 2005; or Milward & Provan, 2000 focus on its applications in PAM. Networks are **“typified by reciprocal patterns of communication and exchange”** (Powell, 2003, 295). In PAM, they are described as growing because of their flexibility, which is a particularly good fit for public actors facing increasing complexity. The key rationale promoting them as a superior organisation form is (in the eyes of network theorists) their flexible and adaptable properties rendering them more successful than hierarchies or markets (the two other main forms of organisation). The NPG subscribes to these views and can on this basis be perceived as a logical transition from NPM (a mixture of both hierarchies and markets). Osborne advances a

similar view, considering “*somewhat provocatively, that the NPM has actually been a transitory stage in the evolution from traditional [public administration] to what is here called the New Public Governance*” (2006, 377).

This transition is also visible in the prescribed collaboration mechanisms to manage “*NPM-era devices*” such as public-private partnerships (PPPs) under NPG (Bovaird, 2006; Teicher, Alam, & Gramberg, 2006). In this new paradigm where “reciprocal patterns of communication and exchange” are the norm, sustained trust and relational capital become key attributes of successful partnerships. It is therefore important to understand that NPG does not reject the core of “the hollow state” in the sense of Milward & Provan (2003, 2000): “*the devolution of power and decentralisation of services from central government to subnational government and, by extension, to third parties non-profit agencies and private firms - who increasingly manage programmes in the name of the state*” (Milward & Provan, 2003, 2). It however advocates for ways to manage it differently, ensuring a **co-creation** of value among a multiplicity of stakeholders. These are important considerations for our thesis, that are highly reminiscent of the account we make of public value theory below.

Building on this perspective, Greve and Hodge advance the idea that, under NPG, “*PPPs could also be seen as a genuine new way of bonding between public-sector and private-sector organizations*” (Greve & Hodge, 2010, 153). This argument is made on the basis of the spread of responsibilities (Kettl, 2011) and authority (Heinrich & Lynn, 2000) that governance mechanisms such as the ones promoted by NPG allow. It is opposed to privatisation and the PPP mechanisms seen at play in NPM where an emphasis on efficiency could be detrimental to outcomes. Greve and Hodge, who elaborate their contribution about PPPs in the infrastructure sector, highlight five challenges NPG still has to contend with when developing new partnerships. In the context of our research and our focus on partnerships with social enterprises (whether we qualify these partnerships of PPPs or not)²¹, we choose to retain three of them. We aggregate them in two dimensions and expound on them as followed:

²¹ Here we choose not to actively differentiate between PPP and outsourcing. This position is sometimes observed in the literature (e.g. Klijn & Teisman, 2005), but we acknowledge the existence of concurrent views.

1. The first challenge of PPPs under NPG comes from the increasingly complex nature of the cooperation dynamics (or “*strategic interactions*”, cf. 2.4). It stems firstly from the shared nature of the generated value (i.e. a public value as we define it further in our thesis – the value for the citizenry - and private, profitable economic value). And secondly, complexity is also increased through the multiple roles that public actors can endorse. In the case of WISEs for instance, they are often both funders and client.
2. The second challenge is that even under the more open, trust-promoting relationships favoured by the NPG paradigm, private actors may still have the temptation to act as agents of their public funder / client’s principal. Actors will therefore continue to pursue distinct strategies, even though they share the common goal of public value creation (Crosby, Bryson, & Stone, 2009)

While these observations may also ring true for other paradigms such as NPM or NWS, their relevance is heightened in NPG due to the fact that network relationships take over more formal market or hierarchical relationships, with which some of this complexity can be addressed through formal channels.

The NPG has also its critics. The most preeminent is summarised by Pollitt and Bouckaert (2017) as a lack of empirical evidence to make it a reliable and stable model: “*it can be hard to decide what is not NPG*” (126). While saying so, the authors go on and use the model as one of three main paradigms to carry out an international comparison of PAM reforms. We find that one of the shortcomings of this lack of empirical evidence is that few countries have been specifically shown as examples of states implementing NPG reforms (Osborne’s 2010 foundational book on NPG is surprisingly quiet on this aspect). This can produce challenges when attempting to define in details the boundaries of NPG. Furthermore, we understand that the most distinctive component of NPG is the emphasis brought on governance. However this focus on governance is also visible in other models (Pollitt & Hupe, 2011) such as Digital-Era Governance (Dunleavy et al., 2006).

We acknowledge this criticism and are careful not to overstretch NPG’s attributes. We centre our use of NPG on the concept of governance, which is the most accepted feature of the model. At its core, NPG’s governance is about civil society inclusion and the coordination of

multiple stakeholders from the private and third sector to achieve its goals. It does so by drawing on network theory and building “*enduring inter-organisational relationships*” from which stem trust and social capital. As such, NPG is a particularly befitting framework for the Nordic administrative profile described in Table 1, which places an emphasis on transparency, accessibility and the inclusion of the civil society in the political debate, while providing a decentralised structure to allow for governance to reach far within the social and geographic fabric of the state. In our research, we therefore assimilate Denmark as a mostly NPG country, while we stress again that we do not take models as preceptual (cf. Figure 3).

2.3.2 The Neo-Weberian State: centralisation, legalism and bureaucracy

Before diving into the Neo-Weberian State (NWS), let us first take a moment to understand how a “*Weberian State*” can be characterised. When enquiring about this question, one may think of Max Weber’s famous quote: “*a state is a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory*” (Weber, 1965, 1). Fortunately, Weber’s conception of the state goes well beyond this rather bleak citation (in Weber’s own words: “*Of course, force is certainly not the normal or the only means of the state --nobody says that--*”).

By its discourse on violence monopoly, Weber is putting an emphasis on the pregnancy of the vertical relationships of power (domination) in the state. From this stems centralization and bureaucracy that rest, in the Western world, on the rational-legal authority wielded by the state (what we refer to as **legalism**). The last key component of the state as described by Weber is a disembodiment of its institutions and their actors in a process of “*political expropriation*”, where the state becomes an intangible figure of authority, whose power is yielded by political leaders (Weber, 1965).

In sociology, political science or PAM literature, “*neo-Weberian*” is a term that is used at least since the 1970s, to designate a variety of models and methods derived from Weber’s writings, and rarely defined (Lynn, 2009). Here, we use it in the way it was coined by Pollitt and Bouckaert, in the second edition of their book *Public Management Reform* (2004). The authors summarise NWS’ core claim as followed: “*To modernize the traditional state*

apparatus so that it becomes more professional, efficient and responsive to citizens. Business-like methods may have a subsidiary role, but the state remains a distinctive actor with its own rules, methods and culture” (cf. Table 2).

Breaking down the “*neo*” from the “*Weberian*” in the sense of Pollitt and Bouckaert, we find four “*Weberian*” attributes and two “*neo*” attributes to frame and explain the NWS.

“Weberian” attributes:

1. The reaffirmation of the **central** role of the state, especially as an enabler of solutions to the new challenges faced by the citizenry and brought upon by technological change, globalisation, demography or environmental concerns.
2. The reaffirmation of the role of the (modernised) legislative law (**legalism**), aiming at upholding the principles governing the relationships between a state and its citizens, chief among which are equality before the law but also legal mechanisms to prevent excessive state actions.
3. The reaffirmation of the role of representative democracy, at all levels of the state, which is a vector of legitimation for the state and its administrations.
4. The preservation of the public service and its characteristic missions and status, oriented towards the greater good (often equated to public value, cf. Chapter 2).

“Neo” attributes:

1. A swing from an inward orientation, characterised by idiosyncratic rules and procedures, developed and followed in a vacuum (**bureaucracy**, inherited from the classical Weberian apparatus), towards a more outwards functioning, focused on meeting citizen’s needs. This can be extended by a focus on results rather than the following of procedures, by an increased use of performance management and *ex-post* evaluations and by a shifting perception of public servants from “bureaucrats” to “managers”.
2. An extension (not a replacement) of representative **democracy** through devices allowing direct consultation and representations of the citizens. This attribute is observed at various scales across the countries implementing NWS reforms and is

typically less stringent in countries presenting a Napoleonic administrative profile such as France or Italy (cf. Table 1).

These attributes clearly show how NWS adopts at its core a Weberian conception of the state as we detailed it above. We find the prominence of the law, the primacy of the central administration and the importance of the traditional state apparatus. Drechsler summarises NWS around the same three pillars: *“to keep the state as the primary framework; to use the law as the steering instrument of the framework; to not experiment with state, administration and other such important issues”* (2009, 95). In contrast, while clearly proposing a modernising framework, the *“neo”* tenets of NWS do not fit the NPM formula of competition and incentivisation. To that extent, Pollitt and Bouckaert emphasise that NWS should not be understood as *“Weber plus NPM”* (2017).

The PAM literature tells us that the NWS is observable mostly in continental Europe, where traditionally *“large states”*²² attempt to self-preserve and reinvent themselves in the face of globalization (Drechsler, 2009; Pollitt, 2009; Pollitt & Bouckaert, 2004). Relating NWS to the administrative profiles presented in Table 1, it is typically found in countries of the Continental European Napoleonic tradition, and to some extent in the Continental European Federal ones. And indeed, among the apparatus the NWS strives to maintain, is the welfare state and an *“European social model”*. Or as Pollitt puts it, in opposition to Anglo-American views of the state: *“The state is not seen principally as a burden on economy and society, or as a necessary evil. Rather it is the guarantor and partner of both a strong economy and a civilized, socially cohesive society”* (2009, 14). Among proponents of NPM, these views make states implementing NWS-type of reforms look like *“laggard or faint-hearted who have been slow to climb aboard the NPM train”* (Pollitt & Bouckaert, 2017, 121).

In our research, and while we reassert that models are not clear-cut templates (cf. Figure 3), we assimilate France, with its preponderant welfare state, a marked preference for

²² In the sense that they concentrate a large amount of national spending and are deeply woven in society's fabrics through the provision of extensive welfare, and often as direct market actors, e.g. France.

centralization²³ (although deploying decentralisation and deconcentration policies since the 1980s), a fad for legalism and a knack for bureaucracy engineering as a mostly NWS country (cf. Part 2).

²³ France has been deploying decentralisation reforms since the 1980s (cf. Chapter 7). While they produced tangible results, they failed to translate into a truly decentralised political power.

2.4 Public value theory: an underlying logic rather than a model

In order to better understand the interplay we see taking shape in the NPG and the NWS environments (i.e. the macro level) between WISEs and their public stakeholders, we adopt the theoretical framework provided by the Public Value Theory (PVT). PVT allows us to go to the micro level, beyond the overall administrative and management dynamics we identify in our PAM models. It enables us to capture the nature of the **strategic interactions** (a term we define below) between WISEs and their public stakeholders, and can offer us a peek at the **intent** behind the observed interactions (i.e. what is the agenda motivating the witnessed behaviours).

In this section, we introduce PVT more formally. We start by going through its genesis and its initial framing by Mark Moore in 1995 while highlighting some key contributions since then, as well as the ongoing challenges and critics of PVT. In a second section, we detail our own take away of two decades of PVT and the way we intend to operationalise this knowledge through PVT's main instrument, the strategic triangle. We isolate three analytical entries in each of the triangle's vertices: deliberation, legitimacy devices and endorsement in the legitimacy vertex; agency, organisational design and activities in the operationalisation vertex; and public value definition, creation and acknowledgment in the public value vertex.

2.4.1 What is public value theory

Public value, or public value theory (PVT) has been used as a concept in PAM literature and elsewhere since the late 1980s in order to shed a new light on the dynamics observed in the public sector in terms of management, strategy and overall interactions (Alford & O'Flynn, 2009). The inception of PVT as it is today commonly understood in PAM is however formally credited to Mark Moore, with the publication of *Creating public value: Strategic management in government* (Moore, 1995). Moore's book is an outcome of the *Kennedy Project*, started by the Harvard Kennedy School of Government in the mid-1980s. Moore and Khagram summarise the focus of the project as followed: "We argued that just as the goal of private managers was to create private (economic) value, the goal of government agencies was to 'create public (social) value'" (2004, 2).

In its original contribution, Moore equates public value to shareholder value in private firms, where public managers have a responsibility to increase public value, just as private managers are responsible for the increase of shareholder value in the market place. Or as Horner and Hazel put it (2005, as quoted by O'Flynn, 2007, 358): *“Think of citizens as shareholders in how their tax is spent. The value may be created through economic prosperity, social cohesion or cultural development. Ultimately, the value – such as better services, enhanced trust or social capital, or social problems diminished or avoided – is decided by the citizen. Citizens do this through the democratic process, not just through the ballot box, but through taking part in (...) consultations and surveys, for example”*.

Reflecting back on the Kennedy Project, Moore and Khagram say *“To determine what constituted public value, and to act to produce it, we developed a concept of ‘strategy in the public sector’”*. (2009, 2). This strategy was conceptualised around the *“strategic triangle”* of PAM (Moore, 1995, 22) designating the three main issues public managers have to address in order to take action (cf. Figure 4, where the arrows symbolise the alignment effort that must be made to carry out strategically relevant activities). These issues are:

1. The public value outcomes: it is the creation of public value itself and aims at elucidating what public value is and how to translate it through the provision of services, the development of infrastructure, or other intangible outcomes (e.g. the diffusion of populations' well-being, a sense of security, etc.) we develop more about this below;
2. The authorizing environment²⁴: where legitimacy and support have to be secured, required to assure a formal and acknowledged authority in the undertaking of actions leading to the creation of public value;
3. The operational capacity (technical, human, financial or others) needed (or to develop) in order to reach the creation of public value through the planned course of action.

Moore and Khagram summarise the application of the *“triangle”* as followed: *“The strategic problem for public managers thus came to be: imagine and articulate a vision of public value*

²⁴ *“Authorizing environment”* and *“legitimacy”* are often seen used interchangeably in the PVT literature. We follow this trend in our thesis but often privilege the use of *“legitimacy”*.

that can command legitimacy and support, and is operationally doable in the domain for which you have responsibility” (2014, p.9)²⁵.

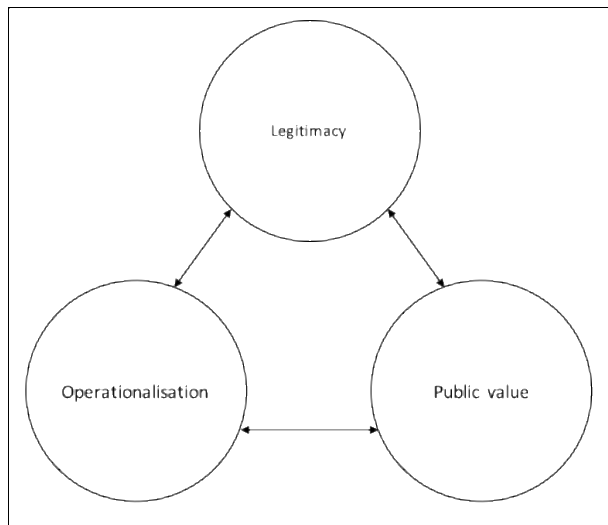


Figure 4: The PV Triangle, as traditionally represented

Source: Moore, 1995

The concept of PVT as a strategic framework so introduced by Moore was welcomed by practitioners (e.g. Coats, 2006; Department of Labour, 2006) who have designed policy frameworks inspired by Moore’s model. PVT also garnered a sustained interest from the academic community. For instance, Bryson et al. (2014), Alford and Hughes (2008) or O’Flynn, 2007 have argued PVT is the next logical step of PAM after NPM (particularly through its focus on collective preferences rather than performance). Others have pushed for the idea of **enlarging the use of the triangle** to stakeholders involved in the creation of public value, other than public managers (e.g. Benington 2011, Bryson, Crosby, & Bloomberg, 2015b; Bryson, Sancino, Benington, & Sørensen, 2017), which is just what we do in our thesis with WISEs. Some like Stoker (2006), Benington (2015) or Bryson et al. (2017) have lauded the democratic virtues of PVT. Further, Alford and O’Flynn (2009) and Hartley et al. (2017) have clarified some of PVT’s concepts and propose new fields of investigation. Lastly, Moore himself continuously built upon its initial proposal in the past two decades (e.g. Benington & Moore, 2010; Geuijen, Moore, Cederquist, Ronning, & van Twist, 2017; Moore & Khagram, 2004; Moore, 2007, 2014) along the lines we just discussed.

²⁵ We develop further about the strategic triangle in section Part 12.4.2, once we have introduced more ideas in the present section.

This flurry of contributions has also generated a few parallel (and often overlapping) understandings of PVT. Authors such as Alford and O’Flynn (2009) or Hartley et al. (2017) offer a good account of these different streams. In contemporary PAM, Hartley et al. (2017) identify “*at least three*” concurrent uses of the terms public value, which can sometimes lead to competing views. We summarise them below and add a fourth one: public values, as framed by Bozeman (2007).

Public value as a contribution to the public sphere:

This conception is primarily attributed to Benington (2011), where public value is seen as a construct capturing the benefits of the citizenry garnered from Habermas’ public sphere (a place where social problems are identified and discussed, 1962). This view therefore perceives public value as a concept reaching beyond the notions of general interests or public good (cf. below). By drawing on Habermas’ work, Benington places public value as a mechanic of democracy: the public sphere is “*where public values are explored and contested and public value created*” (Hartley et al., 2017, 672). Benington summarises this idea by arguing that public value is a “*contested democratic practice*” (2015, 29).

This brings us to the concept of **collective preferences**, which is a core component of PVT. Collective preferences are what is deemed valuable by the citizenry (Alford, 2002; Kelly, Mulgan, & Muers, 2002; Moore, 1995; O’Flynn, 2007). In the words of O’Flynn: “*Public value creation is said to rely on the politically-mediated expression of collectively determined preferences (...). This directly contrasts with the idea that individual preferences can be aggregated to reflect what it is that the ‘public’ wants from government, as has been the tendency in the NPM paradigm*” (2007, 360).

Collective preferences, as well as Benington’s proposal of contested democratic practices go against a conception of a predetermined public value, or one simply reduced to the answer to the question “*what does the public want*”. To that point, we can add Alford and O’Flynn’s words: “*public value is not an absolute standard. Rather (...), it is relative to circumstances, in the ‘task environment’*” (2009, 176). By stressing on how public value is a product of

democracy, Benington suggests that public value has to be discussed, debated and tentatively agreed upon among an array of stakeholders from multiple sectors.

Public value as that which is created through the activities of public organisations and their managers:

This is the original framing by Moore, as an attempt to propose a bottom line to the public service. Equating the value generated by public services to that of private companies' dividends has not been without its critics. Most notably Dahl and Soss, (2014) criticised the democratic implications of public value being in the hand of unelected public managers. However we see in the above that democracy still holds a central place in this approach, where public value must be assessed and discussed.

At its core, this approach is about accounting for the value "*pertinent to social outcomes*" (Hartley et al., 2017, 673), a slant that Moore develops further in his 2014 paper *Public Value Accounting: Establishing the Philosophical Basis*. Even with this contribution however, Moore does not really address what counts as valuable, and even more plainly, what is value, apart from perhaps normative assumptions as per what an ideal society could be.

A conceptualisation of PVT focused on the strategic triangle:

This approach focuses on the operationalisation of the framework provided by Moore. Although presented by Hartley et al. as "*arguably [...] not 'proper' academic theory, rather a pragmatic, heuristic tool enabling public managers to 'get things done' in a strategically smart and practically feasible fashion*" (2017, 674), this slant can generate valuable insight in case studies. This is the idea promoted by De Jong et al., 2017, who argue that the triangle presents the quality of bringing politics to the forefront of PAM, and provide an unmatched granularity in the understanding of strategic behaviours when used properly in case studies.

Public values:

Finally, another aspect of public value worth noting, but that does not build specifically on Moore's initial proposal is the *public values* introduced by Bozeman with his 2007 book *Public Values and Public Interest: Counterbalancing economic individualism*. Bozeman contrast his

view of public *values* (plural) to Moore's public *value* (singular) by anchoring Moore's contribution in strategic management where he is interested in what can be qualified of society's subjective norms. Bozeman defines public values as followed: "*Society's public values are those providing normative consensus about (a) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; (b) the obligation of citizens to society, the state, and one another; and (c) the principles on which governments and policies should be based*" (Bozeman, 2007, 13).

While not building directly on Moore's framework, if we follow Benington's understanding of public value exposed above (i. e. contributing to the public sphere), we can stipulate that Bozeman's public values participate in the identification and in the creation of public value in the public sphere. Since as we already quoted Hartely et al., the public sphere is the space "*where public values are explored and contested and public value created*" (2017, 672). This places public values at the epicentre of the debates, deliberations and discussions supposed to reveal public value.

Having these somewhat concurrent, but also complementary conceptions of public value has hampered the development of PVT to a certain extent (Hartley et al. 2017). It has also offered an additional front for critics of public value. This is for instance the case of Rhodes and Wanna (2007), who are very critical of the lack of clarity and purpose associated to PVT, emphasising that it is unclear whether public value is "a paradigm, a concept, a model, a heuristic device, or even a story ... [As a result,] it is all things to all people" (2007, quoted by Bryson et al., 2014, 449). Critics are not limited to this aspect of PVT. Dahl and Soss (2014, already mentioned above) argue that the focus placed on managers (often unelected) goes against the fundamentals of democratic representations, while Jacobs (2014) promotes the idea that the game of influence in politics dilutes public value and may make managers act against the public interest. We can also mention Power's argument (1999) that PVT, by the precedence it lends to public managers and their strategic action, runs the risk to increase technocracy in state structures. Lastly, and this is more an observation from PVT's supporters than a critic from its detractors, the lack of empirical research is often deplored (Bryson, Crosby, & Bloomberg, 2015a; Hartley et al., 2017; Williams & Shearer, 2011).

2.4.2 Operationalising public value theory

2.4.2.1 An analytical framework to understand strategic interactions between actors

Having introduced all these considerations on PVT, the question for us now is, what to do with them and how we operationalise them in our research. We start by making note of the critics presented above, and particularly of the argument brought about by Rhodes and Wanna (2007) regarding the lack of clarity for what exactly is PVT. This idea is also brought forth by the proponents of PVT (Alford & O’Flynn, 2009; Bryson et al., 2014; Hartley et al., 2017). These critics and observations encourage us to take a stand that is inspired by our own understanding of Moore’s original work: as a strategic framework, PVT does not provide enough substance to be characterised as a paradigm or a “*Big Idea*” (i.e. a model) in the sense that NPM, NWS or NPG can. We therefore distance ourselves from the idea of a *public value management paradigm* such as the one proposed by Stoker (2006) or the conception of PVT taking over as a post-NPM paradigm in the UK or New Zealand (Laegreid, 2017).

We consequently take PVT as a strategic framework more than a model (in the sense of being an overarching idea that translates into identifiable, transboundary patterns and models), following authors such as Hartley et al. (2017) or De Jong et al. (2017). We see that it provides great insight when considered in its initial framing as a strategy device for managers (rather than an overarching paradigm). We particularly see PVT as an acute lens to analyse and understand some of the development we see in NWS and NPG. The strategic behaviour suggested by PVT can for instance provide a justification for the preservation of aging state mechanisms (perceived as custodians of public value) in the NWS or for the inclusive nature of the political debate (as a mean to enhance legitimacy and support through deliberation) in the NPG.

Besides, while the Project Kennedy and Moore’s initial work were concerned primarily with public stakeholders in the application of the PVT, the strategy framework they propose is also very well suited to the analysis of private organisations like WISEs, which cooperate closely with public institutions and contribute actively to the creation of public value. The use of PVT to understand the functioning of such organisations or of the broader private sector in its

contribution to society is for instance suggested by Benington (2011), Bryson et al. (2017), Fisher (2014), or Hartley et al. (2017).

In Bryson et al.'s words: *“while [PVT] places public managers in a proactive and productive role, this tends to downplay the roles of other actors (e.g. elected officials, **social entrepreneurs**, community activists and business leaders). Here we argue that a general theory of public value production must assume (...) that in democratic societies public value entrepreneurs, whoever they may be and in whatever sector they operate, face the same tasks of obtaining authorization and legitimation, building organizational capacity, and discerning, defining and [creating public value].”* (2017, 641-642). Moore himself also recommends to include private actors in the framework he proposes for public value accounting (Benington & Moore, 2010; Moore, 2014), as well as applying the strategic triangle to private corporations (Moore & Khagram, 2004).

While we do not use PVT as we use NWS and NPG to understand **macro** level dynamics in the countries we research, we capitalise on the insight provided by the so-called PVT triangle to inform our analysis of the *“**strategic interactions**”* we observe in our field research (i.e. at the **micro** level). We develop a PVT-informed understanding of the term *“strategic interactions”*, by which we mean all aspects of the relationship developed over time between a WISE and its public stakeholders which pertains to the (co-)creation of **public value**. This encompasses how this relationship materialises (e.g. the nature of agreements passed, the nature of the dialogue, its frequency or the topics covered), what characterises its dynamics (in terms of e.g. bonds, power or agenda) and what is mutually extracted from it (e.g. resource access or value creation). In our research, we focus on the strategic interactions that can contribute to elucidate issues pertaining to performance measurement. We do not have a preconceived map of what they can or should be but, we identify them as we go throughout our data (cf. Chapter 6).

PVT is therefore the tool we use at the micro level, to understand the rapports between actors, the shifting balances of power and the evolving institutional environment. For instance, PVT can help us understand why some WISEs are more interested in performance

measurement to increase their legitimacy towards public stakeholders (thereby securing more public resources in their operational vertex) while other would rather use it in their operational vertex as a management tool to increase the public value they generate.

2.4.2.2 A framework for extending PVT's strategic triangle

On the basis of the considerations exposed above, and out of our literature review, we propose an original framework to build an understanding of our research objects informed by PVT. This framework is summarised in Table 3. Since we are concerned with WISEs and the way they operate with their public stakeholders, we orientated the design of the proposed framework towards these preoccupations. We nevertheless believe this scaffold to be transferrable to other PVT issues, particularly the ones concerned with the role of non-public actors.

In Table 3, we break down the main elements observed in the literature about each vertex of the triangle into key "*themes*" and "*questioning*" and "*analytical entries*". Themes are the broad topics in which can be arranged most of the ideas presented under a given vertex. Questioning is the main type of interrogations each vertex attempts to answer. Analytical entries are the items we retain to investigate each of the triangle's vertex (three entries per vertex). In addition to our own judgement, the main sources informing this table are: (Alford & O'Flynn, 2009; Benington & Moore, 2010; Bryson et al., 2015b, 2017; Hartley et al., 2017; Moore, 1995, 2014; Moore & Khagram, 2004; O'Flynn, 2007). Other contributions such as (Alford & Hughes, 2008; Benington, 2011; Bozeman, 2007; Bryson et al., 2014; Fisher, 2014; Jacobs, 2014; Kalambokidis, 2014; Rhodes & Wanna, 2007; Stoker, 2006) are also influencing our synthesis.

Table 3: Breaking down the triangle

	Public value (PV)	Legitimacy	Operationalisation
Main themes	Collective preferences, Collective choice, co-creation, deliberation, evaluation	Discourse, influence, political mandates, democracy	Competition, cooperation, adaptation, processes, activities, resources
General Questioning	<ul style="list-style-type: none"> - What is the PV of the programme - What are the democratic mechanisms behind the framing of PV - Is PV consensual or contested - How can PV be increased 	<ul style="list-style-type: none"> - What is the source of legitimacy for (stakeholders and institutions) - What are the influence strategies - What is the level of acceptance and / or support 	<ul style="list-style-type: none"> - How are resources secured - Are resources used purposefully and sustainably - Are stakeholders competing and or cooperating for resources
Key analytical entries	<ul style="list-style-type: none"> - Public value (co-definition - Public value (co-creation - Public value recognition 	<ul style="list-style-type: none"> - Deliberation processes - Devices - Endorsement 	<ul style="list-style-type: none"> - Agency - Organisational design - Activities

Source: Author

The analytical entries form the foundation of our PVT analysis. We describe them below, and doing so, we provide a more detailed explanation on the thought process gone into the design of Table 3. Overall, the analysis below is very much informed by the first conception of public value we exposed in 2.4.1 (Benington, 2011), resting on collective preferences (Alford, 2002; Kelly et al., 2002; O’Flynn, 2007) and contributing to the public sphere (Habermas, 1962). Beyond this, Bryson et al. 2015b and Bryson et al. 2017 are instrumental contributions in guiding our reflection, through the way they dissect the triangle. From this basis, we focus on utilising the triangle to scrutinize the strategic interactions between WISEs and their public stakeholders, particularly in terms of performance measurement.

Legitimacy / authorising environment:

In this vertex, we strive to understand how stakeholders build their legitimacy (i.e. in the sense of PVT's authorising environment). We see three main means legitimacy can be acquired through, both internally and externally:

- Deliberation processes: occurring when stakeholders engage in a dialogue, sometimes as part of a democratic process, sometimes simply as part of a negotiation (e.g. PPPs). This is where influence through discourse and narratives is generated, in the imperative of convincing (garnering legitimacy from) the other party. It can be an important part of how WISEs garner legitimacy in the absence of formal evaluation mechanisms (or legitimacy devices, cf. below).
- What we call "*legitimacy devices*": formal or informal mechanisms aiming to assess, understand, evaluate or measure (*ex-ante*, *interim* or *ex-post*) what happens in the operationalisation and / or public value vertices. Devices are leveraged to carry out advocacy work, thereby *proving* legitimacy (e.g. an evaluation shows that a programme reached its outcome, therefore the use of resources was legitimate). Social impact assessment can be a good example of legitimacy device.
- What we call "*endorsement*": received from an authoritative, legitimate figure or stakeholder, or an institution. Endorsement can come from governments, businesses, third sector actors or even a specific group of citizens. Election mechanisms and political mandates are a form of endorsement.

Public value

This is the most challenging piece of the triangle: "*The triangle is fairly silent on the kinds of practices necessary to produce public value*" (Byson et al. 2017, 642). It is also rather vague (to say the least) as to even provide a good gauge for what is public value. In our research, we consider this is fair practice as we already put forward Alford and O'Flynn's view: "*public value is not an absolute standard. Rather, (...) it is relative to circumstances, in the 'task environment'*" (2009, 176).

Having said that, we still attempt to further refine our approach for the public value vertex. While we go along with Moore's idea of a "*bottom line for of the public sector*", we also

endorse a broader definition of public value, such as the one proposed by Alford and O’Flynn (2009), where public value includes:

1. Public goods: jointly consumed goods which are (to a certain extent) non-excludable and indivisible (e.g. public lighting or public education, Shafritz, Krane, & Wright, 1998);
2. **Remedies to market failures** (other than public goods): where market mechanisms fail to maximise citizen’s behaviours (e.g. through imperfect information, negative externalities or monopolies, (Bator, 1958; Zeckhauser & Stokey, 1978);
3. Institutional devices, *“which enable markets to operate and societal orderings to function, such as the rule of law, maintenance of order, and mechanisms for the protection of property rights and enforcement of contracts”* (Alford & O’Flynn, 2009, 175).

Through the multi-stakeholder turn that has taken PVT in the last decade, Social enterprises (and by extension WISEs) have been credited as creators of public value (Bryson et al., 2017; De Jong et al., 2017). This is something we build further upon when introducing social impact assessment more in details in Chapter 4. For now however, we would like to refer to the three dimensions of public value described above and specify that we see WISEs primarily as providers of remedies to market failures (while other form of social enterprises, and also some WISEs, can perfectly produce public value that could be qualified as public good or institutional devices).

Market failures were first theorised in their modern understanding by Bator (1958) where market mechanisms fail to deliver a Pareto-efficient solution for the distribution of welfare outcomes (i.e. an *ideal* situation where anyone’s welfare cannot be improved without making someone else’s welfare worse off). Market failures are producing **net welfare loss** (Medema, 2004) and can be associated to a variety of factors such as information asymmetry (e.g. the information withheld by management in an equity market, Dierkens, 1991) or externalities (e.g. the cost of pollution not being factored in most economic activities, Smulders & Gradus, 1996). In our case, WISEs remedy to market failures as they increase the welfare (i.e.

decreasing net welfare loss) of individuals who are left out of the job market and (in most cases) cannot be efficiently sustained by public welfare services.

Considering the intangible nature of the elements described above, and having recognised the idea of a public value relative to circumstances, we therefore endorse the idea of a “*co-defined*” public value, through deliberative and democratic processes. This sends us back to collective preferences and the great benefit of public value in proposing to surface these preferences through deliberation. This is an important feature of PVT since we know from social choice theory that aggregating individual preferences do not yield any meaningful results in terms of what a collective preference could be. The so-called Condorcet Paradox showing that collective preferences can be cyclic when individual preferences are not (i.e. that under the right conditions, collective preference can be interpreted as virtually any of the expressed individual preferences) and Arrow’s impossibility theorem showing that under specific “fairness” conditions, collective preferences cannot emerge from individual preferences, are excellent examples (Arrow, 1951).

With this in mind, the three entries we choose in the public value vertex are the following:

- The **definition** of public value: which refers to all the “*co-definition*” processes described above of concertation, debate and deliberation that ultimately should lead to the emergence of collective preferences. In this entry, we want to pay attention to whether these processes actually take place within WISEs and with their stakeholders, and if so understand the way they are carried out.
- The **creation** of public value: where we look at how public value is presumably created by a WISE activity, and assumedly co-created through its network of stakeholders. While it is difficult for us to positively assess whether value is created or not, here we want to verify whether the premises for co-creation are present in the settings we observe and identify their core mechanics.
- The **recognition** of public value, which from a performance measurement perspective, is our main entry. The evaluation, assessment or measurement of public value is a key topic we wish to understand, whether it is actually happening, through which means, and whether stakeholders consider their social impact is public value. This entry also

brings us on the terrain of acknowledgement (are WISEs acknowledged for the value they contribute to create, and do they feel acknowledged) and its corollary of credibility, which feeds back into the legitimacy vertex.

Operationalisation

At the heart of operationalisation's questions are resources. They are mostly human (skills and competences) and financial (as stocks of capital or as flows of expenses and revenue streams) but can also be more intangible (e.g. intellectual property or, increasingly, data). They can be exclusive to a stakeholder (a seldom situation, which would typically require a particularly strong legitimacy) or (most likely) several stakeholders have to compete for resources or collaborate to secure them. Our analysis of operationalisation is therefore centred on resources and the way they are shared, leveraged and optimised in a PVT context using the three following analytical entries:

- Agency: the cooperation or competition strategies that are deployed to secure resources both internally and externally. In our case, we stipulate that WISEs are competing for public revenue streams, and that performance measurement can be used as a mean to secure these resources, through an increased legitimacy (i.e. as a legitimacy device) or to fulfil a form of a contractual obligation. It is also through agency that government actors and WISEs attempt to co-produce public value.
- Organisational design: the creation and articulation of processes to optimise the use of resources and to ensure their acquisition. In this space, innovation and co-creation can be important vectors of success. A typical example of organisational design would be a WISE setting up different legal entities to tap into specific beneficiary pools or resources in an attempt to fulfil their mission and generate public value.
- Activities: what the stakeholders do, and especially how resources are prioritised across the organisation to reach defined ends. Here, entrepreneurship is an important vector that can create new activities and contribute to the generation of new resources. The competition for resources is also internal, where activities have to be prioritised to the detriment of other doings. For instance, the literature on social impact assessment shows that carrying out impact assessment (SIA) can represent a

substantial toll on WISEs, which typically are small organisations. There is therefore a clear opportunity cost associated to SIA and performance measurement in general.

Before closing this section on PVT, and now that we have clarified how we intend to operationalise Moore's framework, one last question that remains to answer is what to place at the *centre* of the triangle: "*different kinds of actors may join up in the centre of the triangle in a shared effort to strategically lead and manage public value creation*" (Bryson et al. 2017, 642). While Moore initially put only the public manager at the centre, this question becomes evident when considering the multi-stakeholder approach of PVT we adopt, where complex inter-organisational cooperation can be understood in contributions such as Bovaird et al. (2016), Huxham & Vangen (2013), Pestoff et al. (2006), Pestoff et al. (2013).

On top of stakeholders, Bryson et al. (2017) propose to include other elements in the centre of the triangle, such as arenas or spheres of action, practices, functions and public problems and challenges, suggesting the use of more abstract concepts than "*simply*" actors. In this regard, we focus our analysis (i.e. we place at the centre of our triangle) on WISEs, their main public stakeholders (i.e. the Direccte in France and municipalities in Denmark) and the issues pertaining to the measurement of their performance (whether it is at the initiative of the WISEs themselves or coming from their public stakeholders).

This choice allows us to study how the strategic interactions between WISEs and their public stakeholders play into the way their performance is understood and measured. (We elaborate further on the topic in Chapter 5). We also specify that although we interpret and analyse strategic interactions as bi-directional (i.e. going both from the WISEs towards their public stakeholder and *vice-versa*), when researching the intents behind the observed interactions (in Chapter 10), we report primarily from the WISEs' point of view, as they are our main topic of interest in this research.

Conclusion of Chapter 2:

How can we look at the question of WISEs' performance if we do not understand the premises upon which this performance should be achieved? As our research focuses on understanding

how WISEs' performance can be measured by their public stakeholders, we established that we first need to understand how they work together with them towards the co-creation of public value. This chapter has brought us the tools for this.

We presented the frameworks that we mobilise to understand the relationship between WISEs and their public actors. The **strategic interactions** is the concept we introduce to characterise this relationship at the micro level. They are informed by PVT (Moore, 1995) and focus on aspects pertaining to the creation of public value. More broadly, PVT, its strategic triangle and the nine entries we present for our original analytical framework are used to parse our data at the micro level. For the macro level we introduced PAM models as our primary analytical framework: we use the **NWS** for France (Pollitt & Bouckaert, 2004) and the **NPG** for Denmark (Osborne, 2006). We complement this level of analysis with Kuhlmann and Wollmann's **administrative profiles** (2014), for more granularity when required.

On the basis of these frameworks, a first step of our analysis will consist in confirming whether the traits we lend to each model (macro) are observed in the two countries we research: e.g. do French WISEs must indeed contend with **bureaucracy, centralisation** and **legalism**? Are Danish WISEs able to capitalise on forms of **relational capital**? A second step will be to chart how these traits manifest through the strategic interactions (micro), e.g. how do French WISEs build upon NWS's features to maintain their legitimacy? How do Danish WISEs co-define public value through NPG's open governance?

As we answer these questions, we will refine our understanding of what matters in terms of performance (i.e. which aspects of performance WISEs and their public stakeholders prioritise, how do they translate them into indicators and how do they communicate about them). The next two chapters present the tools that we will use to answer questions of this nature.

Chapter 3 Performance: its measurement and its regimes

This is the first of two chapters that we dedicate to performance measurement. Before diving into the specifics of WISEs' performance in Chapter 4, we focus this section of our thesis on more general considerations about performance and its measurement in public administration and management (PAM). Just as we introduced the scaffold that allows us to make sense of WISEs' environment in Chapter 2, the purpose here is to introduce the key “*universal*” dimensions (i.e. regardless of our focus on WISEs) that structure the analytical framework we put together for our study of performance.

We start with the cornerstone concept of **performativity**. Performativity designates the effect that indicators can have on the phenomenon they aim to quantify. In our first chapter, we explained that performativity matters because if not thought out, it can lead to the creation of indicators detrimental to outcomes (through misguided metrics, see e.g. Salais, 2010 on the use of unemployment statistics) and can impede democracy (through opaque conventions, see e.g. Desrosières, 2008, on the place of statistics in governance).

As it can be a daunting concept to apprehend, our thesis introduces an original framework to unpack performativity into three components, which we present in this chapter:

1. The **socio-technical arrangements** (STAs) introduced by Callon (2006): performativity's matrix, or the mutual interactions entertained between performance concepts, their users and the environment in which they are developed and operationalised.
2. **Retroactions**, as identified by Desrosières (2012): the effect of performance indicators on economic agents' anticipations and behaviour.
3. What we dub the “*indicator-substitution*” issue: the fact that indicators, engineered through opaque conventions, tend to replace social realities, and its implications on democracy.

We then complement our analytical framework by identifying, within the literature (e.g. Pollitt & Bouckaert, 2017 or Rossi et al., 2003), seven performance criteria used in programme evaluation, which allow us to clarify some of the taxonomy about performance (e.g. what is

the difference between outputs and outcomes?) and provide us with a first understanding of how public actors may be tempted to gauge WISEs' performance (e.g. in terms of economy, of efficiency or effectiveness?). Lastly, we introduce Jany-Catrice's four performance regimes (2012) as the last component of our analytical framework. We use it to plot the broader characteristics of the performance measurement models we encounter in the field (e.g. how complex are they, and how much deliberation was involved in their design).

3.1 Performance and performativity

What exactly is performativity? How does it occur? Why does it matter? This first section brings answers to these questions by introducing the three components we identify as the core of performativity: socio-technical arrangements (STAs), retroaction and indicator-substitution. To explore these matters, we start with a definition of performance, where the polysemy we encounter only serves as an introduction to the complexity that underpins our topic of interest. While our second sub-section delves into the details of performativity, the third one exemplifies its three aspects through a metric that is right onto WISEs' topic: the measurement of unemployment.

3.1.1 Defining performance

Attempting to frame the concept of performance, an obvious starting point would be etymology. The word "performance" is a substantive derived from the verb "to perform", itself borrowed from the old French "parformer", which can mean "to execute", "to accomplish" or "to perfect"²⁶. There is therefore a polysemy attached to the word "performance" that can be split essentially between the idea of achieving or completing something, and the idea of reaching perfection. The contemporary meaning of the word "performance" captures both these impressions, as it is used to mostly express:

- the capability of something (e.g. in terms of benchmark, with the idea to determine which option is best, or nearing perfection by current standards). It is an assessment, that can be done either *ex-ante* or *ex-post*.
- The accomplishment of a task (e.g. performing an assignment). This is understood as an action (i.e. *interim*).
- A sports or arts performance: arguably a mix of the two previous ideas: accomplishing an act out of the ordinary, in an attempt to reach perfection.

This dual meaning is also observed by Jany-Catrice (2012). On this basis, she further highlights that an increasing ambiguity stems from the use of the term "performance" in the sense that it can both designate the results and the processes (and their improvement) leading to these results. There is therefore a "mix-up" between the measurements or the indicators we create

²⁶ From the Frédéric Godefroy Dictionary of old French, as digitalised by Hitoshi Ogurisu

to get a sense of a phenomenon's performance, and the actual processes and components of the observed phenomenon, which can be infinitely complex: *"We remark since at least a decade or two, a tendency among indicators to embody concepts and operationalise them in a way that these concepts themselves may be unable to achieve"*²⁷ (2012, 24).

An example of this can be found in the discussion on GDP we offer in Chapter 1. Over the decades, GDP and its growth has become an implicit proxy for overall nations' *"performance"* (i.e. beyond pure economic performance) while some nations with a growing GDP are failing on aspects that can also be considered as performance, such as a citizen inequality, well-being and welfare, or even public finance's sustainability. The idea, introduced by Jany-Catrice's quote, and that we illustrate here with GDP as a performance indicator, can be summarised as the issue of indicators' **performativity**.

As we progress in our characterisation of performance, we have already made the shift from performance to indicators. This is a swing that occurs naturally throughout the literature on the topic, as indicators, whatever their nature, are the primary means by which we express performance (Desrosières, 2014; Gadrey & Jany-Catrice, 2016; Jany-Catrice, 2012). The following segment explores how and why the indicators we use to measure performance are vectors of performativity.

3.1.2 The three aspects of performativity

Authors like Alain Desrosières (2012), Michel Callon (2006), Vatin, (2009) or MacKenzie et al. (2007) have demonstrated the performativity of indicators, or the tendency for the indicators we create and use to shape the reality and *"take over"* the reality they are supposed to represent. Or as Callon puts it: *"How can a discourse be outside of the reality that it describes and simultaneously participate in the construction of that reality as an object by acting on it? To this paradoxical question the concept of performativity provides a convincing and general answer."* (2006, 7).

²⁷ "On note depuis une ou deux décennie au moins, une propension élevée à ce que les indicateurs incarnent les concepts et leur fixent le caractère opérationnel qui leur fait parfois défaut."

In our research, we propose an original way to unpack the notion of performativity in three core components. The following three parts are dedicated to the introduction of these components: Callon's **socio-technical arrangements** (STAs, 2006), Desrosières' **retroaction** (2015) and what we call "**indicator-substitution**" (a notion already explored in the literature but never really singled out).

3.1.2.1 Socio-technical arrangements

The concept of "**Socio-technical arrangement**"²⁸ (STA) is abundantly used by Callon (2004, 2005, 2006, 2007, 2009) to explain the interactions (and especially the entanglement) between a theoretical, conceptual or technical construct (such as performance indicators) and its environment. In *What does it mean to say that economics is performative?* (2006), Callon explains that his conception of STAs builds upon Deleuze and Guattari's ontological concept of *agencement* (1980)²⁹. *Agencement* refers to a combination of various items or elements, of the social and the physical world, that are disposed in a specific manner and interact with each other, in shifting ways. *Agencements* are self-contained mechanisms, meaning that they suffice to themselves: "there is nothing left outside *agencements*: there is no need for further explanation" (Callon, 2006, 13).

On this basis, "*socio-technical arrangements*"²⁸ describe the entanglement between statements and the facts they aim at describing in self-contained realities. When describing this, Callon focuses particularly on economics and the way economic mechanisms, as they are unearthed by economists, contribute to affect and change the very mechanisms they describe. His idea goes as followed: by gathering the knowledge brought about by economists on their own behaviour, economic actors adopt a conduct that encompasses this new information. Therefore, the economic statement is entangled with the device or behaviour it attempts to describe. This is where performativity lies according to Callon.

An example is provided by MacKenzie and Millo (2003) through the use of the Black-Scholes-Merton model in the stock market. The Black-Scholes-Merton model (Black & Scholes, 1973)

²⁸ Often cited by Callon using the French term of *agencement*: socio-technical agencement

²⁹ In *Qu'est-ce qu'un dispositif* (1989), Deleuze explains that his notion of *agencement* builds upon Foucault's "*devices*" (*dispositifs*, 1977)

provides a partial differential equation to prototype financial markets which can help deducing theoretical security prices. Mackenzie and Millo show that while initially providing a way to discover pricing in markets, the Black-Scholes-Merton model went on to modify behaviour and price patterns themselves on the basis of the information it provided: “*a formula that progressively discovers its world and a world that is put into motion by the formula describing it*” (Callon, 2006, 14).

Callon dissociates between the formation of STAs and their *performance*, or the process by which their postulate becomes entangled with reality. It is thus perfectly possible for a STAs to be **unrealised**, and for multiple arrangements to coexist until one becomes **enacted**. In our research, we therefore stress the difference between unrealised and enacted STAs, and we attach a particular importance in identifying unrealised ones (or at least not yet enacted), as they are a source of understanding for how performance is envisioned, before it becomes performative.

Furthermore, in the context of our research, we consider as STAs (typically unrealised and on a small scale) most kinds of performance measurement we come across (whether coming from public authorities or from WISEs). This stance goes along Callon’s work who extends his thoughts from economic theory, where STAs emerge from scientific statements, to the practices of “*accountants, marketers and, more generally, market professionals*” (2006, 29) whose activities can also be vectors of STAs. We argue that performance measurement frameworks, by proposing a normalised frame of indicators (i.e. social constructs) to gauge occurrences from the “*physical world*” are doing exactly what Callon describes.

Lastly, STAs is a concept that goes particularly well with the critical realist approach we adopt (cf. Chapter 6). It indeed implies that empirical reality can be actualised by the constructs emerging from its own description (Callon, 2006, MacKenzie and Mill 2003). That is to say that reality can be *changed* by the devices (such as indicators) we use to make a description of it (e.g. how the Black-Schole-Merton model modified the price patterns it initially aimed at describing). This goes hand in hand with the critical realism’s stratified ontology we explain in Chapter 6.

3.1.2.2 Retroaction

A prolific author on the topic of performativity, Desrosières (2012, 2014, 2015) develops some very pertinent views for our thesis on the increasing use of indicators in our societies, induced especially by the spreading of NPM reforms. One of them is the notion of “*retroaction*”, specifically tied to performance indicators. Retroactions is the effect that evaluation indicators have on the actors they aim at assessing (the “*quantified actors*” as Desrosières puts it). This phenomenon has cascading implications as actors may seek to anticipate on the construction and the use of the performance indicators as the programmes they are a part of and their evaluation roll out: “*Indicators’ retroaction induces adverse effects: actors are focusing on the indicator rather than the action itself*”³⁰ (Desrosières, 2012, 267). Power provides similar views in *The Audit Society* (Power, 1999), where he proposes a detailed analysis of what he dubs “*gaming and cheating*” mechanisms.

Retroactions can also induce looping mechanisms where changing behaviours on the part of the quantified actors may lead to a response from the evaluators’ side and so on (Desrosières, 2006, 2015). Retroaction therefore brings a close focus on the behaviour of economic actors in the face of evaluation against defined indicators. Although this may seem to echo the STAs introduced above, retroaction is in fact expressing the possible consequences associated to the formation of a STA.

On one hand retroaction is primarily a pervasive aspect of performativity that shifts the focus of an organisation from its core activity towards the fulfilment of “satisfactory” performance criteria. On another hand STAs propose a framework to assess and understand the relationships between a social construct (an indicator) and the environment in which it is created. Both are relevant aspects of performativity for our analysis, that we use to inform our comprehension of how indicators in performance frameworks are understood and used by their stakeholders and the kind of effect they might have on their activity and broader environment.

³⁰ “La rétroaction des indicateurs induit des effets pervers : les acteurs se focalisent sur l’indicateur, et non sur l’action elle-même”

3.1.2.3 The indicator-substitution problem, or the democratic stakes of performance measurement

In this last part dedicated to the unpacking of performativity, we elaborate about an ultimate concept, that we dub “*indicator-substitution*”, where indicators tend to overtake and oversimplify the social realities they attempt to gauge. We show that this issue is anchored in a tendency for the realist connotation of measurement to overshadow the conventions (quantification) necessary to the creation of indicators. This matter strongly influences the way indicators are used and understood by the public and thereby carry substantial democratic ramifications.

As a starting point, we take the fundamental distinction introduced by Desrosières (2012) between *quantifying* and *measuring*. *Quantifying* aims at lending a numerical existence to concepts or ideas expressed primarily through words. It consists in reaching a consensus, or at least a form of agreement on how “*things should be accounted for*”, e.g. which elements should be considered, which should be excluded, on what basis, etc. In short: the first stage is about building the **conventions**. On another hand, *measuring* implies that a realist metric can be attached to what we are attempting to gauge (e.g. a physical quantity) and consists in carrying out the measurement itself, which translates into the use of a metrological system, along the conventions that have been designed in the previous stage. Quantifying is therefore different from measuring, as it entails creating equivalence conventions, which to be accepted and understood, should be defined through a transparent process involving negotiations and compromises, leading to replicable codification procedures. Measuring should only intervene after this process, once conventions are established and understood.

Having clarified the distinction between the “quantifying phase” and the “measuring phase”, we want to stress the increased use of the phrasing “measurement” for most matters concerned with a form of evaluation (e.g. the use of *social impact measurement*). This is often applied to public action but extends to other realms such as social enterprises, as we see in Chapter 4. We understand from the above that this new taxonomy overshadows the necessary conventions that are embedded in the production of a quantifiable output. By

focusing on the realist notion of measurement, we forget about the complex network of constructs that lead to the creation of indicators. And from this stems the risk of substituting the “truth” (a complex social reality) for the indicator (a construct built on simplifying conventions). This issue constitutes what we will call from now on “*indicator-substitution*”.

Desrosières elaborates further on the topic: “*The recurrent question of knowing ‘to which extent statistics are an accurate representation of reality’ is a fallacy, contaminated by the metrological realism of natural sciences. Statistics, and more generally all forms of quantification (for instance probabilistic or stemming from bookkeeping) transform the world, by their sole existence, by their diffusion and their usage for argumentative, scientific or journalistic purposes. Once the quantification procedures coded and routinized, their outputs are reified. They tend to become ‘the truth’, in an apparently irreversible fashion. Initial conventions are forgotten, the quantified object is naturalised and the use of the verb ‘to measure’ comes automatically*”³¹ (2012, 268).

This problem sits at the core of the topic we are researching: social impact assessment, often called social impact *measurement*³². When developing about SIA in 4.1, it quickly becomes apparent that in this field too, the use of the term *measurement* is new and follows the concept of assessment. For this reason, and on the premises of the considerations developed above, we settle on using the phrase “*social impact assessment*” rather than “*social impact measurement*”. We show in 4.1 that the two are often used interchangeably in the literature.

In the same section, we also see that the use of the terms “*social impact measurement*” raises a number of issues in France, where stakeholders have been using the wording of “social utility *evaluation*”³³ for more than two decades, and where the realist implications of measurement (as opposed to evaluation) are not looked upon kindly by stakeholders who

³¹ “La question récurrente de savoir si une statistique « reflète plus ou moins bien la réalité » est un raccourci trompeur, contaminé par le réalisme métrologique des sciences de la nature. La statistique, et plus généralement toutes les formes de quantification (par exemple probabiliste, ou comptable), transforme le monde, par son existence même, par sa diffusion et ses usages argumentatifs, scientifiques, politiques ou journalistiques. Une fois les procédures de quantification codifiées et routinisées, leurs produits sont réifiés. Ils tendent à devenir « la réalité », de façon apparemment irréversible. Les conventions initiales sont oubliées, l’objet quantifié est naturalisé et l’emploi du verbe « mesurer » vient machinalement à l’esprit et sous la plume.”

³² The issues exposed by Desrosières echo the notion of “*quantophrenia*” proposed by the sociologist Sorokin in 1961 and brought forth in a context closer to our research by Vincent de Gaulejac (De Gaulejac, 2005, 2011).

³³ Evaluation de l’utilité sociale

want to convey the complexity of their activity. Furthermore, the separation between the quantification phase and the measurement phase explained above is pertinent when considering the practices advocated in SIA, where it is recommended to have stakeholders participating in the definition of the assessment framework and in the design of the indicators (i.e. the quantification phase).

The issue of measurement taking precedence over the conventions used in the quantification process is therefore important. We signal above that these conventions are expected to be the outcome of a transparent process of deliberation. This should allow to put the understanding of the proposed indicators on a common baseline among all stakeholders, and more crucially, it should foster the necessary understanding of the mechanisms used in the quantification procedure to avoid indicator-substitution. However many contributions suggest that this deliberation process rarely happens (Desrosières, 1993, 2012; Fourquet, 1980; Gadrey & Jany-Catrice, 2016; Jany-Catrice, 2012).

In Jany-Catrice's words: "*Seldom the socio-political conditions under which performance indicators are elaborated are made visible, seldom they are questioned, seldom they are debated*"³⁴ (2012, 25). This sentence can send us back to Callon's socio-technical arrangements (STAs, the self-contained devices of entangled facts and descriptions). Considering the crucial influence that enacted STAs can have in the economy (illustrated by the example of the Black-Schole-Merton model above), we understand why having visibility on their components (what Jany-Catrice call "socio-political conditions") can be of relevance to the citizenry, or at least to the stakeholders they will come to affect.

It is therefore important that a "**public space**" is created to discuss and understand the performance and the indicators we want to measure (Desrosières, 1992). There are indeed significant democratic stakes attached to the measurement of performance: on top of helping creating a common understanding of the indicators, as explained above, a deliberation processes can reinforce their legitimacy, as highlighted by Jany-Catrice (2008). The author

³⁴ "Rarement les conditions sociopolitiques de l'élaboration des indicateurs de performance ne sont rendues visibles, rarement ils sont interrogées, rarement ils sont débattues."

writes about the inclusion of different stakeholders in the development of a social health indicator at the regional level in France. She argues that this “*deliberative democracy*” is a legitimisation vector for the developed indicator and its use. Desrosières (1992) makes the argument that public fora where such conversations could take place were more common between the 1950s and the late 1970s, and questions about the indicators of public performance could be discussed more openly. However public reforms and new dogma have changed the dynamics in which was baked the development and the understanding of national accounts. And the NPM is explicitly pointed as responsible for these changes by authors like Jany-Catrice (2012) and Desrosières (2012).

3.1.3 Performativity in action: the measurement of unemployment

To illustrate the three aspects of performativity that we develop in the previous segment of our thesis, and drawing from some of the authors we cited (particularly Jany-Catrice, 2012; Gadrey & Jany-Catrice, 2016; Salais, 2010), we take the example of the measurement of **unemployment**: an indicator typically used at a macro level that nevertheless has a close connection to WISEs’ activities and the nature of their impact (OECD, 2013; Sibieude, 2010).

At the boundary between social and economic matters, measuring unemployment is a popular mean of gauging policy performance (whether rightly or wrongly). In most countries, the conventional definition of unemployment is based upon the delineation offered by the International Labour Organisation (ILO). Statistics using this definition of unemployment are created by the means of statistical surveys. That is to say, they are not based on the actual number of people registered at job centres (this is another indicator). According to the ILO’s definition, is considered unemployed a person who is:

- without a job,
- actively seeking a job,
- immediately available to work

This is a stringent definition that has us acknowledging the importance of conventions (i.e. the quantifying phase). When the statistical survey is carried out (the measurement phase), anyone who has worked just one hour during the reference period (usually a week) is not

considered unemployed. Likewise, someone who gave up looking for employment will then be excluded from the labour force and not be accounted as unemployed.

With this in mind, a declining unemployment rate, while suggesting a reality where more people are at work (and are presumably able to sustain themselves from their work) may in fact conceal more precarious situations. This has become a growing issue in the past decade, with for instance the use of zero-hour contracts in the UK, or the growing number of mini jobs in Germany. Likewise, surveys suggest an increasing amount of discouraged unemployed who “vanish” from the unemployment statistics (by exiting the labour force, see for instance (Desilver, 2017; “State Analysis: How Many Unemployed Have ‘Given Up’?,” 2017). The social “reality” depicted by the unemployment statistics can therefore be deceptive.

Let us clarify though, that we are not arguing against the stringency of the proposed convention, which indeed, by design, always have to “draw a line” somewhere. In the case of unemployment statistics, that is truly what they measure: unemployment, as rigorously poised by the ILO. And having worked for one hour will effectively take anyone out of that statistic. By having this discussion, we however want to stress the risk that poses, in the collective mind, the fact of taking the unemployment metric for a neutral representation of a socio-economic situation (**indicator-substitution**): the chosen indicator (and the conventions used for its quantification) becomes a social reality in the eyes of the populace, no matter how close or far the empirical social situation is from that measurement. Looking at this from a PAM perspective, this risk is augmented by the effects of possible political recuperation (and associated **retroaction**) of this indicator, as highlighted by Jany-Catrice (2012) or Salais (2010).

This issue is not ignored by the institutions producing the indicators. Many countries run concurrent series of statistically surveyed unemployed population (according to the ILO’s definition) and the number of people registered under unemployment schemes. Pôle Emploi³⁵ in France and the Job Centres in Denmark run their own indicators, in parallel to the one tracked by INSEE in France and Danmarks Statistik on the basis of the ILO’s manual for

³⁵ The French Job Centres

the survey of economically active population. The series ran by job centres are typically higher than the ones obtained through surveys, because registration criteria to a job centre are typically less stringent than the ones framing ILO's statistical definition of unemployment. A person seeking a job can for instance be registered as unemployed even though they already have a (inadequate) job. As such, the work done by the ACDC³⁶ network in 2006 in France shows that 4.45 million people were registered at the Job centres³⁷ while 1.17 million were accounted in INSEE's unemployment series.

In 1966, the ILO also introduced an “*underemployment*” category in its manual, where underemployment is described as “*when a person's employment is inadequate, in relation to specified norms or alternative employment, account being taken of his occupational skill (training and working experience)*” (Husmanns, Mehran, & Varmā, 1990, 121). The ILO distinguishes between “*visible underemployment*” and “*invisible underemployment*”. The former reflects an insufficient number of hours worked (i.e. it takes interest in the quantity of work) while the latter is concerned with the underutilisation of skills, low productivity or low incomes (i.e. it takes interest in the quality of work).

In 1982 however, the ILO signalled a sole focus on the tracking of visible underemployment: “*for operational reasons the statistical measurement of underemployment may be limited to visible underemployment*” (Husmanns et al., 1990, 121). France's INSEE tracks underemployment in its quarterly unemployment publications (underemployment concerned 5.6% of the employed in Q4 2017 and unemployment was 8.9%, INSEE, 2018) and Danmarks Statistik only tracks its annual publication on immigrants, where the metric is broken per ethnic groups and sex: in 2015, 12% of western men immigrants were underemployed and 34% of non-western men immigrants were underemployed (for women, these numbers were respectively 21% and 41%, Elmeskov & Bang, 2017).

The introduction of new ILO categories is an example of what **socio-technical arrangements** can produce: as growing concerns emerged about the way unemployment was gauged

³⁶ Les autres chiffres du chômage: The other unemployment figures

³⁷ Agence Nationale Pour l'Emploi (ANPE), former Pôle Emploi – the French job centres

around the world on the basis of the ILO's definition, ways to account for "*alternative*" forms of social realities were developed. The creation of such new indicators goes hand in hand with novel policy frameworks being developed to cater to these newly identified "*issues*" as they gain pre-eminence and legitimacy (see for instance Livingstone, 2018).

The measurement of unemployment offers one of many examples for how performance measurement is often operationalised by indicators that offer undeniable qualities but also some serious limitations induced by performativity. Our point here is not to systematically denigrate the use of indicators, which by design will always be limited to their conventions. It is to show and understand the mechanisms by which they become (sometimes insidiously) legitimate actors and shapers of social reality.

An interesting counterpoint to that is the development of "*progress indicators*" (see e.g. OECD, 2004; UNEP, 2012; Hamilton & Naikal, 2014; Gadrey & Jany-Catrice, 2016), which have been developed specifically with the intent to challenge *enacted* STAs in order to re-centre policies on topics deemed of interest (e.g. favouring the use of the Human Development Index rather than GDP growth to gauge development issues). This can form an example of what Callon calls "*performance's struggle*" (2006, 25), as several STAs coexist. Some of these indicators, such as the ones focused on life outcomes promoted by the OECD are even used to feed social impact assessment frameworks. In Appendix 6, we propose a selection of development indicators that propose methods to quantify the type of outcomes WISEs may find desirable (e.g. measurement of happiness, social health, etc.)

Having introduced and exemplified the original framework we propose to apprehend the complex notion of performativity; the next section will focus on the more tangible questions of performance measurement approaches in a PAM setting.

3.2 Performance in public administration and management

As we are zeroing in on the core of our research topic (WISEs' performance from a PAM standpoint), this section moves from the abstract dimensions of performativity we just introduced to more tangible aspects of performance measurement in PAM. This is a step required in our analysis in order for us to understand how performance is typically framed by public actors. The elements we introduce here come as a complement to the characterisation of the Danish and French PAM environment we introduced in Chapter 2 to answer questions such as how are public programmes evaluated? what type of criteria are used? Or what do these criteria encompass?

In PAM, the performance of a programme³⁸ is often assessed in the light of a practice called “programme evaluation” (see e.g. Fitzpatrick, Sanders, & Worthen, 2017; Heckman, LaLonde, & Smith, 1999; Pollitt & Bouckaert, 2017; Rossi, Lipsey, & Freeman, 2003; Stufflebeam, 2001), and some authors equate “performance measurement” to practices of “programme evaluation” (Gadrey & Jany-Catrice, 2016; Jany-Catrice, 2012; Ogien, 2010). We thus elect to use programme evaluation³⁹ as a starting point for our reflection on performance in a PAM context.

On the basis of the information gathered about programme evaluation and our own research in PAM, we then propose a framework for performance in PAM. In this scheme, we build on existing contributions (particularly Pollitt & Bouckaert, 2017) and show how different elements of a programme are typically conceptualised when attempting to assess performance. We also introduce a set of seven performance criteria in PAM: relevance, economy, efficiency, effectiveness, cost-effectiveness, utility and sustainability.

³⁸ In this section, as we are concerned with a view on performance from a PAM perspective, we call “programme” any PAM object that may fall under evaluation: a policy, a specific programme or an organisation (the latest including WISEs).

³⁹We adopt the conventions we observed in a majority of the literature on programme evaluation and use the terms “programme evaluation”, “evaluation” and “evaluation research” interchangeably.

3.2.1 Programme evaluation

As highlighted by Pollitt (1993, p.353), *“There are many slightly varying definitions of evaluation⁴⁰”*. Scriven (1991, p.139) provides a concise definition: *“judging the worth or merit of something or the product of the process”*. The aspect of **“merit or worth”** appears central in several other definitions (e.g. Guskey, 1999, p.41; Stufflebeam, 2001, p.11) and comes from the Joint Committee on Standards for Educational Evaluation (JCSEE)’s Program Evaluation Standards (Yarbrough, Shulha, Hopson, & Caruthers, 2010). The following definition proposes a relatively good encapsulation of the multifaceted aspects of the discipline: *“Programme evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions.”* (Rossi, Lipsey, & Freeman 2003, p.16).

Programme evaluation’s field of application is virtually limitless, and it is subject to broad discussions covering its theoretical aspects as well as its technical aspects. The topical literature is therefore very extensive in its spectrum, from considerations in econometrics and statistics (from the fundamental work of Rubin in 1973 to the recent developments reported by Imbens & Wooldridge in 2009, including distinctive contributions such as Heckman, LaLonde, & Smith in 1999), to works on evaluation approaches and philosophy (Furubo et al., 2002; Scriven, 1991; Stufflebeam, 2001). This leads to a very wide diversity of evaluation outlooks, stressed by Rossi et al. (2004, p.26): *“Evaluation practitioners are drawn from a wide range of academic disciplines and professions with different orientations and methods, and this mix has contributed significantly to the multiplicity of perspectives”*.

As far as methodology goes, program evaluation has been relatively static since its early developments. Levitan and Wurzburg (1979, p.9) thereby quote William Morrill asserting that the discipline is in its *“pre-Copernican”* stage. This statement is mainly motivated by the progressive and incremental improvements brought to the field since its inception and the

⁴⁰ We adopt the convention we observed in a majority of the literature and use the terms “programme evaluation”, “evaluation” and “evaluation research” interchangeably.

absence of conceptual breakthrough (Appendix 2 provides a short history of the development of programme evaluation).

A common dichotomy when classifying approaches to programme evaluation is the separation of studies proposing an economic angle and the ones adopting a social lens. While it is true that some studies may focus on either of these aspects, the two are in fact complementary of an exhaustive programme evaluation (Poister et al., 1981). A holistic evaluation should indeed both investigate the effectiveness (the social aspects, looking at outcomes) and the efficiency (looking at the use of resources, often the economic angle) of a programme (Borus, 1979).

There is a number of existing contributions cataloguing different approaches to evaluation. Most notably Fitzpatrick et al., 2017; Furubo et al., 2002; Guba, 1990; House, 1983; Madaus et al., 2012; Rossi et al., 2003; R. E. Stake, 1974; Stufflebeam, 2001 and Yarbrough et al., 2010. Our work mostly draws from Stufflebeam's work, which encompasses a large majority of the research on the topic and provides one of the most comprehensive monographs. Stufflebeam catalogues what he identified as 22 evaluation approaches, grouped into four categories, using ten descriptors (the items characterising an approach). The tables presented in Appendix 3 and Appendix 4 summarise Stufflebeam's 90-page monograph.

This review of programme evaluation has three main applications for our thesis:

1. When hatching a framework for SIA in Chapter 4, we partly capitalise on Stufflebeam's classification as presented in Appendices 3 and 4.
2. We use the sub-items listed in the Appendix 3 to inform our coding grid and bring structure to our field data (cf. Appendix 3)
3. We use contributions such as Rossi et al. (2003) or Scriven (1991) to retain that evaluation consists in essentially judging the "merit or worth" of a programme, and to inform our general understanding of performance measurement in the public sector, such as we develop it in the following segment of our thesis.

3.2.2 A framework for performance in in public administration and management

Looking at what NPM (as the paradigm that contributed to the generalisation of performance measurement in PAM) has to say about performance, Hood's third doctrine tells us that for NPM performance translates into a "*Greater emphasis on output controls*". While NPM is indeed characterised by a stringent focus on outputs (Alford & Hughes, 2008; S. P. Osborne, 2006; Pollitt, 2016; Pollitt & Bouckaert, 2017), a recurrent point made in the PAM and programme evaluation literature we just reviewed is the difference between outputs and outcomes and how the latter matters most in attaining a programme's objective (see e.g. Boyne et al., 2003; Rossi et al., 2003, Pollitt & Bouckaert, 2017).

This is the kind of nuance in the understanding of PAM performance that we provide in this section. Figure 5 summarises our view on the topic. Its design and its interpretation are informed by programme evaluation practices as we have just reviewed them and more general concepts of performance borrowed from PAM. Figure 5 provides a short but somewhat painful answer to the question of what is performance in PAM, in the sense that it is clearly multi-dimensional and rather difficult to "*measure*". We nevertheless take it as supplementary building block informing our comprehension of performance. In this representation, we draw the main components coming into play when considering the performance of a PAM object: the evaluated programme, policy or organisation itself (the grey box⁴¹) and its social, economic and political environment (the items surrounding the grey box). As we detail the main constituents of this representation below, and in order to make our descriptions more relatable, we take the example of Lambda, a hypothetical work integration programme that provide training and mentorship to former offenders in order to help them secure a job.

⁴¹ As a side not, we emphasise the resemblance between Figure 5 (particularly in the grey box) as a rather orthodox representation of PAM performance and SIA's impact value chain (cf. Figure 8) we present in Part 1Chapter 4, stressing the similarities in the conceptualisation of both SIA and programme evaluation (also highlighted in Part 1Chapter 4).

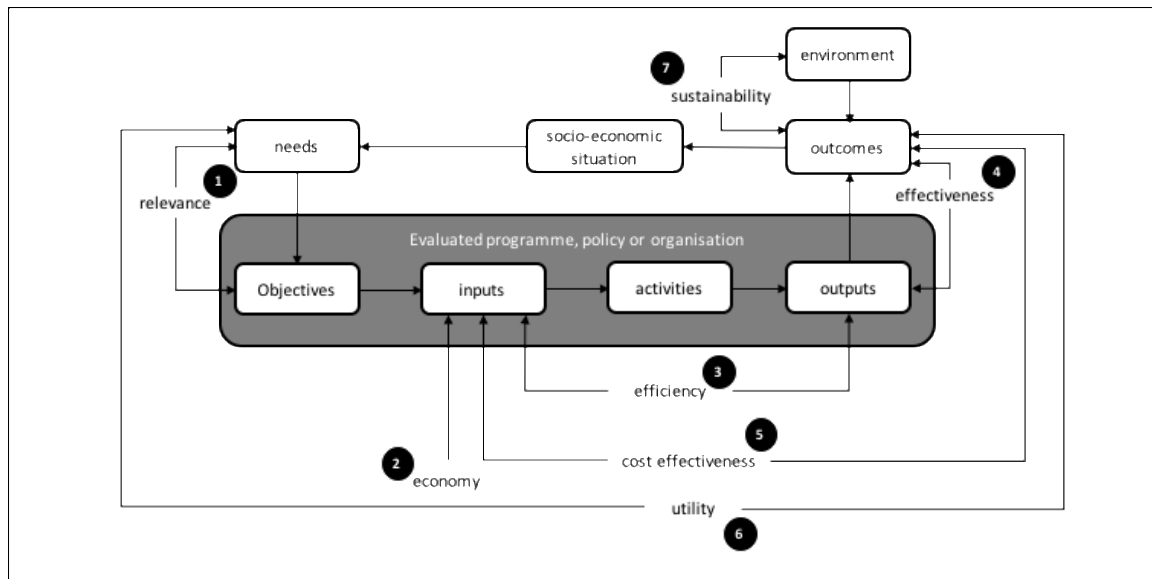


Figure 5: Performance in PAM

Source: Author, adapted from Pollitt and Bouckaert 2017, 15, 136. Similar representations are used throughout the literature on PAM performance and programme evaluation (see e.g. Boyne et al., 2003; Rossi et al., 2003)

In the grey box, the first considered item is the objectives of the evaluated programme. They are in most cases formulated by its managers or its political sponsors. In our example, it is to bring a specific population (former offenders) back into the labour force. Then come the inputs, which are all the resources (financial, material, immaterial, staff) that are necessary for Lambda to function. The actual activities of Lambda are the next item to take in considerations: what actions and processes are carried out with the identified inputs to reach the stated objectives (e.g. job training, mentoring or social skills development).

Now comes the part where most of the attention is placed when measuring the performance: **outputs** and (out of the grey box) **outcomes**. The outputs are what comes directly out of the activity. In our example, it could be the number of previous offenders that end the programme with a job offer. While outputs can provide an easy proxy for what looks like success (the offender got a job, ergo he is back into the labour force), they can also be terribly deceiving: the offender loses its job after three weeks and goes back to criminal activities. This is why outcomes are crucial in the understanding of the actual result a programme has on its target population: outcomes capture the lasting effects generated by the programme (e.g. the upgrade in professional and social skills that will actually allow to sustain a job in the long run).

Because outcomes are so intangible, they are very difficult to capture, and sometimes to even fathom. They are out of the programme's grey box because although the programme's activities have a direct influence on them, outcomes are also shaped by the environment. This is another important aspect of understanding the results of a programme: great skills and professional opportunities may be offered through Lambda, but offenders also have to be preserved from e.g. their former network of delinquents (a factor completely exogenous to the programme) in order to decrease their likelihood of relapse into criminal activities. Focusing on outcomes therefore means both accepting intangible proofs of performance and avoiding the pitfall of monocausal explanation.

The other aspect of Figure 5 we want to explore are the seven approaches to performance measurement that can be derived from such a framework, as numbered in our drawing:

Relevance: the extent to which the programme and its objectives can reliably meet the identified needs. In our example, it could be about making sure that the goals of Lambda are aligned with the local realities of the ex-offenders population (profiling) and of the local job market (job placement and training opportunities). Relevance typically consists in reviewing a programme design and what is call its "theory of change" (i.e. the series of logical statements delineating how it intends to generate positive outcomes) in order to assess its robustness and make sure it is aligned with the identified needs. It is typically evaluated in qualitative terms, *ex-ante*, although it can as well (and often should) be evaluated *ex-post*. Being focused on the alignment between objectives and needs, Relevance does not capture the externalities of a programme, which happen at the outcome stage and will therefore be better encompassed by other dimensions such as utility, sustainability and effectiveness. Surveys, questionnaires and a thorough review of the programme can typically be used to collect the necessary data to capture relevance, with reasonable requirements in terms of competences and time investment.

Economy: the financial gains leveraged by a considered option and the implementation of the solution(s) it proposes. It is a purely quantitative dimension, often an absolute value, where the assessor will seek to put a financial worth on the savings associated with the considered

intervention. In our example, one could compare Lambda to another programme for the work integration of ex-offenders and simply look at the difference in terms of costs. Economy is purely focused on inputs. At the most basic level, its evaluation is the subtraction of the resources invested in the support of the considered programme (or WISE) to the means necessary to sustain other interventions leading to similar outcomes (at best) or outputs (at least). It can be carried out as *ex-ante* or *ex-post*, depending on the aim of the assessor. Due to its focus on inputs, Economy does not encompass outcomes in its valorisation, which means no advanced cost-benefit analysis are required to assess it. Economy can therefore be evaluated without too much effort, provided the information related to inputs is clearly identified and easily accessible.

Efficiency: the ratio of outputs to input for a given programme. For Lambda, it could be the cost corresponding to the amounts of inputs required to get one ex-offender successively out of the programme (expressed as the total cost divided by the total of *successful* outputs). It is a quantitative notion, expressed in relative terms, and meant to measure mostly productivity gains occurring through the delivery of the considered intervention. Both as part of an *ex-ante* assessment or as an *ex-post* evaluation, efficiency can be compared to other existing programmes or organisations providing similar services in order to figure out which is best maximising the utilisation of the provided inputs. Efficiency can also be assessed in the course of an intervention to identify potential ways to improve processes and increase the ratio of outputs to input. As a task, assessing Efficiency can vary considerably in time and complexity, depending mostly on how easily inputs and outputs can be identified and quantified.

Effectiveness: the extent to which the WISE's outputs translate into improved outcomes. While programmes can meet their objectives (often measured in terms of outputs), they might produce adverse externalities resulting in negative outcomes. Concurrently, the transformation of a programme outputs into positive outcomes is where lies the true value added of an intervention. For Lambda, positive outcomes can be no recess into criminal activities, while a negative outcome could be a socialisation with a network of offenders. This is why it is important to assess Effectiveness and investigate the linkage between outputs and outcomes. Due to its focus on outcomes, Effectiveness evaluation is carried out *ex-post*,

sometimes with a substantial amount of time following the considered intervention. It is however possible to perform such an assessment *ex-ante*, as an exploratory study to understand the possible impacts of an intervention. In practice, outputs are usually relatively easy to measure and communicate (often in quantitative forms), however outcomes are most often more challenging to document and sometimes even to comprehend. Carrying out this kind of qualitative assessment therefore requires skilled labour and a substantial time investment.

Cost effectiveness: the extent to which inputs deliver tangible improvements in outcomes. This is a quantitative as well as a qualitative measurement, where one seeks to assess whether the resources invested in a programme translate into verifiable improvements for its target population. In our example, it could consist in looking at how much investment was required to get to long-term, sustained inclusion of ex-offenders. While the needs of the beneficiaries are part of the equation, Cost effectiveness is more about the resources part of the programme, and the extent to which they generate value-added in the considered outcomes. This is where cost-benefit analysis-based approach are very popular, and where the use of financial proxies (quantitative data) to value social outcomes (qualitative data) play an important part. This type of assessment can be done *ex-ante*, as well as *ex-post*, and even during the course of a programme in some instances, in order to provide insight for management or to help secure further funding. A key challenge in the gauging of cost-effectiveness is the problem of attribution (or causality), i.e. how much of the actual outcomes can be credited to the programme's inputs versus how much is caused by exogenous factors. Assessing Cost effectiveness is however often complex and requires skilled labour to carry out research and craft robust counterfactuals.

Utility: the extent to which the outcomes generated by the programme meet the originally expressed needs. For Lambda, it could be checking that the profiled ex-offenders do indeed keep a job and exhibit life patterns of successful social inclusion after a period of time. This dimension relates outcomes to the target population's needs, and considers how they are met through the intervention, as opposed to relevance, which relates needs to objective and is consequently not concerned with outcomes. While utility can encompass quantitative

aspects, it is mostly a qualitative dimension, where one seeks to thoroughly understand the nature of the programme's outcomes (both positives and negatives) and how they relate to the needs initially identified. In most cases, assessing utility requires a thorough investigation carried-out by qualified staff that has a genuine understanding of the challenges associated to the studied issue.

Sustainability: the extent to which the outcomes that are created are sustainable for the society and its environment. Here environment is understood holistically, including the natural environment, but also its economic and social dimensions. For Lambda, sustainability could be concerned with the environmental impact of the activities proposed to ex-offender for their rehabilitation, or simply whether the programme is financially sustainable in its proposed operational framework. Sustainability is therefore another aspect of performance covering both qualitative and quantitative aspects, with a strong emphasis on the former due to the multidimensional nature of the issues at stake, and the relatively intangible forms they can assume. This is therefore another topic that requires substantial time and skills to investigate, and is better carried out both *ex-ante* and *ex-post*.

As conclusion to our presentation of Figure 5, we stress that in a stark contrast with NPM's restricted focus on outputs, these seven criteria for performance provide an understanding of the multidimensional nature of a programme's success in a PAM context. They provide a structure for the interpretation of our field data and they contribute to inform the performance measurement framework we propose in Chapter 4.

Equipped with a more tangible understanding of what performance measurement can entail in a public context, we are now ready to move to the final section of our Chapter 3, where we introduce four broad categories of performance measurement: the performance regimes (Jany-Catrice, 2012).

3.3 The performance regimes

Having introduced a high level of conceptualisation through the notion of performativity in our first section and having moved to more tangible considerations about PAM performance criteria in a second segment, this third section capitalises on these concepts to introduce the last tenet of our analytical framework for performance: Jany-Catrice's four **performance regimes**.

Jany-Catrice (2012) identifies four broad types of performance regimes, in which she classifies different approaches to performance measurement. She uses two main delineators to distinguish her categories: the nature of the evaluation process, which can be negotiated or (centrally) planned and the nature of the evaluation itself, which can be unidimensional or multidimensional. This framework comes as a handy (if not oversimplified) synthesis for some of the consideration brought forth until now. In our analysis, we use Jany-Catrice's framework as a plane to map out the performance regimes we come across in order to gain a high-level insight of the stakes that a given performance measurement model conveys (e.g.: is it prone to performativity because of a planned design? Or can it be vector of retroaction due to its unidimensionality?).

We describe below the four performance regimes using some of the concepts introduced in the first sections of this chapter, and we represent them in Figure 6. Rather than closed boxes or distinct categories, we take these regimes as a continuum, where different measurements or evaluation methods can be placed along two axes of planification versus negotiation and of unidimensionality versus multidimensionality. We therefore take the liberty to adapt Jany-Catrice's presentation to allow for a more fluid analysis of the performance regimes we encounter.

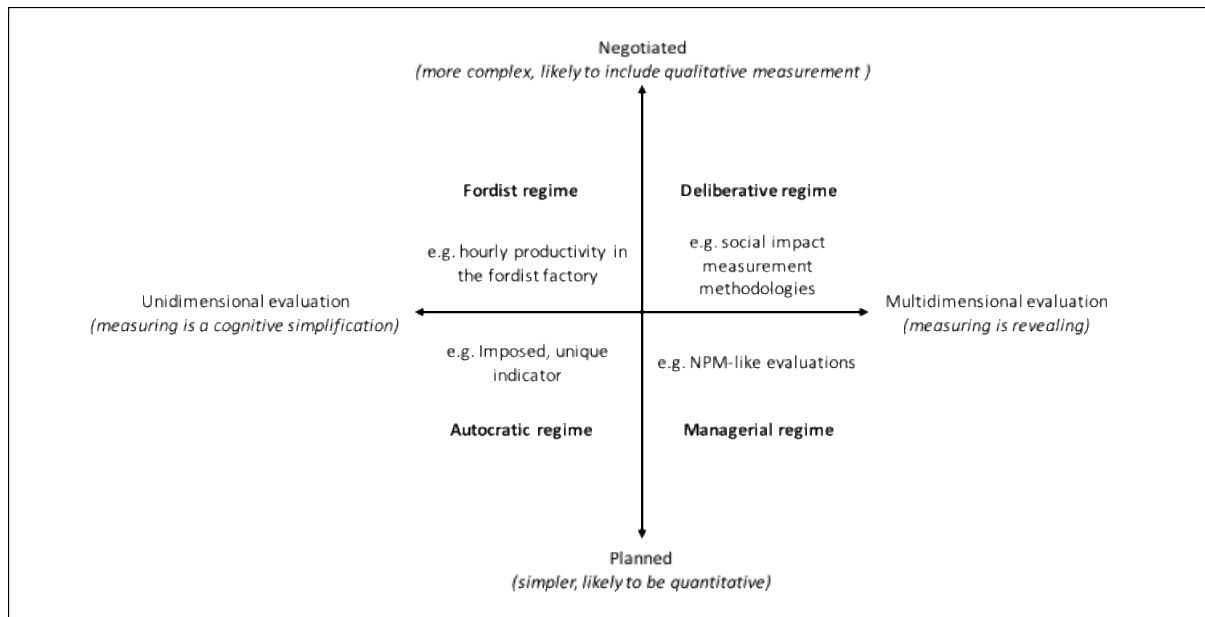


Figure 6: The four performance regimes

Source: adapted from Jany-Catrice (2012, 150)

The autocratic regime: it is characterised by a unique indicator to gauge performance, which is chosen and designed by a central authority, without a concertation of the stakeholders involved in the performance measurement processes or those concerned with the measured performance (top-down approach). An example of such a regime can be found in the performance system that was operationalised in the Soviet Union for its manufacturing industry, through the so-called “*product range*”⁴²: a list of approximately 4,000 items for which a number of units had to be reached within a given period (Mironova & Bovaird, 1999). This system gave rise to all forms of retroaction where product quality was sacrificed in order to get to the desired number of outputs, or where efficiency gains were not realised because targets were too low to justify them.

The Fordist regime: it is still resting on a unique indicator for performance measurement, but this indicator is negotiated through a deliberation between stakeholders. This inclusive process of indicator definition is one way of illustrating Desrosières’ public space and has the substantial benefit of reinforcing the acceptance and legitimacy of the developed metric. This regime is best exemplified by the performance approach bearing its name: the efficiency ratios as they were implemented in Ford’s factories in the first decades of the 20th century.

⁴² nomenklatura izdelyi

These ratios of production per hour were unidimensional in their way to apprehend performance, but they were negotiated collectively and underpinned by shared benefits associated to the productivity increase to be realised (Jany-Catrice, 2012).

The managerialist regime: it is characterised by a plurality of indicators, aiming at capturing multiple dimensions of performance (i.e. conferring an explanatory, or revealing dimension to performance measurement), but designed through a top-down process where stakeholders are not included. Jany-Catrice tells us that this is the kind of performance measurement that is typically applied by the NPM. This is indeed a shortcoming of NPM that we often find highlighted in the literature (Dunleavy et al., 2006; S. P. Osborne, 2010), which can be illustrated through the focus on outputs rather than outcomes that we have mentioned several times already. This often translates into indicators that are developed to serve policy objectives, but with a degree of disconnection from the field and the actual outcomes that are needed for a programme to genuinely be successful (Figure 5 provides a good visual queue to understand these considerations).

The deliberative regime: a regime characterised by a multiplicity of indicators to spell out different dimensions of performance, with a high degree of deliberation (bottom-up approach). This regime places a particular emphasis on stakeholders involvement, ensuring that the developed approach is well accepted, but that it also fully encompasses all aspects of performance that, in the eyes of the concerned participants, should be taken into account. Is it the regime that best allows the development of a public space as conceptualised by Desrosières (2012), where conventions can be debated and agreed upon. By way of consequence, it is also very reminiscent of the way Benington (2015) envisions public value, explored and contested in Habermas' public sphere.

A good illustration of a deliberative regime is what can be observed in most SIA practices, where stakeholders involvement in the definition of performance indicators and the processes used to carry out that measurement is core to the development of frameworks. Jany-Catrice (2012) also uses "*social utility evaluation*" (an approach observed in France that

we equate to SIA) to exemplify deliberative regimes. SIA is precisely the type of performance measurement we focus on in our next chapter.

Conclusion of Chapter 3:

This chapter has equipped us with a detailed analytical framework to analyse the nature of performance, its measurement and the implications of its measurement. Doing so, we built a precise taxonomy. To characterise performance, we established seven performance criteria identified in the literature (Pollitt & Bouckaert, 2017; Rossi et al., 2003): **relevance, economy, efficiency, effectiveness, cost-effectiveness, utility** and **sustainability**. To scrutinise its measurement and its implications we presented three core components of **performativity: socio-technical arrangements**, which when they are enacted become a source of performativity (Callon, 2006); **retroaction** (Desrosières, 2012), or how indicators influence actions and behaviours; and what we call **indicator-substitution**, or the way indicators, as social constructs, can supersede the complex realities they proxy.

With this taxonomy, we can now talk in much specific terms of the issues that our research encompasses: does a focus on economy increase the risk of retroaction? Is there indicator-substitution when performance indicators are expressed in forms of outputs rather than outcomes? What can public actors do to ensure that conventions are understood in order to prevent the enactment of STAs? As we raise these questions, we begin to see how the considerations introduced in previous chapters start to fall into place. For the last question, public value theory (Moore, 1995) and its focus on deliberation and public spaces (the places where open exchanges of ideas occur) can provide interesting hints: can WISEs and their public stakeholders find the public space they need to deliberate measurement issues and prevent performativity?

In the next chapter, we build further upon the taxonomy introduced here (particularly on the seven criteria of performance) as we go from a general conception of performance and its measurement to an application specifically for WISEs.

Chapter 4 The performance of WISEs?

This chapter is the last of our literature review and looks at the characterisation and the measurement of work integration social enterprises (WISEs)' performance. What is so specific about WISEs' performance? How do WISEs themselves measure their performance? How does social impact assessment (SIA) fit into the performance measurement of WISEs, and what are the different methods and approaches to it? These are the questions this chapter seeks to answer. On the basis of what we developed in our previous chapter about performativity and the risks it conveys, especially in terms of sub-optimal outcomes, we understand the importance of elucidating these interrogations as they are what ultimately determine whether WISEs are successful partners of labour policies.

In a first section, we talk about what characterises the performance of WISEs itself. With their focus on putting (back) people into employment, we see that their performance does not stop at outputs but must be determined by the longer term, more complex **outcomes** of their beneficiaries. This "**social performance**" (Gianfaldoni, 2018; Lall, 2017) posits some complex *measurement* issues since simple metrics cannot capture the complex social reality at stake (e.g. a beneficiary that leaves a WISE with a job can quickly stumbled back into unemployment if their soft skills are not up to par).

The practice of **social impact assessment** (SIA) is an answer to precisely these preoccupations. We make a detailed introduction of the concepts underpinning SIA, refining further our taxonomy as concepts such a social impact (the sway an organisation's activities manage to imprint in others' life) and social value (the long-standing needs of a society) are expounded. We show that the social value produced by WISEs can be equated to public value (especially under the form of **market failure remedies**), and doing so, we show that social impact assessment is considered a form of performance measurement for WISEs (Austin et al., 2006; Grieco, 2015).

We then proceed with a short history of the development of SIA. This grounds our understanding of SIA in its historical significance (partly for accountability towards external stakeholders and partly born out of policy frameworks) and lays the foundation for the SIA

review we introduce in the last section of this chapter. We identify 124 SIA models, among which 18 are particularly applicable to WISEs. We also introduce the analytical framework we use to parse these methods and the performance models we come across in our research, made of 13 variables falling into four attributes (boiled from 118 identified in the literature). These are some of the key empirical contributions of our thesis, which we will use to comprehend the performance measurement models we come across in our research.

4.1 Capturing the performance of social enterprises: enters social impact assessment

What is the performance of a WISE? How can it be conceptualised? This first section spells out the performance nexus of social enterprises in general and WISEs⁴³ in particular, as they must perform economically as well socially. On the question of its measurement, we introduce social impact measurement (SIA) and articulate the link we draw between performance measurement and SIA, where the latter can be understood as a form of performance measurement for WISEs.

Building upon these preliminary considerations, as we did for performance in general, we move on to establish a clear taxonomy of the issues concerned with WISEs' performance. Doing so, we clarify what we mean by social impact (the sway the organisation's activities manage to imprint in others' life) and we establish a link between social value and public value (where the former is a subset of the latter).

4.1.1 The performance nexus of social enterprises

As commercial entities, social enterprises and WISEs can be subject to the usual performance metrics observed in the business world, such as, *inter alia*, return on assets, revenue growth or profitability. Due to the social mission of WISEs and a wide range of them adopting non-profit legal forms, these business performance criteria can often be limited in their application (i.e. some SEs may choose to focus on maintaining an equilibrium in their statement of income rather than increasing their margins). They are nonetheless a reality of these organisations, as commercial activities are used to sustain their social mission (Bagnoli & Megali, 2011; Battilana et al., 2015; Nyssens, 2006).

As a primary goal of WISEs, this social mission is a fundamental component of their performance. WISEs are therefore presented with a peculiar performance measurement challenge where they have to account for both economic and social dimensions. To this extent, Gianfaldoni in France (2018) or Lall in the USA (2017) speak of "*social performance*".

⁴³ From now on, we will only refer to WISEs, even though most of our observations are valid for most SEs, and sometimes formulated as such in the referenced literature.

While research in economics and management, as well as current practices in the business world provides plenty of near-consensual guidance for apprehending economic performance of organisations, social performance can prove a lot more difficult to understand and assess. To this effect, social impact assessment (SIA) is often considered as a practice allowing to gauge aspects of the social performance of an organisation.

This association between performance measurement and SIA is found recurrently in the literature (see e.g. Austin et al., 2006; Grieco, 2015; Grieco, Michelini, & Iasevoli, 2015; Maas & Liket, 2011; S. Olsen & Galimidi, 2008). It stands on the double premises that organisations' social performance should be assessed in the light of the outcomes they generate and that SIA aims precisely at evaluating outcomes (Alexander, Brudney, & Yang, 2010; Liket & Maas, 2015; G. E. Mitchell, 2013). We also observe the association between performance and social impact (including social utility in France) among our case studies (both on the WISE and public sides). We subscribe to this view in our research and adopt the perception of SIA as a performance measurement instrument for WISEs. We do so with an aim to spell out the conventions (as exposed in Chapter 3) that underlie such a use of SIA and while acknowledging that WISEs' performance (and therefore its measurement) extends to other dimensions (economic, as explained above, but also environmental and social in forms that may not be fully encompassed by current SIA models).

4.1.2 The basics: social impact and social value

To put it plainly, the social impact of an organisation such as a WISE is the sway the organisation's activities manage to imprint in others' life, primarily among its beneficiaries, but possibly on other stakeholders as well. Impacts can be intended or unintended, as well as negative or positive, although WISEs will strive to create a positive impact. Social impact can hence be many different things to different people. In the context of a WISE, it can be an extended social network, the ability to secure and sustain a job, a raised self-esteem or an increased financial independence for instance.

Burdge & Vanclay propose the following definition for social impacts: "Social impacts include all social and cultural consequences to human populations of any public or private actions

that alter the ways in which people live, work, play, relate to one another, organize to meet their needs, and generally cope as members of society” (1995, 59). All these “life outcomes”, as they are sometimes referred to, are typically of an intangible nature and difficult to express through quantitative means.

Among the literature on SIA, there is a broad tendency to equate “*social impact*” to the creation of “*social value*”, where “*measuring social impact*” will consist in “*measuring*” the share of “*social value*” imputable⁴⁴ to the investigated activities (see e.g. Austin et al., 2006; Grieco, 2015; S. Olsen & Galimidi, 2008; Zappalà & Lyons, 2009). The adoption of this posture in SIA is largely correlated to the development and the widening adoption of the Social Return On Investment framework in the early 2000s and onwards (SROI)⁴⁵, which popularised the idea of a “*measurable*” social value generated by third sector actors. While we find that social value is a compelling entry to apprehend social impact and its assessment, it is another difficult notion to consensually establish (see e.g. Austin et al., 2006; Gibbon & Dey, 2011; Mulgan, 2010; J. Nicholls, 2007; Tuan, 2008).

⁴⁴ This implies a relation of causality between the *amount* of social value created and the activities that can specifically be tied to this *amount*. We elaborate further on this aspect of causality in 4.2.: we accept it as a conceptualisation feature of SIA (i.e. for some models) but reject as a defining feature (i.e. for all models).

⁴⁵ A popular social impact assessment method developed by the Roberts Enterprise Development Fund (REDF). It is based on a cost-benefit approach and consist in computing a return on investment ratio encompassing intangible social outcomes valorised through financial proxies. We provide more information on SROI in Appendix 8.

Textbox 2: Clarifying our taxonomy

the emergence of the phrasing impact “*measurement*” builds upon the considerations brought forth in Chapter 3 about the precedence of measurement over quantification (Desrosières, 2012). For this reason, and to avoid inducing a systematic metrological underpinning to a practice that is much more diverse, we elect to use the phrase “*social impact assessment*” rather than “*social impact measurement*”. The two terms are often found used interchangeably in the literature (see e.g. Grieco, 2015; Maas & Liket, 2011; Olsen et al., 2008).

Furthermore, while admitting to slight lexical deviations, we indicate that we subscribe to the interchangeable use of the terms “*social impact*”, “*social value creation*” or “*social return*” often observed in academia and among practitioners and highlighted by e.g. Maas and Liket (2011). Additionally, in the context of our research, we indicate that for WISEs, we consider social impact as a subset of “*performance*” (the subset we focus on in our research). By extension, we equate “*social impact assessment*” (SIA, as used generically throughout our dissertation) to a subset of WISEs’ “*performance measurement*”: as specified previously, we recognise that WISEs’ performance is not only social (Clark, Rosenzweig, Long, & Olsen, 2004; Elkington, 1997; Emerson, 2003).

In this dissertation, we nevertheless focus on the social aspects of performance for WISEs. This means that (i) by “*performance measurement*”, we mean: a general term for gauging WISEs’ performance, which we use for the frameworks implemented in the public sector (mostly focused on social aspects for WISEs, but not always). And (ii) by “*social impact assessment*” (or SIA) we mean: the performance measurement initiatives focused on social impact and stemming from WISEs and other mission-driven organisations.

More generally speaking, performance measurement extends over a broad corpus of literature (which in this thesis we restrict to economics and PAM). SIA constitutes a distinctive body of literature due its focus on, *inter alia*, social enterprises that is nevertheless encompassed within performance considerations (cf. Figure 7). Consequently, all the knowledge and contributions emphasised in this thesis pertain to performance measurement in general (e.g. performativity, the performance criteria introduced in Chapter 2, etc.) while some can *also* be tied specifically to the SIA literature (e.g. the attributes and variables that we describe in this chapter for our analytical framework).

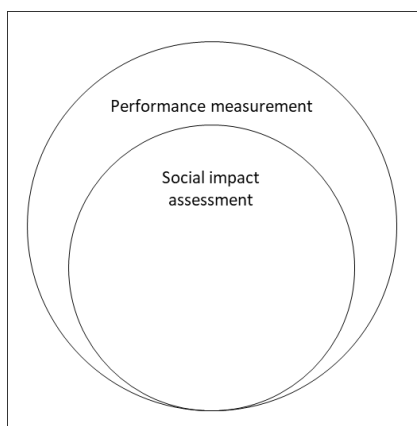


Figure 7: Performance measurement encompasses social impact assessment

Source: Author

When defining social entrepreneurship and social enterprises in 1.1, we already quoted Certo and Miller asserting that social value is *“the basic and long-standing needs of a society”* 2008, 267. This is a good start, but we need to go further to build a robust frame for our research. The two authors elaborate further about social value, by opposing it to personal or shareholder value. Here personal and shareholder values seem mostly implied in financial terms, with the authors pursuing that social value *“has little to do with profit but instead involves the fulfilment of basic and long-standing needs such as providing food, water, shelter, education, and medical services to those members of society who are in need”* (Certo & Miller, 2008, 67). Interestingly, this outlining of social value makes specific references to public goods (such as water or education) as we introduce them in 2.4 when delineating public value (where public value includes but is not limited to public goods). Our understanding is echoed by other contributions such as Austin et al. (2006), associating social value with increased public good.

Another interesting source to shape our understanding of social value is a publication originally written for REDF⁴⁶ and archived by the Harvard Business School, where Emerson, Wachowicz, and Chun define social value creation as *“[The combination of] resources, inputs, processes or policies (...) to generate improvements in the lives of individuals or society as a whole. (...) Examples of Social Value creation may include such ‘products’ as cultural arts performances, the pleasure of enjoying a hike in the woods or the benefit of living in a more just society”* (2001). We extract two main takeaways from this definition and its examples:

1. Social value is co-created by a complex network of *“resources, inputs, processes and policies”*
2. There is an emphasis on the intangible nature of social value, just as we highlighted above for social impact.

More crucially, the authors also state that *“[social value] has intrinsic value, but can be difficult to agree upon or quantify”*. This brings about a key consideration found across the literature on SIA, which is the necessity for providing a concerted, inclusive definition of the

⁴⁶ Roberts Enterprise Development Fund, where the SROI framework was originally developed

social value created by an organisation (see e.g. Grieco, 2015; S. Olsen & Galimidi, 2008; Tuan, 2008; Zappalà & Lyons, 2009). This inclusion is ideally reached through the involvement and participation of all the stakeholders concerned by the impact of the organisation, where they would get the opportunity to provide inputs as per how they perceive the social value created by the endeavour and discuss (i.e. **deliberate**) how it should be accounted for.

Looking at practical applications of social value in a policy making environment, we take the example of the so-called “Social Value Act” in the UK (2012), where the text is non-prescriptive about social value and instead requires, in case of public procurement, for *“the authority [to] consider how what is proposed to be procured might improve the economic, social and environmental well-being of the relevant area, and how, in conducting the process of procurement, it might act with a view to securing that improvement”*. We see that this piece of legislation leaves considerable room to shape one’s own understanding of what can be deemed to be social value. It is consistent with the observations we formulate earlier, and further stresses the likely irrelevance of a *“universal definition”* for social value.

These observations lead us to draw two main parallels between social value and public value as we frame it in 2.4. The first one is that both social value and public value encompass public goods. We saw in 2.4 that public value includes but is not limited to public goods, and we understand here from contributions such as Austin et al. (2006), Certo and Miller (2008) or Emerson et al (2001) that this is also the case for social value. This idea is reinforced by the dichotomic use of *“shareholder value”* to define both public value (Moore, 1995) and social value (Certo & Miller, 2008), as a mean to express the breadth of both concepts. This first parallel emphasises the intangible nature of both public value and social value and the near-impossibility to gauge them solely through financial figures.

The second parallel we want to bring forth is the relative nature of both social value and public value, and the importance for them to be commonly agreed upon by the stakeholders they concern. We saw in 2.4. that Benington talks about a *“contested democratic practice”* for public value (2007, 360), and we understand here how practitioners and academics alike advocate for a participative and inclusive definition of social value. This is evidenced in the

SIA literature, but also beyond, through the example we take of the Social Value Act in the UK, where social value remains undefined. This second parallel emphasises the participative and democratic dynamics pertaining to both public value and social value, and recalls that both types of value are co-created by complex networks of actors and processes.

Although we do not fully equate social value and public value, we use the parallels identified in this section to draw a bridge between SIA and public value, where the social value that SIA aims at capturing can be encompassed within public value (a concept broader than *just* social value). This view is supported by PVT contributions such as Bryson et al.(2017) or De Jong et al. (2017) who explicitly credit social enterprises for the creation of public value.

4.2 A short history of social impact assessment

As a primer for the SIA review we introduce in the third section of this chapter, and to understand where the flurry of methodologies and approaches come from, we propose to walk through a brief history of SIA. This historical view helps us understand how SIA emerged as a practice by actors such as WISEs, and why it can be relevant in a public setting. A thorough review of the literature indicates two historical trends in social impact assessment: one centred on “social accounting and audit” (SAA), and the other on “social impact assessment”⁴⁷. Each strand provides insight to better frame our research topic: SAA gained traction out of a need to demonstrate accountability to external stakeholders, and SIA was developed out of a policy making environment, initially using approaches of programme evaluation such as the ones we introduce in Chapter 3.

4.2.1 Social accounting and audit

Social accounting and audit (SAA) is a commonly used label nowadays for what has been dubbed, among others, “social audit”, “corporate social reporting” or “social responsibility accounting” depending on the zeitgeist. Rob Gray proposes the following definition for SAA: *“the preparation and publication of an account about an organisation's social, environmental, employee, community, customer and other stakeholder interactions and activities and, where possible, the consequences of those interactions and activities. The social account may contain financial information but is more likely to be a combination of quantified non-financial information and descriptive, non-quantified information”* (2000).

The earliest reference to the term “social auditing” is traced back by Carrol and Beiler (1975) to a 1940 monograph by Theodore J. Kreps entitled *“Measurement of the social performance of business”*. In the context of the financial crisis of the 1930s, Kreps was arguing that companies should report on their wider societal responsibilities to the public (Zadek et al., 1997). While this initial venture into social accounting was not met with success, Howard Bowen revived the idea with his book *“Social Responsibilities of the Businessman”* (1953).

⁴⁷ Here, “social impact assessment” designates the evaluation practice born in the late 1960s, in a policy-making setting, before third sector actors started to use similar methods and called them “social impact measurement” -we recall that we make the choice of conserving the term “assessment” as we avoid to induce a systematic metrological underpinning to a practice that is much more diverse (as shown in this section).

Although arguing for the same idea, Bowen's reasoning was at odds with Kreps', with the former highlighting the relevance of social auditing for internal purposes while the latter stressed a need for accountability towards external stakeholders (Carroll & Beiler, 1975; Zadek et al., 1997). Despite George Goyder's contribution which strove a balance between the two approaches (1961), the debate around the use and purpose of social auditing (external accountability versus internal management) remained during the development of the discipline in the 1970s (Bauer, 1973, cited by Hess, 2008; Zadek et al., 1997).

While experimentations with social auditing were conducted by a majority of large companies in the USA during the 1970s, the trend dissipated in the 1980s (Gray et al., 1987; Hess, 2008), before regaining popularity in the 1990s, through rising environmental concerns and the need for accountability mechanisms in this perspective (Gray, et al., 1997; Zadek et al., 1997). This context also gave birth to a more holistic approach to accountability, with the formalisation of the "triple bottom line" accounting concept (Elkington, 1997). While he first published on the topic in 1997, John Elkington is said to have coined the term in 1994 through the activities of his consultancy SustainAbility (Norman & MacDonald, 2004). The triple bottom line designates the three areas in which businesses should be accountable, namely in the economic, social and environmental fields. This is summarised by Elkington's famous quote "People, Planet, Profit" and is often associated with the three pillars of sustainable development, identified by the UN World Summit of 2005 as economic development, social development and environmental protection.

This broader and more inclusive approach to SAA is now taking precedence, and the development of organisations such as the Global Reporting Initiative (GRI) created in 1997⁴⁸ has helped the field reach a considerable momentum. When issuing its upgraded guidelines (G3) in 2006, the GRI was entering formal partnerships with, among others, the UN Global Compact and the World Economic Forum. The GRI has released the G4 version of its guidelines in 2013 and its network of users continues to expand. SAA in its traditional form continues to be used globally and is supported by organisations such as the Social Audit

⁴⁸ The GRI's main activity is the provision of guidelines to help organisations understand and communicate about the impact of their activities.

Network (SAN, created in 2000) but has lost some of its momentum, especially with the development of social impact assessment methods such as the social return on investment (SROI) framework⁴⁵ (Pearce & Kay, 2012).

4.2.2 Social impact assessment

Like several of the topics we cover in our research, social impact assessment⁴⁹ offers several definitions, which are relatively aligned and that we summarise in Table 4. We choose to retain the delineation proposed by Vanclay (2003) and adopted by The International Association for Social Impact Assessment: *“Social Impact Assessment includes the processes of analysing, monitoring and managing the intended and unintended social consequences, both positive and negative, of planned interventions (policies, programmes, plans, projects) and any social change processes invoked by those interventions. Its primary purpose is to bring about a more sustainable and equitable biophysical and human environment”* (2010). While SIA was initially meant to analyse a venture in forehand (*ex-ante*) (Burdge, 2003, 85), its recent widespread made the methodologies of SIA used also as evaluation tools (*ex-post*), contributing to the lack of a consensual definition for SIA highlighted by Rabel J. Burdge (2003).

Table 4: Proposed definitions for social impact assessment

Author	Year	Definition
Freudenburg	1986	SIA refers to assessing (as in measuring or summarizing) a broad range of impacts (or effects, or consequences) that are likely to be experienced by an equally broad range of social groups as a result of some course of action
Interorganizational Committee on Guidelines and Principles	1995	We define SIA in terms of efforts to assess or estimate, in advance, the social consequences that are likely to follow from specific policies or actions
Branch and Ross (as reported by Grieco, 2015)	1997	SIA needs to be project-oriented to ensure that social issues are included in project design, planning and implementation, as well as

⁴⁹ Here, *“social impact assessment”* designates the evaluation practice born in the late 1960s, in a policy-making setting, before third sector actors started to use similar methods and called them *“social impact measurement”*.

Author	Year	Definition
		ensuring that development is acceptable, equitable and sustainable
Goodland	2000	SIA is more than a technique or step; rather, it is a philosophy about development and democracy. As such, ideally, it considers pathologies of development, goals of development, and processes of development
Burdge	2003	The goal of SIA is to help individuals and communities, as well as government and private- sector organizations, understand and better anticipate the possible social consequences for human populations and communities of planned and unplanned social change resulting from proposed policies, plans, programs and projects
Vanclay	2003	SIA includes the process of analysing, monitoring and managing the intended and unintended social consequences, both positive and negative, of planned interventions (policy, programs, plans, projects) and any social change process invoked by those interventions
Council for Social Development	2010	SIA is a process that seeks to assess in advance the social repercussions that are likely to follow from projects undertaken to promote development [...] As an aid to the decision-making process, SIA provides information on social and cultural factors that need to be taken into account in any decision that affects the lives of project area people
Epstein & Yuthas, 2017	2017	Social impact measurement is designed to identify changes in social impacts that result from [an organization's] activities. Most organizations measure the outputs they produce (for example, meals served or jobs created). Social impact measurement assesses the ultimate impacts of those outputs on individuals and the environment (for example the quality of life, or survival of species)

Source: Adapted from Grieco, 2015

The theoretical foundations for the discipline of impact assessment are often attributed to Donald T. Campbell, who laid down the principles for valid experimentations in social sciences (1957). The nascence of social impact assessment itself is usually associated to the introduction of the National Environmental Policy Act of 1969 (NEPA) in the United States of America (USA). William R. Freudenburg highlights the simplicity of NEPA (a five-page piece of

legislation), contrasting it with the considerable impact it had (1986). NEPA required and still requires any federal agency about to implement actions likely to “significantly affecting the quality of the human environment” to assess the potential impacts of the said actions, utilizing “a systematic, interdisciplinary approach which will insure the integrated use of the natural and social sciences” (in section 102 of NEPA). NEPA thereby laid down the foundation for what was to be labelled environmental impact statement (EIS) and environmental impact assessment (EIA).

While NEPA is credited for introducing a formalised approach to impact assessment, it is however worth noting that previous regulatory initiatives already encompassed policy evaluation dimensions. For instance, the Office of Economic Opportunity created in 1964 in the context of the American president Lyndon Johnson’s “War on Poverty” had a department dedicated to programme evaluation: the Office of Research, Plans Programs, and Evaluations (Caro, 1977).

The term “social impact assessment” is assumed to have first been coined in a 1973 EIS produced to gauge the consequences on indigenous populations of the construction of a pipeline in Alaska (Burdge & Vanclay, 1995). This case, together with its impact on Inuit populations’ customs made obvious the need to assess not only environmental factors in EIS but also social factors. SIA was born as a discipline, and C.P. Wolf would commit its first publication in 1974 (Freudenburg, 1986).

The following years saw an enlargement of SIA practices beyond the environmental field, such as the evaluation of the National Supported Work Demonstration programme in the USA (1974). Guidelines for SIA were progressively refined and adopted in different legislations, until 1986 when the World Bank adopted environmental and social assessment as part of their evaluation procedures, paving the way for the widespread adoption of NEPA-inspired assessment models around the world (Burdge & Vanclay, 1995).

Born in a policy-making and legislative environment, the use of SIA shifted to the third-sector in the 1990s. The work of Georges R. Roberts, is often considered pioneering in that

perspective, supporting his philanthropic investments in work and education programmes which led to the creation of the Roberts Enterprise Development Foundation (REDF) in 1997. REDF, under the direction of Jed Emerson, developed a dedicated framework in order to monitor its investments. It documented its method in a publication in 2000, dubbing the approach Social Return On Investment (SROI) (Millar & Hall, 2012). While other philanthropic stakeholders have been developing and implementing their own assessment and evaluation frameworks (e.g. the W.K. Kellogg Foundation's Evaluation Handbook in 1998) SROI is the approach that gained the most traction over the past decade, largely contributing to the spread of social impact assessment in the third sector. Milestones contributing to the widespread of REDF's tool are, *inter alia*, the creation of a working group to provide an SROI framework in 2004, the launch of the SROI Network in 2008 or the publication of a Guide to SROI supported by the Cabinet Office in the United Kingdom (UK) in 2009.

While working on the development of the SROI framework, Emerson rapidly advocated for a holistic approach to value creation, promoting the "blended value proposition". The main idea behind blended value is that social value should not be separated from the economic return generated by an activity or an investment. Instead, they should be integrated in a framework allowing to assess the overall value of a project (Emerson, 2003). This holistic approach to value creation is not unreminiscent of Elkington's triple bottom line. However Emerson's original proposition insists on going "beyond the triple bottom line" (2003, p. 49).

Having the inventor of one of the most popular social impact assessment model (SROI) proposing an approach taking inspiration from SAA (the triple bottom line) contributes to blur the line between the two schools of thoughts we describe. It is however a fact that in practice, both slants come more and more together at a time where the "*measurement*" of social impact is a widespread practice that extends beyond policy-making and beyond the third sector. Global "*mainstream*" corporations have adopted and implemented the GRI guidelines, others have developed their own approach to the issue (see for instance the 2013 initiative Roundtable for Product Social Metrics⁵⁰, involving global companies such as L'Oréal, BASF or

⁵⁰ An initiative driven by leading global companies to measure the social impact associated with the products they manufacture and / or sell.

Philips). Numerous organisations have thereby been very active in this area, each trying to shape solutions to answer their own needs (Gibbon & Dey, 2011; Maas & Liket, 2011; Zappalà & Lyons, 2009).

Through this clarification of SIA's development process, we highlight two key takeaways. The first is the external accountability that underpinned the adoption of SAA and that still bears a significant importance in the way SIA is operationalised today (interestingly, what is credited to be the first contribution to the field of SAA, in 1940, already spoke of "*measuring social performance*"). The second takeaway is an inception of the original practices of SIA anchored in policy making and attached to the use of programme evaluation practices. Does it mean that SIA models can be implemented as is for public performance measurement? Before elucidating this question, our next section looks at what exactly these models are and propose an analytical framework to structure our understanding of them.

4.3 Social impact assessment: a state of the art

The enthusiastic and creative undertakings around SIA described above led to a sprouting in the number of models proposed. For instance, a quick search on the Foundation Center's Tools and Resources for Assessing Impact library (TRASI) across all proposed criteria returned 193 results⁵¹. How to identify the models most relevant to our needs (WISEs measuring their performance in a PAM context)? How to classify what looks like a forest of models?

This section helps answering these questions by first introducing the impact value chain (Clark et al., 2004), an established overarching framework for these models, before proposing a scaffold of 13 variables split across four attributes (going from 118 variables identified in the literature) to parse the content of these models. Doing so, we also propose a review of SIA models best applicable to WISEs. As we do, we will bear in mind that *"There is no single tool or method that can capture the whole range of impacts or that can be applied by all corporations"* (Maas & Liket, 2011, 9).

4.3.1 An overarching framework for social impact assessment: the impact value chain

Facing the early spur in the number of approaches to social impact assessment, and in an attempt to harmonise the taxonomy among different experts coming from political science, economics, management and a wealth of other disciplines, Clark et al. proposed the impact value chain in 2004 (cf. Figure 8) to suggest a common frame of reference. The five stages of the impact value chain cover all the aspects of a programme, starting from the resources used as inputs to the outcomes it creates for the populations taking part in its activities. The key feature of this holistic approach, as stressed by Clark et al. (2004), is to differentiate outputs from outcomes, with the latter allowing to identify the programme's contribution to social change. Considerations about the impact value chain are covered extensively in the existing literature (see e.g. Clark et al., 2004; Grieco, Michelini, & Iasevoli, 2015; Maas & Liket, 2011; Olsen & Galimidi, 2008). It is a relatively consensual notion for practitioners and academics alike and therefore constitutes a useful frame of reference for our reflection on SIA.

⁵¹ This search was carried out in May 2015. TRASI was then retired by the Foundation Center as a stand-alone project and combined with Foundation Center's IssueLab, a centralised knowledge service. IssueLab does not allow to carry out a research similar to the one that was undertaken with TRASI.

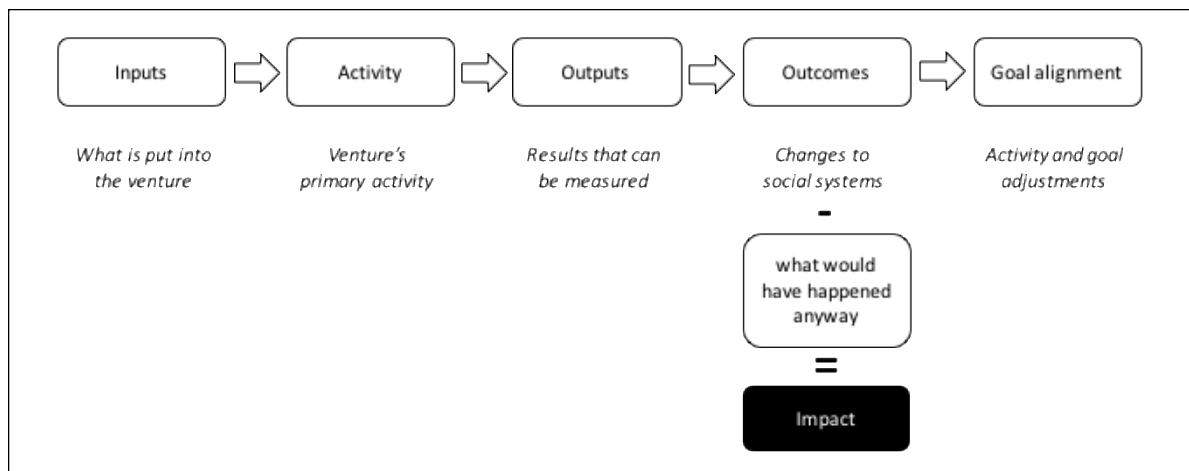


Figure 8: The impact value chain

Source: adapted from Clark et al., 2004

The impact value chain is highly reminiscent of the “grey box” drawn in Figure 5 (representing the programme to evaluate) when introducing a framework for performance evaluation in PAM. In Clark et al.’s original contribution (2004), it is hinted the creators drew heavily on programme evaluation literature to develop their scaffold, which can in part explain this overlap. Interestingly for us, this draws **another parallel** between the performance of WISEs and performance as understood in a PAM context. We see that performance is conceptualised in very similar ways in both situations, which seems like a logical consequence of SIA’s historical development as we describe it in the previous section.

Further, we note that the difference between outputs and outcomes is also a key highlight both in PAM performance and in SIA. In SIA, there is a consensual emphasis of the importance of outcomes, which is where social value (i.e. the social impact) is created. The pre-eminence of outcomes is less evident throughout PAM, where the most recent contributions such as the ones found in NPG or PVT stress the importance of outcomes, clearly highlighting the dichotomy with outputs, but where paradigms such as NPM tend to favour a focus on outputs. Consequently, the use of PVT provides a good alignment with SIA for the conceptualisation of performance across sectors and stakeholders.

Lastly, we want to highlight an important feature of the impact value chain, which is the implied notion of **causality** shown by the “*what would have happened anyway*” box in Figure

8. Here, the aim is to single out the outcomes only attributable to the considered activities. There is a similar intent (however less evident) in PAM performance frameworks by displaying outcomes out of the programme's "grey box" and linked to the environment (cf. Figure 5) In practice, identifying outcomes causality it is a difficult endeavour that involves the use of a counterfactual (i.e. a theoretical scenario for what would have happened without the considered intervention). Despite its presence in the social value chain, causality (or the problem of "attribution" as it is often referred to in the SIA lingo), is present in some SIA models, but not all. It is particularly used in models focused on monetarisation (cf. Table 6) like SROI or the Robin Hood Foundation Benefit-Cost Ratio which draw heavily on cost benefit analysis, where counterfactuals are commonly used. It is however absent from popular models such as IRIS or the Outcome Star (cf. Table 5).

While we present the impact value chain as it is most often conceptualised (Clark et al, 2004; Olsen et al. 2008; Maas & Liket, 2011), we do not adopt the stance of considering causality a defining feature of SIA. This is reflected in the selection of SIA models we introduce below (which do not always encompass causality issues), but is also a tacit assumption of previous contributions cataloguing different SIA models (e.g. Grieco, 2015), including the above-mentioned authors making use of the impact value chain. As a final comment, we want to highlight how causality, in France, seems to be the differentiator between the notion of "social utility" traditionally used in the country since the 1970s and social impact (cf. Chapter 8).

4.3.2 Proposed models for WISEs' performance

In this segment, we provide a likely non-exhaustive list of SIA models that can be applicable to WISEs. What we propose here rests on a literature review from both academia and practitioners. As highlighted previously, SIA has evolved very rapidly in the last decades and academia has often lagged behind when it came to inventory practices in this ever-expanding field. Incorporating resources from practitioners therefore helps achieving thoroughness.

Using three academic (Business Source Complete, EconLit and Google Scholar) and one practitioner databases (the Foundation Center's TRASI), as well as some ad'hoc resources

from other practitioners such as the Bill & Melinda Gates Foundation or the New Economics Foundation, we compiled a list of 124 models for social impact assessment potentially relevant for social enterprises, available in Appendix 7. From there, we refined our sample following the rules described below and based on two anchors: the applicability of the selected models to WISEs (in terms of impact measured and proposed approach) and the level of information available for each model.

Applicability to WISEs:

In order to decide which SIA models are valid for WISEs, we use the definition of WISEs proposed by Davister et al cited in 1.1. Building on this definition, we choose to retain impact measurement models that meet the following criteria:

1. The approach must allow to assess organisations or entities considered autonomous. Here, autonomy is mostly defined by opposition to a control exerted by public authorities. The considered organisation must therefore be of an entrepreneurial nature with a “largely autonomous management” (Davister et al., 2004, p.24).
2. The approach must allow to assess economic entities. This is an important point that will exclude a large share of evaluation models focused solely on programme evaluation.
3. The approach must allow to assess organisations whose activities are focused on professional integration of “people experiencing serious difficulties on the labour market” (Davister et al., 2004, p.3). This will filter for some models having exclusive sectorial approaches such as the environment-focused Trucost or the development-focused Development Outcome Tracking System (DOTS) (S. Olsen & Galimidi, 2008).
4. The approach must allow to assess the outcomes (professional and social integration) of the type of activities offered by WISEs for the targeted population (productive activities and tailored follow-up or qualifying training). This will exclude some models having an exclusive topical approach such as the New Progressive Coalition’s “Political Return On Investment (Olsen & Galimidi, 2008) or the Volunteering Impact Assessment Toolkit by the New Economics Foundation.

Information availability:

Additionally to these criteria inherent to the concept of WISE, we also exclude some specific types of approaches that may have been part of previously available impact measurement catalogues but do not fit the purpose of this thesis. We therefore exclude:

- 1) Approaches for which not enough background and technical information is publicly available from primary sources to carry out a proper analysis. This includes:
 - a) Proprietary solutions, often provided through consultancy services or part of a portfolio management system, that cannot be deployed without their owner's support or outside of their owner's activities such as Acumen Scorecard, Human Impact + Profit (HIP), Shujog, B Lab's GIIR rating (formerly B rating system), the Dalberg Approach or REDF's Real Indicators of Success in Employment (RISE).
 - b) Packaged software solutions that cannot be considered as "stand-alone models" without the considered system or its provider's support, such as Pulse (a portfolio management system developed by App-X), Acumen's Strategic Management Solution or Guidestar's embedded impact measurement solution.
 - c) "Discontinued" solutions that, at the time of the research, were not anymore supported by the organisation that created them, such as Assessment of Social ImpactS (OASIS), SROI Lite or SROI Calculator.
- 2) Approaches that provide conceptual scaffolds and guidelines to apprehend impact measurement but that do not propose operationalizable assessment tools (i.e. a framework or a methodology concrete enough to be applied as is, and for which a systematic approach is applicable with a minimum of adaptation) These include:
 - a) Design guides or seminal approaches, such as Social Accounting and Audit (SAA), Triple Bottom Line Accounting, Logical Frame (Logframe), Social IMPact for Local Economies (SIAPLE) or Participatory Impact Assessment, all of which provide guidelines and high-level frameworks for assessment and evaluation but do not propose tools for impact assessment per se.
 - b) Similarly, we exclude standards and guidelines such as AA1000, SA8000, the Global Reporting Initiative (GRI) or Social Reporting Standard (SRS), as well as quality management systems such as Practical Quality Assurance System for Small

Organisations (PQASSO) or ISO 26000 which are focused on social issues but do not provide operationalizable assessment tools.

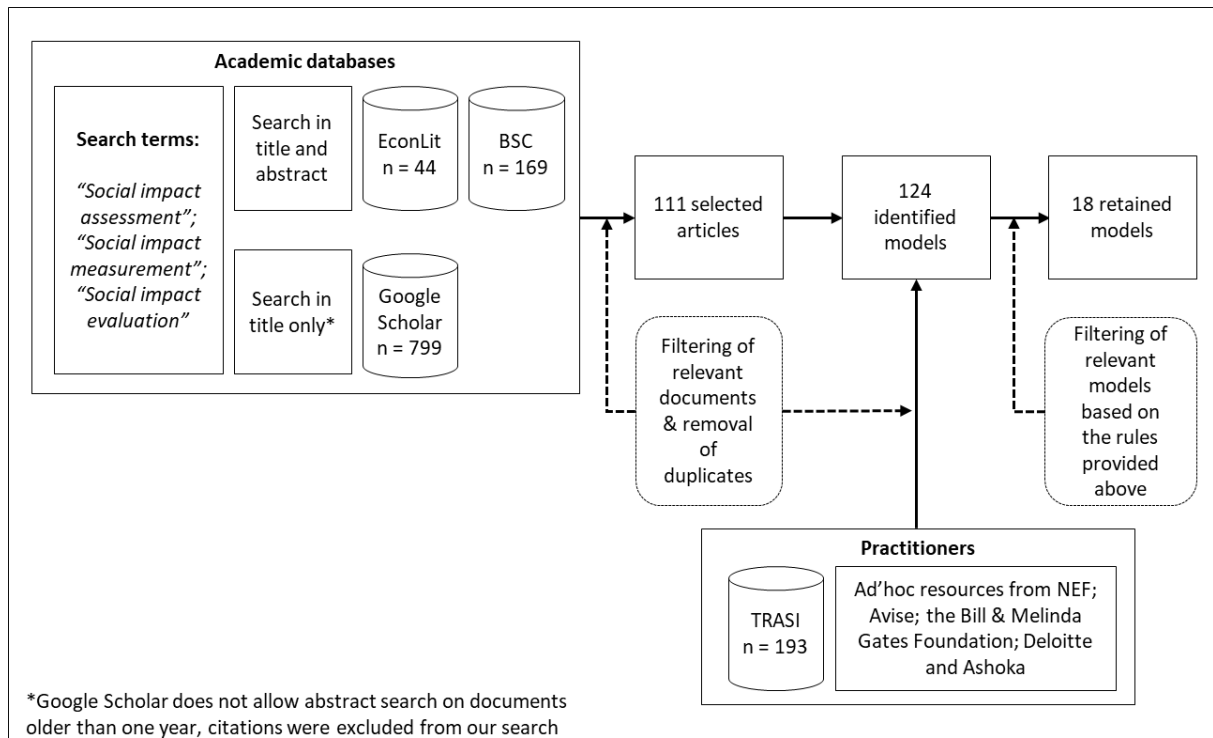


Figure 9: Selection method

Source: Author

Table 5 presents the 18 SIA models we retain as applicable for WISEs following the filtering process described above. We note the presence, among the models introduced, of the Public Value Scorecard, proposed by Moore in 2003. This selection is the result of the filtering process and was not influenced by the framework's focus on public value. When initially framing his approach, Moore did not refer to social enterprises. The model was however inventoried later by Maas and Liket (2011), Grieco et al. (2015) and Grieco (2015) as a SIA. This appears to further validate our stance of assuming an overlap between social value as framed in a SIA context and public value.

We emphasise that, although we ambition to be as thorough as possible in establishing the selection proposed below, due to the increasing number of organisations developing proprietary tools, the list we provide should better be considered a "thorough review".

Table 5: Social impact assessment models applicable to WISEs

Model	Developed by	Selected academic references
Ashoka Measuring Effectiveness Questionnaire	Ashoka	n/a
Atkisson Compass Assessment For Investors (ACAFI)	Atkisson	Olsen & Galimidi, 2008
Balanced Scorecard Modified To Include Impact	New Profit	Olsen & Galimidi, 2008
Basic Efficiency Resource (BER) Analysis	Cugelman & Otero	n/a
Best Available Charitable Option (BACO)	Acumen Fund	Maas & Liket 2011
Charity Analysis Tool (ChAT)	New Philanthropy Capital	Olsen & Galimidi, 2008
Comparative Constituency Feedback	Keystone	n/a
Cost Per Impact	Center for High Impact Philanthropy	Tuan, 2008
Echoing Green Mid-Year And Year End Reports	Echoing Green	Kramer, 2005
Hewlett Foundation Expected Return	William and Flora Hewlett Foundation	Tuan, 2008
IRIS Library	Deloitte, PWC, Acumen Fund, B-Lab, Hitachi, Global Impact Investing Network	Olsen & Galimidi, 2008
Measuring Impact Framework (MIF)	World Business Council for Sustainable Development	Maas & Liket 2011
Outcomes Star	Triangle	Nicholls et al. 2012
Public Value Scorecard (PVSC)	M.H. Moore	Maas & Liket 2011
Robin Hood Foundation Benefit-Cost Ratio	Robin Hood	Maas & Liket 2011
Social Compatibility Analysis	Institute for Sustainable Development	Maas & Liket 2011
Social Cost-benefit Analysis	n/a*	Maas & Liket 2011
SROI Framework	REDF	Olsen & Galimidi, 2008

Source: Author

*Social Cost-benefits Analysis and Social Costs-Effectiveness Analysis are both described by Maas as models derived from "general economic tools" (2011, p.29) that have been adopted to economically assess the social aspects of programmes or interventions. Vardaskoulas (2013) wrote a piece for the New Economics Foundation about Social Cost-benefits Analysis as a social impact assessment tool.

Along with presenting the models listed here, several contributions have focused on providing frameworks to categorise and understand how these approaches can be operationalised and the type of methodology they promote. Appendix 10 offers a detailed account of the 36 attributes and their 118 variables (including redundancies) of SIA models proposed in these contributions. On this SIA basis, and building upon the highlights of Chapter 3 in terms of performance criteria and performance regimes we build our own analytical framework for WISEs' performance models. We apply it to the 18 SIA models we retain for WISEs in Appendix 8 and we detail its entries below in Table 6.

Table 6: key performance models attributes and variables and their contributors

Attributes	Variables	Authors
Purpose	Screening (<i>ex-ante</i>)	Grieco et al., 2015 Krev et al., 2012
	Monitoring (<i>interim</i>)	Olsen et al., 2008 Maas & Liket, 2011
	Reporting (<i>interim</i>)	Mulgan, 2010 Rinaldo, 2010
	Evaluation (<i>ex-post</i>)	Tuan, 2008
	Analytical lens	Economy
Efficiency		Maas & Liket, 2011
Effectiveness		Tuan, 2008
Cost-benefit		Rauscher et al., 2014
Monetarisation		
Data typology	Quantitative	Grieco et al., 2015
	Qualitative	Nicholls, 2005
Design	Planned	Jany-Catrice, 2012
	Negotiated	

Source: Author

Our framework introduces one novel contribution that does not appear in the SIA literature, which is the “*Design*” entry and is inspired by Jany-Catrice’s work on performance regimes (2012, cf. Chapter 3 and Figure 6). It aims at specifying whether the performance measurement process and its indicators are the fruition of a negotiated (bottom up) or planned (top down) approach. This attribute is surprisingly absent from all the existing

classifications we came across, perhaps because most SIA models insist on stakeholders' involvement when outlining their assessment process and a tacit assumption is made that these directives are applied (our data shows otherwise, cf. Part 2). We only explicitly include one of the two delineations proposed by Jany-Catrice (planned versus negotiated performance measurement processes) as we consider that the other demarcation (single indicators versus multiple indicators systems) is reflected in the use of other attributes.

The rest of the attributes are based on our literature review (chiefly Clark et al., 2004; Gibbon & Dey, 2011; Grieco, 2015; Grieco et al., 2015; Kelly et al., 2002; Krlev, Münscher, & Mülbert, 2013; Liket & Maas, 2015; Maas & Liket, 2011; Mulgan, 2010; J. Nicholls, 2007; S. Olsen & Galimidi, 2008; Rinaldo, 2010; Schober & Rauscher, 2014; Tuan, 2008) and informed by the seven performance criteria (relevance, economy, efficiency, effectiveness, cost effectiveness, utility and sustainability) we introduced in Chapter 3. They are understood as followed:

Purpose is an attribute observed recurrently across existing contributions. Most models can be used both for screening (*ex-ante*) and evaluation (*ex-post*) purpose, provided assumptions are formulated for the former. Social impact investment funds such as Den Sociale Kapitalfond in Denmark have carried out such *ex-ante* analysis using the SROI framework to screen investment opportunities ahead of committing capital. A lot of models are however more difficult to operationalise on an ongoing basis for monitoring or reporting purposes (especially those focused on approaches such as monetarisation). Such a goal will be better served by models like the Outcome Star or the Balanced Scorecard Modified To Include Impact.

Although we are interested in the purpose a model can serve, we leave out the potential “*end goal*” for which the model can be used by a WISE (regardless of its initial purpose). End goals include uses such as advocacy or fundraising⁵², which cannot be determined by the design of the method alone and answer managerial needs (i.e. strategic intents) observable at the level of the organisations only. Part of the work undertook in our research, and one of our key

⁵² We further specify that we leave out “*certification*” from our list of purpose despite observing it in some contributions (e.g. Grieco, 2015) on the basis that we assimilate it as an *end-goal* (e.g. B Lab's models are primarily designed to *screen* and *assess* organisations in the view of delivering a certification).

contribution, consist precisely in identifying the strategic intents behind the use of SIA and performance measurement models in general (cf. Chapter 10).

In our framework, the use of the purpose attribute renders the timeframe (prospective, ongoing, retrospective) and accountability (external, internal) attributes observed in other contributions redundant: screening is by essence prospective; monitoring and reporting are ongoing (respectively towards internal and external stakeholders) and evaluation is retrospective. Similarly, we also left out related attributes such as length of timeframe (short term or long term, e.g. in Maas & Liket, 2011) as they do not bring enough value in the analysis of our empirical material.

Analytical lens

This attribute parses the analytical approach of different models. It is informed by our SIA literature review, but also by the seven indicators we identify for PAM performance, which helps us bring a granularity in conceptualising the approaches that is not often observed in SIA.

Economy brings an emphasis on raw costs, without necessarily relating them to outputs (cost efficiency) or outcomes (cost effectiveness). It draws on the “input orientation” proposed by Maas and Liket (2011) as well as on the costs analyses proposed by Tuan (2008) and Rauscher et al. (2014). Based on the discussion we provided on social impact and social value, we understand that economy is not a focus area for “true” SIA. It is hence rather normal to observe that none of the 18 models we retained for WISEs offer an emphasis on economy, even though some such as the Measuring Impact Framework may touch upon the topic while assessing other economic areas.

Efficiency relates inputs to outputs, gauging the capacity with which a WISE can use resources with parsimony to carry out its activities. This attribute is found at the intersection of Maas and Liket’s input and output orientations (2011) as well as in Tuan’s (2008) and Rauscher et al.’s (2014) work. It can also be paralleled to Clark et al.’s process function (2004), where SIA serves as a management tool to drive, *inter alia*, efficiency. We note that although always

expressed numerically, efficiency does not necessarily carry an economic dimension. For instance the “*positive graduation rate*” used in France by public actors is an efficiency measure of the rate of beneficiaries successfully completing a programme without any economic measure behind it. Efficiency is found quite pregnant in models such as Basic Efficiency Resource (BER), Best Available Charitable Option (BACO) or Cost per Impact.

Effectiveness relates outcomes to outputs and is interested by how effectively the latter translate into lasting effects as captured by the former. This is a dimension that is seldom explored as such in available SIA models although some of its aspect can be found in processual oriented approaches (Clark et al., 2004; Maas & Liket, 2011) such as the Measuring Impact Framework (MIF) but also in some implementations of monetarisation-oriented models such as SROI where the transformation of output into outcomes is modelled in order to get a realistic quantification of the assessed impact.

Cost-benefit encompasses the relation between inputs and outcomes (cost efficiency), but also possible wider economic impacts that are typically identified through the use of a counterfactual (i.e. a scenario of what would have happen without the WISE’s intervention, with its estimated costs – assumed to be saved by the intervention). We see such an approach in models like the Robin Hood Foundation Benefit Cost ratio, the Measuring Impact Framework or the adequately called Social Cost-Benefit Analysis. While we made the choice to create a variable combining a certain diversity of approaches, contributions such as Rauscher et al. (2014) and Tuan (2008) provide more granularity, proposing a distinction between cost efficiency, cost-benefit and cost utility.

Monetarisation

Although not an analytical approach as such, we choose to include monetarisation as a variable in our analytical lens attribute due its heavy reliance on cost-benefit analysis and on complementary economic and financial research. Monetarisation consists in providing a monetary value to outcomes, using financial proxies. These quantitative substitutes are supposed to reflect the value of these outcomes to society and / or the stakeholders concerned with them. Monetarisation is a variable initially proposed by Clark et al. (2004) and

is a fundamental component of the SROI approach, where stakeholders get to participate in defining the proxies they wish to use to value a venture's impact.

Monetisation is used in the valuation of non-market resources such as environment preservation or indeed, social impact. Although these resources provide a degree of utility to the population, the fact that they cannot be exchanged through customary market mechanisms make it difficult to value them or provide them with a market price. Monetisation has been a popular approach to value a wide range of public goods (R. C. Mitchell & Carson, 2013). We discern two main methods to carry out monetisation and thereby outline financial proxies for intangible outcomes:

Stated preference: It consists in estimating the value of a good by surveying a population and asking them how much they would be willing to pay to access or maintain the good that is being valued (Ciriacy-Wantrup, 1947). Alternatively, contingent valuation can also ask how much one would be willing to accept to be compensated for the loss of a good. In SIA, this technique can be mobilised by asking stakeholders how much value they place on a given outcome. Stated preferences valuation is typically carried through contingent valuation methods (i.e. valuing a good as a whole) or choice modelling method (i.e. valuing specific attributes of a good).

Revealed preference: It consists in inferring the value of a good by observing consumption patterns around it. Pioneered by Paul A. Samuelson (1938), the approach here is to address the gap that may be seen between statements and actions (a weakness often highlighted in stated preference) by deriving value directly from observed consumer behaviours. One main approach to revealed preference are hedonic pricing, where the value of the good is inferred by the market mark-up observed on goods with similar characteristics. It can for instance be used in environmental impact assessment, where environmental features (e.g. proximity to a forest) are valued according to the market premium associated to housing offering that feature. Another common approach is the travel cost method, which aims at valuing a good (often a site) through the price consumers are willing to pay to travel and access that good.

Revealed preference approaches can be found in environmental impact assessment, but are seldom observed in SIA.

Data typology refers to the type of data used to inform a SIA model. We stress the notion of informing, as some models known to present purely quantitative yields such as the SROI framework are in reality constructed from a wealth of qualitative information collected as part of their design. Consequently, we rate in Appendix 11 models such as SROI or the Social Cost-Benefit Analysis as primarily quantitative while acknowledging a degree of qualitative approach.

Our focus on WISEs 'SIA is by essence concerned with a micro level. We hence made the choice to exclude the perspective attribute (micro, meso, macro) presented for instance by Maas & Liket (2011), although we acknowledge the meso and macro perspectives that some SIA models may allow.

Conclusion of Chapter 4:

Through this chapter, we introduced the last piece of our literature review. We focused on the performance of WISEs and showed how complex it is to capture due to its heterogenous and often intangible nature. Doing so, we also highlighted how the **social value** (the long-standing needs of a society) created by WISEs aligns with the public value (the value generated for the citizenry) of Moore's public value theory (1995). While this would seem to confirm that public authorities have legitimate grounds to question WISEs' performance, this hardly provides any answer as per how the measurement of this performance should be carried out.

Social impact assessment (SIA) offers an interesting and pertinent scaffold to bring some answers to this question, through the angle of **social performance** (Gianfaldoni, 2018; Lall, 2017). The last section of this chapter was therefore dedicated to the presentation of the analytical framework we built from our review of the SIA literature. We enriched this scaffold with some of the findings from Chapter 3 (e.g. the design attribute derived from Jany-Catrice's work). While containing some of our main empirical contributions, this review and the

framework it informs will be our main instrument to parse and comprehend the performance measurement models we come across (whether they are actual performance models coming from the public or SIA-type models coming from WISEs themselves).

Having the last piece of our investigative puzzle in place, we are now ready to frame our research question and present an overarching analytical grid, encompassing the public administration and management (PAM) environment in which Danish and French WISEs operate (our Chapter 2), laying out the key issues of performance we will focus on (our Chapter 3) and using the analytical lens of public value theory to make sense of it all.

Chapter 5 A PVT-informed, critical realist analytical framework

In this chapter, we collect the key considerations we brought forth across our review of the literature in Public Administration and Management (PAM), in performance measurement, and in social impact assessment (SIA). We use these theoretical entries to further develop our initial questioning: how is the development of social impact assessment (SIA) changing the way that public actors are measuring work integration enterprises (WISEs)' performance?

We start by recalling that WISEs are co-creating public value with their public stakeholders, as we evidence the central role of the strategic interactions they entertain with public actors. We develop on the role of performance measurement in these interactions and show that PVT's attributes (such as deliberation and the focus on shared preferences) can bring answers to the performativity issues we tie to performance measurement.

We continue by showing how our three entries allow us to build on the paradigmatic approach we adopt in our research (critical realism, which we define in Chapter 6). Critical realism is characterised by three layers of reality: the empirical (the facts as they are experienced), the actual (the facts as they occur) and the deep real (what causes the occurrence). We highlight how different parts of our literature review cater to the different layers, and we show how PVT acts as the structuring vector of our analysis, allowing us to identify explanatory patterns for the behaviours we observe associated to performance measurement.

We conclude by delineating our research question, breaking it down into three sub-questions.

5.1 The complementarity of our three theoretical anchors

In this section, we indicate how the different analytical entries we develop in Chapter 2 through Chapter 4 contribute to an overall understanding of the issues we research. We start by outlining how these theories shape an apprehension of the strategic interactions observed between WISEs and their public stakeholders. We then develop about how they mutually reinforce each other to form the overarching framework we use in carrying out our research.

We start by recalling some of the fundamental elements covered in Chapter 1 around the cooperation between work integration social enterprises (WISEs) and public actors, using the insight provided by our public administration and management (PAM) literature review (Chapter 2). In a second segment, we develop about how public value theory (PVT., Moore, 1995), our central analytical device, builds relevantly on the reflections we introduced on performativity (our Chapter 3) and in the measurement of WISE's performance in particular through social impact assessment (our Chapter 4).

5.1.1 A starting point: the strategic interactions between WISEs and public actors

We have seen that as they propose remedies to market failures (Bator, 1958; Zeckhauser & Stokey, 1978), WISEs can be qualified as active creators of public value (Bryson et al., 2017; De Jong et al., 2017). This perspective ascribes our research to the most recent trends in PVT where multi-actor public value co-creation is stressed and where the use of the PVT triangle across all PVT stakeholders (and not only the public sector) is increasingly recognised (see e.g. Benington, 2011; Bryson et al., 2017; Fisher, 2014; Hartley et al., 2017).

WISEs are private actors, which are mission-oriented (i.e. driven primarily by their social mission as opposed to the accumulation of profit). This social imprint (Battilana et al., 2015) is what motivates them towards the creation of public value (Bryson et al., 2017; De Jong et al., 2017). This is also part of what makes them interact with public stakeholders (either voluntarily through contracted work; or unavoidably to channel beneficiaries or receive non-market incomes such as subsidies). Through their mission, both WISEs and public actors are assumed to pursue similar goals of public value creation (Moore, 1995).

These strategic interactions are also mandated on the public side. First under the reforms inspired by the NPM paradigm and its increased recourse to market resources to create public value such as the one delivered by WISEs (see e.g. Defourny & Nyssens, 2010; Hulgård, 2011; A. Nicholls, 2010; Nyssens, 2006; V. A. Pestoff, 2008). And second in the context of new post-NPM paradigms such as the NWS or the NPG, conducive to cooperation networks and the co-creation of public value (Alford & Hughes, 2008; Bryson et al., 2014; O’Flynn, 2007).

5.1.2 Articulating PVT, performance measurement and our reflection on performance indicators

Further building on the relationship between WISEs and their public stakeholders, performance measurement finds relevance on several levels. It can be one of the monitoring and performance evaluation instruments advocated by NPM’s doctrinal components (Hood, 1991) and recuperated by NPM’s offspring such as NWS and NPG (Pollitt & Bouckaert, 2017). Drawing on PVT, it can also be e.g. a device to build or reinforce legitimacy by “proving” the creation of value to WISEs’ stakeholders. This approach would fall into the extended use of the PVT strategic triangle advocated by e.g. Bryson et al., (2015b) or Hartley et al. (2017). The encouraged practice of stakeholders involvement in the definition of impacts, found in a number of SIA models, also builds relevantly on PVT’s ideas of collective preferences (see e.g. Moore, 1995; O’Flynn, 2007) and contested democratic practice (Benington, 2015), and aligns with Benington’s view on how to reach tentative agreement on public value (2015).

The reflection we introduce on performance measurement and the creation of indicators in Chapter 3 enhances the considerations brought forth through the other entries of our theoretical framework. We show that one of the core issues linked to the performativity of indicators, indicator-substitution (see e.g. Desrosières, 2012; Jany-Catrice, 2012), carries important democratic stakes as per how indicators are defined and why their delineation (i.e. the quantification process) should involve the wider public (Desrosières, 2012; Jany-Catrice, 2012; Ogien, 2010). This issue finds possible answers in the PVT’s collective preferences as well as in SIA’s inclusive definition of impacts.

The two other matters we relate to performativity, namely retroaction (Desrosières, 2015) and socio-technical arrangements (STAs, e.g. Callon, 2006) also find applications in the PVT triangle (with e.g. retroaction implying actions to optimise operationalisation, possibly to the detriment of public value creation). More crucially, they contribute to refine our understanding of how performance measurement models and indicators can be developed and evolve (through STAs) and how they can be utilised as legitimacy-generating devices and generate possible retroactions.

Ultimately, Jany-Catrice's performance regimes (2012) and their two delineations (the planned or negotiated nature of processes and the unidimensionality or multidimensionality of evaluations) are also aligning well with the PVT triangle: the planned versus negotiated variable can indeed be seen as an almost immediate proxy for PVT's legitimacy (i.e. a negotiated performance criterion will carry more legitimacy and will benefit from an increased acceptance). This is illustrated in e.g. Jany-Catrice (2008) through the development of local composite development indicators and constitute an interesting entry to understand how SIA can gain acceptance among its stakeholders through the inclusion practices it promotes.

5.2 Operationalising our three theoretical anchors in our research

In this section, we prime the methodology presented in Chapter 6 by introducing how each of our three theoretical entries are operationalised to deliver insight on performance measurement. We relate them to the critical realist (CR) posture we adopt in Chapter 6, which is characterised by three “layers” of reality (the empirical, the actual and the deep real). As CR strives to uncover the deep real, we introduce PVT as our main charting tool to probe the “*hidden mechanisms*” resulting from strategic interactions that our research aims at surfacing. Inquiring the deep real is about looking for causal mechanisms: are WISEs measuring their performance only to comply with regulation? Are they guided by a wish to improve their public value?

5.2.1 Charting the empirical: SIA attributes, PAM performance and performance regimes

The work carried out on SIA in Chapter 4 provides us with a first entry door into our empirical material. As we focus on performance measurement and the way it can mediate the interactions between WISEs and public actors, the four attributes and 13 variables we retained from the literature allow us to chart our cases’ own understanding of what constitutes their impact. The performance framework we presented for PAM in Chapter 3 is also used to reinforce our understanding at this first level of analysis, as we seek to comprehend how WISEs see themselves as providers of public value and how they believe they are perceived by their public stakeholders. We follow a similar process for the WISEs’ public partners, where we delineate their understanding of what constitutes a WISE performance using the same tools.

In parallel, we use Jany-Catrice’s four performance regimes (Chapter 3) as a map to plot the different measurement methods we come across in our empirical material. Here, our decision to use her framework as a plane rather than a table allows for more granularity in our analysis (cf. Figure 6). While some of the information encompassed in the performance regimes is already encapsulated in the way we designed our SIA attributes, Jany-Catrice’s framing of wider models (the regimes) allow to gain more insight on the topics at hand and fit particularly well when applied at the macro level of PAM models (the NWS and the NPG).

Applying our findings from Chapter 3 and Chapter 4 as described above caters to the empirical (primarily) and the actual levels of the critical realist ontological framework (cf. Chapter 6), informing about how WISEs' performance is self-perceived as well as interpreted by their public stakeholders.

5.2.2 The role of public value theory

PVT is the scheme that allows us to dive deeper in the *actual* level of the critical realist's ontological framework and provides a scaffold to apprehend what we assume is happening in the deep real (cf. Chapter 6 and Figure 11). PVT and its focus on strategy provides a unique schematic to understand what is going on in the minds of the stakeholders we research and explain the dynamics (the strategic interactions) we see unfold in the field.

We recall that in Chapter 2 we identified three entries to each of the PVT strategic triangle vertices: deliberation processes, legitimacy devices and endorsement for the legitimacy vertex; agency, organisational design and activities for the operationalisation vertex; and public value (co-)definition, (co-)creation and recognition for the public value vertex. The use of these analytical entries (at the micro level), completed by the literature presented in the rest of Chapter 2 (PAM models and administrative profiles at the macro level) help us unearth valuable insight from our field data.

We follow Bryson et al.'s recommendation to figuratively place a variety of elements (such as actors, practices, arenas and spheres, public problems, or challenges and functions) at the centre of the PVT triangle (2017; 2015b). By the "*centre of the triangle*", the authors mean the topical focus on which the PVT analysis is deployed and invite researchers to consider more analytical topics than *just* public managers (PVT's traditional area of investigation).

In our research, our PVT analysis is focused on WISEs (as an actor) and it seeks to understand the role of performance measurement (a practice) in their relation to their public stakeholders (a second type of actor: municipalities in Denmark and the Direccte in France). In Bryson et al.'s words, these three elements are what we put at the centre of our triangle, as we seek to analyse the strategic interactions WISEs entertain with their public stakeholders

around performance measurement issues. We use the PVT triangle and the nine analytical entries we defined to identify “*observable regularities*” (Kaplan, 1964) that help us chart the actualised reality and unravel deep real mechanisms.

For instance, our research shows that the use of PVT helps us understand that some WISEs may intend to use performance measurement as a device to increase their legitimacy, expecting an improvement (or at least a stabilisation) of the operational environment provided by their public stakeholders in return. Concurrently, other WISEs may use it primarily as a management tool to monitor their activities, with implications mainly in the definition and the identification of public value. Others may want to use performance measurement to garner a competitive advantage, translating into a more favourable operational vertex. These **intents**, identified through PVT, help us understand how performance measurement is thought out and operationalised by WISEs and how it can influence strategic interactions with their public stakeholders.

The above illustrates how PVT provides us with the tools contributing to generate the narratives that allow us to make sense of our data and understand the dynamics at play in the field, beyond what is empirically observable. Combining this reading to the NWS and NPG PAM models deepen our comprehension of our observations in their context. For instance in Denmark where NPG is dominant, emphasising a culture of dialogue and consensus, we understand that devices may not be the best option to reach legitimacy, which may be better achieved through the deliberation processes allowed by the openness of the NPG environment.

The ambition here is to use the PVT triangle as a prism to understand and break down into cohesive bits of information (Kaplan’s observable regularities) the strategic interactions around performance measurement issues observed in the NPG and the NWS between WISEs and their public stakeholders. This process is operationalised mainly by the means of a coding grid, presented in Chapter 6.

Before proceeding to the delineation of our research questions, we present in Figure 10 the overarching analytical framework of our thesis, representing the different scaffolds we use at the different levels of our analysis (micro, and macro) and for the different strata of the CR ontology. We note that this depiction is indicative and is not aimed at providing a precise representation of the levels at which each single concept is operationalised (e.g. the performance regimes are all assessed at the three levels of the CR reality but they are primarily used at the macro level; and the public performance criteria are also marshalled at the empirical level).

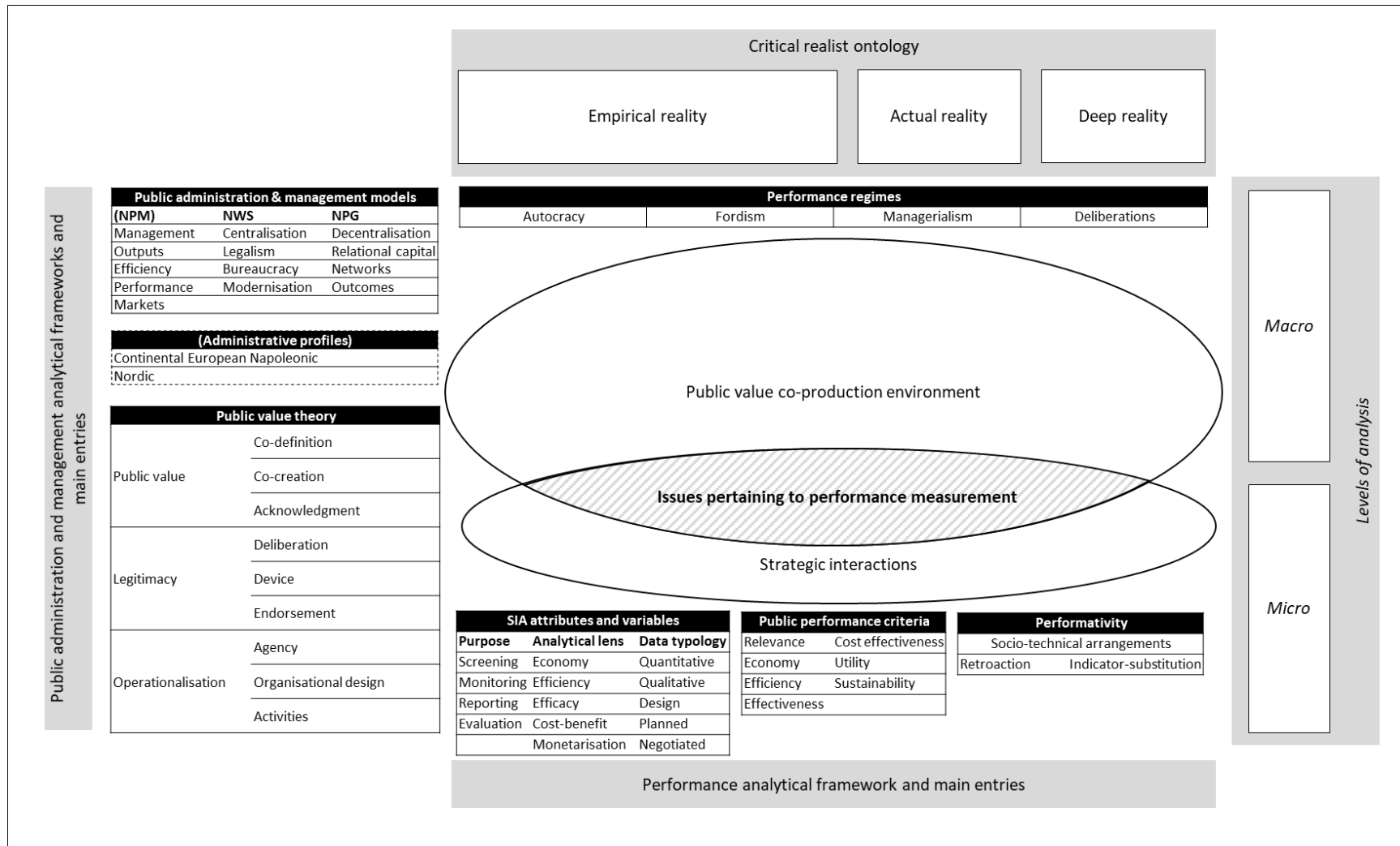


Figure 10: Analytical framework

Source: Author

5.3 Delineating our research questions

Several existing contributions are exploring the relationships between social enterprises (SEs) and their public stakeholders. E.g. Defourny & Nyssens (2010), A. Nicholls (2010), Nyssens (2006) or V. A. Pestoff (2008), are describing the role of partner that SEs are taking in the execution of labour policies, presenting some of the mechanisms government deploy to foster their development. Some write on the place of SEs in identified PAM paradigms such as NPM (e.g. Nicholls, 2010; Mason, 2012, describing SEs as welfare subcontractors) or NPG (e.g. Defourny, Hulgård, & Pestoff, 2014; Pestoff et al., 2013; Ridley-Duff & Bull, 2015, presenting SEs as contracted long-term partners). Some other develop about these issues with a specific focus on WISEs (e.g. Battilana et al., 2015; Petrella & Richez-Battesti, 2016).

A popular approach in such contributions is to use the theoretical frameworks offered by institutional theories such as isomorphism (DiMaggio & Powell, 1983), institutional work (Lawrence, Suddaby, & Leca, 2009) or institutional pluralism (Kraatz & Block, 2008). Publications offering this kind of slant include Battilana et al. (2015), Curtis (2008), Mason (2012), A. Nicholls (2010) or Osborne (2010).

We endorse these contributions' approach and build upon them in our research. And while we also acknowledge the relevance of the institutional frameworks to bring further insight into the topics we are investigating, we wish to explore other, less travelled alleys in our thesis. As such, and as we demonstrated above, PVT and the strategic angle offered by the public value triangle provide an interesting and relevant platform for exploring the issue we are interested in: the strategic interactions of WISEs with their public stakeholders and their incidence on performance measurement within a PAM context.

While some existing PVT contributions nod at social enterprises and include them in the list of organisations that could benefit from applying the PVT triangle (e.g. Bryson et al., 2017; De Jong et al., 2017), their number remains limited. We attribute this scarcity to the initial framing of PVT by Moore (1995), who restricted it to the public sector. Suggesting the use of PVT outside of the public sector is a more recent trend (see e.g. Benington 2011 or Bryson et al. 2015b). Consequently, the approach we are taking ascribes our thesis to some the most recently suggested research agendas: as proposed by Bryson et al (2017, 2015b), we widen PVT's scope of inquiry to new actors (WISE) and practice

(performance measurement), while making use of the techniques (comparative case studies, cf. Chapter 6) and research purpose (range of applicability) suggested by Hartley et al (2017). We carry out this novel approach while resting on the solid foundations of PVT and the rest of the entries we detailed above in PAM, SIA and performance measurement in general.

The context we brought forth in Chapter 1 shows how increasingly relevant SEs are to the delivery of public good and therefore become **co-creators of public value**. For WISEs, this translates into a heightened importance for the delivery of welfare services. While this trend took shape in the midst of the reforms inspired by NPM, recent findings tend to show that it will endure as NPM is tapering off and as new paradigms such as NWS or NPG have emerged (see e.g. Bryson et al., 2014, 2017; Defourny & Nyssens, 2013; Howlett, Kekez, & Poocharoen, 2017; S. P. Osborne, 2010; Ridley-Duff & Bull, 2015). SEs and WISEs indeed still have a relevant place in these paradigms as these new forms of PAM seek to devolve more space to civil society (each in their own way).

A first step in our investigation is to understand how different PAM contexts may translate into different modes of public value co-production. Variances are already hinted in our Chapter 2 when introducing the NWS and the NPG. On this basis, with our focus on performance management, we can raise the question of how different PAM contexts and public value co-production environments may translate into different approaches to performance measurement (**RQ1**). Here our line of enquiry is concerned with the way PAM models, associated to administrative profiles (the imprint of local culture and traditions in the public administration, Kuhlmann & Wollmann, 2014), are conducive to specific approaches of performance measurement at a macro level. We ask this question both for public stakeholders assessing WISEs' performance but also for the way WISEs envision their impact. Jany-Catrice's performance regimes (2012) are a particularly valuable tool to answer this question.

Because public actors play such a crucial role in WISEs' authorizing environment, and by way of consequence in their operational capacity and in the public value they are able to co-create, we want to understand better the strategic intents that animate the WISEs providing performance measurements to their public stakeholders, whether formally or informally (**RQ2**). This perspective ascribes to the micro level, where we wish to comprehend how strategic interactions shape the

purpose of performance measurement, what that implies for the measurement carried out and whether this leads to common or disjoint understandings of performance and its measurement among WISEs and their stakeholders. To answer this question, we make an intensive use of PVT (Moore, 1995) to characterise the strategic intents and we mobilise the framework we propose in Chapter 4 to identify the key components of the observed performance measurement instruments.

This brings us to our last inquiry: as co-creation occurs between WISEs and their public actors, and as these interactions generate questions about performance and its measurement, one may wonder whether these mechanisms, fostered by a wavering NPM and the dissipation of purely output-based performance measurement, can be conducive to the limitation of issues that plagues performance measurement (**RQ3**). Here we of course refer to our considerations on performativity, which encapsulate a wide array of problems that may arise with performance measurement, from democratic considerations (Desrosières, 2015) to behaviours focused on indicators and hindering the production of public value (Ogien, 2010). Building on the macro considerations of our first question (how the PAM environment shapes the general approach to performance measurement) and the micro reflections of our second question (the strategic intents explaining performance measurement behaviours), we look for whether strategic interactions can limit retroaction or indicator-substitution and how it affects the enactment of existing and emerging STAs (Callon, 2006).

These three questions help us reach a degree of thoroughness in our comparative study by (i) providing a macro understanding of performance measurement across contexts and stakeholders, (ii) offering a micro level of analysis, focused on WISEs and their strategic intent when measuring performance, and (iii) by delivering a view of how these two levels of considerations play out in terms of performativity issues in different contexts.

Research question: How strategic interactions between WISEs and public actors are contributing to frame the measurement of performance in different PAM contexts?

RQ1: How different PAM contexts translate into different performance regimes?

RQ2: What are the strategic intents of WISEs measuring their performance and what are the implications for the way this measurement is done?

RQ3: Are strategic interactions related to performance measurement conducive to the avoidance of performativity issues?

Conclusion of Chapter 5

In this chapter, we show how the three entries we adopt from the literature (contributions on PAM, on performance measurement and on SIA) build upon each other to provide an original frame of analysis that allows us to delve further into our initial questioning.

The following chapter describes the critical realist ontology which we allude to in this chapter and thereby provides the paradigmatic and methodological framework that we use to anchor our research and answers the questions we raised above.

Chapter 6 Methodological framing and empirical material

This chapter introduces the way we frame our research methodology and presents our case studies as well as the other sources that inform our work. Numerous contributions to economics insist on the relevance and the importance of carefully framing all the methodological considerations pertaining to a research piece (see e.g. Backhouse, 1994; Dow, 2002; Lawson, 2003; Starr, 2014). Choosing a methodology, and anchoring it in a clear ontological and epistemological framework is an invaluable compass to carry out the analysis of our data and identify pertinent findings.

We therefore start this chapter with the introduction of our paradigmatic approach (i.e. we articulate clearly the world view that guides our research and our analysis). In this section, we develop about our choice of the **critical realist** (CR) framework (Bhaskar, 1975; Lawson, 2003). We introduce critical realism's key concept of layered realities: the empirical (the facts as they are experienced), the actual (the facts as they occur) and the deep real (what causes the occurrence). As CR seeks to identify causal mechanisms in the deep real, we explain how we do so using a retroductivist reasoning method (often equated to abduction, where we go back and forth between our field data and interpretative hypotheses).

In a second section, we present how we operationalise this paradigm in our methodology. We explain the **societal approach** (Maurice, 1989) through which we decide to carry out our international comparison (a balanced approach ensuring a certain linearity between our units of analysis, while keeping them rooted in their local context). We develop about our initial choice of carrying out a mix of quantitative and qualitative research (a particularly good fit for CR) and how we ended up focusing more on qualitative research as a result of challengingly low response rates in France.

After we provide details on the way we collected and analysed our data in NVivo 12, we move on to present our empirical material in a last section: four case embedded case studies and one survey carried out in Denmark.

6.1 Paradigmatic approach

This first section provides the details motivating the choice of our paradigmatic approach: Bhaskar's **critical realism** (CR). We start by contextualising the place of the research's paradigm, before introducing CR, its layered ontology and its retroductive reasoning (a form of abduction). As we present an overview of CR's application to the field of economics, we also dedicate a segment of this section to the place of mixed research and qualitative research in the field of economics: powerful but underutilised approaches (Starr, 2014).

6.1.1 Choosing a paradigm

The paradigm sets out the scientific framework in which the research is undertaken. Guba and Lincoln (1994, p. 105) define a paradigm as *“the basic belief system or worldview that guides the investigator, not only in choices of methods but in ontologically and epistemologically fundamental ways”*. Articulating the paradigmatic approach of a project allows the researcher to structure its work in a coherent manner while being aware of any implicit hypotheses that might impede the clarity of the research (Allard-Poesi & Perret, 2014).

Guba and Lincoln break down paradigms into three questions (1994, p. 108):

1. The ontological question-or what is reality
2. The epistemic question –or what is knowledge
3. The methodological question –or what is the protocol for valid knowledge

While these three questions represent the core dimensions covered by most of the research investigating research paradigms, it is not uncommon to find extra dimensions, such as axiology (or the effects of knowledge) (Allard-Poesi & Perret, 2014).

Choosing a paradigm allows to answer the questions raised in these three dimensions. Guba and Lincoln (1994) summarise the perspectives brought by the main paradigms through the opposition of the positivist and constructivist perspectives (cf. Table 7). Positivism is characterised by a fact-based production of knowledge, where the assumption is that reality can always be described by logical and / or mathematical relations (Beitone, Cazorla, Dollo, & Draï, 2013). Positivism therefore stipulates that

the understanding of reality (as the object of knowledge) is independent from the researcher (the subject of knowledge) (Allard-Poesi & Perret, 2014). It is opposed to constructivism in the sense where constructivists consider that reality is unreachable by the human mind and that knowledge is a construct shaped by the many forces at play in a studied phenomenon.

Table 7: Paradigmatic dimensions

Dimension	Positivism	Critical realism	Critical theory	Constructivism
Ontology - What is reality	"Real" reality, but apprehendable	"Real" reality but only imperfectly and probabilistically apprehendable	Virtual reality shaped by social, political, cultural, economic, ethnic, and gender values; crystallized over time	Local and specific; constructed realities
Epistemology - What is knowledge	Dualist / objectivist; findings are true	Modified dualist / objectivist; critical tradition / community; findings are probably true	Transactional / subjectivist; findings are value-mediated	Transactional / subjectivist; findings are created
Methodology - What is the protocol for valid knowledge	Experimental / manipulative; verification of hypotheses; chiefly quantitative methods	Modified experimental / manipulative; critical multiplicity; falsification of hypotheses, may include quantitative methods	Dialogic / dialectical	Hermeneutical / dialectical

Source: adapted from Guba and Lincoln (1994, p.109)

Seeking a balance between the positivist and the constructivist approaches, our research will adopt a **critical realist** (CR) posture (Archer, Bhaskar, Collier, Lawson, & Norrie, 2013). According to Alvesson and Sköldbberg, CR *"seeks to identify those deeper lying mechanisms which are taken to generate*

empirical phenomena” (2009, p.40). CR shares the essentialist view of reality stipulated by positivists (i.e. that reality indeed exists in its own essence) (Allard-Poesi & Perret, 2014), but contrasts it starkly with the assertion that reality goes beyond empirical events as experienced by individuals (i.e. by opposition to the Humian-informed view of reality held by empirical realists).

When introducing CR in 1975 through *“transcendental reality”*, Roy Bhaskar stresses that the scientific focus has to shift from the epistemological standpoint towards the ontological standpoint. Within ontology, Bhaskar insists on the exploration of the underlying mechanisms rather than the observable events. This view is summarised by Bhaskar as the *“epistemic fallacy”*, or the blending of ontology and epistemology that tends to favour the epistemic view and shade ontological considerations. Figure 11 provides an overview of the stratified ontology (the layered realities) suggested by CR: the empirical are the facts as they are experienced (e.g. a leave is flying), the actual is the facts as they occur (the wind is carrying the leave) and the deep real is what causes the occurrence (air masses of different temperature are causing a depression).

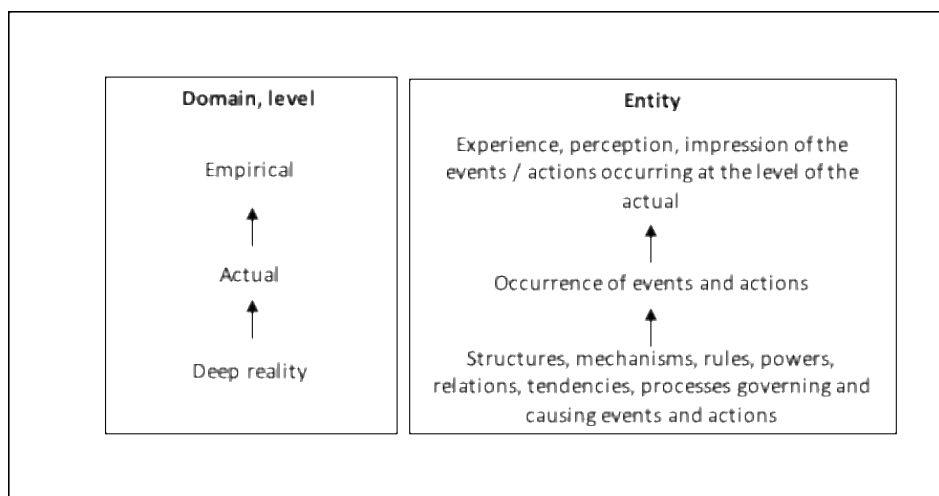


Figure 11: Stratified ontology of critical realism

Source: Adapted from Robert Delorme (1999)

Tony Lawson has championed the application of CR in economics (Hédoin, 2010). Lawson’s reflexion on the application of critical realism to economics is two-pronged (Lawson, 2003):

1. He challenges the view of standard economics theory and its tendency to formalise findings into mathematical models, arguing that models are often expressed as closed systems whereas in

reality the observed systems are open and have to be closed artificially for the purpose of modelling.

2. He proposes an alternative approach to social ontology, informed by the principles of layered realities illustrated above, stressing that social reality is structured dynamically and can therefore not be captured in the way positivist economic models assume.

From an epistemic standpoint, placing our work in a CR paradigm implies that our reflexion will be at the crossroads between objectivist and relativist standpoints. An overarching consideration in realist paradigms, including CR, is that knowable reality has a meaning of its own, which is independent from the preferences of its observer (Le Moigne, 1995). We therefore acknowledge our interest in finding causality and identifying patterns in the deep reality (Archer et al., 2013), but we bear in mind that causal mechanisms *“exist only in virtue of the activities they govern and cannot be empirically identified independently of them”* (Bhaskar, 1979, 48). This means that because of the openness of social systems, CR’s causality is not understood in its Humean regularity of *“if x then y”*. Rather, with CR we strive to find the (fallible) best explanation of reality (Fletcher, 2017).

The CR approach in economics is thus very much defined by opposition to orthodox economic theory. We acknowledge that this causes CR to be challenged by a number of economists (see for instance Hodgson, 2004 or Mearman, 2006). The analytical framework provided by CR however offers unique features that make it especially relevant to investigate economics from a perspective such as the one we adopt out of public value theory. The defining feature of this framework lies in the non-atomicity of economic agents and the necessity to consider their actions in *“open systems”*, in conjunction with the social fabric that surrounds them and the inherent layers of reality that underpin their behaviour (Fleetwood, 1999). This framework will guide our methodological choices as set forth below.

6.1.2 The research object in the critical realist paradigm

Following a CR approach, we will focus on unearthing the deep reality, or the mechanisms and structures embedded in the deep reality, which are responsible for the observed occurrences (Fleetwood, 1999). While this approach differs from positivism in the ways and means research will be carried out (Sayer, 2004), it is still compliant with the essentialist way of constructing a research object

(Allard-Poesi & Maréchal, 2014) as illustrated in Figure 12. In the CR view, the emphasis is put on the third level (bottom) in the illustrated framework, where the valid scientific knowledge is produced (Archer et al., 2013; Fleetwood, 1999).

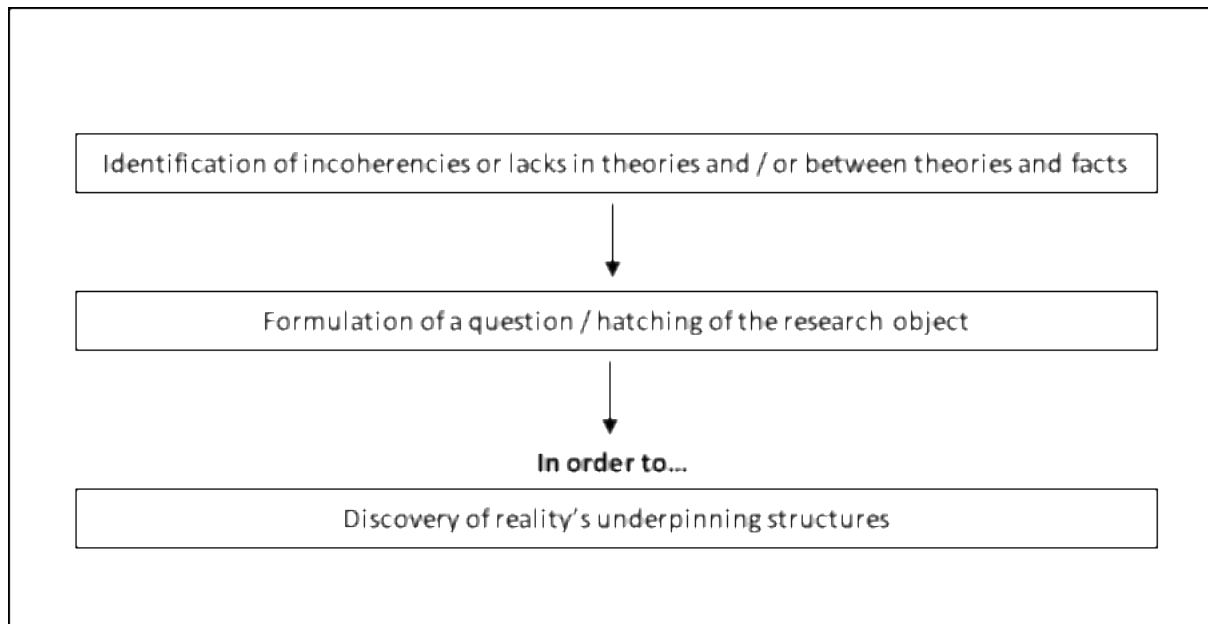


Figure 12: The research object in the critical realist approach

Adapted from Allard-Poesi and Maréchal, 2014 p.55

6.1.3 The reasoning method

Logical reasoning can be characterised by two different approaches: deductive reasoning and inductive reasoning (Charreire Petit & Durieux, 2014). Deduction is presented by Grawitz (1993) as mean to demonstrate. It essentially consists into a top down approach, where the researcher goes from the formulation of general laws or hypotheses towards testing them against reality and see how true they hold.

Induction is characterised by a reversed approach, where the formulation of hypotheses and general laws is inferred from the observation of reality (Chalmers, 2013). Here, the researcher goes from the observation of isolated phenomenon to the coining of hypotheses aiming to outline a general explanation for the observed phenomenon.

While one might be tempted to see these two approach to logical reasoning as opposed, they often appear complementary in the research process (Charreire Petit & Durieux, 2014). Figure 13 provides a good overview of how the two paradigms complement each other.

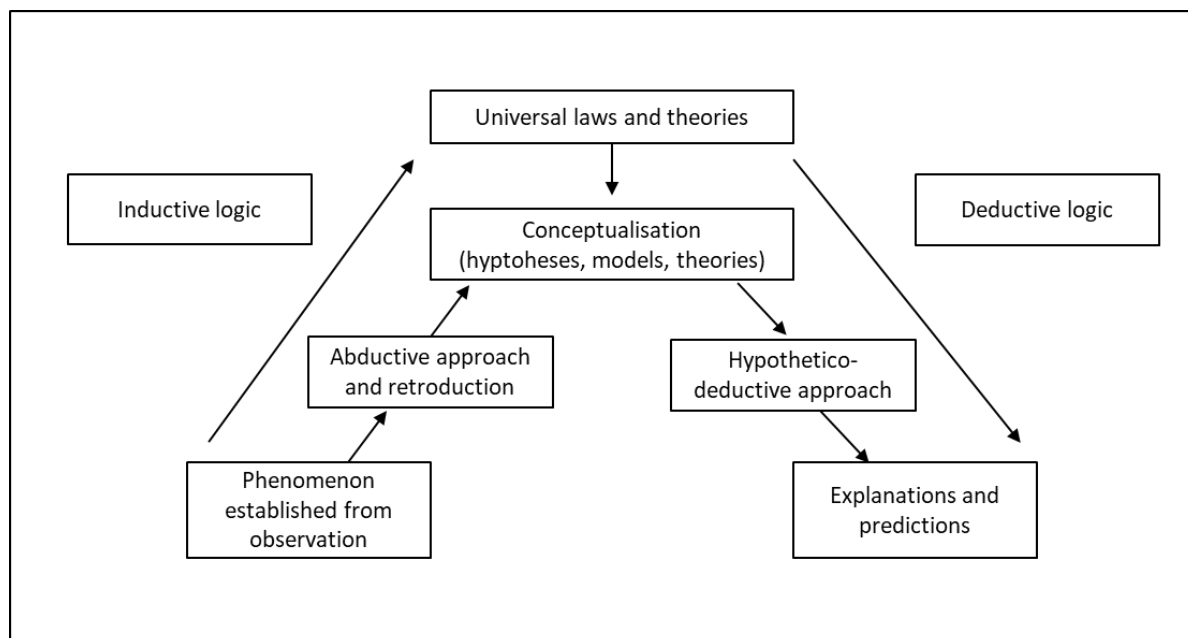


Figure 13: Logical reasoning

Adapted from Charreire Petit & Durieux (2014, p. 82)

Due to our anchoring in CR, our research will adopt retrodution. According to Lawson (2006), retrodution is a reasoning method consisting in inferring conjectures from the observation of reality; these conjectures being subject to ulterior discussion and trials. Retrodution is the reasoning method favoured by critical realists. In the CR literature, the terms “*retrodution*” and “*abduction*” are often used interchangeably (Lawson, 2006 p.24; 2003, 145; Roy, 2015)⁵³.

Retrodution allows to go beyond empirical observations, towards the identification of the mechanisms and structures underpinning them. To illustrate this, Lawson uses the following image: “*If deduction is illustrated by the move from the general claim that ‘all ravens are black’ to the particular inference that the next one seen will be black, and induction by the move from the particular*

⁵³ Some critical realists such as Fletcher (2017) or W. Olsen (2007) differentiate between abduction (an inference of facts to likely theories explaining the facts) and retrodution (the conditions under which the inferred theories seem to apply) while Lawson’s definition of retrodution seem to cover both aspects. Lawson takes inspiration from the American philosopher Charles Sanders Peirce, who used abduction, retrodution as well as “*hypothesis making*” interchangeably in its writings.

observation of numerous black ravens to the general claim that “all raven are black”, retroductive or abductive reasoning is indicated by a move from the observation of numerous black ravens to a theory of mechanism intrinsic (and perhaps also extrinsic) to ravens which disposes them to be black” (Lawson, 2006 p.24).

6.2 Methodological approach

Our methodological choices echo the paradigmatic approach we have adopted and are aligned with the features of our research topic. We first explain our choice of an international comparison carried out under the angle of the societal approach (Maurice, 1989). We then develop about our initial choice of a mixed qualitative and quantitative method (and the limits we encountered), before moving to the research design and the data collection. We conclude by explaining how we will carry out our analysis with the use of NVivo 12.

6.2.1 Aligning the purpose of the research with the case design: the societal approach

Our research investigates how the strategic interactions between WISEs and their public stakeholders are contributing to frame the measurement of performance in different PAM contexts. Since we are preoccupied by the interplay between WISEs and public actors, the context in which it occurs (i.e. the Public Administration and Management, or PAM, setting) is crucial to make sense of our research. For this reason, we choose to make an international comparison (Dupré, Jacob, Lallement, Lefèvre, & Spurk, 2003; Hassenteufel, 2005; Maurice, 1989) and undertake our research in two different countries, characterised by two different PAM models: Denmark, a country applying a New Public Governance framework (Osborne, 2006), and France, a country applying a Neo-Weberian framework (Pollitt & Bouckaert, 2004). Carrying out this research in two different PAM settings aims at identifying potential variances in the way strategic interactions between WISEs and their public stakeholders may affect performance measurement.

Marc Maurice (1989) identifies three types of international comparison. The first one, which he dubs “*cross-national*” is characterised by a functionalist approach, where research topics are compared from one country to the other with a form of implied continuity. That is to say the researched phenomena are scrutinised on the basis of their functions (what do they, how do they operate), but without linking these observations to the national context in which they are made. Maurice speaks of a disconnection between the macro level (the context) and the micro level (the research phenomena), where the phenomena are compared directly (continuity). International rankings such the OECD’s PISA, which leaves little room for contextual information, epitomise this type of approach.

A counter point to this approach is the “*inter-cultural*” comparison, which focuses on understanding the researched topic in its context and derives a characterisation of this topic almost exclusively from its context. Here the focus is on the linkage between the macro and micro levels. Because of this emphasis on embeddedness, there is little to no continuity on the researched phenomena, which can paradoxically hinder the initial effort of comparison. Maurice illustrates this approach with the work of Abegglen (1959) on Japanese industrial organisation, adopting a culturalist standpoint.

Lastly, and this is the approach proposed by Maurice, the **societal approach** (1989) aims at realising a synthesis between the two extremes presented above. Striking a balance between the embeddedness of the micro in the macro and the continuity (at the micro level) of the research object, Maurice talks about “*actors and spaces*” that jointly develop within societies. As the author insists on the interdependencies between the macro and micro levels, he develops about how actors and environment mutually influence each other in a way that is close to the strategic interactions we are interested in.

We highlight that while Maurice focuses on the micro and the macro levels, our research also encompasses stakeholders that do not clearly belong to one or the other level (e.g. network organisations, consultancies, certain government agencies and WISEs other than our case studies). These stakeholders are not part of our case studies, but they are part of “*actors and spaces*” that provide additional contextual information in our research and allow us to better understand the interaction between the macro and the micro levels. In France, this is particularly important as WISEs networks play an important interface role with central authorities.

Lastly, Maurice also introduces elements of interest in terms of methodological considerations: “*From a methodological standpoint, the interdependency processes by which actors and spaces develop exclude the existence of a linear causality, rather, they suggest the existence of a multiple and interactive causality*”⁵⁴ (1989, 183). This indicates a compatibility between the societal approach and

⁵⁴ « D'un point de vue méthodologique, les processus d'interdépendance à travers lesquels se construisent acteurs et espaces excluent l'existence d'un principe de causalité linéaire, évoquant plutôt l'existence d'une causalité multiple interactive »

CR's ontology, which is interested in causal mechanisms while being acutely aware of their fluctuating nature.

6.2.2 Qualitative, quantitative and mixed research

Traditionally, research methodologies are separated in two approaches: qualitative methods and quantitative methods. While these two perspectives have often been opposed in the way they address scientific research (Hammersley, 1992; Silverman, 2006), there is a growing tendency to acknowledge the complementarity of the two, and embrace them both for their respective slants (Glaser & Strauss, 1967; Hammersley, 1992).

Baumard and Ibert (2014) highlight that an essential part of the distinction between the two approaches stems from the kind of data (qualitative or quantitative) each method typically deals with. At the most basic level, Miles and Huberman (1994, p. 246) describe qualitative data as "*words rather than figures*". Evrard et al. (2009) refine the boundaries of this description and define qualitative data as variables measured on nominal and ordinal scales while quantitative data is variables measured on interval scales and in proportions. It should also be highlighted that it is often possible to apply a statistical treatment to qualitative data (Baumard & Ibert, 2014).

Another distinction factor between qualitative and quantitative approaches is the research orientation: exploratory research will in general be more prone to adopt a qualitative approach while confirmatory research will be more likely to tap into quantitative methods (Silverman, 2006). However, separating the two approaches entirely would often prove irrelevant, as a mixed approach is frequently employed to carry out research (Baumard & Ibert, 2014; Glaser & Strauss, 1967). The combined use of the two is even encouraged in exploratory research: "*The triangulation made possible by multiple data collection methods provides stronger substantiation of constructs and hypotheses*" (Eisenhardt, 1989, p. 538).

6.2.3 Mixed research design

The concept of triangulation (cf. Figure 14) is further explained by Bouchard (1976) as a way to safeguard research against discoveries that would only be the reflection of the method used. Mixing qualitative and quantitative approaches both allow to extend the exploratory field with a larger breadth

Onwuegbuzie, & Turner, 2007 p.119). Figure 15 presents an overview of the different overall designs for mixed research.

		Time order decision	
		Concurrent	Sequential
Paradigm emphasis decision	Equal status	1. QUAL + QUAN	2. QUAL → QUAN QUAN → QUAL
	Dominant status	3. a. QUAL + quan b. QUAN + qual	4. a. QUAL → quan b. Qual → QUAN c. QUAN → qual d. quan → QUAL

Figure 15: Mixed research design

Adapted from Johnson & Onwuegbuzie, 2004, p.22

“Qual” stands for qualitative, “Quan” stands for quantitative, capitalisation indicates dominance, “+” indicates concurrence, “→” indicates sequence order

Our research design is to carry out qualitative research and quantitative research in sequence, with an emphasis put on qualitative research (dominant) but a quantitative research carried out first. This corresponds to configuration 4.d. in Figure 15. In the triangulation configuration (cf. Figure 14) insights from both research strands inform each other, with the cross fertilisation of the data happening once all field research has been carried out.

6.2.4 Research design in economics

In the field of economics, mixed and qualitative research are more the exception than the rule. While a growing number of economists argue for an increased use of qualitative tools, this kind of approach is, to this day, still considered marginal in the discipline (see e.g. Bewley, 2002; S. C. Dow, 2007; Lawson, 2006; Starr, 2014). The place for more qualitative work in the field of economics is however not a new debate, where a lot of early arguments were concerned with the quality of the observations made to

feed quantitative analysis and modelling (see for instance Dewey, 1910; McGoun, 1936). While not advocating directly for qualitative research, these contributions worry of the excessive focus on the quantity of indicators produced rather than on their quality: *“As a nation we are altogether too credulous, and have yet to learn that columns of figures are not statistics.”* (Dewey, 1910, 44).

This is core to the value that qualitative research can bring to the discipline of economics: qualitative research can convey an intimacy with the research object that will allow a meticulous and thorough understanding of the phenomenon observed that quantitative research, typically desk-based, cannot offer. In fact, some seminal economic ideas have been developed out of what can easily be qualified as qualitative research, such as Adam Smith’s famous pin factory visits, which allowed him to articulate pivotal notions about the division of labour. Out of this example, qualitative research is remarkably fitting to explore research subjects and identify new leads for theory building.

Starr illustrates particularly well the place of qualitative research in economics: *“Although economists often think of qualitative research as involving words and quantitative research as involving numbers, a more valuable way of thinking of the distinction is in terms of open- vs. closed-end approaches to gathering data. Thus, in quantitative studies, researchers gather or use data with the expectation that they know in advance a fixed set of dimensions along which the data should be characterized (e.g. income level, daily share prices, given instruments of monetary policy); in qualitative studies, researchers ‘proceed to the field’ with clear and detailed guidelines as to what issues they want to investigate and how, but expecting their interaction with research subjects and/or their broad review of relevant data to provide the basis for constructing a sound characterization of the phenomena of interest”* (2014, 239).

Adopting Starr’s stance, we fully acknowledge the usefulness of quantitative research to advance some areas of economics. We also find that due to (i) the critical realist perspective we adopt and (ii) the absence of *“fixed set of dimensions”* available (or even readily identified) to frame our research subject, a degree of qualitative approach is preferable to carry out our work.

6.2.5 Qualitative case design

Qualitative research is mostly carried out through case studies, to a point where the terms “qualitative” and “case studies” are often used interchangeably (Eisenhardt, 1989, p. 538). Yin (2009) identifies four main types of case study designs, defined by two dimensions: the number of cases and the number of units of analysis (cf. Figure 16). Holistic case studies are characterised by a systematic approach of the studied phenomenon and are useful to approach situations where sub-units of analysis cannot be identified. On another hand, embedded case studies are valuable to confront rival interpretations or to strengthen the internal validity of the hypotheses developed.

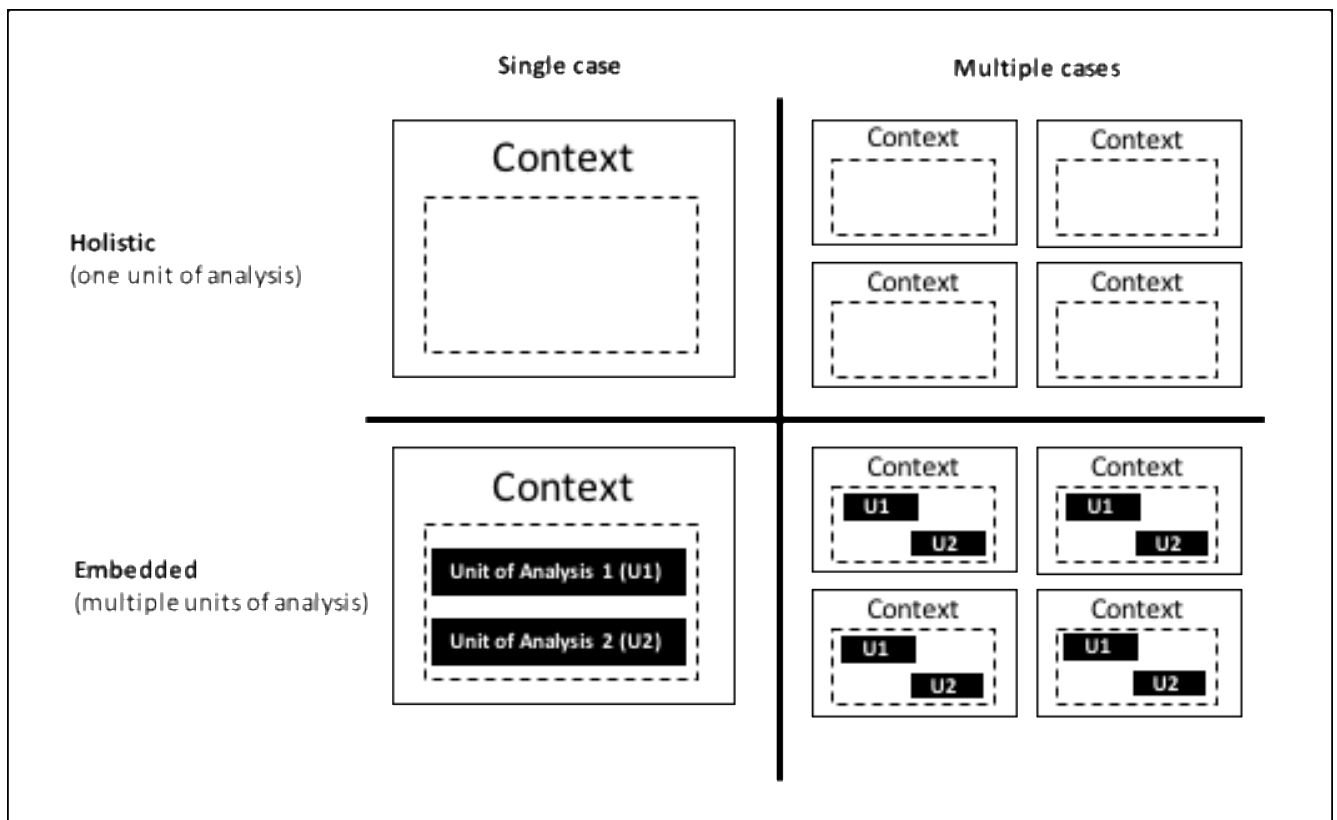


Figure 16: Qualitative case design

Source: adapted from Yin (2009)

As far as the number of cases goes, single-case case studies are valuable in a different array of situations (longitudinal cases, representative cases or unusual cases for instance). Multiple case study designs are useful in comparative studies and help providing a bigger picture of the research object (Robert E. Stake, 2013). Regardless of the case study design, the context is always taken into consideration.

6.2.5.1 Designing the qualitative cases:

As highlighted in the previous chapters, social enterprises are still considered an emergent field in academia and its definition and boundaries are still being actively discussed. Similarly, the discipline of social impact assessment is the object of continuous experimentation from practitioners and academics alike. On another hand, Yin (2009) highlights that case studies are particularly suitable to understand contemporary phenomenon in the actual context they operate. Qualitative case studies therefore seem especially appropriate to carry out research on the topic of WISEs and social impact assessment.

Based, on the presentation of the different case designs drawn in Figure 16, and on the research objectives stated above, we follow an embedded case study design. Table 8 summarises the high-level design of the cases. We aim to study four cases of WISEs assessing their social impact and communicating about it to their public stakeholders. Two of the cases are studied in France and the two others in Denmark. Each case has two units of analysis: the WISE itself and the public stakeholders with which they have performance measurement-related strategic interactions. We recall that in addition to our cases, we carry out a number of contextual interviews: while introducing a degree of useful information, these interviews pursue a goal of environment characterisation and do not constitute a separate unit of analysis in our design.

Table 8: Overall qualitative case design

Country / context	Case	Unit of analysis 1	Unit of analysis 2
Denmark	Case 1	WISE 1	Public stakeholder 1
	Case 2	WISE 2	Public stakeholder 2
France	Case 3	WISE 3	Public stakeholder 3
	Case 4	WISE 4	Public stakeholder 4

Source: Author

6.2.5.2 Qualitative cases selection

The cases were selected using a “purposive sampling” method (Patton, 2005). This approach is often used to study phenomenon such as our research object where existing structured database are not

available (Royer & Zarlowski, 2014). The sampling methodology followed the steps described in Table 9.

Table 9: Qualitative sampling methodology

Stage	Description
1. Identify WISEs in each country	Using local repositories of social enterprises and national associations and federations, we selected a number of entities fulfilling the four criteria to be considered as a WISE according to Davister et al. (2004, 3).
2. Reach out to selected WISEs	Contact was established with the selected WISEs (unit of analysis 1), preferably via email, in order to determine: <ul style="list-style-type: none"> a. Whether they had a form of SIA experience b. Whether they were interested in taking part in academic research.
3. Apply selection filters	Once contact was established, a preliminary interview determined whether the WISE was suitable for the case study. The selection filters were as followed (all questions to be answered positively): <ul style="list-style-type: none"> a. Confirm that the enterprise's activity falls into the WISE's scope b. Establish that the WISE has carried out SIA (preferably) or have a strong interest in doing so c. Determine whether the WISE is open to provide access to its public stakeholders d. Ensure that the WISE is willing to invest enough time for the research to be carried out thoroughly
4. Find out and contact the public stakeholders relevant to the case	Contact was established with the public stakeholders (unit of analysis 2) relevant to the case in order to determine whether they were interested in taking part in the case study.
5. Confirmation	Both units of analysis saw their participation in the case study confirmed.

Source: Author

Stages one through four were to be repeated iteratively until the desired number of cases was reached. In practice, stage four was never a cause for case rejection (i.e. public stakeholders never refused to take part in the research). Here, we however have to highlight the difficulties experienced in identifying cases, in both countries. Our case selection phase ran for one full year, between January 2017 and January 2018 (including the sampling of our quantitative survey described below). We experienced a challengingly low response rate to our solicitations, despite a relatively strong support from existing networks in both countries (although support was variable from one network to another in France). We attribute these difficulties mainly to two factors:

1. A limited number of WISEs have carried out SIA. The main cause for rejection occurred at stage 2.a., where we realised that the majority of the WISEs we contacted had not carried out any form of SIA⁵⁵. At first our criteria here was that WISEs should have implemented a form of SIA. (which was also reflected at stage 3.b.). We had to “soften” this criterion to work on a second Danish case for which we deemed a strong interest expressed towards SIA sufficient to carry our research (bringing a relevant perspective on “why” some WISEs that have a marked interest for SIA do not necessarily execute on that vision).
2. A limited amount of resources and time from WISEs to take part in “non-core business” activities. This is something we observed also with some of the cases we studied: most WISEs are small organisations with strained resources, and participating in research, albeit a compelling idea, is seen as an additional stress on the organisation. This was particularly felt in Denmark, where case rejection occurred at step 3.d. of the selection process in two instances (i.e. where an initial interview took place but where the WISEs decided that they did not have enough time to spare for the research).

6.2.5.3 Qualitative data collection

Our qualitative data was collected through semi-structured interviews (Eisenhardt, 1989; Yin, 1981). We developed three main templates for our interview guides: one for WISEs, one for their public stakeholders and one for contextual interviews, with public or private actors. We modified our templates according to the seniority level of our interviewees (three levels for WISEs and two for public actors). We provide a sample interview guide (WISE, management level, the one we used the most) in Appendix 17.

For each case study, we had a first introductory interview over the phone with the WISE’s top management (CEO or operations’ manager). We then held physical interviews with the top management (CEO and operations’ manager or equivalent) as well with some of the operational staff

⁵⁵ In France, while screening our candidates, we did not associate the reporting mechanisms to the Direccte (e.g. of graduation rates) to a form of SIA, even though these mechanisms are considered a form of performance measurement by the Direccte. WISEs therefore had to have an approach off of their own initiative to be considered as carrying out SIA.

(case workers and intermediate management) as well as with beneficiaries (except for FR1, where we did not have a chance to interview beneficiaries). We then carried out interviews with each WISEs' main public stakeholders (always after the WISEs interviews). In both our French cases, at least one staff member of the WISE was present during our interviews with public stakeholders (we did not perceive this to affect the nature of the exchanges).

Table 10 summarises our interview schedule, which is presented in full in Appendix 9. Our case research spans 67 hours of 62 interviews. 43 hours were recorded and transcribed (phone conversations, as well as some conversations and broader meetings were not recorded). Complimentary interviews were held, mostly over the phone, to clarify some points in our analysis. Sources informing our qualitative research include corporate publications, press clippings, audited accounts and websites as specified in 6.3.3.

Table 10: Summary of interviews schedule

Country	Case / level	Stakeholder	Count	Hours	
France	Contextual	Other WISEs	2	1:00	
		Network	11	5:00	
		Public	5	3:30	
		Investors & consultancy	3	2:00	
	FR1	WISE	4	9:00	
		Public	2	2:30	
	FR2	WISE	2	7:30	
		Public	4	2:30	
	Total France			33	33:00
	Denmark	Contextual	Other WISEs	4	2:00
Network			2	2:00	
Public			4	3:00	
Investors & consultancy			3	2:30	
DK1		WISE	8	10:00	
		Public	2	2:00	
DK2		WISE	4	9:00	
		Public	2	2:00	
Total Denmark			29	32:30	
Europe		Networks	2	1:30	

Country	Case / level	Stakeholder	Count	Hours
Total			62	67:00

Source: Author

6.2.5.4 Qualitative data Analysis

We carried out the analysis of our qualitative data in NVivo 12. This solution allowed to perform our coding with flexibility, while leveraging the statistical treatment capability of the software (we tried the automated patterns recognition but were not pleased with the results). We used statistical treatments to derive word frequencies and correlations on ad’hoc topic. An example of our use of word frequency treatment can be found in Table 32, when presenting some of our final results. While this does not take precedence over the researcher’s reasoning, NVivo brings an ability to centralise all the case data in one place (including secondary sources), structure our entire analysis using a streamlined coding and parse dynamically through large amounts of data with relative ease.

Our data analysis was done following Miles & Huberman (1994) and adopted a semi-structured coding grid in order to identify “*observable regularities*” (Kaplan, 1964). Observable regularities are the basic unit of a scientific inquiry, the foundational bits of information we use to build our analysis and reach our conclusion. We defined most of our codes before our field research (*ex-ante*), on the basis of our literature review, and out of the analytical grid we presented in Chapter 5. Our semi-structured approach however allowed us to adapt our grid to our findings as patterns emerged during our interviews and as we were parsing through the data. This method, allowing us to go back and forth between our empirical material and our theoretical framework, follows the retroductive reasoning we adopt in our critical realist ontology.

We followed the conception of coding expressed by (Allard-Poesi, 2003), where labelling our data is “*only*” one interpretation of our raw material. We thereby embrace the existence of several interpretations for our data, and we endorse the stance that rather than “*decoding*” hidden secrets, we are “*encoding*” our material using our own language. On this basis, we reject the idea of a realist or unique coding.

This brings us to the next item regarding the treatment of our data: as we parsed our interview transcripts, we adopted a plurinomial coding (Ayache & Dumez, 2011). I.e. we applied multiple codes to single quotes when relevant. We can exemplify this process with e.g. the following quote from a Danish municipality, on how they assess WISEs: *“Yes, it [performance] is assessed only through the report, and all the conversation I have had about the beneficiaries as well. Hearing about what they can do, reading, or being more social, or whatever.”* We coded this piece of conversation as part of evaluation logics attached to performance (code 3.1.4) and deliberation as vector of legitimacy (code 2.1.1). Table 11 provides further coding examples, and we provide our full code list in Appendix 16. In both tables, emergent codes (codes appearing during the analytical process are identified *in grey*. NVivo has been particularly useful in this process, allowing to easily keep track of data with multiple codes, updating them as we go without disrupting our data set.

Table 11: Coding examples

Abstract	Coding
<i>“We escalate the information. We try to explain that the track record counts, that the work integration scheme matters...”</i>	1.1.2: Neo-Weberian centralisation logics 2.1.1.3: Emergence of shared preferences through deliberation
<i>“Normally the process is that the job centre provides some goals, some things that they want us to clarify – we want to clarify these competences for this individual. (...) But in some occasion they don’t (...) and then we put our own goals (with the beneficiaries).</i>	2.1.1.: Co-defined public value 2.2.1.4.: Emergence of shared preferences through deliberation
<i>“So that is why the municipality is willing to pay us to execute on these programmes. Because we can document everything. And the last report we fill out includes a dedicated piece for their rehabilitation team. It is the last questionnaire on the document. So that’s why. If we didn’t have that we would not be able to have the contracts. And that is where the money is for us, to allow us to survive as a social enterprise.”</i>	2.2.2.2.: Performance measurement as a legitimacy device emerged from inside 2.3.1.3.: Performance measurement as a competitive instrument
<i>“This is why we stopped the covenants with the county. Because they promise 2,000 euros, we will</i>	1.1.1.1.: Neo-Weberian Bureaucratic logics 4.2.2.1.: Deliberate retroaction

Abstract	Coding
<i>spend three days on the paperwork and at the end we will get 500 euros 18 months later”</i>	
<i>“In our contract we say we can make an evaluation of the programme, ask the candidate. We take the right to do it. So the provider knows that maybe we are going to ask the candidate about the programme.”</i>	3.1.4.: Performance measurement through evaluation 2.1.1.: Public value co-definition 2.2.1.3.: Phasing of deliberation in public value definition
<i>“The job centre has a programme called “young ones on their way”, and if I cannot get them to work they will take them back and put them there, but I think our programme is much better for their development.”</i>	1.2.2.1. Competition against public services in NPG

Source: Author

Miles and Huberman (2003) indicates that first-level codes are meta-codes that encompass the rest of the developed coding through synthesized themes. In order to provide a high-level view of our coding work, and as an introduction to Appendix 16, we describe below the four main codes structuring our analysis, which are hatched around the considerations we bring about in our literature review. The codes are:

1. Public administration and management environment
2. Public value strategies
3. Performance measurement approach
4. Conceptions of performance

In **PAM environment**, we use the characteristics we identified attached to NPG and NWS to typify the environment in which WISEs co-create public value with their public stakeholders. We look for characteristics of each country’s PAM model and its administrative profile, but we are also wary to identify any unexpected features though our emerging coding (e.g. emergence of relational capital in the NWS and occurrences of bureaucracy in the NPG).

In **Public value strategies**, our sub-codes are structured by the PVT triangle and the three entries we identify for each vertex in Chapter 2. We use these nine entries and refine them on two to three

additional levels to parse the observed strategic interactions and understand the role of performance measurement in these interactions.

In **Performance measurement**, we use the four attributes and 13 variables of SIA identified in Chapter 4, in combination with the seven PAM performance criteria identified in Chapter 2 to structure the sub-levels of our coding. We do this with the intent to map out and understand performance practices in the way they are operationalized (with which indicators, through which processes, and with which kind of intent).

In **Conceptions of performance**, we explore the three aspects of performativity we highlighted in Chapter 3. Coding this item requires an additional level of analysis since elements related to performativity do not emerge evidently through conversations (by essence, performativity is rarely a conscious process). Here we rely on the critical realist framework and look for information beyond the empirical level, at least at the level of the actualised. In this entry, we also code the data that we can relate to Jany-Catrice's four performance regimes.

The analysis we make of our data capitalises heavily on our coding and our understanding of its associated content (the quotes behind the codes, of which Nvivo allows us to make trackable notes). We use quantitative information such as occurrence frequencies, word or phrase counts and topical correlations (calculated in NVivo) mostly marginally, to confirm our finding or assure no further interpretations can be drawn from our data set. When delivering the results of our analysis in the second part of our thesis, we combine the findings of our qualitative findings together with the quantitative ones (when applicable, cf. qualitative dominant design in Figure 15).

6.2.6 Quantitative research design

Through the adoption of a mixed-method research, we aimed at carrying out a triangulation following a dominant sequential configuration (setup 4.d. in Figure 15) with our qualitative research as the dominant strand. Although we could not carry out this vision to its term across all our units of analysis, this section describes the initial intent of the quantitative research (as an online survey), how it was

carried out (sampling and data collection) and how the results we collected were analysed and interpreted.

Although we attempted to carry out our survey both in France and in Denmark. We only yielded exploitable results in Denmark. Therefore, our description of the research protocol below reflects the sequence of events in our research, and we provide the full procedure for Denmark while we stop at the sampling and data collection stage in France.

6.2.6.1 Purpose of the quantitative research

In both countries, the quantitative survey was to deliver a first view of the field, providing a descriptive snapshot of SIA practices among WISEs, and procuring insights to be tested while carrying out our qualitative research. This survey did not seek to test any pre-defined hypotheses, but instead to provide clarification on the four following areas:

1. Provide an understanding of how widespread the use of SIA is among national WISEs populations and what are the key practices for those who carry out SIA
2. Understand whether public stakeholders have an influence in the development and application of SIA models
3. How do WISE self-perceive the way their performance should be conveyed, and how do they perceive their public stakeholders' expectations in that matter
4. Is there a gap between WISEs' expectations when doing SIA (expressed by WISEs interested in SIA but not yet having a functioning SIA model) and the perceived results they obtain from SIA (expressed by WISEs having a functioning SIA).

The answers provided to these questions (mostly through descriptive statistics) were then to be tested and tried in our qualitative case research. Chapter 9 provides the answers to these questions for Denmark.

6.2.6.2 Sampling and data collection

Denmark:

In Denmark, we aimed at contacting the largest number of SEs possible, with the idea to filter and identify WISEs *ex-post*, based on filter questions using the criteria we lay out in section 6.2.5.2. As our

survey was hosted online, our primary mean of communication was through email, although we also used social networks and in-person announcements (cf. below). There is no publicly accessible directory of social enterprises in Denmark. A list of Danish organisations registered as social enterprises under the 2014 law is accessible on Virk.dk, the Danish public “business portal”⁵⁶, but email addresses are not always available. Furthermore, in June 2017, 208 social enterprises were registered, which suggests that at least one third of social enterprises in Denmark did not apply for, or were not granted the social enterprise status.

A collaboration with Social Entreprenører i Danmark (SED) was therefore set up to reach out to a maximum of social enterprises in the country. SED is an association that was founded in 2010 by alumni of Roskilde University’s social entrepreneurship track. The association acts as a forum and a networking place for social entrepreneurs in Denmark, with multiple events organised throughout the year. SED has identified 218 social enterprises in Denmark of which 169 have valid email addresses.

Using SED’s database, we reached out to a total of 169 social enterprises, representing at least 171 entities. Of these, 114 were registered social enterprises under the 2014 law. The survey response rate is 18% (31 valid questionnaires filled out, of which 24 present a full dataset). Due to (i) the way SED’s database has been created (personal network and multiple desk researches over several years of operations) and (ii) the restricted number of respondents, there is no guarantee that the sample is representative of the Danish social enterprises population (cf. Table 12). Relatively to the estimated Danish WISEs population, our sample should have a representativeness of approximately 20%. While this may seem like an acceptable ratio, the absolute number of WISEs in our sample (26) is still relatively low to derive a meaningful quantitative analysis.

⁵⁶ A public repository managed by, among others, the Danish Business Authority and The Danish Chamber of Commerce

Table 12: Survey population and sample data

Population subset	Data	Total	Of which are registered social enterprises	Of which are WISEs
National population	Count	300 ⁵⁷	208 ⁵⁸	129 ⁵⁹
	%	100%	69%	43%
SED data	Count	218	155	n/a
	%	100%	71%	n/a
SED data - contacted	Count	169	114	n/a
	%	100%	67%	n/a
Sample (respondents)	Count	31	24	26
	%	100%	77%	84%
	% of population	10%	12%	20%
	Response rate	18%	21%	n/a

Source: Author

The data was collected online using LimeSurvey. The data collection process lasted for 37 days, between May 20th 2017 and June 27th 2017. Although we had the full support and cooperation of SED, it was a rather tedious exercise requiring multiple reminders (one every second week in average) despite the use of SED's social networks accounts (Facebook and Tweeter) and public announcements during SED gatherings. We attribute these difficulties to factors similar to the ones identified as responsible for the limited participation in our qualitative research (cf. 6.2.5.2).

France:

In France, due to the existence of segregated institutions for SEs and WISEs, and thanks to well-established networks for the latter, we chose to rely primarily on these organisations to disseminate our survey, with an objective to target WISEs directly, without the necessity to filter them out from other SEs *ex-post*. French WISEs networks have layered organisations, with a national entity and several regional entities functioning with a varying degree of autonomy. For our research, we worked at the national level with the three principal WISEs networks in France covering 1,700 of the 3,600 French WISEs (cf. Chapter 7): The Fédération des Entreprises d'Insertion (FEI), Coorace and Chantier Ecole.

⁵⁷ Based on The Committee on Social Enterprises, 2013

⁵⁸ Based on the Danish Business Authority / Virk.dk in June 2017

⁵⁹ Based on Thuesen et al., 2013

When requesting support for the dissemination of our survey, there was a shared concern among these networks about a risk of perceived “*information overload*” from their members which, in some networks, have complained about the overabundant email communication or multiplicity of surveys being carried out. This second grievance was particularly aggravating for the FEI and Chantier Ecole, to a point where the former declined to actively promote the survey through e.g. an email campaign and instead passively advertised it on the members’ intranet. The two other networks carried out one emailing without reminders.

In a period of four months, we collected 12 completed surveys, among which two were only partially filled out. Assuming a dissemination to all the networks’ members (which is doubtful), that gives us a response rate of 0.7%, or 0.3% of the French WISEs’ population. We deemed both the absolute and relative values of our sample too low to be exploitable and chose to not use the data of our French survey. Background information to prepare our qualitative research was instead provided by secondary sources such as the ones described in 6.3.3.

6.2.6.3 Data analysis

The data collected through LimeSurvey for the Danish survey was cleaned and analysed in Stata/IC 15. Due to our reduced sample size, we limited our analysis to simple descriptive statistics in order to get a primary understanding of the questions we sought to answer through our quantitative research. The key descriptive highlights of our findings are exposed in Chapter 9.

6.3 Presentation of our empirical material

This section introduces the empirical material upon which the analysis developed in this thesis rests. We present our four case studies and outline the relationship they entertain with their public stakeholders⁶⁰. We then proceed by specifying the complementary resources used to carry out our research, both in terms of primary sources (primarily through interviews of stakeholders external to our cases) and by consultation of secondary sources.

Table 13: Overview of case studies

Case	FR1	FR2	DK1	DK2
Creation date	1992	1968	1999	1984
Annual revenue*	30.6	1.4	2.7	8.0
Headcount	3,977	388	189	250
Staff on integration scheme	3,857	375	160**	202
Assess impact	Yes	Yes	Yes	No

Source: Author

* Million euros

** Some of the management on special “flexi-job”⁶¹ contracts

As an introductory note, we recall the difficulties mentioned in section 6.2.5.2 related to the identification of our cases.

6.3.1 Empirical material in France: FR2 and FR1

In this section, we introduce our French empirical material, two case studies of FR1 and FR2 and their public stakeholders. We take a look at the activities of these WISEs, their legal structure, and briefly touch upon their approach to performance measurement through SIA and the relationship they entertain with their public stakeholders (developed further in Part 2).

⁶⁰ In order to preserve the anonymity of our case studies, we avoid mentioning specificities (such as e.g. names and locations) and we adopt a naming convention: FR1, FR2, DK1 and DK2 identify each WISE, while FR1PS, FR2PS, DK1PS and DK2PS refer to their main Public Stakeholders: The regional Directces in France and the municipality of the city where the WISE is headquartered in Denmark

⁶¹ Fleksjob

6.3.1.1 French case study number one: FR1

Founded in 1992, FR1 grew into one of France’s largest WISEs, grouping 30 establishments across the country. The company focuses primarily on construction and waste management, mixing several of the covenants⁶² available in the French work integration legal environment (cf. Figure 17). In 2016, FR1 had a revenue of 31 million euros and employed 3,977 people, among which 3,857 were following an integration scheme. 90% of FR1’s revenue comes from market incomes, while 10% comes from public subsidies pegged to the beneficiaries the WISE supports through its covenants.

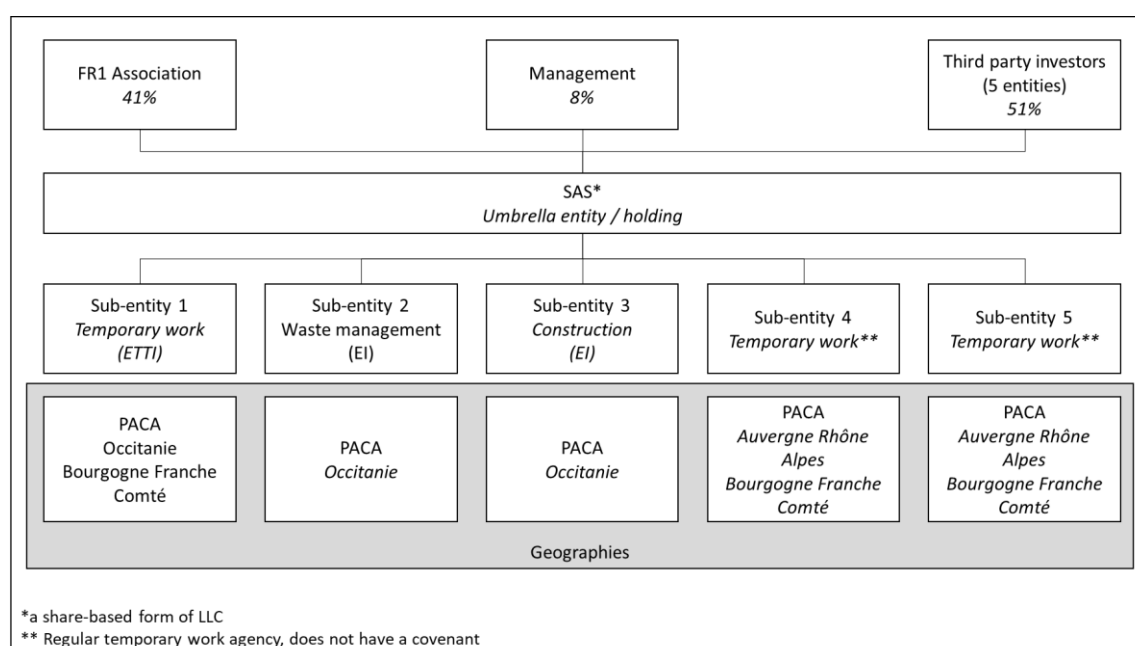


Figure 17: Structure of FR1

Source: FR1 / Author

FR1 was initially founded as an association near one of France’s largest cities, after one of the city’s boroughs had received a “disadvantaged district”⁶³ label from national authorities. The initial aim of FR1 was to warrant the economic and social inclusion of marginalised individuals through the economic development of the *territoire*⁶⁴ they live in. In less than a decade, the association created four sister organisations, among which two were covenanted (i.e. benefiting from a special agreement from the

⁶² Covenants are a formal agreement between a WISE and the regional authorities granting access to referring channels (e.g. job centres) and the associated public subsidies pegged to a referred beneficiary. Under our interpretation of the PVT, we associate covenants as a form of institutional endorsement (cf. Chapter 7 for more insight on covenants).

⁶³ Quartier sensible

⁶⁴ *Territoire* in French, used in the sense of “local area” or “region”

local authorities to carry out work integration activities, cf. Chapter 7). It started accepting third-party capital in 2008.

This rapid transition from a quasi-non-commercial association to a multi-activities business entity with third-party investors was reported troubling for some of the association's founding members, who had difficulties accepting the business dynamics of a project founded on social intents. FR1's current CEO, who has been in place for 20 years and is at the origin of most of these transformations nevertheless considers that they were indispensable drivers for the breadth of the impact the organisation is currently capable of achieving at a national scale. The association still exists: it has 41% shares of the group's umbrella LLC and can exercise a veto on board decision.

At the time of the research, FR1's principal activities were in waste management (waste sorting and recycling) and construction. For the latter, one of the defining activities of the WISE is the repurposing of old maritime containers into housing or office space. FR1 carries out its activities through a variety of legal entities, some covenanted, some not (cf. Figure 17). The main covenants are one ETTI (business to business temporary work) and two EIs (production of goods and services, mostly services in FR1's case)⁶⁵. In addition, few small ACIs (sheltered workshops with 30% capped commercial revenue, production of goods and services) are nested in the group's organisation. Because sponsorship agreements⁶⁶ do not allow to work with beneficiaries for more than two years, FR1 has also created "*regular*" (i.e. non-covenanted) temporary work agencies in order to be able to prolong some of its beneficiaries' integration path out of the covenants system (an example of organisational design strategy in our PVT analysis, cf. Chapter 8).

As a multi-covenanted WISE, FR1 has numerous interactions with several public stakeholders. The primary partners are the Regional Authority for Business, Competition, Consumption, Labour and Employment (Direccte)⁶⁷ and the local job centres. The former is the authority administrating covenants and the associated reporting devices (cf. Chapter 7). Managing covenants is one person's

⁶⁵ Cf. Chapter 7 for a description of each covenant

⁶⁶ Agrément in French, sponsorships are necessary for WISEs to hire beneficiaries and receive corresponding subsidies

⁶⁷ Directions régionales des entreprises, de la concurrence, de la consommation, du travail et de l'emploi. It is a direct emanation of the central power (of both the Ministry of Economy and the Ministry of Labour) at the regional level and boasts a broad portfolio encompassing competition, consumption, employment and labour.

full-time job at FR1. The latter is any WISE's main referring channel (i.e. the public entity formally channelling beneficiaries towards WISEs in France). Under the French work integration framework, the main public authority with which FR1 discusses its performance is the Direccte. While it does so continuously throughout the year, FR1 must also submit to a yearly "*management dialogue*"⁶⁸ where performance is discussed more in details. These conversations often involve other stakeholders (e.g. municipality or "*metropole*" – an intermediary level of authority found in some of France's largest cities) as they track their own performance indicators (cf. Chapter 8).

This means that just like any other covenanted WISE in France, FR1 is subject to public reporting mechanisms to track its performance (what our PVT analysis labels as legitimacy **devices**). These instruments are very much focused on outputs and we introduce them in details in Chapter 7 and analyse them in Chapter 8. FR1 runs its own SIA model in parallel. This tool appears in a continuous development. While there is a genuine concern for outcomes on the part of FR1, its SIA model is very focused on quantitative indicators and outputs and the WISE uses its metric abundantly in both its internal and external communication. While it is not an instrument that can benefit from a formal recognition in the covenant system, FR1 leverages it during management dialogues to advocate about its performance (a **deliberation** process according to our PVT analysis). Chapter 8 details FR1's SIA model and its use through strategic interactions with public stakeholders.

Third-party investors detaining shares of FR1's holding are all considered "impact investors" by FR1's management. That is to say that the group's social mission was an investment criterion for all five investors. Interestingly, there does not seem to be a structured way through which these investors are measuring their "*social return*", at least none that FR1's management could tell us about. These partners appear satisfied with the way FR1 tracks its own impact. We interviewed one of these investors during our research (the group's first financial partner in 2008), who indicated not having any formal tracking device for social return but was working towards the development of such instruments.

⁶⁸ Dialogue de gestion

6.3.1.2 French case study number two: FR2

FR2 was founded in 1968 in a rural region of Eastern France. It was incepted as an association by a group of unemployed people wishing to experience a form of self-employment through the provision of paid-per-hour service such as cleaning or gardening for private individuals. The WISE's creation happened at the very beginning of the structuring of the work integration sector in France, as we describe it in Chapter 7. FR2 was a founding member of Coorace⁶⁹, a national network of (today) AI and ETTI covenanted WISEs. FR2's current CEO was Vice President of Coorace for a period of time and is still very active in the network.

After the *Intermediary Association* (AI) covenant was created in 1987, FR2 was among the first organisations to adopt this framework which allowed to reach out to beneficiaries⁷⁰ more easily and granted access to a level of public funding. At the same time however, the AI covenant limited FR2 to the sole activity of "*labour provision*" only in a business to consumer setting⁷¹ (cf. Chapter 7). As FR2 started to want to cater to both different beneficiaries (e.g. people who cannot operate autonomously on customer's premises) and customers (especially business to business, where more commercial options are available, it created three daughter organisations, each covenanted differently, and became an Economic Solidary Grouping (GES)⁷² a legal form open to covenanted WISEs willing to consolidate under one structure (effectively a form of "*social holding*", cf. Figure 18).

⁶⁹ Created in 1985 as (COORDination des associations d'Aide aux Chômeurs par l'Emploi) while associations with a purpose similar to FR2 gathered in a national network to advocate for their activity. Coorace's activity led to the creation of the AI covenant in 1987 and ETTI in 1991.

⁷⁰ The covenant allows public stakeholders to identify AIs more easily, and therefore ease the channelling process. However AIs' beneficiaries do not require a sponsorship agreement as in EIs' or ACIs' do.

⁷¹ AIs are limited to the business of provisioning labour to private individuals for defined (often domestic) tasks. Consequently they do not sell services as such (customers pay for access to labour hours, not for the execution of a service) and cannot carry out business with moral entities.

⁷² Groupement économique solidaire

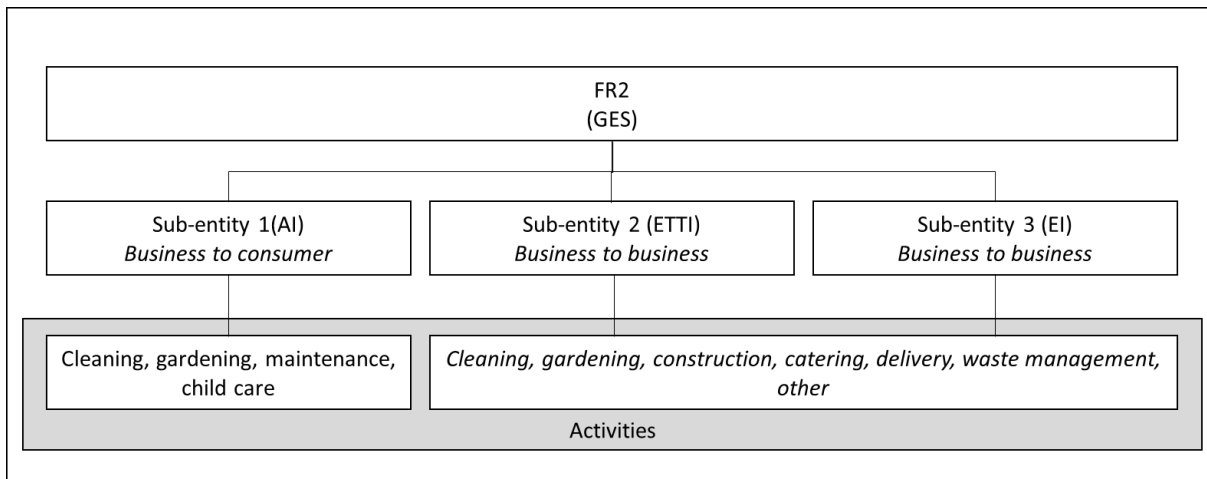


Figure 18: Structure of FR2

Source: FR2 / Author

Just like with FR1, we see strategies of **organisational design** in order to optimise FR2’s operational vertex by diversifying its customers and beneficiaries base while still benefiting from the **institutional endorsement** of covenants. To that effect, while FR2 started as an organisation dealing exclusively in a business to consumer setting and still had 80% of its turnover coming from private individuals in 2010, it now derives 75% of its revenue from business to business activities. At the time of the research, FR2’s activities were still primarily oriented towards cleaning (mostly of office space), with complementary businesses in waste management, logistics, construction, hospitality, child care and other types of services. These services are provided either via temporary staff through Sub-entities 1 and 2, or *via* packaged services using mentored staff through Sub-entity 3. Table 14 provides a breakdown of FR2’s 2016 revenue between its activities of temporary work (Sub-entities 1 and 2) and its service activities (Sub-entity 3).

Table 14: FR2 revenue break down

Entity	Revenue 2016	Public subsidies 2016
Sub-entity 1&2	€ 1,340,000	€ 182,000
Sub-entity 3	€ 90,000	€ 66,000
Total FR2 (GES)	€ 1,430,000	€ 248,000

Source: FR2 2018

Through its three covenants, FR2 experiences a situation very close to FR1, where it has recurrent conversations with the Direccte and where it is asked to report on the same performance criteria

(“*positive graduation rates*”, a measure of the WISE’s outputs, cf. Chapter 7). FR2 also interacts with some other local public stakeholders during its annual management dialogues. During our interviews, the WISE’s leadership complained about the recurrent staff changes at the Direccte (every year in the last three years), which does so that FR2 needs to re-educate their partners about their activities and the value they generate every time. This appears detrimental to the emergence of **collective preferences** and the generation of a co-defined public value, which in turn may favour a form of **indicator-substitution** where the indicators from the reporting imposed by the Direccte becomes a “*source of truth*” by default (cf. Chapter 8).

It is precisely in order to be able to better communicate about what it does that FR2 has developed its own SIA model. It is based on the French approach of “*social utility assessment*”, also called “*Territorial social utility assessment*”(UST)⁷³, which promotes the use of a holistic method to the overall apprehension of impacts and tend to avoid metrological tools (we explore this notion as a distinct socio-technical arrangement from impact measurement in Chapter 8). As an active member of Coorace, FR2’s SIA initiative was primed by the network, which in 2012 adopted a methodology developed by H el ene Duclos (a pioneer of the social utility approach) for “*Territorial social utility assessment*”. FR2’s staff followed several workshops organised by Coorace in order to develop its own SIA model based on the approach developed by Duclos. We describe this model and the use that FR2 makes if it in Chapter 7.

6.3.2 Empirical material in Denmark: DK1 and DK2

In this section, we introduce our Danish empirical material, two case studies of DK1, DK2 and their public stakeholders. We take a look at the activities of these WISEs, their legal structure, and briefly touch upon their approach to performance measurement through SIA as well as the relationship they entertain with their public stakeholders (developed further in Part 2).

⁷³ Evaluation de l’utilit e sociale,  valuation de l’utilit e sociale territoriale

6.3.2.1 Danish case study number one: DK1

DK1 was founded in 1999 near one of Denmark's largest cities. It was originally created on an idea from the head teacher of a "Day High school"⁷⁴ to offer job opportunities to some of his disabled students. Day High Schools offer vocational education for adults experiencing employment difficulties. They are based on the model of the Danish Folk High Schools⁷⁵ which are informal education institutions often presented as forming a cornerstone of the Danish society (cf. Chapter 7 for more details and their history and their influence on the Danish social economy). Initially focusing on the reselling of small IT hardware and IT reparations, DK1 soon started to expand the scope of its activity, and in 2006 it began to open affiliates throughout the country.

At the time of the research, DK1 had a revenue of 2.7 million euros and employed 189 people (cf. Table 13). It had four establishments (cf. Figure 19) grouped in two legal entities (two more closed shortly after their opening, due to mismanagement). The two legal entities are stand alone, with no equity or form of capital other than intellectual property linking them. They have two separate boards but which are made of mostly the same individuals. The main entity is still located in DK1's headquarters and our case work is based on its activities, although we also visited Sub-entity 4 located near Denmark's capital, as a complement to our research. The different branches of DK1 are operated separately from one another.

With the years, DK1 has considerably extended the range of its activities, while its IT hardware business has nearly come to a stop, the main office (Sub-entity 1) branched out into communication (e.g. design and printing of advertisement) and other entities are running their own commercial operations (cf. Figure 19). DK1 did not provide a breakdown of its revenue per activity, it was nevertheless stated by the staff during our interviews that the income from commercial productive activities is marginal, and that most of the WISE's revenue is generated through the work integration services sold to the municipality (this seems particularly true in Sub-entity 1). Productive activities are mostly there to support work integration service by providing opportunities for job placements (a strategy also observed in DK2).

⁷⁴ Daghøjskoler

⁷⁵ Folkehøjskoler

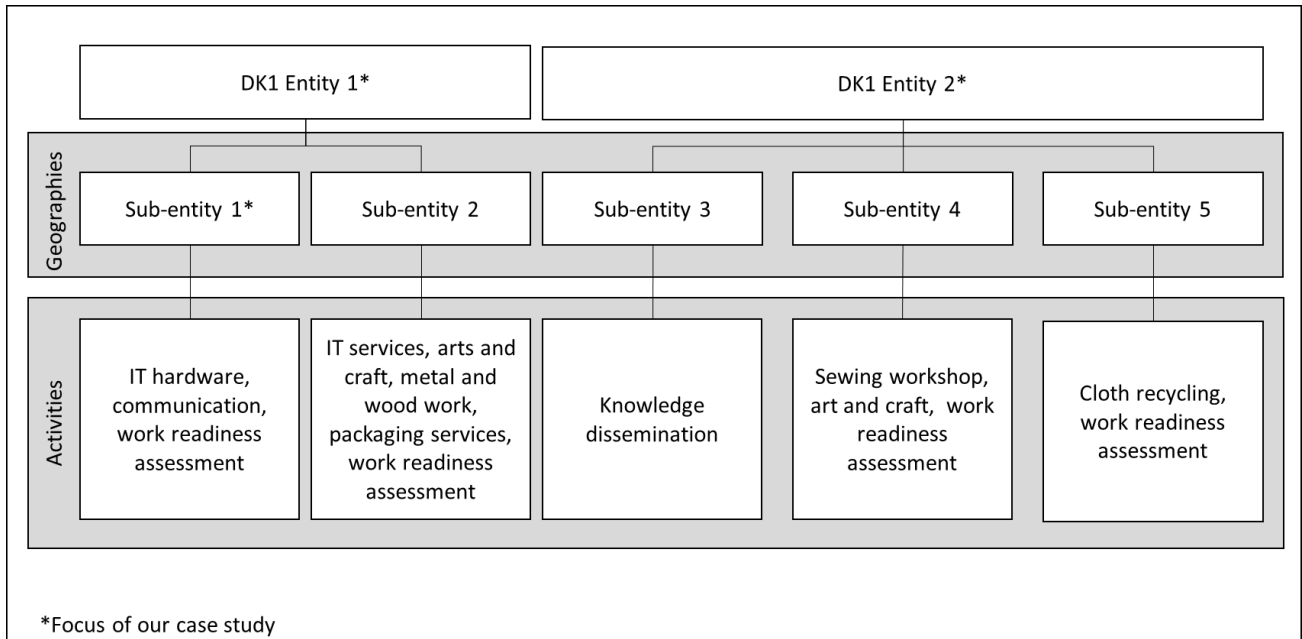


Figure 19: Structure of DK1

Source: DK1 / Author

DK1 is a peculiar WISE in Denmark as it earned a spot in the Danish finance law in 2008, as an “*enterprise project employing people with functional limitations*”⁷⁶. This translates into a pool of money (approximately one million euros) from the Ministry of Labour that is earmarked to DK1 yearly. DK1’s management has direct interactions with the Ministry of Labour in relation to this funding, and has to submit bi-yearly activity reports to ensure the funding is maintained. However due to the exceptional nature of this funding from the central government, we chose to not include it in our analysis of strategic interactions (i.e. the relationships with the Ministry of Labour connected to funding issues were not analysed) as it would introduce a considerable bias in our results: in 2018, out of the *circa* 300 WISEs operating in the country, only one other than DK1 was in the finance law.

Compartmenting the strategic interactions between our area of focus (the local government, represented by the municipality and the job centre) and the central government was relatively painless: in ten years, the government funding became a relatively routinized part of DK1’s operations that generates little attention in the conduction of daily business. Throughout our interviews, this topic only

⁷⁶ Projektvirksomhed for personer med funktionsansættelse

surfaced once, with the CEO, when specifically asking about eventual dealings with the Danish central government. That is not to say that this funding is without importance For DK1 since it represents a sizable contribution to its operational vertex and has entered into consideration when framing initial SIA reflections (we account for that in Chapter 9). However the clear-cut separation between the annual direct funding from the Ministry (which is rather unique in Denmark) and the daily business operations with the municipality and the job centres (which is common to all Danish WISEs) has helped us keeping focus on relevant observations applicable to the Danish environment and not contaminating our data with idiosyncratic noise.

In terms of SIA, DK1 has a fully-fledged model since 2016. It was developed in-house, with a focus on public customers. Its gestation lasted for three years, which is a process providing interesting trajectories in terms of case use for SIA and choice of indicators that we analyse in Chapter 9. In a stark contrast to the French environment, the absence of formal legitimacy-generating mechanisms such as endorsement or devices means that legitimacy is secured through other means such as the actual creation of public value and its recognition by public stakeholders (in the public value vertex), and the activities of the WISE as well as its price level (in the operational vertex), which must be appealing and befitting the requirements of public customers. Ultimately, we observe that legitimacy is also generated directly through deliberation, when DK1 and their public stakeholders attempt to surface collective preferences. In this context of loose evaluation, where relational capital can play an important role, we explore the use of DK1's SIA in Chapter 9 and uncover market-driven motivations.

6.3.2.2 Danish case study number two: DK2

DK2 was founded in 1984 and is focusing on supporting young people through different aspects of their life, including vocational training and work integration. DK2 was initially thought out as a school for young people with special needs. As it developed vocational training programmes, it rapidly branched into developing its own business activities in order to create opportunities for job placement for its students.

In 2017, the WISE had a revenue of eight million euros (cf. Table 15) and was employing 250 people. The enterprise is headquartered in in the outskirts of Copenhagen. DK2 also has a site in the city's

periphery where some of its work integration workshops are installed, as well as glasshouses and arable lands to grow crops as part of the company’s productive activities. It is this second site that DK2 also installed the “Copenhagen Dome” in 2017. The Dome is an architectural piece that was put on display for several years in Aarhus and Copenhagen, and that DK2 repurposed to be a “knowledge centre for social economy” in Denmark, with the idea to use the construction as a meeting and exchange place for actors of the social economy. This installation does not impact DK2’s business activities directly but places the enterprise as one of the front runners in the Danish social economy.

Table 15: DK2 revenue breakdown

Items		Yearly revenue*
Public contracts	Housing services	€4 Millions
	Sales of work integration & education services	€2 Millions
Commercial activities**	Sales of good and services	€2 Millions
Total		€8 Millions

Source: Author

*Information collected during interviews, no formal revenue documentation was provided

** Commercial activities include goods and services sold to customers in the public sector

DK2 is structured as a collection of different legal entities, falling under the umbrella of a foundation (cf. Figure 20). Its activities are structured as followed:

Accommodation services:

The enterprise offers an accommodation solution, providing places to live for young people having social difficulties. This is accessible to individuals between 17 and 26 years old, displaying social challenges either as autonomous persons or within a family household, whether linked to a medical condition or not. The accommodation service is provided as a social service carried out on behalf of the municipality, which pays for the service. DK2 garners approximately half of its revenue from accommodation services, or four million euros.

Youth education:

DK2 proposes a 3-year education programme targeted to people with special needs between 18 and 25 years old. This programme is provided as an equivalent to a high-school education, with an emphasis

on vocational training as well as personal and social development. This offering is grounded in a 2007 law (the Act on education for youth with special needs), upon which was developed a framework for privately-operated specialised institutions dubbed “Specially Adapted Youth Education” (STU)⁷⁷. At the time of their inception, STUs came as a natural piece to add to many WISEs’ offering, which saw a way to reach out to their target groups at a young age, for an extended period of time through a generous finance mechanism (about 3,000 euros per month per individual, compared to an average of 1,000 euros for individuals on the Afklaring work integration scheme). In the first years of the scheme, and before it got reformed, STUs would also represent a sizeable share of WISEs’ revenue. For DK2, it reached approximately half of its revenue, while it is now accounting for less than a quarter.

Work integration:

DK2 offers an array of work integration activities under the form of workshops grouped in a social enterprise. These cover a wide range of business sectors, with the main ones being: a bakery, blacksmithing, carpentry, organic farming and catering services. Just like at DK1, these activities offer opportunities for job placements, training and actual employment.

In its diversity, the entire offer of DK2 is explicitly and almost exclusively geared towards bringing young individuals into employment. While DK2’s social enterprise offers many different opportunities, the primary purpose of its workshops is not to provide lasting jobs but to be used as places to develop skills and competencies that match the aspirations of the individuals the WISE is supporting. There are cases where individuals can be offered a job at DK2 after having gone through the assessment period sponsored by the municipality, but these are the exceptions rather than the rule. DK2’s management is straightforward about their goal of having the youth they support working out of the organisation and in a sustainable fashion.

Our research focuses on DK2’s work integration programme, and therefore on the activities carried out through Sub-entity 3. The rest of the activities (accommodation and education) however enter the scope of our research when considering DK2’s operational vertex and the **organisational design** strategies that are behind the observed setup, securing both a supply of beneficiaries and a steady

⁷⁷ Særligt tilrettelagt ungdomsuddannelse, named after the legal framework allowing this kind of offering out of public funding

income (work integration is a minority activity for DK2, even though it reaches employment-related outcomes through other means, especially in the STU).

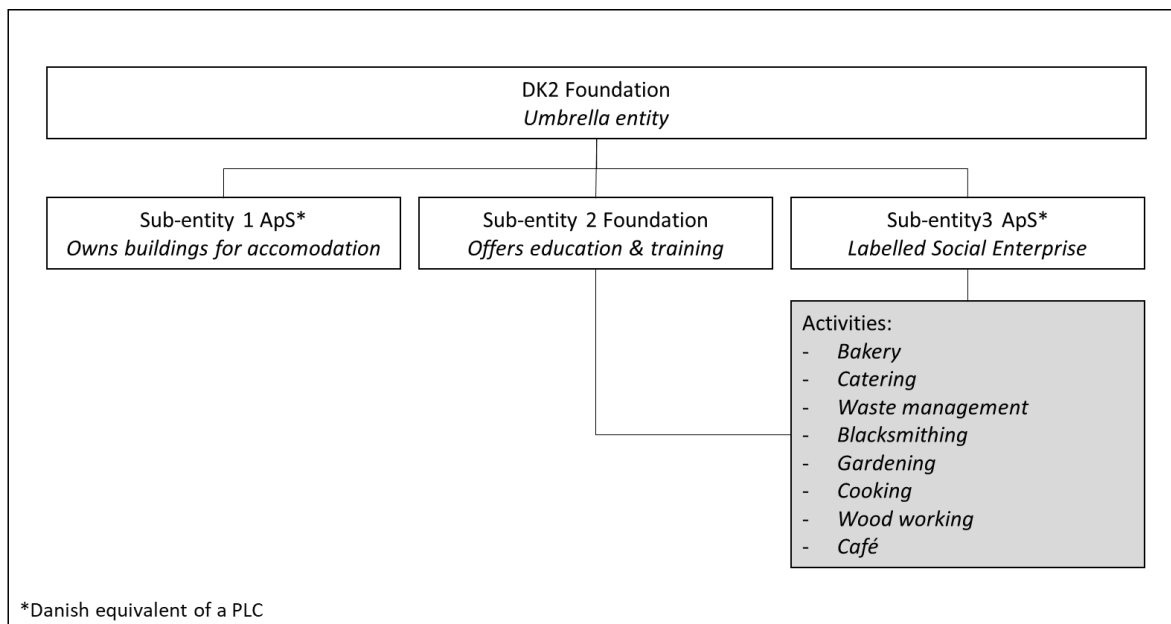


Figure 20: Structure of DK2

Source: DK2 / Author

DK2's relationship with its public partners follows the Danish conundrum of **relationship-based quasi-market** between a private provider of employment services and a public customer paying to get these services executed (cf. Chapter 7). For its work integration activities, DK2's main public interlocutor is therefore the local job centre. It also has direct interactions with the municipality through its housing and education schemes. Differently from France, the partnerships Danish WISEs can form with public actors are not bound by geography, and although DK2 is running projects with other municipalities than the one where it is headquartered (DK2PS, DK2's main public stakeholder). While day-to-day business interactions in regard to integration schemes is carried out through the job centre, the municipality is still implicated in the back office, and particularly on the evaluation aspects (cf. Chapter 9).

In the case of DK2PS, the administration has set up an evaluation tool dubbed “Effect and progress”⁷⁸. When interviewing the municipality and the job centre’s staff, this tool was equated to a performance measurement model⁷⁹. It was developed internally and attest of the very loose approach that Danish municipalities have of performance measurement for WISEs (cf. Chapter 9). At the time of the research⁸⁰, DK2 did not have any processes in place to measure their social impact. They nevertheless appeared as an interesting case to study because of their strong interest in the topic and because they offered an opportunity to explore what prevents organisations from taking action towards concrete SIA practices.

6.3.3 Complementary resources

We completed our four case studies and our survey by an array of resources obtained either through primary or secondary research. This brief section describes the main sources used we used to inform our analysis.

6.3.3.1 Primary sources

Most of our primary material is made of our case research and of the survey we carried out in Denmark. To complement these views, we carried out 36 interviews to better understand the French and the Danish context in which WISEs evolve (cf. Table 9 and Appendix 9). In both countries, interviewed stakeholders included:

- Work integration or SEs networks as well as other WISEs: they have provided a view that complemented the *reality* that was described by our case studies and have allowed us to cross check facts. For instance, using DK2 and two other WISEs we met in Denmark, we realised that DK1’s funding from the central government had indeed no impact on the running of day-to-day operations.
- Advocates, consultants and investors: they provided us with an insight on how stakeholders from the private sector were supporting the development of WISEs and were sometimes advising them

⁷⁸ Effekt og progression, where the Danish equivalent of the English word “effect” is used to designate what we would qualify as outputs in impact measurement terms

⁷⁹ Use of the wording “effektmåling” in Danish, or “*measurement of the effect*”

⁸⁰ December 2017

on SIA (we observed this particularly with Den Sociale KapitalFond, a Danish investor and KPMG, a consulting company in France)

- National central authorities (ministries and sub-ministries): they provided us with central governments' take on WISEs' performance measurement, confirming a strong interest for SIA in both countries (however translating in little to no local initiatives).

Additional insight was also collected through participation in local events like the inauguration of the "Copenhagen Dome" repurposed as a "knowledge centre of social economy" in Denmark, SED meetings or participation in Chantier Ecole's SIA commission meeting. Such events allowed to exchange with a more complete panel of field actors and gain further understanding on the way they apprehend their environment and the role they attribute to performance measurement in this perspective.

6.3.3.2 Secondary sources

As developed earlier in our empirical section, the efforts involved in carrying out a survey in France similar to the one we executed in Denmark did not yield exploitable results. For France, we therefore used secondary sources that could provide us with a view similar to the one we developed for Denmark. Our most useful resource in this endeavour has been the 2018 edition of KPMG's "social impact measurement's bellwether"⁸¹. Unfortunately, KPMG declined to share the data of its report for the use of this thesis. This means that the exploitation we make of the data is solely based on what is communicated in KPMG's report. The report obviously does not offer data fully comparable to the ones we collected in Denmark and suffers from some caveats (particularly the absence of a definition for what is qualified as social impact measurement⁸²). It nevertheless constitutes a valid resource to anchor our reflexion on some of SIA practices in France. It should be noted that an interview with KPMG revealed challenges very similar to the ones we encountered in getting survey responses, with a response rate of 4.4% despite KPMG's well qualified contacts base and extended network. To

⁸¹ Baromètre de la mesure d'impact

⁸² A design choice clearly stated by the report's authors but which renders the interpretation of some of the data difficult

complement our view of SIA in France, we also abundantly used the resources provided by Avise⁸³, as summarised in Table 16.

Table 16: Main social impact assessment-related reports referenced by Avise

Title (translated from French)	Author	Year of publication	Content:	
			SIA models	Cases
Short guide to social impact assessment	Avise, ESSEC	2013	Yes	Yes
Measuring social impact	CSESS	2011	Yes	No
Suggested approaches to measure social impact	EU /GECEs	2014	No	Yes
Analyzing, following and assessing one's contribution to social change	AFD	2016	No	Yes
Measuring Impact ⁸⁴	G8	2014	No	Yes
Guide to social impact measurement	Rexel Fondation, Improve	2014	No	No
A Guide to Social Return on Investment (SROI)	ESSEC	2011	No	Yes
A practical guide to measuring and managing impact	EVPA	2015	No	Yes
Assessing your activity's social utility	Avise	2007	No	No
Passeport Avenir, a programme that counts double	Accenture / ESSEC	2012	No	Yes
Assessing the social impact of a social enterprises: benchmarks	Stievenart and Pache	2014	Yes	No
Assessing the social value of social and solidary economy	Branger, Gardin, Jany-Catrice, Pinaud	2014	Yes	Yes
Should we measure everything? Reflection about the usefulness of social impact measurement	Mortier	2013	No	No
Guide to best practices	ESF	2015	No	Yes
Barometer of social impact measurement	KPMG	2017	No	No

⁸³ Agence de valorisation des initiatives socio-économiques : an association created by the French Deposits and Consignments Fund and several stakeholders from the third sector. It acts as an FR2iate Body for the European Social Fund and benefits from the sponsoring of several ministries. Its main role is to support the development of social economy in France.

⁸⁴ In English

Title (translated from French)	Author	Year of publication	Content:	
			SIA models	Cases
User Guide: Assessing Social Impact	Avisé	2017	No	Yes
Social and Solidary Economy and Value Creation: A prospective approach to social impact measurement	Avisé, La Fonda, Le labo de l'ESS	2017	No	Yes
The experience of social impact assessment	Agence Phare	2017	No	No

Source: Author

In Denmark, our main complementary resources were made of the documentation produced by the National Centre for Social Enterprises before it closed down, resources from Den Sociale Kapitalfond (Denmark's sole impact investor), papers from Mandag Morgen, a Danish think tank which was instrumental in raising awareness about social enterprises among the country's elite as well as the documentation made available by municipalities having developed a *"strategy for social enterprises"* (cf. Table 21).

In both France and Denmark, we used the corporate publications of our case studies as well as their websites to inform further our understanding of their activities and their overall strategy. We also carried out topical research on each case studies on the Internet (using Google as our main search engine) and in Factiva. Further, we used the audited accounts of all four cases to better gauge their situation and deliver additional insight on some of our observations. The accounts were obtained directly through FR1 and FR2 in France. In Denmark, for DK1, they were obtained directly through the WISE and checked against the information available publicly in Virk, the Danish Business Register⁸⁵; for DK2, they were obtained solely through Virk.

In both countries, we also used documentation produced by the European Commission, and particularly the Map of social enterprises and their ecosystem in Europe, which was issued for France and for Denmark. In France, publications from institutions framing the social economy and work integration such as the National Council of the Regional Chambers of Social and Solidary Economy or

⁸⁵ <https://datacvr.virk.dk/>

France Active have provided considerable insight to understand and gauge our research topic in its national context.

Conclusion of Chapter 6:

This chapter has made clear the fundamental world view (our paradigmatic approach: **critical realism**) through which we will use the analytical grid presented in Chapter 5, and how that view is articulated in a defined methodology to investigate a selection of case studies and complementary material.

Critical realism (CR, Bhaskar, 1975, Lawson, 2003)' **layered ontology** (the empirical, the actual and the deep real) is the philosophical cornerstone of our investigation, as we will probe the deep real to explain the phenomena observed in the empirical (the facts as experienced and reported by our cases). With CR, we will elucidate the questions we raised in Chapter 5 with context-sensitive explanatory answers, but we will not frame them as universal rules: what is the intent of WISEs behind the measurement of their performance? Can these intents explain the phenomena of performativity?

Since we explained in Chapter 2 and Chapter 5 that answering such questions (at the micro level) necessarily involves a thorough comprehension of the local context (the macro level), Maurice's societal approach (1989) appears particularly relevant to our research. It indeed focuses on the interactions between these two levels while ensuring the comparability of the units of analysis, and it is also underpinned by an ontology admitting "*multiple and interactive causality*".

Lastly, we showed the interest of our mixed research design (although leaning heavily on the qualitative side due to field constraints) for the societal approach through an embedded case studies (i.e. ensuring the balance between comparability and embeddedness). We also showed the relevance of mixed methods and qualitative methods for CR, where retroduction (a form of abduction characterised by back and forth cycles between the data and explanatory hypotheses) is the primary mode of inference.

Having presented the general context of our research (Chapter 1), our literature review (Chapters 2 through 4), the analytical grid we built out of it (Chapter 5), and how we operationalise this analytical

grid (this chapter), we are now ready to move to Part 2, where we explore our field data and answer all the questions we raised in this first part.

Part 2 Empirical results

The second part of our thesis presents our findings, where we seek to bring an answer to our research questions, building upon the analytical and theoretical framework we introduced in the first part. We do so by following the mental pathway presented in Figure 21, that shows how the different aspects of our research flow into the observed practices of performance measurement (with potential performativity and its influence illustrated with dashed arrows). In order to better grasp the phenomena at stake, our analysis does not trail the flow indicated in Figure 21 but instead follows the numbering of the items. Doing so, we apply the critical realist retrodution, by altering between different layers of reality (e.g. performance measurement is in the empirical, as the concrete models and indicators we observe, but the strategic intents are in the deep real) and the analytical frameworks we use to explain them.

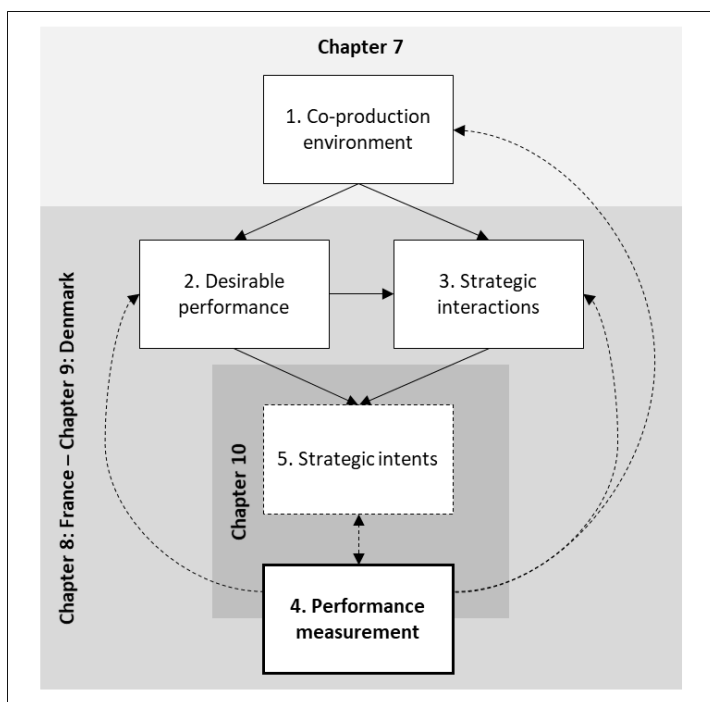


Figure 21: Conceptualising our approach

Source: Author

In Chapter 7, we therefore start by articulating what France's Neo-Weberian State (NWS, Pollitt & Bouckaert, 2004) and Denmark's New Public Governance (NPG, Osborne, 2006) models mean in terms of public value co-creation. The key difference between the two countries resides in the **legalist** environment observed in France and what we identify as a **relationship-based quasi-market** in

Denmark. The French environment is the result of NWS mechanisms which establish that work integration social enterprises (WISEs) must benefit from an institutional endorsement (the covenant) to operate and access beneficiaries and funding (but not necessarily commercial activities). In Denmark, the place of **relational capital** induced by NPG allows for public value co-creation to rest on underlying market mechanisms (WISEs are commercially selling a work integration service to municipalities, their main public stakeholder) while maintaining a close collaboration and a high level of **deliberation** (deliberation is the principal vehicle through which performance is assessed) among WISEs and their public stakeholders.

These two environments are crucial to understand because they are what sets the foundation of *desirable performance* (i.e. the expectations of stakeholders in terms of performance in a given environment, box 2 in Figure 21) and conditions the strategic interactions between WISEs and their public stakeholders (box 3). We explore these two items in Chapter 8 and Chapter 9, where we develop the analysis of our cases (first in France, then in Denmark), establishing how both WISEs and their public stakeholders measure performance (box 4). We focus on observing how the public value co-creation mechanisms we identify in Chapter 7 translate in strategic interactions, and how they tend to shape approaches to performance measurement.

It appears that NWS' legalism in France places a heavy emphasis on the public value theory (PVT, Moore 1995)'s **legitimacy** vertex, which can be detrimental to the two other vertices. We also see the effects of NWS on performativity: bureaucracy can favour **retroaction**, while centralisation has a contrasting effect on **socio-technical arrangements** (STAs) and **indicator-substitution**. In Denmark, WISEs deploy their strategies in a more **balanced** manner across the PVT triangle, as they must maintain a competitive positioning where legitimacy, operationalisation and public value creation are intertwined. The downside is that competitive considerations can strain WISEs' **operational** vertex. We also see that the loose approach to performance measurement translate into an environment of important **deliberation**, which contributes to a co-definition and an acknowledgement of the created public value, but also favours the emergence of a multiplicity of **socio-technical arrangements** that present a capacity to become enacted (i.e. become performative).

In Chapter 10, we bring these field observations together to answer the research questions formulated in Part 1. Doing so through the use of our PVT analytical framework, we reveal the **strategic intents** tied to performance measurement (box 5). We conclude that the strategic intents and the public value co-creation environment are two causal mechanisms that can explain the formation of performativity through strategic interactions. By doing this, we also unearth potential ways to prevent performativity. These last aspects are further developed in our final chapter, where we focus particularly on the role of STAs, and the way they develop. There, we introduce the concept of **socio-technical arrangements' feedbacks**, which suggests that STAs can actually be harnessed to **prevent** performativity.

Chapter 7 Mechanics of public value co-creation in France and in Denmark

In Chapter 5, we identified strategic interactions between work integration social enterprises (WISEs) and their public stakeholders as a key aspect to investigate in order to gain an understanding of issues pertaining to performance measurements. Using the two models introduced in Chapter 2 (the Neo-Weberian State -NWS- and the New Public Governance -NPG-), this chapter provides the foundations to comprehend how the Danish and French public administration and management (PAM) models shape these interactions.

Doing so, we answer questions such as how do the NWS and the NPG translate into frameworks for public value co-creation? What are the rules that WISEs must comply with in order to operate in this environment? What can be said of the relationship towards their public stakeholders? Are they partners? Customers? Or even competitors? By elucidating these questions, we progressively gain an understanding of what characterises WISEs' performance for public actors and how it differentiates in France and in Denmark.

We start by introducing the overall administrative and policy architecture of each country, in order to ground the choices we make in terms of PAM associations. We then proceed to provide a historical overview of the development of social economy, social enterprises (SEs) and work integration social enterprises (WISEs) in each country. This topic is developed further for Denmark, where the lack of formal institutional framework for social economy and work integration encourages us to look further into the historical and cultural roots of the observed phenomena. In both countries, we highlight the role of cooperatives as founding movements for the development of social economy and the branching out into work integration.

We then dedicate a substantial space to the description of the actual institutional framework in which WISEs operate in both countries and how these scaffolds define the way public value is **co-created** by WISEs, thereby driving the **strategic interactions** between WISEs and their public stakeholders. In France, the **legalism**, the **centralisation** and the **bureaucracy** of the NWS generate multiple compliance mechanisms (covenants and detailed mandatory reporting) executed at the county level but by direct representative of the state authority. This leads to a centrality of state actors as providers of a

legitimacy indispensable for WISEs to operate. To this extent, Gianfaldoni (2018) speaks of “*quasi-hierarchic rapports*” and “*subordination relations*” between WISEs and their public stakeholders.

In Denmark, the loose, network-driven NPG organisation commands an environment where public value co-creation occurs through what we call a **relationship-based quasi-market**: a place where market mechanisms are waned by NPG’s network governance and deliberation. In this setting, WISEs compete with public services providers, but have the opportunity to build long-term relationships with their public customers and main public stakeholders (the municipalities, through the job centres), thereby building **relational capital**, a vector of trust and a key feature of NPG.

7.1 France: work integration and social economy in a Neo-Weberian state

How do French WISEs co-create public value with their public stakeholders? We start by introducing the French public PAM framework at a high level, showing how its features of reformed **legalism**, **centralisation** and **bureaucracy** match the description we make of the NWS in Chapter 2. We then continue with a brief history of social economy in the country, before proceeding to describe the different institutional frameworks mobilised for social economy on one hand and for work integration on another hand.

As we introduce the complex network of stakeholders and rules governing work integration in France, we explain how public value co-creation takes place in the NWS framework. We thereby describe the key mechanisms that articulate this system: the **covenants** system, an institutional **endorsement** that provides a legalist basis for the operation of WISEs; the centralised reporting system, a **legitimacy device** to which WISEs must submit in order to maintain their covenants; and the bureaucratic intricacies that arise from this framework. Lastly, we also see how the reforms imprinting NWS's "neo" tenet impact work integration activities.

7.1.1 Characterising the French public management framework

In Chapter 2, we indicated our choice of ascribing France to a mainly NWS model. In this short introductory section, we present the main features of the French PAM framework and we explain how we tie them to a NWS model, mainly through the defining characteristics that we highlight in Chapter 2: legalism, bureaucracy and centralisation. This section is by no mean a thorough reflection on the French political system and its organisation. It is merely a high-level overview designed to equip us with the main keys to navigate France's institutional environment through our analysis of WISEs and their strategic interactions with public stakeholders.

In the PAM literature, France can often be found epitomising the Napoleonic state (cf. Table 1), with a **central** authority infused in all the strata of the administration and complex **bureaucratic** systems based on a daunting body of administrative **law**. Or in Hammerschmid et al.'s words: "*The French state has been recognised as the archetype of a 'unitary and centralised state' – a 'Napoleonic state' – with*

strong national integration through a territorially based and vertically integrated administrative system” (2016, 53).

The French **bureaucracy** is best illustrated by Table 17, presenting the four main layers of sub-national authority in the country where the smallest unit, the municipality, is found in 35,416 occurrences. This number changes every year, as new municipalities are created and others are merged. To this end, one of the aims of the last French territorial reform, enacted in 2016, was to bring that number below 36,000 (with the main goal to bring the regions from 27 to 18). With an average ratio of inhabitants per municipality below 2,000, this figure provides an illustration of how intricate bureaucracy can be in France. Besides, additional layers than the ones presented in Table 17 can be found under certain conditions. For instance one of our case study (FR1) falls under the jurisdiction of a “*metropole*”, a level combining a portfolio of attributions ranging from the municipality to the region.

Table 17: Layers of authority: the French sub-national onion

Type	Number ⁸⁶	Main Functions
Regional council	14	Economic development, infrastructure, non-urban transportation
County council ⁸⁷	101	Social welfare (child and elderly care, disability, minimum income) and some infrastructure
Inter-municipal cooperation (EPCI) ⁸⁸	1,268	Similar to municipalities, for services across municipalities
Municipality	35,416	Urbanism, housing, environment and wastes

Source: National Assembly⁸⁹

The French centralisation can be illustrated by the role of the “*préfet*”, a direct representative of the central government who has a jurisdiction at county⁹⁰ level (deconcentration of power). Incepted by Napoléon in 1800, Préfets are invested with state prerogatives. They have a dedicated authority on police forces and are able to overrule (through an administrative court) decisions made by “*territorial*

⁸⁶ In 2017

⁸⁷ Conseil départemental

⁸⁸ 1,268 with their own tax levy. Adding other inter-municipal entities, the total is 12,610

⁸⁹ www.assemblee-nationale.fr

⁹⁰ Département

*authorities*⁹¹ (i.e. municipalities, county or regional councils). Of course the role of the *préfets* have changed considerably since 1800, and NPM-inspired reforms focusing on decentralisation have largely contributed to this, by for instance removing the *préfets*' executive power and locating it at the regional council in 1982, and then at the county council in 2015.

The same NPM-inspired reforms also attempted to promote managerial techniques, efficiencies and a culture of evaluation (see e.g. Bezes, 2016; Hammerschmid et al., 2016) in the French PAM. While the country has often been presented as resisting NPM (Hammerschmid et al., 2016), the 2001 Organic Law on Budget Legislation⁸ (mentioned in Chapter 1) is generally seen as a turning point, instating a NPM toolbox focused on performance and cost management (see e.g. Bezes, 2010 or Corbett, 2010), through a **legalist** framework: evaluation became tied to the French constitution through a dedicated organic law. Another iconic reform is the General Public Policy Review⁹² between 2007 and 2012, attempting to address both bureaucracy and centralisation.

Despite such reforms, and through the reluctance they were met with, France conserves a label of Weberian state as presented in Chapter 2 (a centralised, rule-based power conducive to bureaucratic intricacies). But because of the new PAM mechanisms induced by these reforms, such as an increase awareness of performance, France is considered a Neo-Weberian State, displaying a unique blend of reformed Napoleonic centralised authority, bureaucracy and an increasing share of market mechanisms.

This environment sets the scene in which WISEs **co-create public value** and has several implications in the framing of the **strategic interactions** we research. First, while WISEs' primary stakeholder, the *Directe*⁹³, has regional prerogatives, our cases have to contend with multiple stakeholders having a multiplicity of targets and their own views on how WISEs should perform. Second, the rule-based, centralised nature of power in France (despite the *decentralisation* started in 1982) generates a

⁹¹ Collectivités territoriales

⁹² Révision générale des politiques publiques (RGPP)

⁹³ Directions régionales des entreprises, de la concurrence, de la consommation, du travail et de l'emploi. It is a direct emanation of the central power (of both the Ministry of Economy and the Ministry of Labour) at the regional level and boasts a broad portfolio encompassing competition, consumption, employment and labour. The *Directe* is a form of deconcentrated power: they represent the central authority, are subordinated to it and administrate a jurisdiction without a legal entity (different from e.g. a region, which is a form of decentralised power, although limited in scope).

number of idiosyncrasies that have turned work integration activities into a regulated sector, as we detail it below. Additionally, and as we see in Chapter 8, centralisation and the layers of sub-national governance also tend to generate distortions in the understanding and application of the norms that regulate WISEs' activities. Finally, this passion for rule-making and officialdom has France boasting a well-structured social economy, with dedicated resources at all government levels.

7.1.2 Historical developments of social economy

France has a long tradition of “social economy”, with the term credited to have been coined in the country at the beginning of the 19th century (European Commission, 2015). Before that, the genealogy of social economy is traced to authors such as Dunoyer (1830) or Du Buat-Nancay (1773), who put a strong emphasis on how the economy should contribute to the general welfare of a nation's population: “In the name of humanity, refrain from making the love for wealth the general spirit of a nation”⁹⁴ (Du Buat-Nancay, 1773, 24). However, it is not before 1884 and the repeal of the Le Chapelier Act of 1791 that France saw the first tangible applications of some of the ideals put forth by the promoters of social economy. Cooperatives and mutual insurances then started to develop in the country, largely inspired by the cooperatives movement started in the UK after the founding of the Rochdale Society of Equitable Pioneers in 1844. The most iconic of the French forerunners, the Albi glass factory, opened in 1986, is still in operation today, manufacturing state of the art glass bottles.

This historical context makes France a land where social economy has been striving throughout the 20th century, especially through cooperatives, mutuals and associations. Despite that, the term “social economy” reappeared on the political agenda only recently, in 1981, under France's socialist government of President François Mitterand. The topic gained traction progressively in the following decades as the “social and solidary economy” (ESS)⁹⁵ construct, where solidary economy was meant to introduce a wider range of initiatives not limited to traditional cooperatives and mutual insurances (Fraise, Guérin, & Laville, 2007).

⁹⁴ Au nom de l'humanité, gardez-vous bien de faire de cet amour des richesses l'esprit général d'une nation

⁹⁵ Economie sociale et solidaire. Throughout this thesis, we most of the time refer to the French “social and solidary economy” as “social economy” and we seldom use the ESS acronym.

7.1.3 Pursuing the Weberian tradition: current structuring of the social economy and work integration sectors

In France, work integration and social enterprises are two concepts that have developed in parallel. WISEs are therefore not systematically associated to social economy and are even opposed in some instances. In this section, we introduce these two aspects of the French context. We also detail how public value co-creation is structured in the sector of work integration, providing the basis to understand strategic interactions between WISEs and their public stakeholders.

7.1.3.1 The social and solidary economy

In 2018, France has a well-defined social economy sector that benefits from an established legal framework since 2014 and the passing of a dedicated act. The act on social and solidary economy outlines four general principles any entity has to comply with in order to be considered part of ESS:

1. Having an explicit purpose other than profit generation
2. Observing a democratic governance, enshrined in the entity's statutes, where decision making is not only conditioned by stake ownership or other financial contributions
3. Using a majority share of the profits to maintain or develop the entity's activity
4. The amounts constituted as reserve requirements cannot be shared

The act then refines further the type of entities that can be considered members of ESS:

- Associations, foundations, cooperatives and banking or insurance companies that follow a mutualist management model.
- Commercial entities that respect the following:
 1. Being aligned with the four general principals outlined above
 2. Seeking a social purpose as outlined in the law (i.e. primarily about social, health or cultural inclusion, education, integration, sustainable development and international solidarity)
 3. Locking at least 20% of yearly profits in a dedicated "development fund"
 4. Locking at least 50% of yearly profits as reserve requirements
 5. Having the entity's capital constrained by an asset lock

A High Council for Social Economy⁹⁶ was created after “*social economy*” became a political topic again in France in 1981. In 2012, a Deputy Minister was appointed for social economy and the High Council for Social and Solidary Economy (CSESS)⁹⁷ was founded, at the same time as parliamentary discussions started on the ESS Act. The 2014 Act is therefore the result of a process started in the 1980s, that saw the progressive development of a state apparatus dedicated to social economy. In 2018, the institutional framework for social and solidary economy in France is as followed:

- A High Commissioner for social economy, falling under the authority of the Ministry for Ecological and Solidary Transition. He has a competency across ministries to implement governmental initiatives about ESS.
- The CSESS, which was revamped in the 2014 Act, aims at facilitating a dialogue between members of the ESS and national and European instances, providing consultation on new regulations and promoting ESS.
- ESS France, the French Chamber for Social and Solidary Economy, tasked with representing and promoting the ESS towards the government and internationally, and with following up on the implementation of the 2014 Act.
- The National Council of Regional Chambers for Social and Solidary Economy (CNCRESS)⁹⁸ as well as its network of regional chambers for ESS (CRESS)⁹⁹, tasked with coordinating regional policies and the structuration of local networks for the development of ESS.

This considerable institutional framework provides France with one of the most advanced ecosystems for social enterprises in Europe (European Commission, 2015). The CNCRESS also maintains an observatory for ESS, that delivers detailed statistics on the topic. In the 2017 iteration of its statistical review, the CNCRESS counts 164,173 enterprises as part of the ESS in France (CNCRES, 2017), including 96 commercial companies (equivalent to PLCs and LLCs). Incorporating commercial entities in the scope of the ESS in France is new and stems from the 2014 Act. Data on these companies is therefore scarce and was reported for the first time in 2017. Excluding these companies, the ESS in France represents

⁹⁶ Conseil Supérieur de l'économie sociale

⁹⁷ Conseil Supérieur de l'économie sociale et solidaire

⁹⁸ Conseil national des chambres régionales de l'économie sociale et solidaire

⁹⁹ Chambre régionale d'économie sociale et solidaire

221,136 establishments and 2,373,812 employees (data at year-end 2012). These numbers are gauged to represent 10% of France’s GDP and 13% of private employment.

These figures are again placing France ahead of the pack, and well in front of Denmark in terms of how developed the ESS appears to be in the country. However, the scope of the data makes it difficult to compare internationally: all associations, foundations, cooperatives and mutual insurances are automatically accounted as part of the ESS in the published statistics. By comparison, the Danish accounting explicitly target entities having “significant commercial activities” (cf. 7.2), thereby likely to leave out most foundations and associations captured by the French statistics.

Table 18: The weight of different entities in ESS

Type of organisation	Value added in billion Euros	Number of employees in thousand
Associations and foundations	65.1	1,886
Cooperatives	14.7	307
Mutuals	10.1	131

Source: Data from the Conseil National des Chambres Régionales de l’Economie Sociale et Solidaire, 2017

7.1.3.2 Work integration

Work integration is a well-defined activity in France where it is considered a business sector in its own right. The first private initiatives in this field are traced back to the end of the post-war economic boom (a 30-year period known as “The Glorious Thirty”⁵ in France) when some welfare state logics denounced as passive started to lose ground. Private actors, often non-profit, then began to step in to fill a void left by both the changing state support schemes and evolving social and family structures (Mauger, 2001). The creation of FR2, one of our case studies, by a group of unemployed individuals in 1968 is a good illustration of this zeitgeist. The 1970s gave rise to more innovative practices offering vocational training and other service close to work integration as conceived today.

By the end of the decade, as these initiatives expanded, structured themselves in networks and gained tractions amongst policy makers, a whole administrative structure was put in place to oversee and fund

what soon got dubbed “*integration through economic activity*” (IAE)¹⁰⁰. Now a truly institutionalised activity through the ways of the Weberian state, work integration is framed centrally by a stringent set of rules and regulations designed by the DGEFP¹⁰¹, a central administration that belongs to the Ministry of Labour. A worthy observation is that the precedence of work integration (or IAE) over social economy in the political agenda does so that the two are rather distinct objects in France. This is observed on two levels:

- The way policy and legal schemes have been developed around work integration on one hand and social economy on another hand. These are two different policy setups, equipped with their own institutional ecosystems.
- The way operators perceive the two. This was observed during our field research, especially in the case of FR1, where both the interviewed staff and their public stakeholders were critical of some of the work carried out in the context of social economy and were stressing the distinction between social economy at large and work integration.

The cornerstones of the French regulatory framework for work integration are the **covenants**. They are awarded by the *préfet* (a direct representative of the central state authority for a specific territory) and administrated by the Regional Authority for Business, Competition, Consumption, Labour and Employment (Direccte)⁶⁷. Covenants set the scene for how WISEs can **co-create public value** by signalling the right for an enterprise to provide work integration services for the public, in a clearly delimited territory (often defined at county-level). Effectively, covenants are a form of institutional **endorsement** that creates a *de-facto* **legitimacy** for WISEs (as long as the covenants are in place) in a pure NWS tradition. This legitimacy translates into two main implications for the WISEs’ operational vertex: they are granted access to referring channels (mostly the job centres, through which individuals must obtain sponsorship⁶⁶ to access a WISE) and they are able to access subsidies from the Direccte for each beneficiary they support. Covenants can run between one and three years and have to be renewed by the Direccte upon expiration. The legal framework for work integration was reformed in 2014 and currently rests on four main types of covenants:

¹⁰⁰ Intégration par l’activité économique

¹⁰¹ Délégation générale à l’emploi et à la formation : General Delegation for Employment and Training : the central administration tasked with designing and developing the legal and financial framework to implement the central government’s work and training policy orientations. Work integration is ca. 7% of the DGEFP’s budget.

Integration workshops and building sites (ACI)¹⁰² can be an association, and in some rare instances affiliated to a public entity. ACIs target individuals the furthest away from the job market, with a heavy emphasis on soft skills development. ACIs produce goods and services that they sell on market terms, but these incomes cannot represent more than 30% of their revenue. The rationale behind this cap is the avoidance of competition distortions through subsidised labour. However, we noticed during our research, that with the decrease of subsidies allocated to work integration, this cap is becoming a growing challenge for ACIs who struggle to remain financially sustainable.

Work integration enterprises (EI)¹⁰³ can either be an association or a commercial entity (in more than two third of cases). EIs focus on the production of goods and services that they sell at market terms, using staff following a work integration path. EIs' staff carry out their assignment under direct supervision of the WISE employing them, which gives the individuals they are supporting an opportunity to develop the social and professional competencies necessary to sustain a job on the long run.

Temporary work integration enterprises (ETTI)¹⁰⁴, can either be a commercial entity or an association. ETTIs are temporary work agency that support the integration of their staff through temporary work assignments at customers' premises. Their operation is restricted to a "business to business" environment where they provide staff to their customers. ETTIs operate on the same standings as regular temporary work agencies and propose similar terms to their customers. Staff therefore operates at customers' facility without supervision from the WISE, which presupposes a high level of autonomy.

Intermediary association (AI)¹⁰⁵, can only be associations. AIs operate in similar terms to ETTIs, providing temporary labour to their customers as a work integration exercise. At the difference of ETTIs, AIs are allowed to operate in a "business to consumer" environment and are particularly active

¹⁰² Ateliers et chantiers d'insertion

¹⁰³ Entreprise d'insertion

¹⁰⁴ Entreprise de travail temporaire d'insertion

¹⁰⁵ Association intermédiaire

in this sector with services such as cleaning or child care. AIs also have a wider mission of outreach and advising vulnerable populations, with a legally defined territorial jurisdiction. Just as with ETTIs', AIs' staff operate directly at customer's premises without WISE supervision.

Each covenant has its own national network, mainly to promote members' activities, raise awareness and carry out advocacy work. WISEs covenanted as ACIs belong to the Chantier Ecole network (700 members), EIs to the Fédération des entreprises d'insertion (FEI, 512 members), and AIs and ETTIs to Coorace (548 members). Together, these networks include 1,750 of the 3,600 French WISEs and have all been involved in our research where they have provided support in finding relevant field data (cf. Chapter 6).

As evidenced above, covenants provide the framework in which WISEs can **co-create public value** in France and are therefore a key driver structuring the strategic interactions in the NWS environment. They determine part of the public value a WISE will co-create, by targeting different types of beneficiaries and having different aims. For instance, on one hand an ACI will primarily aim for a form of social rehabilitation, where work activities are just a pretext to re-socialise marginalised individuals. On another hand, an ETTI will target professional immersion at clients' premises and aim at long-term, sustained employment. These are two distinct forms of public value that also entail different approach to understanding and measuring performance.

Covenants also have a considerable influence on the **operational** vertex as they will contribute to define the type of commercial activities a WISE can have and, in some cases, (like ACIs) the amount of commercial revenue it can generate. Crucially for some WISEs, the type of covenant also defines the amount of subsidies they can claim, as shown in Table 19. Ultimately, and as we already pointed out, covenants are also an **institutional endorsement** that warrants the necessary legitimacy to operate. This legitimisation mechanism comes with an additional instrument.

Table 19: Summary of the four main work integration covenants

Covenant	Legal form	Minimum subsidy per FTE*	Count	Staff on inclusion scheme
ACI	Association or public body	€ 19,655	1,877	48,900
EI	Association or Commercial entity	€ 10,237	906	13,600
AI	Association	€ 1,331	2,690	50,600
ETTI	Association or Commercial entity	€ 4,351	260	13,300
Total covenanted WISEs		n/a	5,733	126,400

Source: France Active, 2018

* Full Time Equivalent, on inclusion scheme

The Direccte (and other public stakeholders such as the municipality in certain instances) are using monitoring mechanisms to track WISEs' outputs. Since these mechanisms are used (through what is achieved in the operational vertex) to maintain covenants in the legitimacy vertex (and thereby to maintain the WISEs' "*right*" to access funding and beneficiaries), we label them as **legitimacy devices**: a formal mechanism aiming to assess, understand, evaluate or measure what happens in the operationalisation and / or public value vertices, in order to *demonstrate* legitimacy (as defined in Chapter 2). Enforced by the Direccte locally, these devices are designed by the DGEFP.

These reporting mechanisms (devices) function as followed: once a covenant is setup, WISEs can negotiate their yearly goals with the Direccte. The main goal is defined in terms of "*dynamic graduations*"¹⁰⁶ ratio. Two other goals are added to the graduation ratio to form three criteria that are used in order to determine a "*modulation*" rate. This rate can range from 0% to 10% (also negotiated with the Direccte) and is applied to the basis amount of FTE subsidies (cf. Table 19) in order to define a bonus amount that covenanted WISEs can claim upon achieving their yearly goals. Table 20 summarises the modulation criteria and their rationale. This extended goal-based bonus system is one of the latest mechanisms (2014) introduced by recent reforms, further imprinting the "*neo*" in Neo-Weberian State.

¹⁰⁶ Sorties dynamiques : they encompass what is considered positive outputs when beneficiaries "*graduate*" from a WISE: long or short term employment contract or the pursuit of training or education.

Table 20: Modulation criteria

Criterion (ponderation)	Details	Rationale
Dynamic graduation rate (25%)	Dynamic graduations are either: - Towards sustainable employment (work contract for a duration of more than six months) - Towards transitional employment (work contract for a duration of six months or less) - Positive (staff hired by another WISE or pursuing training or education)	Aim at assessing whether WISEs meet their primary objective (i.e. to durably integrate marginalised individuals back into the workforce, or place them on a trajectory to achieve that goal)
Entry situation (35%)	Share of individuals on public benefits schemes	Aim at avoiding cherry picking, ensuring that WISEs hire the population targeted by the local authorities and
Integration effort (40%)	FTE ratio of supervisory staff (for training, mentoring and skills development)	Aim at being a proxy for the quality of the integration scheme delivered by the WISE.

Source: Ministry of labour¹⁰⁷

Covenanted WISEs must submit to a yearly “*management dialogue*”¹⁰⁸ involving chiefly the Direccte as well as some of its other local public stakeholders (essentially the local Job Centre and other relevant local authorities). This yearly conversation is used to assess a WISEs’ performance during the past year, which translates into the renewal of their covenant, determines how much of the modulation (bonus subsidies) will be paid to them, and serves to negotiate the next year’s modulation criteria. The dialogue takes place on the basis of a standard report that WISEs are due to provide prior to the conversation, but also based on any other piece of information the WISEs wish to put forward for their performance appraisal.

This means that, if a WISE has done any kind of social impact assessment (SIA) work, it is typically a place where they will try to leverage that work in an attempt to demonstrate to their public stakeholders the impact they have on the individuals they support (i.e. increase their legitimacy by

¹⁰⁷ <https://travail-emploi.gouv.fr/>

¹⁰⁸ Dialogue de gestion

documenting the public value they contribute to generate). The purpose of the management dialogue is to go further than the legitimacy device of indicator-based performance appraisal. It is therefore a place for **deliberation** where WISEs get a chance to argue for their case and still get a share of their modulation if they did not achieve their targets. Our cases provide compelling evidence that these yearly conversations play a crucial role in the strategic interactions between WISEs and their public stakeholders, contributing to a common definition of public value and therefore to the emergence of **collective preferences** (cf. Chapter 8).

The four covenants presented in this section cover the vast majority of work integration in France and lay the relatively stringent framework in which our case studies operate in their NWS environment. There are two additional statutes WISEs can apply for in order to carry out work integration activities in partnership with public actors: the Grouping of Employers for Integration and Qualification (Geiq)¹⁰⁹ and Neighbourhood Enterprises¹¹⁰. WISEs operating under these schemes have to be associations and may or may not be covenanted. In either case, they can still work in partnership with public stakeholders such as the job centres to identify candidates for their integration scheme.

Finally, there is also a very small number of entities operating out of the formal IAE (work integration) regulatory environment, as regular commercial companies. These entities function without a covenant but will almost always operate in conjunction with an ecosystem where some stakeholders are covenanted. Among the cases we researched, FR1 has a similar setup, with two “regular” temporary work agencies. These two companies are used to prolong the integration path of the individuals they support in the case where an individual’s sponsorship expires before they can reach sustainable employment. This is a case of **organisational design** strategy (in the operational vertex) that we explore further in Chapter 8.

The Direccte is the pivotal public stakeholder that structures most of the strategic interactions WISEs have with the public sector in France. A number of other public actors nevertheless contribute to shape the operational environment of WISEs. We describe them briefly below:

¹⁰⁹ Groupement d’employeur pour l’insertion et la qualification

¹¹⁰ Régies de quartier

- The Local authorities (municipality, county, region and others), which can:
 1. Offer ad'hoc subsidies to WISEs to cover a wide variety of needs, ranging from operational expenses to capital expenditures, or tied to the individuals they take on their integration schemes. In this regard, they may also contribute to set goals in terms of the individuals a WISE should target (e.g. in terms of social benefits beneficiaries) and might take part in the management dialogue.
 2. Offer public procurement contracts incorporating mandatory social provisions¹¹¹. These provisions stipulate that a defined percentage of the contract's value has to be carried out by staff on integration schemes. WISEs can tender directly to these contracts or act as a subcontractor for third-party tendering entity, allowing it to fulfil the contract's social provision. These provisions were created as a mean for WISEs to have a guaranteed access to public procurement contracts. While this is the case, our findings indicate that WISEs, even large ones like FR1 struggle to submit a competitive offer and often end up as subcontractors, which plays an adverse effect on their business development activities.
- The job centres, which is the institution signing off the sponsorship agreements that allows individuals to join a WISE's integration scheme (for a duration up to two years). The collaboration set up between WISEs and job centres implies that all job offers proposed by the former have to be brought to the knowledge of the latter. The job centre supports the recruiting effort or approve for sponsorship the candidates brought directly by WISEs. They are also involved in the follow-up of the individuals throughout their integration path and participate in the final appraisal at the end of the scheme. Job centres are also involved in the management dialogue.
- The County Committee for Work Integration¹¹² has more of a strategic role and has seldom direct interactions with WISEs. It outlines a work integration action at county-level and has an advisory role in the management of covenants and in the administration of related work integration public funds.
- The National Council on Work Integration (CNIAE)¹¹³ seldom has direct interactions with WISEs either. Its main role is to advise the government at the national level on questions pertaining to work integration. It also acts as a facilitator across the national networks, promoting and disseminating best practices.

¹¹¹ Clause sociale

¹¹² Conseil départemental de l'IAE

¹¹³ Conseil national de l'IAE

These relationships are substantially more formal than the ones described in the context of social economy. This is mostly driven by the magnitude of public funding that is poured into work integration (whereas no particular subsidies are available to operators of the social economy solely based on their belonging to the ESS) and intensified by the bureaucratic and legalist inclines of the NWS. During an interview at the Ministry of Labour as part of our field research, these amounts were gauged to the tune of 800 million euros, raising to one billion euros if reduced social contributions for employees on integration schemes are included.

These financial stakes are part of the considerations on performance we observe in Chapter 8 and contribute to shape strategic interactions in all three vertices of the PVT triangle, particularly in the context of public funding stringency that is described by field operators. However the formal mechanisms of endorsement (covenants) and control (legitimacy devices) introduced by the NWS tend to shift strategies towards the legitimacy vertex, where WISEs need to comply to legalist rules in order to exist. To this extent, Gianfaldoni speaks of “quasi-hierarchic rapports which develop on the basis of subordination relations between public actors and WISEs”¹¹⁴ (2018, 86). It is on these premises that we develop our analysis of performance measurement in Chapter 8.

¹¹⁴ « rapports quasi-hiérarchiques s'établissent sur des relations de subordination entre acteurs publics et SIAE »

7.2 Denmark: work integration and social economy in the New Public Governance

How do Danish WISEs co-create public value with their public stakeholders? In this section, just like we did for France, we start by introducing the Danish PAM framework at a high level, showing how its decentralised design builds on the tenets of the NPG. We then proceed with a short history of the development of social economy in the country and the particular place occupied by work integration in the conception of social economy and SEs. We conclude by looking at the current institutional framework laying down the foundation for the cooperation between WISEs and Danish municipalities (via the job centres), and how that plays into strategic interactions through the considerations we attach to PVT.

7.2.1 Characterising the Danish PAM framework

In Chapter 2, we indicated our choice of ascribing Denmark mainly to a NPG model. In this short introductory section, we present the main features of the Danish PAM framework and we explain how we tie them to a NPG model through three main features: **decentralisation**, the place of **networks** (and the forms of cooperation they convey) and the role of **relational capital** (and trust as its corollary). This section is by no mean a thorough reflection on the Danish political system and its organisation. It is merely a high-level overview to equip us with the main keys to navigate Denmark's institutional environment through our analysis of WISEs and their strategic interactions with public stakeholders.

Just as France epitomises centralisation (albeit a degree of decentralisation in recent decades, Hammerschmid et al., 2016), Denmark is one of the most **decentralised** country in the world (Greve, 2006), both in terms of geography and institutions. This means that in the central government, ministries typically enjoy a high degree of independence while autonomy is also experienced by local governments, who have a great latitude in implementing national policies and in developing their own initiatives, levying their own taxes (while still depending on block grants from the central government). Municipalities¹¹⁵ are the central unit of the local government and enjoy jurisdiction across a wide portfolio encompassing the areas of interest for our research: social inclusion and employment services

¹¹⁵ Kommuner. Interviewees tend to refer to municipalities and job centres indifferently, as they have interactions with both stakeholders and the subordination links between the two organisations (from the municipality to the job centre, cf. Figure 22) blur the differentiation.

(i.e. training, mentoring or work experiences)¹¹⁶. Since a reform enacted in 2007, there are 98 municipalities in Denmark, grouped in five regions (270 before the reform). Regions are mostly an administrative aggregate, with no tangible power and few policy prerogatives apart from healthcare.

The Danish decentralisation is synonym with “*popular control*” (Greve, 2006), where citizens are expected to contribute actively to their policy environment. Existing research show that interactions with interest groups, actors from the third sector or looser citizens’ movement are institutionalised (Christiansen & Nørgaard, 2003). Similarly, and more centrally, collective bargains are recognised as a defining feature of the Danish welfare system (Greve, 2006). This leaves a substantial place for the development of **networks** and for their capacity to influence policy making (Sørensen, 2002). Or as Greve puts it: “*In short, the [Danish] political system can be characterized as a consensus democracy with strong emphasis on co-operation and integration of interest groups, organizations and citizens*” (2006, 163).

We showed in Chapter 2 how networks are a core element of NPG and how they help shape the cooperation mechanisms that characterise NPG (Osborne, 2006, 2010). In Denmark, and in the context of our research, this translates into the close collaboration we describe in this chapter between WISEs and municipalities (through the job centres). This relationship is founded on market mechanisms that have been inspired by NPM reforms in the 1990s, as we show below. But it is also rooted in the traditional culture of consensus and collaboration that is a characteristic of the Nordic states (cf. Table 1), which translates into a high level of **relational capital** and is reflected throughout the historical development of the Danish social economy as it is described below. Greve summarises this idea as followed: “*Denmark has ‘modernised’ to a great extent and ‘marketized’ to a lesser extent. Radical marketization strategies have been held back, as there has not been a majority in Parliament for such a strategy, and there is a consensus on developing the public sector rather than dismantling or minimizing it*” (Greve, 2006, 168).

¹¹⁶ Employment services (servicearbejde) is a generic term used in the Danish legislation for the provision of training, mentoring or the provision of work experience to the general unemployed population. This includes “*work integration services*” as we use it in our thesis for particularly vulnerable individuals. Because of our focus on WISEs, for the Danish context, we use “*work integration services*” and “*employment services*” interchangeably, although the former is a subset of the latter.

7.2.2 Historical developments of a social economy in Denmark

Denmark has made a formal entrance in the space of social enterprises rather late, with no concrete policy framework around the issue before 2013. However the national discussion about social enterprise started before that, with for instance a consulting entity called the Centre for Social Economy¹¹⁷, founded in 2008 partly on public funds (and closed in 2014) or the creation of Social Entrepreneurs in Denmark (SED)¹¹⁸, an association started in 2010. This is still rather late when compared to neighbouring Scandinavian countries like Sweden, where social enterprise is an item on the policy agenda since the 1990s.

When the concept of social enterprise started to gain traction in Denmark in the 2000s, it was primarily in relation to **active labour market policy**, with a view to make small and medium “*mainstream*” enterprises more geared towards integrating a marginalised labour force. These companies would therefore earn the name of “*sociale virksomheder*”¹¹⁹ for aiming to address unemployment issues. No specific definition was associated to this phrasing. But at the time of its emergence, the terms “*social*” and “*enterprise*” were used in conjunction with each other more to colloquially characterise firms displaying unusual behaviours rather than as a fully fledged concept of social enterprises in the breadth of its current understanding. Social enterprises in Denmark have therefore been very intimately associated to work integration activities, whether or not these activities were carried out by organisations befitting the definition of a social enterprise.

Hulgård and Bisballe (2004) attribute this misconception to a lack of interest among institutional actors to understand and foster the potential of social enterprises. They link this apparent apathy to four potential factors:

1. The lack of a legal framework for cooperatives or similar type of bodies in Denmark, which could allow socially-driven organisation to develop productive and commercial activities without risking of compromising their identity.

¹¹⁷ Center for Socialøkonomi

¹¹⁸ Sociale Entreprenører i Danmark

¹¹⁹ Social enterprises

2. The lack of a specific discourse that could differentiate social enterprises and social economy from civil-society based organisation with a wider purpose on democracy or the provision of welfare services.
3. A reluctance to associate the notions of “social” and “economy” together in a fruitful manner. *“Especially from people within the social sector, because they consider ‘social enterprise’ and ‘social capital’ as being oxymoron”* (Hulgård & Bisballe, 2004, 4).
4. A perception that social enterprises are irrelevant in a Scandinavian country where the welfare state provides for the population.

However, and perhaps paradoxically, Denmark has a rich history of social economy, dating back to the middle of the nineteenth century, with the development of a strong cooperatives movement, that played a key role in shaping the modern Danish national identity (Andersen, Gawell, & Spear, 2016; Mordhorst, 2005). This movement is traditionally split between the farmers’ cooperative movement (andelsbevægelsen) and the workers’ cooperative movement (koorporation). The former is the most influential front and is strongly associated to the philosophy of Nikolaj Frederik Severin Grundtvig and the development of the Folk High Schools¹²⁰, which both have a substantial legacy on the contemporary Danish society and political landscape.

A pastor who turned to inter alia poetry, philosophy and politics, Grundtvig’s ideas are indeed associated with many of the Danish’s society peculiarities (as well as with some of the Nordic countries’ in general) and his influence has certainly contributed to shape the Danish welfare society in the way it functions today. The development of the Folk High Schools, started in 1844, is a good illustration of his philosophy, as he was concerned over the enlightenment of the populace, who did not have the money or the time to spend on education. Grundtvig started the folk school movement on these premises, preoccupied with personal development and providing education mostly through oral language and **informal** means (no exams held or diploma awarded). Such schools still exist in Denmark (approximately 70 of them, typically established in rural areas) and focus mostly on providing informal education to adults on topics like arts, religion, philosophy, but also sports or design.

¹²⁰ Folkehøjskole

Authors like O. D. Campbell (1928), Davies (1931) or Hegland (1994) stress how the ideas underpinning the folk schools (an equal, free and unformal access to the public good that is education) underpin aspects of the contemporary Danish welfare society. They are also credited for having generated the necessary amount of social capital and trust to allow for the creation of the Danish farmers' cooperatives (Horton & Kohl, 1997), which design was informed primarily by Grundtvig's teachings. The cooperation and solidarity dynamics established by the farmers' cooperatives and later by the workers' cooperatives (the latter were initially very reluctant towards cooperatives) are also part of a fundamental imprint contributing to shape the modern Danish welfare state.

From the 1960s, as the public sector was progressively taking over social support functions that would typically be fulfilled through cooperatives in the past, private organisations, often linked to the historical cooperatives movements, would then establish themselves as "*self-owning*" entities¹²¹ to carry out this work in **conjunction with public actors**. In parallel, other organisations, not linked with historical cooperatives started to emerge locally as bottom-up social experimentations towards the late 1960s. Hegland (1994) attributes these partly to Grundtvig's legacy. This trend became fully fledged in the 1980s and the 1990s when public funds were poured into the experimentations of local projects in the health care, social and cultural areas. The most iconic initiative, the Social Development Program (from 1988 to 1992), facilitated the launch of numerous initiatives that Hulgård and Bisballe consider as SEs (2004) and its lineage to Grundtvig's ideas was explicitly articulated by the Danish government. These two decades of experiments have contributed to create a culture of experimentation in the social space, with a particular focus on the development of local solutions. This in turn fostered a specific Danish discourse on local, experimental projects and innovation that can still be observed today (Andersen & Hulgård, 2008).

A final element that contributes to the historical development of work integration activities as we observe them today in Danish WISEs is the 1994 labour market reform, which turned around labour market policies from passive measures to active measures. The underpinning idea was to move from a rule-based system to a needs-oriented system focused on individuals' needs. Typically this would

¹²¹ Selvejende institution, a Danish legal form where the entity is self-governed, with its own assets and no formal owner able to influence activities. It is typically adopted by organisations from the third sector such as associations or foundations.

translate into a wider array of activation opportunities (training schemes, job placements, etc.) available at any point during the unemployment period rather than after a determined amount of time in order to allow a more tailored and faster return to employment. This reform also allowed more regional freedom, allowing local employment strategies to be defined at the municipal level.

This framework also favoured the cooperation with private sector organisations such as WISEs, by e.g. initially mandating that 20% of employment activities should be carried out by non-public organisations. If, like Hulgård and Bisballe (2004), we consider the workers' and farmers' movements, together with the folk high schools, as the first founding pillar of modern WISEs; and the wave of bottom-up, social experiments from the late 1960s and onwards as a second pillar; then the labour market reform of 1994 (and the ones having pursued the "*flexi-security*" framework instituted then) is certainly the third foundational element contributing to explain the nexus observed today in the Danish WISEs and the way they collaborate with local public actors.

7.2.3 The fluidity of NPG? Current structuring of the social economy and work integration sectors

Contributions such as Defourny and Nyssens (2008) or Hulgård and Bisballe (2004) highlight that Danish SEs are heavily dependent on the state, relying on the public sector for support (e.g. through subsidised staff) or commercial activities (i.e. by having little diversification in their customer base, other than public clients). Furthermore, the tasks carried out by some SEs can be of the same nature than what is already proposed through public welfare services. This is especially the case for WISEs in the framework we describe above as laid down by labour market reforms, where they provide work integration services similar to those available through public welfare services. The differentiation is often that WISEs focus on a specific target of beneficiaries (e.g. ex-offenders or visually impaired people), and thereby offer services tailored to a specific population.

These are traits we observe in our field research, where a large majority of both our Danish case studies' revenue stems from public sources, with public stakeholders (the job centres or the municipalities) as clients. Although garnered from public sources, this revenue is a form of **commercial activity**, since WISEs supply their services at market terms. WISEs can also benefit from direct public support under the form of subsidised staff, where part of the employees' salary is supported by the municipality.

Among such mechanisms, the “*flexi-job*” scheme is one that is used most. Under this scheme, employees with a hampered productivity can work for a limited number of hours, paid by their employer, and get a salary complement from the municipality compensating for non-worked hours (within set boundaries). This form of subsidy is common among WISEs’ staff as they seek to provide employment to individuals who may not be able to sustain a job under normal conditions. Since it is tied to the individual and not to the employer, it is however not a form of subsidy that is exclusive to WISEs.

Such support schemes play an important role in most WISEs’ operational vertex (e.g. in our cases, this is what allows DK1 to reach an 85% ratio of staff on integration scheme, with an important part of its management also employed on flexi-jobs). However, contrary to contracted work integration services, subsidies schemes are a form of resource WISEs have little control upon. Subsidies schemes are tied to the persons benefiting from them, on the basis of an individual evaluation, ran by the local job centre (sometimes through a WISE) and which can be ongoing for years. While some WISEs get involved with conversations between the municipality and the job centre to try and weight-in in the decisions concerning their employees, the burden of securing access to a scheme remains largely on the individual.

As an example, in our empirical material, we find one case where a person suffering from a handicap preventing her from working full time got offered a job at DK1, provided she could secure a flexi-job scheme to ensure a decent revenue. Facing repeated decline to access a flexi-job scheme in her local job centre, she relocated to another municipality in order to get a different assessment from another job centre and finally start on her job at DK1. In this instance, as in a vast majority of cases, DK1 simply could not intervene in order for this person to enter the flexi-job scheme (her admittance on the scheme being considered purely on the basis of her assessed capacity to sustain a fulltime job, which in this example proves to be quite subjective).

Moreover, while municipalities and job centres have the authority to decide how to enact subsidies schemes, their design and the rules governing their implementation is decided centrally through the law and parliamentary debates. Here again, although WISEs and other interest organisations can lobby

in order to influence legislative outcomes, their bandwidth is very limited. The 2013 reform of the flexi-job scheme (mostly tightening access rules and changing the subsidies payment methods) is an example of how little influence WISEs manage to exert on the final text, despite being strongly opposed to the reform.

As a consequence, and although we acknowledge the importance of subsidy schemes for some WISEs' operational vertex, we keep the questions tied to the issues of subsidies out of the scope of **strategic interactions** in Denmark. The WISEs taking part in our research simply have too little influence on this aspect of their operations for us to consider it a relevant strategic dimension.

However the relations WISE have as third-party providers of work integration services constitute the nexus of our analysis. When proposing their services, Danish WISEs can compete with one another (although seldom observed in practice), and with local public work integration services. Contrary to France where work integration is a very regulated space, there is no formal endorsement mechanisms such as covenants in Denmark. This means that actors operate, to some extent, in a **quasi-market**, supplying competitive offerings to public clients (the municipalities)¹²² having competing services of their own.

Quasi-markets are typically a feature associated to NPM (cf. Table 2 and see e.g. Bartlett & Le Grand, 1993; Le Grand, 2009; Ranerup & Norén, 2015). They aim at providing public services through supply and demand mechanisms, allowing to deliver public good while presumably reaping efficiency induced from market mechanisms (Le Grand, 2009). This is the type of framework that was instated in Denmark through the 1994 reform, encouraging the use of non-public providers for the provision of employment activities.

¹²² Here we emphasise that market mechanisms are specifically concerning the delivery of work integration services: Danish WISEs provide these services to customers (municipalities), while French WISEs must seek an administrative authorisation (the covenant). WISEs in both countries are competing for public contracts (e.g. construction or gardening) in order to generate commercial, productive activity and revenue. In Denmark, the market dimension is **also** observed for the provision of employment services (and the access to candidates), which is a paid-for service provided by the WISEs to the municipality (over and above any contracts warranting commercial revenue through activities such as construction or gardening). This is not the case in France. Our topic of interest is around the provision of work integration services, not around the securing of public contracts, for which similar or other performance measurement tools than the ones considered in this thesis may be used, depending on the contractor and the type of activity.

While a number of non-SEs operate in that space, catering to the needs of the general population (i.e. providing job placement and training services to individuals with no specific employment challenges identified), Danish WISEs manoeuvre in a subset of this quasi-market where their social imprint sets them apart as actors specialised on niche beneficiaries. This niche environment, together with the close proximity WISEs entertain with public stakeholders while co-creating public value, is what leads us to the qualification of **relationship-based quasi-market**. We believe this is a fitting description for what we observed in the field, where market mechanisms are active to a degree, but tamed by the common objective of public value co-creation and stakeholders shared values (cf. Chapter 9). It is also very reminiscent of Greve's statement about radical marketizing strategies being held back in Denmark (2006).

The specific public value WISEs offer is recognised by their public stakeholders, who provide them with a dedicated space to operate, but still under supply and demand mechanisms offering the premises for competition (especially against existing public offerings) and market instruments such as negotiation processes between actors. It is this mix of market forces and administrated co-creation of public value that we place behind the term of relationship-based quasi-market. We tie this concept to the way Hulgård & Andersen frame social entrepreneurship in Scandinavia: *"The welfare state makes an active contribution to initiating and developing social innovation and social entrepreneurship, and does not simply leave this to market forces and entrepreneurial citizens"*. (2015, 42). This type of setting is highly reminiscent of the values of trust and sustained cooperation and networks carried by the NPG (cf. Chapter 2) and is consistent with Osborne's view of NPG's **plural** complexities (2006, 381).

It is therefore the strategic interactions occurring in this relationship-based quasi-market setting that are the object of our enquiry: they frame the way that WISEs co-create public value by ensuring the level of legitimacy required to secure favourable operational conditions amid market forces. We of course focus on the role of performance measurement as an influence in these interactions.

7.2.3.1 Current institutional framework for work integration and social enterprises

The institutional framework currently in place in Denmark for WISEs is on the opposite side of the spectrum of what is observed in France. As a primer, we noticed throughout our field interviews an

enduring confusion between social enterprises and work integration activities, even at the Danish Business Authority¹²³, which is currently the Danish state authority having a jurisdiction on questions related to social enterprises. There, as our conversation touched upon the impact of SEs in general, respondents kept developing about the importance of reintegrating individuals at the margin of the labour market, clearly limiting their perception of social enterprises' activities to that of WISEs. This is consistent with the facts highlighted by Hulgård and Bisballe (2004) that we present earlier in this chapter, about SEs being perceived primarily in terms of labour market policy. It is nevertheless surprising to find these conceptions in the state's strata dedicated to the issue. The reciprocal to this first observation, marking a stark contrast with France, is that work integration as a sector is closely intertwined to the conception of social economy and social enterprises.

A second difference with France is the quasi-absence of regulatory and institutional frameworks for neither social economy nor work integration. Table 21 summarises the main milestones that can be associated to the creation of a formal framework for SEs in Denmark. The Socialist government elected in 2011 drove some of the most emblematic initiatives. These include the creation of the National Centre for Social Enterprises¹²⁴ in 2013 and the passing of the law on registered social enterprises in 2014. The Centre was the first state administration dedicated to the development of social enterprises in the country. It followed a series of local initiatives such as municipalities' strategies for social enterprises (cf. examples in Table 21) and other semi-public funded initiatives such as the Centre for Social Enterprises (created in 2008) with an aim at impulsing a national agenda for the development of SEs. A 10-people strong organisation, the Centre was running programmes to support SEs in their development (including on questions about SIA) and provided them with a voice within the government (the Centre was affiliated to the Ministry of Social Affairs).

¹²³ Erhvervsstyrelsen

Table 21: Main institutional milestones for the social economy in Denmark

Year	Development
2008	- Creation of the Centre for Social Economy (private consultancy, partly funded by the public, closed down in 2014), started with a grant voted by the Danish Parliament
2009	- The municipality of Kolding adopts a strategy for social enterprises, aiming at establishing 10 SEs
2010	- The municipality of Copenhagen adopts a strategy for social Enterprises - Creation of a National Strategy for Social Entrepreneurship (an initiative from the think tank MandagMorgen) - Publication of a National Civil Society Strategy (government)
2012	- The municipality of Aarhus adopts a strategy for social Enterprises - Creation of the Social Capital Fond, a social impact investor funded through public grants and local foundations
2013	- Creation of the Committee on Social Enterprises - Creation of the National Centre for Social Enterprises ¹²⁴ (affiliated to the Ministry of Social Affairs, closed down in 2015)
2014	- Creation of a law on registered social enterprises (L 148 Forslag til lov om registrerede socialøkonomiske virksomheder)
2015	- Silkeborg strategy on social enterprises

Source: Author

After a majority shift to the right of the political spectrum in the 2015 general elections, the Danish government did not renew the Centre’s budget, which closed down at the end of the same year. The social economy portfolio was passed on to the Danish Business Authority which, at the time of our research, had one person managing it, together with several other attributions. Admitting to the fact that *“the Business Authority’s activities with SEs is perhaps 1% of all the things we are doing”*, the personnel we interviewed acknowledged that the development of SEs was not a priority despite the value they may contribute to generate. This lack of state engagement was felt across our case studies, which all deplored the closing down of the Centre and interpreted it as a state signal of the lesser importance attributed to SEs.

¹²⁴ Vækstcenter for socialøkonomiske virksomheder, referred to as “the Center” in the rest of this section

As per the 2014 law on registered social enterprises, it establishes a framework for what can formally be acknowledged as an SE in Denmark, giving right to an official “SE mark”. However this mark is not associated to any kind of tangible benefits. According to the law, in order to be acknowledged as an SE, an organisation must:

1. Have an explicit social purpose
1. Be business-driven¹²⁵ (the idea here is that revenue must be derived, to a certain extent, from commercial activities. This extent is however not specified in the law)
2. Be independent from public authorities
3. Have an inclusive and responsible governance: it must involve employees, customers, partners and stakeholders. In addition, the company must be managed responsibly in accordance with the social objectives
4. Manage profits with a social purpose (i.e. re-invest a majority of profits in the enterprise itself or in other social projects). However, a limited share of profits may be distributed as dividends.

Interestingly, when outlining its definition of what is a social enterprise in Denmark the Committee on Social Enterprises chose to explicitly exclude specific criteria (Committee on Social Enterprises, 2013):

1. The enterprise having to be operating as a specific corporate entity
2. The enterprise having to employ vulnerable individuals.
3. The enterprise having to be innovative

These were deemed too restrictive, but they also inform us on the fact that there was a clear intent from the Danish administration to not restrict its conception of SEs to work integration. At the time of the research⁵⁸, 208 SEs were registered under the 2014 law, while it is generally considered that there are around 300 SEs in Denmark, with roughly half being WISEs (cf. Table 22).

¹²⁵ Erhvervsdrivende

Table 22: Social enterprises census in Denmark

Author	Year	Scope / chosen definition for SE	Identification method	Social enterprise count
The Committee on Social Enterprises ¹²⁶	2013	<ul style="list-style-type: none"> - Has a social purpose - Has a commercial activity - Reinvest most of its profits in the company or the cause - Is independent from the public sector - Has an inclusive governance 	Unknown	“Close to 300” (p. 21)
Thuesen et al. (SFI) ¹²⁷	2013	<ul style="list-style-type: none"> - Has a social purpose - Has a commercial activity - Reinvest most of its profits in the company or the cause - Employs marginalized individuals - Has a tax registration number 	“Various Sources” (p. 47)	115 registered companies, representing 129 entities
Jensen (CABI) ¹²⁶	2012	<ul style="list-style-type: none"> - Has a social, health or environmental purpose - Has a commercial activity - Reinvest profits in the company or the cause - Is independent from the public sector - Has a tax registration number 	Questionnaire sent to job centres and municipalities, ad’hoc interviews with stakeholders	“Minimum 300” (p. 2)
Jacobsen ¹²⁸	2013	<ul style="list-style-type: none"> - Has a social purpose - Has a commercial revenue - Has a positive revenue - Has a positive value-added - Has a positive income - Has a tax registration number - Employ vulnerable individuals - Has a minimum of 3 FTEs¹²⁹ 	Query on Denmark Statistics’ micro data (2006 data)	409

¹²⁶Jensen (2012) and the Committee on Social Enterprises (2013) adopt a definition that is very close to the one proposed in the Danish law, and both suggest a number of social enterprises of “around 300”. We therefore take this number as our default hypothesis for the Danish social enterprises population, bearing in mind this is not a perfect measure.

¹²⁷Thuesen et al. (2013) adopts an approach that seems too restrictive compared to EMES’ guidelines or the definition provided by the Danish law, which clearly appears to be a focus on WISEs (i.e. employment of marginalised individuals as a formal criteria).

¹²⁸Jacobsen (2013) adopts a statistical definition of social enterprises that appears quite far from EMES’ guidelines or the definition provided by the Danish law. The number of social enterprises that comes out of his research (409) also appears higher than in the rest of the literature.

¹²⁹ Full time employees

7.2.3.2 How Danish WISEs operate in this institutional framework

The institutional environment we describe here is very loose, and very much in contrast to the French one. The absence of political agenda for the development of SEs seems to be mainly the result of political forces. However it is still a strong marker of the lack of a sustained interest for developing the sector at the national scale. We saw that the Literature sometimes associate this lack of interest to a traditionally robust welfare state where the place of private sectors is not felt as much needed (e.g. Hulgård & Andersen, 2015; Hulgård & Bisballe, 2004). We also found a similar explanation delivered to us at the Danish Business Authority.

Things are however different at the local level, where “*local strategies*” (cf. Table 21) are being designed to support the development of SEs (here again largely understood as WISEs). These often consist in quantitative targets (in terms of how many SEs should be developed in an area) and may provide support in different ways, through subsidies, discounted space or supplies, etc. We observed that these strategies are primarily developed to answer a demand for work integration services (i.e. for the development of WISEs). As a side comment, we note that the first municipality which proposed a strategy for SEs (Kolding, in 2009) is also where the first Folk high school was open.

These local initiatives on SEs are clearly a consequence of Denmark’s decentralised administration, with a considerable independence granted to municipalities in the tackling of their issues. Leaving social economy aside, and focusing on work integration and WISEs, we saw that labour policy reforms enacted in the 1990s also granted a high degree of liberty at the municipal level to organise active labour policies. The 2016 Act on active employment services¹³⁰ is the latest addition to that corpus and is summarised as followed through a correspondence with the Ministry of Labour:

“The municipalities have overall operational responsibility for employment services, and decide for themselves how to facilitate their active employment services aimed at citizens and enterprises. The Act on active employment services permits the municipalities to outsource tasks and decision-making relating to external providers.”

¹³⁰ Bekendtgørelse om en aktiv beskæftigelsesindsats

The Act provides the legal framework through which municipalities can contract WISEs (and other providers of employment services) for their services. The contracting occurs through job centres, which are under the municipality's authority and are WISEs' main interlocutors. Job centres are tasked with welcoming all unemployed individuals and place them on a track towards employment. Tracks are individualised and job centres can decide to which employment services provider (if needed) a particular individual should be sent, depending on their training needs and personal profile. This is where WISEs come into the picture, as providers of employment services for particularly marginalised people.

Just like in France, Danish WISEs are therefore dependent on job centres as a referring channel to access the individuals they wish to support. However in the framework laid down by the Danish legislation, WISEs are also executing a paid-for employment service. This service can consist in either assessing the working capability of an individual¹³¹ or in equipping them with basic skills to sustain a job¹³². In our research, this constitutes the two basic components of public value that Danish WISEs create: for each beneficiary, WISEs and the job centre co-define a desirable outcome before WISEs start their activities. This is also what constitutes the premises for the evaluation of WISEs performance: whether the agreed-upon public value has been reached.

These two types of services are performed against a price negotiated with the **municipality** (not the job centre), usually through yearly framework agreements. The individual is being followed by a **job centre's** social worker during the duration of the contract, who interfaces with a counterpart within the WISE to be kept abreast of progress and potential challenges. WISEs maintain a direct liaison with the municipality to foster commercial conversations and discuss performance, while the job centre's case workers are also providing feedback to the municipality on WISEs' performance (cf. Chapter 9). These strategic interactions between WISEs and their public stakeholders are depicted in Figure 22.

¹³¹ Referred to as "*afklaring*" (clarification) in the Act on active employment services, which can run for up to 13 weeks

¹³² Referred to as "*ressourceforløb*" (resource development) in the Act on active employment services, which can run for up to five years

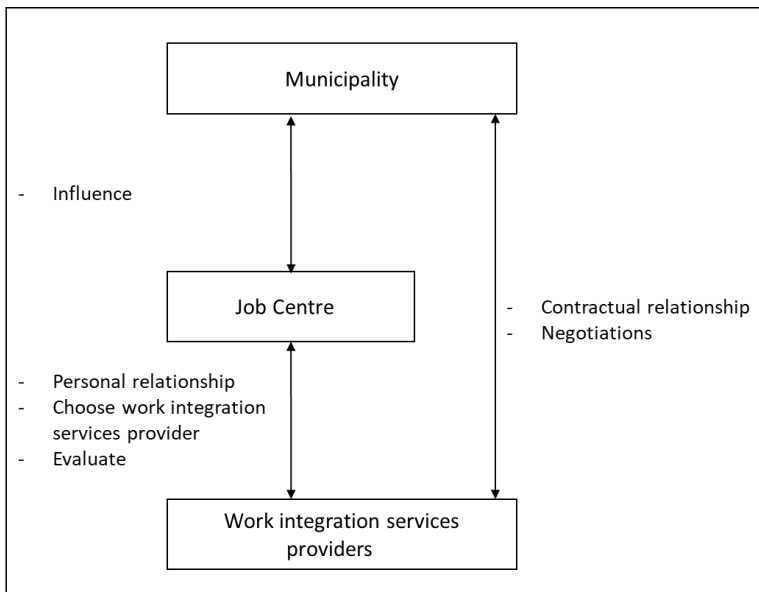


Figure 22: Strategic interactions between Danish WISEs and their public stakeholders

Source: Author

The Danish job centres, empowered by the municipalities, therefore have a dual, entangled role of **customer** and **referring channel** for Danish WISEs. In this setup WISEs can establish ongoing relationships with the job centres and the municipalities, where they can establish themselves as “preferred suppliers” (Osborne, 2006, 383) in a relationship-based quasi-market. We observed this in our Danish cases: they both have a historical relationship with the job centre and the municipality, public social workers know the WISEs well and have a natural incline towards repeating business with them. This contributes to build the **relational capital** that translates into an increased trust, leading to the “negotiation of values [and] meanings” (Osborne, 2010, 10) observed in NPG.

It is on the premises instituted by the relationship-based quasi-market for public value co-creation that we develop our analysis in Chapter 9. We investigate further how this environment translates in identifiable strategic interactions in the PVT triangle, and how these interactions contribute to understand the way the performance of Danish WISEs is measured in the NPG.

Conclusion of Chapter 7

By linking the conceptual models presented in Chapter 2 to the reality of France and Denmark’s PAM, this chapter answered the overarching question: how do WIES **co-create public value** in Denmark’s NPG and France’s NWS?

Equipped with this concrete understanding, we are ready to move on to our two next chapters that explore, in the two environments we just described, the nature of the strategic interactions occurring between WISEs and their public stakeholders and the resulting performance measurement models. What do French WISEs do with the centralised performance measurement imposed by the Direccte? Are these rigid indicators a source of performativity such as retroaction? Do WISEs manage to convey their own sense of performance through management dialogues? And how does that compare to Denmark, where no defined measurement is used? What is then the place of market mechanisms?

As Chapter 8 (France) and Chapter 9 (Denmark) will bring answers to these questions through practical case studies, we will gain an understanding of how strategic interactions influence performance measurement.

Chapter 8 Performance measurement in France: between autocracy, managerialism and deliberation

Having explained the mechanisms through which public value is co-created in the French **Neo-Weberian State** (NWS, Pollitt & Bouckaert, 2004) in the previous chapter, here we explore the concrete aspects of the cooperation between French work integration social enterprises (WISEs) and their public stakeholders. Doing so, we expose the different **performance measurement models** we observed (including WISEs' social impact assessment -SIA- models) and we characterise the **observed strategic interactions** using public value theory (PVT, Moore, 1995).

As we develop this chapter, we answer the following questions: What are the strategies deployed by WISEs to contend with the legalist framework imposed for work integration in France? How does that translate in performance measurement practices? We also explore the implications of having concurrent performance measurement approaches running in parallel: the Direccte's graduation-based monitoring tool and WISEs' own SIA models. Is it a source of performativity? Or does it prevent it? Does it have an effect on deliberation? These interrogations start to address the core of our topic, and they are the last stage of our research before we explore them under the angle of the strategic intents they are underpinned by in Chapter 10.

We start by providing an overview of social impact assessment (SIA) practices in France. Our first entry is what we identify as two competing socio-technical arrangements (STAs) in France for measuring the performance of work integration social enterprises (WISEs): "*social utility*" and "*social impact*". We then proceed by describing some key trends identified through the documentation provided by Avise and KPMG (in the absence of our own quantitative survey). We continue with a presentation of the performance measurement models observed.

In the next section, we use the PVT strategic triangle to chart the strategic interactions between WISEs and their public stakeholders. We find that strategies lean heavily towards the **legitimacy** vertex, which we tie to the way the NWS structures WISEs' public value co-creation in the country. We identify legalism (through mechanisms such as covenants and compulsory output reporting) as the main driving

force behind the emphasis of a legitimacy that is not directly connected to public value creation but more to how WISEs comply with the rules.

In this context, we identify **bureaucracy** and the layered levels of local governance as a vector of **retroaction**, as local administrations develop their own targets for WISEs to pursue. We however see that the “*neo*” tenet of the NWS introduces **deliberation** mechanisms at the local level that prevent **indicator-substitution** issues which could otherwise easily happen as a consequence of centralisation. We also observe that the same mechanisms can be a driver for the emergence of unrealised STAs, such as SIA, which can attempt to **circumvent** the centralised STA of performance measurement.

We close this chapter by building further upon the tension between the “*neo*” and the “*Weberian*” components of the NWS and explaining how the sometimes-diverging mechanics at play contribute to position work integration in France in a **hybrid performance regime**, trying to move away from its autocratic anchor, where centralisation gets it stuck between managerialism and deliberation.

8.1 Observed practices

In this first section we introduce the general performance measurement practices observed in France. This initial approach is first characterised by a focus on what we identify as two competing STAs in the country for measuring the performance of WISEs: one based on “*social utility*”, and the other on “*social impact*”. We then provide an overview of the French SIA landscape informed by Avise’s documentation and by KPMG’s “*social impact measurement’s bellwether*”⁸¹ (2018). In the absence of exploitable results from our quantitative survey in France, the latter provides a basis to answer some of the questions we raise in Chapter 6 when introducing the objectives of our quantitative research.

8.1.1 Two competing socio-technical arrangements

When talking about social impact, France is characterised by a dual approach to the issue that is not found in Denmark, neither to our knowledge, in most European countries where SIA is currently being discussed (with the exception of Belgium). Before the use of the phrase “*social impact*” started to spread in the country in the last decade, most WISEs in France were talking about “*social utility*”¹³³. Social utility is a notion that started appearing in the country in the 1970s, in the same context of the ending post-war economic boom and social changes that prompted the emergence of WISEs and the structuring of a broader third sector. Social utility progressively became part of policy discussions and eventually ended up in the French legislation, primarily around fiscal matters (social utility is explicitly mentioned in the 1977 and 1998 tax instructions) and later in the 2014 law on social economy.

In this context, private entities such as associations gained legitimacy in catering to the needs of the public and contributing to the general interest, an area which was historically perceived as a state monopoly (Engels, Hély, Peyrin, & Trouvé, 2006). Mathieu Hély’s phrase of “*denationalised general interest*”¹³⁴ (Hély, 2006) to qualify “*social utility*” captures this idea and provides a good illustration for the historical meaning of social utility in the context of its emergence. As per a definition of social utility, Engels et al. tell us that “*Social utility is an equivocal notion that resists all attempts at a universal definition*”¹³⁵ (2006, 14) and insist for its meaning to be collectively agreed upon, through the

¹³³ Utilité sociale

¹³⁴ Intérêt général désétatisé

¹³⁵ “*L’utilité sociale est une notion équivoque qui résiste à toute tentative de définition universelle*”

deliberation of the actors involved in it. This is something that is also reminded by Nadine Richez-Battesti: *“Social utility is more a social belief than a collective belief, in the sense that different stakeholders reach an agreement upon a cluster of values rather than a unique conception”*¹³⁶ (2006, 124).

This conception of social utility brings us very close to what we describe in Chapter 2 for the deliberation mechanisms allowing the emergence of collective preferences surrounding public value, but also in Chapter 4 about the social value that SIA models attempt to capture. However a number of French actors insist on differentiating social utility from social impact (see e.g. Hély, 2010; Stievenart & Pache, 2014; Studer, 2017). The main argument driving this distinction is the implied causality of SIA that we describe in Chapter 4, where social impact is perceived as something that is directly imputable to an organisation or to individuals, whether social utility is a broader conception of an organisation’s contribution to society, involving its networks of stakeholders (collective externalities) and therefore where the precise attribution of outcomes to specific actors is not sought after.

Although resting on similar conceptual considerations (i.e. capturing the public value created by an activity), the differentiation we articulate above translates into distinctive operationalisation procedures. For instance, impact is a neutral concept (impacts can be both negative and positive, and organisations strive to produce a positive net impact – cf. Chapter 4), whereas *“social utility”*, by the use of the notion of utility, is inherently positive. In our analysis, we therefore take this distinction as two competing STAs operating in France, where actors having a similar concern of understanding and documenting the positive effects their activities have on society have created different concepts and models to reach that goal.

Due to its historical precedence and its established legal recognition, social utility could be considered a STA closer to enactment than social impact is. In 2014, the law on social enterprises refers to social utility to define a social enterprise and not to social impact. However social impact is gaining traction as a STA (cf. Figure 23). In our case research, a concern that appeared repetitively among the promoters

¹³⁶ *“l’utilité sociale est plus une croyance sociale qu’une croyance collective partagée, dans le sens où les différents acteurs des réseaux s’entendent sur un faisceau de valeurs plus que sur une conception unique”*

of social utility was the explicit reference to social impact made in the 2018 report on work integration from the Ministry of Labour¹³⁷. The same report however mentions social utility three times more than social impact (twelve counts versus three). In our case studies, FR1 is adopting a social impact view, where “impact” is the sole terminology that was used during our interviews to qualify their “social performance”. FR2 on another hand has an approach based on social utility and its staff was trained by H el ene Duclos, an authoritative figure in France on the question of social utility. In our interviews, “social utility” and “social impact” was used interchangeably by FR2. But when asked about the difference, our interlocutor (operations manager) could articulate it spontaneously:

“Well, the impact is more at the level of the individual. It is what really happens to a person, for instance the increased competencies, or the fact that we allow our beneficiaries to socialise more, or the fact that as a part of all this, they end up with an increased self-confidence. And social utility is more... It is broader, it is what I showed you, the effect we have on all our stakeholders, there are many aspects like empowering, developing social economy, etc.”

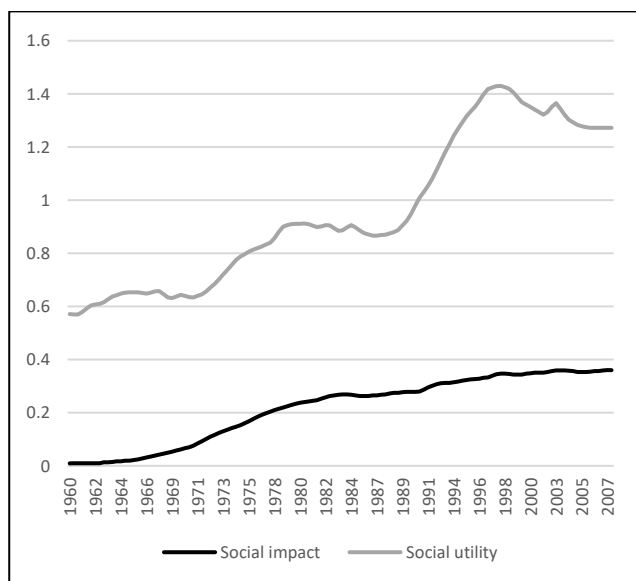


Figure 23: Ngrams of social impact and social utility

Source: Google Ngrams, unit is number of phrase occurrence per million words in books indexed by Google (search carried out in the French corpus for [utilit e sociale, impact social])¹³⁸

¹³⁷ Borello & Barfety, 2018,

¹³⁸ At the time of the research, Google Ngrams were available only up to 2008. We would expect the “social impact” curve to catch up substantially in the 2010s decade.

These two competing STAs is an ongoing issue in France and an illustration of how vivid the questions about performance measurement are for WISEs. On the premises of Callon's argument that preformation occurs when STAs are enacted, this situation can also be seen as an instance of "*performance's struggle*" (2006, 25): as social utility became a widespread STA in France (i.e. potentially reaching enactment, considerations on social impact came along, driven by the managerialist zeitgeist and an increased attention to metronomic measurements, thereby bringing a new STA to challenge the first one.

In this thesis, we acknowledge the difference in the use of social utility and social impact in France. However, in order to preserve consistency in the way we tackle the topical focus of our research (SIA) and because "*social impact*" is a term well accepted and understood in the international literature, we use SIA as a shorthand to designate any form of attempt from WISEs to measure, assess or evaluate the social externalities associated to their activities. We mark the difference between "*social utility*" and "*social impact*" only when relevant to our analysis.

8.1.2 General practices in France

As exposed in Chapter 6, we were unable to yield significant results from the quantitative survey we carried out in France, making the establishment of an overall understanding for the use of SIA by WISEs somewhat challenging. The abundant documentation provided by Avise (cf. Table 16) allowed us to bridge part of that gap, by providing an understanding of which SIA approaches were primarily promoted in France. KPMG's "*social impact measurement's bellwether*"⁸¹ (2018) provided a quantitative basis to apprehend some aspects of the practices adopted in the country. As a side note, and building upon our identification of two competing STAs in France, we highlight that KPMG's report only talks about "*social impact*" and never about "*social utility*". Furthermore, the term "*impact assessment*"¹³⁹ is used about five times more than the term "*impact measurement*"¹⁴⁰ by the report's writers, who highlight about the concurrent use of these terms, a need for "*stabilising the terminology*" (page 14).

¹³⁹ Evaluation de l'impact

¹⁴⁰ Mesure de l'impact

8.1.2.1 General observations from the field and from secondary data

As highlighted when presenting our methodology in Chapter 6, finding WISEs implementing a form of SIA (i.e. performance measurement off of their own initiative⁵⁵, excluding the reporting made to the Direccte) was rather difficult. The practice therefore appeared to us as less spread than what suggests KPMG's latest report (53% of SEs, cf. below). This was confirmed by our interactions with national WISEs networks (such as Coorace, Chantier Ecole or the Fédération des entreprises d'insertion, cf. Chapter 7), which only knew of a very limited number of members applying active methods to track their social impact.

Another aspect that emerged from our interactions with these networks is the role they play structuring and promoting approaches to SIA (whether it is social utility or social impact). In one of our case studies (FR2), the SIA model that was implemented was the result of a national initiative from Coorace. Our repeated interactions with Chantier Ecole also demonstrated the implication of the network in SIA, with a full-time researcher working on developing a framework to be used by members across the country, as well as a pilot study carried out with a specialised consulting company.

The documentation we consulted through Avise are evidencing the two competing STAs we highlight in the previous section more plainly, with reports available both on social impact and social utility. In its documentation, Avise never really spells out the difference between the two and even suggest an interchangeable use: *"There is no consensus on the vocabulary to be used in order to qualify the contributions that organisations with a social utility bring forth to their beneficiaries and to the society at large. Some will speak of social utility, social value or externalities. Here we retain the term of 'social impact'"*¹⁴¹. We also note that, supporting what is evidenced in Figure 23, contributions on social utility made available by Avise tend to be older than the ones on social impact.

¹⁴¹"Il n'existe pas de consensus sur le vocabulaire à employer pour qualifier l'apport des structures d'utilité sociale à leurs bénéficiaires et à la société dans son ensemble. Certains parleront d'utilité sociale, de valeur sociale ou d'externalités. Nous retenons ici le terme 'impact social'". Retrieved from <https://www.avise.org/articles/impact-social-de-quoi-parle-t-on> in September 2018

8.1.2.2 An overview of KPMG's social impact measurement survey

KPMG's survey focuses on social economy at large, and its sample is made of 289 respondents, split between 27 funding entities (impact investors) and 262 field actors. It is not possible to formally identify WISEs in the survey's sample, however we can see that 30% operate in the "education, training and work integration" sector. The data in the report make it possible to bring a level of answer to some of the questions our own survey in Denmark is concerned with: how widespread the use of SIA is, what are the key practices and what are the perceived benefits associated to it.

Among the main takeaways of the survey, we observe that 53% of social enterprises (i.e. excluding funders) have adopted a form of SIA in the 2018 report (compared to 41% in the previous and first edition). This number however includes the use of reporting tools imposed to WISEs by the Direccte, so the ratio of SEs implementing SIA out of their own initiative (i.e. what we ask in the Danish survey) is unknown, but has to be equal to at least 23% (i.e. when subtracting all potential WISEs from the sample). This also means that through the scope of this survey 100% of covenanted WISEs carry out a form of SIA, while we know from our field research, and particularly from our difficulty in identifying relevant cases (cf. Chapter 6), that only a very limited number of WISEs have SIA models of their own.

KPMG splits SIA approaches into three broad categories: management devices (based on operational reporting tools, adopted by 16% of SEs), research devices (based on tools such as ad'hoc surveys, experiments or cost-benefit analysis, 62%) and hybrids (a combination of the two, 23%). We reproduce the reported benefits of SIA by SEs, as communicated by KPMG, in Table 23. Although the methodology used to rank the data may skew some of the reported results¹⁴², we identify three highlights:

1. We see two items primarily pertaining to the legitimacy vertex ranking in the top three: accountability towards stakeholders (demonstrate accountability helps maintain a degree of legitimacy) and the securing of covenants and marks (both are forms of endorsement). While the concerns for covenant is likely a natural consequence of SEs using institutional reporting devices (like the Direccte's graduation ratios, considered as SIA in this survey), this overall preoccupation for legitimacy is very much aligned with what we observed in the field.

2. We notice the relative prominence of “Describe our ambition and purpose”, ranked second. The underlying pedagogical or educational intent that can be associated to this item is also observed in our field research.
3. We see that management-related items, such as the ones concerned with activity development, decision making or partnership development are ranking low (all below one). This is a finding that comes at odds with our case studies, in which both organisations are actively using their SIA for management purpose and are reporting a high level of satisfaction in this regard. We suggest that one explanation for this deviation could be found in the bias introduced by WISEs using the Direccte’s reporting instruments (considered as SIA by KPMG): our research evidences that WISEs do not consider this model as a useful management tool.

Table 23: Reported benefits of SIA in France

Reported benefits	Average rank¹⁴²
Accountability towards stakeholders	1.30
Describe our ambition and purpose	1.09
Obtain covenants or marks ¹⁴³	1.09
Differentiation of our organisation	1.01
Mobilising staff and stakeholders	1.01
Better understand the evaluated activity	1.01
Operational or strategic decision making	0.97
Develop or renew operational partnerships	0.90
Communicate	0.86
Business / activity development	0.81
Develop or renew financial partnerships	0.70

Source: KPMG, 2018

8.1.3 Practices observed in our case studies

This section introduces the performance measurement practices observed for our two French cases presented in 6.3. For each case, we focus on their SIA approach (i.e. the models they developed internally) and highlight the way through which their public stakeholders measure their performance

¹⁴² KPMG uses a three-step scale in its survey, to which scores are attached, such as Target fully reached = 2, Target partially reached = 1, Target not reached or No answer = 0. An arithmetic average determines each item’s rank. The association of the score 0 to both No answer and Target not reach appears to us as an important caveat that limits the validity of the observed scores and rankings (i.e. No answer negatively skew the average).

¹⁴³ “Label” in French, referring to institutional (e.g. SE mark as proposed in the law) and non-institutional (e.g. B-Corp) certification

(briefly, as it results in large part of the environment described in Chapter 7). We then use the performance model framework presented in Chapter 4 to summarise and provide a high-level comparison of these approaches.

We start by introducing FR1's SIA, which has a model based on the *"social impact"* STA we highlight previously in this section and makes a heavy use of output indicators. We then move on to FR2's which is based on the *"social utility"* approach. As we attempt to capture the motives and the concepts associated to WISEs' performance measurement, this section makes an abundant use of verbatim, *in italic grey*.

8.1.3.1 FR1

FR1's SIA initiative was inspired by some of the other large WISEs in France, which had annual reports displaying selected metrics like the number of FTEs employed throughout the year. FR1's first indicators (no specific period communicated) were developed around some of their practices (*"I want the same, but nicer"*), before evolving into a series of metrics that keeps developing (both in nature and in quantity). This leads to a flurry of indicators, that seems to be primarily requested by the executive management, as illustrated by the manager of training programmes (whose tasks include SIA tracking):

"[The CEO] was talking about the social data table that we have to fill out. Me, I call it the 365-cell table, one per day... Because it is what specifies, for each entity, how many permanent employees we have, how many women, how many men, how many worked hours, and the same for our workers on integration path, with how many hours in a company, on training, where they are from, what prevents them for working, etc. And [the CEO] has me adding criteria every year..."

The most detailed documents we were provided for our research shows 48 indicators tracked by FR1, grouped into seven categories (demographics, employees, full-time beneficiaries, part-time beneficiaries, graduation rates, training and education and social challenges, cf. Appendix 14). We know however that due to local policy initiatives in some of the locations where the WISE operates, more metrics are used across the organisation. All these indicators are designed quantitatively, providing counts and ratios for the different variables in use. These metrics serve primarily as an

internal management tool, as explained by the CEO: *“The goal is not always to show that there are no gaps between what we say and what we do. But it is to take the time to think every year and see how our project evolves.”*

This management motive has two components. The first one is the maximisation of FR1’s impact, through the identification and follow up of targets: *“It allows us to create goals. For instance migrants, we talk a lot about them, but we don’t do much about it”*. The second one is the internal follow-up of the performance criteria negotiated with the Direccte (the graduation rates) and often with other public stakeholders as well (e.g. at municipality level, or county level, and others). While FR1 has to use an IT system provided by the French administration to log its progress, it also has its own internal tools to follow-up on the targets and make sure they are met.

The second aspect appears very linked to the way FR1’s indicators have evolved over time, as new public policies were rolled out on integration issues, and as the different layers of governments were attaching work integration targets to these policies. The covenant manager explains: *“there was a time where our stakeholders were looking at which town [our beneficiaries] were living in. But when the Metropole was established, we had to change our territorial coverage. (...) On the same topic, we have a ‘Gypsy’ criterion (...). There was also a time where we had the ‘Romani’ criteria. Or the ‘isolated women’ criteria... So these criteria, they evolve with politics.”*

The above *verbatim* illustrates how policy initiatives can be tied to either a general topic (e.g. isolated women) or to a location (e.g. city districts), or both. This creates additional complexity in FR1’s dashboards. All these metrics are integrated in the daily management of the WISE, through different layers of the organisation (*“I look at the criteria every day, (...) and use the metrics when I go in meetings”*). Their combined use on impact matters and compliance with targets proposed by public stakeholders appear to have a strong influence on FR1’s commercial development as the WISE seeks to meet internal and external targets.

A more restricted set of indicators are also informing the long-term development strategy of FR1 and are used by the WISE’s board of directors, its investors, as well as for general-purpose communication

through a synthesized “social report”. Here we identify a second motive behind the use of SIA which is more guided towards communicating and **demonstrating** FR1’s impact, through a restricted sample of indicators. One metric that is particularly used by the management team is the number of employees per FTE, as stressed by the CEO: *“if I say that I provided 1,000 hours of training, all the while I trained 1,000 people, it means I trained each person for one hour, my impact is 0. If I do 1,000 hours for one person, it’s better.”*

Also framed around demonstration motives, and separated from management concerns, we identify a strong interest for cost-benefit approaches, as evidenced by the CEO: *“One euro invested in our companies returns three or four euros to the state coffers. Every year. We know that it is a profitable investment even before the end of the first year. Heck, there is a point where this is simply the best investment the state can make.”* FR1’s executive management has made its own, back-of-the-envelope calculations in this regard, inspired by a practice from the Fédération des Entreprises d’Insertion (the national WISE network it belongs to) but it does not use it in its written communication (such as their annual report or their social report). During our interviews, the motives associated to this kind of initiative were clearly associated to the **tightening of public budgets**.

Lastly, during our interviews with the executive management, we observed an acute awareness about the importance of outcomes. However FR1’s SIA in its current state fails to address this issue. The CEO said it was working on initiatives to address this gap: *“So there are two series of indicators that we should create: one about the secure access of a path towards sustained employment, and one about what happens once they [the beneficiaries] get out”*. The result is a SIA which, in its current form, appears very geared towards outputs.

Based on the information gathered, Table 24 structures our understanding of FR1’s SIA using the framework we introduce in Chapter 4. Using the same framework, the table also shows how FR1’s public stakeholders (PS) typically try to assess its performance through the mechanisms we describe in Chapter 7. The WISE’s approach appears very focused on monitoring since SIA is primarily used as a management instrument. Our interaction with both the executive management and the staff carrying out the tracking of the indicators reveals that indicators are created in a very top-down fashion

(planned design), often driven by policy issues. Lastly, the SIA model appears focused on quantitative indicators. Quantitative metrics are also used abundantly to traduce qualitative aspects. For instance, the CEO talks about indicators as followed:

“Another qualitative indicator is that, when we realized that we had 15% of illiterates, we developed training offers that we call ‘basic foundations’, (...) and we track these offers. We provided 58,000 hours of training last year and 103,000 this year. So people become increasingly competent, more qualified, and so that is a qualitative indicator. (...) As another qualitative indicator, (...) we also keep track of the number of work injury resulting in leave of absences.”

Table 24: FR1's performance measurement and SIA

Attribute	Variable	FR1	Public stakeholders
Purpose	Screening	0	2
	Monitoring	2	1
	Reporting	1	2
	Evaluation	1	1
Analytical lens	Economy	0	0
	Efficiency	2	2
	Effectiveness	0	0
	Cost-benefit	1	0
	Monetarisisation	0	0
Data typology	Quantitative	2	2
	Qualitative	1	0
Design	Planned	2	2
	Negotiated	1	1

0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus¹⁴⁴

Source: Author

8.1.3.2 FR2

FR2's approach to SIA was fostered by Coorace, the national network for Intermediary Associations (AI), one of the covenants the WISE benefits from. Coorace has had an extensive collaboration since 2012 with H el ene Duclos (one of the supporters of the “social utility” approach in France, and a major

¹⁴⁴ The rating is the product of our own judgement based on (i) elements of our interview, (ii) the performance measurement model itself and (iii) the communication material developed about the model's information. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

contributor to its operationalisation) in order to apply social utility assessment to its members. It is in the context of this cooperation that FR2 came to develop its approach to SIA¹⁴⁵.

The WISE developed its model over the course of several months, through seven workshops hosted by Coorace in Paris with several other members. It then refined its model through the involvement of its stakeholders, with multiple meetings where public stakeholders, beneficiaries and FR2's staff all provided input for how impact should be gauged.

The primary use of FR2's SIA that was evidenced during our interviews was communication, with a clear connotation on stakeholder education. The operations manager tells us: "This is really the communication medium we use on a day-to-day basis to explain what we do to our stakeholders. Because this is really complicated. Whether we talk to a consumer, a company, an association or a public stakeholder, we won't use the same arguments." Interestingly, this pedagogical intent clearly applies to public stakeholders: *"What is difficult is... how do we explain them what we really do in our organisations? For example in three years we are getting our fourth interlocutor with the Direccte. And every year during the management dialogue, we have to start from scratch and explain them what we do."*

The pedagogical undertone that we find in FR2's approach to communication through SIA provides an insight behind the outreach intent that is not necessarily explicit with e.g. FR1, which is primarily concerned with proving they make a difference (SIA as a legitimacy device). However we do find this aspect of SIA with FR2 as well, and particularly towards public stakeholders. The operations manager pursues: *"But actually I think that the approach of social utility aims at convincing. So one can argue that this is communication, but is important to be able to show, to demonstrate."*

FR2 also associates the work carried out on SIA to a source of improvement in the way it communicates: *"This work on social utility, it led us to distance ourselves from the day to day context and to wonder about what it is that really interest people"* FR2 management reports they now communicate more

¹⁴⁵ We remind what we signalled in Chapter 7: we acknowledge the difference in the use of social utility and social impact in France. We use this differentiation ourselves to foster our analysis of socio-technical arrangements. However, we use SIA as a shorthand to designate any form of attempt from WISEs to measure, assess or evaluate the social externalities associated to their activities.

often (they launched a newsletter), in a more compelling fashion for their audience (they use testimonies, case stories, etc.) and leverage these improvements internally as well as externally.

FR2 also demonstrates an interest in SIA tied to management aspects. This appears strongly entangled with the quality management processes started a decade before SIA (to the point where a confusion between the two were observed during our interview with the operations manager). Incidentally, we observe that a large part of the indicators used in the SIA model as management tools (particularly the ones monitoring outputs) are coming from the quality management system *“For me it truly comes as a complement to our quality insurance processes”*.

Furthermore, we identify a clear differentiation between outputs and outcomes. And despite the characteristics of the social utility approach typically less concerned with indicators, we identify a noticeable effort in attempting to identify metrics to capture FR2’s impact. The operations manager tells us: *“[We have] output indicators. For instance how many people we welcomed, and especially how many we managed to put to work. And then impact, it’s complicated. What is difficult to measure, it’s when we send people on a training, and then we see them exchanging phone numbers, hugging... All this is an impact (...) but how can we manage to evaluate that? And so here we are on things that remain to be created.”*

Lastly, we observe a use of cost-benefit analysis, with a simplified model similar to the one observed at FR1. It was developed separately from FR2’s social utility approach but is now integrated to it, as a contribution of the WISE to local economic development. FR2’s CEO indicated using this measure with their public stakeholder:

“CEO: It allowed us to compute the generated wealth for one euro of public funding. (...)

Researcher: And do you use these ratios during the management dialogue?

CEO: Ho yes!”

FR2’ SIA model is constantly evolving. At the time of the research, it was structured through 19 criteria, spanning 30 indicators, grouped in five main themes (cf. Appendix 15): empowerment, fighting exclusion, provision of transitional employment, generation of local wealth and enacting the social

economy. Indicators are a mix of qualitative (e.g. testimonies, satisfaction surveys) and quantitative measures (e.g. ratio of beneficiaries following through initial meetings). Although quantitative indicators are a majority, they are operationalised to build a qualitative, narrative arc in FR2's communication material (similar to FR1, but more developed).

Based on the information gathered, Table 25 structures our understanding of FR2's SIA using the framework we introduce in Chapter 4. Using the same framework, the table also shows how FR2's public stakeholders (PS) typically try to assess its performance through the mechanisms we describe in Chapter 7. With its emphasis on communication, FR2's SIA is rather focused on evaluation (i.e. to provide a picture of what has been done to stakeholders). We perceive the model as less oriented towards management since most indicators monitored are coming from the quality assurance processes, that are anterior to SIA. Lastly, we highlight the highly negotiated design, with a deep stakeholder involvement.

Table 25: FR2's performance measurement and SIA

Attribute	Variable	FR2	Public stakeholders
Purpose	Screening	0	2
	Monitoring	1	1
	Reporting	0	2
	Evaluation	2	1
Analytical lens	Economy	0	0
	Efficiency	1	2
	Effectiveness	1	0
	Cost-benefit	1	0
	Monetarisaton	0	0
Data typology	Quantitative	2	2
	Qualitative	2	0
Design	Planned	0	2
	Negotiated	2	1

0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus¹⁴⁶
 Source: Author

¹⁴⁶ The rating is the product of our own judgement based on (i) elements of our interview, (ii) the performance measurement model itself and (iii) the communication material developed about the model's information. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

8.2 Effects of Neo-Weberian State mechanisms on strategic interactions

In this section, we observe how the NWS environment influences the **strategic interactions** between WISEs and their public stakeholders. We do so while focusing on interaction related to performance measurement, in a context where both WISEs and their public actors agree that WISEs' performance should be gauged in terms of public value creation. We observe that the executed strategies display a heavy emphasis on the generation of legitimacy, which is connected to the WISE's operational vertex (through its activities and its organisational design) but not directly to the public value it contributes to co-create (i.e. not to the outcomes of its activities).

We explore this situation by way of three NWS mechanisms. The first is **legalism**, which through covenants (a form of institutional endorsement) and their imposed reporting (a legitimacy device) on graduation rates (an efficiency ratio of output to inputs) creates a framework where legitimacy is primarily acquired through operational **compliance** rather than actual public value creation. The second is **bureaucracy**, which through the implementation of outputs-oriented indicators within different layers of local governance favours performativity under the form of **retroaction**.

The third NWS mechanism we explore is **centralisation** and how it creates layered STAs: one centralised approach to performance measurement through graduation rates, and other local STAs that can emerge through the **deliberation** between WISEs and their local public stakeholders. SIA can thereby be considered a form of emerging, unrealised STA that may attempt to **circumvent** the central one. We attribute this paradoxical situation to the "*neo*" tenet of NWS and shows that it allows to avoid another form of performativity: **indicator-substitution**.

8.2.1 Legalist legitimacy and bureaucratic performance

In Chapter 7, we showed how the French NWS environment creates a rigid framework for the provision of work integration services and the co-creation of public value. In the first sub-section below, we show how this translates into a focus on **legitimacy**, which must be secured through **compliance** to regulation (part of which is performance measurement) rather than actual public value creation. The second sub-section shows how NWS-induced bureaucracy can be a vector for **retroaction**.

8.2.1.1 Legalism puts an emphasis on strategic interactions in the legitimacy vertex

The regulatory framework that we describe in Chapter 7 for work integration activities is a good illustration of NWS mechanisms: a “Neo” recourse to private actors in order to co-create public value (work integration services) with an openness to foster deliberation (the yearly management dialogue) in a “Weberian” framework of rigid regulation (the covenants and the different rules governing them in terms of access to referring channels and subsidies) administrated through stringent control instruments (graduation rates).

As a result, in order to have access to the resources required to operate and generate public value (access to beneficiaries and public funding), WISEs have to generate legitimacy primarily through the mechanisms emerging from the NWS’ legalism: the institutional endorsement of covenants, and the use of legitimacy devices (the tracking of graduation rates) in order to maintain their covenants (and therefore their operations). Legitimacy can also be acquired less formally through deliberation during the yearly management dialogues, where WISEs get a chance to argue for their case and justify potential deviances from the graduation rate targets.

This heavy emphasis on **legalist legitimacy** tends to minimise the two other vertices of the PVT triangle (operationalisation and public value), where both activities and actual public value creation become secondary issues in the face of **endorsement** and **legitimacy devices**. FR1’s CEO tells us: *“It turns entrepreneurs into managers, and they are less and less entrepreneurial. They go ‘me, I respect all the rules, I have a covenant, I have provided the welcome brochure, I provide vocational training, I develop soft skills, I transfer know-hows, etc.¹⁴⁷ But we don’t care about this! What’s the result?”*

WISEs can therefore appear as following a strategy of **compliance** with the requirement imposed by the burden of legalism, where parts of their operationalisation vertex (resources and activities) becomes utilized solely towards the generation of legitimacy, to the potential detriment of public value creation. As FR1’s CEO puts it: *“We spend more time justifying how we did a good job as a work*

¹⁴⁷ All these terms are key words of the French legislation on work integration

integration provider than we spend time face to face with people, to tell them – ‘Wait, you’re in trouble, you want to work, here is what we’re going to do about it’”.

As covenants are the key for both funding and access to referring channels (at least for ACIs, EIs and most ETTIs), these observations are true also for WISEs like FR1 or FR2 which both show a low reliance on public subsidies but still need referring channels to access beneficiaries. This reminds us of Gianfaldoni’s suggestion that WISEs have a **subordination** relation to public stakeholders (2018), where we can see the observed compliance as an expression of this quasi-hierarchical relation.

Furthermore, we observe through our complementary research (i.e. outside of our case studies) that the legalist constraints created by covenants on WISEs’ operationalisation vertex can condition the use of SIA models developed or envisioned by WISEs. We observed this with ACIs, which covenant caps their commercial revenue to 30%. On this premise, in the Chantier Ecole network SIA is considered as a strategic legitimacy device to prove the creation of public value and thereby secure the necessary legitimacy warranting the 70% of public funding required to operate the organisation. During a meeting at Chantier Ecole, a regional delegate said about SIA: *“I think we are talking about our survival”*.

A last aspect of how Weberian legalism influences the articulation of the legitimacy and operationalisation vertices is through the use of **organisational design** strategies. We observed this in both our cases studies, where the management has created multiple legal structures (cf. Figure 17 and Figure 18) in order to make sure it is able to secure the right opportunities for its beneficiaries while respecting the law. In FR2’s case, it allowed the WISE to operate in more scalable markets. In FR1’s case, it allows to hire beneficiaries once their sponsorship agreement expires and provide a balanced source of revenue across unequally profitable covenants.

8.2.1.2 Bureaucratic performance and its retroactions

Throughout our interactions with field actors during our case studies, we confirmed the linkage we establish in Part 1 between performance measurement and SIA, where SIA is perceived as a way to both drive performance (when using it as a **management tool** to improve impact) and communicate about performance (when using it as a **legitimacy device**). We also observed that performance (or

“social performance”, used by both FR2 and FR1) is conceived as public value creation by both WISEs and the Directe along the lines of what we describe in Chapter 2, in terms of remedies to market failures (*“We manage to put to work some unemployable dudes!”*)

On the public sector’s side, the way this translates into applicable performance measurement is mainly through a device tracking graduation rates (cf. Chapter 7). In our cases studies, we find that the intricacies of the NWS bureaucracy are particularly conducive to the complexification of this *“simple”* device based on three main indicators (cf. Table 20). As the management dialogues where WISEs’ yearly targets are defined in terms of graduation rates include a wide of array of public stakeholders, we observed that the different layers of local government impose indicators tied to their jurisdiction, often in relation to a policy area they wish to develop. For instance FR1 has targets in terms of beneficiaries coming from disadvantaged *“priority districts”*¹⁴⁸ in some urban areas, while it also has to contend with a certain number of beneficiaries benefiting from social benefits (requested by the county), as well as other ad’hoc targets (Romani, isolated women, disabled people, etc.).

This setup appears particularly prone to retroactive behaviours where WISEs will mobilise their resources to achieve the targets specified by their public stakeholders. (FR1’s covenant manager tells us: *“we also tackle this internally, going – ‘you’re missing disabled workers’, or ‘we don’t have enough women’, etc.”*) rather than pursuing other opportunities potentially more favourable in terms of public value creation. A form of retroaction is also observed in the way FR1 develops its SIA, adding indicators in large part driven by the targets requested by public stakeholders. These indicators end up embedded in the WISE’s management dashboards and suggest that performance indicators implemented by public stakeholders have a normative influence.

Retroaction also has an effect on how WISEs channel their beneficiaries, as public stakeholders strive to reach their own objectives. We observed that NWS bureaucracy tends to produce administrations working in silos, following their own performance indicators often expressed, in the work integration space, in terms of *exits* (of beneficiaries, from specific support schemes). While WISEs generally benefit in being considered as *“exits”* by public stakeholders as this is how they take beneficiaries in (e.g. this

¹⁴⁸ Quartiers prioritaires

is the case for job centres: unemployed individuals are considered *exiting* job centres' support while on integration scheme), they are not considered as "exits" by all the stakeholders involved in work integration. Some public stakeholders may thereby actively avoid to channel beneficiaries towards WISEs.

Furthermore, the formal performance measurement induced by NWS bureaucratic mechanisms (i.e. through delimited quantitative indicators) tend to restrict the evaluation of performance to outputs only since they are much easier to operationalise as metrics than outcomes (FR1's CEO: *Here again, we are following an administrative logic: we are paid to generate 'positive graduations', but not to follow up after three months, six months, one year, two years, etc.*"). Outputs thereby tend to become a proxy for public value, which is at odds with what we describe in Part 1, but also with the views of all the stakeholders we have interviewed (including from the public sector), who agree outcomes are the real added value of WISEs.

This implementation of operationalizable indicators favoured by the NWS bureaucracy leads us towards our next section where we explore the role of centralisation in framing the NWS approach to performance measurement.

8.2.2 Centralisation and the layers of socio-technical arrangements

Chapter 7 evidences the role endorsed by France's central authority (the DGEFP) in defining how performance should be measured for WISEs, through the use of the graduation rates as a device. We see the importance of these indicators confirmed throughout our case studies, as WISEs and local Direccte administrations are using them to evaluate covenanted WISEs, with the concrete implications we describe above on all three vertices of the PVT triangle. To this end, we associate the monitoring system developed by the DGEFP as a form of STA rooted in the central state.

Whether they are accepted as a relevant indicator (seldom) or not (often), graduation rates have become the structuring metric of the work integration sector in France. We see in our research that WISEs are actively pursuing them as an objective (even though they are fully aware of their limitation as a performance metric) and tend to embed them in the daily management of their organisation. To

this extent, graduation rates display the features of entangled statements and reality that Callon attributes to enacted STAs (2006). The forms of retroaction we highlight in the previous section would then be a symptom of this (at least partial) enactment.

Interestingly, we also observe in our data the formation of other STAs in France, which are fostered by the strategic interactions between WISEs and their local stakeholders. Interactions with the local Direccte are particularly compelling evidence of this phenomenon, as we observe a shared view that graduation rates are far from being an optimal performance measurement mechanism. The Direccte says: *“The performance target, from a state perspective, the indicator for work integration it is the positive graduation rate. So the graduation rate remains [important]... Indeed we are part of a labour policy (...). But it is true that because of this, we do not valorise all the actions that are carried out to induce progress towards beneficiaries’ integration.”*

In parallel, we also observe how the management dialogue, which is also a form of legitimation mechanism anchored in NWS’ legalism, appears as a particularly conducive forum to exchange on these issues. (*“Direccte: “But it is only during the management dialogue or during general assemblies that we can realise the work that is done, however that is not quantifiable with numbers”*). It fosters a conversation between all local public stakeholders involved in a WISE’s activity (i.e. not only the Direccte) and contribute to deliberation. As we took part in a meeting involving these stakeholders (albeit not a formal management dialogue meeting), we observed how collective preferences are deliberated and agreed upon.

By doing this, WISEs and their stakeholders provide deliberated co-definitions of public value (e.g. a covenant manager at the Direccte: *“I consider that if the beneficiary is being catered to, then the job is done, even if the quantitative or qualitative targets are not met”*) which translate into emerging STAs, competing with the established, centralised performance measurement approach. Management dialogues are therefore a central place where WISEs having a SIA model can present their results, proposing an alternative STA in an attempt to **circumvent** the centralised one. This is further illustrated by the following quote, from FR2,’s CEO in a meeting with Direccte and the City Council:

“I am glad to hear you say that (about the inadequacy of the graduation rates as a performance indicator)! I realise that I am not the only one to be frustrated about these questions (...). And I think it is maybe through some work on the modulation¹⁴⁹ that we could reshape these indicators because it is a concern that we all share. For instance, the RSA¹⁵⁰ is not a good criterion (...). So I personally don’t know how much liberty a region can have to undertake a work of this nature...”

The last sentence of this quote takes us to the last point we want to emphasize about centralisation, which is its relative inability to enact those STAs emerging from the base. The Direccte’s long-winded reply to FR2’s question summarises in its regulatory incapacity to formally implement alternative performance measurement approaches. In essence, NWS displays a form of centralization where a degree of deliberation can exist at a local level (the “neo” element of NWS) and can translate into the identification of shared preferences and co-defined public value. It however prevents these mechanisms to carry any form of tangible weight in the centralised (Weberian) legitimacy vertex.

The downsides of this situation are (i) the obvious shortfall of democracy in the way the NWS measures the performance of its WISEs, and (ii) the occurrence of retroactions, as we explain them through bureaucratic mechanics above. We however also find one positive externality to the observed PAM dynamics: the relative absence of indicator-substitution issues between WISEs and their public stakeholders. Even though all parties seek to meet their targets, expressed in outputs-oriented metrics, most of them, if not all at the local level demonstrate an acute awareness that outcomes are what matters most and that WISES’ impact extends well beyond what is captured by the central framework for performance measurement.

¹⁴⁹ The term “modulation” refers to the variable part of the subsidies (maximum 10%) that is allocated to WISEs based on the graduation rate targets they are meant to achieve (cf. Table 20)

¹⁵⁰ A social benefit whose beneficiaries are often part of defined targets for WISEs

8.3 Between autocracy, managerialism and deliberation: how performance measurement evolves under the Neo-Weberian State

In this last section, we summarise the findings we highlighted so far and use them to characterise the French NWS performance regime. We articulate how the tensions between the “*neo*” and “*Weberian*” tenets of the NWS translate into a **hybrid** approach to performance, stuck between autocracy, managerialism and offering signs of deliberation (Jany-Catrice, 2012).

We develop about how in this context, local, unrealised **socio-technical arrangements** can emerge and how they are contained by centralisation. And we show how this mechanism contributes to the avoidance of **indicator-substitution** in a centralised configuration that would otherwise be very prone to it. We then close this section with the conclusion of Chapter 8.

8.3.1.1 The *Weberian* components of NWS, anchoring autocratic practices

Our highlight of the influence of bureaucracy, centralisation and legalism are all illustrations of the pregnancy of rigid state structures in the NWS paradigm. We showed in Chapter 2 that these are attributes mostly associated to the Weberian side of the NWS. In our particular case, these aspects of the NWS are well anchored in the French Work integration environment and are tied to most mechanisms through which WISEs can generate the necessary amount of legitimacy in order to secure the means to operate. This includes the way WISEs’ performance is framed by the central state, which we find partly rooted in an **autocratic regime** (Jany-Catrice, 2012).

According to Jany-Catrice, an autocratic performance regime is planned and unidimensional in its conception of performance (cf. Figure 6). Arguably, it can be difficult to fit the performance measurement practices observed by public actors (even centralised) entirely into the “*autocratic box*” since we for instance discern three different criteria associated to the modulation. However, we observed in the field that when referring to WISEs performance, all field actors (i.e. WISEs’ managers and staff, as well as their public stakeholders) spontaneously mention the graduation rates. To this extent, our interview coding shows that whenever performance is evoked in relation to the Direccte,

“graduation rate” (whether positive or dynamic) comes systematically as the first (and often sole) indicator quoted (100% of occurrences¹⁵¹).

While the modulation indeed includes three criteria (among which the graduation rate carries the lowest weight, cf. Table 20), the main metric upon which WISEs are gauged since 2008 is the graduation rate. At the time, the legislation introduced for the first time a formal performance measurement instrument for WISEs – which is used as a legitimacy device. This could explain why it so often appears in the field portrayed wearing autocracy’s mask: as a planned indicator, imposed by the central government which, through its unidimensionality, fails to capture the breadth and the impact of WISEs activities.

These are considerations that indeed are based on the perceptions of the interviewed field actors and not actually anchored in the reality of the legislation as it currently exists, which actually paints a more complex picture. However these interpretations, observed among all interviewed field actors, do have an influence on the identified strategic interactions which, about state-sponsored performance measurement, will consist mostly in **complying** with devices that have an impact on WISEs’ operational vertex but that they do not associate to the creation of public value, as we highlight above.

8.3.1.2 The *Neo* component of NWS: stuck between managerialism and deliberation

While we show in Chapter 7 that the sector of work integration in France has been in constant flux, we can define 2008 as the year where the “Neo” tenet of the NWS made its mark on WISEs, through the DGEFP Circular of December 10, introducing a formal system of performance measurement (under the form of graduation rate targets) tied to the covenant system. Another important reform was the one voted in 2014, introducing the modulation through the use of several criteria (i.e. not only the graduation rate) to account for more “*qualitative*” aspects of the work done by WISEs (i.e. the integration effort and the entry situation, cf. Table 20).

¹⁵¹ 15 observations

This attempt at capturing a multidimensional aspect of performance would then place us in a managerial regime according to Jany-Catrice's framework. To corroborate this observation, we observe that (i) Gianfaldoni & Rostaing (2010) qualify the recent legislative changes we highlight as a set of regulations aimed at operationalising, within the field of work integration, the NPM toolbox fostered by the 2001 Act on Budget Legislation (LOLF)⁸ and that (ii) Jany-Catrice cites the LOLF as exemplifying the application of a managerialist performance regime. In our particular case, we subscribe to this idea to the extent that the proposed indicators were indeed planned by the central legislator. This trend towards managerialism is also clearly identified as a "neo" attribute by Pollitt and Bouckaert (2017) and presented in our Chapter 2.

Based on our field observations, we however make the argument that the local actors of this managerialist performance regime are creating their own deliberative spaces. It is the mechanism we describe in the previous section (especially the "*management dialogues*"), through which stakeholders get a chance to identify shared preferences in an attempt to co-define public value and foster their own STAs. This process happens to benefit particularly well the second component of the NWS' "neo" attribute we stressed in our Chapter 2: an extension of democratic mechanisms where the state can facilitate the emergence of collective preferences.

It is nevertheless quite clear that the performance regime to which French WISEs are subject cannot be labelled as a deliberation regime either. It is, after all, in the nature of the NWS to retain an amount of centralised, top-down power. Furthermore, Pollitt and Bouckaert (2017) make the remark that these particular features of extended democracy do not foster well in Napoleonic countries like France. In our case, they however appear sufficient to limit some aspects of performativity, like **indicator-substitution** (but not retroaction).

Conclusion of Chapter 8

As we introduced the performance measurement models observed in our case studies, this chapter characterised the strategic interactions that take place between WISEs and their public stakeholders in the French NWS environment. We summarise these findings in Table 26 using our nine PVT entries. We

see that in NWS’s rigid legalist framework, strategic interactions are focused on the generation and the conservation of **legitimacy**. We also see that performance measurement, under the form of devices (the graduation rates monitoring system), is an important part of maintaining that legitimacy. This causes WISEs to develop their own SIA models to capture what they consider their “*real*” performance.

Having a clearer picture of the strategic interactions and of how the question of performance is dealt with in these interaction, we can now start to unearth **strategic intents** associated to performance measurement. We for instance see that French WISEs **comply** with a performance measurement system that is imposed upon them by the NWS mechanisms and in which they do not believe; in a way that is reminiscent of the relation of subordination that is described by Gianfaldoni (2018). At the same time, they also **circumvent** that very system both for management and advocacy purposes. What do these different intents mean for the measurement of performance? Do they have an influence in terms of performativity? And how does that compare to Denmark? These questions will be answered in Chapter 10 as we consolidate the strategic intents on the basis of our findings in Denmark, which we expose in the following chapter.

Table 26: Summary of strategic interactions in the French NWS

PVT entry	Summary
Public value	
(Co)-definition	The co-definition of public value is marginal. It can happen primarily through the management dialogue, but never formally.
(Co)-creation	Work integration’s institutional setup allows for a true co-creation of public value, where WISEs are instrumental in the delivery of employment services.
Acknowledgement	Public value must be acknowledged by public actors in order for WISEs to be able to maintain their covenant. This is however done primarily through output-focused devices and deliberation (i.e. two legitimacy instruments).
Legitimacy	
Deliberation	There is a degree of deliberation at the local level with the deconcentrated power (Direccte). It allows to negotiate some aspects of the graduation rates and permits to maintain a level of legitimacy even if the measured performance is deemed unsatisfactory by the Direccte.

PVT entry	Summary
Legitimacy devices	The Direccte's performance measurement model (the graduation rates) is a form of legitimacy device: WISEs have to prove that they create a form value - regardless of the actual value captured by the indicators. These devices are central elements of the French covenant system, and WISEs' SIA is not considered a viable alternative by their stakeholders. We noted that when associated to devices bureaucracy tend to generate retroaction (i.e. WISEs try to achieve specific targets to please specific stakeholders)
Endorsement	Institutional endorsement, under the form of covenants, is instrumental for French WISEs to operate. Securing the renewal of the covenant becomes a leading strategic objective.
Operationalisation	
Agency	Most forms of agency are curbed by legalist and bureaucratic mechanics. There is however a degree of competition in certain areas for accessing covenants (they are in limited supply for a given territory).
Organisational design	Organisational design can play an important role since covenants may limit the type of legal entity or revenue WISEs can have access to.
Activities	Some covenants may also limit the type of activities WISEs may engage in. Furthermore, WISEs can have an advantage on some public contracts displaying " <i>social provisions</i> " mandating the use of covenanted resources.

Source: Author

Chapter 9 Performance measurement in Denmark: deliberation or chaos?

Having explained the mechanisms through which public value is co-created in the Danish **New Public Governance** (NPG, Osborne, 2006) in Chapter 7, here we explore the concrete aspects of the cooperation between Danish work integration social enterprises (WISEs) and their public stakeholders. Doing so, we expose the different **performance measurement models** we observed (including WISEs' social impact assessment -SIA- models) and we characterise the **observed strategic interactions** using public value theory (Moore, 1995).

As we develop this chapter, we answer the following questions: What are the strategies deployed by WISEs to contend with the Danish relationship-based quasi-market for active labour policies? How does that translate in performance measurement practices? We also explore how the void created by the absence of a structured way to measure WISEs' performance among Danish municipalities drives WISEs to consider social impact assessment (SIA) as a competitive edge. Is it a source of performativity? Or does it prevent it? Does it have an effect on deliberation? Like in Chapter 8, these interrogations address the core of our topic, as they allow us to fathom the strategic intents that we will outline in Chapter 10 to answer our research questions.

We start by providing an overview of SIA practices in Denmark. We go through practices identified at the national level (very few) and then reconstitute the key information delivered by the quantitative survey we carried out in the country. We then move on to provide a detailed view of how the performance of our case studies is measured where we focus on their own approach of SIA. In this last section, we identify the key aspects surrounding the motives associated to the development of a SIA model and we highlight some of the hindrances through DK2's case.

In the following section, we use the PVT strategic triangle to map out the **strategic interactions** between WISEs and their public stakeholders, as they surface through our data. As we observe strategies demonstrating a relative **equilibrium** across the three vertices of the triangle, we highlight how the creation of public value is a marker of **performance** in the Danish relationship-based quasi-market of work integration services. We also show that underlying market mechanisms create pressures on WISEs' operational vertex, where SIA can become a vector of **competitive** advantage.

In a dedicated sub-section, we show how what is realised in the public value vertex through the means of the operational vertex can be leveraged in the **legitimacy** vertex in order to keep commercial relations on a good standing with public actors. We demonstrate the important role of **deliberation** in this configuration and tie our observations to some of the NPG's characteristics we highlighted in Chapter 2 such as **reciprocal patterns of communication and exchange** (Powel, 2003) and the role of **relational capital and relational contracts** (Osborne, 2006).

We conclude this chapter by summarising the main takeaways of our analysis and using them to characterise the way performance measurement is evolving under the Danish NPG. We equate the abundance of **deliberative mechanisms** and the complete absence of centralised procedures to what can sometimes be perceived as chaotic processes, and we highlight the benefits of deliberation in terms of low performativity.

9.1 Observed practices

How is the performance of WISEs measured in Denmark? In this first section we introduce the practices observed in Denmark. This initial approach is informed mostly by stakeholders having a national visibility on the subject, namely the former National Centre for Social Enterprises and the impact investor Den Sociale Kapitalfond. We then provide an overview of the Danish SIA landscape by presenting the highlights of our quantitative survey.

We continue with the detailed presentation of the performance measurement considerations surrounding our case studies, which translates into the introduction of the SIA model used by DK1 and the performance measurement framework used by DK2PS. Seen as DK2 does not carry out formal SIA activities, we also propose a hypothetical SIA model for the WISE, based on the information collected during our field research. As we attempt to capture the motives and the concepts associated to SIA for these two WISEs, we use selected verbatims *in italic grey* which best represent our coding work.

9.1.1 General practices in Denmark

9.1.1.1 Observed practices in the Danish environment

In Denmark, questions related to the performance of social enterprises started to gain traction in the early 2010s. A popular approach at the time was SROI, which is a SIA model that was initially embraced by the National Centre for Social Enterprises (the government institution that was tasked with the development of social economy in the country, before closing down) and Den Sociale Kapitalfond (Denmark's sole impact investor). This initial interest in monetarisation-oriented approaches quickly evolved, with the National Centre proposing workshops on more qualitative methods and Den Sociale Kapitalfond developing initiatives around “*classical*” cost-benefit analysis (i.e. excluding the monetarisation of outcomes seen in SROI).

While the National Centre is now closed, Den Sociale Kapitalfond developed an online platform¹⁵² in 2017 where WISEs¹⁵³ can assess the “*social value*” they create through predefined indicators in terms

¹⁵² <http://socialletal.dk/>

¹⁵³ Interestingly, Den Sociale Kapitalfond actually targets “*social enterprises*” in general, but the indicators proposed are only tied to work integration – another instance of the confusion between work integration and social economy observed in Denmark

of avoided public spending and additional taxes garnered, depending on their target group. Apart from these “*high profile*” initiatives, few are particularly noticeable for providing SIA guidelines in the Danish environment. As evidenced in the following sections of this chapter, the public sector does not propose a harmonised approach to gauge the performance of WISEs, although isolated initiatives exist. Similarly, only a limited number of Danish WISEs have implemented a form of SIA. We explore some of their initiatives through the quantitative survey we carried out in Denmark in the following sub-section.

9.1.1.2 Key takeaways from our survey

In this section, we highlight some of the key findings from the survey we carried out in Denmark on SIA practices among SEs. While introducing our methodology in Chapter 6, we emphasised that although our sample’s relative size is acceptable when considering Denmark’s SE population size (10% of the SE population and 20% of the WISE population is represented), the absolute sample size is still relatively small (31 SEs, among which 26 were identified as WISEs, cf. Table 12), which makes the relevance of any advanced statistical analysis limited. The findings we highlight below, which are mostly descriptive, can however provide us with a first level of understanding as per actual SIA practices among WISEs in Denmark. We articulate their presentation around the four objectives we formulated in Chapter 6 for our quantitative survey.

1. How widespread is the use of SIA among WISEs and what are the key practices identified:

We find that six (21%) of the WISEs which took part in our survey have undertaken a form of SIA. The ratio remains unchanged when considering the overall SE population of our sample. Among the ones which do not use SIA, ten (50%) do so because it is not something they have been prioritising in their activity. While it can be a reason on its own, it is mostly associated to a lack of human resources (50% of cases). 17 (85%) do not know about methods for SIA. For the WISEs that indeed carry out SIA, four out of six have developed their own model. The two that used existing models adopted SROI and *virkningsevaluering*, a model developed by a network of Danish stakeholders coordinated by Aalborg University for which little documentation was found available. In their SIA, four out of six WISEs used a form of monetarisation of their impact (observed on two who developed their own approach, and the two which used “*off-the-shelf*” models, cf. Table 27).

Table 27: Selected features of SIA models

Observation	SIA model	Outputs	Self-reported outcomes	Monetarisatation	Case studies
8	Own	Yes	No	No	Yes
45	Own	Yes	No	Yes	Yes
51	SROI	Yes	Yes	Yes	No
56	Own	Yes	Yes	Yes	No
60	Own	Yes	Yes	No	No
42	virkningsevaluering	Yes	No	Yes	Yes

Source: Author

Most of these findings are consistent with what we observed in our case studies, where a combined lack of resources and knowledge explain why DK2 has not carried out any SIA despite entertaining a strong interest in the topic. DK1 developed their own SIA model, which was a time consuming and expensive process, spanning two years, and where they first attempted to carry out monetarisatation, before abandoning the idea.

2. Do public stakeholders have an influence in the development and application of SIA models:

Public customers have been a factor in how the approach to SIA was designed in five cases out of six: three where they provided some form of advice on the methodology or the chosen indicators, and two where they were identified as a target audience for SIA (which had an influence in the model’s design). Interestingly, the level of influence is not correlated to the amount of revenue garnered through public sources: the one case where public stakeholders have no influence has between 50% and 75% of revenue from public customers, which is more than some of the WISEs having worked with public actors for their SIA. These findings are corroborated by our case studies: in DK1’s case, public customers have been instrumental in the development and operationalisation of their SIA model. For DK2, although the WISE does not use a SIA model, we observe that most of its considerations towards SIA are influenced by its interactions with public stakeholders.

Our survey tells us that among the six WISEs that had carried out a form of SIA, four had taken inputs from their employees for the design of the SIA model and its indicators, whereby beneficiaries provided inputs in three cases (50%). This last aspect comes as a surprise since stakeholders’ involvement (and

particularly beneficiaries’) is advocated by most SIA methodologies. This observation seems to temper assertions that SIA is always a vector of increased democracy and deliberation mechanisms in the design of performance indicators (cf. Part 1). However, the one SIA we could investigate through our case studies (i.e. DK1) demonstrated an involvement of all the WISEs’ stakeholders in its development.

3. Differences between WISEs’ expectations when doing SIA and how they perceive the results:

We sought to establish a gap analysis between expectations and reality of the benefits that SIA can bring for WISEs. We did so by comparing the expectations of WISEs interested in SIA (but not yet using it), to the perceived benefits of WISEs having carried out a form of SIA. We highlight three findings (cf. Figure 24):

- a. The effect on public customers is rated stronger on experienced results than on expectations: the experienced effect on public customers rank two notches higher than in expectations (we even find one occurrence of negative effect on the expectations’ side). This finding is however contrasted by the effect on public funders (i.e. providers of funding tied to projects or subsidies, but not to commercial activities) that is weaker in perceived effects than in expected ones.
- b. Perceived effects rank the *“identification of leads to increase impact”* substantially higher than in expectations (by five notches). This suggests SIA as a form of management tool to **improve** the impact that WISEs are generating.
- c. There is a strong alignment between expectations and reality when it comes to SIA as an awareness generation mechanism (mainly, of the WISE’s mission, or the cause it supports): the item ranks first both on expectations and perceived effects.

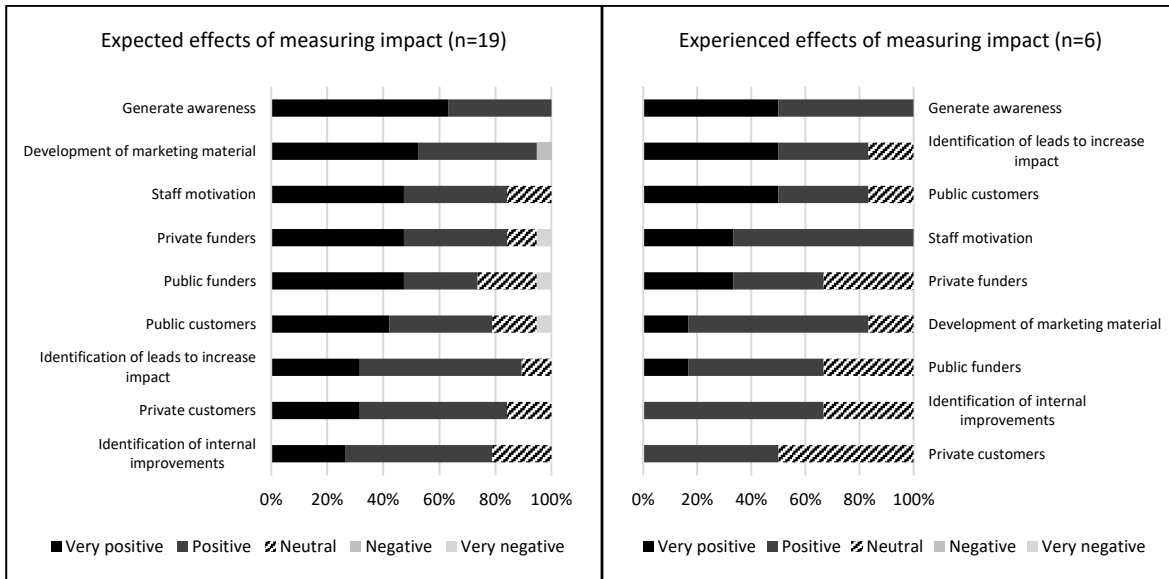


Figure 24: Expectations and perceived effects of SIA

Source: Author

This kind of analysis was difficult to replicate through our case research since (i) we only had one case that actually uses a SIA model (DK1) and (ii) judgement and experience tend to cloud the interviewees' restitution of initial expectations, prior to using SIA. Our conclusions from the research carried out at DK1 however seems to align well with the findings of the survey and provides a much clearer indication of positive effect towards public customers than suggested by the present data.

4. WISEs' perception of performance:

Through our survey, using the seven PAM performance criteria we identify in Chapter 3, we sought to compare which criteria WISEs think should be used by their public stakeholders to gauge their performance and how they perceive these stakeholders are actually assessing their impact. Here again, we highlight three main findings (cf. Figure 25):

- a. Economy (the absolute cost of the services provided) is ranked as the least important criteria by WISEs to gauge their performance, but is perceived as the most important one for their public stakeholders. This is a very clear externality of the **NPG** environment and the (relationship-based) quasi-market it creates for the provision of work integration services.
- b. Utility (how outcomes align to the needs of the society) ranks very close in both cases: it is the first criteria through which WISEs think they should be judged and they perceive it as the second most important factor for their public stakeholders. This suggests a form of alignment about

the importance of outcomes, which is also a defining feature of NPG as we introduce it in Chapter 2.

- c. This observation is however contrasted by the gap observed on effectiveness (whether outputs translate into outcomes) and cost-effectiveness (the efficiency in the realisation of outcomes). This seems to indicate that although WISEs feel their public stakeholders are concerned with outcomes, they perceive that the way by which these are reached matters less to their public stakeholders.

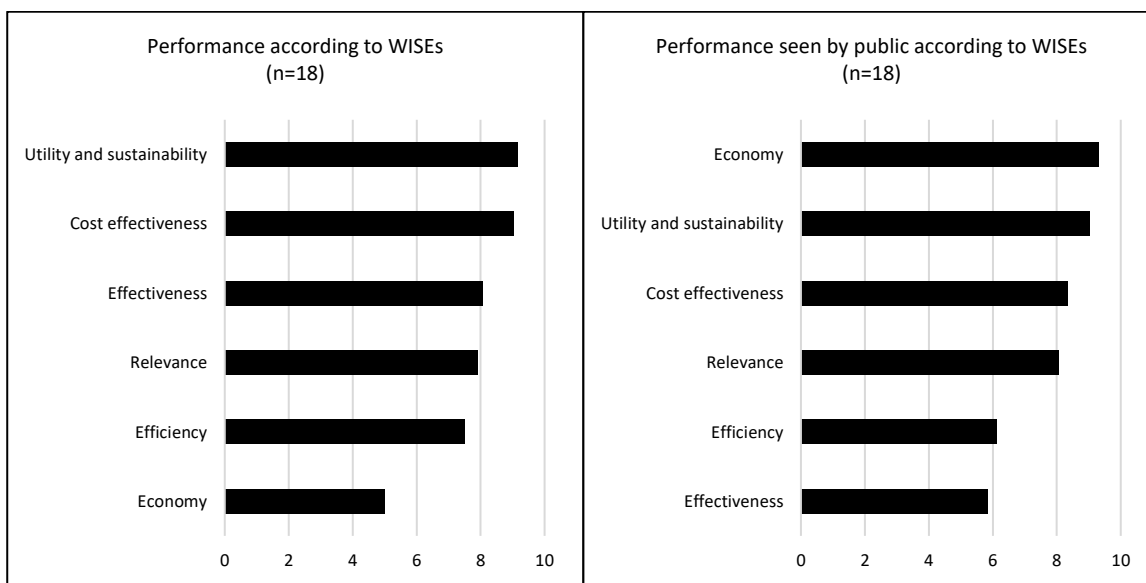


Figure 25: Preferred performance indicators and perceived performance assessment¹⁵⁴

Source: Author

Observations one and two are confirmed through our case studies on the WISEs' side. The picture appeared more contrasted when interviewing public actors, but a focus on economy appeared clearly, while outcomes were cited as a priority.

The following sections of this chapter build upon these findings while bringing further insight from our two case studies in Denmark.

¹⁵⁴ Scores are calculated based on items' ranking order and then normalised on a 0 to 10 scale using the following formula: $\frac{n^2 - Mo \times \tilde{x}}{n^2 \times 10}$. The use of nonparametric statistics (mode and median) allows to avoid some of the caveats associated to the statistical treatment of ordinal data (Siegel, 1957)

9.1.2 Practices observed in our cases

This section introduces the performance measurement practices observed for our two Danish cases presented in 6.3. For each case, we focus on their SIA approach and highlight the way through which their public stakeholders measure their performance (briefly, as it results in large part of the environment described in Chapter 7). We then use the performance measurement framework presented in Chapter 4 to summarise and provide a high-level comparison of these performance measurement approaches.

We introduce DK1's SIA model, a fully-fledged methodology developed in-house. As DK2 has not implemented any formal SIA model, we summarise the WISE's perspectives on the matter and propose a "*hypothetical SIA model*" based on the information collected to inform our analysis. For DK2PS, DK2's main public stakeholder, we present two concurrent approaches to performance measurement. As we attempt to capture the motives and the concepts associated to WISEs' performance measurement, this section makes an abundant use of verbatim, *in italic grey*.

9.1.2.1 DK1

DK1 has gone through a long process (two years) to develop its SIA model. This process was primed by the executive management, who wanted to "*make sure we are doing the right thing*" and wanted to demonstrate a form of performance ("*What are we achieving*"). DK1 started its reflection on the topic through a workshop carried out at the National Centre for Social Enterprises (the now closed government institution tasked with the support and the development of social economy in the country) while it still existed. DK1 first turned towards SROI. We originally attributed this initial choice to DK1's place on the finance law (cf. 6.3), whereby funding from the central government conditions most of its operational vertex. However this hypothesis was discarded by the findings from our interviews:

"Knowledge manager: [The municipalities] are not interested [in SROI], they are only interested in what they are paying for, which is the programme, and the central government is paying for other things. We are already on the finance law, so we don't need that.

(...)

CEO: But if we needed to argue our case with the government, so SROI would be relevant. But we do not. And for the municipalities, it is not relevant. That is why we ditched SROI, we would put too much work in something not of immediate use.”

After almost a year of work on SROI, the WISE turned to a radically different approach where it would attempt to assess the progress made by its beneficiaries purely in qualitative terms. The driver behind this decision was primarily to answer a demand from DK1’s public customers (i.e. the municipalities, through the job centres¹¹⁵: “[We started our SIA] because that was the only way the job centre would say yes to our competence development programme. (...) It started only to be a measurement of our competence development programme but then they wanted us to measure all the programmes that we do.”

Here we see that DK1’s SIA is also the fruit of a form of **competitive** positioning towards the municipality, answering a customer request. While this was initially a demand for performance measurement framed primarily around outputs by the job centre, it is interesting to note that DK1 managed to shift this request towards something that was more meaningful to the WISE, through the development of a fully-fledged SIA, oriented towards outcomes and primarily informed by qualitative input. This signals two things: (i) that DK1 was able to impose a form of **socio-technical arrangement** (STA) to its stakeholders and (ii) that the initial competitive dynamic has morphed into a form of **cooperation to co-define** a desirable SIA on both sides:

“Knowledge Manager: They had this expectation in DK1PS, that we should demonstrate a success rate of 70% or 80% in making people capable to get a job on special terms. And we could not say that, How could we measure that? And that is also why we created this [SIA] because we had to find a way to measure that we were able to enable people to have a job”

CEO: But they wanted it in figures, in numbers, because that was the way of thinking, it was quantitative. But our group [of beneficiaries] is a grey area, you cannot treat them with figures and numbers. So that is the reason why we pushed for this [qualitative] approach.

Knowledge Manager: And also because we believe that if you are able to have success as a [WISE], you have to be able to show that you make a positive impact. So that is also why we talk about it."

While stressing the influence of public actors in the way they frame their approach, DK1 also repeatedly highlighted the importance of SIA as an internal management tool, in order to achieve a better impact: *"No, not proving what I was doing was working, more finding out: is it working? That is the reason. The main reason in this thing (SIA) is to find out: Is it working, is it not working. And if it is working by that prove and provide evidence on what I have done. And if it is not working, allowing me to make better decisions about how to do to improve other people's life."*

We noticed that part of these motives was tied to a perception of a tighter budgetary environment in the public sector, as highlighted by the CEO: *"Sometimes it does feel like we have to do more with less..."*. This is experienced through the pressure that municipalities exercise on prices while negotiating the WISE's services.

The final outcome of DK1's SIA effort is a tool that tracks the progress of DK1's beneficiaries at different stages of the programme they follow. Its components are described in Appendix 12. It assesses beneficiaries on four personal competences (psychological stability, personal appearance, self-awareness and punctuality) and seven professional competences (focus, capacity to follow instructions, work intensity, quality of performed tasks, autonomy, flexibility and adaptability). DK1's SIA structure trails the typical requirements laid out by the job centre in terms of follow-up associated to the delivery of work integrations services. This design was adopted both as a work tool for DK1's staff and as a means to foster the collaboration with the WISE's public stakeholders.

There is a strong focus on outcomes in a qualitative way, that translates into a focus on effectiveness, as well as a degree of attention brought to efficiency in terms of successful assessments (as per the municipality's requirements). Criteria for success are discussed individually between the job centre and DK1, on the basis of the beneficiary's wishes (which may or may not be met). While this process is not exempt of tensions (e.g. when a beneficiary is refused access to a programme of their choosing), this seems to serve as a good basis for an initial **deliberation** on collective preferences (i.e. what is the

preferable outcome for the beneficiary), and *ergo*, a **co-definition** of the public value to be created. These dimensions of deliberation are explored further in this chapter. Lastly, DK1's SIA is characterised by its negotiated design, involving the municipality, DK1's staff and its beneficiaries.

Based on the information gathered, Table 28 structures our understanding of DK1's SIA using the framework we introduce in Chapter 4. Using the same framework, the table also shows how DK1's public stakeholders (PS) typically try to assess its performance.

Table 28: DK1's performance measurement and SIA

Attribute	Variable	DK1	Public stakeholders
Purpose	Screening	0	1
	Monitoring	2	1
	Reporting	1	1
	Evaluation	2	2
Analytical lens	Economy	0	1
	Efficiency	1	2
	Effectiveness	2	1
	Cost-benefit	0	0
	Monetarisaton	0	0
Data typology	Quantitative	1	1
	Qualitative	2	2
Design	Planned	1	0
	Negotiated	2	2

0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus.¹⁵⁵

Source: Author

9.1.2.2 DK2

As seen in 6.3, DK2 does not have a SIA model in use. The closest the WISE has to a SIA is a form of internal quality management where it carries out a systematic assessment of the individuals it supports through a structured approach dubbed KRAP (Cognitive, Resource-Focused and Acknowledging Education)¹⁵⁶. This approach was developed by PsykologCentret, a third-party organisation, and DK2's

¹⁵⁵ The rating is the product of our own judgement based on (i) elements of our interview, (ii) the performance measurement model itself and (iii) the communication material developed about the model's information. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

¹⁵⁶ Kognitiv, ressourcefokuseret og anerkendende pædagogik

staff are all trained to use the KRAP framework. Through KRAP, DK2 is able to document the progress made by the individuals it supports. While this may sound close to DK1's SIA, KRAP is different in the sense that it is purely a work tool with no intent to be used as a performance gauge or to communicate with external stakeholders the way DK1's SIA is designed.

Despite the absence of formal SIA mechanisms, DK2's management has a strong interest in the topic. This interest appears primarily driven by a will to *"show we do good for society"*. When talking about SIA, DK2's executive management was primarily concerned with the fact that they know their organisation has an impact on individuals' lives, but that they do not have the data to rationalise it. When talking about the importance of SIA, the CEO referred several times to her past career as a journalist, and to the fact that she was therefore considering herself a good communicator, able to convey powerful messages. But at the same time, she felt that without impact data, all she could tell would simply remain *"a good story"* in the eyes of her interlocutors:

"Yes, I would love to [have impact data], and not just sound like a story teller. Today I am so much a journalist that I can hear myself when I don't have any facts, of course I can hear that. So I can get the stories and I can also say that I believe that between 60% and 70% of people actually continue in, sort of a job, and then I need facts for the rest. So it would be nice to have some facts."

This concern for communication appears later (cf. below) as a clear focus on generating influence, where SIA is used to **prove** the impact (i.e. as a **device**), ensuring legitimacy (*"[SIA] is a way to demonstrate, yes."*). Similarly, an additional potential benefit is perceived through the capacity to prove that the WISE proposes a cost-efficient approach to equipping youth with the competences required to sustain a job. This idea is brought forth explicitly with an objective to compete with public services: *"I think... I am sure... but I don't have the facts. I think if I took the cost for five years, three years of education, two years of working, so for five years, then it would be very very cheap to take them (the beneficiaries) in DK2 because we bring them in jobs and they don't, all those public places."* Ideas along that line of thoughts were often expressed in conjunction with references to a context of **budget tightening** on the part of municipalities.

Down one notch in the hierarchy, when interviewing the operations manager, communication is still one of the key expectations regarding SIA. She however also brings internal management into consideration as well, placing both elements on an equal footing: *“[We have an interest in SIA] to make sure that we do our job well, that we succeed with what we want to succeed with but also externally to be able to make people understand in a language that they also understand, that we succeed, that we are good at what we do.”* For the external aspects, the operation manager is primarily concerned with proving that DK2 does a good job at work integration, and showing that, in her opinion, they can achieve better results than the programmes ran by the municipality.

Another benefit that DK2 attributes to SIA is a potential for reinforcing business ties with its private customers. There are no expectations that customers will come to DK2 solely on the basis of the social value the WISE is offering, but this is seen as a meaningful addition to the commercial products and services delivered, with DK2’s CEO perceiving a demand for impact documentation. DK2 has for instance recorded interviews with some of their staff placed to work at Ikea, which then used the media as CSR material. Still in order to cater to customers’ interest, DK2 has tried using the UN’s development goals as a framework to talk about their impact, adding, referring to the lack of concrete impact data: *“we can make a small story about how we contribute to [the UN development goals], but we cannot do more because we don’t know”.*

We therefore identify two main drivers for DK2’s interest in SIA: external communication (tied to influence) and management (tied to impact improvement). For the former, we find that there is the underlying assumption, shared by all parties within DK2, that the WISE does indeed create an impact and that it is not something to prove internally. This is especially true at the executive level where management considerations are merely an afterthought:

“Researcher: (summarising previous interactions) It sounds like you would use it [SIA] as an influence tool?”

CEO: Yes, I would. I would.

Researcher: So that’s the primary benefit you see. Do you see more?”

CEO: (laugh) what could it be?... Of course it is important for me to know where we don't succeed, I mean... It would be a management tool for me. Because one thing is that I have the feeling that [it works] and now I would know more precisely... Yes..."

The value in terms of management appears therefore as secondary and was not developed much. This can potentially be attributed to DK2 having its own methodology (KRAP) to measure the progress made by the individuals they support. As KRAP allows to get an understanding of whether what DK2 does translates into an improvement for their beneficiaries or not, the need for validating the impact internally may foster less urgency. As an aside, and to inform our hypothetical SIA approach for DK2, we also note that despite the lack of formal SIA work in the organisation, using their own terminology, the stakeholders interviewed had a clear perception of the difference between outputs and outcomes, and the importance of focusing on the latter rather than the former.

In terms of SIA obstructions, we identify two main factors: the first is a limitation in terms of resources. Resources have been expressed by different stakeholders interviewed as time (man hour), money or knowledge, often in an intertwined way (i.e. having more money would allow to have additional staff to spend time investigating SIA and build knowledge on the topic). This is consistent with the findings of our survey where, for social enterprising prioritising SIA in their agenda, the lack of financial resources is cited as the primary reason why they do not measure their impact (50% of respondents). Lack of human resources and lack of skills respectively come second (45%) and third (30%).

"I think it is a lack of knowledge, and also a lack of money. (...) You know, we are always on the verge, always about to die. – Now this is working – now we don't make money here but I mean... it's such a balance... They don't know but we have to fire three or four persons here after new year because we did not get enough STU¹⁵⁷ spots. (...) Yes, we have to cut down. And then where do you want the people to be? In the research department (i.e. working on SIA)? Or together with the young people out here?"

¹⁵⁷ Særligt tilrettelagt ungdomsuddannelse: an entity providing vocational education to youth with special needs. DK2 has founded such an entity to complement its work integration offer (cf. Section 6.3)

The second obstruction we identify is more implicit and was not directly formulated during the interview process. We however find that due to a lack of consensus and direction on how to do SIA (especially from the public sector) and DK2's own lack of knowledge on the topic, there appears to be a form of "management paralysis" on the topic where not knowing where to start, and not having a body of knowledge readily available, interviewed stakeholders prefer not to take action at all rather than experiencing or starting progressively at a small scale: *"it's funny how everyone has to develop their way of measuring (laugh). And I think that's why it is a big problem that we don't have that... The National Centre for Social Enterprises –you remember?"*

An illustration of that is how the statistics of "60% to 70%" recurrently came up in conversations as the success rate for positive outputs, but was not documented in any way. The main rationale for not documenting was the legal hindrance to carry out a follow-up on individuals after they exit DK2 (the Danish legislation makes this difficult) and therefore the impossibility to verify the transformation of that output into a potential outcome. However, the prominence with which the success rate is used among DK2's staff indicates that this ratio is a meaningful metrics for the organisation, regardless of the superior value attached to the outcomes. But because the approach to measure outcome is not framed yet, DK2 does not take the first step of measuring its outputs.

In DK2's case, and as we touched upon in 6.3, we bring forth additional considerations from the way DK2PS (DK2's main public stakeholder and customer) handles the assessment of its work integration services providers. The administration has set up an evaluation tool dubbed "Effect and progress"¹⁵⁸. As the name indicates, it is interested in both the "effect" of the programme (which corresponds to outputs in the way it is tracked) and the progressions that the programme has allowed on the concerned individuals. Outputs are tracked immediately at the end of the programme in terms of the individual's job or educational situation, and then checked again after two months to confirm the "sustainability" of the outputs. No further checks are made in this regard to validate outputs.

¹⁵⁸ Effekt og progression, where the Danish equivalent of the English word "effect" is used to designate what we would qualify as outputs in impact measurement terms

Appendix 13 provides an overview of the 102 indicators the municipality's framework uses across five categories of challenges experienced by beneficiaries (27 unique indicators since most are used in several categories). For any given indicator, a five-point Likert scale with predefined detailed statuses is used to grade individuals. Individuals are assessed before they take part in the programme, to form a baseline. They then undergo the same assessment at the end of the programme, where their score is measured against the baseline in order to establish their progression.

In the municipality's lingo, it therefore appears that "progression" is close to what one would consider "outcomes" from a SIA standpoint. Besides, while the municipality easily identifies the outputs (or effects), it reports the difficulties widely documented in the literature when it talks about outcomes (or progression): *"The effect is the easy part, but the progression is the difficult part... Because in what way, what to measure, what are the parameters?"*. There is therefore a clear interest geared towards outcome, and a grounded understanding that outputs only provides a limited understanding of a programme's performance.

The municipality's staff described three main goals for this tool (with no order of importance):

1. To create a system that will help identify relevant service offerings: DK2PS deals with 20 WISEs that provide services to put marginalised individuals back towards an employment path, and also hosts its own internal work integration programme. The evaluation framework is set out to help identify which programme works best for what kind of individuals (i.e. screening).
2. To create a frame for a systematic and consistent reporting for the job centre's staff when interacting with WISEs: current interactions are very unstructured and do not follow a standard approach (cf. below). The proposed framework intends to provide a structure for that, making sure that all the relevant aspects of an individual's development are covered.
3. To create a frame to measure the progression made by individuals throughout their work integration path: similar to point 2, the progression made by individuals is itself not well documented. Some WISEs use their own assessment framework and communicate it to the municipality or the job centre, but the municipality does not have a formal or centralised way to process this information.

In its essence, the framework proposed by the municipality aims at structuring and formalising information that is already captured informally by the job centre's staff in their interactions with WISEs. There is therefore an intend to introduce a **device** as a rationalization process in the collaborations with WISEs, in an attempt to go beyond the role that interpersonal relations play in securing long standing collaborations (observed both for DK2 and DK1). We therefore found a clearly articulated attempt by the DK2PS to move towards a more normative approach of assessing the legitimacy of their providers of work integration services. According to our interview data, this approach is not unique in Denmark and was adopted by at least two other municipalities, using their own evaluation framework.

While the measurement framework was designed by the municipality, it is intended to be used by the case workers from the job centre while they are following individuals on their work integration path. At the time of our interview with the project manager leading the measurement effort¹⁵⁹, the tool had been completed and ready to be deployed for a year **but its use was not enforced among the job centre's workers**. On the basis of our interview, this seems attributed to a plurality of factors:

- No mandatory use of the tool instructed by the municipality or the job centre's management;
- The job centre's staff does not consider the tool brings tangible value to their job (i.e. they cover what is scoped by the tool through their regular interactions with individuals and programme managers);
- A reluctance from the job centre's staff to use new tools.

DK2PS's framework was designed over the course of a summer, involving internal stakeholders only. This appears to be a relatively short span of time when compared to the work done by e.g. DK1 on their own SIA framework. Besides, the weak stakeholders involvement runs contrary to most SIA guidelines and to the best practices we identify in Part 1 both in terms of indicator development and of public value strategy (collective preferences cannot emerge). Incidentally, one could associate the poor adoption rate of the tool among the job centre's staff as an externality of the limited stakeholder involvement.

¹⁵⁹ May 2018

As a consequence of this state of affairs, the municipality had no results to share from this measurement framework. This approach to performance had therefore no tangible influence in the relationship between DK2 and their public stakeholders. Instead, the collaboration is very much driven by the personal relationships between the Job Centre's staff and the WISE's management. This **relational capital** is established over the course of several years and turn out to be largely built on interpersonal exchanges and feelings. The performance of the programmes that are ran, and the results they obtain, are encompassed in the scope of these exchanges, but they are never formalised and are left to the appreciation of the job centre's staff. With experience, and as relationships build up with surrounding work integration providers, staffers learn which WISE works best with which kind of beneficiaries.

This is evidenced in the following quote from a job centre's staffer: *"We have worked with DK2 for such a long time, we know that Kate for instance does a good writing and we have talked to her about, many times, what we need to know, and what she has to write to us. We have to know what has been done and tried with a beneficiary. And she knows. So the report we get from her, that's quite perfect. So we can use her when we are going to ask for a flexi-job, because she knows what we need to get it through."*

Based on the information gathered, Table 28 structures our understanding of DK2's SIA using the framework we introduce in Chapter 4. Using the same framework, the table also shows how DK2's public stakeholders (PS) assess its performance (PS1) as well as the new framework for performance measurement designed by DK2PS (PS2).

DK2's hypothetical model appears primarily focused on evaluations aspect, as the management is interested in having *ex-post* results to **prove** that what they do work. This goes hand in hand with the focus on effectiveness (i.e. proving that the 60%-70% ratio of successful output translates into sustained outcomes) and points towards using SIA as a legitimacy **device**, especially in terms of DK2's competitiveness. This is also evidenced by the focus on cost-benefit, whereby the WISE intends to demonstrate that it achieves better outcome at a cheaper price than its public counterparts (here again to reinforce the legitimacy of its services in the eyes of its customers).

The two performance measurement frameworks used by its public stakeholders have the same focus on evaluation but with a greater focus on screening and efficiency for PS2. The design of PS2 is also purely planned, at the opposite of PS1. PS1 is very closed to what is observed at DK1, for which the main difference is an increased focus on efficiency on the part of DK1' municipality.

Table 29: DK2's performance measurement and SIA

Attribute	Variable	(DK2)*	DK2PS1**	DK2PS2***
Purpose	Screening	0	1	2
	Monitoring	1	1	0
	Reporting	0	1	1
	Evaluation	2	2	2
Analytical lens	Economy	0	1	1
	Efficiency	1	1	2
	Effectiveness	2	1	1
	Cost-benefit	2	0	0
	Monetarisatation	0	0	0
Data typology	Quantitative	2	1	2
	Qualitative	1	2	1
Design	Planned	1	0	2
	Negotiated	1	2	0

0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus.¹⁶⁰

Source: Author

*hypothetical SIA based on observations

**performance measurement practices currently in place

*** performance measurement as designed by DK2PS

¹⁶⁰ The rating is the product of our own judgement based on (i) elements of our interview, (ii) the performance measurement model itself and (iii) the communication material developed about the model's information. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

9.2 Effects of New Public Governance mechanisms on strategic interactions

In this section, we explore the effects attributed to the Danish NPG environment on the strategic interactions between WISEs and their public stakeholders. Overall, we observe that the executed strategies must observe a relative **equilibrium** between all three vertices of the PVT strategic triangle, as WISEs must generate legitimacy to secure contracts from their public customers, while this legitimacy builds upon their achievements both in their operational and public value vertices.

We start by investigating the public value and operationalisation vertices, which we tie more specifically to the underlying market mechanisms found in the Danish provision of work integration services. In the public value vertex, we highlight how these mechanisms facilitate the association between public value creation and WISEs' **performance**. In the operationalisation vertex we show that SIA tends to be used as a means to improve **competitive** positioning.

We then proceed with analysing the legitimacy vertex, tying it to some of the key NPG components that are Powell's **reciprocal patterns of communication and exchange** (2003) and **relational capital and relation contracts** (Osborne, 2006). We see how these aspects allow abundant **deliberation**, which in turn facilitates a co-definition of public value. We also see the limits of these mechanism, especially towards the regulation of principal-agent behaviours.

For each vertex, we focus on how SIA, as a possible proxy for WISEs' performance, is used by WISEs to strengthen their position in strategic interactions. We thereby deepen our analysis of the motives associated to SIA that we first surfaced in the previous section.

9.2.1 Market mechanisms and their implications on public value and operationalisation

In Chapter 7, we showed how the Danish NPG environment creates a **relationship-based quasi-market** for the provision of work integration services, where municipalities contract WISEs for the provision of these services, and where the job centres assure the role of daily interface between WISEs and their public stakeholders (cf. Figure 22). In this section, we analyse how these dynamics play out in the public value and operationalisation vertices.

We see that market-inspired mechanisms, by creating a customer – provider relationship, contribute to instigate a clear linkage between public value creation and performance, which is expressed both by WISEs and their public stakeholders. These same mechanisms are also generating considerable pressures on WISEs’ operationalisation vertex as they need to be competitive while performing (i.e. creating public value). To this end, SIA appears to be a useful instrument to release some of these pressures, as it enhances the WISEs’ competitive position.

9.2.1.1 Public value creation and performance

In Chapter 7, we highlighted that Danish WISEs deliver two main types of services for municipalities and job centres: the assessment of work capability of individuals (afklaring) and their training to sustain jobs (ressourceforløb). The main deliveries that come out of WISEs activities for the job centres are evaluation reports (and conversations with WISEs’ staff) stating the beneficiaries’ capabilities, while the desirable outcome is the securing of an education or a job, which can also happen while the beneficiaries are under WISEs’ care. From our interviews, we understand that both public stakeholders and WISEs see this as a form of public value creation, although they do not use the phrase “*public value*” but rather “*contribution to society*”, “*social value*” or “*impact*” in contexts that agree with the framing of public value that we propose in Chapter 2.

We also observe that in the Danish relationship-based quasi-market of employment services, WISEs have to satisfy their public customers in order to ensure business repetition. This is especially true in the high relational capital environment fostered by the NPG’s relationship-based “*component*” of the market, that favours the establishment of “*preferred suppliers*” (Osborne, 2006, 383). On the basis of the above considerations, customer satisfaction occurs mainly through the delivery of the agreed service and therefore through the creation of public value. Public value therefore constitutes a marker of performance for WISEs since it is what their customers measure their satisfaction against. On this, DK2’s CEO tells us: “*I think we have a very market-oriented way of measuring performance because performance is when we succeed with the beneficiaries, and succeeding is when the municipality decides that they got what they paid for.*”

In most cases public value is co-defined between the WISE, the beneficiary and the job centre through deliberation (cf. below). Although we find exceptions, the rule is generally that no performance criteria are formalised. Performance is therefore loosely assessed through the beneficiaries' evaluation reports produced by the WISEs as well as through the cooperation process and the case workers' overall experience with the WISEs. A case worker from DK2PS says: *“Yes, it [performance] is assessed only through the report, and all the conversation I have had about the beneficiaries as well. Hearing about what they can do, reading, or being more social, or whatever.”*

In the Danish environment, we therefore identify a strong relationship between public value creation and performance that is determinant to WISEs' commercial success. That brings an emphasis on the public value vertex where social value has to be **co-defined** through **deliberation** and where activities must indeed contribute to the **creation** of social value, in the sense that it is commonly agreed upon. The former (co-definition through deliberation) rests on a utilisation of the legitimacy vertex, and is explained in detail later in this section, when developing about relational capital. The latter (activities creating public value) happens through the operational vertex and is described in the sub-section below.

9.2.1.2 Pressures in the operationalisation vertex

Throughout our interviews with both WISEs and their public stakeholders, we discern important pressures in WISEs' operational vertex, as they have to secure the resources (mostly economic) to carry out their activities. In both our case studies, as for a majority of Danish WISEs (78% in our survey), revenue is derived primarily from public sources, mainly through commercial activities. Commercial revenue is obtained through price negotiations (often through yearly framework agreements) in a market where WISEs can compete with each other (seldom observed in our research), and where they also compete with the employment services provided by the public sector. In this complex environment, we focus on the role that performance measurement can have to structure some of the observed strategic interactions and release some of the pressures in the operational vertex.

Our survey suggests that SIA leads to an improvement of perception from public customers. Findings from our research at DK1 tend to support this observation. The CEO tells us: *“For us it is a very big*

*selling argument. And for the municipality it is a very big buying argument. Because of the idea of measuring and making better the way you work, while proving your results along the way". Here we clearly see that SIA is conceived as a **device** that **proves** the creation of public value ("results", in the quote), but also as a management tool that helps **improve** the impact. Both increase the legitimacy (and therefore the **competitive** position) of DK1 as a work integration services provider towards Danish municipalities, as the knowledge manager tells us: "And it is easier for us to scale in new locations. Now we are opening in Kolding, and the head of employment at the municipality is very interested [in DK1's SIA]. He already called me, before we even started the measurement, to ask for it."*

Our survey also highlights that pricing is an important factor structuring the strategic interactions in the operational vertex. This is corroborated by our field research, from both sides of the table. The chief economist at Arhus Municipality tells us: *"But at the end of the day it is the price and quality".* We find that SIA is also used to influence that aspect of the contractual relationships. DK1's CEO: *"We are a social enterprise, we do not work for money, we work for a purpose. So we can use this tool too to raise the prices if we need it according to our cost".*

While our research cannot assess whether this attempt of price negotiations through SIA is successful, it provides additional insight on how Danish WISEs envision SIA as tool that can support their strategic interactions also in the operational vertex, and contribute to provide them with a **competitive** edge. This competitive slant is also observed at DK2, where the management is interested in showing that the WISEs' services are cheaper than the ones provided by the municipality (while providing better outcomes and therefore generating more public value).

9.2.2 Reciprocal communication, relational capital and their effect on legitimacy

As we highlighted multiple times while presenting our field data, the way work integration services are provided in Demark, through a relationship-based quasi-market structure, befits particularly well the description of a NPG PAM environment. In Chapter 2, among the key features of the NPG we emphasize are the **reciprocal patterns of communication and exchange** (Powell, 2003, 295) observed in network organisations and the place of **relational capital and relation contracts** (Osborne, 2006, 384).

In the two sub-sections below we delve into the way these aspects of the NPG affect WISEs' legitimacy vertex. We find that reciprocal communication and exchanges favour a strategy of ongoing deliberation where common preferences can be realigned to co-define public value (and therefore performance as seen above) iteratively as WISEs carry out their programmes.

9.2.2.1 Patterns of communication and exchange: legitimacy through ongoing deliberation

In the Danish relationship-based quasi-market of employment services, WISEs have to secure a form of legitimacy with their public customers in order for them to (i) access referring channels and (ii) obtain contracts from the municipalities remunerating them for the services provided. In the absence of formal mechanisms generating legitimacy, like e.g. the French covenants (a form of institutional endorsement), Danish WISEs are left to garner legitimacy on the basis of **deliberation**.

In an attempt to generate a sufficient amount of legitimacy, deliberation capitalises on the two other vertices of the PVT triangle: WISEs have to convince they carry out the right kind of activities at the right price (operationalisation vertex, as we saw above), and that these activities lead to the generation of public value agreed between WISEs and the job centre (performance in the public value vertex, as we saw previously). These are the premises upon which Danish WISEs can position themselves towards municipalities and job centres as a rightful provider of services for the beneficiaries they target.

Based on our field observations, we distinguish between three types of deliberation:

- *Ex-ante*: the deliberation occurring before the signing of a contract, where the aim for WISEs is first to be considered as a potential provider of employment service, and second to co-define a common baseline for the value creation that is to happen through its services, This last step consists in identifying shared preferences, which in turn will set expectations for the services to be delivered. This is illustrated by a quote from a job centre's case worker:

"We talk about [beneficiaries' wishes]. and often it is about what do you dream of doing when you get old. If the candidate says, 'I dream about (...) doing that kind of job', then we look at what DK2 does and see if there is a match. But if they would rather work in an office for instance, I would rather choose another provider."

- *Interim*: when WISEs updates the job centre's case workers, often on a monthly basis, of the progress realised by the beneficiaries. It is an opportunity to potentially recalibrate preferences and realign expectations. This process aims at avoiding eventual unwarranted outcomes, or at least to ensure that they do not come as a surprise for the stakeholders involved, thereby maintaining trust and legitimacy. We can illustrate this with another quote from a job centre's case worker:

“So we just had talks about two beneficiaries, and I have had several conversations with both [DK2's social worker] and the candidates. We are talking about – ‘ok where are you right now, and what are you working with’... or... ‘ok... you have managed to do this new assignment, you have managed 20 hours, now we can try 22 hours’... and we talk about what to practice now, and understand where the limits are, if a mentor is needed, etc.”

- *Ex-post*: when debriefing after the completion of a contract. This is an essential part of making sure the public value created is recognised by the job centre and therefore by the municipality. In cases where the WISE could not reach its objectives, it is an important stage to preserve its legitimacy where it can argue and explain the reasons the desired outcomes were not reached. The centre piece of the closure phase is the delivery of an evaluation report, that details what the beneficiary is capable of, in accordance with what was agreed at the start of the programme. A social worker from DK1 tells us about this process:

“And sometimes in the month two or three they (the job centre) come here, we have a meeting and if the person is finished the case worker comes here and I write a report of two to three pages about the progress, or the lack of it. (...), But then we always talk about the specifics of each case.”

These three forms of deliberation are not systematic. For instance, the interim stage can be handled mostly through written documentation. However WISEs always have the opportunity to meet case workers and discuss the status of beneficiaries if needed. We observe that WISEs tacitly capitalise on these interactions which, combined to the relatively free-form of performance measurement taking place between job centres and employment services providers, allow to reassess the co-defined value creation in the *ex-ante* phase and deliberate on a new baseline for the public value to be created by the WISE.

This has direct implication on the WISEs' performance measurement since through the mechanisms of **deliberation**, they get to modify the agreed-upon targets against which they are supposed to deliver. This is of course not a simple tell tales process and as we explain above, WISEs have to capitalise on their operational and public vertices in order for this deliberation to carry weight. This is however a distinctive feature of the Danish environment that we attribute to the NPG's **reciprocal patterns of communication and exchange** and their capacity to create the necessary amount of trust and transparency to allow for the observed mechanisms to realise. This is also a good antidote to avoid the performativity issue we identify as **indicator-substitution**. We illustrate this last point with the following abstract of our interview with DK2's operations manager:

"Researcher: So what happens if you don't reach the performance criteria?"

DK2: Then we have to explain them really well why we did not... I think worst case scenario is that they don't want to work with us in the future. But if we can tell them that the beneficiaries we got were not the target group they expected... Hum... And, yeah, there could have been other challenges, for instance the challenge with the information that we get, when we get a young person here, that is not consistent with the profile they (the job centre) provided, or when there is not any information... It requires a lot of different things to perform. And in the project, they provide [evaluation] criteria... This is what we deliver, from the job centre, and this is what you have to deliver. And if one part is not delivered, of course that's... So this is the evaluation... But I don't know maybe we will reach the target."

The place occupied by all these deliberation processes and overall patterns of communication and exchanges brings us to the role of relational capital.

9.2.2.2 The role of relational capital

One of the most evident manifestation of relational capital we observe in our empirical material is the **enduring inter-organisational relationships** (Osborne, 2006, 384) we identified between WISEs and their public stakeholders. This is largely fostered by the job centres' case workers, who have the day-to-day relationship with WISEs through the beneficiaries they follow. This relationship, as well as the

de facto evaluator role they endorse towards the WISE's services positions them as pivotal players with which a substantial amount of **relational capital** is built over time. Relational capital is of course not concentrated on case workers and is also found at management levels of the job centres and the municipalities. WISEs are in limited supply, which tends to foster a "*microcosm*" where actors know each other well, hence the **relationship-based** quasi-market.

A first implication of this institutional setup is that the situation of "*preferred suppliers*" (2006) described by Osborne as characteristic of relationship to non-public actors in a NPG environment is observed rather evidently. A corollary to this observation is that, as described in the literature the accumulated relational capital translates into an increased trust between stakeholders (Bovaird, 2006; Teicher et al., 2006). In the context of our research, this increased trust seems to translate into a looser approach to performance measurement on the public side, where indicators are seldom formalised and where perception and feelings play an important role. On the topic, DK1PS says:

"It also depends on how they [WISEs] describe the elements in their programmes and all that. And the experience, and what the case workers feel is the right thing, and what case workers see because obviously our staff sometimes go to their places to observe and talk to them, how they are actually doing it so that impression is part of the evaluation."

We also observe that as experience builds up, case workers tend to channel beneficiaries almost automatically towards specific WISEs. We associate this phenomenon to legitimacy gain. And while we partially attribute this to indubitable public value created through past interactions (i.e. the successful execution of contracts), we also observe that as trust increases through the accumulation of relational capital, legitimacy becomes easier to acquire and maintain through daily interactions. The process of ongoing deliberation we document above can illustrate this idea.

This brings us to the limitations of relational capital. Greve and Hodge (2010) raises the issue that even under the trust-promoting framework of NPG, actors still tend to reproduce principal-agent behaviours. This is a risk that is also highlighted by Crosby et al. (2009), stressing that that even when trying to achieve a common goal of public value creation, actors (private or public) will tend to pursue distinct agenda. We observed this through the entangled roles of customer and referring channel

endorsed by the municipalities, where some WISEs behaviour can appear detrimental to their mission of public value creation because they are concerned with preserving their relational capital. DK2's CEO says:

"It is very difficult for us because we cannot go in supporting the family to fight the municipality [to take beneficiaries in]. We cannot because it (the municipality) is our customer. So the only thing that we can advise them is to find some support through associations that can potentially help. But we cannot go ourselves, because then the municipality would hate us..."

Another limit to NPG and its reliance on relational capital is that, upon the premises of principal-agent behaviours described by Greve and Hodge (2010), one can easily see how the process of ongoing deliberation we describe above may lead to potentially undue rebasing of performance. Leveraging their relational capital, WISEs can attempt to compensate for a weak execution in their operational vertex by attempting to re-establish legitimacy through deliberation, or by capitalising on a high trust environment. DK1's CEO tells us about past abuses (not relating them to WISEs):

"Some years ago, many of the 'mainstream companies' were working like this [providing work integration services] but it was only a way to milk public money"

The underlying market mechanisms at play in the Danish environment suggests that such a strategy would likely not be viable in the long run, as the depletion of public value would reveal the inadequacy of WISEs attempting to follow such a path. It is however under the assumptions that performance measurement is indeed too loosely conducted that some municipalities like DK2PS attempt to introduce formal evaluation mechanisms such as the one we presented above. We also see a similar pattern in DK1PS, where one of the key drivers for DK1 to develop their SIA as it is today was a clear ask from their public stakeholders.

It is by following this lead towards a potentially more stringent evaluation of WISEs that we start the next section of this chapter, investigating the evolution of performance measurement under the NPG framework

9.3 Deliberation or chaos: how performance measurement evolves under NPG

This section summarises the main takeaways from our case studies and uses them to characterise the Danish NPG performance regime. While on the basis of the above analysis, Denmark is clearly in a **deliberative** regime (Jany-Catrice, 2012), our title “*deliberation or chaos*” is a provocative queue to point at some of the tensions we observed and find interesting to deepen in this section.

As such, we look at how existing performance measurement frameworks allow SIA to position itself as an emergent and innovative STA filling a form of void. We also develop further about the benefits of continuous deliberation, especially in terms of performativity avoidance. We temper this analysis with the role that prices can play through the underlying market mechanisms at work in the Danish environment. This section ends with a short conclusion for Chapter 9.

9.3.1 98 ways to measure performance

While establishing in Chapter 7 that Denmark is one of the most decentralised countries in the world (Greve 2006), we insisted on the autonomy enjoyed by the 98 Danish municipalities. This autonomy is fully exercised on active labour policies, and thus on the way to measure their performance. There is therefore, as we see through our case data, no unified way to evaluate the activities carried out by WISEs in Denmark, which leaves each of the 98 municipalities free to devise their own performance measurement scheme, would they be concerned by it. We nevertheless observe in our research that both municipalities and WISEs spontaneously associate considerations on work integration services’ performance to what we qualify as public value creation.

The general rule of performance measurement in Denmark is that WISEs’ deliveries are gauged informally, through the evaluation reports they produce for their beneficiaries and through the impressions of the case workers following these beneficiaries. In this context, having a formalised SIA model such as the one developed by DK1 appears to be a **competitive advantage** since municipalities can suddenly be offered a more formal way to track the performance of one of their providers. To this effect, DK1’s management has been discussing whether or not to license their SIA model for use by other WISEs and municipalities (potentially against payment).

We evidenced that WISEs such as DK1 adopting a fully-fledged SIA can propose a form of STA towards their local stakeholders. We prolong this analysis by arguing that taking advantage of the void created by the absence of structured approach to performance measurement, first movers such as DK1 could attempt to scale these STAs nationally. The observed circumstances of DK1's development in Kolding support this argument. Although this has not been discussed directly, it is an opportunity the executive management seems tacitly aware of, as the CEO tells us: *“CEO: Because it [SIA] is like, I would say, it is like potatoes. In the 14th century: nobody expected potatoes in Denmark because they came from South America. But today everybody expects potatoes, and they could not in the 14th century. It is the same here [with SIA] (laugh).”*

In a similar line of thought, we can mention the indicators developed by Den Sociale Kapitalfond (Denmark's sole impact investor), which we touched upon at the start of this chapter. While offering a free and simple way for WISEs to measure their impact, this stakeholder is also implicitly promoting its own STAs informed by cost-benefit analysis. Municipalities foster their own performance measurement initiatives too. In our cases, it is particularly evidenced by DK2PS, while several other local governments are attempting to implement similar mechanisms (we were unable to obtain precise data as per how many). In DK2PS' case, the underlying intention behind this initiative is to *“rationalise”* the existing evaluation process that is carried out by case workers from the job centre.

We observed that after one year of use, the municipality did not have any results to share because the adoption rate of its model was too poor among the job centre's staff. The main factor associated by our interlocutor to this poor adoption rate is the lack of formal sponsorship among the senior management of the municipality. This sheds an interesting light on the inner workings of NPG; where the flexibility of governance can be detrimental to the implementation of new initiatives, which thrust us back to the remark we highlight in Chapter 2, where NPG can lack mechanisms observed in other paradigms (such as hierarchy in NWS) to maintain the complexity of its network organisation under control (Greve and Hodge, 2010).

In a similar fashion, we were informed during our interviews at DK1PS that the largest Danish municipalities have access to DREAM, a database maintained by the Ministry of Labour that keeps track

of all individuals under any kind of public benefits in Denmark through their social security numbers. Using this data, DK1PS had attempted to create a system to keep track of individuals' benefits use, relating them to the employment services providers they had access to. This failed multiple times in the past three years because of recurring changes in data architecture and systems. This gave us an impression of unfocused and poorly structured approach to defining formal performance measurement frameworks, as illustrated in the following interview abstract:

“Researcher: How long have you had the interest in measuring the outcomes of your providers?”

Chief Economist: I have no idea (laugh)... I came in five years ago and that is one of the topics I started on because I had worked on it before. I wouldn't say that there has been a lot of systematic... sort of measures... People have been talking about it, but I am not sure if it has been actually measured.”

We retain three highlights from these considerations. The first one is that the absence of a centralised process for measuring the performance of WISEs results in a main trend of **loose and informal evaluations**, paralleled by a multiplicity of initiatives attempting to rationalise this process (both from the WISEs side and the municipalities side). The second is that these multiple rationalisation attempts translate into **several STAs that coexist and compete to fill a void**. The third is that the governance features of the NPG, associated to the high form of decentralisation observed in Denmark, seem **detrimental to the development of structured performance measurement**. This can be attributed to a lack of formal hierarchical channels as in DK2PS or to misaligned, unfocused processes as seemed to be the case in DK1PS.

While the above may sound like we are advocating for a centralised or more formal approach to performance measurement, we are simply identifying some of the weaknesses that could hinder the development of performance measurement practices in the Danish NPG environment. This environment also provides a number of benefits that we describe in the following sub-section.

9.3.2 No indicators, no problem? A low performativity environment

Generally speaking we observe that an environment where no formal performance measurement processes are implemented leaves very little room for the undesirable side effects of performativity. In

the absence of identified performance indicators, there is obviously no risk of indicator-substitution. And while retroaction could take place even through the informal evaluation mechanisms at play in Denmark, we could not witness any obvious cases during our research. However, as evidenced in the section above, such an environment is particularly conducive to the formation of STAs, which can be quite diverging in their nature.

In the case of DK1, the STA surrounding the development of its SIA model is inclusive, with the WISE having taken inputs from all its stakeholders and being open to ongoing feedback and modifications. In the case of the model developed by DK2PS, we witnessed a top-down approach that experiences difficulties to be adopted at all by the job centre's staff.

More generally however, we observe that the informal evaluation mechanism encouraged in the Danish environment favours deliberation and the emergence of collective preferences. We see this through the ongoing deliberation we identify between WISEs and the job centre: provided that it is not abused, it fosters trust and avoid situations in which a form of retroaction could take place whereby WISEs would carry out activities inadequate to the beneficiaries' needs simply to comply with the job centre's initial ask.

One limitation we introduce to this analysis is the role that prices play in the relationship-based quasi-market setup of work integration services, where the absence of shared indicators to assess performance (other than economic) can lead to a focus on the wrong metric (i.e. the price) instead of what matters in terms of public value creation (in the eyes of both the municipalities and the WISEs): the generation of positive outcomes. This appears to be at the core of the perception that WISEs have of how their performance is assessed by municipalities (both in our cases and in our survey). While municipalities do acknowledge the importance of pricing when prompted on the topic, they still place the creation of public value as the most relevant performance item.

Conclusion of Chapter 9

As we introduced the performance measurement models observed in our case studies, this chapter characterised the strategic interactions that take place between WISEs and their public stakeholders in

the Danish NPG environment. We summarise these findings in Table 30 using our nine PVT entries. We see that in **NPG’s relationship-based quasi-market**, strategic interactions are kept balanced between all three vertices of the PVT triangle, as WISEs have to operate competitively while creating public value. In this environment, **relational capital** and **reciprocal patterns of exchange** foster a form of trust. This contributes to **deliberation**, leading to the emergence of **shared preferences** and the co-definition of public value. In turn, these elements limit the pervasive effects that pure market mechanisms could have on public value creation and seem to present little ground for performativity to take place.

Having a clearer picture of the strategic interactions and of how the question of performance is addressed through them, we can now start to unearth **strategic intents** associated to performance measurement. We see that Danish WISEs are stuck in a dilemma of **cooperation** and **competition** with municipalities as they strive to create public value for their beneficiaries. Performance measurement, and particularly the development of their own SIA model appears as a key in maintaining an edge in this relationship: both to foster the co-creation of public value and to *prove* that they can be a better alternative than public integration services. But how do these intents influence the measurement of performance? Are they conducive to performativity? And how does that compare to France? As we consolidate the findings of Chapter 8 pertaining to France, we elucidate these interrogations in Chapter 10 by answering the three research questions we formulated in Chapter 5.

Table 30: Summary of strategic interactions in the Danish NPG

PVT entry	Summary
Public value	
(Co)-definition	Co-definition of public value is evidenced throughout the value creation (i.e. <i>ex-ante</i> , <i>interim</i> and <i>ex-post</i>) process and is used to set customers (municipalities)’ expectations.
(Co)-creation	Work integration’s institutional setup allows for a true co-creation of public value, where WISEs are instrumental in the delivery of employment services.
Acknowledgement	Public value creation must be acknowledged for a WISE to “ <i>repeat business</i> ”, it is however seldom evaluated formally (DK2PS is an exception, and its model is not being adopted).
Legitimacy	

PVT entry	Summary
Deliberation	Deliberation is an important part of the negotiation process as it contributes to the emergence of shared preferences, to the co-definition of public value and therefore participate in setting expectations.
Legitimacy devices	There is little use of legitimacy devices in Denmark on the public side, but Danish WISEs certainly consider SIA as legitimacy device to prove their worth.
Endorsement	There is no formal institutional endorsement for WISEs, however tacit endorsement from job centres is required and is often secured through deliberation, the generation of acknowledged public value and sustained relational capital.
Operationalisation	
Agency	WISEs have to balance cooperation and competition strategies with their public stakeholders. This is an important factor that has repercussions across their operational vertex. It can also be a vector of retroaction if WISEs and their customers become focused on prices.
Organisational design	Some forms of organisational design strategies can be observed: e.g. DK2 opening a special education programme (STU)
Activities	The nature of WISEs activities can be the main attraction factor for both customers (municipalities) and beneficiaries.

Source: Author

Chapter 10 How strategic interactions frame the measurement of performance in different contexts

This last chapter of our analysis brings together the elements we need to answer our main research question: how strategic interactions between work integration social enterprises (WISEs) and public actors are contributing to frame the measurement of performance in different public administration and management (PAM) contexts? We recall that in order to research this phenomenon, we chose to focus on the strategic interactions between WISEs and their public stakeholders as the nodal point where performance measurement questions are addressed. We also recall that we place a particular attention on the effect these interactions may have on issues pertaining to performativity (or the influence that performance instruments can have on what they *measure*).

We start by summarising the different models of public value co-creation in which WISEs operate in the French Neo-Weberian State (NWS) and the Danish New Public Governance (NPG) and the nine approaches to performance measurement we observed in these two environments. We use this recap as a foundation to answer our three research questions.

How different PAM contexts translate into different performance regimes (**RQ1**)? At the macro level, our second section draws parallels between each PAM environment and how the different performance measurement models we encountered can be characterised. We identify two delineations that structure our view: (i) the market logics observed in NPG and how they oppose the legalist and administrative logics found in NWS and (ii) the antagonism between centralisation and decentralisation in the two environments. Doing so, and following our critical realist (CR) paradigm, we evidence a first array of macro factors contributing to explain the observed performance measurement practices.

What are the strategic intents of WISEs measuring their performance and what are the implications for the way this measurement is done (**RQ2**)? In a third section, we move to the micro level and summarise the strategic interactions we observed between WISEs and their public stakeholders in the French NWS and the Danish NPG. We do so using the nine entries of the PVT triangle we developed in Chapter 2. From there, we identify seven crystallisation points that contribute to explain WISEs' behaviour in

terms of performance measurement (whether it is their attitude towards the instruments used by their public stakeholders or their own social impact assessment model). We call these crystallisation points **strategic intents**. As we identify and characterise them, we build the foundation to answer our third research question.

Are strategic interactions related to performance measurement conducive to the avoidance of performativity issues (**RQ3**)? In a last section, we use the strategic intents to frame the diffuse nature of strategic interactions and understand how, when related to performance measurement, they may have an effect on performativity. We describe the mechanisms through which some strategic intents are more likely to foster deliberation, which in turn favour the emergence of shared preferences, which can evolve into the reinforcement of socio-technical arrangements (STAs). We then show the effects, both positive and negative, that competing STAs can have in terms of performativity. Doing so, we identify the strategic intents as a second array of explanatory factors for the observed performance measurement practices.

10.1 Summarising the observed practices

This first section summarises the key findings of our research. We start by articulating the two different models of public value co-creation we observed in France and in Denmark. Second, we recapitulate all the performance measurement models we came across, identifying patterns across countries and organisations. The overview we provide in this section serves as a foundation for the following three segments of this chapter, which answer our three research questions.

10.1.1 Different models for co-creating public value

We establish throughout our thesis that WISEs are creators of public value (both theoretically and through the views of field actors). We also show that in France as well as in Denmark, WISEs have to generate a form of legitimacy with their public stakeholders in order for them to access the means (referring channels and access to economic resources, under the form of subsidies or market incomes) to create this public value. In effect, WISEs are therefore **co-creators** of public value, together with their public stakeholders, in both countries.

The different PAM environments in which France and Denmark operate however translate into two distinctive co-creation mechanisms which we summarise in the two sub-sections below.

10.1.1.1 Legalist and centralised co-creation in France:

In France, the NWS' legalism inherited from the "*classical*" Weberian state turns the co-creation of public value into an exercise of rules-following where WISEs must seek to meet the requirements needed to preserve the holy grail that are the covenants (i.e. what we identify as **institutional endorsement** in our PVT-informed analysis). Without covenants, WISEs cannot access the resources allowing them to co-create public value. The NWS traits are probably amplified in a Napoleonic country like France (by opposition to e.g. a Continental European Federal country like Germany, also traditionally associated to NWS).

This rigid structure, combined to the NWS' centralisation, generates a very top-down relationship with public stakeholders: the main interlocutor of WISEs in France is the Direccte, an emanation of the

French central authority, and the person administrating the covenants is the Préfet, a direct representative of the executive power. The corollary to this state of affairs is a high level of **compliance** among French WISEs, which adopt the behaviour they need in order to maintain the **legitimacy** they require to operate. In our analysis, we echo this to Gianfaldoni's subordinate relations (2018). These views are articulated in both of our case studies: FR1 explains about the opportunity cost of compliance, which represents substantial man-hours and even capital investment (in terms of e.g. software) and FR2 describes at length the discrepancies in the way they are asked to report their outputs.

This compliance does not mean that French WISEs remain passive in their environment. We see that the modern tenets of NWS (the "*neo*"), has been conducive towards more deliberation mechanisms, particularly at the local level, where WISEs can have a direct exchange with their public stakeholders and re-instate a dynamic of co-creation through for instance the management dialogues, which allow to negotiate objectives and therefore to loosen the legalist stranglehold, but also to foster the build-up of a relational capital and mutual trust.

10.1.1.2 Market mechanisms kept in check by network governance in Denmark:

In Denmark, the observance of the NPG instates a sensibly different dynamic, characterised by an increased openness, accessibility and transparency, but where WISEs act in a **relationship-based quasi-market** through which they have to commercially sell their services to public customers (the municipalities). WISEs are competing against other WISEs (although this specific form of competition did not appear as a defining part of their operational vertex) but also against public stakeholders. The latter relationship is for them more difficult to manage as they cannot go head-to-head with a stakeholder that is also their main customer and their referring channel.

In this complex setting, WISEs are co-creator of public value through a dual relationship of competition and collaboration with their public stakeholders. The NPG mechanisms of relational capital, trust and exchanges seem to bring a form of balance in what could otherwise turn into an environment potentially destructive for public value through pervasive rent seeking or misguided agent-principal behaviours. In this model, abundant deliberation leading to an **iteratively co-defined public value**

along the process of co-production, allows to produce the social glue of shared values that prevent “pure” market mechanisms to be potentially detrimental to public value creation.

However this is not sufficient to create an “ideal” environment. In our research, we observe how DK2 attempts to preserve its legitimacy towards the municipality and prevents itself from interfering in the referral process of potential beneficiaries. This is due to the entangled role of customers and referring channels that municipality have in the Danish environment. In France, we see the exact opposite where WISEs can become actors of the referral process. For instance we saw that with its legitimacy protected by the NWS’ legalist framework, FR1 takes an active part in supporting their beneficiaries to get a sponsorship, confronting the reluctance of job centres.

Moreover, the absence of formal evaluation mechanisms in Denmark seems to foster isolated performance measurement initiatives which, in the absence of centralised guidelines, can translate into local attempts at developing non-inclusive reporting instruments (as observed in DK2PS). We see such initiatives rooted in an attempt to bring structure to existing informal processes. Their implementation nevertheless tends to be chaotic and not followed through.

10.1.2 Observed performance measurement practices

The two different models of public value co-creation we describe above translate into different performance measurement practices, as observed in our research. This short section summarises these practices. We remind of the models implemented by both WISEs and their public stakeholders. We then highlight the most meaningful patterns we can identify across countries and organisations in each of the five attributes we use to structure our analysis of performance measurement models.

The first conclusion we draw is that there is no *silver bullet* for the measurement of WISEs’ performance. It is a key highlight of the existing social impact assessment (SIA) literature (Clark et al., 2004; Grieco, 2015; S. Olsen & Galimidi, 2008; Tuan, 2008), and a point that emerges very clearly from our material: a common framework for performance to be used across all social enterprises would be extremely difficult, if not impossible to implement. Our research shows that just within the sub-segment of WISEs, which is fairly homogenous in terms of vision and mission (and therefore in terms

of ambitious outcomes), organisations adopt very heterogeneous approaches to performance measurement, as we summarise in Table 31, together with the performance measurement approaches adopted by the WISEs' public stakeholders.

Table 31: Summary of encountered performance measurement approaches*

Attribute	Variable	French NWS				Danish NPG				
		FR1		FR2		DK1		DK2**		
		WISE	PS	WISE	PS	WISE	PS	(WISE)	PS1	PS2
Purpose	Screening	0	2	0	2	0	1	0	1	2
	Monitoring	2	1	1	1	2	1	1	1	0
	Reporting	1	2	0	2	1	1	0	1	1
	Evaluation	1	1	2	1	2	2	2	2	2
Analytical lens	Economy	0	0	0	0	0	1	0	1	1
	Efficiency	2	2	1	2	1	2	1	1	2
	Effectiveness	0	0	1	0	2	1	2	1	1
	Cost-benefit	1	0	1	0	0	0	2	0	0
	Monetarisatation	0	0	0	0	0	0	0	0	0
Data typology	Quantitative	2	2	2	2	1	1	2	1	2
	Qualitative	1	0	2	0	2	2	1	2	1
Design	Planned	2	2	0	2	1	0	1	0	2
	Negotiated	1	1	2	1	2	2	1	2	1

0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus¹⁶¹

Source: Author

* For each WISE, the WISE column defines the model used by the enterprise and the column PS identifies the performance measurement practice of their public stakeholders.

** The use of brackets for WISE indicates the use of a hypothetical model. PS1 is the performance measurement in place at the time of the research and PS2 is the one developed by DK2PS but that was not implemented yet.

Comparing practices observed across our cases in Denmark and in France, we highlight the following patterns for each performance attribute:

Evaluation is a common purpose to all performance measurement systems:

In the purpose attribute, we see that evaluation (i.e. assessing an activity *ex-post*) is a variable common to all models, on the WISEs side as well as on the public side. It is applied with a different emphasis by

¹⁶¹ The rating is the product of our own judgement based on (i) elements of our interview, (ii) the performance measurement model itself and (iii) the communication material developed about the model's information. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

different organisations. For instance, DK2 sees it as a main purpose while FR1 is more concerned with monitoring its activity (i.e. an ongoing use of performance measurement as a management tool), while still worry of ex-post evaluation.

In the analytical lens, patterns of monetarisation, efficiency and economy:

Monetarisation (i.e. using financial proxies to place a monetary value on outcomes) is a practice that we did not observe in any of our cases. Our survey in Denmark however finds evidence of WISEs using monetarisation approaches. Findings from our cases suggest that this approach was more popular when social impact started to become a “*hot topic*” in the early 2010s than it is now (which may be the timeframe in which some of the surveyed WISEs have developed their own model).

Different than monetarisation, but still about placing a financial value on impacts, we observe an interest for cost-benefit approach shared by all cases except for DK1. We find that this kind of approach is often associated to the intent of proving its impact, indicating the use of performance measurement as legitimacy device, which is mostly used towards stakeholders from the public sector.

Economy is only considered in Denmark, and solely by public stakeholders. This is something we also observe in our survey and is attributed to the relationship-based quasi-market model in which Danish WISEs carry out their activities. We however see efficiency across the board which is the result of outputs always being captured to a certain extent, whatever the purpose is (monitoring for management purposes or as a proxy for performance).

More focus on quantitative indicators in France:

We observe an overall focus on quantitative indicators more important in France than in Denmark. This is driven by (i) the systematic implementation of quantitative performance measurement by the centralised state and its local emanations and (ii) a tendency observed in cases like FR1 to tacitly adopt this quantitative logic in their own SIA model. FR1 is the WISE that puts the heaviest emphasis on quantitative indicators, although the management stresses the qualitative phenomena they want to assess.

Denmark much more conducive towards negotiated design:

In Denmark, the emphasis on negotiated performance measurement design is stronger than in France, and is observed also in the models implemented by the public administration (except for the model we saw developed at DK2PS, where it was weaker). Here we see the mark of the NPG, more conducive to deliberation, but we cannot abstract this observation from the Nordic administrative profile and its tradition of accessibility and transparency (cf. Table 1).

On the basis of the variable approaches to performance measurement and the different public value co-creation environments we summarised in this section, the following three sections answer our three research questions by looking at the articulation of the elements we just recapitulated.

We start at a macro level, where we look at how the two PAM contexts and the different mechanisms for public value co-creation contribute to explain the diversity of performance measurement approaches we presented here. We then explain these differences further, at a micro level, by looking at the strategic interactions between WISEs and their public stakeholders, and by characterising WISEs' strategic intents associated to performance measurement. Equipped with this understanding of what is driving the observed differences in performance measurement practices, we conclude by looking at whether they can contribute to decrease the issue of performativity.

10.2 How PAM models shape the practice of performance measurement?

In this section, we answer our question **RQ1**, which is:

RQ1: How different PAM contexts translate into different performance regimes?

Both in France and in Denmark, our field observations establish that WISEs, as well as their public stakeholders associate the performance of work integration activities to the creation of public value. In both countries, we also observe a consensus among all stakeholders (WISEs and public actors, both at the central level and at the regional level) that public value is better captured through outcomes than through outputs. Furthermore, we also see a common environment of public budget tightening, where an increased accountability is asked from WISEs. We observed that both in France and in Denmark, public stakeholders are indeed on a trajectory of increased control which translates into the implementation of various mechanisms to gauge the performance of their providers of work integration services.

From this common baseline, we however observe a number of differences in the actual approach to the measurement of performance. These differences are observed among WISEs as well as among their public stakeholders. We structure this divide along two defining vectors: the **market logics** of NPG that we oppose to the **administrative and legalist logics** of NWS, and NPG's decentralisation that runs contrary to NWS' centralisation.

10.2.1 Market logics versus administrative legalist logics

The first demarcation line we identify, market logics versus administrative and legalist logics, is rooted in the different mechanisms leading to the co-creation of public value that we highlighted in the first section of this chapter. We see that this divide can act as an explanatory variable for the unidimensionality versus multidimensionality nature of performance measurement as theorised by Jany-Catrice (2012) – i.e. the tendency for performance measurement to be centered on a single indicator or to encompass several.

In France, the NWS' legalist approach, together with its bureaucratic apparatus, will favour the adoption of *operationalizable* metrics. That is to say indicators that must be simple (in terms of

definition) for WISEs to implement and easy for the administration to track. Metrics designed in this way should not leave room for interpretation and can therefore only be narrow in meaning. They exemplify the *“measuring is a cognitive simplification”* from Jany-Catrice’s framework.

The graduation rates introduced by the DGEFP (the centralised administration) and used by the Direccte (the deconcentrated authority) are a good illustration of this. They are calculated using simple rules leaving no space for interpretation and are a typical form of output metric. As we emphasise throughout our thesis, the literature tells us that outputs are a rather poor means to assess public value (Bryson et al., 2014). The use of indicators with such a restricted meaning is however mandated by the NWS’ legalist approach to work integration: the devices built around these metrics become the instrument that determine whether a WISE can carry out its activity within the framework of the law and must therefore enable clear-cut decisions.

Graduation rates were adopted in 2008 as the single metric through which WISEs should be assessed. One unique, centrally planned indicator to measure WISEs’ performance brings us into the **autocracy** regime described by Jany-Catrice (2012). We however assimilate the introduction of modulation criteria in 2014 to an expression of the outward swing of NWS (i.e. a move from inward-centric bureaucracy to more outward **managerial** practices, more concerned with results, cf. our Chapter 2 and more generally Pollitt & Bouckaert, 2017). This reform allowed for a few more indicators to be taken into account for WISEs’ performance, although still very unidimensional in their meaning (cf. Table 20).

In Denmark, the relationship-based quasi-market in which WISEs operate pushes the observed dynamics towards the opposite side of the spectrum, where WISEs are judged through a plurality of dimensions. In most cases, there are no indicators to speak of, and performance is assessed along the delivery of services, which is favoured by the relationship-based feature of the work integration quasi-market.

In this particular form of market, our findings show that the criteria for performance (even though not formally expressed through indicators) are indeed multidimensional, as WISEs have to be price

competitive (Denmark is the only place where we observe the use of economy, as a performance indicator, albeit only on the public sector's side, cf. Table 31), at the same time as the activities they propose have to compel to both public actors and beneficiaries, and indeed have to co-create public value amid these constraints.

In the Danish NPG environment, we are therefore getting away from the cognitive simplification that measurement can be in the NWS, and closer to the *"revealing"* nature Jany-Catrice associates with **multidimensional** performance measurement. We however find that (rather unsurprisingly) market mechanisms, even contained by shared values, can have a tendency to bring performance back towards cognitive simplification, especially towards the economy criteria, which indeed has a prominent use among municipalities.

This is where the relationship-based features of the quasi-market, fostered by the NPG's conduciveness towards relational capital and associated trust bring tangible value-added to the practice of performance measurement (particularly towards the assessment of the co-created public value). The way this part of WISEs' performance is assessed is best explained using the second delineation we identify to explain the different performance approaches we observe between France and Denmark.

10.2.2 Centralised NWS versus decentralised NPG

The second line of demarcation that emerges from our research is the centralised versus decentralised approach to performance measurement, which we find can be proposed as an explanatory variable for the planned versus negotiated performance in Jany-Catrice's framework (2012).

In Denmark, this is evidenced by how the degree of liberty enjoyed by municipalities in defining their approach to performance leaves space for negotiated, if not indicators, at least targets. These targets are left to be decided between the job centre, the beneficiaries and the WISEs as the desirable outcome for a particular individual. It is co-defined through **deliberation** which occurs, as we show in our research, iteratively along the delivery of the work integration services.

The Danish NPG therefore appears to follow Jany-Catrice's description of a **negotiated** regime: complex, and including qualitative considerations. The NPG's decentralisation appears to be a key element in this process, avoiding the use of centralised indicators detached from the stakeholders they have to be applied upon. Decentralisation is also what allows municipalities to experiment with performance measurement. We witnessed this trend particularly through what we observed at DK2PS. Through this example, and through what we observed at DK1PS, we also realised that NPG can sometimes lack the mechanisms (Greve & Hodge, 2010) to ensure the proper implementation of its performance measurement.

We tie these local experimentations to the environment of budgetary tightening that is often part of our interviewees' narrative. The intent is to normalise the existing evaluation process by the use of metrics defined and adopted within the municipality (very obvious in DK2PS). While such initiatives bring us closer to the realm of planned metrics, they remain the exception rather than the rule, and what we observe in of DK2PS shows that they meet resistance in their implementation.

In France, we observe an opposite trend where the NWS' centralisation translates into the creation of a top-down performance measurement model, based on a restricted set of indicators. These indicators are tied to the national *"employment policy"*¹⁶² and are **centrally** designed by the DGEFP. These premises therefore seem to anchor us in Jany-Catrice's planned regime of *"simple, likely to be quantitative"* performance measurement.

However, in the same outward swing of the NWS that we highlight in the previous sub-section (i.e. the *"neo"* in NWS), we also observe some degree of deliberation associated to this form of centralised performance measurement. Although WISEs do not get to deliberate about the indicators' design, they are offered a degree of negotiation with the Direccte as per the value these indicators should be set to. More crucially, the centralised system also plans space for deliberation through the management dialogue, where WISEs get a chance to deliberate with the plurality of local stakeholders shaping their bureaucratic operational environment.

¹⁶² Politique de l'emploi

This leads us to a contrasted performance measurement approach in France, that we identified as predominantly anchored in a planned design (at the macro level), but with features that allow us to get it closer to a negotiated performance regime (at the micro level).

10.2.3 What the public context of performance measurement means for WISEs' SIA

Our research shows that WISEs, while submitting to the performance measurement initiatives of their public stakeholders, also develop their own understanding of how their performance should be measured and operationalise this understanding through SIA models.

We observe that the developed models will, to some extent, echo parts of the processes or indicators used by their public stakeholders to measure performance. For instance, FR1 makes an intensive use of all the output indicators that have been used (sometimes on isolated initiatives) throughout the history of its development. In Denmark, DK1 has framed the processes of its approach around the process of follow-up and evaluation that structures its cooperation with the job centre.

However, we do not discern any clear pattern of how WISEs position their SIA against the performance regime adopted by their public partners. A general trend would seem to indicate that WISEs tend to use more indicators in their SIA to better reflect the multidimensionality of their impact. However this is contradicted by the coding of our conversations with DK2, which indicates mainly a preoccupation for selected outputs indicators and for cost-benefit metrics. Based on the process observed with DK1, our other Danish case, and out of the fact that DK2 had not actually developed a SIA model at the time of the research, we could consider that this initial focus on restricted metrics is part of the development process of SIA.

Nevertheless, as we chart our conclusions using our interpretation of Jany-Catrice's performance regime framework, we stick to the facts emerging from our analysis and plot the different approaches to performance measurement we observed in Figure 26.

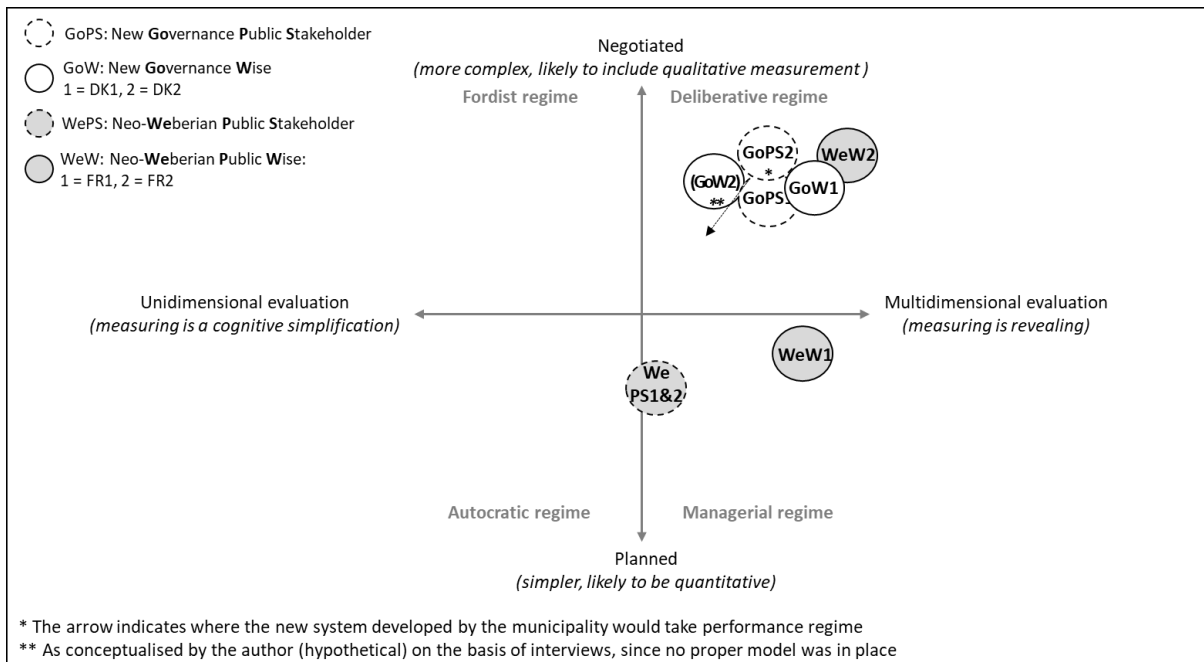


Figure 26: Observed performance regimes¹⁶³

Source: Author

Figure 26 synthesises our answer to **RQ1**:

RQ 1: How different PAM contexts translate into different performance regimes?

The French NWS and the Danish NPG do indeed translate into different performance regimes for public stakeholders. These regimes appear in flux, where we observe that a common environment of stringency in public funding translates into a “re-centring” of performance measurement in Denmark (i.e. towards less negotiations and an increased unidimensionality) and a shift towards the opposite direction in France as the “neo” pendant of NWS injects more managerialism and favours a form of decentralised negotiation.

The influence of the environment is less obvious with WISEs’ SIA, where we see them adopting some indicators and processes from their public stakeholders, but where we also observe them following

¹⁶³ Using Jany Catrice’s framework (initially proposed as a table) as a plane allows us to introduce more granularity in our analysis. We used our own judgment to plot the different models encountered. The models are first plotted according to their absolute characteristics and then positioned relatively to one another. E.g. we first position GoW1 (DK1’s SIA) relatively high in the *deliberative* section of the chart, before comparing it to the other models in this box and adjusting its position relatively to that of the others: WeW2 (FR2’s SIA) is higher than GoW1 on the vertical axis because it includes more stakeholders in its design, and it is considered more multidimensional because it has *inter alia* economic indicators that are not encompassed by GoW1. WISEs’ SIA are assessed independently of their environment since they are WISEs’ own models and are operationalised on their own terms (i.e. WeW2 is highly deliberative despite coming from a covenanted entity).

their internal agenda and needs in the hatching of their models. We see that most evidently with the difference between FR1 and FR2: the former makes an abundant use of top-down designed outputs to steer managerial decisions whereas the latter uses inclusive processes to define multiple indicators in order to inform and educate its stakeholders. These internal agenda and needs are the object of this chapter's next section.

10.3 Strategic intents and what they tell us on the conceptions of performance and its measurement

This section answers our second research question:

RQ2: What are the strategic intents of WISEs measuring their performance and what are the implications for the way this measurement is done?

Building upon the two public value co-creation models we summarised in this chapter's first section, we start by recapitulating the strategic interactions we observed in our cases using the nine entries we developed for the PVT triangle. We do so by exhibiting the coding we made of PVT-related topics in our interviews.

We then focus more particularly on the strategic interactions concerned with performance measurement. Still resting on the PVT triangle, we identify seven crystallisation points that can explain WISEs' motives for performance measurement in France and in Denmark. We call these crystallisation points **strategic intents**. We describe these intents and use them to structure our understanding of how WISEs' performance measurement (through whether their own SIA or the tools used by their public stakeholders) is carried out.

10.3.1 Different strategic interactions in France's NWS and in Denmark's NPG

We started this chapter by recapitulating how the differences in the Danish NPG and the French NWS PAM models contribute to the implementation of distinctive co-creation mechanisms for public value. These mechanisms are structured by a legalist centralisation in France and by a relationship-based quasi-market in Denmark. Throughout our thesis, we evidence that these mechanisms translate into different behaviours on the part of WISEs. We identify these behaviours as strategic interactions and we used the PVT triangle (Moore, 1995) to chart them.

In this section, we summarise the main characteristics of the strategic interactions we observed in our research. We do so by exhibiting the frequency table of the level 3 coding we realised with NVivo on interview items pertaining to PVT (level 1 code 2 in our coding grid, cf. Appendix 16). Level 3 codes

reflect the three sub-components of PVT that we developed in Chapter 2. In Table 32, for each level 3 code, we provide:

- The frequency at which the code is observed in our interview corpus¹⁶⁴ (often a good gauge of the importance of the item)
- An order of magnitude (low to high importance), based on coding content, and completing the information provided by the code frequency
- A short description summarising the main items behind the coded interview content.

Table 32: Strategic interactions paradigms in France and in Denmark

Code ¹⁶⁴	France	Denmark
2 PVT	100%	100%
2.1 Public value	23%	32%
Code 2.1.1 (Co)-definition	3% -- Low importance: Little verification carried out other than through devices focused on outputs	7% -- Medium importance: Required to set the bar on the service to be delivered
Code 2.1.2 (Co)-creation	13% -- High importance: The necessity of public value creation is shared and understood among stakeholders at the local level	14% -- High importance: The necessity of public value creation is shared and understood among stakeholders at the local level
Code 2.1.3 Acknowledgement	7% -- Medium importance: Required to maintain endorsement (covenants), but secured through output- focused devices and deliberation (i.e. two legitimacy instruments)	11% -- High importance: Needed for “repeated business”, but no formal assessment method
2.2 legitimacy	56%	31%
Code 2.2.1 Deliberation	12% -- Medium importance: Allow to negotiate the content of devices and a	13% -- High importance: Where the negotiation processes take place, which

¹⁶⁴ Frequencies are computed by NVivo and are expressed as the percentage of total interview material coded in our Public Value Strategies main code (2). E.g.: 56% of the interview material coded in Public Value Strategies in France (code 2) was coded under the legitimacy sub-code (2.2) and 12% of the interview material coded under code 2 in France was coded under the deliberation code (2.2.1). Frequencies are aggregated at level 3, i.e. they encompass sub-levels (cf. full coding grid in Appendix 16).

Code ¹⁶⁴	France	Denmark
	channel to maintain endorsement	sets the premises for successful contract execution
Code 2.2.2 Legitimacy devices	22% -- High importance: Institutional devices allow to maintain the endorsement mechanisms	10% -- Low importance: No or limited formal devices in place
Code 2.2.3 Endorsement	22% -- High importance: Institutional endorsement (covenants) are required to operate	8% -- Medium importance: No formal institutional endorsement for WISEs, however tacit endorsement from job centres is required (often secured through deliberation, the generation of acknowledged public value and sustained relational capital)
2.3 Operationalisation	21%	37%
Code 2.3.1 Agency	1% -- Low importance: WISEs may compete for access to institutional endorsement (covenants) in some geographical areas	12% -- High importance: WISEs have to compete and / or cooperate with peers and public services
Code 2.3.2 Organisational design	8% -- Medium importance: Some covenants limit the type of activities or of revenues	6% -- Medium importance: Does not affect work integration directly but enter in considerations for diversification (e.g. STUs)
Code 2.3.3 Activities	12% -- Medium importance: Legally acceptable activities are partly guided by covenants, which can also provide access to some public contracts	19% -- High importance: Enters into consideration for beneficiaries channelling through the job centres

Source: Author

The code frequencies are generally a good proxy for the relative importance of each item. A notable exception is the weight occupied by legitimacy devices in Denmark, which is more a result of

conversations focused on performance measurement than of the actual importance of devices for WISEs in Denmark. Overall, we see that French WISEs show a heavy emphasis on the legitimacy vertex (which concentrates 56% of the PVT coding), while Danish WISEs demonstrate an emphasis on their operationalisation vertex (37%). The emphasis is however a lot less pronounced in Denmark than it is in France, as WISEs deploy strategies that are more balanced across the PVT triangle (we see that each vertex reaches almost a third of the coding).

10.3.2 Identifying seven strategic intents for performance measurement

Table 32 provides an overview of what determines the strategic interactions between WISEs and their public partners occurring through the co-creation of public value. It enables us to get a general sense of what matters most to WISEs depending on the context in which they operate. Using this strategic nexus as a foundation, we focus on understanding what drives WISEs' performance measurement (both by using the instruments of their public stakeholders and by developing their own SIA).

Doing so, we recognise seven crystallisation points that can justify the use of performance measurement in the contexts of the Danish NPG and the French NWS: four in Denmark, and five in France, with two observed in both countries. We call these crystallisation points **strategic intents**. Strategic intents show how WISEs attempt to use performance measurement (both their own SIA and the performance models used by their public stakeholders) in one or more vertices of the PVT triangle to leverage another vertex.

Before proceeding with our presentation of the strategic intents we stress that they are not mutually exclusive and are in fact often observed occurring concurrently. We use these intents as structuring vectors to characterise the strategic interactions we observed. Figure 27 provides a visual queue for apprehending the seven intents observed in the two countries by depicting them in the PVT vertices they mobilise. In this figure, we intentionally distance ourselves from the *"traditional"* representation of the PVT triangle (cf. Figure 4) by using a Venn diagram. This allows to illustrate in a more compelling fashion (and in a way that reflects the coding of our data) the dynamics occurring when specific vertices of the triangle interact with each other.

Our identification of seven strategic intents builds upon the coding of our data in Nvivo. Since we adopted a plurinomial coding (i.e. one bit of text can be encoded using several codes, Ayache & Dumez, 2011), we retained all coded segments pertaining to both public value (2) *and* performance measurement approach (3) *or* conceptions of performance measurement (4). Our selection, covering 56% of our data, allowed us to identify how some aspects of performance measurement had influence in several vertices of the PVT triangle (for instance maintaining legitimacy for both FR1 and FR2 takes a toll on operations because of the investments and staff required to meet administrative requests, or DK1's SIA is used internally to understand whether activities are generating social value). As we identified themes that seemed to justify the observed associations, we used information coded under PAM environment (1) to further structure our understanding of the observed regularities and boil our results down to seven intents.¹⁶⁵

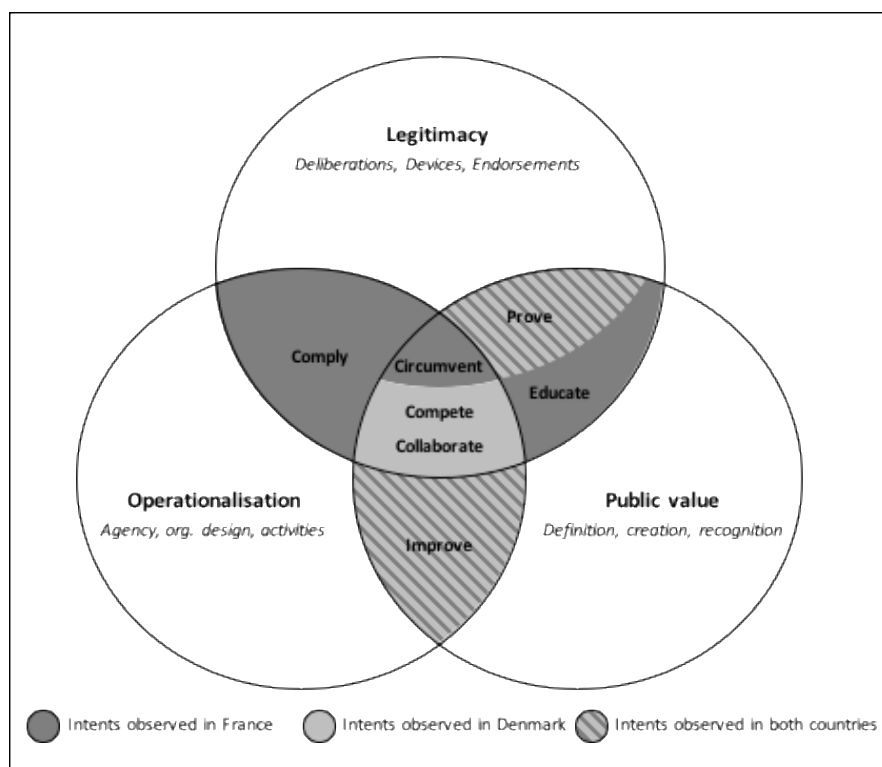


Figure 27: The seven strategic intents

Source: Author

¹⁶⁵ An additional step of our analysis consisted in generating correlation tables between the identified intents and the characteristics of the observed performance measurement models as detailed in Table 31, in the hope to identify performance traits pertaining to specific intents. However this initiative did not yield any meaningful results: no significant correlations could be established.

The two sub-sections below describe the identified seven strategic intents and highlight how they contribute to the choices made in relation to performance practices, in terms of indicators and processes.

10.3.2.1 Two strategic intents found in both France and Denmark

Prove:

Proving appears as an important motive for carrying out performance measurement, as WISEs are eager to inform their stakeholders about their impact. This intent is at the interstice between the social value and the legitimacy vertices, as WISEs seek recognition and acknowledgment on the basis of the value that they generate for the society. With this intent, performance measurement is therefore perceived as a legitimacy device.

We tie this intent to a form of communication (as the “*educate*” intent can be), which was found particularly important for WISEs in our Danish survey: “*generate awareness*” is the item that ranks first both in terms of expectations and in terms of experienced effect for performance measurement. In both countries, the “*prove*” intent appears rooted in large part in an environment where public budget tightening are creating an increasing demand for accountability from WISEs (see e.g. Borello & Barfety, 2018 for France or Hulgård & Andersen, 2015 for Denmark).

As we find most motives associated to this intent anchored in debates over the use of public funds, cost benefits approaches are privileged. Our two French cases use it, and DK2 is interested in such an approach. In France, FR1 is also very much geared towards using the quantitative output indicators it has developed over the years to prove its impact (while insisting on the qualitative information conveyed by these indicators). Furthermore, “*prove*” is the only intent where we find a degree of acceptability for methods oriented towards monetarization (although we have not observed them in practice), as they may help convey a cogent message under the form of simple yet compelling quantitative metrics.

Improve:

We find that both in France and in Denmark, WISEs use performance measurement with an intent to “*improve*” their impact. This strategy is nested in the overlap between the operational and the public value vertices as it aims at bettering the creation of social value by making changes in the operational environment, using performance measurement as a form of management tool. In our Danish survey, although this intent appears as clearly producing results for WISEs implementing SIA (ranked as the second most positive effect), it does not seem to be a particularly obvious motive for the ones who do not (ranks seven out of nine). Furthermore, we observe a surprisingly low ranking for the reported management-related benefits by KPMG in France (cf. Table 23).

In our cases, DK1 and FR1 are strong proponents of using their own SIA model with the Improve intent. We find a similar motive with FR2, which is characterised by the entanglement observed between SIA and quality processes. At DK2, management considerations associated to SIA were not expressed spontaneously at the executive level, but they were by the operations’ manager.

Approaches concerned with impact improvement appear to seek a mix of quantitative measures and qualitative evaluations, with an incline towards the prominence of the latter. Unsurprisingly, approaches oriented towards cost-benefit analysis are deemed of little relevance for this intent. In FR1’s case, we see this intent translates into the use of relatively “*actionable*” metrics that can be tied very precisely to the activity (i.e. mostly outputs).

10.3.2.2 Four idiosyncratic intents

Based only on four case studies (even completed with a degree of quantitative research), it would be unreasonable to assert that these strategic intents cannot be observed in a country other than the one we attribute it to. We however see a magnitude of alignment between the intent and the PAM environment for at least the first four of them. The last one, “*educate*”, seems less tied to the environment and more to the WISE’s own motive and the choice of a specific SIA approach.

Practices observed in Denmark:

In Denmark, we observed what we call a relationship-based quasi-market for the delivery of work integration services. This particular environment translates into strategic interactions that capitalise on all three vertices of the PVT triangle, as Danish WISEs have to secure legitimacy through both their operational (particularly through compelling activities) and public value (which must not only be actually generated, but also acknowledged) vertices in order to access beneficiaries.

Compete:

As WISEs evolve in a (relationship-based) quasi-market environment, we find, rather unsurprisingly, that one strategic intent behind the use of performance measurement (and the development of their own SIA) is the enhancement of their competitive position. We see that evidenced in our survey, but also in both of our case studies. This intent is characterised by an equal footing on all three vertices of the PVT triangle, where WISEs have to effectively generate public value through relevant and appealing (to the customers and beneficiaries) activities and capitalise on these two aspects to maintain their legitimacy as work integration services providers.

Generally speaking, this intent is more characterised by an aim at rendering the offer of services more compelling to public customers rather than to take contracts away from other WISEs (a behaviour that we have not observed during our research). This is however not true for employment services procured by the public sector, from which WISEs are actively trying to differentiate and where they see performance measurement as a particularly relevant way to do so (especially through cost-benefit approaches). While cost-benefit analysis finds its proponents to justify expenses towards the public, monetarization is generally not seen as a good way to approach performance for competitive purposes.

Cooperate:

The use of performance measurement models also shows that there is a clear intent of establishing more robust cooperation frameworks with municipalities, going well beyond a simple compliance to a demand for performance metrics, as evidenced in DK1's case. Here the performance model proposed by the WISE's SIA encompasses again all three vertices of the strategic triangle, as deliberation in the

legitimacy vertex is used primarily to co-define public value, as well as to co-create it through WISEs' operational vertex, in conjunction with the job centres' case workers.

As a result, WISEs like DK1 are expecting a more streamlined cooperation with job centres, with a SIA model that follows the structure and the requirements of the follow-up carried out by case workers. Due to the relatively new implementation of DK1's SIA, there was no return on experience at the time of the research considering this specific strategic intent.

Practices observed in France:

The mostly-planned nature of the NWS performance regime, and its perception by WISEs as a system anchored in a form of autocracy (cf. Chapter 8) favours performance measurement approaches based on the generation and maintenance of legitimacy. The central authority wants to ensure that WISEs comply with the rules that, it considers, should lead to the creation of public value. In turn, WISEs make sure they are able to meet these requirements by producing the necessary legitimacy.

Comply:

In our two French case studies, as well as through our complementary interviews carried out with national WISEs networks, we identified a pattern of disengaged compliance towards the devices that are deployed by the centralised state to measure WISEs' performance (*FR2: "This is why we don't care about the graduation rates"*). This strategic intent appears to be partly rooted in the WISEs' perception of facing an autocratic performance regime (Jany-Catrice, 2012), which we also tie to Gianfaldoni's idea of WISEs subordination to their public stakeholders (2018).

In this configuration, the targets negotiated with the Direccte become a means to an end: WISEs demonstrate through imposed devices that they comply in order to maintain their legitimacy endorsements (covenants), which in turn warrants the stability of their operational vertex. This is true even for WISEs which do not rely heavily on public subsidies since they need their covenants to channel beneficiaries. For this reason, we locate the compliance strategy at the junction between the operational and the legitimacy vertices: it is disconnected from the public value vertex (the intent itself,

not the actual results), even though the Direccte's performance targets are supposed to provide a representation of the public value creation expected from WISEs.

Circumvent:

Our two case studies show that WISEs can attempt to circumvent the performance metrics that are asked for them to observe. This intent can also echo Gianfaldoni's idea of subordination, where a subordinate attempts to circumvent the injunctions of a superior, (here the superior is the centralised state). We noticed this motive through the deliberation processes occurring during (essentially, but not only) the management dialogues, where WISEs get a chance to instate a conversation that can lead to the emergence of shared preferences and contribute to form alternative, local STAs that can compete (without being fully enacted) with the existing centralised one.

The aim of such a strategy is still to maintain legitimacy, but rather through deliberation than through institutional instruments. For this reason, we place this strategy more towards the centre of the PVT triangle as WISEs must hold a fact-based negotiation about the nature of their activities and how they believe their social value should be recognised. Both our cases demonstrate such an intent by making use of their own SIA model. In the case of FR2, we note the use of a cost-benefit approach for this purpose, even though the WISE articulates its SIA around the social utility framework.

Educate:

In the case of FR2, we identify a clear pedagogical intent attached to its SIA model, concerned with educating all their stakeholders (not only in the public spheres, and including their own staff) about the kind of impact that their activities have on the lives of their beneficiaries. For this purpose, FR2 has developed a comprehensive communication material that encompasses both quantitative and qualitative data.

FR2 is the only WISE we interviewed with a fully-fledged SIA model based on the social utility approach. The very inclusive and holistic nature of the social utility approach makes it a particularly well-fitted instrument for the pedagogical motives associated to FR2's intent. We furthermore observed that although cost-benefit metrics were part of the series of indicators used by FR2 in the context of its

social utility assessment, this particular set of indicators was left out of the material that is primarily used for educational purpose.

The “*educate*” motive could be seen as close to the “*prove*” intent we describe above because of its emphasis on communication. However we find “*prove*” to be much more about generating legitimacy whereas “*educate*” can also have a (marginal and indirect) component of public value creation (through the pedagogical nature of the generated knowledge). We therefore place the two intents at two different locations of our Venn diagram.

The elements above allow us to answer our second research question:

RQ2: What are the strategic intents of WISEs measuring their performance and what are the implications for the way this measurement is done?

We identified seven different strategic intents associated to performance measurement. Two of them (“*prove*” and “*improve*”) were observed in the two countries in which we carried our research, while the five others were specific to France (“*comply*”, “*circumvent*” and “*educate*”) or Denmark (“*compete*” and “*collaborate*”). In associating some intents more specifically to one country or the other (and therefore to one model of public value co-creation or the other), we show that the PAM environment influences WISEs’ strategic behaviours in relation to performance measurement. We also acknowledge that some strategies (“*prove*” and “*improve*”) can somewhat be considered as generic.

By highlighting the patterns of performance measurement attributes and variables mobilised through each strategic intent, we also provided a view of how these intents translate into concrete measurement practices (in terms of privileged approaches, metrics and processes). Having reached a relatively granular level of understanding of the strategic interactions driving performance measurement in France and in Denmark, we can now move on to our final research question.

10.4 The role of shared preferences and socio-technical arrangements

In the previous section, we characterised strategic interactions through their intents in terms of performance measurement. In this section, we build further on these conclusions, using the strategic intents as a charting tool to answer our last research question:

RQ 3: Are strategic interactions related to performance measurement conducive to the avoidance of performativity issues?

Our first analytical segment below establishes that the strategic interactions that occur in France and in Denmark between WISEs and their public stakeholders can translate in deliberation, which in turn can lead to the emergence of shared preferences. On the premises that the development of a SIA model typically increases deliberation, we establish that the extent to which this process leads to more shared preferences can be linked to WISEs' strategic intents.

In a second segment, as we associate performance measurement models, including SIA models, to a form of STA, we prolong this reflection by presenting the benefits and drawbacks that multiple competing STAs can have in terms of performativity.

10.4.1 Strategic intents and shared preferences

In this first part of our analysis, we highlight how deliberation occurs both in the French NWS and the Danish NPG (although through different mechanisms). As deliberation can turn into a "*public space*" in the sense of Desrosières (2012), we show that it contributes to the emergence of shared preferences.

We then focus more particularly on the role of performance measurement. Building on the observation that the development of a SIA model typically translates into increased deliberation, we enquire whether additional deliberation leads to further shared preferences. We find that a relation can be established and explained, at least in part, by WISEs' strategic intents.

10.4.1.1 Deliberation and shared preferences

With the central place occupied by public actors in WISEs' co-creation of public value (as providers of beneficiaries and of financial resources), we observed throughout our research that strategic interactions between them are numerous and diverse. The particular focus we brought on performance-related interactions found that this type of exchange, because of its nature related to the "*merit or worth*" (Scriven, 1991) of the activity, was especially conducive for deliberation.

We saw deliberation happening in France as well as in Denmark and in both places mostly at the local level. This deliberation however arises through different means. In Denmark, it is institutionalised by the relationship-based quasi-market configuration of work integration services delivery. It occurs naturally as co-production happens, and performance measurement is an open, non-normative conversation between two parties who entered a contract underpinned by shared values.

In France, deliberation benefits from a degree of institutionalisation through the place given to the management dialogue in the legalist, legitimacy-based system of covenants. But although management dialogues are taking place through legitimacy devices, on the basis of the normative performance metrics defined by the central state (primarily the graduation rates), we see that deliberation oversteps the boundaries of performance set by these indicators and ventures on more uncharted grounds where performance is debated in what, for the involved actors, resembles a "*public space*" (Desrosières, 1992).

In Denmark, deliberation is embedded in the "*compete*" and "*cooperate*" intents where WISEs seek to meet the expectations of their customers while co-creating public value in a relationship-based and high-trust environment. Deliberation is occurring throughout the process of co-creation, allowing for a rebasing of expectations if needed. However this process is not without tensions since we see that WISEs do not consider that municipalities always assess their performance on the basis upon which it should (i.e. they deplore a perceived focus on economy).

Despite these tensions (stemming from the centralised devices in France or from underlying market mechanisms in Denmark) the different mechanisms of deliberation lead in both countries to the

emergence of shared preferences, as WISEs and their public partners debate how performance should, or could be measured (e.g. Direccte to FR2: *“Yes, we agree [about the value created by FR2], but I am asking the question, how do we valorise this?”*)

10.4.1.2 The role of SIA in the emergence of shared preferences

Although not a systematic outcome, we observed recurrently that the deliberation processes we describe above can lead to the emergence of shared preferences between actors. We observed that particularly in France where actors are together rejecting the centralised devices (the *“circumvent”* intent). For instance, FR2’s CEO during a meeting with the municipality council and the Direccte says: *Municipality Council representative: “When elected representatives take a look at the modulation¹⁶⁶ reports, they go ‘ho but there has only been few graduations’, and then we have to explain: the soft skills, long term employment, etc. all this, this is not quantifiable. And me, I am trying to show our representatives how WISEs work (...).”*

DK2’s CEO: About the modulation... I am glad to hear you say that!”

This brings us to the role of SIA. We observed that SIA, or the use of a performance measurement model proposed at the initiative of the WISEs, is injecting a degree of deliberation over and above what is mandated through the existing performance measurement mechanisms, in both of the studied co-creation contexts.

We distinguish two means by which SIA contributes to deliberation: *ex-ante* and *ex-post* SIA model design. *Ex-ante* deliberation occurs in the SIA’s design phase where stakeholders are involved in the architecture of the processes and indicators. This is conform to what is advocated by the literature (see e.g. Clark et al., 2004; Grieco, 2015; S. Olsen & Galimidi, 2008) and ascribes to a process of **co-definition** of public value. *Ex-post* consists in using the SIA’s results to inform deliberation, regardless of stakeholders’ involvement in the design (and therefore co-definition) phase of the model. This ascribes to a process of **recognition** of public value.

¹⁶⁶ The term *“modulation”* refers to the variable part of the subsidies (maximum 10%) that is allocated to WISEs based on the graduation rate targets they are meant to achieve (cf. Table 20)

On these premises, an interesting question to ask is therefore, how much of the additional deliberation attributed to SIA translates into extra shared preferences? Whether the deliberation occurs *ex-ante* or *ex-post* (or both) has of course a strong incidence on the degree to which shared preferences appear: unsurprisingly, we find that *ex-ante* deliberation has a more beneficial effect on shared preferences, observed at both FR2 and DK1 (in the former it is however hampered by the NWS bureaucratic game of musical chairs, which makes it difficult to maintain interactions with the individuals having contributed to the model's design). *Ex-post* deliberation can participate to shared preferences as well, but the information presented by the SIA first has to be accepted as valid by the stakeholders. In the case of FR1, with a planned SIA design where deliberation occurs *ex-post*, we find little to no effect in terms of additional shared preferences.

We suggest that an explanatory answer to this observation is best found by looking at strategic intents. In our case studies, we find that three strategic intents contribute the most to the emergence of shared preferences: “*compete*” and “*collaborate*” for DK1 and “*circumvent*” for FR2. These are the two WISEs that have used *ex-ante* deliberation, involving stakeholders in a co-definition of the public value they co-create. These intents, which aim at providing alternative frameworks for the acknowledgment of public value, act as a catalyst to make sure that public stakeholders' expectations are taken into account in the SIA developed by WISEs. However in the case of FR1, mainly focused on “*improve*”, with a planned SIA design where deliberation occurs *ex-post*, we find little effect in terms of additional shared preferences.

Overall, we observed that although SIA does not take precedence on the existing performance measurement processes, it is conducive to an **acknowledgment** of value creation attributed to the WISE that is broader than what is encompassed through the existing performance measurement instruments. These findings apply equally in the different NWS and NPG environments.

10.4.2 The role of socio-technical arrangements

In this last segment of our analysis, we stress how deliberation and ensuing shared preferences can reinforce STAs. Although enacted STAs are a source of performativity (Callon, 2006), we show that competing, unrealised STAs have the benefit of actually preventing some forms of performativity such

as indicator-substitution. In such a configuration, various STAs can foster a debate in the public space, by proposing alternative performance measurement methods to the one currently enjoying prominence.

We however show that such a setup is not easy to maintain, and that emerging STAs can also be a source of disruption. Moreover we observe that even maintained in a balanced *competition*, STAs have limited effects on the avoidance of retroaction. Lastly, we also see that STAs are vulnerable to the PAM environment, where in France, centralisation prevents the spread of arrangements emerging locally.

10.4.2.1 Shared preferences and the strengthening of socio-technical arrangements

Following Callon's considerations, we determined that performance measurement instruments we observe for WISEs can be considered as STAs, largely unrealised (i.e. not fully performative). We include the SIA initiatives we describe above as part of a universe of STAs created to measure the performance of WISEs.

We observed during our research that the emergence of shared preferences was conducive to the strengthening of STAs. As actors align their views, through deliberative processes, on what is the social value that should be co-created out of their activities, they tend to propose, develop and adopt their own framework for the measurement of performance. In the case of FR2 and DK1, *ex-ante* deliberation with their stakeholders has allowed to discuss which indicators the WISE should use to gauge its impact, which translated into a mix of quantitative, output-oriented metrics combined with competency evaluation, satisfaction surveys and a collection of testimonies.

We saw that these STAs arise through different mechanisms in France and in Denmark. In France, they tend to appear as a **resistance** against the centralised performance measurement system, in order to "*circumvent*" it. In Denmark, they tend to manifest to **fill the apparent void** left by the absence of a common reference frame for performance measurement. We then see through both our Danish case studies that new STAs such as SIA become instruments to "*compete*", a vector of differentiation in an environment where there is no clearly established way to measure performance.

Nevertheless, and just like shared preferences in the previous sub-section, the STAs thereby created do not appear to take precedence over the existing institutional performance measurement mechanisms (even if very informal like in Denmark). By going beyond shared preferences and actually proposing their own measurements, STAs however foster the deliberation about both what is the public value that WISEs co-create with their public stakeholders (co-definition) and how performance should be apprehended out of that co-creation process (acknowledgement).

Therefore, even if what stems out of such initiatives does not manage to formally alter the way performance measurement is carried out, it has the benefit of proposing **competing** STAs, thereby preventing the (full) enactment of a dominant one. In France, this has the considerable advantage of creating (at least at the local level) views alternative to the ones promoted by the centralised devices (based on graduation rates and two other ratios).

To this extent, Callon talks about the competing existence of STAs as “the struggle between worlds that are trying to prevail” (2006, 28). In our case, and in the light of the considerations we bring forth on the topic of performativity in Chapter 3, the main benefit we see associated to that struggle, is the quasi-absence of visible **indicator-substitution** issue. As stakeholders deliberate to offer their views on performance, at times managing to surface some shared preferences, and eventually a form of STA, they avoid the pitfall of misrepresenting reality through the dominant arrangement. As the Direccte puts it: *“But everything can be explained, they (WISEs) can explain to us the reasons why they did not achieve their targets”*.

10.4.2.2 The limits of socio-technical arrangements

At the local level, the observations above appear true in both countries on the basis of our findings. In France, the situation is more contrasted at the central level, where emerging STAs among local stakeholders seem to not circulate well towards the upper levels of authority. This appears to create pockets in the higher levels of the administration where indicator-substitution is likely to occur. The Direccte tells us: *“We have our own tools [indicators] and we have the targets. And good or bad, we cannot contest them anyway”*.

Moreover, we also see in our research that shared preferences or competing STAs are not enough to prevent a form of **retroaction** to occur. We see that particularly through the intricacies of the French NWS bureaucracy and its tendency to work in silos, where a multitude of stakeholders are all proposing their own targets (e.g. the county authority targets individuals benefiting from minimum income welfare, the municipal authority targets specific districts, etc.). As a result, WISEs end up chasing these targets to preserve their legitimacy (as expressed in the “*comply*” strategy), which can represent a substantial opportunity cost.

Besides, an obvious downside to the development of competing STAs is the risk for one of them (even coming from a WISE) to become enacted, which is where Callon places the trigger of performativity: *“We can agree to call performance the process whereby STAs are enacted, to constitute so many ecological niches within and between which statements and models circulate and are true or at least enjoy a high degree of verisimilitude. This constantly renewed process of performance encompasses expression, self-fulfilling prophecies, prescription and performance”* (2006, 25).

We observe this risk even in an environment like Denmark where shared preferences emerge easily as part of the interactions between WISEs and their public stakeholders. We see for instance in the case of DK1 that the model developed spreads to new municipalities, and eventually to other WISEs as DK1’s management is considering an eventual commercialisation of the approach and essentially think of its SIA as a competitive asset.

In France, we describe in Chapter 8 two competing grand STAs on the performance of SEs: one promoting social impact assessment and the other social utility evaluation. We observed that the “*social impact*” approach has emerged recently and is quickly taking precedence in the public spheres (see e.g. Borello & Barfety, 2018). This creates a sense of urgency among some WISEs (we observed this in the Chantier Ecole network) where they see this change of STA as a paradigm shift in their operations. As some other WISEs (particularly within the EI covenant) are at ease with the “*social impact*” STA, we argue that preserving both approaches on an equal footing would foster the public conversation on the conventions to be attached to WISEs’ performance, while allowing actors to position themselves in the framework that aligns best with their values.

10.4.2.3 The intensity of socio-technical arrangements

Callon (2006) proposes a distinction between enacted and unrealised STAs. The former designates those that have become performative, while the latter are “*competing*” ones, struggling to become enacted. Although a cornerstone of our analysis, this distinction does not offer the granularity we need to convey what we observed in our research. In order to better interpret our results on the basis of the observations above, we want to introduce an additional distinction between what we call *low intensity* and *high intensity* STAs.

Low intensity STAs are formed on the basis of shared preferences, when actors co-develop a collective understanding of performance out of their common preferences. We observe such an arrangement between for instance FR1 and its main interlocutor at the Direccte, where performance is understood to go above and beyond the graduation rates tracked by public actors (e.g. by exposing beneficiaries to a work environment for a prolonged period). A low intensity STA does not need to go any further than this stage of common understanding, i.e. it does not translate into indicators or measurement methods, even informal (there is nothing of the sort between FR1 and the Direccte, which actually tends to refuse to use FR1’s own SIA as a performance measurement while admitting that its own performance measurement approach is ill-fitted).

These are the additional steps that form what we call a **high intensity** STA. This is for instance what we observe in Denmark with DK1, where the WISE proposes an established framework for measuring its performance, under the form of a SIA, which is adopted to an extent by its stakeholders. As high intensity STAs propose a frame for performance measurement (even rudimentary), they can (although not necessarily) convey a degree of performativity, through retroaction. It is what we observe in our research with FR1 and FR2 actively seeking to meet some of their stakeholders’ targets (e.g. inhabitants of “*priority districts*”). Such STAs are however not *fully* performative (i.e. in the sense of Callon’s enacted arrangements), in which case they would also offer ground for indicator-substitution (but we have not observed such occurrences).

By introducing this additional distinction, we effectively differentiate between three stages of STAs that are illustrated in Table 33. In this table, indicator-substitution appears as an exclusive feature of enacted STAs (as their metrics become normative), but retroaction can be observed in unrealised arrangements.

Table 33: Three states of socio-technical arrangements

Other STAs	Many	Few	None
Intensity	Low	High	High
Status	Unrealised	Unrealised	Enacted
Characteristics	Cluster of ideas	Ideas and metrics	Normative ideas and metrics
Performance	None	Possibility of retroaction	Retroaction and indicator-substitution

Source: Author

On the basis of our findings, and on the premises that, as suggested by Callon, only enacted STAs are fully performative, we argue that competing STAs can be beneficiary to the preservation of non-performative views. They favour a form of deliberation and debates that are advocated by public value theorists. We also evidence that the best stage for deliberation, i.e. the one where shared preferences are most likely to emerge is *ex-ante* to the formation of a STA such as SIA. We show that performance-related strategic interactions are a good vector for that kind of deliberation, and that they can favour democratic spaces (reminiscent to Desrosières’ public space) where WISEs and their stakeholders co-define their public value (debate of indicators’ conventions). We can therefore answer **RQ3**.

RQ 3: Are strategic interactions related to performance measurement conducive to the avoidance of performativity issues?

In this section, we show that strategic interactions related to performance measurement allow to limit performativity through the mechanisms of increased deliberation, emerged shared preferences and an eventual formation of unrealised STAs, at least of low intensity. We find that “*circumvent*” (in France), “*compete*” and “*cooperate*” (in Denmark) are particularly conducive for that kind of processes to happen.

As long as STAs remain unrealised, these processes contribute to the decrease of indicator-substitution and to the creation of public spaces where indicators' conventions can be deliberated. These processes however do not solve all of performativity issues, since they cannot fully prevent retroaction, whereby performance indicators are still pursued to reach legitimacy, even if they are not perceived as a valid representation of public value. This is what we observe in the “*comply*” strategy in France. Similarly, a downside of the “*compete*” strategy in Denmark can be that a developed STA fills the void allowed by decentralisation and reaches enactment.

Conclusion of Chapter 10

This chapter brought forth the elements to elucidate our research question: how strategic interactions between WISEs and public actors are contributing to frame the measurement of performance in different PAM contexts? Using PVT's strategic triangle, our research sought to shed light on this phenomenon through the prism of the strategic interactions between WISEs and their public stakeholders, looking for implications in terms of performativity. Following our critical realist (CR) paradigm, we identify two causal mechanisms (CMs) that participate to the structuring of strategic interactions and which can be tied to manifestations of performativity. We illustrate them in Figure 28.

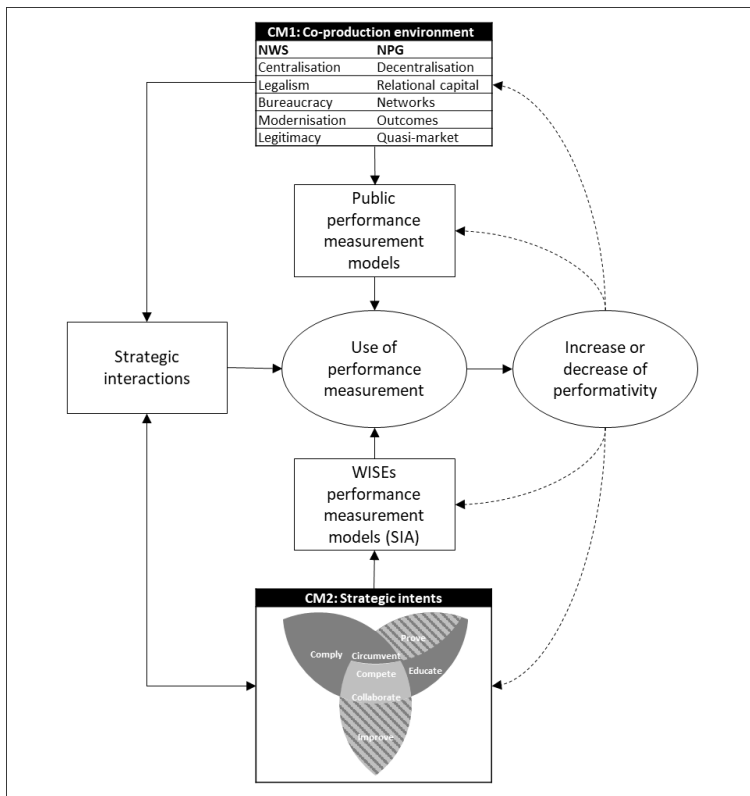


Figure 28: Two explanatory factors for performativity

Source: Author

We started our reflection at the macro level, showing that the frameworks for public-value creation induced by PAM models translate into different performance regimes (**RQ1**): public actors showed **deliberative** approaches in Denmark, while France is stuck somewhere between **autocracy**, and **managerialism**, with a hint of deliberation. Although the performance regimes provide us with the tool to organise our empirical knowledge, the underlying mechanisms we identified (e.g. centralisation and legalism for France or the relationship-based quasi-market for Denmark) allow us to identify a first cluster of explanatory factors from the deep real (the structures and rules of different PAM models) to interpret strategic interactions.

As we are interested in the potential changes occurring in performance measurement models related to the development of SIA, we moved to the micro level and sought to understand the strategic interactions attributed to performance measurement. The first step we took towards this outcome was the identification of the **strategic intents** underpinning these interactions (**RQ2**): using our PVT analytical framework and its nine entries, we identified seven intents: **“compete”** and **“cooperate”** in

Denmark, “*comply*”, “*circumvent*”, and “*educate*” in France and “*prove*” and “*improve*” in both countries. These intents provide us with a framework to better comprehend the diffuse nature of strategic interactions.

Lastly, since we deemed performativity as a particularly relevant aspect of performance measurement to investigate, we established whether some types of **strategic interactions** can be more or less conducive to performativity (**RQ3**). Using the strategic intents we identified in **RQ2** to guide our reflection, we saw that intents like “*compete*”, “*collaborate*” or “*circumvent*” are more likely to favour the emergence of **competing, unrealised STAs** and that these can foster the development of a public space to discuss indicators’ conventions. By way of consequence, we showed they can serve to avoid **indicator-substitution issues**, which in France is limited by the NWS’ centralisation. We also observed that competing STAs do not prevent **retroaction**, especially in cases where the “*comply*” intent is observed. The strategic intents therefore provide us with an additional cluster of explanatory factors for how strategic interactions lead to performance measurement models that can convey or hinder performativity. These factors can only be fully comprehended if put in perspective by the co-production environment in which they are found.

These findings, and the processes leading to them, have driven us to formulate a number of scientific and empirical contributions as we developed our research. We expose them in the following and last chapter of our thesis.

Chapter 11 General discussion

Having provided the answers to our research questions in the previous chapter, this last segment of our thesis revisits some of the concepts we introduced in our literature review and shows how our research builds upon them to introduce new scientific and empirical insight. In this chapter, we show that the combined use of our theoretical entries, together with the use of the critical realist (CR) research paradigm (Bhaskar, 1975; Lawson, 2006) yields novel ideas to address questions such as: how to understand the effects of performativity? What contributes to their limitation? And what can be done to adopt performance measurement models that avoid these effects?

One of these novelties is the contribution we bring to the understanding of the notion of performativity (the influence that performance instruments can have on what they *measure*). We focus on socio-technical arrangements (STAs, Callon, 2006) for which we conceptualise **feedback mechanisms** and show that, for work integration social enterprises (WISEs)' performance measurement, they can actually be capitalised upon to tame performativity. This is an original take on STAs, with the potential to hinder performance measurement's adverse effects such as sub-optimal outcomes (through poorly defined indicators) and detrimental policy choices (through opaque conventions). Another entry we emphasise to limit the negative externalities of performance measurement is the **deliberative design** promoted by most social impact assessment models (SIA, as a distinct practice of performance measurement undertaken by WISEs).

We also show that the original use we make of public value theory (PVT, Moore, 1995) as our main analytical framework to understand the strategic interactions (the aspects of the relationships pertaining to the creation of public value) between (WISEs) and their public stakeholders yields meaningful results that allow us to shed a new light on performance measurement through the understanding of its **strategic intents** (i.e. why do WISEs measure their performance the way they do it?).

As we deliver these insights, we organise this chapter in four sections. We start by providing our theoretical, methodological and empirical contributions, before highlighting some of the limits of our thesis and suggesting future research alleys.

11.1 Theoretical contributions

This section exposes the theoretical contributions of our thesis. We start by highlighting how our approach of performativity, differentiating STAs from retroaction and indicator-substitution replacement, brings about a more granular understanding of the notion. Doing so, we highlight the role of **STAs' feedbacks**, which we identify as either pervasive (i.e. STAs are vectors of performativity) or virtuous (STAs contribute to the avoidance or the limitation of performativity). This last stance is unobserved in the literature and goes against most conceptions of the phenomenon of STAs.

We pursue the restitution of our theoretical contributions by emphasising the different aspects of our contribution to PVT. As our research proposes additional entries to the strategic triangle, it also expands its use to new actors (WISEs) and practices (performance measurement) while presenting novel empirical entry to PVT. As we recapitulate these contributions, we recall how they are all aligned with the most recent research agendas proposed for PVT. We then go over the seven strategic intents we identified for performance measurement. With our CR objective in mind of proposing causal mechanisms, we recall how these intents tie into WISE's local environment (mapped mostly through PAM models) and how they can influence performativity. Lastly, we highlight how our research finds relevance in the broader discussion pertaining to the creation of performance indicators.

11.1.1 Refining and operationalising the understanding of performativity

11.1.1.1 Identifying three entries to explain and operationalise performativity

Performativity as we first outline it in our thesis (the influence that performance instruments can have on what they *measure*) may seem like a daunting notion to apprehend. One contribution of our research is to propose an original way to frame this concept through three distinct components. By distinguishing between STAs, retroaction and indicator substitution, we allow future research work to explore the problem of performativity from a more structured angle. This approach can also be used for pedagogical purposes to break down a notion that can be challenging to frame as a whole.

While STAs and retroaction are well identified and explored in the literature (see e.g. Callon, 2006 and 2009 for the former and Desrosières, 2015 for the latter), indicator-substitution is part of our

contribution to a refined definition of performativity. It is an idea that is already well represented in the literature on performativity (see e.g. Desrosières, 2006), but we find that identifying it as a distinct observable feature helps refine our view of performativity.

For instance, while we observe the formation of STAs (e.g. the different SIA models developed by WISEs, or the performance measurement used by the Direccte) and a degree of retroaction in several of our case studies (e.g. FR1 stretching resources to work with beneficiaries from specific districts, or DK2 acting on prices to the detriment of its staffing), we do not observe cases of indicator substitution (where e.g. the graduation rates in France would be considered as an accurate proxy for the public value created). This distinction therefore allows to present a more granular analysis of performativity: yes, FR1 focuses part of its resources to “tick the boxes” of the NWS bureaucracy (occurrence of retroaction as a form of performativity), but its stakeholders know that the public value FR1 generates is not captured by the metrics it uses (the indicator-substitution aspect of performativity is avoided). These considerations permit to shed light on the role of mechanisms such as deliberation: by allowing to discuss performance by other means than through devices, deliberation is the main driver behind the avoidance of indicator substitution issues.

11.1.1.2 A better understanding of the role of socio-technical arrangements

While Callon has written at length on the role of STA in economics, to our knowledge, our work is the contribution in the field of economics that empirically researches the role of STAs on performance measurement. As such, our thesis proposes a window for seeing and understanding how STAs emerge and operate through strategic interactions between field operators.

We observe that as WISEs and their public stakeholders interact about performance issues, they form their own understanding of what is the performance of a WISE and how it should be measured. This is most often an unformal development that is part of the normal process attached to the co-creation of public value (i.e. the regular interactions between public actors and WISEs occurring through the delivery of employment services).

We show that this process occurs in both of the PAM contexts we research. It originates from different mechanisms in the two countries we surveyed, such as the management dialogues in France's NWS or an ongoing co-definition of public value in Denmark's NPG. But deliberation and the ensuing emergence of shared preferences stemming from these mechanisms are the common processes that lead to the creation of STAs (such as FR2's SIA model which is informed by deliberation with local public actors) or to the reinforcement of existing ones (such as the tacit understanding between FR1 and the Direccte that the WISE generates value through many other aspects than the ones captured by the graduation rates).

This leads us to another of our contribution, which is a refined delineation of STAs' development stages. When introducing the concept, Callon splits STAs between unrealised and enacted ones. We recall that the former are STAs that have not yet become performative while the latter are a source of performativity. In our research, we identify an additional level of granularity since we propose to distinguish between **low intensity** STAs and **high intensity** STAs (cf. Table 33).

In the examples we just provided, the tacit agreement between FR1 and the Direccte is a low intensity STA: both stakeholders have shared preferences that inform their view of performance (e.g. FR1 creates value by upgrading the soft skills of increasingly marginalised individuals), but they do not operationalise them in the form of a common performance measurement model: the Direccte's instruments fail to do so and public actors take FR1's SIA results as inconclusive. Concurrently, FR2's SIA is a high intensity STA since it is built on shared preferences (agreed upon ideas such as e.g. FR2 contributes to the development of the local economy) and propose indicators to capture these preferences (corporate expenses carried out locally, *inter alia*). We operationalised this view of STAs in Table 33, showing the three stages of STAs, which use to build Figure 29 below.

As we refine our understanding of how STAs emerge and gain intensity until they eventually become enacted, we also highlight the benefits associated to the co-existence of several STAs. This is a key contribution on the topic of performativity since until now, because of their contribution to performativity, STAs were considered rather undesirable (cf. e.g. Bracking, 2012, 2015 or Pakh, 2017).

While Callon’s seminal paper (2006) does indeed propose situations where STAs compete, we go a step further by suggesting that this co-existence can work towards preventing performativity.

In France, this is evidenced by the way the STAs that form over time between WISEs and the Direccte (even the low intensity ones) are used to rationalise deviations from devices designed by the central authority to measure performance (the graduation rates, which is the competing STA) and prevent them from becoming performative. In Denmark, we see an opposite situation where no pre-existing STAs leave way to the emergence of arrangements displaying a potential for enactment (such as the SIA proposed by DK1) or to non-concerted approaches such as the one adopted by DK2PS.

11.1.1.3 Three distinct feedbacks of socio-technical arrangements

We build upon these observations to propose a simplified model for the different kind of **feedback mechanisms** we attribute to STAs. In Figure 29, we identify and label these feedbacks. Dashed lines are said feedbacks, while solid lines indicate how interactions lead to the formation of STAs. Label (a) shows the kind of virtuous feedback that can emerge from multiple, unrealised STAs (regardless of their intensity). These feedbacks tie directly into deliberation mechanisms, feeding debates and discussions between stakeholders. Type-a feedback is the virtuous trait we lend to STAs.

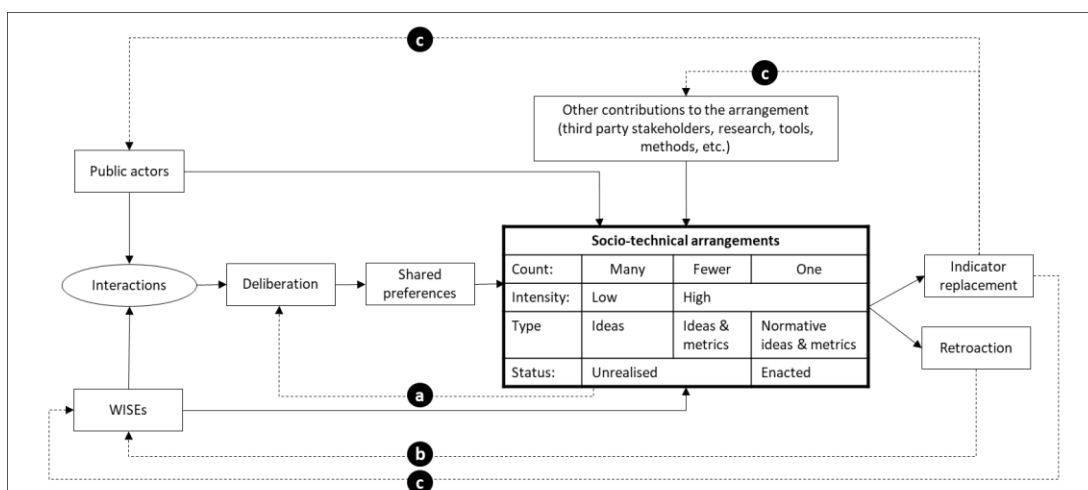


Figure 29: Feedbacks of socio-technical arrangements

Source: Author

Labels (b) and (c) identify feedbacks that contribute to performativity. Both tie into stakeholders directly rather than the deliberation occurring between them. Type-b feedback is the result of retroaction. We observe that STAs do not need to be “fully” enacted to produce that kind of feedback. However they need to be of high intensity since they require a formalised performance measurement model with the indicators necessary to generate a degree of retroaction. The recurring example in our analysis for this type of feedback is the indicators used by public stakeholders in France that propose “targets” (e.g. isolated women, Romani, priority district residents, etc.) to WISEs which do not necessarily align well with their operational model.

Type-c feedback stems from indicator-substitution and tie into all stakeholders. We did not observe true indicator-substitution in our research, and based on works such as Callon (2006, 2009) and the three states of STAs that we propose in Table 33, such feedback seems only to occur from “fully” enacted STAs, (i.e. enjoying a monopoly, at least temporarily). Type-b and Type-c feedbacks are the pervasive effects of STAs that Callon stresses and that are retained by most contributions citing his work (e.g. Bracking, 2015 or Desrosières, 2015).

As Callon emphasises the position of STAs as matrices for performativity, a conclusion from his considerations could be to say that one should avoid STAs. The conclusion that emerges from our research would rather be to say that quite the contrary, STAs should exist, and they should exist in numbers, precisely to avoid performativity. This is particularly evidenced in France, where both low intensity STAs (such as the tacit understanding between FR1 and the Direccte) and high intensity STAs (such as FR2’s SIA) foster a form of deliberation and avoid the Direccte’s graduation rates to become performative indicators. In a way, we expand on Callon’s notion of competing STAs, by specifically identifying the virtues of this coexistence (or “struggle”, as Callon puts it).

11.1.2 Contributions to public value theory

From our initial question of how the development of SIA may influence WISEs performance measurement as undertaken by public actors, we focused our interest on strategic interactions between WISEs and their public stakeholders as the nodal point where this influence could be observed. To make sense of our observations, we chose PVT as our analytical framework because of its

focus on strategy, its embeddedness in PAM issues relevant to our research (public value creation, of which the social value generated by WISEs is a subset) and its originality as a theoretical instrument to research WISEs.

Below, we present our contributions to PVT, which we structure around some of the research axes identified by Hartley et al. (2017, about a refining of the concepts used in PVT) and Bryson et al. (2017, about an expanded use of the PVT triangle). A last sub-section focuses on how our work is relevant as an empirical piece of PVT research.

11.1.2.1 Refining the conceptualisation of public value theory

Hartley et al. (2017) suggest that one interesting research alley for PVT is the deepening of the strategic triangle. This is what we do in Chapter 2 by proposing three distinct entries for each of the triangle's vertex. In the public value vertex, we identify (i) (co-)definition, (ii) (co-)creation and (iii) acknowledgement of public value. In the legitimacy vertex, we separate the legitimacy acquired through (i) deliberation, (ii) what we call legitimacy devices (such as SIA) and (iii) forms of endorsements (France presents a potent form of institutional endorsement). Lastly, in the operationalisation vertex, we separate issues pertaining to (i) organisational design, (ii) the organisation's activities and (iii) agency (the forms of cooperation or competition with other stakeholders to secure the resources necessary to operate).

In our research, a key benefit of this refinement is that it allows us to use the concepts proposed by PVT more effectively by (i) segmenting the information that each vertex should capture and (ii) assisting us in scanning for the relevant data in our case studies. The entries we propose were particularly useful in fine-tuning our understanding of the interactions we observed between WISEs and their public stakeholders. We thereby find that the refinement of the triangle alleviates some of the critics addressed to PVT, formulated around its lack of clarity (Rhodes & Wanna, 2007).

11.1.2.2 Expanding the use of the strategic triangle

Bryson et al. (2017) encourage the development of a multi-actor PVT theory and advocate for the use of the strategic triangle on other stakeholders than public managers. A similar angle of PVT extension

to a broader array of actors and concepts is deemed a relevant contribution by Hartley et al. (2017), and even by Moore and Khagram (2004), who emphasise the pertinence of the PVT framework for actors from the private sector contributing to the creation of public value. By its focus on WISEs and performance measurement using a PVT lens, our research is a clear contribution to the agenda promoted by these authors.

As we extend the field of PVT analysis to new horizons, our contribution can be placed along the line of works from e.g. (Page, Stone, Bryson, & Crosby, 2015) or (Cook & Harrison, 2015). While these papers propose an application of PVT to respectively network management and IT planning, we leverage PVT in order to better comprehend the measurement of WISEs' performance. Furthermore, the performance measurement tenet of our work can be paralleled to other contributions such as Moore's (2014) on public value accounting or Wang and Christensen's (2017) on public value measurement. From these two papers, our contribution differs by its level of focus level (micro, versus macro) and its stakeholders (WISEs rather than governments).

Our research particularly follows Bryson et al.'s guidance, who stress the variety of *inter alia*, actors and practices that can be "*placed at the centre of the triangle*". By the centre of the triangle, the authors refer to the area of focus the PVT analysis should be placed on. As we position WISEs and performance measurement at the *centre* of our triangle, our research is a pioneering work on WISEs and their performance using PVT as an analytical framework. Furthermore, our thesis confirms that PVT analysis yields meaningful results outside a pure focus on public managers (chief among which are the seven strategic intents we discuss as a distinctive contribution in section 11.1.3.2, but also the importance of deliberation to avoid issues pertaining to performativity)

11.1.2.3 An empirical research making use of public value theory

Lastly for PVT, our thesis is also a contribution towards resorbing the lack of empirical research deployed by e.g. Bryson, et al. (2015a); Hartley et al. (2017) or Williams & Shearer (2011). As we mobilise the PVT triangle in a way that extends from its original design (by providing additional entries to the triangle and by focusing on WISEs and their performance) we operationalise it concretely to derive results from both quantitative and qualitative field data.

Doing so, our work follows the path of contributions such as De Jong et al. (2017), encouraging the use of the strategic triangle as a tool, that we utilise to make sense of our case studies and our survey. Importantly, we also contribute to bring an answer to the questions asked by e.g. (Stoker, 2006) or (Dahl & Soss, 2014) by specifying further what PVT can be used for and showing how its tools can be mobilised in research.

11.1.3 Decrypting the intents of performance measurement

11.1.3.1 Different types of strategies for different PAM environments

As a tool, PVT's triangle has allowed us to chart the strategic interactions we observed between WISEs and their public stakeholders. This brings us to a first contribution we highlight in regard to the way WISEs co-create public value. We evidence that while pursuing similar goals underpinned by similar values, they operate differently in different PAM contexts. More crucially, we are able to characterise these differences through the use of the strategic triangle.

Our contribution here is to show that NWS mechanisms (particularly legalism) tend to displace WISEs' strategies towards the legitimacy vertex, while in the NPG environment, strategies are more balanced across the triangle to answer all-encompassing market demands. In the former, WISEs use an amount of operational resources to wrestle regulation in order to maintain what Moore (1995) calls their authorising environment (or their legitimacy). In the latter, customer satisfaction is paramount, and while relational capital helps, WISEs have to deliver public value and demonstrate an ability to respond to price pressures.

By describing these mechanisms, we build upon the literature depicting how public value can be co-created by different actors (e.g. Bovaird, 2006; Bovaird et al., 2016; Pestoff et al., 2013). By characterising them with the use of the PVT, we bring the distinctive contribution of spelling out these processes specifically for WISEs and showing how they are constrained by the PAM environment in which they operate. Doing so, we bring an answer to Battilana et al.: "*Subsequent studies of WISEs will need to explore the role of public authorities across different contexts and, more generally, the heterogeneity between different institutional contexts*" (2015, 1680). While being a distinctive

contribution on its own, uncovering these fundamentals strategic drivers is also needed to proceed to our topic of performance measurement.

11.1.3.2 Seven strategic intents tied to performance measurement

A key contribution of our work is the seven intents that emerge from the coding of our data in NVivo and that we associate to performance measurement in the context of the strategic interactions described through the PVT triangle: “*prove*”, “*improve*” (both observed in PAM and NPG), “*compete*”, “*cooperate*” (observed in NPG) and “*educate*”, “*comply*” and “*circumvent*” (observed in NWS). With our focus on WISEs, this represents a relevant contribution to the SIA literature (i.e. the performance measurement approaches specifically focused on mission-driven actors such as WISEs) such as Grieco (2015); Maas & Liket (2011) or Olsen and Galimidi, (2008). The strategic intent can bring an additional attribute to the ones used in these publications to classify SIA models and can also be used to narrow down the “*purpose*” attribute (i.e. the attribute specifying to which end SIA is used).

Linking the identified intents to areas of the PVT triangle allows to provide a clear characterisation of these motives, which makes our contribution more accessible and more actionable for future research. More fundamentally, it also shows that the issue of WISE’s performance measurement can raise specific questions when it is brought in a public context (e.g. the Danish conundrum of cooperating while competing with public partners and the French focus on maintaining institutional endorsements). These questions are left aside when considering performance measurement models from e.g. impact investors.

As we refine our understanding of performativity through the three components that we recall above (STAs, retroaction, and indicator-substitution), we are also able to see how these intents influence the occurrence of performativity. We for instance see that the circumvent, compete and cooperate intents are more likely to generate deliberation, which in turn will limit performative effects of the measurement models at work. On these premises, strategic intents are instrumental in answering our research question and showing the role that strategic interactions have in either shaping (especially in the Danish NPG) or challenging (especially in the French NWS) measurement performance models.

Pursuing this line of thought, and using our CR lens, the strategic intents appear as possible **causal mechanism** for the apparition of performativity.

11.1.4 Contributing to the debate on performance indicators

A last contribution of our thesis is towards the academic discussion about the design and the creation of performance indicators that we describe in Chapter 1. We highlighted that critical issues lie around the clarity of conventions (i.e. the assumptions pertaining to the scope and arithmetic of indicators): poor conventions can lead to suboptimal outcomes (e.g. putting individuals into part-time, low quality jobs instead of equipping them with the skills to secure better opportunities, e.g. Salais, 2010) while murky ones can lead to detrimental policy choices (e.g. concluding that a programme fails to deliver on its objectives because unsatisfactory performance indicators hide a more complex reality, e.g. Desrosières, 2007).

To the question of how future indicators should be developed to better capture progress, our research shows the critical role of deliberation in the design of indicators. This is a fact already highlighted in existing publications (e.g. Desrosières, 2012; Gadrey & Jany-Catrice, 2016; Jany-Catrice, 2008). Our thesis builds upon these observations by describing the work we carry out on performativity and by bringing additional empirical evidence (echoing the ones brought forth by Jany-Catrice, 2008) of the role of deliberation in shaping indicators meaningful and useful for measurers and measurees.

Looking at our research and the practices we identified among WISEs, we show that the development of SIA can be especially conducive to processes of deliberation and democratic performance indicator design. This builds upon practitioners' advice (S. Olsen & Galimidi, 2008; Tuan, 2008) and academic observations (Dufour, 2018) and we observe that DK1 in Denmark and FR2 in France have carried out *ex-ante* deliberation for their SIA models. As such, we suggest that SIA practices can be an inspiration for the research on progress indicators undertaken by initiatives such as e.g. the OECD's Better Life Index or the 2007 Istanbul Declaration, and for the identification of "a new system of values and a new hierarchy of the 'items that count'" (Gadrey & Jany-Catrice, 2016, 7)¹⁶. We also bring a limit to this observation, as we find that empirically, when operationalised, all SIA models do not observe the

inclusive design practices advocated by their champions. This is evidenced by FR1 in our research and comes in supports of similar findings by Kleszczowski (2016).

Lastly, we draw a parallel between the role we identify for deliberation and that of “*reference points iterations*” identified by André et al. (2018), which is the process by which stakeholders iterate on performance metrics through repeated interactions. The authors are, to our knowledge, the first to investigate the issue of performance measurement’s performativity for “*social ventures*”, through a case study on a joint venture between a multinational company and a Bangladeshi non-profit (Ruff, 2017 does a similar job on the accounting concept of *materiality* for charities). Following these two contributions, our thesis is in turn a pioneering research exploring performance indicators and performativity for WISEs and social enterprises in general. We add to a restricted set of contributions including Dambrin and Robson (2011, who explore the phenomenon in the pharmaceutical industry) and Vollmer, Mennicken and Preda (2009, offering a take from an accounting standpoint).

11.2 Methodological contributions

In this section, we develop about the two methodological tenets of our thesis that contribute to its originality. The first one is the use of CR and the account we make of the way we operationalise it as a research framework. The second is our use of Maurice's "*societal method*" (1989) for carrying our international comparison. We emphasise how, as we carry on the work of the *Aix group* (cf. below), PVT appears as particularly fitting tool to Maurice's framework.

11.2.1 Methodological development in critical realism

Our research design and the process we follow for the analysis of our findings (particularly in the coding and in the interpretation of our data) follows the CR philosophy. Going from the empirical reality captured by our interview data, we structure our understanding of the actualised and query the deep real using the scaffolds offered by the PVT and the considerations we develop on performativity. We do so using a retroductive reasoning, evidenced by the way our coding grid was actualised as our research progressed.

As such, and as we detail how we address the different layers of CR's stratified ontology in Chapter 5, our thesis offers a contribution to the development of CR methodology. Our research therefore provides an answer to scholars deploring the "lack of methodological development" in CR (Yeung 1997, 52). These include Oliver (2011) and Karlsson and Ackroyd (2014) (with the latter deploring "*serious lack of appealing and accessible material on CR-informed methodology*", 45) and Bhaskar, CR's inceptor, who expressed the importance for its philosophy to be "*applicable*" (2013).

The work we do in describing how CR informs our research can therefore be placed along the lines of contributions such as Edwards et al. (2014), Fletcher (2017) and Parr (2015), although not as exclusively focused on methodological concerns.

11.2.2 An application of the "societal method"

In Chapter 6, we introduce a differentiation between three strands of international comparisons: cross-national (a functionalist approach where analytical units are characterised by a continuity from one

researched country to the other), cross-cultural (where analytical units are essentially considered a product of their cultural environment) and societal¹⁶⁷ (a sort of intermediary stance). We ascribe our work to that last posture, as it is promoted and described by Maurice (1989). With an ontology based on “*multiple and interactive causality*”, the societal approach appears to be a particularly compelling framework for a critical realist work such as ours

Our research is therefore an additional contribution to the societal method, bore by the so-called “*Aix group*” (Hantrais, 2008; Lendaro, 2012; Rose, 1985). To our knowledge, our thesis is a first application of the societal approach using PVT. We argue that it is here another key methodological contribution of our work as PVT’s strategic triangle delivers a powerful analytical matrix for the micro level and allows to create insightful connections between what is observed at the macro level (and attributed to PAM models and administrative profiles) and across countries. On the latter, PVT enables the use of a consistent analytical frame while still capturing the idiosyncrasies of national contexts: an ideal tool for the societal approach.

Furthermore, we find that the use of PVT, in combination with our work on performativity, allows to shed a particular light on the “*joint construction of actors and spaces*” (Maurice, 1989), a process at the core of the societal approach. Our work on strategic interactions, co-production processes (as the fruit of interactions between the macro and the micro levels) and how these matters translate in performance-related questions (at the micro level) are a new application of the insight provided by the societal approach and can be seen as an answer to Reynaud’s call to pursue Maurice’s epistemology (Reynaud, 2011). The seven strategic intents, comparable from one country to another but rooted in their local context are the outcome of this approach.

The same interactions and the role played by STAs at the micro level, as well as the role we associate to PAM reforms at the macro level paint a picture that fluctuates over time. This analysis meets the critiques of the societal approach deploring an excessive structural coherence of national observations (see e.g. Friedberg, 1997; Lendaro, 2012). Lastly, we highlight the relevance of our research in the more recent trends of the societal approach, as we lend a central place to the action of the public sector in

¹⁶⁷ Or “*inter-national*”

our analysis. We thereby build on contributions such as Lendaro (2012) or more crucially Maurice and Sorge,(2000).

11.3 Empirical contributions

The following section highlights how some of our findings can be actioned by field actors interested in the virtuous implementation of performance measurement models. We structure these contributions around two axes: the first one is the state of the art of SIA that we propose and the analytical framework we derive from it. SIA appears to offer a diversity of approaches to measure WISEs' performance with a potential to overcome some of the limitations of *traditional* program evaluation. The second axis is built around more general guidelines to avoid performativity issues and make performance measurement more accessible to those interested in it.

11.3.1 A state of the art for social impact assessment

11.3.1.1 A review of existing methods

One of our key empirical contribution is to bring forth a state of the art for SIA (or at the very least, a thorough review of currently existing models). With the speed at which practitioners and academics come up with new models, there is little doubt that what we propose in this thesis will soon become outdated in terms of exhaustivity. It should nevertheless constitute a robust foundation for anyone wishing to get a detailed overview of the depth and breadth of SIA.

In total, we identified 124 models that can be used to assess a form of social impact, which we list in Appendix 7. From that rather lengthy list, we isolated 18 models that we deem particularly relevant for WISEs, and which also benefit from a sufficient level of documentation in the public domain to be operationalised relatively easily. These 18 models are described in Appendix 8 and they are gauged through selected variables and attributes in Appendix 11.

11.3.1.2 Analytical frameworks for performance models and their indicators

The framework we use in Appendix 11 includes four attributes that we structure through 18 variables. This is the result of a literature review having identified 118 variables (with some redundancies). Our review is based out of the SIA literature, with the idea to use these contributions to structure a view of broader social performance measurement, including models not typically considered as SIA such as the ones used by the Direccte in France. Here, practitioners and academics can find a way to structure an

understanding of the approaches they come across through an analytical framework covering the models' purpose, their analytical lens, the type of data they contain and the type of design they follow. To our knowledge, this last information is a novelty of the framework we propose and was not found in other analyses.

This entry is inspired by Jany-Catrice's performance regimes (2012), who proposes a distinction between performance indicators observing a deliberative design (i.e. following a bottom-up construction taking into account stakeholders' inputs) from indicators exhibiting a planned design (i.e. top-down). As a general rule, the vast majority of SIA models advocate for deliberative designs, which is perhaps why we do not find this attribute in other analytical frameworks (i.e. deliberative design may be considered as granted in SIA). However, as we pointed out already in our theoretical contributions, it would seem that not all WISEs who claim to use a form of SIA apply a deliberative design (we found the ratio to be 50% in a restricted sample in Denmark).

Another of our contributions that goes towards building a robust analytical framework for performance and performance indicators is the seven criteria for performance that we introduce in our Chapter 2: relevance, economy, efficiency, effectiveness, cost-effectiveness, utility and sustainability. These criteria, and the way they flow in a "*performance model*", build mainly on Pollitt and Bouckaert's performance framework (2017), but they are also informed by other sources in program evaluation such as Rossi et al. (2003). Here we contribute by further delineating the framework originally proposed by Pollitt and Bouckaert and, mostly, by explaining clearly what we put behind each criterion (which is unfortunately lacking in the original material).

11.3.2 Guidelines for performance measurement

11.3.2.1 The power of deliberation

Using PVT as an analytical framework in our research has pointed to the importance of deliberation for performance measurement. To this extent, WISEs and policy makers should make sure that whatever the performance measurement model they wish to implement, it should include **public spaces**, or places where conventions (the assumptions pertaining to the scope and arithmetic of indicators) can

be deliberated (Desrosières, 2012; see also Benington's considerations on Habermas' public sphere, 2015).

Our research finds that the importance of deliberation goes beyond the sole design phase of a performance measurement model and its indicators, as is often emphasised (and as we just highlighted above) in SIA literature. In a configuration of public value co-production such as the one experienced by WISEs both in France and in Denmark, we find that ongoing deliberation is conducive to a performance measurement approach that is more focused on actual outcomes than (potentially, as is the case in France) on outputs.

This is evidenced in both countries, despite the very different co-production mechanisms that characterise them: in Denmark, as deliberation occurs naturally to constantly redefine public value on the basis of beneficiaries' development, both WISEs and public stakeholders appear mostly focused on outcomes (even though price considerations can impede that focus). In France, the management dialogue (together with other forms of exchanges, when opportunities arise) is the only platform that WISEs get to deliberate about the centralised performance measurement device with which they must comply. We observed that public value, well beyond the graduation rates followed by the Direccte, is systematically discussed during these conversations and often tacitly co-defined through what we call low intensity STAs (i.e. shared preferences about what the created public value should encompass, but without any indicators attached to it).

11.3.2.2 Fostering different performance measurement models to avoid STAs' enactment

An interesting option to explore for policy makers, local authorities or WISE managers would be the creation of formal **democratic channels** to facilitate the fostering of several STAs. In France, such a mechanism could be embedded in the dialogue that associations and federations representing the different covenants (e.g. FEI, Coorace or Chantier Ecole) host between their members and the government. The explicit aim would be to maintain several competing approaches to performance measurement that WISEs and their stakeholders could adopt as they see fit. On this basis, we could envision the creation of an open library of performance measurement models.

Such solutions would address two recurring issues attached to WISEs' performance measurement. The first one is that it would help circumvent one of the main barriers preventing WISEs from measuring their performance, which is a lack of knowledge. This gap would be bridged by a repository of accessible performance measurement models, vetted by stakeholders, which could be discussed and co-developed freely among interested parties. The second issue our proposal would address is STA enactment, or when a specific model becomes preeminent and performative.

Building upon one of our key theoretical contributions, we argue that both WISEs and public actors should purposely seek the concurrent existence of several performance measurement models. We assimilate the existence of several performance measurement models to competing STAs (Callon, 2006), which we found to be a situation conducive to the occurrence of the deliberation we emphasise above, and more crucially, a configuration where no STAs become enacted and therefore not fully performative. Multiple STAs feed concurrent views on performance and can be a protection against the pervasive effects of performativity. In our research, we see that the SIA models developed by both FR1 and FR2 provide arguments to their management to contest the Direccte' assessment of performance and prevent it from becoming performative.

11.4 Limits and future research alleys

In this last section of our thesis, we first develop about the limits of our research in two sub sections: our application of PAM models and the nature of our empirical material. In the latter, we identify potential weaknesses in our qualitative sampling, which has potential repercussion on aspects of our analysis. We pursue by highlighting possible future research opportunities first around the limits of our own work and second around the broader application of our contributions.

11.4.1 Limits: PAM models and considerations on our empirical material

11.4.1.1 PAM models and their application

A first limit of our research is tied to the often-necessary concessions common in international comparison, as we strive to equate actors and behaviours emerging in different countries, contexts and cultures. While we adopt Maurice's societal approach precisely in order to avoid some of international comparison's pitfalls, we still had to make compromises in order to operationalise our research in a manageable timeframe.

One such compromise is the adoption of NPG and NWS to characterise the two countries we research. Although we state multiple times in our thesis that we do not take these models as preceptial, this choice may still appear as a degree of over simplification in how we analyse each country's institutional environment and the way it shapes strategic interactions between WISEs and public actors. PAM models are not universal recipes, and when relevant, we illustrate this point by balancing our analysis with the use of administrative profiles (Kuhlmann & Wollmann, 2014), which can provide a complementary analytical grid.

Another pitfall associated to the use of PAM models is that their application to one country or the other is not consensual. While the associations we make with France and Denmark is backed by the literature and our own analysis, we also find contributions that could challenge the countries we associate with one regime or the other. One such contribution from Greve and Ejersbo (2016), establishes a link between Denmark and the NWS. Interestingly, the authors do so through Denmark's use of Digital Era Governance mechanisms (Dunleavy et al., 2006), a PAM model that Pollitt and Bouckaert associate to NPG (2017)... Our takeaway form this observation, and our emphasis here, is that PAM models tend to

coexist rather than be mutually exclusive (Iacovino et al., 2017), and our thesis should be understood with this important remark in mind (see e.g. Christensen and Laegreid' reforms layering, 2009).

11.4.1.2 The limits of our empirical material

In addition to the *truncated* nature of our mixed research (i.e. the lack of a quantitative research in France) for the reasons we present in Chapter 6, we highlight two limits in the composition of our empirical material. The first is one is tied to the difficulties we emphasise in Chapter 6 related to the identification of our cases. These challenges have no doubt translated into a *de-facto* creaming of the selected cases, as we ended up researching only WISEs doing *well enough to spare the time and resources* required to take part in our work. This is evidenced particularly in the feedback we got from non-participating WISEs in Denmark.

The second limit associated to the composition of our cases is the consequence of the first one: our cases are all *well-established* in the sense that they are decades-old WISEs and carry a regional weight. This is particularly true in France, where both cases have more than 20 years of operations and are rooted in the political landscape, with ties going to the national level. Our intuition is that management dialogues with the Direccte could likely be less flexible with younger, smaller WISEs that do not benefit from the accumulated legitimacy of our two cases. While not turning our analysis around, this could make some of our conclusions less glaring, with for instance WISEs unable to implement a true "*circumvent*" intent.

We also highlight two limits in the analysis we make of our empirical material. The first one is that while our embedded case study design encompasses stakeholders at the macro level to better comprehend WISEs' environment, our research remains informed primarily by the local, micro level (WISEs and their public stakeholders). We nevertheless find evidence that interactions occurring between levels (e.g. national and local, especially in France due to the weight of WISEs networks) generate relevant interactions that we fail to capture in their entirety. To this extent we note that Bryson et al. (2017) stress the relevance of *levels* in the PVT analysis.

Another point we emphasise is the way we refer to a (relationship-based) quasi-market in Denmark but do not use the term *market* in the French context. Before specifying this limit, we restate the clarification made in Chapter 7: WISEs in both countries are competing for public contracts (e.g. construction or gardening) in order to generate commercial, productive activity and revenue. In Denmark, the market dimension is **also** observed for the provision of employment services (and the access to candidates), which is a paid-for service provided by the WISEs to the municipality (over and above any contracts warranting commercial revenue through productive activities such as construction or gardening). This is not the case in France.

Nevertheless, because covenants are in limited supply for any given geographic area, French WISEs often consider they compete in order to get or preserve their covenant. Such considerations can also be observed in some academic contributions such as Gianfaldoni (2012), who speaks of a “*quasi-market*” (139) for covenants. While we admit this is a limit to the semantics of our thesis, it does not seem to invalidate our analysis founded on the different nature of the strategic interactions observed in France and in Denmark. We also stress how much more of a defining features market-like mechanisms are for the Danish environment than they are in France (particularly price formation, which is absent in the French covenant system).

11.4.2 Future research:

We segment potential research alleys stemming from our work in two sections. The first one introduces leads to address the limits of our thesis or refine its outcomes. The second presents ideas to extend our contributions to broader research topics.

11.4.2.1 Addressing the limits of this research and refining its outcomes

We see three main areas where some limits of our research could be addressed and where our results could be deepened by future contributions.

Researching other levels of interactions

A first opportunity we identify for future research ties into the limit we highlight for how our research fails to capture some of the interactions between levels (especially in the public administration

between the national and the local level). While the absence of these elements is a consequence of our research design, purposely oriented towards the micro level in order to understand strategic interactions, we notice a disconnect between the narratives at the national level and practices at the local level. While both in France and Denmark, national stakeholders state a strong interest for changing the existing approaches to performance towards something more inspired by SIA, very little seems put in practice at the local level. Investigating the disconnects between the levels, using a PVT lens, would contribute to an unexplored area of PVT emphasised by Bryson et al. (2017). In France, deepening the understanding of the role that WISEs networks such as Coorace, FEI or Chantier Ecole can play on the question of performance (and undertaking a research at the meso level) would constitute another very pertinent entry.

Researching younger and smaller WISEs

Future research could also test our findings against WISEs presenting a different profile than our case studies. Smaller and younger organisations should be of a particular interest in order to assess the robustness of our results. Such WISEs would indeed not benefit from the particular sway our case studies seem to enjoy due (chiefly) to their longevity and (to various degrees) to some political connections. Additionally, such work could test hypotheses about the importance for WISEs of building relational capital with public partners and be locally embedded with a network of public stakeholders.

Testing, refining and completing the seven strategic intents

Lastly another area of our own research we deem worth deepening is our identification of seven strategic intents associated to WISEs' performance measurement. Here two particular questions seem worth investigating. The first one is whether more intents can be identified and how these would play out in terms of performativity. A second question is to which extent the intents we associate to a particular PAM environment are not observed in the other (e.g. can the "*educate*" intent be found in NPG?).

11.4.2.2 Capitalising on our thesis: broader research alleys opened by our contributions

We see three broader research areas in which the contributions of our thesis could be built upon:

Revisiting of socio-technical arrangements

While some contributions on the topic of performance measurement and its performativity refer to the notion of STAs (Ambroise, Salle, & Sobel, 2015; Desrosières, 2015), most do not delve too much into the topic. The emphasis we put on STAs brings some new insight on these issues, especially around the different roles they can assume in the face of performativity (i.e. not only pervasive). Revisiting some of the seminal work on performativity in economics, expanding on the ambivalence of STAs as we frame them, could yield valuable knowledge to address performativity-related issues.

Using the extended PVT triangle

Another of our contributions that can be put to a broader use is the nine entries we identify for PVT's strategic triangle. As we offer more granularity to understand and operationalise PVT in research projects, our contribution could be capitalised on to bring answers to some other questions raised by public value scholars. These include the identification of "*larger frameworks*" within the PVT triangle for which the concepts we propose can be seen as "*building blocks*" (Hartley et al., 2017). Bryson et al. (2017, 2015a) also point specifically to the use of an "*expanded PVT triangle*" as a step towards a more general theory of public value creation.

Applying an SIA philosophy to economic indicators

Lastly, we suggest exploring some of the aspects observed about the way SIA models are designed for the creation or the update of broader economic indicators. Such work would tie directly into the question of progress indicators that have gained a considerable traction over the last decade. Here, deliberative designs seem particularly interesting. Jany-Catrice (2008, 2009) shows that similar designs have been used (e.g. for a regional social health index), but always at the micro scale. Here we argue for macro-scaled projects adopting the principle of deliberated conventions (i.e. the assumptions and arithmetic behind the indicator).

Conclusion

Our motivation for this thesis

Work integration enterprises (WISEs) are here to stay. They propose to train and employ vulnerable individuals at the margin of our society, which in the last decades have been services in increasing demand. The reasons for this high demand are multiple and complex but we summarised them as the collapse of the Fordist employment model in the 1970s and a growing attention to performance (often understood as efficiency) in public administration and management (PAM). The latter is frequently epitomised by the New Public Management (Hood, 1991), and while the philosophy of this paradigm is not found in its full extent in the two countries we researched, we still encountered a distinctive emphasis on performance measurement. This preoccupation for performance of course extends to WISEs, and even more so as their relevance increases in changing welfare societies.

It is this question of performance that our thesis focused on. Assessing the performance of mission-driven organisations like WISEs is a complex endeavour, because the outcomes they create (that we assimilate to a form of public value) is intangible, diffuse and do not display clear causal mechanisms. Performance measurement is also a vivid discussion in the literature, as authors such as Gadrey and Jany-Catrice (2016) are advocating for the creation of new performance indicators. In this debate, the notion of performativity (the influence that measurement instrument can have on what they *measure*) has retained our attention, with an emphasis on the democratic stakes of performance measurement and its potential influence on outcomes. In parallel, we also observed that WISEs were developing their own performance measurement frameworks, as social impact assessment (SIA) models, and that these models were used in their interactions with public actors. This prompted us to formulate the following research question:

How strategic interactions between WISEs and public actors are contributing to frame the measurement of performance in different PAM contexts?

To arrive to the notion of strategic interactions, we used public value theory (PVT, Moore, 1995) as the overarching analytical framework for our thesis. With its focus on the creation of public value (i.e. the value generated for the citizenry), PVT proposes a particularly relevant investigative tool to research

WISEs which, as providers of remedies to market failures, are seen as co-creators of public value. PVT is often epitomised by its strategic triangle, or the three vertices that have to be aligned in order to create public value: legitimacy, operationalisation and public value itself. The strategic interactions we problematise are therefore derived from PVT's strategic triangle and refer to the aspects of the relationship between WISEs and their public stakeholders that pertain to the creation of public value, where performance and its measurement are encompassed.

PVT provided a fitting lens for micro analysis allowing us to propose an original framing of performance issues. In order to root our analysis, and following Maurice's societal approach to international comparison (1989), we completed our analytical scaffold with tools to chart the macro environment of public value creation we encountered in each country. For this purpose, we primarily used PAM models: the New Public Governance (NPG; Osborne, 2006) for Denmark (a relationship-based, open and inclusive model) and the New Weberian State (NWS, Pollitt & Bouckaert, 2004) for France (a modernised version of Weber's state resting on bureaucracy, legalism and centralisation).

Our adoption of the societal approach, and our focus on performativity as a phenomenon not directly observable and difficult to capture in traditional economics' closed systems led us to adopt a critical realist (CR) approach (Bhaskar, 1975; Lawson, 2003): CR echoes Maurice's "*multiple and interactive causality*" and its stratified ontology provides a method to identify invisible mechanisms that contribute to explain performativity. This analytical and epistemological framework allowed us to identify two causal mechanisms that contribute to the structuring of strategic interactions and which can be tied to manifestations of performativity.

Two causal mechanisms for performativity

The first one is the co-production environment in which WISEs co-create social value with their public stakeholders. This environment is defined at the macro level and is the translation of PAM models' features to the work integration sector. For instance, the NWS' legalism is a structuring force of work integration in France where WISEs derive their legitimacy from securing a covenant, whereas in Denmark the NPG creates a relationship-based quasi-market where WISEs' legitimacy is assessed both on market terms but also through the build-up of relational capital. Because they define the

mechanisms through which WISEs can fulfil their mission, PAM models and the co-creation environments they induce are a fundamental driver for the nature of the observed strategic interactions. For instance, we see that in France, both FR1 and FR2 spend an amount of resources simply to make sure that their covenant (a form of endorsement, source of legalist legitimacy) is maintained and that they meet their graduation rate targets (a form of legitimacy device). Concurrently in Denmark, we see that WISEs are preoccupied by the prices at which they can propose their service, which stresses out their operation.

The second causal mechanism we identified is the strategic intents that can be tied to performance measurement. We surfaced seven distinct intents that translate the use that is made of performance models by our case studies. These intents are not mutually exclusive and can co-exist (e.g. FR1 pursuing an *“improve”* intent with its own SIA and a *“comply”* intent with the Direccte’s graduation rates), also through a single model (e.g. DK1 which uses its SIA model to *“compete,” “cooperate”* and *“improve”*). There is of course an interdependence between the co-production environments and the strategic intents, which shows by the simple fact that some intents are observed only in one environment, and which we evidence in the representation that we make of our findings in Figure 28.

It is however important to distinguish between the two causal mechanisms simply because the same factors in the environment can contribute to different intents, sometimes towards the same stakeholders (e.g. *“compete”* and *“collaborate”* in Denmark). These observations are consistent with our CR approach which stipulates that because of the open nature of the systems that we research, causal mechanisms *“exist only in virtue of the activities they govern and cannot be empirically identified independently of them”* (Bhaskar, 1979, 48).

Our contributions: implications for performance measurement and performativity issues

Our research shows the complex interplay between PAM models at the macro level, which translate into idiosyncratic public value co-production environments, and strategic interactions at the micro level, where performance measurement takes place. The strategic intents synthesise the way these phenomena influence performance measurement and serve as a gateway to apprehend the occurrence of performativity.

Our work on performance measurement's intents yields new knowledge that capitalises on the original framework we introduced in Chapter 3 to characterise and apprehend performativity (one of our theoretical contributions). We see for instance that the "*comply*" intent is particularly conducive to retroaction, where WISEs act in certain ways not driven by their mission but by the criteria they must meet to maintain their covenant. E.g. we see that FR1 hires some beneficiaries simply to "tick the boxes" of the French bureaucracy, in disconnect from the motivation of public value creation. In Denmark, we see that the "*cooperate*" and "*compete*" intents encourage WISEs to deliberate about what would constitute indicators conducive to the ends they seek (e.g. DK1 taking input from the municipality about the kind of approach that would make a difference to them). We show that as deliberation increases, performativity decreases since actors understand the conventions behind the indicators and do not take them as a replacement of the reality they are meant to proxy. The latter is what we call indicator-substitution: we did not observe it in any of our cases at the local level but it cannot be excluded at the central level in France.

Our research also considerably extends on the notion of socio-technical arrangements (STA, Callon, 2006), or the entanglement between measurement concepts, the environment in which they are developed and what they measure. As we associate performance measurement models to STAs, we see that the existence of multiple (or concurrent) STAs prevent the enactment of one of them (i.e. the process by which STAs become performative). This is already highlighted by Callon as "*performance's struggle*". In our thesis, we however take the extra step of arguing that multiple STAs is a desirable situation. We support this idea by introducing the concept of STAs' feedbacks. Feedbacks can be pervasive and conducive to performativity as the literature already shows (e.g. Callon, 2006, Bracking, 2015, Desrosières, 2015) but they can also be virtuous and a vector of sustained deliberation.

We provide a dichotomic evidence of such a feedback: in Denmark, the absence of formally defined measurement models translates into actors developing their own STAs and essentially advocating for their enactment. For instance DK1 envisions to make its model available to Danish municipalities and DK2SP developed a model in a vacuum, without deliberation, and attempts to force it upon its stakeholders. In France, actors develop their own STAs to *circumvent* the one imposed by the central

power, creating ground for deliberation and for challenging the central model. This has created a national discussion on the topic that is being entertained by national networks as well as (increasingly) by public actors (e.g. the so-called Borello report we mentioned in our introduction).

Expanding on the role of deliberation

We therefore established quite clearly that no matter the type of performance model or the actors supporting its development, deliberation seems key to implement virtuous instruments. This goes in the direction of contributions advocating for new performance indicators such as e.g. Jany-Catrice (2012) or Gadrey and Jany-Catrice (2016). But how can such a deliberation be operationalised? SIA models convey a promotion of deliberative designs that WISEs begin to implement (although our research shows that this is not systematic). But they certainly do not transform the way public actors frame performance (yet?). In France, their emergence has contributed to deliberation, and have brought fresh arguments to argue against the output-based model enforced by the Direccte. However no tangible effects were visible as per a transformation of that model. In Denmark, in the absence of an established performance measurement approach to challenge, SIA remains an attractive idea upon which stakeholders act in isolation.

Another question is the level at which deliberative mechanisms should take place. Our research shows a disconnect between the local and the national level in both the PAM environments we studied where we observe a quasi-systematic deliberation of performance at the local level but almost none at the national level. In France, the system of national advocacy networks (FEI, Coorace, Chantier Ecole) designed specifically to channel local preoccupations towards the centralised actors of the national level seems to yield limited results in that matter given the protruding of the Direccte's output-based approach in the country. At another extreme, the Danish situation where no entity exists to foster a dialogue on the topic beyond the local level is no more fruitful. It would therefore seem that neither the local nor the national levels seem appropriate for fully conducive deliberation on performance measurement.

In between these two levels, one may ask about intermediary echelons (regions, counties, etc.) with a capacity to combine the views of many local actors and confront them to stakeholders closer to the

central state. This question would likely tie into the future research alley we suggest at the meso level (Bryson et al., 2017) in our discussion. Expanding on that thought, one may also wonder about the international level, where we have seen a number of reflections being developed on the topic of social enterprises' performance in recent years (e.g. European Commission / GECES; 2016 G8, 2014; OECD, 2015): are these fora more likely to deliver conclusive deliberation processes pertaining to performance measurement issues? Authors such as Jany-Catrice (2012) and Desrosières (1992) emphasise the role of international deliberation in the design of performance indicators.

More broadly, and as we show that deliberation is favoured by concurrent STAs, we conclude that the co-existence of several models or approaches is preferable to no model at all, or to that of just one single model. This is an idea that aligns well with SIA's philosophy ("*there is no silver bullet*", Tuan, 2008), and this sends us back to the fundamental importance of raising the questions we asked about WISEs' performance: the public value they contribute to is complex, proteiform, intangible and multicausal. Restricting the understanding of this performance to specific systems or beliefs, or tying it to the wrong metric carries a real risk of depriving welfare systems from valuable actors. On the opposite hand, bringing relevant answers to the WISEs' performance (and the gauging of public value) contribute to "a new system of values and a new hierarchy of the 'items that count'" (2016, 7) and could impact much more than the sector of work integration...

Bibliography

- Abdallah, S., Thompson, S., Michaelson, J., Marks, N., & Steuer, N. (2009). The Happy Planet Index 2.0: Why good lives don't have to cost the Earth.
- Abegglen, J. C. (1959). *The Japanese factory: Aspects of its social organization*. Asia publishing house.
- Abramsky, J., & Drew, S. (2014, May 29). Changes to National Accounts: Inclusion of Illegal Drugs and Prostitution in the UK National Accounts. Office for National Statistics.
- Agranoff, R. (2007). *Managing within networks: Adding value to public organizations*. Georgetown University Press.
- Albæk, E. (1998). Knowledge, interests and the many meanings of evaluation: a developmental perspective. *Scandinavian Journal of Social Welfare*, 7(2), 94–98.
- Albæk, E., & Rieper, O. (2002). *Evaluation in Danish Governance*. Transaction Publishers.
- Alexander, J., Brudney, J. L., & Yang, K. (2010). Introduction to the symposium: Accountability and performance measurement: The evolving role of nonprofits in the hollow state. *Nonprofit and Voluntary Sector Quarterly*, 39(4), 565–570.
- Alford, J. (2002). Defining the client in the public sector: A social-exchange perspective. *Public Administration Review*, 62(3), 337–346.
- Alford, J., & Hughes, O. (2008). Public value pragmatism as the next phase of public management. *The American Review of Public Administration*, 38(2), 130–148.
- Alford, J., & O'Flynn, J. (2009). Making sense of public value: Concepts, critiques and emergent meanings. *Intl Journal of Public Administration*, 32(3–4), 171–191.
- Allard-Poesi, F. (2003). Coder les données. *Conduire Un Projet de Recherche: Une Perspective Qualitative*, 245–290.
- Allard-Poesi, F., & Maréchal, G. (2014). Construction de l'objet de la recherche. In *Méthodes de recherche en management-4ème édition*. Dunod.
- Allard-Poesi, F., & Perret, V. (2014). Fondements épistémologiques de la recherche. In *Méthodes de recherche en management-4ème édition*. Dunod.
- Altschuld, J. W., & Engle, M. (2015). The inexorable historical press of the developing evaluation profession. *New Directions for Evaluation*, 2015(145), 5–19.
- Alvesson, M., & Sköldberg, K. (2009). *Reflexive methodology: New vistas for qualitative research*. Sage.
- Ambroise, B., Salle, G., & Sobel, R. (2015). L'économie entre performativité, idéologie et pouvoir symbolique. *L'Homme et La Société*, (3), 13–30.
- Andersen, L. L., Gawell, M., & Spear, R. (2016). *Social Entrepreneurship and Social Enterprises: Nordic Perspectives*. Routledge.
- Andersen, L. L., & Hulgård, L. (2008). Workplaces with an Inclusive Profile-and their Everyday Strategies. Social Enterprises and Work Integration in Denmark. In *International Workshop on Ethnographies of the Social Economy*.
- André, K., Cho, C. H., & Laine, M. (2018). Reference points for measuring social performance: Case study of a social business venture. *Journal of Business Venturing*, 33(5), 660–678. <https://doi.org/10.1016/j.jbusvent.2017.12.002>
- Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (2013). *Critical realism: Essential readings*. Routledge.
- Arrow, K. J. (1951). Social choice and individual values. *Cowles Foundation*.
- Assessment, S. I. (1995). Guidelines and principles for social impact assessment. *Environmental Impact Assessment Review*, 15(1), 11–43.

- Austin, J., Stevenson, H., & Wei-Skillern, J. (2006). Social and Commercial Entrepreneurship: Same, Different, or Both? *Entrepreneurship Theory and Practice*, 30(1), 1–22.
- Ayache, M., & Dumez, H. (2011). Le codage dans la recherche qualitative une nouvelle perspective? *Le Libellio d'Aegis*, 7(2-Eté), 33–46.
- Backhouse, R. (1994). *New directions in economic methodology* (Vol. 3). Psychology Press.
- Bagnoli, L., & Megali, C. (2011). Measuring performance in social enterprises. *Nonprofit and Voluntary Sector Quarterly*, 40(1), 149–165.
- Bartlett, W., & Le Grand, J. (1993). The theory of quasi-markets. In *Quasi-markets and social policy* (pp. 13–34). Springer.
- Bator, F. M. (1958). The anatomy of market failure. *The Quarterly Journal of Economics*, 72(3), 351–379.
- Battilana, J., & Dorado, S. (2010). Building sustainable hybrid organizations: The case of commercial microfinance organizations. *Academy of Management Journal*, 53(6), 1419–1440.
- Battilana, J., & Lee, M. (2014). Advancing research on hybrid organizing—Insights from the study of social enterprises. *The Academy of Management Annals*, 8(1), 397–441.
- Battilana, J., Sengul, M., Pache, A.-C., & Model, J. (2015). Harnessing productive tensions in hybrid organizations: The case of work integration social enterprises. *Academy of Management Journal*, 58(6), 1658–1685.
- Baumard, P., & Ibert, J. (2014). Quelles approches avec quelles données? In *Méthodes de recherche en management-4ème édition*. Dunod.
- Beaud, J. P., & Prévost, J. G. (2000). L'ère du chiffre/The Age of Numbers. *Montréal: Presses de l'Université Du Québec*.
- Beitone, A., Cazorla, A., Dollo, C., & Draï, A.-M. (2013). *Dictionnaire de science économique*. Armand Colin.
- Bellamy, R. P., & Palumbo, A. (2010). *From government to governance*. Ashgate.
- Benington, J. (2011). From private choice to public value. *Public Value: Theory and Practice*, 31–49.
- Benington, J. (2015). Public value as a contested democratic practice. In *Creating Public Value in Practice*.
- Benington, J., & Moore, M. H. (2010). *Public value: Theory and practice*. Macmillan International Higher Education.
- Bewley, T. (2002). Interviews as a valid empirical tool in economics. *The Journal of Socio-Economics*, 31(4), 343–353.
- Bezes, P. (2010). Morphologie de la RGPP. *Revue Française d'administration Publique*, (4), 775–802.
- Bezes, P. (2016). Challenges to French Public Administration. *The Oxford Handbook of French Politics*, 243.
- Bhaskar, R. (1975). *A realist theory of science*. Routledge.
- Bhaskar, R. (1979). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Borello, J.-M., & Barfety, J.-B. (2018). *Donnons-nous les moyens de l'inclusion*. Ministère du Travail.
- Borus, M. E. (1979). *Measuring the Impact of Employment-Related Social Programs. A Primer on the Evaluation of Employment and Training, Vocational Education, Vocational Rehabilitation, and Other Job-Oriented Programs*.

- Borzaga, C., & Galera, G. (2016). *Social Enterprises and their eco-systems: developments in Europe*. European Commission, Directorate-General for Employment, Social Affairs and Inclusion.
- Boston, J. (1996). *Public management: the New Zealand model*. Oxford University Press, USA.
- Bouchard, T. J. (1976). Field research methods: Interviewing, questionnaires, participant observation, systematic observation, unobtrusive measures. *Handbook of Industrial and Organizational Psychology, 1*, 363.
- Bourgon, J. (1998). *Sixth Annual Report to the Prime Minister on the Public Service of Canada*. Privy Council and Secretary to the Cabinet.
- Bovaird, T. (2006). Developing new forms of partnership with the 'market' in the procurement of public services. *Public Administration, 84*(1), 81–102.
- Bovaird, T., Stoker, G., Jones, T., Loeffler, E., & Pinilla Roncancio, M. (2016). Activating collective co-production of public services: influencing citizens to participate in complex governance mechanisms in the UK. *International Review of Administrative Sciences, 82*(1), 47–68.
- Bowen Howard, R. (1953). *Social responsibilities of the businessman*. New York, Harper.
- Boyne, G., Farrell, C., & Law, J. (2003). *Evaluating public management reforms: Principles and practice*. McGraw-Hill Education (UK).
- Boyte, H. C. (2011). Constructive politics as public work: Organizing the literature. *Political Theory, 39*(5), 630–660.
- Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press.
- Bracking, S. (2012). How do investors value environmental harm/care? Private equity funds, development finance institutions and the partial financialization of nature-based industries. *Development and Change, 43*(1), 271–293.
- Bracking, S. (2015). Performativity in the Green Economy: how far does climate finance create a fictive economy? *Third World Quarterly, 36*(12), 2337–2357.
- Brennan, M. (2014). Introduction of ESA 2010 into the national accounts of Ireland. *Eurostat Review on National Accounts and Macroeconomic Indicators -EURONA*, (Special Issue on the Implementation of the European System of Accounts (ESA 2010)).
- Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. *Public Administration Review, 74*(4), 445–456.
- Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2015a). *Creating public value in practice: Advancing the common good in a multi-sector, shared-power, no-one-wholly-in-charge world* (Vol. 194). CRC Press.
- Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2015b). *Public Value and Public Administration*. Georgetown University Press.
- Bryson, J. M., Sancino, A., Benington, J., & Sørensen, E. (2017). Towards a multi-actor theory of public value co-creation. *Public Management Review, 19*(5), 640–654.
- Burdge, R. J. (2003). The practice of social impact assessment background. *Impact Assessment and Project Appraisal, 21*(2), 84–88.
- Burdge, R. J., & Vanclay, F. (1995). Social Impact Assessment: a contribution to the state of the art series. *Environmental and Social Impact Assessment, 14*.
- Callon, M. (2004). The role of hybrid communities and socio-technical arrangements in the participatory design. *Journal of the Center for Information Studies, 5*(3), 3–10.

- Callon, M. (2005). Why virtualism paves the way to political impotence: A reply to Daniel Miller's critique of "The laws of the market". *Economic Sociology: European Electronic Newsletter*, 6(2), 3–20.
- Callon, M. (2006). What does it mean to say that economics is performative? *CSI Working Papers Series*, (5).
- Callon, M. (2007). An essay on the growing contribution of economic markets to the proliferation of the social. *Theory, Culture & Society*, 24(7–8), 139–163.
- Callon, M. (2009). Civilizing markets: Carbon trading between in vitro and in vivo experiments. *Accounting, Organizations and Society*, 34(3–4), 535–548.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312.
- Campbell, O. D. (1928). *The Danish folk school: Its influence in the life of Denmark and the North*. Macmillan.
- Cantril, H. (1965). Pattern of human concerns.
- Caro, F. G. (1977). *Readings in Evaluation Research*, 2ed. Russell Sage Foundation.
- Carroll, A. B., & Beiler, G. W. (1975). Landmarks in the Evolution of the Social Audit. *Academy of Management Journal*, 18(3), 589–599.
- Certo, S. T., & Miller, T. (2008). Social entrepreneurship: Key issues and concepts. *Business Horizons*, 51(4), 267–271. <https://doi.org/10.1016/j.bushor.2008.02.009>
- Chalmers, A. F. (2013). *What is this thing called science?* Hackett Publishing.
- Chandler, J. A. (1991). Public administration: a discipline in decline. *Teaching Public Administration*, 11(2), 39–45.
- Charreire Petit, S., & Durieux, F. (2014). Explorer et tester: les deux voies de la recherche. In *Méthodes de recherche en management-4ème édition*. Dunod.
- Christensen, T. (2012). Post-NPM and changing public governance. *Meiji Journal of Political Science and Economics*, 1(1), 1–11.
- Christensen, T., & Laegreid, P. (2009). Transcending New Public Management—the Increasing Complexity of Balancing Control and Autonomy. *Journal of Public Administration*, 2(1), 1–29.
- Christiansen, P. M., & Nørgaard, A. S. (2003). Faste forhold-flygtige forbindelser. Stat og interesseorganisationer i Danmark i det 20. århundrede.
- Ciriacy-Wantrup, S. V. (1947). Capital returns from soil-conservation practices. *Journal of Farm Economics*, 29(4), 1181–1196.
- Clark, C., Rosenzweig, W., Long, D., & Olsen, S. (2004). Double bottom line project report: assessing social impact in double bottom line ventures.
- Coats, D. (2006). *Reviving the Public: A new governance and management model for public services*. Work Foundation.
- Cobb, C., Halstead, T., & Rowe, J. (1995). The genuine progress indicator. *Redefining Progress*, San Francisco, CA.
- Cobb, C. W., & Daly, H. (1989). The index for sustainable economic welfare. *Daly, HE, Cobb, JB (Eds)*.
- Colwell, R. (1998). A long trail awinding: Issues, interests, and priorities in arts education research. *Arts Education Policy Review*, 99(5), 21–29.
- Committee on Social Enterprises. (2013). Recommendation Report. Committe on Social Enterprises.
- Conseil National des Chambres Régionales de l'Economie Sociale et Solidaire. (2017). *Atlas Commenté de l'Economie Sociale et Solidaire* (Daloz, Juris editions). Paris.

- Cook, M., & Harrison, T. M. (2015). Using public value thinking for government IT planning and decision making: A case study. *Information Polity*, 20(2, 3), 183–197.
- Corbett, A. (2010). Public management policymaking in France: legislating the Organic Law on Laws of Finance (LOLF), 1998–2001. *Governance*, 23(2), 225–249.
- Council for Social Development. (2010). *Report of a Research Project on Social Impact Assessment of R&R Policies and Packages in India*. New Delhi. Retrieved from <https://images.template.net/wp-content/uploads/2016/08/09084715/Social-Impact-Assessment-Template.doc>
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., & others. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.
- Crosby, B. C., Bryson, J. M., & Stone, M. M. (2009). Leading across frontiers: how visionary leaders integrate people, processes, structures and resources. In *The New Public Governance?: Emerging Perspectives on the Theory and Practice of Public Governance*. Routledge Taylor & Francis Group.
- Curtis, T. (2008). Finding that grit makes a pearl: A critical re-reading of research into social enterprise. *International Journal of Entrepreneurial Behavior & Research*, 14(5), 276–290.
- Dahl, A., & Soss, J. (2014). Neoliberalism for the common good? Public value governance and the downsizing of democracy. *Public Administration Review*, 74(4), 496–504.
- Dambrin, C., & Robson, K. (2011). Tracing performance in the pharmaceutical industry: Ambivalence, opacity and the performativity of flawed measures. *Accounting, Organizations and Society*, 36(7), 428–455. <https://doi.org/10.1016/j.aos.2011.07.006>
- Davies, N. (1931). *Education for life: a Danish pioneer*. William and Norgate Ltd, London.
- Davister, C., Defourny, J., & Grégoire, O. (2004). *Work integration social enterprises in the European Union: an overview of existing models*. Working Papers Series.
- De Gaulejac, V. (2005). *La société malade de la gestion*. Seuil Paris.
- De Gaulejac, V. (2011). *Travail, les raisons de la colère*. Le Seuil.
- De Jong, J., Douglas, S., Sicilia, M., Radnor, Z., Noordegraaf, M., & Debus, P. (2017). Instruments of value: using the analytic tools of public value theory in teaching and practice. *Public Management Review*, 19(5), 605–620.
- De Visscher, C. (2005). Le coup dans l'eau de Copernic: réforme de la haute fonction publique, nouvelle gestion publique et participatie en Belgique. *Politiques et Management Public*, 23(4), 33–51.
- Dees, J. G., Ewing, M., Haas, M., & Haas, P. (1998). *The meaning of social entrepreneurship*.
- Defourny, J., Hulgård, L., & Pestoff, V. A. (2014). Introduction to the SE Field. *Social Enterprise and the Third Sector: Changing European Landscapes in a Comparative Perspective*, 1–14.
- Defourny, J., & Nyssens, M. (2008). Social enterprise in Europe: recent trends and developments. *Social Enterprise Journal*, 4(3), 202–228.
- Defourny, J., & Nyssens, M. (2010). Conceptions of social enterprise and social entrepreneurship in Europe and the United States: Convergences and divergences. *Journal of Social Entrepreneurship*, 1(1), 32–53.
- Defourny, J., & Nyssens, M. (2013). Social innovation, social economy and social enterprise: what can the European debate tell us? *The International Handbook on Social Innovation*, 40–53.
- Deleuze, G. (1989). Michel Foucault philosophe: rencontre internationale. In *Qu'est-ce qu'un dispositif?* (Seuil, pp. 185–195). Paris.
- Deleuze, G., & Guattari, F. (1980). Mille plateaux. *Capitalisme et Schizophrénie*.
- Delorme, R. (1999). Realism in Economics: Critical or Complex. Citeseer.

- Denhardt, J. V., & Denhardt, R. B. (2015). *The new public service: Serving, not steering*. Routledge.
- Department of Labour. (2006). *Partnership and Productivity: A Review of the Literature*. Wellington, New Zealand: Partnership Resource Centre.
- Desilver, D. (2017, March). What the unemployment rate does – and doesn't – say about the economy. Retrieved July 19, 2018, from <http://www.pewresearch.org/fact-tank/2017/03/07/employment-vs-unemployment-different-stories-from-the-jobs-numbers/>
- Desrosières, A. (1992). Discuter l'indiscutable. *Raisons Pratiques*, 3, 131–154.
- Desrosières, A. (1993). *La politique des grands nombres. Histoire de la raison statistique*.
- Desrosières, A. (2006). From cournot to public policy evaluation: Paradoxes and controverses involving quantification. *Cournot Centre, Prisme*, (7).
- Desrosières, A. (2010). 2 Words and Numbers. *The Mutual Construction of Statistics and Society*, 41.
- Desrosières, A. (2012). Est-il bon, est-il méchant? Le rôle du nombre dans le gouvernement de la cité néolibérale. *Nouvelles Perspectives En Sciences Sociales: Revue Internationale de Systémique Complexe et d'études Relationnelles*, 7(2), 261–295.
- Desrosières, A. (2014). *Prouver et gouverner: une analyse politique des statistiques publiques*. La Découverte.
- Desrosières, A. (2015). Retroaction: How indicators feed back onto quantified actors. *The World of Indicators. The Making of Governmental Knowledge through Quantification*, 329–53.
- Dewey, D. R. (1910). Observation in economics. Annual address of the president. *American Economic Association Quarterly*, 11(1), 29–45.
- Dierkens, N. (1991). Information asymmetry and equity issues. *Journal of Financial and Quantitative Analysis*, 26(2), 181–199.
- DiMaggio, P., & Powell, W. W. (1983). The iron cage revisited: Collective rationality and institutional isomorphism in organizational fields. *American Sociological Review*, 48(2), 147–160.
- Dow, S. (2002). Economic methodology: an inquiry. *OUP Catalogue*.
- Dow, S. C. (2007). Variety of methodological approach in economics. *Journal of Economic Surveys*, 21(3), 447–465.
- Drechsler, W. (2009). Towards the neo-Weberian state? Perhaps, but certainly adieu, NPM. *The NISPAcee Journal of Public Administration and Policy*, 1(2), 95–99.
- Du Buat-Nancay, L.-G. (1773). *Politique, ou recherche des vrais principes de l'économie sociale*. London.
- Dufour, B. (2018). Social impact measurement: What can impact investment practices and the policy evaluation paradigm learn from each other? *Research in International Business and Finance*.
- Dunleavy, P. (1985). Bureaucrats, budgets and the growth of the state: reconstructing an instrumental model. *British Journal of Political Science*, 15(3), 299–328.
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—long live digital-era governance. *Journal of Public Administration Research and Theory*, 16(3), 467–494.
- Dunn, W. N., & Miller, D. Y. (2007). A critique of the new public management and the neo-Weberian state: advancing a critical theory of administrative reform. *Public Organization Review*, 7(4), 345–358.
- Dunoyer, C. (1830). *Nouveau Traité d'Economie Sociale* (A. Sautet et Cie). Paris.
- Dupré, M., Jacob, A., Lallement, M., Lefèvre, G., & Spurk, J. (2003). *Les comparaisons internationales: intérêt et actualité d'une stratégie de recherche*. Éditions du CNRS.
- Durquety, E., & Baudet, A. (2017). *Baromètre de la mesure d'impact social*. KPMG.
- Durquety, E., & Baudet, A. (2018). *Baromètre de la mesure d'impact social*. KPMG.

- Edwards, P. K., O'Mahoney, J., & Vincent, S. (2014). *Studying organizations using critical realism: A practical guide* (Vol. 17). OUP Oxford.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.
- Elkington, J. (1997). *Cannibals with Forks: Triple Bottom Line of 21st Century Business*. Oxford: Capstone Publishing Ltd.
- Elmeskov, J., & Bang, H. (2017). *Indvandrere i Danmark 2017*. Danmarks Statistik.
- Emerson, J. (2003). The blended value proposition. *California Management Review*, 45(4), 36.
- Emerson, J., Wachowicz, J., & Chun, S. (2001). Social return on investment (SROI): exploring aspects of value creation. *Harvard Working Knowledge Archive*. Retrieved from <https://hbswk.hbs.edu/archive/social-return-on-investment-sroi-exploring-aspects-of-value-creation>
- Engels, X., Hély, M., Peyrin, A., & Trouvé, H. (2006). *De l'intérêt général à l'utilité sociale: La reconfiguration de l'action publique entre Etat, associations et participation citoyenne*. (L'Harmattan). Paris.
- Epstein, M. J., & Yuthas, K. (2017). *Measuring and Improving Social Impacts: a guide for nonprofits, companies and Impact investors*. Routledge.
- European Commission. (2015). *A map of social enterprises and their eco-systems in Europe - Synthesis Report*. Luxembourg: European Commission.
- European Commission / GECES. (2014). *Approches proposées pour la mesure de l'impact social*. European Commission.
- European Commission / GECES. (2016). *Social enterprises and the social economy going forward*. Belgium: European Commission.
- European Commission, & OECD. (2015). Policy Brief on Social Impact Measurement for Social Enterprises. Publication Office of the European Union.
- Evrard, Y., Pras, B., Roux, E., Desmet, P., Dussaix, A.-M., & Lilien, G. L. (2009). *Market-Fondements et méthodes des recherches en marketing*.
- Farnham, D., & Horton, S. (1993). *Managing the new public services*. Springer.
- Fisher, T. (2014). Public value and the integrative mind: How multiple sectors can collaborate in city building. *Public Administration Review*, 74(4), 457–464.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2017). *Program Evaluation: Alternative Approaches and Practical Guidelines* (5 edition). Upper Saddle River, N.J: Pearson.
- Fleetwood, S. (1999). *Critical realism in economics: Development and debate* (Vol. 12). Psychology Press.
- Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*, 20(2), 181–194.
- Flynn, N. (2007). *Public sector management*. Sage.
- Foucault, M. (1977). L'œil du pouvoir. *Dits et Écrits*, 3, 190–207.
- Fourquet, F. (1980). Les comptes de la puissance. Histoire de la comptabilité nationale et du plan. *Paris, Recherches (Encres)*.
- Fraisse, L., Guérin, I., & Laville, J.-L. (2007). Economie solidaire: des initiatives locales à l'action publique. Introduction. *Revue Tiers Monde*, (2), 245–253.
- Frederickson, D. G. (2006). *Measuring the performance of the hollow state*. Georgetown University Press.
- Freudenburg, W. R. (1986). Social Impact Assessment. *Annual Review of Sociology*, 12, 451–478.

- Friedberg, E. (1997). *Le pouvoir et la règle*. Paris: Seuil.
- Furubo, J.-E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. Transaction Publishers.
- G8. (2014, September). Impact investment: the invisible heart of markets.
- Gadrey, J., & Jany-Catrice, F. (2016). *Les nouveaux indicateurs de richesse* (4th ed.). La découverte.
- Geuijen, K., Moore, M. H., Cederquist, A., Ronning, R., & van Twist, M. (2017). Creating public value in global wicked problems. *Public Management Review*, 19(5), 621–639.
- Gianfaldoni, P. (2012). Les attributs de l'innovation sociale dans l'insertion par l'activité économique. *Innovations*, (2), 129–150.
- Gianfaldoni, P. (2018). Les évolutions contraintes des entreprises sociales d'insertion par le travail. *Marché et Organisations*, (1), 81–101.
- Gianfaldoni, P., & Rostaing, F. (2010). L'Insertion par l'Activité Economique comme espace d'action publique: entre politique publique et gouvernance. *Politiques et Management Public*, 27(1), 143–165.
- Gibbon, J., & Dey, C. (2011). Developments in social impact measurement in the third sector: scaling up or dumbing down? *Social and Environmental Accountability Journal*, 31(1), 63–72.
- Glaser, B., & Strauss, A. (1967). *The discovery grounded theory: strategies for qualitative inquiry*. Aldin, Chicago.
- Goodland, R. (2000). Social and environmental assessment to promote sustainability: An informal view from the World Bank.
- Goyder, G. (1961). *The responsible company*. Blackwell.
- Grawitz, M. (1993). *Méthodes des sciences sociales*. Paris: Éditions Dalloz.
- Gray, R. (2000). Current developments and trends in social and environmental auditing, reporting and attestation: a review and comment. *International Journal of Auditing*, 4(3), 247–268.
- Gray, R., Dey, C., Owen, D., Evans, R., & Zadek, S. (1997). Struggling with the praxis of social accounting: Stakeholders, accountability, audits and procedures. *Accounting, Auditing & Accountability Journal*, 10(3), 325–364.
- Gray, R., Owen, D., & Maunders, K. (1987). *Corporate Social Reporting: Accounting and Accountability*. Prentice-Hall International.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274.
- Greenhouse, S. M. (1966). The Planning-Programming-Budgeting System: Rationale, Language, and Idea-Relationships. *Public Administration Review*, 26(4), 271–277. <https://doi.org/10.2307/973298>
- Greve, C. (2006). Public management reform in Denmark. *Public Management Review*, 8(1), 161–169.
- Greve, C., & Ejersbo, N. (2016). 11. Denmark: towards the Neo-Weberian State in the digital era. *Public Administration Reforms in Europe: The View from the Top*, 119.
- Greve, C., & Hodge, G. (2010). Public-private partnerships and public governance challenges. In *New Public Governance?* (pp. 149–162). Routledge.
- Greve, C., Laegreid, P., & Rykkja, L. H. (2016). *Nordic administrative reforms: Lessons for public management*. Springer.
- Grieco, C. (2015). *Assessing Social Impact of Social Enterprises: Does One Size Really Fit All?* Springer.
- Grieco, C., Michelini, L., & Iasevoli, G. (2015). Measuring value creation in social enterprises: A cluster analysis of social impact assessment models. *Nonprofit and Voluntary Sector Quarterly*, 44(6), 1173–1193.
- Guba, E. G. (1990). *The paradigm dialog*. Sage Publications.

- Guba, E. G., Lincoln, Y. S., & others. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, 2(163–194).
- Guskey, T. R. (1999). *Evaluating professional development*. Corwin Press.
- Habermas, J. (1962). The Structural Transformation of the Public Sphere (Cambridge. *Polity*, 7(8).
- Hamilton, K., & Naikal, E. (2014). Genuine saving as an indicator of sustainability. *Handbook of Sustainable Development*, 336.
- Hammerschmid, G., Van de Walle, S., Andrews, R., & Bezes, P. (2016). *Public administration reforms in Europe: The view from the top*. Edward Elgar Publishing.
- Hammersley, M. (1992). Deconstructing the qualitative-quantitative divide.
- Hantrais, L. (2008). *International Comparative Research: Theory, Methods and Practice*. Macmillan International Higher Education.
- Hart, K., Laville, J.-L., & Cattani, A. D. (2010). *The human economy*. Polity Press Cambridge.
- Hartley, J., Alford, J., Knies, E., & Douglas, S. (2017). Towards an empirical research agenda for public value theory. *Public Management Review*, 19(5), 670–685.
- Hassenteufel, P. (2005). De la comparaison internationale à la comparaison transnationale. *Revue Française de Science Politique*, 55(1), 113–132.
- Haveman, H. A., & Rao, H. (2006). Hybrid forms and the evolution of thrifts. *American Behavioral Scientist*, 49(7), 974–986.
- Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3, 1865–2097.
- Hédoin, C. (2010). Le réalisme critique de Tony Lawson: apports et limites dans une perspective institutionnaliste. *Cahiers d'économie Politique/Papers in Political Economy*, (1), 103–131.
- Hegland, T. J. (1994). *Fra de tusind blomster: til en m\ a alrettet udvikling*. Alfuff.
- Heinrich, C. J., & Lynn, L. (2000). *Governance and performance: New perspectives*. Georgetown University Press.
- Hély, M. (2006). *De l'intérêt général à l'utilité sociale: Transformations de l'état social et genèse du travailleur associatif*. L'Harmattan.
- Hély, M. (2010). Le travail «d'utilité sociale» dans le monde associatif. *Management & Avenir*, (10), 206–217.
- Hess, D. (2008). The three pillars of corporate social reporting as new governance regulation: Disclosure, dialogue, and development. *Business Ethics Quarterly*, 447–482.
- Hodgson, G. M. (2004). Some claims made for critical realism in economics: two case studies. *Journal of Economic Methodology*, 11(1), 53–73.
- Hogan, R. L. (2010). The Historical Development of Program Evaluation: Exploring Past and Present. *Online Journal for Workforce Education and Development*, 2(4), 5.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Hood, C., & Jackson, M. (1992). The new public management: a recipe for disaster? *Hazard Management and Emergency Planning: Perspectives on Britain*, 109–125.
- Horner, L., & Hazel, L. (2005). Adding public value. *Work Foundation, London*.
- Horton, M., & Kohl, J. and H. (1997). *The Long Haul: An Autobiography* (1 edition). New York: Teachers College Press.
- House, E. R. (1983). Assumptions underlying evaluation models. In *Evaluation models* (pp. 45–64). Springer.

- Howlett, M., Kekez, A., & Poocharoen, O.-O. (2017). Understanding co-production as a policy tool: Integrating new public governance and comparative policy theory. *Journal of Comparative Policy Analysis: Research and Practice*, 19(5), 487–501.
- Hulgård, L. (2011). Social economy and social enterprise: an emerging alternative to mainstream market economy? *China Journal of Social Work*, 4(3), 201–215.
- Hulgård, L., & Andersen, L. L. (2015). Social entrepreneurship and social innovation. In *Social entrepreneurship and social innovation* (Rosendahls-Schultz Grafisk). Copenhagen: Nordic Council of Ministers.
- Hulgård, L., & Bisballe, T. (2004). *Work integration social enterprises in Denmark*. Working papers series.
- Husmanns, R., Mehran, F., & Varmā, V. (1990). *Surveys of economically active population, employment, unemployment, and underemployment: an ILO manual on concepts and methods*. International Labour Organization.
- Huxham, C., & Vangen, S. (2013). *Managing to collaborate: The theory and practice of collaborative advantage*. Routledge.
- Iacovino, N. M., Barsanti, S., & Cinquini, L. (2017). Public Organizations Between Old Public Administration, New Public Management and Public Governance: the Case of the Tuscany Region. *Public Organization Review*, 17(1), 61–82.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.
- INSEE. (2018). *ILO Unemployment and Labour Market-related indicators (Labour Force Survey results) - fourth quarter 2017* (Informations Rapides No. 38). INSEE.
- International Association for Impact Assessment. (2010). Social Impact Assessment. Retrieved April 11, 2014, from <http://www.iaia.org/iaia/wiki/sia.ashx>
- Jackson, T., & Stymne-Airey, S. (1996). *Sustainable economic welfare in Sweden: a pilot index: 1950-1992*. SEI.
- Jacobs, L. R. (2014). The contested politics of public value. *Public Administration Review*, 74(4), 480–494.
- Jacobsen, R. H. (2013). *Virksomheder med udvidet socialt ansvar*. Copenhagen: Center for Economic and Business Research, Copenhagen Business School.
- Jany-Catrice, F. (2008). Des indicateurs composites de développement humain et de santé sociale: un enjeu de démocratie?
- Jany-Catrice, F. (2009). The french regions and their social health. *Social Indicators Research*, 93(2), 377–391.
- Jany-Catrice, F. (2012). *La performance totale: nouvel esprit du capitalisme?* Presses Univ. Septentrion.
- Jensen, E. (2012). *Fokus på socialøkonomiske virksomheder*. Copenhagen: Center for Aktiv Beskæftigelsesindsats (CABI).
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 602–611.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133.
- Kalambokidis, L. (2014). Creating public value with tax and spending policies: The view from public economics. *Public Administration Review*, 74(4), 519–526.

- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Sciences*. San Francisco: Chandler Pub. Co.
- Karlsson, J. C., & Ackroyd, S. (2014). Critical Realism, Research Techniques, and Research Designs.
- Kaufmann, D. (2007). *Governance matters VI: aggregate and individual governance indicators, 1996-2006* (Vol. 4280). World Bank Publications.
- Kelly, G., Mulgan, G., & Muers, S. (2002). *Creating Public Value: An analytical framework for public service reform*. London: Strategy Unit, Cabinet Office.
- Kettl, D. F. (2006). *The global public management revolution*. Brookings Institution Press.
- Kettl, D. F. (2011). *Sharing power: Public governance and private markets*. Brookings Institution Press.
- Kickert, W. (1993). Complexity, governance and dynamics: conceptual explorations of public network management. *Modern Governance*, 191–204.
- Kickert, W. J., Klijn, E.-H., & Koppenjan, J. F. (1997). *Managing complex networks: Strategies for the public sector*. Sage.
- Kleszczowski, J. (2016). La place des parties prenantes dans l'évaluation de l'impact social des organisations non lucratives: étude empirique au sein d'une organisation française. In *Comptabilité et gouvernance* (p. cd-rom). Clermont-Ferrand, France. Retrieved from <https://hal.archives-ouvertes.fr/hal-01901230>
- Klijn, E.-H. (2005). Networks and inter-organizational management. In *The Oxford handbook of public management*.
- Klijn, E.-H., & Teisman, G. (2005). Public-private partnerships as the management of co-production: strategic and institutional obstacles in a difficult marriage. *The Challenge of Public-Private Partnerships: Learning from International Experience*, 95–116.
- Koenig, G. (1993). Production de la connaissance et constitution des pratiques organisationnelles.
- Kooiman, J. (1999). Social-political governance: overview, reflections and design. *Public Management an International Journal of Research and Theory*, 1(1), 67–92.
- Kraatz, M. S., & Block, E. S. (2008). Organizational implications of institutional pluralism. *The Sage Handbook of Organizational Institutionalism*, 840, 243–275.
- Krlev, G., Münscher, R., & Mülberr, K. (2013). Social Return on Investment (SROI): State-of-the-Art and Perspectives. *CSI Report*.
- Kuhlmann, S., & Wollmann, H. (2014). *Introduction to comparative public administration: Administrative systems and reforms in Europe*. Edward Elgar Publishing.
- Kuznets, S. (1934). *National income 1929–1932: Letter from the Acting Secretary of Commerce*.
- Kuznets, S. (1941). *National Income And Its Composition 1919 1938 Vol I*.
- Laegreid, P. (2017). *Transcending new public management: the transformation of public sector reforms*. Routledge.
- Lall, S. (2017). Measuring to Improve Versus Measuring to Prove: Understanding the Adoption of Social Performance Measurement Practices in Nascent Social Enterprises. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 28(6), 2633–2657.
- Lane, J.-E. (2002). *New public management: an introduction*. Routledge.
- Lawrence, T. B., Suddaby, R., & Leca, B. (2009). *Institutional work: Actors and agency in institutional studies of organizations*. Cambridge university press.
- Lawson, T. (2003). *Reorienting economics*. Psychology Press.
- Lawson, T. (2006). *Economics and Reality*. Routledge.
- Le Grand, J. (2009). *The other invisible hand: Delivering public services through choice and competition*. Princeton University Press.

- Le Moigne, J.-L. (1995). Les épistémologies constructivistes.
- Lendaro, A. (2012). Revisiter l'analyse sociétale aujourd'hui. *Terrains & travaux*, N° 21(2), 109–124.
- Lester, S. M. (2002). The tools of government. A guide to the New Governance. *New York: Oxford University Press*.
- Levitan, S. A., & Wurzburg, G. K. (1979). *Evaluating federal social programs: An uncertain art*. WE Upjohn Inst for.
- Light, P. C. (2008). *The Search for Social Entrepreneurship*. Washington, D.C: Brookings Institution Press.
- Liket, K. C., & Maas, K. (2015). Nonprofit organizational effectiveness: Analysis of best practices. *Nonprofit and Voluntary Sector Quarterly*, 44(2), 268–296.
- Livingstone, D. W. (2018). *The Education-Jobs Gap: Underemployment Or Economic Democracy?* Routledge.
- Lynn, L. (1996). *Public management as art, science, and profession*. Chatham House Pub.
- Lynn, L. (2009). What is a Neo-Weberian State? reflections on a Concept and its Implications. *The NISPAcee Journal of Public Administration and Policy*, 1(2), 17–30.
- Maas, K., & Liket, K. (2011). Social Impact Measurement: Classification of Methods. In R. Burritt, S. Schaltegger, M. Bennett, T. Pohjola, & M. Csutora (Eds.), *Environmental Management Accounting and Supply Chain Management* (pp. 171–202). Springer Netherlands.
- MacKenzie, D., & Millo, Y. (2003). Constructing a market, performing theory: The historical sociology of a financial derivatives exchange. *American Journal of Sociology*, 109(1), 107–145.
- MacKenzie, D., Muniesa, F., & Siu, L. (2007). *Do economist make markets? On the performativity of economics*. Princeton University Press, Princeton.
- Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan*, 80(9), 688.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (2012). *Evaluation models: Viewpoints on educational and human services evaluation* (Vol. 6). Springer Science & Business Media.
- Mair, J., & Martí, I. (2006). Social entrepreneurship research: A source of explanation, prediction, and delight. *Journal of World Business*, 41(1), 36–44.
- Marianthi Dunn, Akritidis, L., & Biedma, L. (2014). The impact of ESA 2010 on key indicators of the national accounts in Europe. *Eurostat Review on National Accounts and Macroeconomic Indicators -EURONA*, (Special Issue on the Implementation of the European System of Accounts (ESA 2010)).
- Marsh, D., & Rhodes, R. A. W. (1992). *Policy networks in British government*. Clarendon Press.
- Mason, C. (2012). Isomorphism, social enterprise and the pressure to maximise social benefit. *Journal of Social Entrepreneurship*, 3(1), 74–95.
- Mathison, S. (2004). *Encyclopedia of evaluation*. Sage publications.
- Mathison, S. (2011). Internal evaluation, historically speaking. *New Directions for Evaluation*, 2011(132), 13–23.
- Mauger, G. (2001). Les politiques d'insertion. *Actes de La Recherche En Sciences Sociales*, (1), 5–14.
- Maurice, M. (1989). Méthode comparative et analyse sociétale: les implications théoriques des comparaisons internationales. *Sociologie Du Travail*, 175–191.
- Maurice, M., & Sorge, A. (2000). *Embedding Organizations: Societal analysis of actors, organizations and socio-economic context* (Vol. 4). John Benjamins Publishing.
- McGoun, A. F. (1936). Observation in economics. *Canadian Journal of Economics and Political Science/Revue Canadienne de Economiques et Science Politique*, 2(2), 128–142.

- Mearman, A. (2006). Critical realism in economics and open-systems ontology: A critique. *Review of Social Economy*, 64(1), 47–75.
- Méda, D. (1999). *Qu'est-ce que la richesse*. Paris, Aubier.
- Medema, S. G. (2004). Mill, Sidgwick, and the evolution of the theory of market failure. Retrieved August, 18, 2007.
- Metcalf, L., & Richards, S. (1990). *Improving public management* (Vol. 1). Sage.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Millar, R., & Hall, K. (2012). Social Return on Investment (SROI) and Performance Measurement. *Public Management Review*, 15(6), 923–941.
- Milward, H. B., & Provan, K. (2003). Managing the hollow state Collaboration and contracting. *Public Management Review*, 5(1), 1–18.
- Milward, H. B., & Provan, K. G. (2000). Governing the hollow state. *Journal of Public Administration Research and Theory*, 10(2), 359–380.
- Miringoff, M. (1995). Index of social health. *Institute for Innovation in Social Policy, Fordham University, Tarrytown, New York*.
- Mironova, A., & Bovaird, T. (1999). Performance management in the Soviet union: A conceptual framework. *RESEARCH PAPER SERIES-ASTON BUSINESS SCHOOL RESEARCH INSTITUTE*.
- Mitchell, G. E. (2013). The construct of organizational effectiveness: Perspectives from leaders of international nonprofits in the United States. *Nonprofit and Voluntary Sector Quarterly*, 42(2), 324–345.
- Mitchell, R. C., & Carson, R. T. (2013). *Using surveys to value public goods: the contingent valuation method*. Rff Press.
- Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard university press.
- Moore, M. H. (2003). The public value scorecard: a rejoinder and an alternative to 'strategic performance measurement and management in non-profit organizations' by Robert Kaplan.
- Moore, M. H. (2007). Recognising public value: The challenge of measuring performance in government. *A Passion for Policy*, 91–116.
- Moore, M. H. (2014). Public value accounting: Establishing the philosophical basis. *Public Administration Review*, 74(4), 465–477.
- Moore, M. H., & Khagram, S. (2004). On creating public value: What business might learn from government about strategic management. *Corporate Social Responsibility Initiative Working Paper*, 3.
- Mordhorst, M. (2005). The Co-Operative Movement in Denmark, between National Identity and Globalization." In *Corporate Images-Images of the Corporation*", *European Business History Conference, Frankfurt am Main, Tyskland* (pp. 1–3).
- Moulaert, F. (2013). *The international handbook on social innovation: collective action, social learning and transdisciplinary research*. Edward Elgar Publishing.
- Mulgan, G. (2010). Measuring Social Value, in "Stanford Social Innovation Review. Summer.
- Nicholls, A. (2010). The legitimacy of social entrepreneurship: reflexive isomorphism in a pre-paradigmatic field. *Entrepreneurship Theory and Practice*, 34(4), 611–633.
- Nicholls, J. (2007). Why measuring and communicating social value can help social enterprise become more competitive. *London: Cabinet Office, Office of the Third Sector*.
- Nordhaus, W. D., & Tobin, J. (1972). Is growth obsolete? In *Economic Research: Retrospect and prospect, Volume 5, Economic growth* (pp. 1–80). Nber.

- Norman, W., & MacDonald, C. (2004). Getting to the bottom of “triple bottom line.” *Business Ethics Quarterly*, 243–262.
- Nyssens, M. (2006). *Social Enterprise: At the Crossroads of Market, Public Policies and Civil Society* (New Ed edition). New York, NY: Routledge.
- Nyssens, M., & Defourny, J. (2012). *The EMES approach of social enterprise in a comparative perspective*. EMES, Working Papers.
- O’Connor, J. (2006). The fiscal crisis of the state. *The Welfare State Reader*, 62–65.
- OECD. (2015a). *Policy Brief on social impact measurement for social enterprises*. Paris: OECD Publishing.
- OECD. (2015b). *Social Impact Investment: Building the Evidence Base*. Paris: OECD Publishing.
- O’Flynn, J. (2007). From new public management to public value: Paradigmatic change and managerial implications. *Australian Journal of Public Administration*, 66(3), 353–366.
- Ogien, A. (2010). La valeur sociale du chiffre. *Revue Française de Socio-Économie*, (1), 19–40.
- Oliver, C. (2011). Critical realist grounded theory: A new approach for social work research. *British Journal of Social Work*, 42(2), 371–387.
- Olsen, J. P. (2010). *Governing through institution building: Institutional theory and recent European experiments in democratic organization*. OUP Oxford.
- Olsen, S., & Galimidi, B. (2008). Catalog of Approaches to Impact Measurement - Assessing social impact in private ventures. *The Rockefeller Foundation*.
- Olsen, W. (2007). Critical realist explorations in methodology. *Methodological Innovations Online*, 2(2), 1–5.
- Ongaro, E. (2009). *Public management reform and modernization: Trajectories of administrative change in Italy, France, Greece, Portugal and Spain*. Edward Elgar Publishing.
- Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, 8(5), 375–387.
- Osberg, L., & Sharpe, A. (1999). *An Index of Economic Well-Being for Canada and the United States*. Presented at the Annual Meeting of the American Economic Association.
- Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming government*. Reading Mass. Adison Wesley Public Comp.
- Osborne, S. P. (2006). *The new public governance?* Taylor & Francis.
- Osborne, S. P. (2010). *The new public governance: Emerging perspectives on the theory and practice of public governance*. Routledge.
- Ostrom, V., & Ostrom, E. (1971). Public choice: A different approach to the study of public administration. *Public Administration Review*, 31(2), 203–216.
- Page, S. B., Stone, M. M., Bryson, J. M., & Crosby, B. C. (2015). Public value creation by cross-sector collaborations: A framework and challenges of assessment. *Public Administration*, 93(3), 715–732.
- Pahk, S. (2017). Misappropriation as market making: Butler, Callon, and street food in San Francisco, California. *Journal of Cultural Economy*, 10(3), 296–308.
- Painter, M., & Peters, B. G. (2010). Administrative Traditions in Comparative Perspective: Families, Groups and Hybrids. In M. Painter & B. G. Peters (Eds.), *Tradition and Public Administration* (pp. 19–30). London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230289635_2
- Parr, S. (2015). Integrating critical realist and feminist methodologies: Ethical and analytical dilemmas. *International Journal of Social Research Methodology*, 18(2), 193–207.
- Patton, M. Q. (2005). *Qualitative research*. Wiley Online Library.

- Pearce, J., & Kay, A. (2012). Brief History of Social Accounting and Audit.
- Perret, B. (2006). De l'échec de la rationalisation des choix budgétaires (RCB) à la loi organique relative aux lois de finances (LOLF). *Revue Française d'administration Publique*, (1), 31–41.
- Pestoff, V. A. (2008). *A democratic architecture for the welfare state*. Routledge.
- Pestoff, V. A., Brandsen, T., & Verschuere, B. (2013). *New public governance, the third sector, and co-production*. Routledge.
- Pestoff, V. A., Osborne, S. P., & Brandsen, T. (2006). Patterns of co-production in public services: Some concluding thoughts. *Public Management Review*, 8(4), 591–595.
- Petrella, F., & Richez-Battesti, N. (2014). Social entrepreneur, social entrepreneurship and social enterprise: semantics and controversies. *Journal of Innovation Economics & Management*, (2), 143–156.
- Petrella, F., & Richez-Battesti, N. (2016). Patterns of social enterprises in France: a focus on work integration social enterprises. *International Review of Sociology*, 26(2).
- Petty, S. W. (1691). *Verbum sapienti*. London.
- Pierre, J., & Peters, G. B. (2000). Governance, politics and the state.
- Piketty, T. (2013). *Le capital au XXI^e siècle*. Le Seuil.
- Poister, T. H., Levitan, S. A., Wurzburg, G., Abert, J. G., Sobell, L. C., Sobell, M. B., ... others. (1981). *Program Evaluation: State of the Art/State of the Practice*. JSTOR.
- Pollitt, C. (1993). Occasional excursions: a brief history of policy evaluation in the UK. *Parliamentary Affairs*, 46(3), 353–363.
- Pollitt, C. (2009). An Overview of the Papers and Propositions of the First Trans-European Dialogue (TED1). *The NISPAcee Journal of Public Administration and Policy*, 1(2), 9–16.
- Pollitt, C. (2016). Managerialism redux? *Financial Accountability & Management*, 32(4), 429–447.
- Pollitt, C., & Bouckaert, G. (2004). *Public Management Reform: A comparative analysis-new public management, governance, and the Neo-Weberian state* (Second). Oxford University Press.
- Pollitt, C., & Bouckaert, G. (2017). *Public Management Reform: A comparative analysis-new public management, governance, and the Neo-Weberian state* (Fourth). Oxford University Press.
- Pollitt, C., & Hupe, P. (2011). Talking about government: The role of magic concepts. *Public Management Review*, 13(5), 641–658.
- Powell, W. (2003). Neither market nor hierarchy. *The Sociology of Organizations: Classic, Contemporary, and Critical Readings*, 315, 104–117.
- Power, M. (1999). The audit society: rituals of verification. *BRITISH JOURNAL OF EDUCATIONAL STUDIES*, 47, 92–93.
- Public Services (Social Value) Act 2012. (n.d.). Retrieved August 20, 2018, from <http://www.legislation.gov.uk/ukpga/2012/3/enacted>
- Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223–234). Springer.
- Ranerup, A., & Norén, L. (2015). How are citizens' public service choices supported in quasi-markets? *International Journal of Information Management*, 35(5), 527–537.
- Reynaud, J.-D. (2011). Marc Maurice (1924–2011). *Sociologie Du Travail*, 53(2), 155–159. <https://doi.org/10.1016/j.sotra.2011.03.001>
- Rhodes, R. A. (1997). *Understanding governance: Policy networks, governance, reflexivity and accountability*. Open university press.
- Rhodes, R. A., & Wanna, J. (2007). The limits to public value, or rescuing responsible government from the platonic guardians. *Australian Journal of Public Administration*, 66(4), 406–421.

- Riccucci, N. M. (2001). The “old” public management versus the “new” public management: Where does public administration fit in? *Public Administration Review*, 61(2), 172–175.
- Richez-Battesti, N. (2006). Evaluer la production associative par les pouvoirs publics : du contrôle à la coproduction des critères ? In *De l'intérêt général à l'utilité sociale: La reconfiguration de l'action publique entre Etat, associations et participation citoyenne*. (pp. 115–128). L'Harmattan.
- Ridley-Duff, R., & Bull, M. (2015). *Understanding social enterprise: Theory and practice*. Sage.
- Rinaldo, H. (2010). *Getting started in social impact measurement: A guide to choosing how to measure social impact*. Norwich: The Guild. Retrieved from https://www.socialauditnetwork.org.uk/files/8113/4996/6882/Getting_started_in_social_impact_measurement_-_270212.pdf
- Roberts, D., & Woods, C. (2005). Changing the world on a shoestring: The concept of social entrepreneurship. *University of Auckland Business Review*, 7(1), 45–51.
- Rose, M. (1985). Universalism, culturalism and the Aix group: promise and problems of a societal approach to economic institutions. *European Sociological Review*, 1(1), 65–83.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach*. Sage publications.
- Royer, I., & Zarlowski, P. (2014). Echantillon(s). In *Méthodes de recherche en management-4ème édition*. Dunod.
- Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.
- Ruff, K. Y. (2017). Materializing Performance: The Interactions that Enact Inclusions, Exclusions and Arrangements in Charity Social Performance Reports.
- Sagawa, S., & Segal, E. (2000). *Common Interest, Common Good: Creating Value Through Business and Social Sector Partnerships*. Harvard Business Press.
- Sahlin-Anderson, K., & Engwall, L. (2002). Carriers, flows and sources of management knowledge. *The Expansion of Management Knowledge*, 3Y32.
- Saint-Martin, D. (2005). The politics of management consulting in public sector reform. *Handbook of Public Management*, 84–106.
- Salais, R. (2010). Usages et mésusages de l'argument statistique: le pilotage des politiques publiques par la performance. *Revue Française Des Affaires Sociales*, (1), 129–147.
- Salamon, L. M., & Sokolowski, S. W. (2016). The size and scope of the European third sector. *Brussels: European Union FP7 (Grant Agreement 613034). Third Sector Impact*.
- Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17), 61–71.
- Sauer, B. (2000). Conceptualizing the German State. *Handbook of Global Social Policy*, 317.
- Sayer, A. (2004). Why critical realism? *Critical Realist Applications in Organisation and Management Studies*, 6–20.
- Schober, C., & Rauscher, O. (2014). Was ist Impact. *Gesellschaftliche Wirkungen von (Nonprofit) Organisationen. Von Der Identifikation Über Die Bewertung Bis Zu Unterschiedlichen Analyseformen*“, WU Vienna University of Economics and Business Working Paper.
- Scriven, M. (1991). *Evaluation thesaurus*. Sage.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation*, 2(4), 393–404.
- Shafritz, J., Krane, D., & Wright, D. S. (1998). *International encyclopedia of public policy and administration*. Westview Press.
- Siegel, S. (1957). Nonparametric statistics. *The American Statistician*, 11(3), 13–19.

- Silverman, D. (2006). *Interpreting qualitative data: Methods for analyzing talk, text and interaction*. Sage.
- Smulders, S., & Gradus, R. (1996). Pollution abatement and long-term growth. *European Journal of Political Economy*, 12(3), 505–532.
- Sørensen, E. (2002). Politikerne og netværksdemokratiet.
- Sorokin, P. A. (1961). Fads and foibles in modern sociology and related sciences.
- Stake, R. E. (1974). Nine approaches to educational evaluation. *Unpublished Chart*. Urbana IL: University of Illinois, Center for Instructional Research and Curriculum Evaluation.
- Stake, Robert E. (2013). *Multiple case study analysis*. Guilford Press.
- Starr, M. A. (2014). Qualitative and mixed-methods research in economics: surprising growth, promising future. *Journal of Economic Surveys*, 28(2), 238–264.
- State Analysis: How Many Unemployed Have “Given Up”? (2017, August). Retrieved July 19, 2018, from <https://www.expresspros.com/Newsroom/America-Employed/11-State-Analysis-How-Many-Unemployed-Have-Given-Up.aspx>
- Stievenart, E., & Pache, A.-C. (2014). Evaluer l’impact social d’une entreprise sociale: points de repère. *Revue Internationale de l’économie Sociale: Recma*, (331), 76–92.
- Stiglitz, J. E., Sen, A., Fitoussi, J.-P., & others. (2010). Report by the commission on the measurement of economic performance and social progress. *Paris: Commission on the Measurement of Economic Performance and Social Progress*.
- Stoker, G. (2006). Public value management: A new narrative for networked governance? *The American Review of Public Administration*, 36(1), 41–57.
- Studer, M. (2017). *L’impact social, objet établi ou projet à co-construire ?* Université de Lille, Lille.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation*, 2001(89), 7–98.
- Teicher, J., Alam, Q., & Gramberg, B. V. (2006). Managing trust and relationships in PPPs: some Australian experiences. *International Review of Administrative Sciences*, 72(1), 85–100.
- The Center for High Impact Philanthropy. (n.d.). How we calculated cost per impact in education.
- Thuesen, F., Bach, H., Albæk, K., Jensen, S., Hansen, N., & Weibel, K. (2013). *Socialøkonomiske virksomheder i Danmark*. Copenhagen: SFI – Det Nationale Forskningscenter For Velfærd.
- Triponel, A., & Agapitova, N. (2016). *Legal Frameworks for Social Enterprises - Lessons from a comparative study*. Washington, D.C: World Bank Group.
- Tuan, M. T. (2008). Measuring and/or estimating social value creation: Insights into eight integrated cost approaches. *Final Paper*. Bill and Melinda Gates Foundation. Seattle, Washington.
- UNEP, U.-I. (2012). *Inclusive wealth report 2012*. New York: Cambridge University Press.
- United Nations. (1993). System of National Accounts 1993. Department for Economic and Social Information and Policy Analysis - Statistics Division.
- United Nations Development Programme (UNDP). (1990). *Human Development Report 1990*. United Nations Development Programme (UNDP).
- Ura, K., Alkire, S., Zangmo, T., & Wangdi, K. (2012). *A short guide to gross national happiness index*. The Centre for Bhutan Studies.
- Vanclay, F. (2003). International principles for social impact assessment. *Impact Assessment and Project Appraisal*, 21(1), 5–12.
- Vardaskoulis, O. (2013). Economics in policy-making -Social CBA and SROI. New Economics Foundation.
- Vatin, F. (2009). *Évaluer et valoriser: une sociologie économique de la mesure*. Presses Univ. du Mirail.

- Veggeland, N. (2007). *Paths of Public Innovation in the Global Age: Lessons from Scandinavia*. Edward Elgar Publishing.
- Vollmer, H., Mennicken, A., & Preda, A. (2009). Tracking the numbers: Across accounting and finance, organizations and markets. *Accounting, Organizations and Society*, 34(5), 619–637.
- Wackernagel, M., & Rees, W. (1998). *Our ecological footprint: reducing human impact on the earth* (Vol. 9). New Society Publishers.
- Walbesser, H. H., & Eisenberg, T. A. (1972). A Review of Research on Behavioral Objectives and Learning Hierarchies.
- Walras, L. (1896). *Éléments d'économie politique pure, ou, Théorie de la richesse sociale*. F. Rouge.
- Walras, L. (1926). *Eléments d'Economie Politique Pure ou Théorie de la Richesse Sociale* (R. Pichon et R. Durand-Auzias). Paris.
- Wang, B., & Christensen, T. (2017). The open public value account and comprehensive social development: An assessment of china and the United States. *Administration & Society*, 49(6), 852–881.
- Weber, M. (1965). *Politics as a Vocation*. Fortress Press Philadelphia, PA.
- Williams, I., & Shearer, H. (2011). Appraising public value: Past, present and futures. *Public Administration*, 89(4), 1367–1384.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage Publications.
- Yeung, H. W. (1997). Critical realism and realist research in human geography: a method or a philosophy in search of a method? *Progress in Human Geography*, 21(1), 51–74.
- Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly*, 58–65.
- Yin, R. K. (2009). *Case study research: Design and methods*. Sage publications.
- Young, D. R., & Salamon, L. M. (2002). Commercialization, social ventures, and for-profit competition. *The State of Nonprofit America*, 423–446.
- Yunus, M. (2006). Social business entrepreneurs are the solution. *Social Entrepreneurship: New Models of Sustainable Social Change*, 39–44.
- Zachariadis, M., Scott, S., & Barrett, M. (2010). Exploring critical realism as the theoretical foundation of mixed-method research: evidence from the economics of IS innovations. *Judge Business School Working Papers*.
- Zadek, S., Evans, R., & Pruzan, P. (1997). *Building Corporate Accountability: Emerging Practice in Social and Ethical Accounting and Auditing*. London: Routledge.
- Zappalà, G., & Lyons, M. (2009). *Recent approaches to measuring social impact in the Third sector: An overview*. Centre for Social Impact Sydney.
- Zeckhauser, R., & Stokey, E. (1978). *A primer for policy analysis*. WW Norton.

Indexes of tables, figures and appendices

Index of tables

Table 1: Administrative profiles in Europe	54
Table 2: Main PAM models used in our work.....	57
Table 3: Breaking down the triangle	75
Table 4: Proposed definitions for social impact assessment.....	122
Table 5: Social impact assessment models applicable to WISEs	133
Table 6: key performance models attributes and variables and their contributors	134
Table 7: Paradigmatic dimensions	157
Table 8: Overall qualitative case design.....	170
Table 9: Qualitative sampling methodology	171
Table 10: Summary of interviews schedule	173
Table 11: Coding examples	175
Table 12: Survey population and sample data	180
Table 13: Overview of case studies.....	182
Table 14: FR2 revenue break down	187
Table 15: DK2 revenue breakdown.....	192
Table 16: Main social impact assessment-related reports referenced by Avise	197
Table 17: Layers of authority: the French sub-national union	208
Table 18: The weight of different entities in ESS.....	213
Table 19: Summary of the four main work integration covenants.....	217
Table 20: Modulation criteria	218
Table 21: Main institutional milestones for the social economy in Denmark.....	232
Table 22: Social enterprises census in Denmark.....	234
Table 23: Reported benefits of SIA in France	247
Table 24: FR1's performance measurement and SIA.....	251
Table 25: FR2's performance measurement and SIA.....	254
Table 26: Summary of strategic interactions in the French NWS.....	265
Table 27: Selected features of SIA models	271

Table 28: DK1's performance measurement and SIA	278
Table 29: DK2's performance measurement and SIA	286
Table 30: Summary of strategic interactions in the Danish NPG.....	300
Table 31: Summary of encountered performance measurement approaches*	308
Table 32: Strategic interactions paradigms in France and in Denmark	319
Table 33: Three states of socio-technical arrangements.....	337
Table 34: DK1's SIA structure and process.....	423
Table 35: The competences assessed by DK1's SIA model	423

Index of figures

Figure 1: Summarised thesis content	20
Figure 2: Key terminology	34
Figure 3: Menus and dishes in PAM.....	56
Figure 4: The PV Triangle, as traditionally represented.....	68
Figure 5: Performance in PAM	101
Figure 6: The four performance regimes	107
Figure 7: Performance measurement encompasses social impact assessment	116
Figure 8: The impact value chain	128
Figure 9: Selection method	132
Figure 10: Analytical framework	149
Figure 11: Stratified ontology of critical realism.....	158
Figure 12: The research object in the critical realist approach	160
Figure 13: Logical reasoning.....	161
Figure 14: Triangulation	166
Figure 15: Mixed research design	167
Figure 16: Qualitative case design	169
Figure 17: Structure of FR1	183
Figure 18: Structure of FR2	187
Figure 19: Structure of DK1.....	190
Figure 20: Structure of DK2.....	194

Figure 21: Conceptualising our approach	201
Figure 22: Strategic interactions between Danish WISEs and their public stakeholders.....	237
Figure 23: Ngrams of social impact and social utility	243
Figure 24: Expectations and perceived effects of SIA.....	273
Figure 25: Preferred performance indicators and perceived performance assessment	274
Figure 26: Observed performance regimes	316
Figure 27: The seven strategic intents	322
Figure 28: Two explanatory factors for performativity	339
Figure 29: Feedbacks of socio-technical arrangements	345

Index of Appendices

Appendix 1: Abbreviations and acronyms	393
Appendix 2: A short history of programme evaluation	395
Appendix 3: Selection of six of ten descriptors provided by Stufflebeam (2001)	398
Appendix 4: 22 evaluation approaches as categorised by Stufflebeam (2001)	402
Appendix 5: Selected program evaluation models.....	404
Appendix 6: Selected alternative indicators	406
Appendix 7: Social impact assessment state of the art - 124 selected models.....	408
Appendix 8: Description of the selected social impact assessment methods for WISEs	412
Appendix 9: Detailed interviews schedule.....	417
Appendix 10: Overview of classification attributes used in social impact assessment literature.....	420
Appendix 11: Retained SIA attributes, variables and SIA models.....	422
Appendix 12: DK1's SIA components	423
Appendix 13: DK2PS evaluation framework	424
Appendix 14: FR1 SIA components	425
Appendix 15: FR2's SIA components.....	427
Appendix 16: Coding grid	430
Appendix 17: Example of interview guide (semi-structured interviews)	435
Appendix 18: Questionnaire used for the quantitative survey (English version).....	441
Appendix 19: Example of interview transcript	452

Appendices

Appendix 1: Abbreviations and acronyms

Acronym	Meaning (original language)
ACDC	The other unemployment figures (Les autres chiffres du chômage)
ACI	Integration workshops and construction sites (ateliers et chantiers d'insertion)
AI	Intermediary association (association intermediaire)
ANPE	Job Center -- former name (Agence Nationale Pour l'Emploi)
ASP	Payment services agency (Agence de services et de paiement)
CEO	Chief Executive Officer
CNCRESS	National Council of the Regional Chambers of Social and Solidary Economy (conseil national des chambres régionales de l'économie sociale et solidaire)
CNIAE	National Council for Work Integration (Conseil national de l'IAE)
CR	Critical Realism / Critical Realist
CRESS	Regional Chamber of Social and Solidary Economy (chambre régionale de l'économie sociale et solidaire)
CSESS	High Council of Social and Solidary Economy (haut conseil de l'économie sociale et solidaire)
CSR	Corporate Social Responsibility
DGEFP	General Delegation for Employment and Vocational Training (Délégation générale à l'emploi et à la formation professionnelle)
DGEFP	Employment and Vocational Training Authority (Délégation générale à l'emploi et à la formation professionnelle)
Direccte	Regional Authority for Business, Competition, Consumption, Labour and Employment (Directions régionales des entreprises, de la concurrence, de la consommation, du travail et de l'emploi)
EI	Work integration enterprises (entreprise d'insertion)
ERP	Enterprise Resources Planning system
ESEA	Elementary and Secondary Education Act
ESS	Social and Solidary Economy (économie sociale et solidaire)
ETTI	(entreprise de travail temporaire d'insertion)
EU	European Union
FTE	Full Time Equivalent
GDP	Gross Domestic Product
Geiq	Grouping of Employers for Integration and Qualification (groupement d'employeurs pour l'insertion et la qualification)
GES	Grouping of social enterprises (Groupement Economique Solidaire)
FR1	FR1
HDI	Human Development Index
HPI	Happy Planet Index
IAE	Work integration (insertion par l'activité économique)
IMF	International Monetary Fund

Acronym	Meaning (original language)
LOLF	Act on Budget Legislation (Loi organique relative aux lois de finance Institutional)
MAP	Modernisation of Public Action (modernisation de l'action publique)
NPG	New Public Governance
NPM	New Public Management
NWS	Neo-Weberian State
OECD	Organisation of Economic Cooperation and Development
PA	Public Administration
PAM	Public Administration and Management
PE	Programme Evaluation
PVT	Public Value Theory
RGPP	General Public Policy Review (Révision Générale des Politiques Publiques)
SAA	Social Accounting and Audit
SE	Social Enterprises
SED	Social Entrepreneurs in Denmark (Sociale entreprenører i Danmark)
SIM	Social Impact Measurement
SNA	System of National accounts
SROI	Social Return On Investment
STA	Socio-Technical Arrangement
STU	School for pupils with special needs (Særligt tilrettelagt ungdomsuddannelse)

Appendix 2: A short history of programme evaluation

The history of programme evaluation can be challenging to piece together. To this extent, Hogan (2010, p.3) highlights that “historical development of evaluation is difficult, if not impossible, to describe due to its informal utilization by humans for thousands of years”. Besides the timeframe issue, Rossi et al. (2004, p.8) acknowledge the existence of “somewhat different historical perspectives” on the matter. With this in mind, we acknowledge that the historical summary below is an attempt to digest a thorough review of the literature, but should not be considered a historian’s take on programme evaluation.

Programme evaluation’s early days are often traced back to the late 1960’s, which in the USA correspond to the period of development of Lyndon Johnson’s “Great Society” and “War on Poverty”. At the time, the Lyndon administration had an Office of Research, Plans Programs, and Evaluations dedicated to programme evaluation and is recognised for having pioneered the use of evaluation on social policy (Caro, 1977). These early American practices, later formulated under the acronym of PPBS (for Planning Programming Budgeting System) are credited as the inspiration source for the introduction of similar evaluation mechanisms (and perhaps for a passion for acronyms) in countries such as France in 1968 with the RCB (“rationalisation of budgetary choices”¹⁶⁸) (Perret, 2006) or the UK in 1970 with the CPRS (Central Policy Review Staff) (Pollitt, 1993).

However, the practice of evaluation itself finds its roots much earlier in history. As previously highlighted, Hogan points to the fact that evaluation has been informally used for thousands of years (2010). Similarly, Scriven tells us that “evaluation is a very young discipline – although it is a very old practice” (1996, p. 395). While Cronbach and associates (1980) are referring to Hobbes’ attempts to assess populations’ social condition in the 17th century as a practice that could be assimilated to early evaluation, Madaus et al. (2012) place the first milestone of evaluation history in 1792, with the introduction of quantitative marks to score examinations in the education sector. Although this may appear as a very incremental innovation, this practice sets the foundation for the field of psychometric, a cornerstone around which a discipline of testing and evaluation would be built (Madaus & O’Dwyer, 1999).

¹⁶⁸ Rationalisation des Choix Budgétaires

From this milestone, and until the 20th century, most of the major steps attached to the history of evaluation praxis as reported by Hogan (2010); Madaus & O'Dwyer, (1999) or Stufflebeam (2001) are related to practices in the education sector. Among the developments of the 19th century, a distinctive trend, pioneered in 1845 in Boston USA (Madaus et al., 2012), consisted in using tests scores to measure education programmes effectiveness as a whole (rather than school-specific or student-specific performance). This is where evaluation started to drift from the sole purpose of assessing education outcomes towards more policy-centred concerns.

The practice was refined over time by key contributors such as Joseph Rice in the 1880s (Colwell, 1998) or Ralph Tyler in the 1930' (Walbesser & Eisenberg, 1972) and even took cues from management concepts such as the ones developed by Fredrick Taylor (Hogan, 2010). Building on Tyler's approach of internally comparing outcomes to program objectives (rather than using economically and socially expensive control groups), a growing number of evaluation methods were developed and tested in the 1950s and 1960s (see for instance Samuel Bloom's Taxonomy of Educational Objectives, which is a distinctive milestone of this era). In the meantime, evaluation became increasingly common in a growing number of fields such as social policies, health and even defence (Rossi et al., 2003).

It is after these advancements that the 1960s policy developments described above occurred. Several pieces of legislations mandating evaluation were passed early in the decade (e.g. the Juvenile Delinquency Program of 1962 and the Manpower Development and Training Act of 1962) (Mathison, 2004). The Elementary and Secondary Education Act (ESEA) in 1965 is thus usually recognised in the literature as the fundamental landmark that pushed evaluation in the legislative framework (Altschuld & Engle, 2015; Mathison, 2011), and is considered to be the "birth of the contemporary program evaluation" (Hogan, 2010, p.6).

From this period, the field of evaluation became a discipline in its own right. The late 1960s marked the widespread adoption of programme evaluation, especially among OECD countries. This sudden enthusiasm for evaluation research is often associated with the development of the welfare state (Hogan, 2010; Madaus et al., 2012; Mathison, 2011). After World War 2, nations across the world

(particularly in the West) engaged in a wide array of social and economic activities they had not been involved with until then. This increased state activity, together with increased expenses and associated tax collection created an information void about the usage of these resources (Albæk, 1998). Facing these interrogations, evaluation soon became a privileged instrument to fill that void with information on the efficiency and effectiveness of the new public activities (Albæk & Rieper, 2002).

Appendix 3: Selection of six of ten descriptors provided by Stufflebeam (2001)

1 Advanced organiser	2 Main purpose(s) observed	4 Characteristic questions	5 Methods typically employed	9 Strengths of the approach	10 Weaknesses of the approach
Programme content/definition rationale	Program Determine whether program objectives were achieved	To what extent was each program objective achieved?	Operational objectives	Common sense appeal	May credit unworthy objectives
Context	Provide constituents with an accurate accounting of results	Did the program effectively discharge its responsibilities?	Criterion-referenced tests	Widely known & applied	May define a program's success in terms that are too narrow and mechanical and not attuned to beneficiaries' various needs
Treatments	Assure that results are positive	Did tested performance meet or exceed pertinent norms?	Performance contracting	Employs operational objectives	May employ only lower-order learning objectives
Time period	Assess learning gains	Did tested performance meet or exceed standards	Program Planning & Budgeting System	Employs the technology of testing	Relies almost exclusively on multiple choice test data
Beneficiaries	Pinpoint responsibility for good & bad outcomes	Where does a group's tested performance rank compared with other groups?	Program Evaluation & Review Technique	Efficient use of standardized tests	May indicate mainly socio-economic status, not quality of teaching & learning
Comparison groups	Compare students' test scores to norms	Is a group's present performance better than past performance?	Management by objectives Staff progress reports	Popular among constituents & politicians	May reinforce & over-emphasize multiple choice test taking ability to the exclusion of writing, speaking, etc.
Norm groups	Compare students' test performance to standards	What sectors of a system are performing best and poorest?	Financial reports & audits	Focus on improving public services	May poorly test what teachers teach
Assessed needs	Diagnose program shortcomings	Where are the shortfalls in specific curricular areas?	Zero Based Budgeting	Can focus on audience's most important questions	Yields mainly terminal information that lacks utility for program improvement
Problem statements	Compare performance	At what grade levels are the strengths & shortfalls?	Cost analysis, cost-effectiveness analysis, & benefit-cost analysis	Defines obligations of service providers	Provides data only on student outcomes
Objectives	Examine achievement trends	What value is being added by particular programs?	Mandated "program drivers" & indicators	Requires production of and reporting on positive outcomes	Narrow scope of skills that can feasibly be assessed
Independent/dependent variables	Inform policymaking	To what extent can students effectively speak, write, figure, analyze, lead, work cooperatively, & solve problems?	Input, process, output databases	Seeks to improve services through competition	May provide too narrow an information basis for judging a program's merit & worth

1 Advanced organiser	2 Main purpose(s) observed	4 Characteristic questions	5 Methods typically employed	9 Strengths of the approach	10 Weaknesses of the approach
Indicators/criteria	Direction for program improvement	What are a program's effects on outcomes?	Independent goal achievement auditors	Efficient means of data collection	May employ many methods because it is the thing to do rather than because they are needed
Life skills	Ensure standardization of outcome measures	Are program activities being implemented according to schedule, budget, & expected results?	Procedural compliance audits	Stress on validity & reliability	May inappropriately & counterproductively mix positivistic & post-modern paradigms
Performance tasks	Determine cause and effect relationships in programs	What is the program's return on investment?	Peer review Merit pay for individuals and/or organizations	Triangulates findings from multiple sources	May oversimplify the complexities involved in assigning responsibility for student learning gains to individual teachers
Questions/hypotheses/causal factors	Inform management decisions & actions	Is the program sustainable & transportable?	Collective bargaining agreements	Uses institutionalized database	May miss important side effects
Policy issues	Assess investments and payoffs	Is the program worthy of continuation and/or dissemination?	Trial proceedings	Monitors progress on each student	May rely too heavily on the expertise & judgment of a single evaluator
Tests in use	Provide balanced information on strengths & weaknesses	Is the program as good or better than others that address the same objective?	Mandated testing	Emphasizes service to every student	May issue invidious comparisons
Formative & summative evaluation	Explicate & illuminate a program	What is the program in concept & practice?	Institutional report cards	Hierarchical analysis of achievement	May produce unhealthy competition
Qualitative & quantitative methods	Describe & critically appraise a program	How has the program evolved over time?	Self-studies	Conducive to policy analysis	May provoke political unrest
Program activities/milestones	Assess a programme's theoretical soundness	How does the program produce outcomes?	Site visits by experts	Employs trend analysis	Accuracy suffers in the face of competing evaluations
Employee roles & responsibilities		What has the program produced?	Program audits	Strong provision for analyzing qualitative information	May undesirably narrow the range of program services
Costs		What are the program's shortfalls & negative side effects?	Standardized testing	Rejects use of artificial cut scores	Politicians tend to press for premature implementation
Evaluator expertise & sensitivities		What are the program's positive side effects?	Performance measures	Considers student background by using students as their own controls	Granting rewards & sanctions may produce cheating
Intra-case/cross-case anal		How do various stake-holders value the program?	Computerized or other database	Considers contextual influences	Inordinate time requirements for administration & scoring

1 Advanced organiser	2 Main purpose(s) observed	4 Characteristic questions	5 Methods typically employed	9 Strengths of the approach	10 Weaknesses of the approach
		Did the program meet all the beneficiaries' needs?	Hierarchical mixed model analysis	Uses authentic measures	High costs of scoring
		What were the most important reasons for the program's success or failure?	Policy analysis	Eliminates guessing	Difficulty in achieving reliability
		What are the program's most important unresolved issues?	Experimental & quasi-experimental designs	Reinforces life skills	High cost
		How much did the program cost?		Focuses on outcomes	Low feasibility
		What were the costs per beneficiary, per year, etc.?		Focuses on a program's strengths & weaknesses	May inappropriately deprive control group subjects of entitlements
		What parts of the program were successfully transported to other sites?		Determines cause & effects	Carries a connotation of experimenting on children or other subjects using unproven methods
		What are the program's essence & salient characteristics?		Examines program's internal workings & how it produces outcomes	Requirement of random assignment is often not feasible
		What merits & demerits distinguish the program from similar programs?		Guides program management	Tend to stifle continual improvement of the program
		Is the program grounded in a validated theory?		Helps keep programs on track	Vital data may be inaccessible to evaluators
		Are program operations consistent with the guiding theory?		Guides broad study & improvement of program processes & outcomes	Investigators may mistake the approach's openness & lack of controls as license to ignore rigor
		Were hypothesized causal linkages confirmed?		Can be done retrospectively or in real time	Evaluators might usurp the program staff's responsibility for program design
		What changes in the program's design or implementation might produce better outcomes?		Documents costs of program inputs	Might ground an evaluation in a hastily developed, inadequate program theory
		What program features are essential for successful replication?		Maintains a financial history for the program	Might develop a conflict of interest to defend the

1 Advanced organiser	2 Main purpose(s) observed	4 Characteristic questions	5 Methods typically employed	typically	9 Strengths of the approach	10 Weaknesses of the approach
						evaluation-generated program theory
		What interesting stories emerged?			Contrasts program alternatives on both costs & outcomes	Might bog down the evaluation in a seemingly endless process of program theory development
					Employs rules of evidence	Might create a theory early in a program and impede the program from redefinition and refinement
					Requires no controls of treatments & participants	
					Examines programs as they naturally occur	
					Examines programs holistically & in depth	
					Engages experts to render refined descriptions & judgements	
					Yields in-depth, refined, effectively communicated analysis	
					Employs all relevant information sources & methods	
					Stresses complementarity of qualitative & quantitative methods	

Appendix 4: 22 evaluation approaches as categorised by Stufflebeam (2001)

Classification	Approach	Description	Typical tools & methods
Pseudo evaluations	Public relations-inspired studies	Intention to use data to convince stakeholders of programme's effectiveness, provides a biased view.	Surveys, Norm table, Narratives (testimonials, anecdotes)
	Politically controlled studies	Studies which conclusions are altered, hidden or the object of selective publication.	Any
Questions and methods oriented (quasi-evaluation)	Objectives-based studies	Typically an internal evaluation to determine whether the programme's objectives have been met.	Performance assessment and tests
	Accountability	Especially relevant for "payment by result" studies. Follows questions-oriented approach, narrowed on outcomes and using external evaluation.	Setting pass / fail standards, audits of procedural compliance and goal achievements, performance-oriented management systems
	Objective-testing programme	Assess the success of individuals pertaining to a programme against pre-defined norms or standards	Standardised tests
	Outcome evaluation	Systematic and recurrent assessment of outcomes (seen as the programme's value-added) through standardised tests and additional processing of tests' data	Standardised tests
	Performance testing	Programme's clients have to perform tasks demonstrating they have acquired what the programme is aiming to equip them with.	Assessment tasks
	Experimental studies	Controlled experiments with randomly assigned beneficiaries to an experimental group and a control group.	Randomised controlled trial
	Managed information systems	Mostly for internal evaluation purpose, follows programme's implementation (focused on processes).	Management systems
	Cost-Benefit analysis	Quantitative procedures to capture the full cost of a programme and compare it to its benefits expressed in monetary value.	Cost-benefit analysis
	Clarification hearing	A judicial approach to programme evaluation, led by a prosecuting evaluator and a defending one.	Trial-like methods (hearing of witnesses, evidence exhibits, etc.)
	Case study evaluations	The programme is not controlled but rather holistically described with great details. It delineates a programme more than it allows to assess it.	Qualitative case study designs
	Criticism and connoisseurship	A highly skilled and specialised evaluator is asked to provide his professional opinion on a programme, not necessarily following any kind of process or method.	N/A
	Programme theory-based evaluation	Aim to answer interrogations related to specific articulation points of the programme theory.	Use of logic models to pinpoint evaluation questions
Mixed-methods studies	Use of mixed methods as a default choice to match audience's expectations, regardless of which approach is best to address the issue.	Qualitative methods (interviews, surveys, cases, etc.) and quantitative methods	

Classification	Approach	Description	Typical tools & methods (statistical indicators and measurements, ad'hoc metrics, etc.)
Improvement accountability oriented	Decision / accountability-oriented studies	<i>Ex-ante</i> and <i>ex-post</i> evaluations carried out independently, mostly in order to improve the programme rather than to prove its efficiency.	Any
	Consumer-oriented studies	Focused on evaluating whether consumers' welfare is increased by a given programme.	Any
	Accreditation / certification approach	Proves that a programme meet a given set of requirements and may identify areas of improvements.	Self-reports and audits.
Social agenda / advocacy	Client-centred studies (or responsive evaluation)	Evaluator works for a heterogonous group of stakeholders (developer, funder, beneficiaries, etc.) involved in the programme. Characterised by relativistic approach seeking no definitive answer.	Action research with a constructivist lens.
	Constructivist evaluation	Paradigm-driven evaluation with subjective epistemology where program's participants are encouraged to help shape the evaluation.	Qualitative methods
	Deliberative democratic evaluation	All stakeholders equally participate in the evaluation process. Findings are debated among stakeholders.	Surveys, interviews, debates.
	Utilization-focused evaluation	Focused on creating an evaluation that makes an impact through tailoring of the evaluation to selected influential stakeholders.	Any

Appendix 5: Selected program evaluation models

We choose to retain 17 of the 22 approaches identified by Stufflebeam (2001). We focus on approaches providing clarity on the methodological choices, as well on the specific questions they aim at answering (i.e. descriptors 2, 4 and 5 in Appendix 3). Therefore, the following approaches are excluded from the selection, as their framing is independent from the methods used or the questions answered:

1. Approaches qualified on the basis of their findings' quality (e.g. the two approaches categorised as "pseudo evaluations" in Appendix 4.
2. Approaches qualified solely on the basis of the stakeholders they are targeting and / or their specific needs (e.g. the mixed-methods studies approach and the utilization-focused approach below).

Approach	Description	Typical tools & methods
Objectives-based studies	Typically an internal evaluation to determine whether the programme's objectives have been met.	Performance assessment and tests
Accountability	Especially relevant for "payment by result" studies. Follows questions-oriented approach, narrowed on outcomes and using external evaluation.	Setting pass / fail standards, audits of procedural compliance and goal achievements, performance-oriented management systems
Objective-testing programme	Assess the success of individuals pertaining to a programme against pre-defined norms or standards	Standardised tests
Outcome evaluation	Systematic and recurrent assessment of outcomes (seen as the programme's value-added) through standardised tests and additional processing of tests' data.	Standardised tests
Performance testing	Programme's clients have to perform tasks demonstrating they have acquired what the programme is aiming to equip them with.	Assessment tasks
Experimental studies	Controlled experiments with randomly assigned beneficiaries to an experimental group and a control group.	Randomised controlled trial

Approach	Description	Typical tools & methods
Management information systems	Mostly for internal evaluation purpose, follows programme's implementation (focused on processes).	Management systems
Cost-Benefit analysis	Quantitative procedures to capture the full cost of a programme and compare it to its benefits expressed in monetary value.	Cost-benefit analysis
Clarification hearing	A judicial approach to programme evaluation, led by a prosecuting evaluator and a defending one.	Trial-like methods (hearing of witnesses, evidence exhibits, etc.)
Case study evaluations	The programme is not controlled but rather holistically described with great details. It delineates a programme more than it allows to assess it.	Qualitative case study designs
Criticism and connoisseurship	A highly skilled and specialised evaluator is asked to provide his professional opinion on a programme, not necessarily following any kind of process or method.	N/A
Programme theory-based evaluation	Aim to answer interrogations related to specific articulation points of the programme theory.	Use of logic models to pinpoint evaluation questions
Accreditation / certification approach	Proves that a programme meet a given set of requirements and may identify areas of improvements.	Self-reports and audits.
Client-centred studies (or responsive evaluation)	Evaluator works for a heterogonous group of stakeholders (developer, funder, beneficiaries, etc.) involved in the programme. Characterised by relativistic approach seeking no definitive answer.	Action research with a constructivist lens.
Constructivist evaluation	Paradigm-driven evaluation with subjective epistemology where program's participants are encouraged to help shape the evaluation.	Qualitative methods
Deliberative democratic evaluation	All stakeholders equally participate in the evaluation process. Findings are debated among stakeholders.	Surveys, interviews, debates.

Appendix 6: Selected alternative indicators

It is difficult to be exhaustive in that matter, but we believe we cover the main indicators that have been the object of scientific inquiries in the past twenty to thirty years (this kind of measurements being a relatively novel field in economics with one the pioneering work attributed to William D. Nordhaus and James Tobin in 1972). Due to the focus of our research on social impact, we excluded indicators with an explicit and sole focus on *inter alia* environment, such as the ecological footprint, (Wackernagel & Rees, 1998). We lack the space here to discuss each of the listed indicators, but the intent is to show that there is a substantial amount of work being done to capture different aspects of welfare and that critics addressed to current socio-technical arrangements are not left without answers (we see for instance that a number of the indicators below have been developed in a response to a form of indicator-substitution by dominant economic aggregates such as the GDP).

Indicator	Overview	Main references
Inclusive Wealth Index (IWI)	It consists in giving a holistic account of wealth, by measuring the stocks of “all forms of capital” (grouped as human, manufactured and natural) and plot a trajectory for the progression of that capital through econometric assumptions.	(UNEP, 2012)
Better Life Index	The index was launched in 2011 in attempt to measure well-being. It measures 24 variables grouped in 11 dimensions: work-life balance, housing (quality and cost), education, income, jobs, environment, community, civic engagement, health, life satisfaction and safety.	OECD ¹⁶⁹
Human Development Index (HDI)	An index made of the geometrical average of purchasing power parity GDP per capita, life expectation at birth and an assessment of education levels using for two thirds the adult alphabetisation rate and for one third the schooling rate.	(United Nations Development Programme (UNDP), 1990)
Social Health Index	It aims to complete or compete with GDP to assess progress. It is made of 16 variables grouped in five categories designed around age categories. For instance, the category “children” includes the variables “infantile mortality” and “child abuse” and the category “elderly” includes the variables “life expectancy at 65” and “poverty rate among people aged 65 and above”.	(United Nations Development Programme (UNDP), 1990)
Index of economic well-being	An index aiming at competing with GDP on the representation of well-being. It computes 24 variables grouped in four categories (adjusted consumption per capita, stock of wealth per capita, equality and economic security) and aggregate them using a combination of weighted averages and monetarised metrics.	(Osberg & Sharpe, 1999)

¹⁶⁹ The detailed indicators and their definition is available on the OECD’s website. The organisation also hosts a website dedicated to the index: <http://www.oecdbetterlifeindex.org> (last accessed July 2018)

Indicator	Overview	Main references
Index of sustainable economic welfare (ISEW)	Another alternative to the GDP, seeking to account for both costs and benefits of production (where GDP would only take benefits into account through increased outputs regardless of externalities). The ISEW is built on the basis of national accounts, as followed: ISEW. = personal consumption + public non-defensive expenditures - private defensive expenditures + capital formation + services from domestic labour - costs of environmental degradation - depreciation of natural capital	(C. W. Cobb & Daly, 1989)
Genuine Indicator (GPI)	Progress The GPI is an extension of the ISEW, framed by Redefining Progress, an American think tank. It introduces minor methodological tweaks mostly focused around better capturing environmental stakes. ISEW and GPI are sometimes used interchangeably. Several American states have promoted the use of the GPI in their legislation.	(C. Cobb, Halstead, & Rowe, 1995)
Adjusted Net Savings	Originally called "Genuine savings" when first introduced, it is a monetarised index that gauges sustainable development through the following equation: adjusted net saving = Gross national saving + education expenditure – consumption of fixed capital – Depletion of energy resources – Depletion of minerals – Net depletion of forests – CO2 damages – Particulate pollution damages. Different financial proxies are used for the environmental indicators to arrive to a final monetary value.	(Hamilton & Naikal, 2014)
Self-Anchoring Scale	Striving Also called "the ladder of life", it gauges people's evaluation of their own life. It weighs an array of questions answered using a zero to ten Likert scale, where the scale symbolizes a ladder where the first step represents the "worst possible life" on a given item, and ten the "best possible life". The scale was popularised through Gallup's World Happiness Report, ranking countries by measured happiness. The scale us also used in composite indicators, such as the OECD's better life index or the Happy Planet Index.	(Cantril, 1965)
Happy Planet Index (HPI)	The index measures "sustainable well-being for all". It is calculated as followed: HPI = (numerator:) wellbeing measured through Gallup using the Self Anchoring Striving Scale x Life expectancy x "inequality of outcomes" (as the distribution in well-being and life expectancy data) divided by (denominator): ecological footprint (a determination of consumption related to land use, cf. Wackernagel & Rees, 1998).	New Economic Foundation (NEF) ¹⁷⁰ , see also (Abdallah, Thompson, Michaelson, Marks, & Steuer, 2009)
Gross National Happiness (GNH)	Born in Bhutan, the GNH intends to measure national happiness. It explicitly proposes a non-western approach to happiness and proposes a framework heavily influenced by the Buddhist dogma (including e.g. spirituality or harmony with nature). Is hatched around nine domains (psychological well-being, health, time use, education, cultural diversity and resilience, good governance, community vitality, ecological diversity and resilience, living standards) grouping 33 variables.	(Ura, Alkire, Zangmo, & Wangdi, 2012)
Social Capital	Social bonds, bridges and linkages that enable trust and contribute to shape individuals' life outcomes. There is no streamlined way to measure social capital, but a preminent method consists in evaluating network sizes (by using e.g. the number of nodes).	(Putnam, 2000)

¹⁷⁰ NEF hosts a website dedicated to the HPI: <http://happyplanetindex.org> (last accessed July 2018)

Appendix 7: Social impact assessment state of the art - 124 selected models

Method	Creator
AA1000	AccountAbility
Acumen Scorecard	Acumen Fund
Ashoka Measuring Effectiveness Questionnaire	Ashoka
Assessment and improvement indicators	Venture Philanthropy Partners
Atkisson Compass Assessment For Investors (ACAFI)	Atkisson
B Lab'S GIIRS Rating	B Lab
Balanced Scorecard Modified To Include Impact	New Profit
Base Of Pyramid Impact Assessment Framework	William Davidson Institute, University of Michigan
Basic Efficiency Resource (BER) Analysis	Cugelman & Otero
Beneficiary Perception Report (BPR)	Center for Effective Philanthropy
Best Available Charitable Option (BACO)	Acumen Fund
Blended Value Framework	Jed Emerson
Board Service ROI Tracker	True Impact, with BoardSource
Brandanomics	Center for Citizenship Enterprise and Governance, Earth
Business ethic excellence model	European business ethics network (EBEN)—GR
C3 perform	C3
CDFI Assessment And Rating System (CARS)	AERIS
Charity Analysis Framework (CAF), former Charity Analysis Tool (CHAT)	New Philanthropy Capital
Charity Assessment Method Of Performance (CHAMP)	Test
Community impact mapping	Development trust association
Comparative Constituency Feedback	Keystone
Co-Operative Environmental And Social Performance Indicators (CESPIS)	Co-operatives Commission UK
Cooperatives performance indicators—CESPI	Co-operatives commission UK
COSA methodology	Sustainable commodity initiative
Cost Per Impact	Center for High Impact Philanthropy
Cost-Benefit Analysis	V. Pareto, J. Dupuit
Cradle to cradle certification Trucost	Cradle to Cradle Product Innovation Institute
Dalberg Approach	Dalberg
Development Outcome Tracking System (DOTS)	International Finance Corporation
Dowjones Sustainability Index (DJSI)	Dow Jones and RobecoSAM
Echoing Green Mid-Year And Year End Reports	Echoing Green
Ecological Foot Print	Global Footprint Network
Eco-Mapping	Heinz-Werner Engel
Eco-mapping EMAS	Heinz-Werner Engel
Environmental Performance Reporting System (EPRS)	Environmental Capital Group
European foundation for quality management	British quality foundation (BQF)
Expected return	Hewlett foundation
Fair Trade Certification	Transfair USA (others available)"
Family of measures	Sawhill and Williamson
Fit for purpose	Development trusts association

Method	Creator
Foundation Investment Bubble Chart	N/A*
GRI Guidelines	Global Reporting Initiative
Handbook For Product Impact Assessment	Roundtable for Product Social Metrics
Hewlett Foundation Expected Return	William and Flora Hewlett Foundation
Human Impact + Profit (HIP)	HIP Investor
Impact reporting and investment standard	Acumen Fund, B Lab, Rockefeller Foundation
Investors in people	UK commission for employment and skills
Iris Library	Deloitte, PWC, Acumen Fund, B-Lab, Hitachi, Global Impact Investing Network
Iso 26000	ISO
Leadership In Energy And Environmental Design (LEED) Certification	US Green Building Council
Local Multiplier (LM3) (Formerly Local Economic Multiplier -LEM)	New Economics Foundation
Logic model builder	Innovation network
Logic Models (Logframe)	USAID
Measuring Impact Framework (MIF)	World Business Council for Sustainable Development
Measuring Impacts Toolkit	Community Development Venure Capital Alliance
Methodology for impact analysis and assessment	Investing for Good
MicroRate	MicroRate
Millennium Development Goal Scan	Business in development, Sustainalytics
Movement Above The \$1 A Day Threshold Project	Microcredit Summit Campaign
Nesta Standards of Evidence	Nesta
Oekom Rating	Oekom
Ongoing Assessment Of Social Impacts (Oasis)	REDF
Outcome Mapping	N/A***
Outcomes Matrix	Big Society Capital
Outcomes Star	Triangle
Participatory Impact Assessment	Feinstein International Center
Political Return On Investment	New Progressive Coalition
Poverty And Social Impact Analysis (PSIA)	World Bank
Practical Quality Assurance System For Small Organisations (PQASSO)	Charities Evaluation Services
Progress Out Of Poverty Index	Grameen Foundation
Prove It!	New Economics Foundation
Public Value Scorecard (PVSC)	M.H. Moore
Quality first	Birmingham voluntary service council
Real Indicator Of Success In Employment (RISE)	REDF
Robin Hood Foundation Benefit-Cost Ratio	Robin Hood
SA8000	Social Accountability International
SCALERS	Bloom and Chatterji
Shujog Impact Framework	Shujog
SIA tool survey	INAFI

Method	Creator
SIMPLE (Social IMPact for Local Economies)	CUBIST Research Group (Socio-Economic Impact Research) Brighton Business School, University of Brighton
Sinzer (Formerly Social E-Valuator)	Sinzer
Social Accounting And Audit (SAA)	T.J. Kreps
Social Added Value Evaluation (SAVE)	Bassi and Vincenti
Social Compatibility Analysis	Institute for Sustainable Development
Social Costs-Benefit Analysis	N/A**
Social Costs-Effectiveness Analysis	N/A**
Social Earnings Ratio	S. Hilton
Social enterprise balanced scorecard	Social Enterprise London (SEL)
Social Enterprise Mark	Social enterprise mark company
Social Footprint	Center for Sustainable Organizations
Social Impact Assessment (SIA)	Global Social Venture Competition
Social Impact Assessment For Local Economies (Simple)	Social Enterprise London
Social investment risk assessment	Hunter consulting LLC
Social performance indicators	CERISE
Social Rating	M-CRIL
Social Reporting Standard (SRS)	Consortium (Ashoka, Auridis, Bonventure, Phineo, Schwab foundation)
Social Return Assessment (SRA)	Pacific Community Ventures
Social Value Metrics	Root Capital
Socio-Economic Assessment Toolbox (SEAT)	Anglo American
Soft Outcomes Universal Learning (SOUL) Record	Norwich City College
SROI Analysis	Pacific Community Ventures
SROI Calculator	Calvert Social Investment Foundation
SROI Framework	REDF
SROI Lite	Global Social Benefit Incubator
SROI Toolkit	SVT
Stakeholder Value Added (SVA)	Schaltegger et al.
Stakeholder value added	Leuphana Universität Lüneburg
Star Social Firm	Social Firms UK
Success measures data system	Neighbor works America
TBL scorecard	Triple Bottom Line Collaborative
The B Impact Rating System	B Lab
The big picture	Scottish council voluntary organizations
The FINCA client assessment tool	FINCA
The values based checklist for social firms	Social firms UK
Theories Of Change	The Bridgespan Group + Carol Weiss+ The New Schools Venture Fund
Third sector performance dashboard	Social firms UK
Toolbox For Analysing Sustainable Ventures In Developing Countries	United Nations Environmental Programme (UNEP)
Triple Bottom Line Accounting	Elkington
Trucost	Trucost
Value Insight	Hact

Method	Creator
Volunteering Impact Assessment Toolkit	Institute for Volunteering Research
Wallace assessment tool	Wallace Foundation
Wellbeing measure	New Philanthropy Capital
Wellventure Monitor	Fortis Foundation Netherlands

* No entity or author is formally credited with the development of the Foundation Investment Bubble Chart Investment method. Tuan (2008) writes *“Some non-profits and foundations are using a bubble chart to display comparative information regarding multiple organizations”*, quoting Sara Olsen, a presentation from Jon Hugget for the Bridgespan Group and Hans Rosling (in a separate document from the Gates Foundation) as sources for her statement.

** Social Cost-benefits Analysis and Social Costs-Effectiveness Analysis are both described by Maas as methods derived from “general economic tools” (2011, p.29) that have been adopted to economically assess the social aspects of programmes or interventions.

*** Krlev et al. (2013) do not credit the development of Outcome Mapping with any specific organisation or person.

Appendix 8: Description of the selected social impact assessment methods for WISEs

1. Ashoka Measuring Effectiveness questionnaire:

This tool is a simple 2-page survey that Ashoka sends to its fellows around the world in order to gauge the overall impact of the organisation and the entrepreneurs it supports. The approach is straightforward and offers low complexity but allows for limited information to be captured.

2. AtKisson Compass Investment for Investors (ACAFI):

This method is based on AtKisson Compass Index of Sustainability, which focuses on four key areas: nature (N), economy (E), society (S) and well-being (W). A fifth element, synergies (+), is added to measure how each area reinforces the others. The approach proposes a rating system for each area, based on pre-selected indicators and criteria.

3. Balanced Scorecard Modified To Include Impact:

This method builds on Balance Scorecards (BSCs), a strategic and management tool popularised in the 1990s. The “impact version” of the BSCs capitalises on its versatile nature encompassing both strategic and operational dimensions to manage and measure social impact. The polytropic nature of this tool however limits its ability to capture in-depth qualitative data.

4. Basic Efficiency Resource (BER) Analysis:

At its core, BER proposes to simplify evaluation and assessment by looking at a condensed set of indicators measured in relative terms in order to facilitate comparison across portfolios of projects. The information is typically presented in a matrix, relating inputs to outputs. While efficient in its approach, this tool lacks the depth of view to properly assess outcomes.

5. Best Available Charitable Option (BACO):

In Acumen’s own word, BACO answers the question “For each dollar invested, how much social output will this generate over the life of the investment relative to the best available charitable option?”. The method proposes to quantify the impact of a considered project to benchmark it against a range of existing programmes in the charity sector proposing to solve a same issue. This method implies impact

quantification through advanced financial computations but is limited in capturing long-term outcomes.

6. Charity Analysis Tool (ChAT):

ChAT builds on several approaches and disciplines, such as the McKinsey Capacity Assessment Grid or the SROI framework to assess projects in three key areas: results, risks and capacity. The method is research-driven and very thorough, investigating qualitative as well as quantitative elements.

7. Comparative Constituency Feedback:

This method proposes to measure a project's impact based on the feedback on its constituents, which it breaks down into three categories: the primary constituents (or the people directly affected by the considered issue), the grantees and investors, the other constituents (policy makers, partners or other ad hoc stakeholders). Being very focused on qualitative data, this method leaves out most quantification to focus on evaluating interventions through in-depth exploration of stakeholder's experience.

8. Cost Per Impact:

Dubbed a "back-of-the-envelope" estimate by its creators (The Center for High Impact Philanthropy, n.d.), Cost Per Impact reduces impact measurement at a single relative metric corresponding to the ratio of the project costs to the project's impact valuation. Here impact valuation rests on "traditional" evaluation approaches such as randomised control trial or quasi-experimental design obtained from previous experiences to assess the cost per impact *ex-ante*. Despite an effort to encompass outcomes in impact evaluation, this method often appears strongly focused on outputs.

9. Echoing Green Mid-year and Year End Report:

Similarly to Ashoka, Echoing Green keeps track of its fellows' work and impact by surveying them with a standard questionnaire. The questionnaire is issued twice a year and touches upon many different aspects of social entrepreneurship, such as impact, management and development plans. It is a straightforward tool to use, and allow to collect some relevant qualitative information, however limited in breadth.

10. Hewlett Foundation Expected Return:

The Expected Return developed by Hewlett Foundation is a single metric based on the estimated output of a programme, its estimated likelihood of success, its total costs and the share of its costs supported by philanthropy. The result is a ratio providing a relative value with the total costs of the programme as a denominator. While this is another model that can provide a good view of an intervention's return in financial terms, its perspectives on outcomes can be limited.

11. IRIS Library:

IRIS provides a library of standard indicators that can be used to measure impact in a wide variety of settings. The standardised nature of the metrics proposed by IRIS, and the wide coverage of sectors it offers allow for a very easy and broad deployment, at the expense of a narrow scope of measurement, often reduced to outputs.

12. Measuring Impact Framework (MIF):

MIF proposes a thorough approach to measuring impact, mostly framed around management processes. It clearly links activities to outputs to outcomes, highlighting "results chain". It appears more oriented towards absolute value measurement than towards the production of ratios, encouraging the use of qualitative information.

13. Outcome Stars:

Triangle proposes more than 20 versions of its outcome star to fit different impact measurement needs. Each star has a number of branches representing the different outcomes it aims to assess, with different levels of progression common to all branches. The Outcome Stars propose a relatively easy to use "out of the box" solution for impact measurement with a good focus on outcomes. Outputs are however often left out and inputs are simply not taken into account.

14. Public Value Scorecard (PVS):

The PVS is another iteration of the BSCs dedicated to impact measurement. Its clear focus on public value means that it brings a pregnant emphasis on outcomes and the ripple effects of the organisation's activities in a society. It also stresses the key differentiations between impact-driven organisations and

for-profit ones, translating into substantial adaptation in the management approach proposed by the scorecard.

15. Robin Hood Foundation Benefit-Cost Ratio:

As its name implies, this method takes a cost-benefit approach to assess social impact. Outcomes of a programme are carefully assessed and pegged to financial proxies in order to assess the total benefit of an intervention over time, discounted for present value. The result is multiplied by a “Robin Hood factor” corresponding to Robin Hood’s funding divided by the total costs of the programme, before being expressed as ratio to Robin Hood’s funding (the benefit-cost ratio). The focus on outcomes in this method makes it complex to implement, with substantial research requirements.

16. Social Compatibility Analysis (SCA):

SCA proposes to look at all components of a programme and to assess them against a range of relevant criteria, ranking them from A (highly relevant issues) to C (low relevance), while additional ranks such as “not relevant” or “possible threat” are also used. Components of a studied issue can be broken down along the impact value chain to then be assessed against impact criteria pertinent to the assessed intervention. The tool can provide a good mapping of the considered issue and highlight benefits and threats, but its potential in terms of proper measurement is limited.

17. Social Cost-Benefit Analysis (Social CBA):

The social CBA rests its principles on the classical economic cost-benefit analysis, extended to include “the full spectrum of costs and benefits” (Vardakoulias, 2013, p. 1) i.e. environmental and social aspects. It does so by focusing particularly on the economically tangible knock-on effects of an intervention, researching how outcomes can be valued through empirical financial estimates. This method requires skills and time to perform ad hoc research, and lack the stakeholder-based approach of SROI.

18. Social Return On Investment (SROI):

While the idea of SROI is inspired by the financial return on investment computed for regular investments, it follows primarily a cost-benefit approach to impact measurement, where outcomes are

valued through financial proxies. The amount obtained, once discounted for present value, is then used as a numerator for the SROI ratio where the denominator is the resources invested in the project. The key difference with other cost-benefit based approach is that SROI is strongly focused on stakeholders and typically involves them in developing the financial proxies used in the impact calculations.

Appendix 9: Detailed interviews schedule

Country	Stakeholder	Case / level	Title	Date
Denmark	Erhvervsstyrelsen	Contextual	Head of Section	06/02/2017
Denmark	Erhvervsstyrelsen	Contextual	Head of Section	16/02/2017
Denmark	Sociale Entreprenører i Danmark	Contextual	Chairman	02/03/2017
Denmark	Den Sociale Kapitalfond	Contextual	Chief Analyst	07/03/2017
Denmark	Sociale Entreprenører i Danmark	Contextual	Chairman	26/04/2017
Denmark	Erhvervsstyrelsen	Contextual	Special Consultant	27/04/2017
Denmark	ComeBack	Contextual	CEO	08/05/2017
Denmark	ComeBack	Contextual	Backoffice Manager	08/05/2017
Denmark	DK1	DK1	Head of Development	19/06/2017
Denmark	DK1	DK1	Head of Development	27/06/2017
Denmark	Foldschack, Forchhammer and Dahlager	Contextual	Partner	29/06/2017
Denmark	Kirken Korshærs	Contextual	Project Manager	29/06/2017
Denmark	Roskilde University / CBS	Contextual	Lecturer	13/07/2017
Denmark	DK1	DK1	Head of Development	15/08/2017
Denmark	DK1	DK1	Social worker	15/08/2017
Denmark	DK1	DK1	Social worker	15/08/2017
Denmark	DK1	DK1	CEO	15/08/2017
Denmark	DK1	DK1	Development Consultant	15/08/2017
Denmark	DK1PS	DK1	Project Manager	25/09/2017
Denmark	DK1PS	DK1	Head of Economics and Statistics Division	25/09/2017
France	Fédération des entreprises d'insertion	Contextual	Chargée de mission Plaidoyer	26/09/2017
France	DIRECCTE Grand Est	Contextual	Chargée de Mission IAE	03/10/2017
France	DIRECCTE Vôsoges	Contextual	Chargée de Mission IAE	03/10/2017
France	COORACE	Contextual	Chargé de mission innovations sociales	09/10/2017

Country	Stakeholder	Case / level	Title	Date
Denmark	DK1	DK1	Head of Development	18/10/2017
France	DIRECCTE Haut-Rhin	Contextual	Directrice Adjointe Pôle Emploi et Formation	20/10/2017
France	Chantier Ecole National	Contextual	Researcher	23/10/2017
Denmark	All Ears	Contextual	CEO	09/11/2017
France	FR1	FR1	CEO	22/11/2017
France	FR1	FR1	Head of covenants & partnerships	22/11/2017
France	FR1	FR1	Head of training engineering	22/11/2017
France	FR1	FR1	Head of covenants & partnerships	22/11/2017
France	Direccte (Region)	FR1	Work Integration Officer	24/11/2017
France	Phitrust	FR1	Investment Manager	28/11/2017
Denmark	Roskilde University	Contextual	Professor	30/11/2017
France	COORACE	Contextual	Vice-Président	01/12/2017
Europe	GECES	Contextual	President	06/12/2017
Europe	ENSIE	Contextual	Director	07/12/2017
France	Hélène Duclos	Contextual	Expert	07/12/2017
Denmark	DK2	DK2	Operations Manager	12/12/2017
Denmark	DK2	DK2	Head of educational programmes	12/12/2017
Denmark	DK2	DK2	CEO	12/12/2017
Denmark	DK2	DK2	Integration Manager	15/12/2017
Denmark	DK2	DK2	Head of partnerships	10/01/2018
Denmark	DK2PS	DK2	Case Worker	10/01/2018
France	Fédération des entreprises d'insertion	Contextual	President	12/02/2018
France	Chantier Ecole IDF	Contextual	Network Promoter	23/02/2018
France	Chantier Ecole Normandie	Contextual	Regional delegate	23/02/2018

Country	Stakeholder	Case / level	Title	Date
France	SAS Formation	Contextual	Director	23/02/2018
France	DEFIS	Contextual	CEO	23/02/2018
France	Chantier Ecole AD	Contextual	President	23/02/2018
France	Chantier Ecole National	Contextual	Natioanl Delegate	23/02/2018
France	Chantier Ecole National	Contextual	Natioanl Delegate	23/02/2018
France	Chantier Ecole National	Contextual	Researcher	23/02/2018
France	KPMG	Contextual	Researcher	13/03/2018
France	FR2	FR2	Operations Manager	03/04/2018
France	FR2	FR2	CEO	03/04/2018
France	Dreccte (Region)	FR2	Work Integration Officer	03/04/2018
France	Direccte (County)	FR2	Employment and Territory Development Officer	03/04/2018
France	Direccte (County)	FR2	Employment and Territory Development Officer	03/04/2018
France	Municipality	FR2	Work Integration and Employment Officer	03/04/2018
France	Collectif 05	Contextual	Manager	07/04/2018
France	National Council for Work Integration (CNIAE)	Contextual	President	08/05/2018
France	Ministry of Labour (DGEFP)	Contextual	Director, Work Integration	17/05/2018
Denmark	DK2PS	DK2	Head of section	30/05/2018
Denmark	Ministry of Labour	Contextual	Special Advisor	04/09/2018

Appendix 10: Overview of classification attributes used in social impact assessment literature

Variables and (variables)	Author	Year
Function (process; impact; monetarisation)	Clark et al.	2004
Impact value chain stage (input; activity; output; outcome; goal alignment)		
Lifecycle stage (start-up; expansion; maturity)		
Purpose (screening; partnership formation; management operations; scaling; external reporting; exit; retrospective evaluation)		
Cost/time		
Time breakdown (management; staff; third-party consultant; investor)		
Qualitative / quantitative	Nicholls,	2005
Methods / framework	Zappalà and Lyons	2009
Type of analysis (cost, cost-effectiveness, cost-savings, cost-benefit)	Tuan	2008
Functional type (general / sector specific; rating system; assessment system; management system)	Olsen et al.	2008
Impact (implied; proven; optimised)		
Category (economic; social; environmental)		
Perspective (external; internal)		
Data management tools (data entry, analysis and report; per investors, company and third-party)		
Feasibility (time by role: investor, management, staff, consultant / third-party)		
Role / purpose (accounting to external stakeholders; managing internal operations; assessing societal impact)	Mulgan	2010
Motivation (assessing effectiveness and/or efficiency; implementing change; quality mark; funder requirement)	Rinaldo	2010
Readiness (defined/undefined social purpose)		
Capacity (small; medium; large)		
Impact type (environmental; economic; holistic; on volunteers)	Maas and Liket	2011
Purpose (screening; monitoring; reporting; evaluation)		
Timeframe (prospective; ongoing; retrospective)		
Orientation (input; output)		
Length of timeframe (short term; long term)		
Perspective (micro; meso; macro)		
Approach (process methods; impact methods; monetarization)	Krlev et al.	2012
Approach (organisational capacity assessment; stakeholder analysis; sustainability ratings)		

Variables and (variables)	Author	Year
Evaluation type (cost analysis; cost effectiveness analysis; cost utility analysis; cost benefit analysis)	Rauscher et al.	2014
Macro-categories (simple social quantitative; holistic complex; qualitative screening; management)		
Data type (quantitative; qualitative; quali-quantitative)		
Impact type (holistic; people; social; environmental; economic)		
Purpose (assessment; screening; management; certification; reporting)	Grieco et al.	2015
Complexity (basic; simple; complex; highly complex)		
Sector (general; specific)		
Time frame (Prospective; ongoing; retrospective)		
Developer (University / research centre; Non-profit network; Non-profit organisation; consulting firm; institution)		

Appendix 11: Retained SIA attributes, variables and SIA models

Model ¹⁷¹	Purpose				Analytical lens					Data typology		Design	
	Screening	Monitoring	Reporting	Evaluation	Economy	Efficiency	Cost-benefit	Effectiveness	Monetari- sation	Quantitative	Qualitative	Planned	Negotiated
Ashoka Measuring Effectiveness	0	1	2	1	0	0	0	2	0	1	2	2	0
Atkisson Compass (ACAFI)	1	2	2	1	1	0	0	1	0	1	2	1	1
Balanced Scorecard	0	2	1	1	1	1	0	1	0	1	1	1	1
Basic Efficiency Resource (BER)	0	2	1	1	1	2	1	0	0	1	1	2	0
Best Available Charitable Option (BACO)	2	1	2	2	1	2	1	1	1	2	0	2	1
Charity Analysis Tool (CHAT)	2	1	1	2	1	1	1	1	0	2	2	2	0
Comparative Constituency Feedback	1	2	1	2	0	0	0	2	0	0	2	0	2
Cost Per Impact	2	1	1	2	1	2	1	1	1	2	0	2	0
Echoing Green	0	1	2	1	0	0	0	1	0	1	2	2	0
Hewlett Foundation Expected Return	2	1	0	2	1	2	1	1	1	2	0	2	1
Iris Library	2	2	2	2	1	2	0	1	0	1	1	1	1
Measuring Impact Framework (MIF)	2	0	2	2	1	2	2	2	1	2	2	1	1
Outcomes Star	1	2	2	2	0	0	0	2	0	0	2	1	1
Public Value Scorecard (PVSC)	1	2	2	1	1	1	1	2	0	1	2	0	2
Robin Hood Foundation	2	1	1	2	1	1	2	1	1	2	1	2	1
Social Compatibility Analysis	2	0	0	1	0	0	0	1	0	0	2	1	2
Social Costs-Benefit Analysis	2	0	0	1	1	1	2	2	1	2	1	1	1
SROI Framework	2	1	1	2	0	1	2	2	2	2	1	0	2

¹⁷¹ 0 indicates no focus on a particular variable, 1 indicates a moderate focus, 2 indicates a strong focus. The rating is the product of our own judgement based on the available literature and application examples. Variables are weighted relatively per attribute, but that does not prevent several variables from being rated similarly when the model design and use justify it.

Appendix 12: DK1's SIA components

Table 34: DK1's SIA structure and process

Scheme	Name	Purpose	Timeframe
A	Baseline	Survey establishing competences baseline	<i>Ex-ante</i>
B	Efforts	A journal of a beneficiary's activities	<i>Ex-ante, interim</i>
C	Ongoing evaluation	A recurring survey assessing competences progress	<i>Ex-ante, interim, ex-post</i>
D	End progress	Survey assessing the beneficiaries' progress against the established baseline	<i>Ex-post</i>
E/F	Final report		<i>Ex-post</i>

Source: DK1, 2017

Table 35: The competences assessed by DK1's SIA model

Personal competences	Professional competences
Meet on time	Focus and ability to work
Psychological stability	Capacity in following instructions
Personal appearance	Work intensity and rhythm
Self-awareness of strengths and weaknesses	Quality in performing of work tasks
	Work autonomy
	Flexibility
	Adaptability

Source: DK1, 2017

Appendix 13: DK2PS evaluation framework

Category	Psychological, light	Psychological, heavy	Personal	Physical	Gang affiliation	Verbal
Description	“Light” disorders such as OCD or stress	Heavier disorders requiring psychiatric treatment	Personal challenges, out of the psychological sphere	Documented physical challenges	Is part of or in the social periphery of a gang	Language challenges
Indicators	Job perspectives Ability to meet on time Professional identity Motivation Experience Professional competences Adaptability Verbal competences Written competences Social competences Conflict management Dropout history Social network Personal hygiene Housing situation Social network Personal economy Illness self-awareness Cognitive competences Abuses	Experience Professional competences Illness self-awareness Abuses Ability to cope with psychological challenges Cognitive competences Ability to meet on time Personal hygiene Conflict management Dropout history Social competences Social network Housing situation Professional identity Motivation Personal economy Verbal competences Written competences	Experience Professional competences Abuses Cognitive competences Job perspectives Ability to meet on time Conflict management Adaptability Motivation Personal hygiene Personal appearance Criminal history Social competences Dropout history Housing situation Personal economy Social network Professional identity Verbal competences	Overweight Musculoskeletal status Ability to cope with physical symptoms Cognitive challenges (brain injury) Abuses Professional identity Job perspectives Experience Professional competences Ability to meet on time Social network Personal appearance Adaptability Motivation Personal economy Verbal competences Written competences	Gang affiliation Social network Personal economy Criminal history Conflict management Ability to meet on time Social competences Abuses Cognitive competences Professional identity Job perspectives Experience Professional competences Housing situation Adaptability Motivation Personal hygiene Verbal competences Written competences	Verbal competences Written competences Experience Professional competences Job perspectives Motivation Ability to meet on time Social network Cognitive competences Personal appearance

Appendix 14: FR1 SIA components

Category	Item
Number of full-time employees	Number of hired
	Number of employees leaving
	Number of female full-time employees
	Number of employed full-time equivalent
Number of part-time employees	Number of hired temporary workers
	Number of temporary workers leaving
	Number of female temporary workers
	Number of hours worked as temporary full time equivalent temporary workers
	Average time worked by temporary workers in hours
	Average time worked by temporary workers in months
	Number of temporary workers per full time equivalent
Number of beneficiaries	Number of beneficiaries onboarded
	Number of beneficiaries graduating
	Number of female beneficiaries
	Number of beneficiaries on minimum income social benefit
	Number of hours worked as temporary
	Beneficiaries full time equivalent
	Number of hours worked
	Number of months worked
	Number of beneficiaries per full time equivalent
Graduations	Average duration of beneficiaries' stay
	Number of sustained employment graduations
	Number of transitional employment graduations
	Number of positive graduations
	Number of dynamic graduations
	Percentage of positive graduations
Training	Training contributions
	Training investments
	Training effort

Category	Item
	Number of training hours
	Beneficiaries' level of training
Non-work-related challenges	Housing (number of cases and percentage)
	Social life (number of cases and percentage)
	Health (number of cases and percentage)
	Finance (number of cases and percentage)
	Mobility (number of cases and percentage)
	Training (number of cases and percentage)
	Legal issues (number of cases and percentage)
	Cumulative challenges
Miscellaneous	Number of beneficiaries without a driving license, per age bracket
	Number of beneficiaries from "priority districts"
	Number of disabled beneficiaries
	Number of beneficiaries undergoing a judiciary process

Appendix 15: FR2's SIA components

Theme	Criteria	Output indicators	Outcome indicators	Impact indicators	
Reinforcing individuals' power		Total gross salary	Number of employees / salary amounts per activity	Survey and testimonies (under development)	
	Accessing financial autonomy	Number of beneficiaries	Number of employment hindrances resolved: (housing, mobility, child care) - under development		
	Self-confidence and self-perception	n/a	n/a		
	Accessing social status		Number of beneficiaries		Number of beneficiaries having opened rights to social security (under development)
			Number of beneficiaries having followed a specific training		Number of beneficiaries having used special benefits (access to microcredit, cinema tickets, etc.) - under development
Fighting exclusion: a citizen value	Respect of differences: unprejudiced welcome		Number of individuals attending welcome meetings	Percentage of welcomed individuals who became beneficiaries	
				Percentage of welcomed individuals claiming social benefits	
				Number of nationalities	
	Number of beneficiaries			Percentage of welcomed individuals claiming social benefits	n/a
				Percentage of welcomed individuals coming from "priority districts"	
				Percentage per age bracket	
	Provide an opportunity to speak out	Number of fora	n/a		
Respect of choices	n/a	n/a			
Create social links	Number of beneficiaries having taken part in collective action or		Number of beneficiaries having created social links with other beneficiaries (survey under development)		

Theme	Criteria	Output indicators	Outcome indicators	Impact indicators	
Transition employment: more than short term contracts	Employment access	training (under development)			
		Number of welcomed individuals (men and women)	Number of individuals put to work (men and women)	Percentage of people graduating with a training or a job	
			Number of beneficiaries per revenue		
	Number of professional fields represented	Weight of different fields in total revenue			
	Training and professionalization (development of competences)	Number of individuals having taken part in a collective action	Percentage of participation	n/a	
		Number of individuals having taken part in a training	Satisfaction of trainees (under developments)	Percentage of beneficiaries having taken part in training related to graduation towards employment or training	
		Number of training hours per type of training	Funder split per total funding		
	Track customisation	Number of interviews carried out	Percentage attending follow-up interviews	n/a	
	Bringing wealth upon the local territory	Local services	Number of permanence hours and places	Number of individuals welcome per type of request	Economic activity in each district
		Avoided costs	Amount of gross salaries	Amount of social contributions	Customer satisfaction (under development)
Financial independence: an inexpensive business model (an association rather than a company)		Annual revenue	Number of beneficiaries FTEs + staff	n/a	
		Annual expenses	Share of public funding	Revenue / expenses	

Theme	Criteria	Output indicators	Outcome indicators	Impact indicators
	An entrepreneurial posture	Number of sectors represented	Weight per customer category	NACE categories
		Number of companies prospected	Number of prospects converted to customers	n/a
		Number of public contracts managed	Number of hours on public contracts	Customer testimonies
	Attention brought to local needs	Number of social provisions answered in public contracts	Number of hours realised under social provisions	Social provisions capture rate
	Sustainable development	Number of actions promoting sustainable development	n/a	n/a
Enacting social and solidary economy	Innovation and projects	Number of projects started	Number of projects carried out	n/a
	Network culture / mutualisation of resources	Number of partnerships	Number of participations to projects meetings	n/a
		Network memberships	Number of participations to development and research meetings	n/a
	Transmitting values: citizen employer	Solidary choices	n/a	n/a

Appendix 16: Coding grid

Code	Definition
1. Public administration and management environment	Data related to the characteristics of the PAM context in which WISEs operate. It concerns the identification of NPG and NWS mechanisms, as well general references to the environment in which these mechanisms are operationalised (e.g. tightened public budgets).
1.1. Neo-Weberian state	Characteristics of NWS as described by Pollitt and Bouckaert (2004, 2017) and how they influence strategic interactions with WISEs.
1.1.1. Traditional state apparatus	Elements of the traditional Weberian state (cf. Weber, 1965) and how PAM actors use them to maintain authority and controls over their stakeholders.
1.1.1.1. Bureaucracy	Evidence of bureaucratic mechanisms and how they influence on strategic interactions.
1.1.1.2. Centralisation	Centralised rules and administration and their influence on strategic interactions.
1.1.1.3. Legalism	Legalist mechanisms in the work integration sector and their influence on strategic interactions.
1.1.2. Modernised state	The modernised tenets (“ <i>neo</i> ”) of the NWS as described by Pollitt and Bouckaert (2004 and 2017).
1.1.2.1. Managerialism	An outward shift, from a focus on processes induced by bureaucracy to a focus on results induced by management-oriented practices.
1.1.2.2. Increased deliberation	Evidences of state structures more open to dialogue, identification of participatory processes and deliberation places.
1.1.2.3. Decentralisation	How decentralisation manifests itself and how it affects strategic interactions.
1.2. New Public Governance	Elements characterising the NPG as described by Osborne (2006, 2010) and encompassing fundamental contributions to network theory such as Powell (2003).
1.2.1. The place of networks	Data evidencing the reciprocal patterns of communication and exchanges (Powell, 2003)
1.2.1.1. Relational capital	Evidences of the formation and accumulation of social capital between WISEs and their public stakeholders and its effect on strategic interactions.
1.2.1.2. Decentralisation	Data directly referring to decentralisation or that can be interpreted as a result of decentralisation. A substantial amount of the information coded under this node refers to considerations on flexibility.

Code	Definition
1.2.2. The role of markets	The place of market mechanisms in NPG, particularly in the delivery of services generating public value such as employment services.
1.2.2.1. Competition	What characterises competition for WISEs operating in and NPG environment and what are its effects. Competition is seen both against other WISEs as well as against concurrent public employment services.
1.2.2.2. Privatisation of employment services	Data referring to the changes subsequent to the introduction of private suppliers for employment services.
1.2.2.3. Agent-principal in relationship-based quasi-markets	Agent-principal relations in a market setting with an explicit goal of public value creation and strong relationships between the agents and their principal
1.2.3. Focus on outcomes	Evidences of the focus on outcomes in NPG (Osborne, 2006)
1.2.4. Bureaucracy	Bureaucratic mechanisms found in the NPG and deplored by WISEs.
1.3. The environment	Elements related to the environment in which PAM actors and WISEs operate.
1.3.1. Tightened public budgets	References to restrictions in public funding (in terms of subsidies or price acceptance).
1.3.2. Increased accountability	References to a perceived increase in accountability from WISEs towards their public partners, but also sometimes towards the general public.
2. Public value theory	Data that relates to elements of public value theory, seen through the PVT triangle (Moore, 1995) and our own analysis of it (i.e. the nine entries we introduce in Chapter 2).
2.1 Public value	The definition, the creation and the acknowledgement of public value.
2.1.1 (Co)-definition	The process by which public value is defined (often tactilely) by WISEs and their public stakeholders. We separate between co-definition (involving deliberation) and a more one-sided definition.
2.1.2 (Co)-creation	The process of creating public value. IT focuses on the way WISEs create value and whether co-creation occurs with public stakeholders, based on the mechanisms we describe in Chapter 7.
2.1.3. Acknowledgement	The acknowledgement of the public value created.
2.1.3.1. Internal acknowledgement	Refers to the internal within the WISEs (e.g. among their staff or their leadership) of the public value created.
2.1.3.2. External acknowledgement	The acknowledgement of the public value created by external stakeholders (public actors in our research).
2.2. legitimacy	The means by which WISEs ensure their legitimacy towards their public partners, in order to secure the operational environment, they need to create public value.

Code	Definition
2.2.1. Deliberation	Deliberation as a means to acquire legitimacy.
2.2.1.1. Deliberation as a means to influence	Processes of deliberation used to generate a level of influence in order to garner legitimacy.
2.2.1.2. Deliberation as means to negotiate	Processes of deliberation used in negotiations and their effect on legitimacy
2.2.1.3. Phasing of deliberation	Difference between <i>ex-ante</i> , <i>interim</i> and <i>ex-post</i> deliberation when referring to an impact measurement process.
2.2.1.4. Emergence of shared preferences	How deliberation processes lead to the emergence of shared preferences among stakeholders.
2.2.2. Legitimacy devices	The creation and the use of legitimacy devices (the instruments such as SIA that are used to ensure legitimacy by “ <i>proving</i> ” the creation of public value.
2.2.2.1. External devices	Devices imposed by external stakeholders or the environment.
2.2.2.2. Internal devices	Devices developed internally: their development and their use.
2.2.3. Endorsement	Endorsement by third parties as a means to generate legitimacy.
2.2.3.1. Informal	Informal endorsement, occurring as a result of interactions rather than through established mechanisms.
2.2.3.2. Institutionalised	Institutionalised endorsement, occurring as a result of mechanisms created to that effect by the PAM framework (e.g. covenants in France).
2.3. Operationalisation	The way by which WISEs secure resources to carry out their operation and create public value.
2.3.1. Agency	The relation between WISEs and their stakeholders and the strategies they shape around them.
2.3.1.1. Cooperation strategies	Cooperation with external stakeholders (especially public actors)
2.3.1.2. Competition strategies	Competition with other stakeholders (other WISES or some public services)
2.3.1.3. Performance measurement to compete or cooperate	Role of performance measurement models in the cooperation or cooperation strategies with public stakeholders
2.3.2. Organisational design	The way WISEs structure their activities.
2.3.2.1. Legal entities	The strategies deployed around legal structures in order to access beneficiaries or resources.
2.3.2.2. Diversification and refocus of activity	The way WISEs develop their organisation in order to be able to cater to the needs they identify.
2.3.3. Activities	The activities carried out by WISEs.
2.3.3.1. Operational activities	The actual description and characterisation of activities.
2.3.3.2. Entrepreneurship	The way WISEs adapt to their environment by proposing new services or new activities in order to keep creating public value.
2.3.3.3. Innovation	The way innovation is channelled and framed around the purpose of public value creation. Often related to entrepreneurship items.

Code	Definition
3. Performance measurement approaches	Data related to the characterisation of the performance measurement approaches observed, based on the variables and attributes identified for SIA in Chapter 4.
3.1. Purpose	Specifying the purpose of carrying out a form of performance measurement
3.1.1. Screening	The purpose of screening of several options in order to choose the best one (typically <i>ex-ante</i>).
3.1.2. Monitoring	The purpose of monitoring activities, on an ongoing basis, and make sure they reach their goal (i.e. their impact, or public value creation).
3.1.3. Reporting	The purpose of reporting results, often to a third-party, and often with an underlying intent of control on the behalf of the party requesting reporting.
3.1.4. Evaluation	The purpose of gauging impact <i>ex-post</i> , in order to find out whether targets are reached.
3.2. Economic lens	The economic approach which influences the way performance is conceptualised and measured in specific models.
3.2.1. Economy	Approaches or discourses interested in raw economy of an option versus another one.
3.2.2. Cost-efficiency	Approaches or discourses focused on outputs, typically relating the amount of input necessary to the creation of outputs.
3.2.3. Cost-benefit	Approaches or discourses focused on outcomes, where inputs are related to broader effects than just outputs, typically encompassing avoided costs.
3.2.4. Monetisation	Approaches or discourses focused on expressing impact or outcomes in monetary units, typically using proxies.
3.3. Impact data typology	The type of data through which impact is or should be characterised.
3.3.1. Quantitative	Focus on quantitative data.
3.3.2. Qualitative	Focus on qualitative data.
3.4. Method design	The way performance measurement methods are designed.
3.4.1. Planned	Top-down design process where stakeholders' inputs carry little weight.
3.4.2. Negotiated	Inclusive design, sometimes even bottom-up.
4. Conceptions of performance	Data related to the understanding of performance and the operationalisation this understanding through the practice of performance measurement.
4.1. Performance regimes	Elements allowing to characterise the performance in which the encountered performance measurement approaches ascribe themselves to.
4.1.1. Autocracy	Very restricted sets of performance indicators (possibly just one), centrally planned.

Code	Definition
4.1.2. Fordism	Restricted set of performance indicators but which design and content can be negotiated.
4.1.3. Managerialist regime	Multiple sets of performance indicators, centrally planned (top-down design).
4.1.4. Deliberative regime	Multiple sets of performance of indicators, which design and content are deliberated by stakeholders.
4.2. Performativity	Elements of performativity observed in the discourse of the interviewed stakeholders.
4.2.1. Socio-technical arrangements	Elements pointing to the formation or the reinforcement of socio-technical arrangements.
4.2.1.1. Low intensity	Shared ideas and concepts among stakeholders (often on the basis of shared preferences) leading to tacit but not formalised agreements of how performance should be measured
4.2.1.2. High intensity, unrealised	Shared ideas and concepts among stakeholders that take the form of a measurement mechanisms, even rudimentary, but where performativity is not observed.
4.2.1.3. High intensity enacted	Shared ideas and concepts among stakeholders that take the form of a measurement mechanisms and where this mechanism seems to generate a degree of performativity.
4.2.2. Retroaction	Elements pointing to the existence of retroaction in the behaviour of stakeholders.
4.2.2.1. Deliberate retroaction	Retroaction occurring deliberately in order to meet targets imposed by third parties, possibly to the detriment of regular activities
4.2.2.2. Unintentional retroaction	Retroaction detrimental to regular WISEs' activities, through a process of which staff or leadership is not aware of.
4.2.3. Indicator-substitution	Elements pointing to the existence of indicator-substitution in the behaviour of stakeholders.

Appendix 17: Example of interview guide (semi-structured interviews)

Transition 1: Let us start with a few general information regarding your company. In this part of the interview, I am seeking to get an overview of your company and confirm some of the information I gathered prior to coming here today.

Section 1: General information & demographics

How many full-time equivalent employees does your company currently employ:

How many people are you supporting as part of your integration / insertion activities? What is your current candidates capacity (i.e. how many can you support at the most with your current staff level)?

In your own words, how would you describe the activity of your company?

In your own words, how would you describe the activities performed by the integration / insertion candidates your company support?

Transition 2: Let us now talk about your initial expectations regarding social impact measurement. In this section of the interview, I am looking to understand your state of mind before you actually measured your impact. I am interested in understanding how you took the decision to measure your impact and what was the thought process associated to this development.

Section 2: Your expectations about social impact measurement

Main question 2.1: Can you tell me about the first time you heard about social impact measurement? What were you first thoughts about it?

Probe 2.1.1: Can you remember who / where you heard it from?

Probe 2.1.2: What was your initial reaction?

Main question 2.2: On hindsight, and without taking into account the actual outcome of your work on impact measurement, can you tell me about the main reasons that first lead you to measure the impact of your organisation?

Probe 2.2.1: Can you elaborate on what precisely triggered your decision to measure your impact?

Probe 2.1.2: Can you elaborate about how funding prospects influenced your decision to carry out social impact measurement?

Probe 2.1.3: At the time, did you think that measuring your impact could change the way your organisation work?

Probe 2.1.4: What were your expectations? How did you think you would use the results?

Main question 2.3: When you took the decision to carry out social impact measurement for your organisation, did you foresee any downsides to it?

Probe 2.3.1: Have you anticipated any challenges related to doing social impact measurement?

Transition 3: Let us now talk about your actual experience with social impact measurement. Here I want to understand how your organisation set out to measure its own impact, with which tools, using which metrics, and the kind of experience it was for the stakeholders involved in that endeavour.

Section 3: Your actual experience with social impact measurement

Main question 3.1: What can you tell me about the actual experience of measuring your impact?

Probe 3.1.1: How was the impact on your organisation's resources (i.e. try to know whether it strained the organisation)?

Probe 3.1.2: How much of your staff was involved, how did you organise yourselves?

Probe 3.1.3: Have you identified best practices (i.e. what went well and what went bad)?

Probe 3.1.4: Would you do it again? Why?

Main question 3.2: Today, taking into account the results of your impact measurement study, can you tell me about the key consequences, intended as well as unintended, associated to having measured your impact?

Probe 3.2.1: Explore gaps between expectations and outcomes

Probe 3.2.2: Has the action of measuring your impact brought to light dimensions of your social impact that you were unaware of?

Probe 3.2.3: Since you measured your impact, have you taken actions to improve your impact? If yes, please specify, if no, please specify why you thought it wasn't appropriate

Main question 3.3: Which methods and metrics did you end up using to measure your impact and why?

Probe 3.3.1: Approximately how many impact measurement techniques have you heard of?

Probe 3.3.2: Was there any methods or metrics other than the ones you used that caught your attention? Why? Why did you end up not using it / them?

Probe 3.3.3: Did you inform your decision with case studies from existing impact measurement methods?

Probe 3.3.4: Do you continuously monitor your impact (i.e. ongoing measurement processes) or do you prefer to make impact measurement an ad'hoc exercise? Why?

Probe

Main question 3.4: In your own opinion, what are the most important criteria to assess the impact of your activity and how can they be translated into metrics?

Probe 3.4.1: How do you feel about the use of metrics to document your activities?

Probe 3.4.2: How comfortable are you with documenting and reporting intangible outcomes?

Main question 3.5: According to you, who is the main audience for the results of your social impact measurement exercise?

Probe 3.5.1: did the actual audience turned out to be different from the one you initially expected before carrying out the measurement? Please Elaborate.

Transition 4: Let us now conclude by talking about the relation between your company and the public stakeholders with which it interacts. In this final section, I am seeking to understand who these stakeholders are and the kind of relationship you have with them. I am more specifically interested in understanding the level of involvement they have or have had with your impact measurement activities.

Section 4: Relations with public stakeholders

Main question 4.1 Who are the public actors interacting with the organisation and what kind of relationship does your company have with them?

Avisé / Chambre française / Conseil Supérieur / Conseil National des Chambres
régionale de l'ESS de l'ESS Régionales de l'ESS

— — — —

Mouves Observatoire National / régionaux de l'ESS Ashoka FISO

Probe 4.1.1: Out of these, which ones have you had interactions with specifically on issues pertaining to social impact measurement?

Probe 4.1.2: What is the level at which you have contact with public stakeholder (i.e. municipal, regional, national)?

Probe 4.1.3: If you have had contacts with some of the following organisations (organisations list above), please describe the interactions

Main question 4.2: How has your relationship with public stakeholders influenced your practices in terms of social impact measurement?

Probe 4.2.1: Were specific methods or metrics put forward by the stakeholders you interacted with?

Probe 4.2.2: Do you perceive any benefits or downsides in using some methods or metrics over others when interacting with public stakeholders about social impact measurement?

Main question 4.3: When interacting with public stakeholders, which aspects of your impact did you feel they were most focused on?

Probe 4.3.1: How could you sense this was a priority for these stakeholders?

Probe 4.3.2: I will now provide you with six potential aspects of performance for a company like yours and will define them for you. I will then ask you to choose and rank three which you thought seemed like a priority for the public stakeholders you talked to.

Main question 4.4: What kind of reaction did you get from public stakeholders when you communicated the results of your impact measurement?

Probe 4.4.1: What do you think motivated these reactions?

Probe 4.4.2: Do you feel that different from your impact measurement would have triggered different reactions?

Probe 4.4.3: What did you perceive was the highlight of your impact from public stakeholders' perspective?

Conclusion: That will conclude our conversation today. I would like to once again thank you for the time you took for meeting me. I may be in touch in the coming weeks if I need to clarify any of the elements we discussed during our time together. In any case, I will contact you again once I finalise my findings to let you know about them.

Appendix 18: Questionnaire used for the quantitative survey (English version)

1: Organisations' activities

This section is about getting to know your organisation and the nature of its activities. Some questions are about characterising your organisation as a social enterprise while others are about understanding the commercial activity (if any) in which it engages.

1. Please, specify the name of your organisation **[FT]**

2. Please, specify the level you occupy in our organisation **[SC]**
 - Executive management
 - Middle management
 - Staff

3. Describe in one sentence what your organisation does **[FT]** – Max. 50 characters

4. Briefly outline the commercial activities (i.e. leave out the social value out of the description) your organisation carries out **[FT]** – Max. 50 characters

5. How would you classify the main commercial activity of your organisation? **[SC]**
 - List: NACE rev.2 / DB07, level 1 (13 items)
 - Your organisation does not carry out commercial activities

6. Do you spontaneously describe your organisation as a social enterprise? **[SC]**
 - Always
 - Often
 - Rarely
 - Never

7. Is your organisation a registered social enterprise under the 2014 Lov om registrerede socialøkonomiske virksomheder? **[SC]**
 - Yes
 - No

8. Kindly specify how the revenue streams of your organisation breaks down

	None	Below 25%	Between 25% and 50%	Between 50% and 75%	More than 75%
Commercial activities with private customers					

Commercial activities with public customers					
Private project funding					
Public project funding					
Subsidies					
Other					

9. Is the creation of social value /social impact a primary objective of your organisation? **[SC]**

- Yes
- No

10. How much of your organisation's revenue is derived from commercial activities (i.e. revenue from goods and services sold to customers, as opposed to contributions such as subsidies, project funding, donations, gifts, etc.) **[SC]**

- None
- Less than 25%
- Between 25% and 50%
- Between 50% and 75%
- More than 75%

11. Is your organisation part of a group?

- Yes, as a subsidiary of a parent organisation
- Yes, as a parent organisation heading one or several subsidiaries
- Yes, as part of a more complex set of relations
- No

12. If your organisation is part of a group, would you say the other organisations in the group belong to the third sector (i.e. social enterprises, foundations, cooperatives, other nonprofits organisations)?

- Yes, all of them
- Yes, some of them
- No
- Not part of a group

13. Which body holds the ultimate decision power in your organisation?

- A board of trustees/directors
- A general assembly/meeting (of members, shareholders, ...)
- A single person (owner, executive director, manager)

- A Parent organisation
- Other (please, specify)

Section 2: About your social value

This section asks about how your organisation generates social value and how you communicate about this value. We also ask you if you have carried out impact measurement.

14. Please, briefly outline the social value your organisation delivers through its activities **[FT]** – Max. 50 characters

15. How important do you perceive the social value you create is for your customers? **[SC]**

- Very: they would probably not do business with you if you were not creating social value
- Important: social value is an important part of their decision to do business with your organisation but is also balanced by other market-driven factors
- Moderate: they would probably do business with you even without the social value your organisation creates

16. How actively are you communicating about your social value to your stakeholders? **[TM]**

Stakeholder	Actively (e.g. dedicated material, sales pitch, specific data)	Passively (e.g. part of regular communication channels but no ad’hoc activities and material)	No formal communication (i.e. information not pushed towards stakeholders but may still be available, e.g. online)
Private customers			
Public customers			
Private financiers			
Public funders			
Employees			
Beneficiaries			

17. Have you attempted to formally measure the social impact of your organisation? **[SC]**

- Yes
- No but I may be interested in doing so in the future
- No and I am not interested in doing so in the future

If Q17 = Yes

18. With what resources did you carry out the measurement of your social value? [MC]

- Internal resource hired temporarily for that purpose (e.g. temp contract, intern)
- Internal resource whose impact measurement is a formal part of job description
- Internal resource whose impact measurement is not part of formal description
- Paid-for external resource (i.e. consultancy or other third-party providing measurement service for a fee)
- Voluntary external resource (i.e. consultancy or other third-party providing measurement service for free)
- Other (please, specify)

19. What is your approach of social impact measurement? [SC]

- You carried out impact measurement once and do not plan to do it again in the foreseeable future
- You are planning on doing at least one other impact measurement in the foreseeable future
- You would like to implement impact measurement as a recurring process
- You already implemented impact measurement as a recurring process

20. What kind of method did you use to measure your social value? [SC] + [FT]

- You built your own impact measurement method
- You used an existing impact measurement method (please specify)

21. Do you know any specific social impact measurement method? If so, please name them. [FT]

22. What kind of quantitative indicators have you used to measure your impact? [MC]

- Account of the outputs generated by your activity (number of individuals supported, number of jobs created, number of products served, etc.)
- Account of the outcomes (i.e. the consequence of the outputs) generated by your activity through the valuing of public incomes and savings (such as extra taxes garnered or avoided social support costs)
- Account of the outcomes (i.e. the consequence of the outputs) generated by your activity through the use of financial proxies valuing intangible social outcomes (such as increased self-esteem, extended social network, etc.)
- Other (please describe)
- You have not used any quantitative measurement

23. What kind of qualitative indicators have you used to measure your impact? [MC]

- Report of self-perceived outcomes (i.e. the consequence of outputs, such as increased self-esteem, extended social network, etc.) as expressed by stakeholders

- Report of perceived outcomes (i.e. the consequence of outputs, such as increased self-esteem, extended social network, etc.) as perceived by your organization or third-parties (experts, thought leaders, etc.)
- Case studies qualitatively documenting impact
- Multimedia files (audio, video, pictures)
- Other (please describe)
- You have not used any qualitative measurement

24. Have any of the stakeholders listed below had an influence on the design or execution of your impact measurement? [TM]

	Has advised on the methodology and / or the choice of indicators	Was a target audience when choosing methodology and indicators	No influence
Private customers			
Public customers			
Private financiers			
Public funders			
Employees			
Beneficiaries			

25. How much the following aspects of your organisations have been influenced by having concrete data to communicate about your impact? [TM]

	Strong positive influence	Mild positive influence	No influence	Mild negative influence	Strong negative influence
Ability to attract public funding					
Ability to attract private funding					
Public customer attraction and retention					
Private customer attraction and retention					
Identify internal management issues					
Develop marketing material					
Generate buzz					

Identify leads to increase impact
Staff motivation

26. If you have communicated your impact data to one or more of the public stakeholders below, please specify the level of interest you perceived from them [TM]

Stakeholder	High interest	Mild interest	No interest	No communication
Job Centre				
Municipality				
Region				
Social Styrelsen or affiliated				
Erhvervsstyrelsen or affiliated				
Styrelsen for Arbejdsmarked og Rekruttering or affiliated				
Other (please, specify)				

If Q17 = No but I may be interested in doing so in the future

27. What has prevented you from measuring your impact until now? [MC]

- Lack of human resources (your staff does not have the time for impact measurement)
- Lack of skills (your staff could find some time for impact measurement but does not know how to do it)
- Lack of financial resources (you cannot afford or do not want to hire a third-party to carry out the impact measurement of your organisation)
- You did not have the opportunity to prioritise it so far
- You want to learn more about impact measurement before deciding whether your organisation should go forth and invest resources in it.
- Other (specify)

28. Do you know any specific social impact measurement method? If so, please name them. [FT]

29. What are the key benefits you envision would be associated with measuring your organisation's impact? [TM]

	Strong positive influence	Mild positive influence	No influence	Mild negative influence	Strong negative influence
Ability to attract public funding					
Ability to attract private funding					
Public customer attraction and retention					
Private customer attraction and retention					
Identify internal management issues					
Develop marketing material					
Generate buzz					
Identify leads to increase impact					
Staff motivation					

30. What kind of quantitative indicators would you be interested in to measure your impact? [MC]

- Account of the outputs generated by your activity (number of individuals supported, number of jobs created, number of products served, etc.)
- Account of the outcomes (i.e. the consequence of the outputs) generated by your activity through the valuing of public incomes and savings (such as extra taxes garnered or avoided social support costs)
- Account of the outcomes (i.e. the consequence of the outputs) generated by your activity through the use of financial proxies valuing intangible social outcomes (such as increased self-esteem, extended social network, etc.)
- Other (please describe)

- You are not interested in quantitative measurement

31. What kind of qualitative indicators would you be interested in to measure your impact?

[MC]

- Report of self-perceived outcomes (i.e. the consequence of outputs, such as increased self-esteem, extended social network, etc.) as expressed by stakeholders
- Report of perceived outcomes (i.e. the consequence of outputs, such as increased self-esteem, extended social network, etc.) as perceived by your organization or third-parties (experts, thought leaders, etc.)
- Case studies qualitatively documenting impact
- Multimedia files (audio, video, pictures)
- Other (please describe)
- You are not interested in qualitative measurement

If Q17 = No and I am not interested in doing so in the future

32. What makes you reluctant in measuring your impact? [MC]

- You do not think that impact measurement could bring value to your organisation
- You think impact measurement may bring value to your organisation but you lack human resources (your staff does not have the time for impact measurement)
- You think impact measurement may bring value to your organisation but you lack skills (your staff could find some time for impact measurement but does not know how to do it)
- You think impact measurement may bring value to your organisation but you lack financial resources (you cannot afford or do not want to hire a third-party to carry out the impact measurement of your organisation)
- Other (specify)

Section 3: About your relations with the public sector

This section explores your relationship with the Danish public sector, particularly in terms of social impact awareness.

33. In the public sphere, which are your primary interlocutors? **[MC]**

- Job Centres
- Municipalities
- Region
- Social Styrelsen or affiliated
- Erhvervsstyrelsen or affiliated
- Styrelsen for Arbejdsmarked og Rekruttering or affiliated
- You have no interlocutors in the public sphere
- Other (please, specify)

34. Has the topic of impact measurement ever come up during your interactions with public actors? **[MC]**

- Yes, to my initiative,
- Yes, to the public actors' initiative as a client
- Yes, to the public actors' initiative as a provider of funding
- No
- Not sure / do not remember
- I do not have interactions with public actors

35. Overall, do you feel that having data to measure your impact would improve your relationship with public actors? **[MC]**

- Yes, it would allow my organisation to have a better access to funding
- Yes, it would allow my organisation to have a better access to public customers
- Yes, it would allow my organisation to access support schemes other than funding (e.g. advisory services, eased access to infrastructure, etc.)
- Yes, it would allow my organisation to get more attention / get its message across
- No, it would not make a difference
- No, it would be detrimental to my relationship with public actors
- I do not have interactions with public actors
- Other (please, specify)

36. Are you operating independently from public authorities (i.e. public stakeholders do not have a significant direct influence on the management of your organisation)? **[SC]**

- Yes
- No

Section 4: Demographics

This last section asks you to specify a few complementary information about your organisation

37. What is your organisation registered as? **[SC]**

- Private limited (ApS)
- Public limited (A/S)
- General Partnership (I/S)
- Limited Partnership (K/S)
- Foundation
- Other (please, specify)

38. How many employees does your company have? **[FT]**

Total number of employees	
Number of full-time equivalents	

39. If you work towards the integration of a specific target group, how many of your employees are part of this target group?

40. How many of your employees, if any, are on a special employment scheme (e.g. fleksjob)? **[FT]**

41. What is the annual revenue range of your company? **[SC]**

- Less than DKK 500 K.
- Between DKK 500 K. and DKK 1 M.
- Between DKK 1 M. and DKK 5 M.
- Between DKK 5 M. and DKK 10 M.
- Between DKK 10 M. and DKK 20 M.
- More than DKK 20 M.

42. In the case where we would be interested in investigating further some of your answers, would you allow us to contact you in the coming weeks? **[SC]**

- I would rather not be contacted again
- Yes (kindly specify your email address)

Appendix 19: Example of interview transcript

45 first minutes :

RESEARCHER : Revenons à vos indicateurs d'impact... Quels sont les principaux indicateurs de performance que vous mobilisez et en quoi vous aident-ils à piloter votre activité ?

INTERVIEWEE 1 : En fait la difficulté quand on gère une entreprise comme la nôtre, réside en plusieurs facteurs: Premièrement, On met le curseur entre le développement économique et un projet social, parce qu'on peut très vite dériver, ça ne se verra pas. Comme moi j'étais très axé sur le développement économique parce que c'était le seul moyen de gagner un peu d'argent, de stabiliser l'entreprise. Moi je me suis dit dans un premier temps, il faut que ce développement profite à des gens en insertion. Donc la première étape de l'impact, on a dit toute croissance économique (mettons 10%) doit permettre une croissance en ETP d'à peu près 10%. Avec une augmentation des résultats nets aussi. Donc ça c'est un critère d'impact : la volumétrie du nombre de personnes que l'on accompagne. Donc dans nos rapports d'activité on va systématiquement s'attacher à montrer les progressions, en CA, en heures travaillées, à c'est 10% sur les deux, donc on voit bien que la croissance économique c'est de la croissance en création d'emplois. Et la croissance en insertion, ben c'est 8%. Donc le fait de suivre ça régulièrement nous permet de dire qu'on travaille bien sur notre sujet. Je dis parce que j'ai aussi des salariés en travail temporaire classiques pour lesquelles on vise un impact différent.

Donc si ma croissance elle n'est faite que sur les entreprises classiques je ne peux pas dire que j'ai augmenté l'impact. Je l'ai peut être maintenu ou abaissé mais il faut être clair. Donc c'est ma première étape: A quoi sert ma croissance économique. Donc ça veut dire que mes commerciaux doivent aller chercher du travail qui est réalisable par des personnes qui ne savent rien faire. Enfin par des personnes en difficulté. Donc c'est indicateur à mon initiative. L'Etat s'en fout de ça. Il me dit – vous êtes conventionné pour 10 postes, est-ce que vous les avez faits ou pas, on me demande pas le CA. Mais nous par rapport à notre gouvernance, on doit montrer que ce que l'on fait sert notre objet social. Et donc si on n'a plus ça on peut se demander à quoi sert l'entreprise. Puisqu'elle

ne sert pas à rémunérer les actionnaires. Donc ça veut dire qu'il y a toujours un arbitrage au niveau des directions. On a 3 directeurs opérationnels par branche d'activité, le DGA et moi. Donc en termes de stratégie, un des moteurs de la décision c'est de se dire – on va se faire chier pour combiner des mecs en plus. Et toute la journée on me dit tes conteneurs c'est sympa, c'est de l'innovation, mais c'est une activité qui est toujours déficitaire et pour un truc qui per des ronds, on n'a que 15 bonhommes en insertion.

Après la réponse c'est qu'avec ça, j'arrive à professionnaliser les gens sur des métiers comme la soudure, la découpe, des métiers très valorisants, et on a 100% de sorties positives sur ce secteur. Donc c'est intéressant par rapport au qualitatif. Là on ne mesure pas la quantité mais la qualité de ce que l'on fait.

Alors ça quand j'en parle à mes partenaires publics, on me dit ok, c'est chouette mais est-ce que t'es dans les clous de la convention ou pas. Donc ça c'est une préoccupation essentiellement du patron, parce que sinon très rapidement on va se retrouver dans une situation où on va se mentir à soi-même, malgré nous. Donc, premier indicateur.

Donc ça c'est des indicateurs non administratifs, c'est des indicateurs qualitatifs. Comment faire pour ne pas perdre en niveau de qualité quand je développe l'entreprise. Et le projet en général est super bien porté par la direction, assez bien par le personnel de terrain, et pas très bien par l'encadrement intermédiaire. Donc là on a un soucis. Parce qu'ils ont entre le marteau et l'enclume. Et donc c'est le revers de médial du développement. Quand on était petit je prenais le café avec tout le monde. Donc pour la mesure de la qualité on a plusieurs facteurs. D'abord on a de la certification. Nos entreprises sont certifiées ISO, mais aussi AFA EI AFAC ETT, donc ça c'est un certificat AFNOR porté par la FEI et qui est expertisé, audité tous les ans par l'AFNOR. Et là dedans on a tous les fondamentaux de ce qu'est une EI de qualité. Donc on va aller voir par exemple, les personnes qui font le suivi social du personnel, quels réseaux de partenaires sociaux ou de recruteurs ou de prescripteurs, elles ont mis en place sur leur territoire. Et donc on contrôle les noms. Donc on va commencer à regarder les points essentiels de notre projet pour voir s'ils sont validés. Donc le gestionnaire de parcours individualisé (GPI / Chargé de recrutement) doit avoir des

réseaux locaux. Ils doivent faire le suivi sur les logiciels dédiés. On doit voir que ce suivi est individualisé. Si on s'aperçoit que c'est toujours les mêmes phrases, ou que les choses sont mal remplies on intervient. On a besoin de connaître le projet professionnel, les freins, l'évaluation technique, comportementale, etc. Donc le fait de se certifier (il n'y a que 10%-15% d'EI certifiées) apporte des garanties...

Et donc ces éléments là vont générer des actions correctives, non conformités, mineures, majeures, et donc chaque année on va réactualiser notre projet et se mettre en conformité avec ce que l'on a décrété. Ca c'est donc un indicateur pour nous.

Autre indicateur quali. Bon on calcule le nombre d'ETP mais ca c'est lié au nombre d'heures, mais on va regarder combien de personnes il faut pour un ETP. Et on mélange tout : EI, travail temporaire, et on fait une moyenne. Par exemple, on voit que le nombre diminue, donc ça veut dire que la même personne travaille plus d'une année à l'autre. Alors ça reste très schématique, c'est une moyenne. Mais par exemple pour les DIRECCTES c'est plutôt le nombre de personnes qui est important plutôt que ce que l'on fait avec, puisqu'ils raisonnent en stats.

Autre indicateur : avec l'évolution des métiers, il y a très peu de métiers non qualifiés. A partir de là, et en se rendant compte qu'on a 15% d'illettrés chez nous, on a mis en place des formations qu'on appelle socle de base, et on fait des formations métier, et on suit cela. Nombre d'heures de formation. 58 K. l'an passé, 103 cette année. Donc les gens deviennent plus compétents, mieux qualifiés, donc c'est également un indicateur qualitatif.

Après comme autre indicateur quali, les sorties positives. Là le taux s'est dégradé. On va aussi regarder le nombre d'accidents de travail avec arrêt. Parce qu'en termes quali, c'est bien beau de dire je fais de l'insertion, mais si les gens ont plus d'accident... Enfin c'est des choses bêtes mais ça personne ne le suit. Donc c'est cette matrice ESG que nous suivons, avec l'accord des actionnaires. Et on fait des projections sur 5 ans. Et on se fixe des objectifs de cette façon. Et ça permet d'apprécier l'entreprise selon différents aspects, en volumétrie et en termes quali.

Bon, nous on a dit que la croissance allait se faire sur les activités environnementales.... Donc on compte les actions mises en place sur les activités environnement, et combien d'heures de travail ça a représenté. Alors on peut aller plus loin, mais c'est ce qu'on suit. Il y a aussi le taux de participation au conseil de surveillance. Et la performance sociale et sociétale de l'entreprise, c'est un peu tout ça.

19 :22

RESEARCHER : Et de quelle façon utilisez-vous ces indicateurs ?

Ca c'est demandé par nos partenaires financiers qui sont des investisseurs à impact et qui consolident cela dans leurs comptes. Mais pas par nos partenaire publics, qui ne s'y intéressent pas. Ceci dit ceux avec lesquels on travaille directement, compte tenu du fait qu'on leur présente nos rapports ils commencent à être sensibilisés à ces questions.

Après en termes quali, on va faire des genres de rapports sociaux où on va regarder pour les permanents par exemple, les effectifs, % homme – femme, cadres non cadres, les flux... Sur les salariés en insertion, on va regarder les effectifs, les ETP, les heures payées, les bénéficiaires du RSA, les niveaux de qualification en dessous de 5, les flux. On identifie aussi les problématiques sociales à l'embauche, on en identifie 5. Déficit de formation initiale (60% en PACA, 33% en Bourgogne), des problèmes de logement, santé, mobilité. Et là on voit les disparités liées au territoire. Et en parallèle on travaille sur l'uniformisation des indicateurs. Et donc ça c'est du qualitatif parce que ça montre bien qu'on traite du personnel prioritaire.

Et là cette année on s'interroge sur nos publics cibles : c'est les personnes en fin de droit, les gens au RSA, les migrants et réfugiés, roms, sortant de prison, les handicapés. Là cette année on s'est attelé à ça : redéfinir à qui on s'adresse. Et là on verra combien de personnes de chaque public on a traité. Et ça nous permet de faire des objectifs. Par exemple les migrants on en parle beaucoup et on en fait pas beaucoup. Donc ça nous permet d'adapter le projet en permanence à la réalité de ce que l'on fait. Le but c'est pas de toujours montrer qu'entre ce qu'on a dit et ce qu'on fait y'a pas

d'écart. Mais c'est de poser tous les ans et de voir comment le projet évolue. Et est-ce qu'on est en phase avec nos objectifs annoncés.

RESEARCHER : Comment cela impacte-t-il votre activité ?

Heu... Ici sur le chiffre d'affaires... Ça c'est les différentes structures... On a 40 millions de CA, mais on a de plus en plus d'activité qui vient d'entreprises traditionnelles que d'activité d'insertion. C'est pour ça que je suis les indicateurs de croissance évoqués avant. Donc là aujourd'hui on se demande si une entreprise classique a sa place dans le groupe. Quelle est leur mission ? On peut dire deux choses : elles ont une mission de stabilité et de consolidation des résultats, ok... Mais disons-le. Soit elles ont vocation à servir l'évolution de notre projet. En l'occurrence nos deux sociétés d'intérimaires ont cette double mission. À la fois elle stabilise notre activité, et à la fois elles font le constat que le simple outil entreprise d'insertion n'est pas suffisant pour mettre en œuvre l'accompagnement que l'on décrit, qui est renforcé, agile, machin. Et donc ces entreprises classiques sont inclusives. Ça va permettre de poursuivre le parcours au-delà de la réglementation des EI. Donc là on se demande comment opérationnaliser cela : comment rendre une société classique inclusive.

J'explique : dans une ETTI, j'ai quelques règles :

- Exclusivité de métier (les permanents ne font que ça, ils ne gèrent pas l'ETT)
- Exclusivité du public (les gens sont tous en insertion, donc avec un agrément de Pôle Emploi)
- Exclusivité de lieu (Mon ETT ne peut pas être dans les mêmes locaux que l'ETTI)

Et donc ça c'est la loi. La durée totale d'une personne en parcours d'insertion, qu'elle ait travaillé avant, chez moi ou pas, c'est 24 mois. Après 24 mois je ne peux plus faire travailler cette personne.

Ça c'est le premier constat.

Deuxièmement, on sait que pour sortir durablement de la précarité c'est un long combat où il faut manier différents instruments. Donc je dis que mon ETT elle a pour mission la sécurisation des parcours. Pour que je puisse reprendre mes personnes en fin de parcours jusqu'à ce qu'elles deviennent réellement indépendantes et autonomes.

Pour pouvoir faire ça, j'ai besoin de mon ETT indépendante avec son propre modèle économique. Je peux y faire le CDI intérimaire, les former. Mais mon but est bien de les faire partir. Et là-dessus on souhaite développer une application pour faire un réseau des anciens de FR1.

Donc ce projet est différent, mais il nourrit toujours un impact social. Tout ça c'est récent, on n'a pas encore l'histoire, mais il faut qu'on développe des indicateurs pour suivre ça. Et là il n'y a pas d'aide publique, c'est 0. Et quand je parle avec la DIRECCTE, si je lui dis mon métier – alors avant notre moto c'était « nul n'est à priori inemployable ». Et puis là c'est plus d'inventer, d'innover pour atteindre nos objectifs. Mais aujourd'hui sur la solution qu'on a créé, avec l'ETT, il n'existe aucun critère de la part de l'Etat. Il faudrait des subventions articulées autour de notre capacité à proposer du travail 6 mois dans l'année à des personnes, mais ça n'existe pas.

Donc on est en train de créer ces batteries d'indicateurs. Et on s'interroge sur la façon dont nous allons structurer nos activités, sur le plan stratégique, dans les 25 années à venir. Je vous ferai parvenir un document à cet égard, que nous sortons pour les 25 ans de FR1.

La vision que j'ai c'est que les entreprises n'arrivent pas à recruter avec 10% de chômage. Moi je le vois, on va chercher des portugais pour notre ETT. Le noyau dur on n'arrive pas à le faire bosser sur ces choses là. Il y a un vrai problème de formation.

Alors on peut aussi demander pourquoi on ne fait pas un partenariat avec Adecco ou Manpower, et pourquoi on investit de l'argent de l'insertion dans le développement d'ETT. Et pourquoi on fait ? Parce qu'aujourd'hui en France in y a le groupe LE, un gros groupe de l'insertion en France, qui ressemble un peu à FR1. Et ils comme actionnaire Adecco, qui détient 40% du travail temporaire d'intérim. Et donc ils n'ont pas leur propre ETT, ils passent par Adecco, depuis 15 ans. Et ça ne marche pas. Parce qu'ils sont en concurrence sur le territoire, parce qu'il y a des prestations sur lesquelles les deux peuvent proposer des services. Et je pense que c'est parce que ces structures ne sont pas sous le même management.

Et nous va créer des coachs transversaux, des sortes de référent uniques. On va résonner en matriciel. De cette façon il est plus facile d'individualiser le suivi et être plus efficace dans ce que l'on fait. C'est donc le parcours hyper rapproché, hyper stratégique pour le noyau dur.

Et donc moi il me semble que les entrepreneurs comme nous, on s'est endormi sur nos lauriers pendant des années. Mes collègues sont contents parce qu'ils respectent la triple exclusivité. Mais est-ce que ça correspond au besoin des publics ? Ben nous on annonce dans les rapports. Et on voit qu'on a dégradé nos sorties.

Qu'est-ce qu'on fait, on maquille ? Non, on doit se demander comment répondre aux nouveaux besoins. On travaille avec des méthodes vieilles de 30 ans. Et si on ne fait rien on perd en pertinence et on se fait remettre en question par l'état, qui préférera créer d'autres mécanismes.

Quand l'Etat ne vient plus, on est les premiers fautifs ! La loi suit les entrepreneurs, elle n'est pas devant. C'est pas parce qu'on fait une loi qu'il y a des entrepreneurs qui naissent. Aujourd'hui on essaye de taxer les GAFAs : la loi elle suit. Et nous on est encore dans des schémas où on veut être les bons élèves de la classe – ha j'ai pas les postes je ne vais pas trop forcé alors. Ou alors je suis content j'ai eu tous les agréments.

Ok, mais qu'est ce que tu as à dire sur les personnes ? Moi ça m'a pris des jours de faire dire à mes employés que certains mecs ne veulent pas rester dans leur poste. Ils ne voulaient pas leur dire, parce que ça veut dire que c'est des mauvais ! Je dis non, vous n'êtes pas mauvais, c'est un constat, il faut donc travailler de façon différente. Et donc ça se traduit dans l'impact, et il faut qu'on soit capable de l'expliquer comme je vous l'explique pour que les gens en face se disent – ha ouais c'est pas con, on va peut être mettre en place d'autres indicateurs que les ETP.

Et moi je dis plutôt que d'avoir un agrément de Pole emploi, donc un contrôle à priori. Faisons un contrôle à posteriori. Audit dur tous les ans. Faites-moi venir vos spécialistes de je sais pas quoi, ils vont regarder les dossiers, recrutement, formation, qu'est-ce que vous avez fait, pourquoi, etc. Et là on est dans le challenge, on commence à flipper ! L'agrément vous savez comment ça se passe ?

Pôle emploi il met... donc.. les agréments sont pour ceux qui sont plus éloignés, le noyau dur. Ceux qui ne vont pas à Pole Emploi. Ils sont inscrit mais on les voit jamais. Qui sont les agents Pole Emploi qui traitent de ça ? Des stagiaires, des emplois aidés, ou des novices. Ils ne comprennent rien à ce qu'on raconte là. Enfin ils en ont pas la pratique. Et après, cas récent : On nous refuse un agrément à Aubagne : une nana au RSA qui ne bosse pas depuis 3 ans... Alors je lui ai dit va à la Ciota. Là le dossier est passé et j'ai eu l'agrément. Y'a toujours un système pour dévier le système. Si on ne raisonne que système, ou que statut, ou que machin... C'est la dictature.

Après est-ce que c'est logique d'envoyer des gens à la Ciota ? On a mieux à faire... Après je ne veux pas faire le négatif, on fait partie des nantis comme on a dit. Mais je pense que le problème est chez nos acteurs, mais il faut aussi que l'Etat se réveille. On va faire mieux, peut être avec moins d'argent. Sur la France il y a peut être 1 000 personnes à Pole Emploi qui traitent de l'insertion, avec un impact 0, ça sert à rien, supprimer moi les 1 000 emplois vous allez faire des économies. Vous mettez l'équivalent du salaire de ces personnes pour nous aider, ça sera plus efficace.

Après c'est du courage politique, amis il faut le susciter ce courage. On ne peut pas demander à un Enarque qui travaille à la DGEFP avec un ministre fraîchement nommé qui ne comprend rien sur ces sujets là, on ne pas leur demander d'inventer ça. C'est à nous de l'inventer, de proposer de faire le lobbying nécessaire. Pour le faire, c'est des indicateurs (tape sur la table). Nous on a la taille on a le droit de ne plus mentir. Quand on est petit on ne peut pas se permettre de dégrader les sorties. Il faut faire des rapport, l'Etat vient s'immiscer dans la gestion... Moi je peux me permettre de les envoyer, mais pas les petits. Donc c'est à nous, un peu leaders, de prendre le flambeau et de dire maintenant stop, on va faire comme ça. Et Je pense que le gouvernement actuel est prêt à l'entendre.

RESEARCHER: C'est quelque chose que vous avez expérimenté ?

INTERVIEWEE 1: Oui. Là je vous fais la déclinaison opérationnelle du discours politique, c'est le rôle de l'entrepreneur. Je sais que lors du congrès, le Haut Commissaire à l'ESS a annoncé officieusement que « l'agrément on a compris. Il vaut mieux un contrôle à posteriori qu'à priori ». Donc quand on

expose clairement les choses de temps en temps on peut être entendu. Donc c'est ça qu'on doit faire. C'est pour ça que la question de la mesure de l'impact... Quand ça vient de nos financeurs, au bilan, avec qui on doit être ne phase sur nos performances et sur notre objectif. Ben ça ça doit se construire avec eux. Là je vous parle d'une entreprise qui est la mienne. Mais l'impact social peut être dans plein d'autres choses, le sport, etc. Et comment un fond qui va investir dans pleins de trucs va pouvoir comparer ? Ça sert à ça... Donc quelle est l'objectivité qu'on a ? Alors là Phitrust, un de nos actionnaires, a monté un système, qui dit « alors là comme je vous ai demandé de planifier à 5 ans »... Là c'est vos prévisions (trace un graphique). Chaque année, je regarde les écarts entre prévisionnel et réalisé. Et je nous positionne sur un graphique avec différents points. Un c'est l'ambition et deux c'est l'atteinte des objectifs. Et donc ils obtiennent un nuage de points correspondants à plusieurs structures, là c'est FR1, et donc ils peuvent comparer. Ça permet de comparer des choux et des navets. J'aimerais bien qu'ils vous expliquent, mais la personne qui est là a remplacé celle qui a créé ce système chez Phitrust.

Et donc notre actionnaire a un comité d'engagement. C'est des privés qui mettent des fonds, et ces privés coachent des entreprises. Donc nous on a quelqu'un au conseil de surveillance et avec qui je suis en contact en permanence et qui me tient informé de là où ils en sont, me demande si j'ai besoin d'aide... Et quand arrive la fin de l'année, il me dit ne met pas d'objectifs trop ambitieux sinon tu ne sortiras pas bon sur le tableau. Soit modeste, parce que les autres font ça. Donc ça veut dire que c'est regardé déjà, et puis si on n'est pas premier, et bien on peut progresser.

Summary

With an aim of providing employment to a growing pool of marginalised individuals, work integration social enterprises (WISEs) have progressively become valuable partners in the provision of employment policies. Today, their performance is coming under increasing scrutiny in the context of public reforms inspired by the new public management (Hood, 1991) and its offspring such as the Neo-Weberian State (NWS, Pollitt and Bouckaert, 2004) or the New Public Governance (NPG, Osborne, 2006).

But how does one measure WISEs' broad goal of inclusion? And how does this measurement affect the partnerships that WISEs form with their public stakeholders? These questions pertain to a strand of economics that encourages new ways to frame performance, particularly in a policy-making context (see e.g. Gadrey and Jany-Catrice 2016). Our thesis investigates this issue through an international comparison between France (a country pursuing NWS-type reforms) and Denmark (pursuing NPG-type reforms), using public value theory (PVT, Moore, 1995) as an analytical framework.

Our thesis brings an emphasis on the notion of performativity (the implications that the measurement of a social phenomenon can have on this very phenomenon). Performativity matters because it shows that the measurement of performance has repercussions on outcomes (i.e. poorly designed indicators can lead to detrimental actions) and on democracy (i.e. poorly understood indicators can mislead policies and the comprehension of their impact).

We ascribe our international comparison to Maurice's societal approach (1989), which allows us to ground our case studies in their environment. We also adopt a critical realist approach (Bhaskar, 1975; Lawson, 2003), whose layered ontology provides a relevant framework to investigate intangible concepts such as performance.

The contribution of our thesis is threefold. First, we offer three entries to better capture the notion of performativity. We expend particularly on Callon's socio-technical arrangements (STAs), or the entanglement between the belief systems surrounding performance and what they measure (2006). While STAs are typically presented as a source of performativity, our research identifies different feedback mechanisms for them, some of which can be

leveraged to reduce performativity. Specifically, we see that the development of multiple STAs (through e.g. WISEs creating their own performance measurement model) is conducive to deliberation, which reduces performativity.

Second, we isolate two causal mechanisms for performativity. The first one is attributable to WISEs' public management environment (NWS or NPG), which establishes the model through which they co-create public value. For instance, the NWS' legalism is a structuring force of work integration in France where WISEs derive their legitimacy from securing a covenant, whereas in Denmark the NPG creates a relationship-based quasi-market where WISEs' legitimacy is assessed both on market terms but also through the build-up of relational capital and shared values.

The second causal mechanism is the strategic intents that underpin WISEs' performance measurement in the context of their partnership with public actors: capitalising on PVT, we identify seven intents which are intertwined with the environment. For instance, in France, we identify a comply intent (in the sense of complying with ill-fitted performance metrics used by public actors) that is conducive to behaviours not driven by WISEs' mission but by the criteria they must meet to maintain their covenant (e.g. in the recruitment of their beneficiaries).

Lastly, we propose a state of the art and a framework to apprehend social impact assessment (SIA) models. SIA is the approach to performance measurement that has been developed in the last two decades by mission driven organisations (such as WISEs) attempting to capture their impact (e.g. SROI, which monetarise impact through the use of financial proxies).

Les structures de l'insertion par l'activité économique (WISE) ciblent l'accompagnement vers l'emploi d'une population croissante d'individus marginalisés. Ce faisant, elles sont devenues des partenaires privilégiés dans la mise en œuvre des politiques de l'emploi. Aujourd'hui, à l'heure où les deniers publics sont comptés, et dans le contexte de réformes publiques inspirées par le New Public Management (Hood, 1991), leur performance fait l'objet d'un examen de plus en plus attentif.

Mais comment mesurer les objectifs d'insertion poursuivis par les WISE ? Et comment cette mesure affecte-t-elle les partenariats que les WISE nouent avec les acteurs publics ? Ces questions s'inscrivent dans un pan de la littérature économique qui encourage de nouvelles façons d'appréhender la performance, particulièrement dans le contexte de l'élaboration des politiques publiques (voir, e.g. Gadrey et Jany-Catrice, 2016). Notre thèse examine ces questions à travers une comparaison entre la France (un pays poursuivant des réformes de type néo-Webériennes NWS) et le Danemark (réformes de type nouvelle gouvernance publique NPG), en utilisant la théorie de la valeur publique (PVT, Moore, 1995) comme cadre analytique.

Notre thèse met l'accent sur la notion de performativité (les implications que la mesure d'un phénomène peut avoir sur ce même phénomène). Ce concept met en exergue les répercussions que la mesure de la performance peut avoir en termes de résultats (des indicateurs mal conçus peuvent conduire à des résultats négatifs) et sur ses implications pour les processus démocratiques (des indicateurs mal compris peuvent induire en erreur les politiques et la compréhension de leur impact).

La contribution de notre thèse est triple. Tout d'abord, nous proposons trois entrées pour mieux saisir la notion de performativité. Nous nous intéressons particulièrement aux agencements socio-techniques (AST) de Callon, ou l'enchevêtrement entre les systèmes de croyances entourant la performance et ce

qu'ils mesurent (2006). Bien que les AST soient généralement présentés comme une source de performativité, notre recherche y attribue différents mécanismes de *feedback*, dont certains peuvent être utilisés pour réduire la performativité. Plus précisément, nous constatons que l'élaboration de multiples AST (e.g. des WISE créant leur propre modèle de mesure de la performance) est propice à la délibération, ce qui réduit la performativité.

Deuxièmement, nous isolons deux mécanismes de causalité associés à la performativité. Le premier est attribuable à l'environnement de gestion publique des WISE (NWS ou NPG), qui établit le modèle par lequel elles co-crée de la valeur publique. Par exemple, le légalisme du NWS est une caractéristique structurante de l'intégration professionnelle en France où les WISE tirent leur légitimité du conventionnement, alors qu'au Danemark, la NPG crée un *quasi-marché de réseau* où la légitimité des WISE s'apprécie à la fois par des mécanismes de marché mais aussi par la constitution de capital relationnel et de valeurs partagées.

Le deuxième mécanisme causal est l'intention stratégique qui sous-tend la mesure de la performance des WISEs : en se basant sur la PVT, nous identifions sept intentions qui sont liées à l'environnement. En France, nous identifions e.g. une intention de conformité (aux indicateurs de performance peu adaptés utilisés par les acteurs publics) qui conduit à des comportements non pas motivés par la mission des WISEs mais par les critères qu'elles doivent remplir pour maintenir leur conventionnement (e.g. dans le recrutement de leurs bénéficiaires).

Enfin, nous proposons un état de l'art et un cadre pour appréhender les modèles d'évaluation de l'impact social (SIA). SIA est l'approche de la *mesure* des impacts qui a été élaborée au cours des deux dernières décennies par des organisations dites d'intérêt général (comme les WISEs) qui tentent d'évaluer leur impact (par exemple SROI, qui monétarise l'impact).