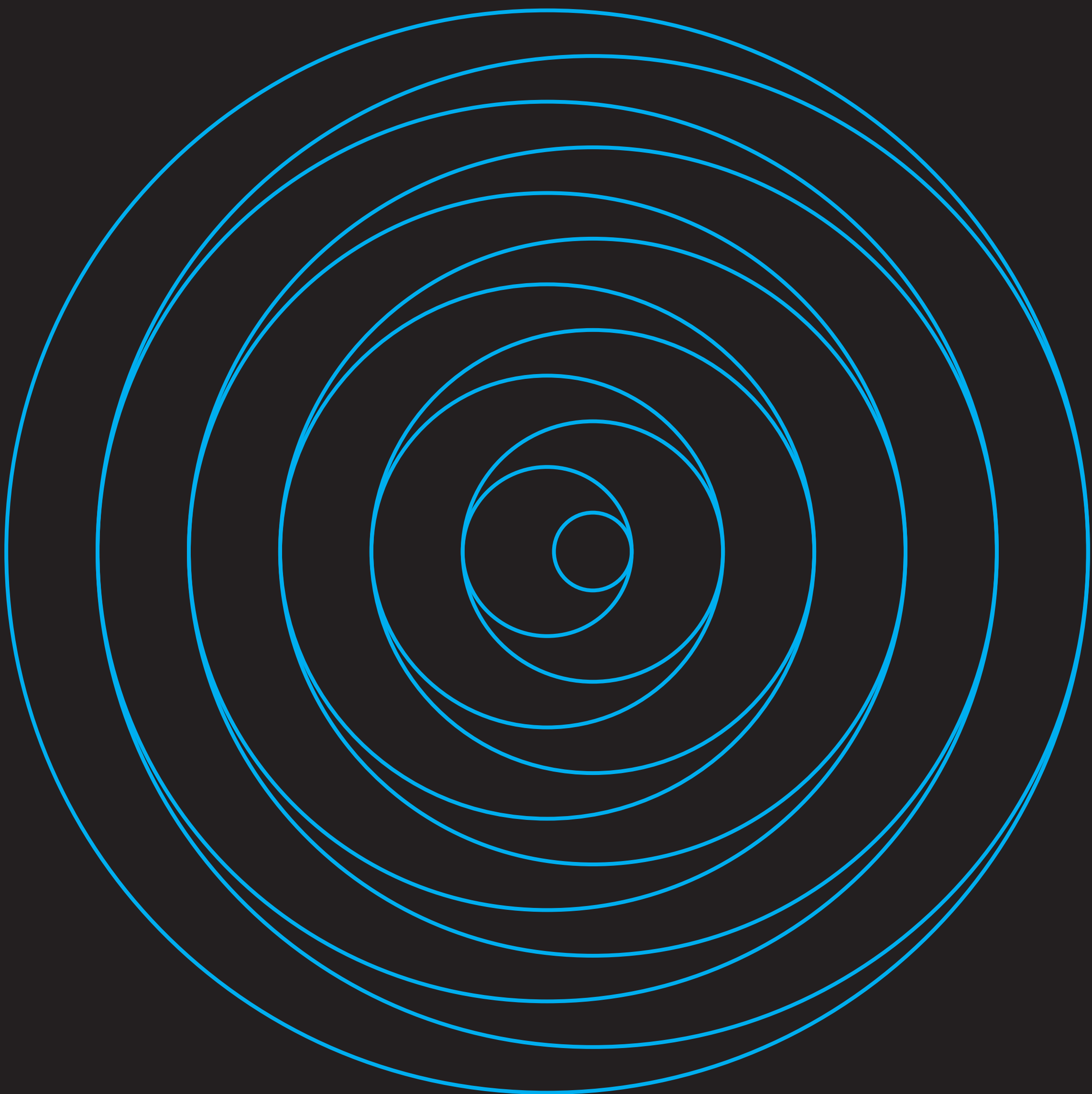
Designing in data-centric policymaking

An exploration of data for policy and policy learning in data ecosystems

Doctoral Dissertation - PhD in Design
Phd Candidate: Francesco Leoni



POLITECNICO
MILANO 1863

DIPARTIMENTO DI DESIGN

POLITECNICO
MILANO 1863

DOTTORATO
DI RICERCA
IN DESIGN

Francesco Leoni

Designing in data-centric policymaking

An exploration of data for policy and policy learning in data ecosystems

Politecnico di Milano
Department of Design

Doctoral program in Design
Coordinator: prof. **Lucia Rampino**
XXXIV Cycle

Supervisor: prof. **Stefano Maffei**
Co-Supervisor: prof. **Grazia Concilio**

This doctoral research has received an interdisciplinary scholarship from Politecnico di Milano, jointly funded by its PhD programmes in "Design" and "Urban Planning, Design and Policy".

POLITECNICO
MILANO 1863

PhD Candidate: **Francesco Leoni**
francesco.leoni@polimi.it

Politecnico di Milano
Department of Design

Doctoral Program in Design
XXXIV Cycle

Editorial & cover design: Andrea Saccavini

Index

Abbreviations

CDS — Critical Data Studies
CR — Critical Realism
EC — European Commission
EU — European Union
DDI — Data-driven Innovation
ICTs — Information and communications technologies
OECD — Organization for Economic Co-operation and Development
QCA — Qualitative Comparative Analysis

Figures

Tables

Annexes List

Abstract (ENG)

A considerable part of the contemporary scientific and political debate sees the digitalization and digitisation of today's sociotechnical systems as an unprecedented possibility to create innovation from digital data. In this view, the relationship between data and innovation is mainly based on a paradigm of extracting value from data. Accordingly, it is possible to reduce uncertainty among possible scenarios through the analysis of large amounts of data, enabling decisions to improve the efficiency of operations in a system.

This innovative value proposition has been progressively examined with respect to the public sector and policymaking, in an emerging field of discussion called 'data for policy'. In this field, some authors noted that the value of data for policymaking cannot be discussed with the same logic of efficiency that has characterised the data debate.

Policymaking is in fact a process that implies a normative and partial view of public issues, and is therefore connected to mechanisms of political judgement and public acceptability, dimensions that lie outside the logics of efficiency. Rather than in the data itself, for policymaking there seems to be value in the processes centred on the collection and use of these data, which can constitute new forms of experimentation and collective learning on policy problems. From these considerations, the thesis conducted a qualitative exploration of the current discourse in the emerging field of "data for policy"; and a comparative analysis of data use practices within data ecosystems in the public sector in four different European countries. The thesis proposed the concept of **data-centric policymaking** to conduct a comparative analysis and developed this concept through a theoretical-conceptual framework based on policy learning. The comparative analysis shows that those involved in data-centric policymaking practices perceived greater individual cognitive learning on policy issues. However, there does not seem to have been a fundamental change in their opinions through participation in the process. The conditions for this learning seem to have depended not on structural enabling condition for data sharing, but on conditions at the organisational and individual level. From this knowledge, the thesis proposed three areas of convergence between 'data for policy' and 'design for policy', articulating the potential contribution of design in data-centric policymaking. The thesis has thus contributed to a better understanding of the topic of "data for policy" and data practices in the public sector, while offering an interpretation of the phenomenon in relation to policymaking and design.

Abstract (ITA)

Una considerevole parte del dibattito scientifico e politico contemporaneo considera la digitalizzazione e digitizzazione dei sistemi socio-tecnologici odierni come una possibilità senza precedenti per creare innovazione a partire dai dati digitali. In questa visione, la relazione tra dati ed innovazione si basa principalmente su un paradigma di estrazione di valore dai dati. Di conseguenza, è possibile ridurre il grado d'incertezza tra gli scenari possibili attraverso grandi quantità di dati, permettendo decisioni volte a migliorare l'efficienza delle operazioni in un sistema. Questa proposizione di valore innovativo è stata progressivamente presa in esame rispetto al settore pubblico ed al policymaking, in un campo di discussione emergente detto "data for policy". In questo campo alcuni autori notano come il valore dei dati per il policymaking non possa essere discusso con le stesse logiche di efficienza che hanno caratterizzato il dibattito sui dati.

Il policymaking è infatti un processo che implica una visione normativa e parziale dei problemi pubblici, ed è quindi connesso a meccanismi di giudizio politico e all'accettabilità pubblica, dimensioni che esulano le logiche dell'efficienza. Piuttosto che nei dati in sé, per il policymaking sembra esserci valore nei processi che si incentrano sulla raccolta e l'utilizzo di questi dati, i quali possono costituirsi come nuove forme di sperimentazione ed apprendimento collettivo sui problemi di policy. A partire da queste considerazioni, la tesi ha condotto un'esplorazione qualitativa dell'attuale discorso nel campo emergente del "data for policy"; ed un'analisi comparativa delle pratiche di uso dei dati all'interno di data ecosystem nel settore pubblico in quattro diversi paesi Europei. La tesi ha proposto il concetto di **data-centric policymaking** per condurre un'analisi comparativa e ha sviluppato questo concetto attraverso un framework teorico-concettuale basato sul policy learning. L'analisi comparativa mostra come coloro coinvolti nelle pratiche di data-centric policymaking abbiano percepito un maggiore apprendimento cognitivo individuale su tematiche di policy. Tuttavia, non sembra esserci stato un cambio fondamentale delle loro opinioni attraverso la partecipazione nel processo. Le condizioni di questo apprendimento sembrano essere dipese non da fattori strutturali abilitanti per la condivisione dei dati, ma da condizioni a livello organizzativo e individuale. A partire da questa conoscenza la tesi ha proposto tre aree di convergenza tra il "data for policy" ed il "design for policy", articolando il potenziale contributo del design nell'ambito del data-centric policymaking. La tesi ha quindi contribuito ad una migliore comprensione del tema del "data for policy" e delle pratiche di utilizzo dei dati nel settore pubblico, offrendo una lettura del fenomeno in relazione al policymaking ed al design.

Acknowledgements

To get a doctorate degree takes much more than just doing your job right. As for any other life's adventure, it requires tons of human resilience, a sharp critical sense, and every bit of support that one can get. Of the many people I have encountered during the last four years, many are those which intellectual and practical contribution led me to the end.

I am firstly and foremostly in profound debt to my supervisor and mentor, prof. Stefano Maffei, who entrusted me with this research and allowed me to work at an intersection of topics which I am deeply passionate about, design, technology, and policymaking. Our "robust" disagreements on the thesis shook the university's walls and will long remain unforgotten. However, I would dare to say that, rather than on literature or methodology, we might have arrived to a mutual (uncompromising) worldview on the human, intellectual and political material that makes a scholar and a public servant. In that respect, you provided me with a rare and generous example in practice, rather than theory.

The same applies to all our remarkable colleagues, who saw me growing as researcher since my graduation. To the remarkable, gritty, and skilful bunch that makes our research group, shall go my immense gratitude: prof. Massimo Bianchini, prof. Beatrice Villari, Barbara, Carla, Francesca, Patrizia. From you I learned how to dirty my hands in doing what needs be done in research and teaching work. A great deal I have also learned from the colleagues of Polifactory and the Specializing Master in Service Design (both who stayed and who left): Raffaela, Laura, Enrico, Diego, Bryan, Andrea, Lisa, Veronica, Luca, Aurelié and Leonardo. A special part of my gratefulness for my research team goes to Erin (which generously shared her experience from years of practice in government) and to Martina; the two fellows with whom, within the Design Policy Lab, I shared the passionate yet daunting challenge of building PhDs at the intersection of design and policymaking. I reach this far, guys. Now is your turn to taking up the baton.

To my co-supervisor, professor Grazia Concilio, and the the Department of Architecture and Urban Studies of Politecnico di Milano, shall also go my deepest gratitude for supporting this research. Grazia, thank you for trusting my choices, being supportive, sharing your knowledge on data for policy, and making me part of insightful projects, such as Safari Njema. My hope is that we have just began to work and think together. In DASTU, I would like to dedicate a special thanks to Giovanni Lanza, always willing to help and share with a PhD fellow puzzling on similar topics. My gratitude goes to the board of PhD program in "Urban Planning, Design and Policy", who welcomed a "disciplinary stranger" during the official reviews.

Going back to the place that became my second home during these years — the Department of Design — many are those toward which I feel thankful, and these include all the professors whose work and dedication improved my research, either directly or indirectly, through formal or informal exchanges. Hence, I am grateful to prof. Daniela Sangiorgi, prof. Cabirio Cautela, prof. Valentina Auricchio, prof. Marzia Mortati, prof. Lucia Rampino, prof. Paola Bertola, prof. Barbara Camocini, prof. Venere Ferraro, prof. Fiammetta Costa, prof. Anna Anzani, prof. Alessandra Spagnoli, prof. Marco Quaggiotto, prof. Valentina Rognoli and many others.

Many thanks are also in order for the administrative staff of the PhD Program in Design, Simona Murina and Paola Mosca, for their constant support.

Then, I wish to thank my peers in the PhD Program in Design. Together, we went through several reforms in the PhD program in Design. I suspect that one day we might be regarded as prototypes. For now, let them simply remind us as the semi-mythical 34th cycle.

Thanks then to Francesca, Ingrid, Emilio, Lily, Erminia, Umair, Beatrice, Claudia, Clorinda, Flavia, Gabriele, Marta, Mila, Jack, Zhang for the time spent together during lessons, workshops, and outside the university. My special gratitude goes to Pamela, the only other PhD of 34^{th} working on "design for policy". Pamela, we certainly learned how hard it is to deal with complexity, and I am in debt with you for our exchanges and for making me a part of the Designing Policy Network, whose members, Verena, Sofia, Federico, Ahmee, Frithjof, Luis, Natalia, Beatriz, I would like to thank for the same reasons. I am looking forward for our next conversations.

Looking outside Politecnico — and Italy — gives me the occasion to express my gratitude to other people who supported me. I am grateful to the Interaction and Service Design Research Group of Linköping University for accepting me as visiting PhD at the end of 2020; a very difficult moment for international mobility. Thanks to Stefan, Mattias, Johan, Vanessa, Eva, Ludwig, and many others I met at the Division of Human Centered Systems. Despite of the pandemic restrictions, you made me feel welcomed in winter-time Linköping with you human and intellectual generosity. Another round of thanks I wish to send to all scholars I met at the University of Padova during the IPPA Summer School in 2018. Many thanks to Giulia Bazzan, prof. Maria Stella Righettini, prof. Claudio Radaelli, prof. Eva Thomann, prof. Susana Borras and all the PhD students in political science that I could interact with in the beautiful frame of Venice and Padova, in a time that seems now so far, and thanks to who I learnt theoretical and methodological aspects later to become central to my research, as QCA and policy learning.

I desire also to thank those who directly contributed to the painstaking craft of research and dissertation writing. Thanks to prof. Nicola Morelli and prof. Joep Crompvoets for their essential reviews and patience. Thanks to Elena Cipressi for her tips on descriptive statistical analysis. Thanks to my brother Guido, for his practitioner insights from the world of business analytics. Thanks to my friend Gabriele, for the technical support with Zenodo API scraping. Thanks to Andrea Saccavini for curating the cover design and the graphic layout of this manuscript.

I am in debt with the several experts, practitioners, civil servants, and scholars which I met (often virtually) across Italy, Sweden, Belgium, United Kingdom, Netherlands and Australia — they are the actors of the field of “data for policy”, who made this research possible.

Finally, I am grateful to all my friends across Milan and Modena for their friendship and support, and in particular five individuals I hope to see more often from now on. Gratitude and affection shall also go to all the members of my big and scattered family: from your support, I derived the strength to endure until the end.

Above all, is to my two younger brothers, Guido and Elia, whose hearts keep me forward, that I desire to dedicate this book. And to Camilla — finally and always at the end — which heart stayed behind me the whole time, and will always remain beyond my language.

Publications

Maffei, S., Leoni, F., & Villari, B. (2020). Data-driven Anticipatory Governance. Emerging scenarios in data for policy practices. In L., Kimbell & L., Vesnic-Alujevic (Eds.) The Future of Government: Exploring Futures through Design and Foresight [Special issue]. Policy Design and Practice. https://doi.org/10.1080/25741292.2020.1763896

Leoni, F. (2020). Design for policy in data for policy practices. Exploring potential convergences for policy innovation. Data for Policy 2020 5th International Conference. Zenodo. https://doi.org/http://doi.org/10.5281/zenodo.3988176

Introduction

The potential innovation that data might represent for public policymaking constitutes the area of investigation of this thesis. This inquiry is increasingly concretizing into a field identified as data for policy, that discusses experimentations centered on data within public settings (see Section 1.4.1). Before dwelling on this area as part of the thesis's problem setting (see Chapter 1), this introduction wishes to take a step back and ask: *why has it become nowadays imaginable to put digital data and innovation in relation to each other? Further, how has this relation mainly been interpreted by governments?* These are not to be considered the thesis' research questions (see Section 1.8), but intend to summon the broader phenomenological background against which this doctoral inquiry develops.

Governments worldwide today consider digital data as central drivers of public sector innovation agendas (OECD, 2019, pp. 145–155; UN, 2020, pp. 145–175). The importance given to data might be the result of a longstanding political discourse that proposed, as part of a narrative of positive progress, a relationship between *digital data* and *innovation*. It is thus not hard to find, in recent times, explicit expressions of this narrative as part of political statements and official documents produced by governmental bodies[1]. A major cause behind the current proposition of data as source of innovation could be found in the impact of Information and Communication Technologies (ICTs), which constituted a disruptive factor within the historical interplay between technological innovation and governments[2]. During the last decades, ICTs[3] greatly affected not only the public sector but society at large (Castaldi & Dosi, 2010) through a technological revolu-

1 In Europe, for example, data was proposed already in 2013 as "*the fuel for innovation*" (Kroes, 2013) by Neelie Kroes (Rieder, 2018), the former Vice-President of the European Commission responsible for the Digital Agenda under the second Commission of José Barroso (2010-2014). However, a similar narrative can be noted in Europe politics both in earlier political documents, as the Bangemann report (1994) (that instead of data talks about public information) or recent ones, as the European Parliament Directive for open data and the re-use of public sector information (European Parliament, 2019).

2 To understand the relationship between technology and governments I here refer to the historical reviews of Margetts (1999), Agar (2003) and to the timeline of Garson (2010).

3 Hamelik (1997) defines ICTs as an encompassing group of technologies: "[…] that enable the handling of information and facilitate different forms of communication among human actors, between human beings and electronic systems, and among electronic systems." (Hamelink, 1997, p. 3).

tion that brought global scale sociotechnical transformation[4].

The economist Carlota Perez suggested that we now live in "*the Age of Information and Telecommunications*" (Perez, 2010)[5]. According to Perez, technological revolutions are characterised by a series of technological innovations[6] of global impact, which transcended their original contexts because being identified as a pivoting element for new models of economic development (Perez, 2010). As a *general-purpose technology*[7], ICTs came to signify more than single technologies and formed the basis of an original techno-economic paradigm upon which national economies[8] could shape their capitalist models of production and growth (Castaldi & Dosi, 2010).

ICTs transformation did not only diffuse through many technological advancements and innovations[9] but thanks to social and cultural change. Today, the commercial diffusion of devices with unprecedented high computational/storage capacities (Hilbert & Lopez, 2011) matches with societies' expectations of always accessing these technologies (Floridi, 2016). This sociotechnical transformation led to growing digitization of telecommunications and the digitalization of services and activities in

4 In using the concept of socio-technical systems or socio-technological systems, I refer to the work of Trist (1978), Mumford (2006) and Fuchs (2005). These concepts define the underlying perspective of socio-technical studies (STS), an action research approach that highlights the embodiment of technology in societies and the interplay between technical systems and social systems (Trist, 1978).

5 Technological revolutions are defined by Carlota Perez as: "*a set of interrelated radical breakthroughs, forming a major constellation of interdependent technologies; a cluster of clusters or a system of systems*" (Perez, 2010, p. 8).

6 According to Perez (2010), the current technological revolution, the *Age of Information and Telecommunications,* would represent the fifth of such events in human history.

7 In other words, those technologies capable of bringing innovation across different application domains (Bresnahan & Trajtenberg, 1995).

8 It must be noted that ICTs might be only partly a global phenomenon. In the words of Castaldi and Dosi: "The ICT- based techno- economic paradigm is occurring within a regime of globalization of international economic exchanges, but not of globalization of technological capabilities [...] The capabilities of mastering new technologies are unevenly distributed across countries, and technological leaders explore possible applications of ICT- based technologies." (Castaldi & Dosi, 2010, p.52).

9 Perez (2010) proposes the announcement of microprocessors in California in 1971 as the first of these technological innovations. Others noteworthy development were: TCP/IP protocol and packet switching in the early 70s (Roberts, 1978); first sensors network at DARPA in the 80s (Chong & Kumar, 2003); the first generation of the telecommunications network, also started in the 80s (Jia et al., 2018).

the industry and market (Hilbert & Lopez, 2011)[10]. The exchanging and processing of digital data thereby became an intrinsic counterpart for almost all social and organizational practices (Parisi, 2018), thus ubiquitous computing appears as a trait of our contemporary societies (Hirschheim & Klein, 2011).

All these factors constituted the fundamental premises of any radical innovation proposition based on digital data, as these data are now recorded, transmitted, stored and processed as never happened before (Kitchin, 2014a) in a digital landscape of interconnected sociotechnical systems[11]. On these premises, many started to propose a revolution centered on data (Kitchin, 2014) whose revolutionary factor would essentially pertain the unprecedented possibility to obtain and analyse these data (Kitchin, 2014). In the last years, there has been a substantial body of literature that, starting from this argument, debated the urgency and meaning of building innovation on data (see Section 1.1.1).

However, many questions remain unanswered about how to concretely reach this innovation in contexts — as the public sector — where scarce data capacities and culture seem to exist (Giest, 2018; Klievink et al., 2017). Moreover, the value entailed with the innovation of data shall remain dependent by the competing social groups visions[12] of what constitutes a desirable future state of things (Jasanoff & Kim, 2009) (see Section 1.2.1). In Europe in particular, the political discourse seems to have conceptualized data as an economic asset and consequently equating innovation with increased economic growth and market advantages (Rieder, 2018). This way of seeing data might be explained by the way the topic entered the political debate, as public decision-makers realized that public sector data and information are an important economic asset within the growing

10 *Digitization* refers to the purely technical conversion of analog signals into digital ones (e.g., audio recording on a smartphone). *Digitalization* refers to the central role that digital technologies increasingly have on social and economic activities (Brennen & Kreiss, 2016).

11 These systems in practice define either groups of individuals actively interacting through computers networks, or even software and hardware, semi-automatically or automatically exchanging and recording data as part of their programming.

12 Here, I adhere to the idea that any technological innovation is not given but shaped after the perspectives and interests of social actors, which will define the innovation that concretely impacts societies (Akrich et al., 2002).

digital market[13], that requires regulations[14] (Gray, 2015; Janssen & Dumortier, 2003). This perspective seems well-captured by the Organization for Economic Co-operation and Development (OECD) that defines data-driven innovation (DDI) as: "*the use of data and analytics to improve or foster new products, processes, organizational methods and markets*" (OECD, 2015, p. 21). Perhaps due to the abiding "Big Data" narrative (see Section 1.2.3), the aggregation and analysis of large amounts of data seems to remain a central aspect of value creation in DDI. Accordingly, data analytics data could reduce uncertainty and inform decision-making to increase productivity (Brynjolfsson et al., 2011) (see Section 1.2.3).

However, an innovation scenario wherein DDI is equally and successfully applied to the private and public sector, appears rather problematic. The public sector has a long tradition of large-scale data collection, going as far as the birth of the modern State, where collecting data to administrate was a historical defining function of governments (Desrosières, 1998, p. 147; Hand, 2011). Also in modern times, with the advent of digitalization, governments had pioneered several information technology systems through dedicated flagship initiatives (Agar, 2003; Margetts, 1999). It must be reminded, however, that governments operate under conditions of absent market competition. Innovation in the public sector thus often means improving the quality of the citizens-government relationship (typically through services) under a frame of expectations and social acceptance (Bekkers et al., 2011; Borins, 2001; Fuglsang & Pedersen, 2011). New disciplines or techniques centred on data (e.g., data science) or advanced analytics might enhance how the public machine works (see Section 1.2.3.4), but using data in a logic of pure optimization, intrinsic to the private sector, might lead to social and political disasters and waste of public moneys, as already happened in the past (Margetts & Dorobantu, 2019).

Given the central role of government in the public value creation (Moore, 1995), data-driven innovation ought to be consider under other dimensions of public sector innovation, such as the governance innovation emerging from data exchanges (Micheli et al., 2020). This perspective appears only marginally touched, but might become increasingly relevant, especially for the field of data for policy (see Section 4.1). Of all the dimensions and governing functions of the public sector, the process of

13 According to the European Parliamentary Research Service (2019) "the total direct economic value of PSI is expected to increase from a baseline of €52 billion in 2018 for the EU-28, to €194 billion in 2030." (Negreiro, 2019, p. 2).

14 Gray (2015) noted how the disputes over PSI and data regarded, for example: *"[…] questions about who pays for what and who is able to use what, technical debates about standards, licenses and formats, and economic arguments about enabling commercial innovation using information generated by the public sector*" (Gray, 2015, p.5-6).

public policymaking might represent the one whose political dimension will make it resistant to a logic of data-driven optimization (see Section 2.4; 2.5). Having more data or having them faster would not reduce the ambiguity of policy problems that policies are sought to address. The recent COVID-19 pandemic offers a painfully clear example of how purely techno-rational solutions cannot be given for granted in policymaking, as national governments would respond very differently to the same problem for different political judgement (Head, 2022, p. 17). However, rather than more data, policymakers could benefit from insights emerging from appropriate, trustable sources (Verstraete et al., 2021, p. 74). The critical factor in this scenarios seems to remain the government capacity to adopt useful knowledge for policy (Borrás, 2011).

This thesis assumes as rationale the importance of understanding how data can enable innovation by considering not only efficiency and effectiveness as public values (Bekkers et al., 2011, p. 6). It is proposed that data might be the central element of innovative processes of *social learning* in which policy actors address in new ways policy problems thanks to the practices that new data enable. The collection of data not traditionally used for policymaking shall remain at the centre of an innovative governance where a *data ecosystem*[15] acts toward a policy problem and learns about it — in a process of data-centric policymaking.

This perspective will be crafted and argued for in the first chapters, and then brought into the main empirical investigation. The theoretical perspective of data-centric innovation represents the thesis' original contribution to the contemporary discourse on data for policy. Further, it is a perspective explicitly and conveniently used to close the gap between the field of data for policy and design for policy. The discipline of design reached governments with its paradigms, approaches and practical methods, which application in public settings is seen (and yet much discussed) as conducive of innovation both for public sector and policymaking (Ansell & Torfing, 2014; Hermus et al., 2020). Design for policy values might integrate to the use of data in policymaking, as both fields might valuably complement but still appear distant (Leoni, 2020).

The thesis has thus been motivated by the three-fold desire to propose the original perspective of data-centric policymaking, substantiate it with theory and empirical exploration and starting to bridge the gap between data and design in policymaking innovation.

15 A system of actors united by the willingness to collect, share and explore data concerning a policy problem (Oliveira & Lóscio, 2018).

Glossary

Data-centric policymaking	Data-centric policymaking is proposed in the context of this research after the initial Research Setting, as a sensitizing *concept*. As such, it was proposed to guide a purposeful exploration into the field of *data for policy,* while maintaining the specific hypothesis and interests that are peculiar to this doctoral research. Implicit to the notion of data-centric policymaking remains an interest of investigating data for policy as a contextual innovation in which non-traditional data have an impact on existing practices of policymaking and the knowledge of policy workers. This perspective is driven by an interest in understanding the innovation of non-traditional data in policymaking with respect to "design for policy". The concept has been the basis to develop a theoretical-conceptual framing, later employed in the empirical investigation (Chapter 2; Chapter 4).
Data-driven innovation	Data-driven innovation is a term encountered in several official documents used to illustrate the innovation that non-traditional data could bring to governments and advocate for its value. The term is generally used to imply several forms of innovation that, in the public sector, mainly regard three aspects: public governance, public services and policymaking. In respect to policymaking, data-driven innovation is expected to support better short- and long-term decision-making for policy. In essence, non-traditional data, either collected within public sector organizations (e.g., for administrative functions) or by private companies, are seen as a new source of evidence for policy. In this line of argumentation, the achievement of data-driven innovation implies a better use of data analytics in government, which would provide relevant data faster and could drive more granular analysis that are relevant for policy. Governments seem to support data-driven innovation by intervening on the re-use of public sector data and information. This concretely translates in guidelines and frameworks that attempt to mitigate internal technical and organizational barriers hampering data sharing and to regulate data governance (i.e., who is responsible for data across public organizations). Data-driven innovation appears to have developed in close connection with other terms of the data debate and their underlying message, in particular the term Big Data.
Data for Policy	Data for Policy is used in this research to refer to what appears to be an emerging field that attempts to specialize on the innovation brought in policymaking by non-traditional data, data analytics, and other topics of the broader data debate. In the Research Setting several themes were identified that seem to be specific to this discourse. Based on these specificities, the research proposed "data for policy" as a field, and sought to investigate its discourse to understand data-centric policymaking. The field is named after the International Conference "Data for Policy", which is regarded as an explicit example of this broader field, while should not be considered coincident with it.
Non-traditional data	Non-traditonal data is used to refer to data which are not originally collected for being an evidence base for policy. A classification of data in the tradtional/non-tradtional spectrum is provided in the first Chapter.
Data debate	The data debate is used here to identify the broad discussion about digital data and their potential for innovation. The data debate encompasses multiple narratives and perspectives on this relationship. The data for policy field is consided and internal discussion within the data debate.

Structure of the thesis

The thesis is divided into six chapters.

The first Chapter, **“Research Setting”**, introduces and problematizes the thesis’ phenomenological background, critically describing its internal themes and perspectives. It further identifies and reviews its research field, labelled “*data for policy*”. After reviewing the specific discourse in this field, it explicits the research interests that motivate the chosen angle of inquiry. The chapter concludes with the Research Questions.

The second Chapter, **“Theoretical Background and conceptual framework”**, draws from the theoretical literature in policy studies with the intention of conceptualizing policymaking in connection with the use of digital data in the public sector. This review ultimately provides the theoretical-conceptual framework guiding the empirical investigation.

The third Chapter, **“Methodology”**, describes the two-phase research design employed to collect data within the field of investigation. It articulates the broader epistemological stance of the research, in which two different methodological approaches (one for each phase) are accommodated. The related methods employed are then described and substantiated.

The fourth Chapter, **“Results and Discussion”**, describes the data collection and analysis performed in line with what was planned in the Methodology. The findings for each phase are discussed. An overall discussion about the investigation closes the Chapter.

The fifth Chapter, **“Designing in Data-centric Policymaking”** advances three original conceptual areas of convergence connecting the field of “*data for policy*” with the field of “*design for policy*”. These areas are articulated bringing elements of both fields together and with the intention to provide pragmatic recommendations for ground work.

The sixth Chapter, **“Conclusions”**, contains the thesis conclusions, discusses the research limitations and the outlook for future research.

Chap. 1. Research setting

The Research Setting presented in this chapter (Fig. 1) illustrates the thesis' background and its elements. Central to this background stands the peculiarity of contemporary societies' sociotechnical systems and the prominent presence of digital data within them. These aspects appear undoubtable because of the several objective trends illustrated in the Introduction. On the other hand, what data-driven innovation entails seems a rather unsettled question which stands at the centre of an ongoing **data debate** (see Section 1.2).

That debate will be treated as a living element of the research background and research area. Once the internal perspectives of the data debate and their propositions are made clear, the chapter continues by considering what actually seems to be at stake for policymaking in terms of data-driven innovation, according to the authors that have focused on this aspect.

Doing so will offer the opportunity to isolate that specific part of the data debate that concerns policymaking. It will be argued that this specific discourse constitutes an increasingly autonomous field of research and practice, labelled as ***data for policy,*** and taken as field of the research (see Section 1.3). The field will be reviewed to understand what its specific **themes** are (see Section 1.3.3). The **interpretations of these themes** (see Section 1.3.4) will be then read through the **research's interests** (see Section 1.4) to shape the proposal of *data-centric policymaking*, a sensitizing concept used to bundle together the first understanding of the field and the specific interests this research has on it. The concept will be used to develop **hypotheses** (see Section 1.5) upon which the **research questions** (see Section 1.6) and **research goals** are built (see Section 1.7).

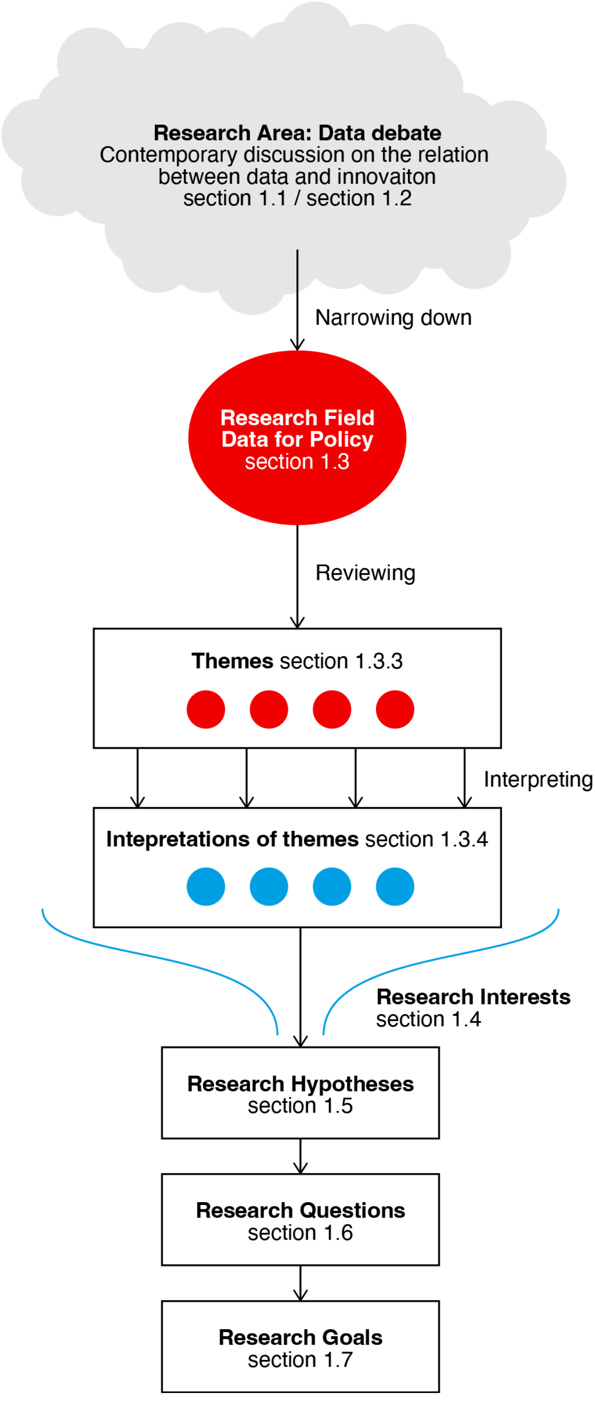


Fig 1. Scheme of Research Setting of Chapter 1

1.1. Research relevance

It is proposed that the research is relevant on the basis of two notions. First, a lively contemporary debate exists on the topic of data and innovation; secondly, this subject also appears high on today's governments' political agendas and priorities.

1.1.1. The centrality of data in the contemporary discussion: the data debate

From the acknowledgment that most contemporary societies are pervaded by sociotechnical systems which continuously exchange and compute digital data, during the last decade a great number of authors started to explicitly proposing that data may potentially be the central element of innovation in several domains of application; and developed research on this topic (Brynjolfsson et al., 2011; Kitchin, 2014b, 2014a; Mcafee & Brynjolfsson, 2012). Among these domains we can mention marketing (Erevelles et al., 2016); management (Mcafee & Brynjolfsson, 2012; Ross & Short, 2012); urban environment analysis and urban planning (Batty, 2013; Bettencourt, 2014; Engin et al., 2020); agent based and complexity modelling (Bansal et al., 2016); general social science research and the humanities (Berry, 2011; Foster et al., 2016). This list could further extend and suggests the existence of an emerging discourse that, in line with Vydra and Klievink (2019), could be called a *data debate,* in which the potential of data to innovate is progressively discussed by its internal perspectives.

Agreement in the data debate exists on the unprecedented nature of the contemporary data exchange and processing and about several enabling factors that make it potentially disruptive in a historical perspective[16]. These factors include *digitisation* and *digitalisation*, both being recognized as objectively trending; and the unparalleled processing capacity reached by modern computers and consumers electronics[17] (Hilbert & Lopez, 2011; CISCO, 2020; Manyika et al., 2011). The *existence and extensiveness* of this *data debate* is here considered the necessary condition for the thesis's relevance. The interest on digital data and on its potential effect on societies appears widely shared (Gandomi & Haider, 2015). Bibliometric analyses, taken as one possible metric of that interest,

16 These factors are themselves dependent by the sociotechnical transformation enabled by ICTs during the last years (see Introduction).

17 A trend recognizable, for example, by the spread of speech recognition software in smartphones or computer vision in cameras network (cf. Mitchell, 2006), both only recently made possible by technological advancements in hardware and software.

indicate that publications relatable to this area[18] have grown considerably during the last ten years (Ekbia et al., 2014; Grover & Kar, 2017; Suominen & Hajikhani, 2021). Also, several scientific venues are now specifically dedicated to this subject[19].

It might be argued, however, that the current existence of a large and lively data debate only represents a *necessary condition* for this thesis' relevance, while a *sufficient* condition would suggest that this topic is not only timely and important now, but it will remain so in the long term. The doubt would be fair, given many have already advanced this debate as biased toward an uncritical and alluring narrative about the potential of data (Boyd & Crawford, 2012; Kallinikos, 2013). Sensibly, it might be asked whether the data debate develops over an important topic or if it is just a scientific fad that momentarily[20] acquired many acolytes. Some asked if we might just be witnessing a change of terminology, with data becoming a new fashionable term and replacing others that populated the discussion on ICTs impact in our societies in the past (Floridi, 2012).

Regardless of the undisputable hype surrounding this topic (Kallinikos, 2013), a discussion on data and innovation might remain relevant in the future because of the growing political interest of governments on this topic.

1.1.2. The relevance of data-driven innovation for governments

As already mentioned, innovation in ICTs and digital data has become a stronghold in governments' agendas worldwide for several years. Data was taken as the centre of ideal and innovative models of government and governance meant to change the relationship between the public sector and citizens[21]. These models were pursued through several rela-

18 These analyses were based on queries through the phrase "big data" that conveys a certain narrative (see 1.2.3), but it is nonetheless an appropriate probe for this debate given its central role.

19 Examples are: Journal of Big Data (https://journalofbigdata.springeropen.com/) and Big Data & Society (https://journals.sagepub.com/home/bds).

20 In the past, unfunded claims related specific technological innovations to broader societal change. Consider, for example, Webster (2014), who noticed how claims of "Information Society" could be seen as overly-enthusiastic accounts emerging as responses to micro-chips innovations (1970s/early 1980s) and the Internet and computer communications (during the mid-90s).

21 For example, the values of transparency, accountability and public participation were articulated as part of the Open Government Model and possible to achieve thanks to several open data initiatives (cf. Gray, 2015, Ubaldi, 2013). Although the main principles of Open Government were discussed in the US long before digitalisation (Yu & Robinson, 2011), the contemporary version of those

tive public initiatives worldwide (Gray, 2015; Ubaldi, 2013). Moreover, public sector information and public data[22] had been increasingly considered as a driver of economic innovation and well-being (Gray, 2015; Janssen & Dumortier, 2003; Rieder, 2018). In line with that, the concept of *data-driven innovation* was proposed as: "*the use of data and analytics to improve or foster new products, processes, organizational methods and markets*" (OECD, 2015, p. 21). In the context of the public sector, data-driven innovation is now expected to enhance several aspects of policymaking, public services and governance, as data are increasingly considered in various ways an asset to improve these aspects (Ubaldi et al., 2019).

The interest of governments seems also testified by the tentative to develop regulations and guidelines dedicated to public data sharing and use (Deloitte & the Lisbon Council, 2021; OECD, 2019). These *data strategies* were recently published by the European Commission (2020); the United States (2019); the United Kingdom (2019); the Netherlands (2019) and Canada (2018), with the aim to enhance current legal and operational frameworks of data sharing and data governance, with the explicit intention of guiding the use of data to improve policymaking and public service provision (Deloitte & The Lisbon Council, 2021).

These policy documents suggest that developing research in connection with the data debate will remain relevant in the long term because the use of data and digital technology in the public sector is currently being extensively featured in public agendas (Ubaldi et al., 2019). However, the use of data for policymaking seems still far from being something fully realized or settled (see Section 1.3.3.2). Therefore, research in this space might be also particularly needed considering the many past flagship government initiatives focused on technological adoption (especially ICTs) that have struggled to find a balance between public administration goals and legitimacy (Margetts & Dorobantu, 2019). Finally, the issue seems even more timely given how COVID-19 pandemic further exposed the potential and challenges of responses based on data to public health issues (Oliver et al., 2020).

principles was articulated in 2007 and then officially adopted in the "Open Government Directive" during the first Obama US Administration in 2009 (Gray, 2015). Similar initiatives proliferated worldwide after the US Open Data Portal opened (Ubaldi, 2013).

22 Public sector information and data are " *[…] obtained or created upon performance of public duties provided by law or legislation issued on the basis thereof*" (UN, 2020, p. 147).

1.2. Research area

The data debate mentioned above discusses the potential innovation represented by and attributed to digital data (Vydra & Klievink, 2019). The presence of this debate demonstrates in the first place that the relation between data and innovation is not to be considered a neutral or uncontroversial phenomenon. The relation between government and ICTs technologies could be reconnected with a well-established and decades-long history (Garson, 2010; Margetts, 1999). Although, this relation only more recently resurfaced with a great emphasis on the role of digital data (Diebold, 2012). Data-driven innovation has been framed and promulgated as the capacity of digital data to create value in innovative ways (OECD, 2015).

However, many authors advance that data-driven innovation has been presented as an unproblematic subject and that the whole discourse around it constitutes a misleading narrative of the potential innovation represented by data. Hence, as part of this research setting, it has been decided to consider the different views of the data debate as a necessary step in the process of clarifying and contextualizing the potential of data for policymaking. In order to clarify how the different views in the data debate stand in respect to the subject of data-driven innovation, it seemed useful to start from the concept of data and its meaning.

1.2.1. The concept of data

Data essentially define entities used to abstract the world into representational forms; as numbers, symbols, images, sounds, bits (Kitchin, 2014) (fig.2). Digital data are, in turn, these representational forms but in a digitalized format, i.e., discrete binary sequences of digits stored on a computer memory. In this section, it has been decided to analyse the concept of data, to first understand the innovation narratives that recently developed around it and only later[23] offer some concrete examples with respect to policymaking (see Section 1.3.2.2).

23 Not least because of the obvious tremendous amount of possible empirical examples of this concept, if not contextualised in respect to something specific.

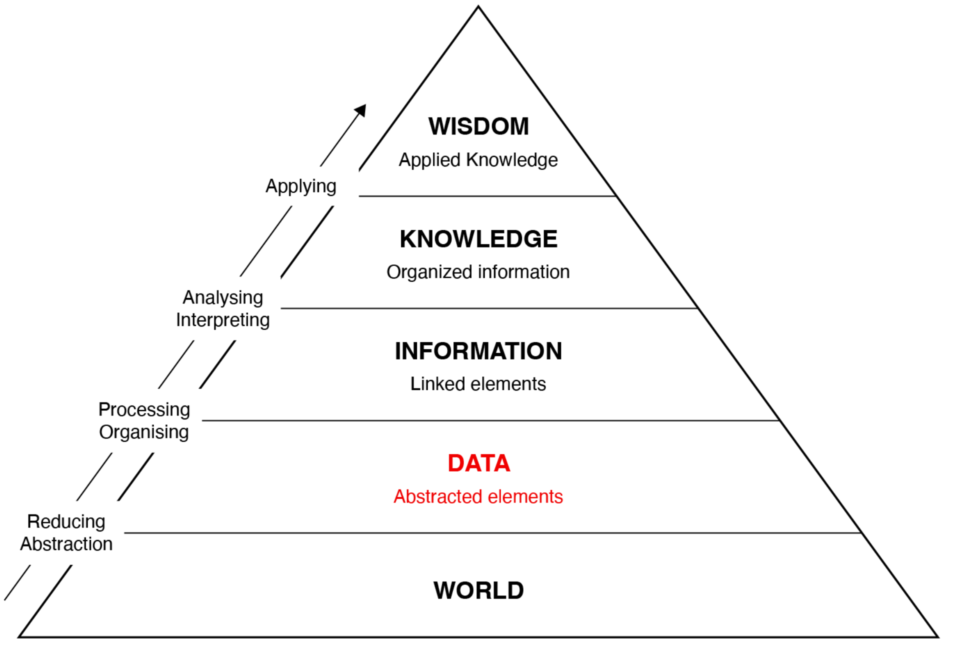


Fig 2. The knowledge pyramid and the traditional role of data within it (adapted from Kitchin, 2014b)

According to the philosopher of information Floridi (2008), the concept of data has several commonly used meanings depending on the interpretative filter adopted (Table 1). Three possible interpretations he proposes are: the *epistemic*, i.e., data regarded as neutral facts or evidence; the *informational,* i.e., data being equated with information (as in "personal data"); the *computational*, i.e., data considered as binary elements or digits. Floridi (2008) further argues for a fourth, more universally applicable interpretation that individuates the ontological essence of data as *the difference between variables.*

Floridi advances that a "*datum*" would exist even when data are virtually absent or zero. He makes the example of a page written with symbols, in which both the symbols on the page and the page itself erased by symbols would constitute a *"datum"*. He then maintains that "data" essentially is not the symbols neither the page, but the variance between symbols and the page, or "*a matter of a relation of difference*" (Floridi, 2008, p. 9).

Table 1. Different interpretation and meanings of data (compiled from Floridi, 2008)

Intepretation	**Data as...**
Epistemic	Collections of facts
Informational	Information about something
Computational	Composition of binary elements
Diaphoric	Differences between uninterpreted variables

This perspective might sound as a conceptual overstretching, but it satisfactorily captures data as a wide concept that could be applied to a vastity of empirical cases:

"The actual format, medium and language in which data (and hence information) are encoded is often irrelevant and hence disregardable. In particular, the same data may be analog or digital, printed on paper or viewed on a screen, in English or in some other language, expressed in words or pictures, quantitative or qualitative." (Floridi, 2008, p. 6)

Given this explanation of the concept of data at the ontological level, it is worth considering data also through its etymology to understand the meaning commonly associated with the word. *Data* derives from the Latin "*datum*", which translates as "*to give*". The term semantic conveys the idea of abstract entities or neutral principles that shape the premises upon which the argumentation could then be built (Rosenberg, 2013). However, the meaning of data as "*given*" elements would had contradictory historical development:

"[...] "data" was especially used to refer either to principles accepted as a basis of argument or to facts gleaned from scripture that were unavailable to questioning. By the end of the century, the term was most commonly used to refer to facts in evidence determined by experiment, experience, or collection. It had become usual to think of data as the result of an investigation rather than its premise. While this semantic inversion did not produce the twentieth-century meaning of data, it did make it possible. Still today we think of data as a premise for argument; however, our principal notion of data as information in numerical form relies on the late eighteenth-century development." (Rosenberg, 2013, p. 32)

The contradiction appears clear if we consider the role of data within scientific progress and methods, thanks to which the concept of data has been associated with the act of measurement. This perspective would make the original meaning of "*datum*" rather counterintuitive (Kitchin, 2014a). In fact, in the scientific tradition, data are not "*given*" entities, but "*captured*" abstraction of the world, (e.g., through observation, experiments), meant to systematize the empirical complexity and to be interpreted for gaining relevant information (Kitchin, 2014a, 2014b).

Thanks to this definition of data, and particularly to the notion of data as "*captured*" entities, it is possible to start understanding the data debate's internal perspectives. In fact, many of them advance that the contemporary ubiquity of data exchange and processing enabled by contemporary sociotechnical systems will offer the possibility of a radical discontinuity with the canonical process of knowledge creation, as proposed by the scientific method (Berry, 2011; Floridi, 2012; Kitchin, 2014a; Wagner-Pacifici et al., 2015).

Taken to the extreme, this argument implies digital data could radically transform science and knowledge. It is a divisive point that generated diverse reactions in the data debate during time.

1.2.2. The mainstream narrative in the data debate: datafication

The traditional scientific method implies that data is "created" as part of experimentations or collected through purposeful inquiries. Following the line of argument exposed above, the *purposefulness* of data collection essentially defines data. In the traditional view, data could be considered as such only if it is collected to become information.

This point could be regarded as the core element that separates perspectives in the contemporary data debate. In fact, in a world of socio-technical systems permeated by ICTs, data is not only collected on purpose to become information, but it is simply created, processed and stored by digital technologies. The innovative scenario that emerges, discussed in the data debate, pertains to the possibility of obtaining knowledge by simply accessing or collecting digital data that already exists as a by-product of interactions in digital infrastructures. Data that once were only handled by scientists and researchers, "*are now being aggregated and made easily accessible to anyone who is curious, regardless of their training.*" (Boyd & Crawford, 2012, p. 664).

A major perspective on this point emerged on the early stages of the data debate, when some authors, through non-scientific publications (Anderson, 2007; Mayer-Schönberger & Cukier, 2013), proposed a radical innovation in science through the concept of *datafication,* that*:*

> "[...] refers to taking information about all things under the sun — including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge— and transforming it into a data format to make it quantified." (Mayer-Schönberger & Cukier, 2013, p.36)

On the premise that our societies are now constantly interacting through digital sociotechnical systems and that a vast quantity of digital data is constantly created as a by-product of these interactions (Xindong, Xingquan, Gong-Qing, & Wei, 2014), the main proposition of datafication thus is that almost any aspect of large-scale social systems could be quantified and represented by the data they produce (Mayer-Schönberger & Cukier, 2013). The epistemological revolution proposed in light of this situation essentially implies that knowledge creation would no longer be dependent on theories, hypotheses, empirical acquaintance with a phenomenon, or careful sampling (Mayer-Schönberger & Cukier, 2013), but would arise directly from data mining/analytics, i.e., the direct algorithmic exploration of large datasets searching for patterns and trends to transform into information and knowledge (Floridi, 2012; Kitchin, 2014; Raghavan, 2014).

The "*end of theory*" proposed by datafication has been extensively criticized (Boyd & Crawford, 2012; Couldry, 2017) and even its proponents admit this notion should not be considered *sensu stricto* (Mayer-Schönberger & Cukier, 2013, p. 151). Datafication appears to have shaped the data debate by large[24] with its own narrative, and its underlying paradigm have found several researchers endorsing it (van Dijck, 2014). Through the spread of the term *big data*, the datafication narrative seems to have dominated the data debate from its early phases and to persist also today within the more recent discussion concerning public sector and policymaking (Suominen & Hajikhani, 2021; Vydra & Klievink, 2019) and in official policy documents that discuss the potential (European Commission, 2020; OECD, 2015).

1.2.3. Datafication and the myth of Big data

The datafication argument seems to have diffused mainly throughout the term *big data* (De Mauro et al., 2015; Philip Chen & Zhang, 2014), which became pervasive in this debate (Ekbia et al., 2014) and remained at the forefront of the datafication narrative, to the point the two became almost synonymous (Sadowski, 2019; van Dijck, 2014). The term *big data* arose abruptly only about a decade ago becoming the focus of a huge transdisciplinary strand of publications since then (De Mauro et al., 2015; Ekbia et al., 2014), while being virtually not existent before that moment (Ekbia et al., 2014; Gandomi & Haider, 2015; Grover & Kar, 2017). Hypothesis on the origin of the term *big data* traces it back to a U.S. software company that during the 90s coined the term (Diebold, 2013). The business intelligence sector used *big data* to indicate data mining performed on data generated by multiple data sources; and is considered to have influenced

24 At the time this thesis is written, the book of Mayer-Schönberger and Cukier (2013) has more than ten thousand references on Google Scholar.

its meaning and diffusion (Gandomi & Haider, 2015). Big data became an encompassing concept that identifies both the potential data availability of contemporary digital sociotechnical systems and the nature of these data sources.

More broadly, the term has been used to describe techniques, practices, and issues connected with the collection of these types of data (Grover et. al, 2016). In line with this, *big data* has been almost unanimously deemed as an unclear concept (De Mauro et al., 2015; Gandomi & Haider, 2015; Grover & Kar, 2017; Ward & Barker, 2013). Big data initial definitions attempted to synthesise the unprecedented characteristics of contemporary data sources. These characteristics[25] include great *volume*, in terms of quantity*; great variety,* in terms of typology of data (e.g., images, videos, unstructured text) and *sources* (e.g., sensors, social media); and *great velocity*, referring to the time of update (Gandomi & Haider, 2015; Laney, 2001). This definition based on features was repeatedly integrated and criticised by showing, for example, that several data sources associated with *big data* will not necessarily feature all the main characteristics suggested by the term (Kitchin & McArdle, 2016). It has also been suggested that the term seems to have been instrumentally used to advance a narrative where the potential of data is depicted within an alluring and compelling narrative, in line with datafication (Couldry, 2017) (cf. Villars et al., 2011). In fact, the datafication narrative repeatedly proposed *big data* as the digital by-product of complex social systems interacting through computers networks, that could be investigated through the *big data* they produce (Bettencourt, 2014; Engin et al., 2020; Kitchin, 2014c; Matheus et al., 2020). The abundance of data and high computational power, in the big data view, form the binomial foundations for an epistemic revolution (Xindong Wu et al., 2014).

The narrative of datafication, through big data, seems to have resisted until present days, despite having caused much criticism (Boyd & Crawford, 2012; Couldry, 2017). That criticism added to the data debate.

1.2.4. Counternarratives in the data debate

Many authors started to call out big data and datafication as biased by techno-enthusiast storytelling (Barocas & Selbst, 2018; Boyd & Crawford, 2012; Dalton & Thatcher, 2014; Gandomi & Haider, 2015; Schroeder, 2014; Ward & Barker, 2013). Some had moved their critiques starting from concerns on the social implications of *big data*. They believed, in fact, that a narrative of overall positive change driven by data was obscuring the

25 These are summarized in the famous three Vs of Big Data, originally presented in a research note of analyst Doug Laney in 2001 for META Group (later part of Gartner) (Diebold, 2013).

potential endangerments of large scale data analysis, in particular how this could affect privacy or bring biased decisions caused by untransparent algorithmic decision-making (Barocas & Selbst, 2018; Zarsky, 2013).

Datafication seemed to assume the total absence of influence by social groups and the neutrality of infrastructures involved in the data collection processes (van Dijck, 2014). The widely cited phrase, credited to British mathematician Clive Humby, "*data is the new oil*", synthesises the underlying metaphor conveyed by the datafication narrative: data is equated to a raw material that needs to be extracted and refined to create value. To its critics, that narrative appears as a self-fulfilling prophecy that presents a potential scenario as a reality while compelling a sense of urgency toward it (Couldry, 2017). The acquisition of large quantities of data is presented by the datafication narrative as a progressive action that entails a positive change for the whole society, while in reality it demands relevant investments which would be unlikely disjointed by pre-existing economic and political interests (Boyd & Crawford, 2012). For these critical authors, through this narrative, the datafication advocates would be advancing, either consciously or unconsciously, an ideology and set of social norms that demands the social life to be turned into data within a logic of surveillance or monetization. For these reasons, datafication was also associated with the concepts of *dataism*, *data capitalism, data extractivism or dataveillance* (Sadowski, 2019; van Dijck, 2014).

These critical accounts coalesced in what might be seen as a unique front that we might call the *data revolution*[26], a counter-narrative of datafication. The data revolution's view does not deny the importance and pervasivity of new sources of data and data analytics, but problematized the view of *datafication* as simplistic and utilitarian. This reframing had been central to the birth of a field called Critical Data Studies (CDS) (Dalton & Thatcher, 2014; Iliadis & Russo, 2016). CDS scholars substantially refused the deterministic technological instances (Wyatt, 2008) of datafication and big data discourse, essentially the reification of technology and data (Schroeder, 2014). They thus criticized the idea of data to be considered as entities independent from the systems that generate them and developed a different narrative to reframe the innovation represented by data, to be then intended as a socio-technical and socio-material assemblage of practices, techniques and technologies (Iliadis & Russo, 2016). Rather than data per se, in the data revolution perspective, the revolutionary element of data would be how actors discursively reconfigure their previous routines of information and evidence building as an effect of new technologies and data sources (Kitchin, 2014a).

26 "Data Revolution" is the title of the book by Rob Kitchin (2014), a relevant figure in Critical Data Studies. The book contains the thesis here associated with datafication and big data critics.

1.3. Research field

Starting from the concept of data, the past section tried to show the different facets of the data debate and its internal perspective without assuming a value-laden perspective. To address the different views of the data debate was deemed necessary to understand the research background. The datafication and big data narrative, in fact, seems to be particularly influencing.

It is suggested by the type of language[27] used in official documents and reports (OECD, 2015; UN, 2020; World Bank, 2017) and specific meta-analysis (Rieder, 2018) suggests that the political debate currently discussing data is oriented toward the datafication narrative. Arguably, this is reflected in the concept of *data-driven innovation*, already mentioned before and often used in these documents. The influence of datafication seems to apply as well to that research branch which more recently started to investigate data-driven innovation in relation to policymaking. In this body of literature, the limits of the term big data and its underlying narrative are often reported (see Section 1.3.3).

This research branch is individuated by this doctoral research as the research field, to be treated distinctly within the data debate, and labelled as "*data for policy*". The proposal is based on noting a series of thematics and issues that seems to pertain to this field (see Section 1.3.5).

1.3.1. Data for policy as an autonomous field in the data debate

The core discussion of the data debate seems to be increasingly translated and applied to policymaking — or said differently, there seems to be a specific discussion on data-driven innovation in the context to policymaking. It is proposed, as the thesis' reading of the research background, that "*data for policy*" constitutes a research area by itself. "Data for policy" seems to be increasingly gathering several epistemic communities interested in exploring, starting from different perspectives, the thematics and issues of "*data for policy*". To advance *data for policy* as a field requires to briefly digress into the concept of field, that I use with reference to the *theory of the social fields* or *field theory*, developed by Pierre Bourdieu.

The notion of field was originally conceived in the '60s as a theory of the practice of cultural production (Boschetti, 2006) and originally intended

27 Consider for example the phrasing used in a European Commission document recently published in support for a European Data Strategy (2020): "Data will reshape the way we produce, consume and live. Benefits will be felt in every single aspect of our lives, ranging from more conscious energy consumption and product, material and food traceability, to healthier lives and better health-care" (European Commission, 2020; p.2).

to highlight the interplay between the individual agency and dynamics of power and positionality within cultural *milieus*. Field theory intended to shed light on how implicit social norms produce regularities in what apparently seem an unregulated social space (Maton, 2008). Today field theory is considered "*a transposable tool capable of explaining the logics specific to each differentiated space of relationships and practices [...] by definition applicable to all areas of sociological research [...]* " (Dubois, 2015, p. 199). Foundation to field theory is the growing specialization of the social world, that can be interpreted as made of relational and autonomous domains of human activity (Hilgers & Mangez, 2015; Salö, 2017; Thomson, 2008). Consequently, field internal composition, activities and roles are structured by self-regulating dynamics based on the sharing of cultural capital between its elites and newcomers[28] (Hilgers & Mangez, 2015).

The notion of field appears useful for this thesis as it recognizes as its background not only concrete realizations of data-driven innovation, but also the broader system of knowledge production that surrounds this subject. The central tenet of field theory, *the struggle for autonomy,* could be then adopted as a guiding principle under the hypothesis that *data for policy* is emerging as separate discourse within the data debate. The proposal of *data for policy* as field also starts from the author's direct involvement with one specific network linked to a scientific conference called "*Data for Policy International Conference*"[29]. This network cannot be held as fully representative of the contemporary discourse investigating data-driven innovation and policy, but its existence suggests that this area is an increasingly emerging and explicit field for research and practice.

28 The field perspective has been used, for example, to describe the contending of knowledge domains among disciplines, and how these are shaped after scholars struggling for assets and reputation in the academic world and outside (Salö, 2017). As a disciplinary community becomes hegemonic within a field, it establishes the "right" set of practice and protocol for knowledge (i.e., what is right/useful to know) (Greckhamer et al., 2008).

29 The conference has been managed by "Data for Policy CIC", an independent (not-for-profit) Community Interest Company based in London and reached its sixth edition in 2021 (cf. https://dataforpolicy.org). It was originally held in 2015 at the University of Cambridge and included a network of U.K. stakeholders from university and government interested in the potential of data science for policymaking (Meyer et al., 2017).

1.3.2. Key clarifications

1.3.2.1. "Policy" in data for policy

Before moving forward into the field of data for policy, it seems necessary to anticipate here a first definition of policy. Policy studies — arguably the disciplinary field that investigated this topic for the longest time — clearly indicate how difficult and crucial it is to define policy (H. K. Colebatch, 2005). A large part of this research has been dedicated to face this theoretical/conceptual task and this will be shown extensively in Chapter 2. For the scope of presenting the field of data for policy in this Research Setting, the definition provided here will be the one more widely accepted and associated with the policy process model called policy cycle. In fact, this model appears to be the one most used by authors in the field of data for policy (Höchtl et al., 2016; Poel et al., 2018; Tsoukias et al., 2013).

The definition of public policy is still debated and occupies a central place in many handbooks (Howlett & Cashore, 2020; Howlett & Giest, 2013; Jann & Wegrich, 2007). The notion of "policy" already represents an analytical lens to interpret the complex activities of government (H. K. Colebatch, 2005). In this sense, policies can include whatever government chooses to do or not (Dye, 2013). This broad vision suggests that public policies are in first place defined as an account of authority, which is the main element characterising governments over the other social actors (Pierre & Peters, 2020). Accordingly, in representative democracies, governments possess the authority to act within the law for making policies — which regards investing public resources, issuing and enforcing regulations and building programs that favour certain classes of social actors in place of others (Lowi, 1964). In this sense, policies are politico-administrative acts through which government mobilizes available resources toward issues perceived as relevant (Howlett & Cashore, 2020; Howlett & Giest, 2013).

Holding this vision, policies can be interpreted theoretically as the outputs of political systems and institutions (Easton, 1957; Knill & Tosun, 2012). Although, the abstract construct of policy also bundles together elements pertaining to earlier stages (Page, 2006) (see Fig. 3). These stages usually regard the dynamics of political discussion about the policy ends/aims, including the goals and principles and the operational plans to attain these goals. The *policy instrument*, i.e., the mean used by government to achieve its goals[30], while in common sense tends to be equated with policy as a whole, is but one element used to affect the other

30 The concept of policy tool in itself has been used to conceptualize and categorize the different possible programs and administrative tools governments use to achieve their goals (Hood & Margetts, 2007; Salamon, 2002).

social actors in the attainment of policy goals. These instruments might consist of a mix of measures and rules that affect public organizations structures, the redistribution of public money, existing regulations and public enforcement rules and governmental communication (Cairney, 2019; Howlett & Cashore, 2020).

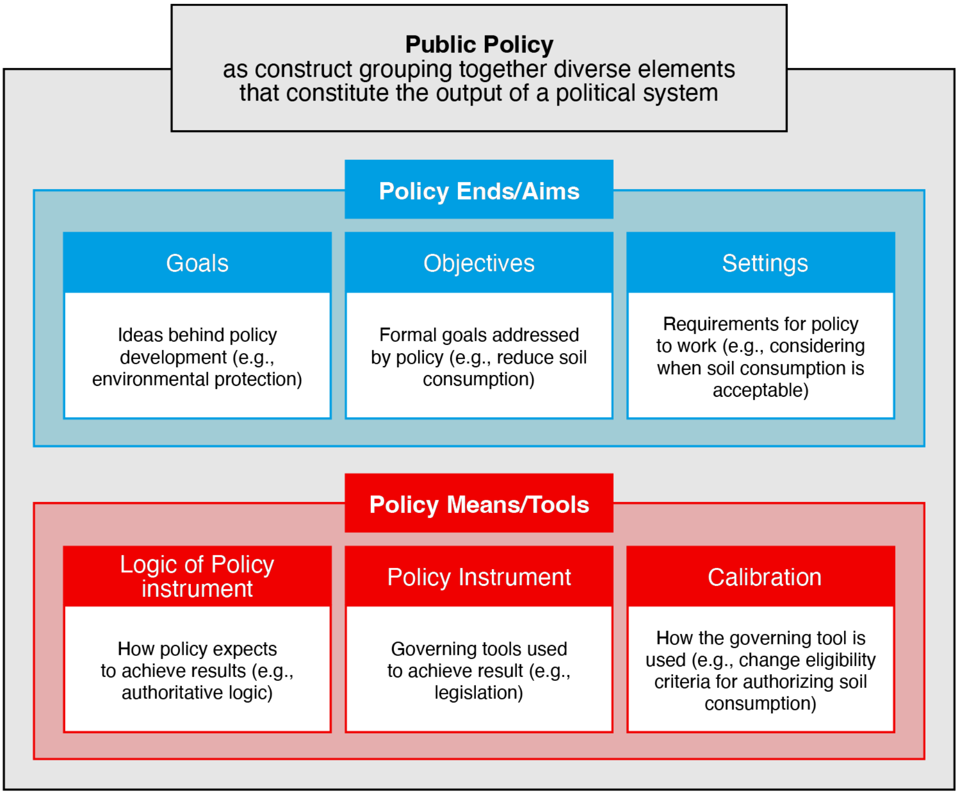


Fig 3. The components of policy as abstract construct (adapted from Cashore & Howlett, 2007).

By maintaining this view, several models and theoretical frameworks have been proposed to analyse the process of policymaking (Weible & Sabatier, 2018). Among them the policy cycle might be regarded as one of most widely used. The policy cycle conceives public policymaking as a problem-solving activity (Jann & Wegrich, 2007). Accordingly, it conceptualizes policymaking as a process in which government addresses public policy problems throughout discrete stages; firstly by recognizing the problem and then acting upon it (Fig. 4).

These stages — in the modern version[31] — are usually indicated as five, each contemplating a different behaviour and role of social actors involved in the policy life cycle (i.e., the policy actors) (Howlett & Giest, 2013).

- *Agenda Setting*; where different societal actors attempt to influence governments' agenda.
- *Policy Formulation*; where government experts and a limited group of societal actors influence policy problems prioritization (i.e., government bureaucracies, interest groups, legislative committee rooms, special commissions, think tanks).
- *Decision-making*; where authoritative government decision-makers (typically experts) take a decision on the right course of action within a policy option.
- *Implementation*; where public administration officials and street-level bureaucrats implement and adjust the measures.
- *Evaluation*; where a range of policy actors evaluates policies/programs and their outputs/outcomes.

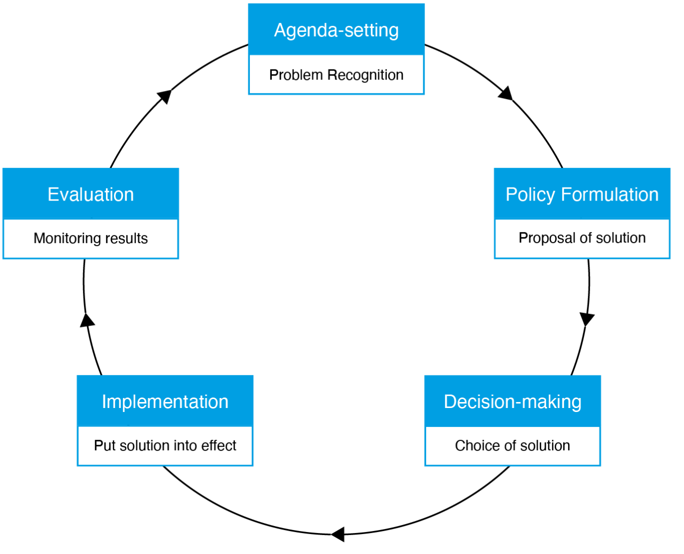


Fig. 4. The policy cycle model that presents the policymaking process as a staged model of public problem solving (adapted from Howlett & Cashore, 2020).

31 The policy cycle here presented a reworking of "*seven sequentially ordered decisional functions*" of governments, originally proposed in 1971 by Harold Lasswell, the father of policy sciences (Dunn, 2017, pp. 43–44). Other reworkings with different numbers of stages exist.

As Chapter 2 will further expose, the policy cycle model is often meant as a device for heuristics, rather than a realistic depiction of the process of policy (Bridgman & Davis, 2003; Wegrich & Jann, 2006). Because of its capacity of simplifying, the policy cycle has been widely employed by scholars and public sector practitioners alike; and, arguably, it might be regarded as the most well-known model of policymaking (H. K. Colebatch, 2005).

1.3.2.2. *"Data"* in data for policy

In light of the first definition of policy proposed, we can proceed to clarify what is meant by data. A classification of data sources potentially employable for policymaking is proposed (see Table 2). The classification is based on reviewing other classifications and data sources mentioned in literature (Connelly et al., 2016; Durrant et al., 2018; Kitchin, 2014; Kitchin & McArdle, 2016; Poel et al., 2018; UN, 2019, 2020). Also, it integrates knowledge of the type of data used in cases found through desk research (see Section 4.2.1). This list should not be considered as exhaustive[32], as building such a comprehensive list was not the scope of this thesis. Instead, the classification aims to provide an overview that integrates the conceptual definition previously provided with concrete examples (see Section 1.2.1)[33].

The classification explores the notion of *non-traditional data* for policymaking (Weber et al., 2021). A central element of data-driven innovation for policymaking seems to be using data not originally collected as a base for evidence across various sources (see Section 1.3.3.3). Therefore, the main rationale of this classification is to classify data sources while sorting them between established sources of evidence for policymaking and those that traditionally are not (Connelly et al., 2016; MacFeely, 2018). Moreover, since the ownership of data also represents a relevant criteria for data-driven policymaking (see Section 1.3.2.3) the data sources are divided considering if they are collected within or outside government (this aligns with other data classifications; cf. UN, 2019). The groups of data emerging from this classification are discussed below and offer a useful exploration for the thesis' research setting.

32 To build such a list might prove highly challenging since not even a representative sample of uses of data for policymaking seem to exist yet (Poel et al., 2018).

33 The list also shows the variety of data sources potentially available today in the contemporary landscape of sociotechnical systems, partly explaining why this landscape has been considered unprecedented in a data perspective (see Introduction).

Government data

Sometimes also referred to as public or public sector data[34], the group encompasses all data " […] *obtained or created upon performance of public duties provided by law or legislation issued on the basis thereof*" (UN, 2020, p. 147). These are data intended for the government to be used as a base of evidence for public decisions. They include data and statistics produced by dedicated public institutions, as well as surveys or statistical registers (UN, 2019). These data possess high coverage and representativeness (e.g., the national scale) and are collected with systematic procedures and standards for being used as an evidence base (Connelly et al., 2016). The census data offer probably the most clear-cutting example of data collected to be used as evidence for policy decisions and upon which several other statistics are derived (Baffour, 2013)[35]. Governments might also rely on statistical services provided by supra-national bodies, like OECD[36], ESPON[37] or EUROSTAT[38].

The data collected by public agencies or public-owned research bodies should also be included in government data. For the most part, these are geospatial environmental data either collected by the local governments for urban planning and territorial administration (e.g., park and areas, cultural heritage sites) or by state-level environmental agencies (Crompvoets et al., 2018). They might be regarded as traditional since their data collection procedure features a long timeframe and quality standards, and remains within the boundaries of the public sector.

In Italy, for example, environmental monitoring performed at city and region levels allows the transmission of environmental data (e.g., air pollution) to other national-level subjects on a daily basis[39]. Whenever government data are publicly released in re-usable formats and web-portals, they might be labelled as Open Data or Open Government Data (UN, 2019).

34 This labelling is adopted for example by the Public Data Policy Statement of Australian Government: https://www.pmc.gov.au/public-data

35 Census data are traditionally and still largely based on the universal enumeration of population every ten years, and provide an essential baseline for a number of national and supra-national statistics (Baffour, 2013).

36 Cf. https://stats.oecd.org

37 Cf. https://database.espon.eu

38 Cf. https://ec.europa.eu/eurostat

39 Cf. Istituto Superiore per la Protezione e la Ricerca Ambientale at: www.isprambiente.gov.it/it/banche-dati

Administrative data

Administrative data (sometimes also called service or micro-data) (Crato & Paruolo, 2019) are "*data sets created primarily for administrative purposes by government agencies or other entities working on behalf of the government*" (UN, 2019, p. 58). Administrative data record the public agencies performances and the interactions of citizens with public administrative systems or services (e.g., education, healthcare, taxation, housing, or vehicle licensing) (Connelly et al., 2016). They are generally used to compile national registries (e.g., house price register) therefore they are closely related to national statistics (Kitchin, 2021). These are sets of data with noticeable volume, regular and frequent update periods (sometimes even daily) and high granularity (to the level of individual citizens or households) (Crato & Paruolo, 2019; Kitchin, 2021).

Administrative data have not been traditionally used as an evidence source for policymaking, but are increasingly adopted for this purpose (Baffour, 2013) (more below). It should be noticed that administrative data use is often considered dependent from successful data linkage and increasingly supported. For example, in the United Kingdom, several public initiatives work to ensure linkage and accessibility of administrative datasets for research purposes[40].

Data exhaust and crowdsourced data

This group encompasses data sources that could be labelled as totally *non-traditional data*. These basically are data which potentially describe a relevant phenomenon for policy but are not intended by default to be transformed into information or evidence, and therefore are collected without the traditional data quality standards. These are mostly generated or collected outside the public sector, in particular by private services providers. Here this group is divided into data exhaust and crowdsourced data.

Data exhaust includes the passive recordings of users' interactions in digital networks and infrastructures (Kitchin, 2014). In the datafication perspective, this group epitomizes the concept of *big data*. For example, data stored on cellular networks database as users interact with them while logging to Internet or telecommunication networks (e.g., log of personal devices as they connect to routers) or recorded by scripts as users interact with websites (e.g., clicks, mouseover, scrolling time, text, images and files uploads). Sensors, cameras, satellite and laser data[41] could also fall under

40 Notable examples are: the Administrative Data Research UK (www.adruk.org) and the OpenSAFELY platform (www.opensafely.org).

41 As said, some territorial/public institutions might collect these data and then

this category (e.g., sound data, pollution data, thermal, light and radiation scans) (Batty, 2013, 2017; Kitchin, 2014c).

Instead of being only passively recorded, non-traditional data can also be actively generated, as in the case of *crowdsourced data* (also called citizens-generated data) (Ponti, 2020), which are data collected through the active involvement of citizens by mobile devices or sensor as part of grass-root mapping, sensing, or citizens science initiatives (Milan & Velden, 2016).

Table 2. Classification of data types potentially available for policymaking according to traditional/non-traditional classification and ownership/origin

Group	Sorting criteria		Types (examples)
	Traditional or non-traditional	**Ownership and origin**	
	With respect to policymaking	With respect to government	
Government data	**Traditional** Collected for policy and public functions and originally intended as form of evidence for policy	**Internal** Collected by government agencies or on behalf of	Survey data, Census
			Geo-spatial, building, infrastructure and environmental data (e.g., digital cadastre)
Administrative data	**Traditional and Non-Traditional** Created primarily for administrative purposes and not originally intended as form of evidence for policy but increasingly used in this sense	**Internal** Government agencies or other entities working on behalf of the government	Primary and secondary health care data (e.g., in-patient admissions, out-patient appointments, comorbidities, clinical test results)
			Social care data (e.g., social benefit claims, attendance at community care services, disability status, employment status, etc)
			Human resource management (e.g., age of public employees)
			Education data (e.g., school attendance, exams score, etc.)
			Law enforcement processes (e.g., taxation data, inspections, police reporting data)

they might be used as evidence for policy decisions. Outside these cases, and particularly in the private sector, these data are originally collected for routinary recording functions, so they would be non-traditional in a policymaking perspective.

Data exhaust and crowdsourced data	**Non-traditional** Created as by-product of active or passive interaction on a digital infrastructure and not originally intended as form of evidence for policy	**External** Everyone else outside the government sector (e.g., private companies, industry, citizens)	Internet and Telecommunication (Call Detail Records, Network data, client data, connections to hotspots)
			Cameras (e.g., CCTV, satellite images, LIDAR, Automatic Number Plate Readers)
			Web (e.g., clickstream, data on clients requests, Web Page Text, Search engines queries, Images)
			Sensors (e.g., smart meters, electricity grid systems)
			Logistics and supply-chain data (e.g., scanning of products)
			Transaction data (e.g., credit card transactions)
			Crowdsourced (e.g., citizen sensing campaigns, etc)

1.3.2.3. The innovation of non-traditional data for policy

The proposed classification tries to make a point for considering the non-traditional dimension of data as a relevant and specific aspect to the data for policy debate (see also Section 1.3.3.3). The data needed by a public authority for issuing decisions must have certain features, which normally are those to be found in government data. This means that data which are regarded as non-traditional could be considered and become increasingly relevant for policymaking, for example by being integrated with more traditional data sources (Durrant et al, 2018).

Administrative data, for example, while not being traditionally collected to inform policy, are increasingly used to integrate official statistics[42] (MacFeely, 2018). Many authors suggest that administrative data are extremely promising for policymaking (Connelly et al., 2016; Crato & Paruolo, 2019; Harron et al., 2017) and are increasingly becoming one of the most widely used source in public sector (Klievink et al., 2017; Malomo & Sena, 2017; Poel et al., 2018). In fact administrative data can provide an accurate and updated picture of citizens' actual activities at the individual level (e.g., the use of public services) (Hand, 2020).

Non-traditional data might offer several advantages. In most cases they can be collected quicker and with a more granular aggregation level

42 In particular, the use of administrative data to integrate or substitute traditional statistics have since years been explored by those countries who implemented a citizen code (i.e., a unique identification number to map single individuals across all administrative services) (Baffour et al., 2013).

(e.g., the single individual) than traditional data. One concrete example is mobile phones data for mobility policymaking. In a traditional data collection process, urban planners would statistically infer mobility patterns on a territorial scale through a mathematical model called origin-destination matrix, based on aggregated data collected through travel surveying that registers transport habits of a population sample (Pucci, 2013). These traditional data for urban policymaking are normally updated every five to ten years.

Cellular networks, instead, might record mobile phones' trace data on a daily basis (e.g., position) from hundreds of thousands devices every time they connect to the network, and can be used to understand mobility patterns (Calabrese et al., 2013).

Despite their potential, administrative (and non-traditional data in general) have many pitfalls exactly because they are not collected to inform policy decisions *by design*, nor have they been extensively used for that purpose. Three main issues of non-traditional data can be highlighted. First, representativeness cannot be guaranteed (Giest & Samuels, 2020). Following the example above, while a travel survey would ensure that the population sample is representative in terms of demographics (e.g., age, gender, occupation), mobile phones trace data could describe only those users interacting on that specific provider network with their personal devices, with obvious problems in terms of digital divide. This connects to a second problem, which is privacy (Athey, 2017). If we consider, again, mobile data, these can reveal much sensible information on networks' users. Thirdly, non-traditional data are often not owned or collected by the public sector. Again, following our example, mobile data are collected by the private companies who provide the service and own the infrastructure. For these reasons, public-private agreements and collaborations for data sharing started to become a relevant topic (Susha et al., 2017).

1.3.2.4. Data ecosystems for policymaking

From the existing attempts of listing technologies for collecting and analyzing non-traditional data for public sector and policymaking (Kamateri et al., 2015), it can be derived that these seem not to imply the adoption of a single technological product (e.g., softwares) but a wide range of non-technological organizational and process arrangement, intended to integrate existing information systems (Barbero et al., 2016; Deloitte & The Lisbon Council, 2020). In fact, given the variety of potentially available non-traditional data sources, as shown in the previous section, the challenge for leveraging on them seems primarily to regard data access and integration. Connelly et al. (2016) indicate how administrative data sources could be treated as they were traditional data for social research in the analysis phase. However, they entail multiple challenges in terms of accessibility, quality and management, since they are produced "outside" the traditional method of data collection (Connelly et al., 2016).

Since some years, public administrations have faced this challenge — possibly under the influence of the open government paradigm — by bringing together different valuable data sources (e.g., spatial data) into ICT infrastructures, often publicly accessible (Crompvoets et al., 2018). More recently, two main logics of data management (Giest, 2017a), that might be connected to two different logics of information systems, are emerging to address the issue of accessibility/interoperability. On the one hand, specific public organizations are following a *data warehouse* logic (Kuonen, 2004), i.e., developing systems to connect multiple data points from multiple data collection sources within a unique architecture for advanced analytics (e.g., predictive analytics) (Athey, 2017; Deloitte & The Lisbon Council, 2021). A clear and recent example of that approach could be found in the application of analytics to detection of non-compliance or frauds to welfare state schemes — what is increasingly referred to as fraud analytics (Atto et al., 2015). In this approach, data are made accessible into one infrastructure.

On the other hand, governments are facing the accessibility/integration challenge by focusing on the enhancement of internal capabilities, what might be labelled the *data mining* approach (Kuonen, 2004), which implies ad hoc integration and analysis of several disconnected data sources to search for patterns and insights. This is reflected into the establishment of specific groups dedicated to data and evidence (i.e., data labs) (Ubaldi et al., 2019) and most notably from the hiring of statutory employees assigned with the duty of data management (e.g., Chief Data Officers) (Desouza & Jacob, 2017) or data science (van der Voort et al., 2019).

Data scientists, in particular, are professional with a multidisciplinary background that mixes computing and statistical techniques exactly to perform data mining, i.e., the gathering, processing and identification of information patterns within a variety of data sources (Crisan et al., 2021). The comparison of the two approaches might highlight two different logics of information systems design, from systems to services (Dahlbom, 2002). The data warehousing approach attempts to engineer a system of data utilization, which can then provide insights for a specific set of goals within a given administrative area. In this approach, the chain of data value-creation and its actors are known; and the data warehouse system will optimize the former to serve the latters

Conversely, the data mining approach does not rely on infrastructures but deploys ad hoc information intelligence services, starting from specific questions, but through an explorative approach that might include multiple actors. The data value-creation chain could become complex, as several actors get involved: data scientists, data manager, data providers, domain experts, decision-makers or political representatives. To describe this approach from creating value from non-traditional data, the concept of data ecosystem has been advanced (Lanza, 2021; Linåker & Runeson, 2021; Oliveira & Lóscio, 2018; Van Loenen et al., 2021).

A data ecosystem defines a self-organized group of individuals[43] from several organizations connected by a common intention or need, and by a series of data-related activities, e.g., data collecting, processing and analyzing (Oliveira & Lóscio, 2018). Most essentially, a data ecosystem works as an abstract category that identifies a networked multi-actor process of value-creation from data in respect to a linear value-creation chain (Oliveira & Lóscio, 2018). In the public sector, it is suggested that the degree of formalization of a data ecosystem might vary depending, for example, on the involvement being due to official data-sharing initiatives, the use of a technological platform or a set of technical standards (Linåker & Runeson, 2021). Regardless of their technological or organizational means, it seems clear that data ecosystems emerge from seeking value from data, which, for the public sector perspective, means deriving public value from data (Oliveira & Lóscio, 2018). Data ecosystems are driven by a demand for data which starts from a problem and imagines a purpose and outcome of the data used (Van Den Homberg & Susha, 2018). Data ecosystems (or data-related ecosystems) have also been identified in relation to policymaking, as groups of actors in the public sector engage with the experimental use of non-traditional data as part of complex public decision-making processes (Lanza, 2021). Preliminary research on this topic suggests that the type of network enabled by data ecosystems (which actors and how many), the nature of policy problem it addresses and the institutional contexts in which it unfolds are all elements influencing how a data ecosystem might affect the institutional level of public decision-making processes (Lanza, 2021).

1.3.3. Reviewing the research field

The following sections report a literature review that seeks to summarize and report the main elements of this discussion (Cooper, 1988, pp. 107–112). The review here presented did not result from a systematic search of the literature database through queries[44]. The topic in fact appears still too new and the relevant publications are scattered across several publi-

43 These individuals can at the same time belong to several data ecosystems (Oliveira & Lóscio, 2018).

44 The choice was also considered sub-optimal since prompting literature databases with general queries containing “data” and “policymaking” would result in thousands of elements not really connected, that would need to be analyzed individually. Moreover, for the systematic literature review method to provide fruitful results, the keywords used should address terms with a consolidated meaning. This seems not to be the case for the data debate, given how it is riddled with fuzzy terms such as “big data” (see Section 1.2).

cations[45]. The review aims to synthesise the main themes of the data for policy field, as expressed by several articles individuated during the three years of research and focused on data-driven innovation in policymaking.

1.3.3.1. Overview of the main stances and themes in the data for policy debate

The exact boundaries of the contemporary discussion on the use of data for policymaking are not easy to define (Suominen & Hajikhani, 2021). Data-driven innovation in relation to public sector and policymaking appears to have only recently coalesced into a dedicated, yet fragmented, field (Mureddu et al., 2020; Suominen & Hajikhani, 2021), with growing political support (mentioned above), reports (Ubaldi et al., 2019), scientific venues and European research projects[46]. The innovative use of data for policymaking and in the public sector seems at its infancy with only few examples going beyond experimentations (Arnaboldi & Azzone, 2020; Durrant et al., 2018; Giest, 2017; Klievink et al., 2017; Poel et al., 2018; Verhulst et al., 2019). The empirical evidence of data-driven application or the use of non-traditional data for policy appears scarce (Poel et al., 2018; Verhulst et al., 2019). Nonetheless, the topic is expected to remain relevant both for the practice of policymaking in government (Giest, 2017; Mergel, 2016) and for policy research (El-Taliawi et al., 2021). Perhaps due to this unsettled landscape, this debate seems currently polarised between the overly optimistic views and bluntly realistic (almost pessimistic) ones (Vydra & Klievink, 2019).

We can group authors between *prospective views*, who describe the potential of using non-traditional data in relation to broad areas of policy intervention and conceptual models of policy action (Dunleavy, 2016; Hagen et al., 2019; Maciejewski, 2017) and *contextual views*, seeking to investigate the same subject but relating it to specific public institution settings, policy areas or tools (Giest & Mukherjee, 2018; Malomo & Sena, 2015).

Prospective views usually advance broad overviews on the innovative impact of data for policymaking by referring to technological applications within potential use scenarios. For example, the use of sentiment analysis for tracking how politically relevant topics gets discussed online, thus informing agenda setting (Kamateri et al., 2015); or the use of data-driven

45 For example, the journal "*Data & Policy*" published by authoritative Cambridge University Press and arguably the one most focused publication on the topic today, only opened in 2019 and, at the moment of writing this thesis, is not yet indexed on Scopus or Scholar.

46 For example: Data4Policy (2014-2016) (www.data4policy.eu); Big Policy Canvas (2017-2019) (www.bigpolicycanvas.eu); PoliVisu (2017-2020) (www.polivisu.eu); BigProd (2019-2022) (www.bigprod.eu).

simulations and data visualisation for computer-generated scenarios to aid policy decision-making (i.e., policy modelling) (Mureddu et al., 2014); or the real-time monitoring of public management processes and public services (Maciejewski, 2017). While generally remaining on a broad level, these authors provide different degrees of empirical evidence to support their points: purely theoretical (Höchtl et al., 2016); showcasing illustrative examples (Maciejewski, 2017) or desk research cases presented as part of model/frameworks of policy action (Dunleavy, 2016; Studinka & Guenduez, 2018), or a small-N case studies sample (Athey, 2017; Azzone, 2018).

Conversely, the *contextual views* focus on specific people and institutional elements in which the innovation of data could unfold. The contextuality can be given by their focusing on a certain level of government/governance, e.g., local authorities (Durrant et al., 2018; Lanza, 2021; Malomo & Sena, 2015); one particular policy area, e.g., energy or education (Giest & Mukherjee, 2018; Williamson, 2016); or a specific policy tool (Giest et al., 2021). Data science in policymaking is also considered in a contextual view (Arnaboldi & Azzone, 2020; van der Voort et al., 2019).

The "*prospective vs. contextual*" lens might help to identify two different stances in the debate on data and policymaking. While a clear demarcation line remains hard to draw, this reading might provide a graspable way to sort differences within the variety of perspectives (Suominen & Hajikhani, 2021). Conversely, shared topics of interest, considered important, seem more easily identifiable. These include:

- *Data governance*, the formal collaboration that regulates data sharing among public or public-private organisations (Susha et al., 2017)
- *Data Culture*, the ensemble of individual, organizational and institutional capacities that allows public organizations to collect, merge and utilize non-traditional data (Giest, 2017b, p. 368).
- *Data science and analytics,* the processes, techniques and softwares applied to data for gaining policy-relevant insights (Arnaboldi & Azzone, 2020)
- *Data quality*, the property of data of being not faulted from a technical and representational point of view (Giest & Samuels, 2020)
- *Privacy*, the ensuring of sensitive information about individuals or households (Zarsky, 2013). For example, administrative data often contain sensible information on citizens' use of public services (e.g., health, social welfare data) which can be further inferred when diverse administrative datasets are linked (Harron et al., 2017).
- *Data ethics*, the ethical ways of using data, making sure they are equally representative of all social groups (Hasselbalch, 2019).

1.3.3.2. Theme (A): The importance of data quality

Within the data debate, the narrative of datafication and big data seems to be prevalently advocating for large scale data collection and the acquisition of great volume of data. In the data for policy perspective these aspects do not appear to be the most innovative factors (Giest, 2017; Klievink et al., 2017; Vydra & Klievink, 2019). In fact, to deal with great volumes of data hardly represents an innovation for governments[47] (Connelly et al., 2016; Giest, 2017). The relevant innovation element instead seems to be the *use of data not originally created as evidence for policy decisions* (Connelly et al., 2016; Durrant et al., 2018; Giest, 2017; Klievink et al., 2017; MacFeely, 2018). Rather than the volume of data acquired, data for policy could be collected from the many heterogeneous *non-traditional* data sources available (Weber et al., 2021) (see also Section 1.3.2.3).

The label *non-traditional* is here stressed to suggest that *data quality* might be the most relevant dimension in data for policy. Data (and information) quality are defined by several aspects among which the system of collection (i.e, who and how is the data collected) and the *fitness of use* (i.e, the scope of data collected in respect to an intended use) (Floridi, 2014) seem of particular relevance for policymaking. As explained (see Section 1.3.2.3), non-traditional data represent both a relevant opportunity for policymaking, but also a challenge because of many factors ultimately attributable to data quality. Traditional data sources used in policymaking are in fact characterized by high quality and reliability standards. Data for policymaking is considered high-quality when it is representative, impartial, collected with methodological rigor and made publicly available over a long period of time (MacFeely, 2018). Policy decisions usually rely on evidence sources collected and published as statistics, aggregates, or indicators by national or international statistical offices. These public or publicly-owned bodies are held accountable for the data quality. Data used for policy decisions must also be trustable, an aspect dependent by the *data backrun,* i.e., how long a given data source has been used in connection to one or a set of specific indicators (Vydra & Klievink, 2019).

1.3.3.3. Theme (B): A lack of empirical cases

On June 2019, the journal "Data & Policy" opened as a new scientific venue under the publisher Cambridge University Press "*to promote a new theory of policy-data interactions by publishing work that considers*

47 As said in the introduction, governments traditionally collected large quantities of data (e.g., census) to support governing functions (Hand, 2011). Policy analysis and evaluation also have a long standing role in using quantitative methodologies and large datasets to advise policymaking (Dunn, 2017; Mintrom & Williams, 2013).

systems of policy and data and how they relate to each other"[48]. In that publication, authors highlighted that:

> "Policy–data interactions or governance initiatives that use data have been the exception rather than the norm, isolated prototypes and trials rather than an indication of real, systemic change. [...] we believe that there should be a sound under-pinning a new theory of what we call Policy–Data Interactions [...] we need a consistent, symmetric approach to consideration of systems of policy and data, how they interact with one another" (Verhulst et al., 2019b, pp. 1–2)

Several authors seem to be agreeing on the fact that the use of data for policymaking appears at an early stage, with only few examples going beyond the experimental stage (Arnaboldi & Azzone, 2020; Durrant et al., 2018; Giest, 2017; Klievink et al., 2017; Malomo & Sena, 2017; Poel et al., 2018; Suominen & Hajikhani, 2021; Verhulst et al., 2019a). Many impute this state of things to many barriers within public organizations. These regard both technological infrastructures and internal capacities and competences. Existent information technology infrastructures in public administration are traditionally of siloed-type (Niederman et al., 2016; Yildiz, 2007), as they were designed to optimize internal decision-making within the single departments (Giest, 2017). The lack of technical standard and infrastructures is a factor hampering data interoperability, accessibility and sharing[49]. Moreover, there is also a general lack of internal competencies and expertise for working with data and in particularly non-traditional sources (Desouza & Jacob, 2017; Giest, 2018)[50]. Another obstacle rests in the capacity of public organizations to develop intra-organizational collaborations for data sharing and to align the value of exploiting data with their statutory task (i.e., developing a data culture) (Klievink et al.,

48 The editors-in-chief who authored the launching editorial (Verhulst et al., 2019b) were coming from a multi-annual experience of "The Data for Policy Conference". In the editorial, authors expressed the need of establishing a permanent venue to host and systematize a cross-disciplinary scientific debate on the use of data and data science in government. Cf.: https://www.cambridge.org/core/journals/data-and-policy

49 For example, these can include common interoperability standards, base registries, shared ICT infrastructure and services, common data architecture/infrastructure (OECD, 2019).

50 There seems to be a trend of hiring data scientists as in-house staff and to appoint Chief Data Officers (i.e., public managers with a mandate of overviewing the governance and sharing of government data) (Desouza & Jacobs, 2017), the vast majority of these competences are still outsourced to private sector (Giest, 2017; 2018).

2017). Finally, at the broader institutional level, a well-known barrier is the absence of legal and policy frameworks to share data among public institutions[51]. The OECD reports only 12% of the 29 member countries possess "*a single dedicated data policy (or strategy)"* (OECD, 2019a, p. 31).

Regardless of these barriers noted by authors, the use of data or data analytics in government and public sector seems not totally absent. Some public agencies developed systems of analytics for public service provision, law enforcement or fraud detection (Athey, 2017; Deloitte & The Lisbon Council, 2021; Dunleavy, 2016). For example, The HM Revenue and Customs (HMRC) of United Kingdom Government, uses administrative data (e..g, historical data on compliance, income, age, occupation) and predictive analytics models to anticipate the probability of certain categories of taxpayers being non-compliant or committing fraud (Atto et al., 2015).

1.3.3.4. Theme (C): The limits of current theoretical frameworks

A general agreement seems to exist on the current limited theorisation of the relation between data and policymaking (Verhulst et al., 2019). Most studies on data for policy seem to be discussing data-driven innovation and the use of data by using or proposing theoretical frameworks from organization and management aspects rather than focusing on the process of policymaking (Dunleavy, 2016; Klievink et al.2017; Androutsopoulou & Charalabidis; 2018; Janssen & Kuk; 2016).

Among the policymaking theoretical frameworks used, the policy cycle (see Section 1.3.2.1) seems to be a prevailing choice[52] (Höchtl et al., 2016; Mureddu et al., 2012; Poel et al., 2018; Studinka & Guenduez, 2018). However, the policy cycle also seems to be increasingly emerging as a limited analytical lens (Höchtl et al., 2016; Concilio & Pucci, 2021; Longo & Dobell, 2018; van der Voort et al, 2019). The main limitation is that the policy cycle depicts policymaking as a linear sequential process (see Section 1.3.2.1) and seems to indicate that data can straight-fowardly feed into the policy cycle, an assumption which has been deemed as wrong (Giest & Ng, 2018). The model simplification is widely recognized in policy studies (Wegrich & Jann, 2006). It also seems acknowledged in data for policy field (Höchtl et al., 2016), where it appears to be extensively used

51 In Europe, recent legislation from the European Commission is specifically addressing this barrier, e.g., with the directive on open data and the re-use of public sector information (European Parliament, 2019).

52 An exception is Dunleavy (2016) that applies the NATOE framework (Hood & Margetts, 2007), thus distinguishing the policy action of government across informational, regulatory, financial, organizational and expertise-mobilization capabilities.

nonetheless instead of other theoretical frameworks.

By noting these limits, some authors suggested the policy cycle should not be used rigorously, but as an ideal unifying model of the different types of decisions happening in the policymaking process, to which data can contribute differently depending on different time frames and scopes (Concilio & Pucci, 2021).

1.3.3.5. Theme (D): The specific nature of policymaking

There seems to be emerging agreement that the use of data for policymaking should acknowledge the specificity of policymaking process (Durrant et al., 2018). In the first place, the use of data for policymaking should not be taken for granted, as it depends on the capacity and willingness of public organizations to collect them as part of their statutory mission (Klievink et al., 2017). Evidence is in fact but one of the many drivers behind policymaking (Marchi et al., 2012). Whenever policymaking happens with the intention of using evidence, the incorporation of non-traditional data sources will have to be subdued to existing political dynamics, contextualities and ideas of policymaking processes and actors (Durrant et al., 2018).

In policy studies, the fact that evidence and knowledge for policy are not given elements but mobilized by the actors involved in the policymaking processes, is a well-established fact (Radaelli, 1995) (see Chapter 2). The process of collecting non-traditional data for policy could not be regarded as neutral nor linear, because evidence in policymaking remains a public and contested matter, inevitably central to a process of collective social interaction (Höchtl et al., 2016; Kettl, 2016). This high level of interaction depends on the nature of the policymaking processes. In fact public policies, if intended as long-term decision-making processes, should happen under public legitimization and deliberation, which themselves implies high interactions among social actors (Marchi et al., 2012). As it will be further argued in Chapter 2, the necessity of legitimization (thus of interaction) also depends on the nature of policy problems, which often are developed out of uncertainty and ambiguity of what solutions should be applied to them (Hoppe, 2011, pp. 71–75). The specificity of policy problems also affects the possibility of using data and analytics, as the underlying cause is often more important to know than forecasting based on available data (Höchtl et al., 2016).

Additionally, the use of non-traditional data sources for policymaking requires the involvement of several actors (e.g., data owners, technical profiles) (Giest, 2017; Lanza, 2021). This implies that public bodies are willing to experiment with their decision-making processes and open them even further (Lanza, 2021).

1.3.4. Interpretation of themes found in data for policy

The themes that seem specific to the field of data for policy (see Sections 1.3.3.2-1.3.3.5) were critically interpreted (Table 3) to create the terrain for the research questions.

Table 3. Summary of themes/issues found in data for policy field and interpretations

Data for policy themes summary	Interpretations of themes
A) The importance of data quality Data-driven innovation for policymaking pertains to the collection and sharing of non-traditional data sources. This implies challenges that could be attributed to *data quality*.	Quality in data for policy depends on how non-traditional data will be re-used, which is itself dependent on contextual factors of policymaking process.
B) A lack of empirical cases The literature reports a scarcity of empirical cases in data for policy, although some cases of data analytics in public agencies exist.	A conceptual difference should be made between "data for public administration" and "data for policy", based on the complexity of policy problems.
C) The limits of current theoretical frameworks There is poor variety in the theory used to read the interaction between the use of data and policymaking. The policy cycle seems limited.	The field would benefit from conceptualizations of policymaking that acknowledge the role of policy actors at the implementation stages, and their role in determining policy problems and in mobilitizing the useful data.
D) The specific nature of policymaking The policymaking process is a political and public process which demands that the use of data is subdued to existing logics, contexts and ideas. It also implies high interaction between actors involved.	The contextual dynamics of policymaking processes and the agency of policy actors (as the main subject collecting and using data) are increasingly suggested as central aspect to consider in data for policy.

Interpretation data for policy theme (A)

Summary: quality in data for policy depends on how non-traditional data will be re-used, which is itself dependent on contextual factors of policy-making process.

The public sector is well-accustomed to collection of data on a large-scale, which are either used for statistical or administrative purposes. This implies that a great quantity of data is potentially already available for policymaking within the public sector (Malomo & Sena, 2015). Rather than collecting new data, it seems that integrating the data already collected by different public and private subjects represents the first way to make them valuable for policy (Durrant et al 2018). However, public bodies are suggested to have little capacity of using data that are collected by other subjects, whether public or private. On the other hand, data which are shared across governments bodies (e.g., data for statistics), are collected with certain methodological standards and by trusted public subjects (see

Section 1.3.3.3).

The innovative challenge of data-driven innovation in the field of data for policy seems to regard how to leverage non-traditional data (which are more granular and updated more frequently than traditional data) while maintaining the data quality dimension of traditional data for policy. This means resorting to data sharing and linking across public bodies and outside, and, in this perspective, the concept of data quality could be reconsidered as being dependent not only on factors concerning data collection but also on *data use* which is largely driven by contextual aspects. Data originally collected for a specific purpose could be re-used in other policy situations and needs, on the basis of contextual considerations. This *fitness of use* of data (Floridi, 2014) ought to be evaluated in respect to goals and settings which are inevitably *contextual* factors (i.e., bounded in time and space) to policymaking processes and its policy actors. This reinforces the idea that to create value of data for policymaking, the use of data must be germane to various specific contexts of which the policymaking process is made (Harron et al., 2018).

Interpretation Data for Policy Theme (B)

Summary: a conceptual difference should be made between "data for public administration" and "data for policy", based on the complexity of policy problems.

The limited capacity of the public sector to adopt data analytics and leverage non-traditional data sources, while being largely recognized, might only partially explain the paucity of data for policy cases reported in literature (Poel et al., 2018). This lack of cases also seems to contrast with the recognition that advanced data analytics systems have increasingly been developing and integrating within the functions of public agencies, e.g., in the United Kingdom [53] (Athey, 2017).

It is here proposed that an additional cause might reside in the absence of a clear conceptual distinction about which uses of data and data analytics should be considered cases of "*data for policy*", and which ones instead only represent cases of "*data for public administration*". In fact, among the few cases reported, most seem to describe uses of the employment of data analytics for supporting relatively linear processes of administrative action (Athey, 2017; Dunleavy, 2016), typically law enforcement (e.g., tax collection, policing, health inspections, etc).

While undoubtedly relevant in the overall governing process, these

53 It should be noted that the United Kingdom might represent one of the most advanced nations worldwide in terms of public data utilization, as this is suggested by OECD's Digital Government Index (2020).

cases basically describe the employment of data analytics to support the operationalization of clear goals and courses of action. Arguably, this linearity could not be featured by the majority of policy problems (Dunn, 2017; Colebatch, 2010), which can be of different types, depending on the level of uncertainty about the "right" solution; and the level of ambiguity on the "right" normative core values that drive solutions (Hoppe, 2011, pp. 71–75) (see more below, see Chapter 2).

Cases such as data analytics for addressing tax fraud (Atto et al., 2015) already present a clear structuration of a public issue (e.g., public money wasting) into a clear policy problem, for which both the cause (i.e., fraudulent citizens) and the right solution (i.e., financial sanctions) are defined. In line with that, these cases appear to lend themselves to rational problem-solving, corroborated by data analytics based on administrative data. Again, in the cases of data analytics for tax-fraud detection, which are increasingly emerging in many national governments (The Lisbon Council & Deloitte, 2021), the use of data appears to drive an optimization goal, i.e. to increase the percentage of tax collection on the rate of citizens' tax returns. Moreover, it should be considered that the public agencies involved in these cases already own the data fitted to their specific needs (Dunleavy, 2016), whereas the same data either do not exist or cannot be reused automatically for broad policy problems (Durrant et al, 2018).

Interpretation Data for Policy Theme (C)

Summary: the field would benefit from conceptualizations of policymaking that acknowledge the role of policy actors at implementation stages, their role in determining policy problems and mobilitizing the useful data.

In the group of authors that are increasingly giving shape to the field of data for policy, the conceptualizations of policy adopted seem to follow a view of policymaking conceptualized as a staged model of rational problem-solving, epitomized by the policy cycle model (see Section 1.3.2.1). While being a valuable model for taming the complexity of the policy process (Bridgman & Davis, 2003), the policy cycle carries the limits of this view into the data for policy field. Consequently, analyses tend to connect the potential of data analytics technologies/techniques to the policy stages in a prospective way (e.g., sentiment analysis used for agenda setting, monitoring used for policy implementation) (Höchtl et al, 2016), rather than discussing the context where the data is used. By doing so, the policy cycle fails to acknowledge the support-independence of data (and information in general) (Floridi, 2014), which means that data sources are not bounded to a given policy stage, but could be used differently to inform the various types of decisions of which a policymaking process is made (from the operational to the long-term planning) (Concilio & Pucci, 2021). Data collected at one stage of policymaking can also serve another stage. For example, data on student outcomes could be used either for monitoring purposes at the level of schools (Williamson, 2016) or for long-time planning (Studinka & Guenduez, 2018, p. 17). This perspective view on data in the policy cycle risks conflating the stages (Höchtl et al, 2016), thus invalidating the very essence of the model.

An additional risk here proposed is that the policy cycle implies an authoritative instrumental logic of policymaking (Hoppe, 2010; Turnbull, 2018), that maintains policymakers would first issue authoritative decisions later implemented by public agencies or other implementation actors (see Chapter 2). Contrary to this logic, rather than just being involved in realization tasks, the role of policy actors involved in policy implementation has been recognized as essential in determining policy problems (Turnbull, 2018). These actors, in fact, mobilitize the information and data considered relevant to re-problematize given policy problems into problems that are politically and operationally manageable and feasible in their contexts (Turnbull, 2018). In data for policy, not recognizing their role would imply to obfuscate the micro-level dynamics of data aggregation and use in which they are involved, because of both their role in determining policy problems and their closeness with non-traditional data sources (e.g., administrative data).

Interpretation data for policy theme (D)

Summary: the contextual dynamics of policymaking processes and the agency of policy actors (as the main subject collecting and using data) are increasingly suggested as central aspect to consider in data for policy.

The specificity of the policymaking process — in comparison to other types of decision-making processes — seems to be increasingly recognized in data for policy field (Kettl, 2016; Hochtl et al, 2016; Marchi et al., 2012). Data in policymaking, it is suggested, cannot be equated to an asset used to reduce the uncertainty of decisions, as it might happen in the private sector (Brynjolfsson et al., 2011) or in cases of "data for public administration" (see above). The ambiguous nature of policy problems, as said above, makes the use of data necessarily a matter of interaction among the policy actors involved. Accordingly, agreement on the "right" problem would result from bargaining, closed or public discussion, or imposition from a higher authority (Hoppe, 2010). Arguably, the collection of "right" data depends on the understanding of the policy problem.

This point should not surprise given that how evidence could be ignored, selected, and manipulated by policy actors is not new in policy studies (Caplan, 1975). As for any other type of evidence in policymaking, for most policy problems it would be hard to expect that one data source will present itself as absolutely superior to others (Head, 2014) or that it will speak by itself in relation to the policy problems (Durrant et al, 2018). It is suggested that, for policymaking, the use of data is influenced by the relational and public nature of the public decision-making processes, in which policy actors might create a process centred around data, but they also will — in a not neutral way (Marchi et al., 2012) — decide which data matters to them (Head, 2014) and will mobilize to integrate it (Radaelli, 1995). Arguably these policy workers could be expected to integrate data sources based on their knowledge of the policy problem they are facing. These aspects appear largely neglected in the broader narrative of data-driven innovation in the data debate. Instead of encouraging us to think that data drives policymaking, the contextual works of several authors in the data for policy field (Durrant et al, 2018) invite us to consider the innovation of data in policy as primarily driven by the dynamics of policymaking processes and policy actors (as agents behind data collection and use).

1.4. Research interests

The contemporary discussion on data and innovation (the data debate) and the data for policy field both look like topical areas to which different disciplinary perspectives and theoretical backgrounds converged (Suominen & Hajikhani, 2021). Each discipline seems to be adopting its own filter on the phenomena in these areas (Mergel, 2016). This doctoral research, too, intended to bring the disciplinary perspective of design into the exploration of the field of data for policy and, for reasons that will be explained, purposefully took interests in local government scale. These interests are not here presented as neutral lenses, but they actively shaped choices of how to frame the investigation in the field of data for policy. The lens of design has been adopted in line with those that are the present disciplinary interests of design toward policymaking. However, this interest might be not self-evident to those not acquainted with the discipline, therefore will be contextualized as part of the broader evolution of design discipline (hereinafter simply called *design*).

1.4.1. The evolution of the design disciplinary interests

The domain of professional practice and academic research of design had shifted gradually in the last decades from artifacts to complex systems (Buchanan, 1992; Jones, 2015); and from the business and commercial sector to the public and social sphere (Hillgren et al., 2011; Markussen, 2017). Until the 80s, most specialisms of the design profession in Western countries regarded only graphic, textile and industrial product design (Julier, 2017). However, because of design being a profession with no rigid curricula, as new domains of professional practices emerged (e.g., because of digitalization) the traditional specializations of design got hybridized and designers started to work in new fields and with new labels (e.g., interaction design, web design, design management, service design, etc). The teaching activity and research of universities expanded accordingly to cover these areas (Cooper, 2019; Cross, 2018; Margolin, 2016).

Two main changes might be said to have characterised the evolution of design during time (Buchanan, 1992). First, the field witnessed a shift of interest and conception about the area of design action, expanding from the tangible world (i.e., artifacts) to the intangible dimensions of this tangible world (e.g., the objects semantics, end-user experiences, systems of production) (Göransdotter, 2021; Maffei, 2022). Secondly, there has been a change in the idea of design as problem-solving activity, mostly depending on the individual agency and talent of the single designer, to designing as problem setting and inquiry activity to be developed collaboratively (Maffei, 2022; Julier, 2017; Kimbell, 2011). With design expanding into other domains, the paradigms of design action were re-articulated by new designers specializations into these new domains[54]. It should be noted that this disciplinary evolution was not driven solely by internal intradisciplinary dynamics and cultural change, but mostly happened as a response to historical factors, e.g., market changes, technological development (Julier, 2017; Maffei, 2022).

Among these external factors, some have pushed design toward the public sphere and public sector (Bason, 2017) in discontinuity with its tradition of a practice supporting the commercial sectors (Markussen, 2017).

1.4.2. The interest of design in policymaking: "Design for Policy"

In the contemporary design research discourse, "design for policy" investigates the potential of design to drive experimentation and innovation in governments (Mortati et al., 2018) . Design for policy as an area of research has mostly been shaped by reflections on the potential innovative impact of design approaches, methods and tools within the public sector, often through the reflections of actors directly involved in experimental practices that involved these elements (Bason, 2017; Bailey & Lloyd, 2016; Legeby et al., 2018; Kimbell, 2015; McGann et al, 2018; Mortati et al., 2018; Junginger, 2017). It is widely agreed that the phenomenon of *public sector innovation labs* (PSI) has been a main driver behind design for policy (Bason & Schneider, 2014; Junginger, 2014; Kimbell, 2015). The initial origins of PSI could be retraced in the aftermath of the economic

54 An exemplification case of this re-articulation of paradigms might be found in the evolution of the field of design for sustainability, driven by the vision of design as the responsible driver of social and environmental issues (Papanek, 1972). Designers interested in this vision rearticulated the paradigm of sustainability during time, changing the emphasis of design practice from the artefacts designed (e.g., focusing on the recycling of components or product life assessment) to the systems of relations and activities that would ensure sustainability goals (e.g., responsible sustainable practices) (Ceschin, 2014). As a result of this evolution, the field of design for sustainability branched into new design fields (e.g., design for social innovation) (Jégou & Manzini, 2008).

crisis of 2008, when governments struggled with improving service provision while coping with budget cuts (Julier, 2017). Incentivizing innovation became an important point in government agenda for maintaining public service quality under challenging circumstances; and several small design consultancies with expertise in service design started to offer consultancy services in the field of public sector innovation (Julier, 2017). Many were externally hired or got embodied into governments and entitled with supporting public sector innovation agenda, becoming established as *public sector innovation labs* (Julier, 2017; Tõnurist et al., 2017). Often, PSI employed design methods and tools for their activities, as these could be easily employed within a standardized and replicable approach (McGann et al, 2018). During time, design for policy — that emerged as a scholarly and practitioner reflection on those experiences[55] — tried to understand if the value of these practices could outbound the dimension of service provision (Bailey & Lloyd, 2016; Buchanan et al., 2017). The main guiding question that seemed to drive these investigations was whether these experimental settings and practices could contaminate their institutional settings and enable a more experimental form of policymaking and governance (Kimbell, 2019; Kimbell & Bailey, 2017; Legeby et al., 2018). Underlying that question, there seemed to be a normative view about the desirability of transferring paradigms of design activities into policymaking and driving public sector organizational change — for example, by enabling forms of policymaking more centred on the citizens' needs (Junginger, 2013, 2017; Vaz & Predeville, 2019). In that view, design could highlight the paradigms of established design practices in policymaking and drive new ones (Bailey & Junginger, 2014).

Parallel to a more internal disciplinary perspective, other disciplines (e.g., policy and public administration studies) inquired the relation between design practices in public sector and policy/governance (Considine, 2012; Hermus et al., 2020; Howlett, 2014; van Buuren et al., 2020), in particular to understand design contribution to forms of collaborative governance (Ansell & Torfing, 2014). Among these accounts, some questioned the limited political view of design in policy and governments in regard to political and governance structure (Clarke & Craft, 2019). These perspectives seemed to be matched by voices "on the ground", witnessing the limits of design to scale beyond experimental practices due to public

55 Service design seemed to have taken a central stage in these experiences if compared with other design approaches, such as systemic design (Jones, 2014), which nonetheless were used in the public sector (Feast, 2018; Nohra et al., 2020). The reason for this apparently prevalent role of service design in the "design for policy" discourse might be found in the importance of public service co-creation in the contemporary paradigm of public sector innovation and in the distinctive set of methods and tools that had come to characterize service design (Wetter-Edman et al., 2018).

sector culture and the political nature of policymaking (Bailey & Lloyd, 2016; Blomkamp, 2018).

1.4.3. Data for policymaking at the level of regional and local governance

Together with the disciplinary perspective of design, justified by the design for policy discourse, this thesis also takes the regional and local governance scale as its interest (Hooghe & Marks, 2016). Great emphasis has been lately put on local governments (cities, regions, local institutions) possibility of leveraging non-traditional data for local policymaking, management and planning (Bettencourt, 2014; Engin et al., 2020; Durrant et al. 2018; Giest, 2017; Malomo & Sena, 2015). Cities, in particular, had been considered an environment densely populated with people, infrastructure and services (Batty, 2015; Shapiro, 2018). Cities started to be regarded as coherent data-rich environments also for having been earlier adopters of ICTs (Kitchin, 2014). While cities or municipalities in particular seemed to have gained much of the attention in the discourse on non-traditional data[56], also other types of local governments (e.g., counties, regions, local institutions with statutory authority) were chosen as units of analysis in recent research focused on the potential of non-traditional data for policymaking (Durrant et al. 2018; Malomo & Sena, 2015; Walravens et al., 2021). It was highlighted that local governments approach data collection either by organizing them in one unique database (a data center) or by connecting existing databases through the support of data scientists in different departments (Giest, 2017).

The former approach entails the idea of building a one-stop urban analytics and simulation platform for administration, an ambition toward which local governments strived in the last years through several dashboard projects (Kitchin, 2014) and that today seems resurfacing thanks to the concept of digital twins (Notcha, 2021). This approach, however, was criticized for being highly focused on technology and thus overlooking the social and governance processes of local contexts, hardly representable in the existing data collection (Notcha, 2021). The one-stop platform scenario also presents many feasibility issues. Data relevant to local authorities are often not collected, and it requires a great technical and economical investment to aggregate and make available those who are

56 Cities in particular have been deemed as the ideal venue of technological-driven innovation, mainly branded via the buzzword *smart city* (Walravens et al., 2021). The smart city rhetoric is largely centered on the idea of collecting data from various sources in the urban environment, and has been advocated in strict connection to the big data narrative (Giest, 2017). The two views also prompted the same type of critique and lately smart city seemed dismissed (Morozov & Bria, 2018).

(Papyshev & Marime, 2021). On the other hand, several positive remarks exist for the second approach, i.e., the *ad hoc* linking of existing databases or datasets. It should be noted that local governments are often in charge of providing direct welfare to citizens and environmental services (Giest, 2015). Administrative data on the local government scale, particularly if integrated with other types of data (e.g., demographics, contextual data) represent a tremendous asset for policymaking (Durrant et al., 2018). The administrative data can give insights on the behaviour of a certain population and about the influence of certain policies (e.g., grants schemes) (Giest, 2021; Malomo & Sena, 2012).

In line with this, what seems to make local governments scale interesting in a data for policy perspective is its relatively close relation between decision-making and service delivery[57]. Consequently, local governments could use data from monitoring not only for real-time management but also for medium-term decision-making in an experimental fashion (Concilio & Pucci, 2021) (Fig. 5), in a way that is possibly unfeasible on other governance scales (e.g., national government).

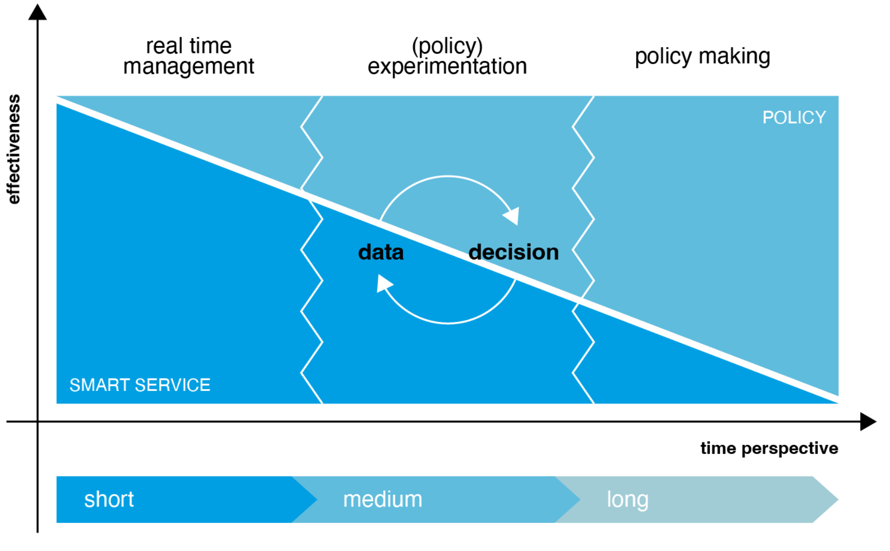


Fig 5. Schematic representation of the experimental space for local government policymaking thanks to data, between long term policy planning and real time management of digitalized services (Concilio & Pucci, 2021).

57 This should not suggest that cities act as autonomous entities. Instead, they are affected by influences from national and supranational bodies in a complex system of multi-level governance (d'Albergo, 2010).

The local government scale is also an interesting scale for data for policy because local authorities (in particular cities) have been encouraged during the last years to experiment with their governance models (Gelli, 2001; d'Albergo, 2010). This willingness to experiment with governance might be a central condition for using non-traditional data (Lanza, 2021). In fact, the use of non-traditional data in a way that is germane to currently perceived policy issues requires sense-making activity across diverse departments, authorities and stakeholders interested in these issues (Durrant et al., 2018). On the contrary, the availability of administrative data per se could not be a sufficient condition, since these are often collected for other purposes (e.g., reporting) than understanding a policy-relevant phenomenon (Malomo & Sena, 2012). However, as the other side of the coin that this scenario brings, these processes greatly complexify governance structures, in turn possibly hampering decision-making (Giest, 2017).

The conjunction of all the aspects described above makes the local government scale an interesting, perhaps ideal, level of government/governance in which developing an explorative analysis on data for policy. For reasons partly touched above, the use of non-traditional data at this scale, according to some authors, would benefit from socio technical perspectives in research (Notcha, 2021) and bottom-up approaches in practices (Bettencourt, 2014; Engin et al., 2020; Papyshev & Marime, 2021).

1.4.4. Convergences between data for policy and design for policy

Arguably, developing and innovating how governments learn about policy issues, thus improving policy and public services design and implementation, stands as an innovative proposition both in *data for policy* and *design for policy*. Both fields seek to inquire about new approaches, methods, and tools to understand public issues and orient governmental interventions through policymaking.

However, the two fields had quite different development. The data for the policy field (as shown in the previous section) seems to be emerging by developing its topics from the broader data debate. Therefore, it started from the proposition of the data debate — i.e., the proposal that non-traditional data of current socio-technical systems can drive innovation. Data for policy can be seen as a reaction to the structural conditions of our socio-technical systems, which resulted from the evolution retraced in the Introduction. Design for policy, on the other hand, seems to have developed for its value to contribute to public sector "non-technological" innovation, driven mainly by the logic of value about the quality of public services and public legitimization (Bekkers et al., 2011) to which it has added in terms of enhancing problem-setting activities, user-centeredness and co-creation (van Buuren et al., 2020; Vaz & Predeville, 2019).

The different drivers behind these fields might also explain the different approaches to enabling policy-relevant knowledge (see Section 2.5) through their practices in the public sector. D*esign for policy* seeks an inspirational approach to gather contextual/local knowledge by probing users' needs and expectations (Hermus et al., 2020); while data for policy seems to be struggling (see Section 1.3.3) with the notion of "letting the data speak", inherited by the datafication narrative in the broader data debate. These different approaches emerge from the attempt to integrate methodologies to bridge design and the use of non-traditional data (Kunneman & Alves da Motta Filho, 2020; Ricci et al., 2019) and resonate in a few experimentations in public sector innovation labs[58]. It has been suggested that a knowledge gap exists about how design can contribute to the use of data and evidence for public sector activities of formulating policies (Mortati, 2019).

The perspective above signals that interest in connecting the field of "data for policy" and "design for policy" exists. Considering that, this work is developed under the belief that there is value in closing the gap between the two fields. At the same time, to divide them by their approaches to policy-relevant knowledge presented above seems to the author to imply an artificial view of policymaking — where policy-relevant knowledge and practice are separated (see Section 2.2). If this work were to follow that vision, the challenge for bridging the two fields would be integration and cooperation (among disciplines). Both worlds will then risk sitting like well-educated guests at the formal dining table that is policymaking and keeping a distance between themselves and the actual dynamics of policymaking. That might be not advisable for none of the two fields. Design for policy is being questioned in its real capacity to engage with political aspects of the public sector and drive change behind services (J. Bailey & Lloyd, 2016; McGann et al., 2018). The review of data for policy presented above shows how the field seems to struggle with the (somehow positivist) notion of using non-traditional data for policymaking (Longo et al., 2017) — a struggle that resonates with studies on knowledge utilization in policy (see Section 2.4).

With the help of different visions of policymaking (see Section 2.2), this thesis will try to move away from the notion that policy-relevant knowledge can be separated from policy practice or developed in experimental innovation spaces. Instead, it will highlight how policy-relevant knowledge emerges from the work and activity of policy workers and that

58 For example, the Policy Lab in UK Government Cabinet Office, a notable subject in design for policy landscape, claim to use big data and "thick" data (from ethnographic research) to support a human-centred approach to policymaking (Siodmok, 2020). On the other hand, public sector innovation labs are traditionally more focused on data innovation, as United Nations Pulse Lab Jakarta showcases in their project portfolio on their website explicit use of service design methods and tools (Pulse Lab Jakarta, 2019).

activity defines policymaking (see Section 2.3.2). For data for policy, it is suggested that implies a conceptual shift: from data-driven innovation in policymaking to data-centric policymaking (UN, 2020, p. 150). Implied with that view is that data can become the centre of social activities — within the boundaries of the public sector and outside —oriented toward the resolutions of public problems. It is a perspective that resonates with the tradition of studying policy change through policy learning rooted in social learning (Hall, 1993). Under that view, competencies in data and competencies and knowledge in design should converge into practices and settings whose goal is to develop policy-relevant knowledge from policy problems. Data-centric policymaking would then be an experimental and learning process, where evidence is constructed organically through practice centered around data. The bridging of the two fields would then be a matter of collaboration and converge (fig. 6). These areas of convergence will be proposed in Chapter 5.

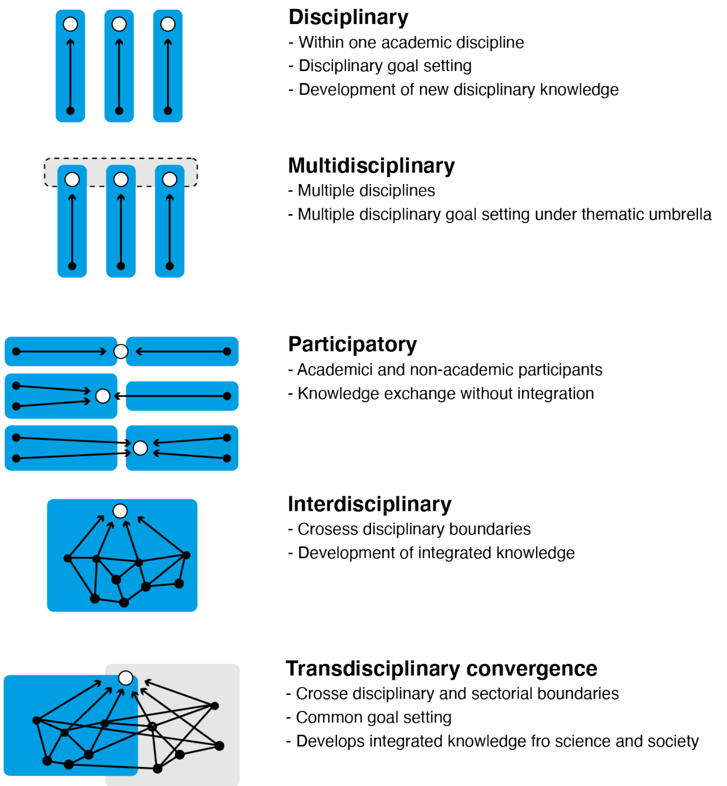


Fig 6. Representation of the concept of transdisciplinary convergence (redrawn and adapted from Morton et al., 2015)

1.5. Research hypotheses: investigating data-centric policymaking

The research hypotheses[59] proposed in this section result from the themes individuated in the data for policy field and their interpretation (see Section 1.3.4), considered in light of the specific interests of the doctoral research (see Section 1.4). Thus, the hypotheses are presented in the form of statements that, based on a first understanding of field, point at specific aspects that the research aims to investigate (Maxwell, 2012). Both these aspects (i.e., the preliminary understanding and the research interests) are conveyed through the notion of *data-centric policymaking* used as a *sensitizing concept* (Bowen, 2019; Schwandt, 2007). Sensitizing concepts are not completely operational concepts that can help to collate information upon which research hypotheses and questions are crafted, thus functioning as a synthetic and rough guide to interact with the research subject while keeping on the investigation (Bowen, 2019).

Data-centric policymaking is proposed as the process through which actors involved in policymaking create *data ecosystems* (Oliveira & Lóscio, 2018; Parsons et al., 2011) by practices of sharing, aggregating, analysing non-traditional data and interacting with them to fulfill their policy-related needs. In Chapter 2, data-centric policymaking will be detailed into a conceptual framework through theory for the analysis stage. For now, the concept is crafted out of comparison with the concept of data-driven innovation[60] and intends to stress the research's view on the subject addressed, resulting from problem setting. Hence, data-centric policymaking is used to include the following stances about the data for policy field (Table 4):

- **Stance (A)**. The *locus of innovation* in data for policy has to be searched in the practices enabled by non-traditional data sources (Micheli et al., 2020).
- **Stance (B)**. The *innovative value* of data for policy is the possibility to develop open and richer processes for the construction of policy-relevant knowledge through data practices.
- **Stance (C)**. The *potential innovation impact* is enhancing the quality and legitimacy of public problems construction (Dunn, 2017).

59 Contrary to what is often advanced, and following Maxwell (2012, pp. 81-82), this research does not consider hypotheses incompatible with the qualitative research approach, and hypotheses are kept distinct from research questions.

60 As defined (See Introduction, Section 1.1.2) and discussed above (Section 1.2).

Table 4. Stances toward innovation of data for policy: data-driven innovation in policymaking compared to data centric policymaking

Stances on innovation of data for policy	Data-driven innovation in policymaking	Data-centric policymaking
Locus	Data sources and data analytics	Practices centered on data sources
Value	Linear: Problem-solving Turning data into evidence for policy decisions	Relational: Problem setting Improve policy-relevant knowledge, in particular on policy problems
Expected impact	New decisions based on data	More collective understanding of policy problems, more dynamics of multiple evidence seeking

Data-centric policymaking essentially presupposes a subversion in the innovation perspective as this is suggested by data-driven innovation in policymaking (see fig. 7). Accordingly, the creation of value through data should not be seen as an absolute value resulting in new outputs (e.g., new policy decision), but as affecting contexts and processes. In particular, the value could be created by improving the capacity of building policy-relevant knowledge (Dunn, 2017).

More specifically, since the attempt to use non-traditional data in policymaking implies the inclusion of more actors in the policymaking process (Giest, 2017), data practices increase interactions within the policymaking processes, connecting subjects across public organizations, or even externally, with private sector subjects and citizens. While this risk to complexify policymaking and hamper the capacity to decide and act (Giest, 2017), it could also be beneficial for enacting a richer exploration and structuring of public problems (Hoppe, 2010) by the inclusion of different forms of evidence and knowledge (Colebatch & Hoppe, 2018).

It follows that data should not be considered primarily as an asset to be turned into evidence for policy decisions, but as a mean to develop more open and richer processes for the construction of policy-relevant knowledge through data practices. In essence, instead of being linearly applied to problem-solving to improve the efficiency and effectiveness of public activities and operations, data could bring public value by creating quality and legitimacy in problem setting (Bekkers et al., 2011).

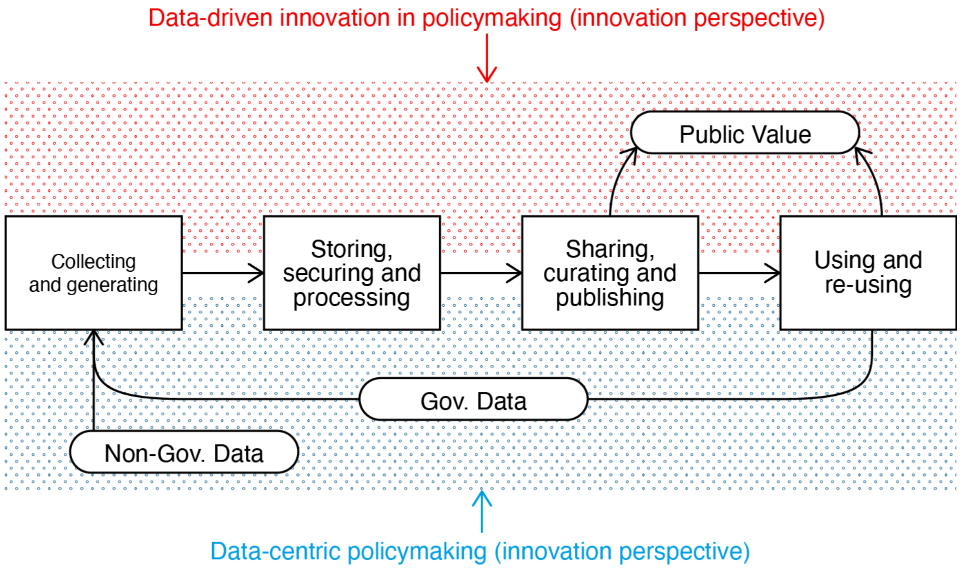


Fig 7. Perspectives on innovation between data-driven innovation in policymaking and data-centric policymaking (adapted from the Government Data Value Cycle in van Ooijen, et al. 2019, p. 11).

The concept of data-centric policymaking supports the thesis in formulating the followings hypotheses:

- H1. Data-centric policymaking would demand that the research discussion on data and policymaking (data for policy) is less tilted toward the dimension of data and technology, and more toward the one of policy processes and actors. It is expectable[61] that any explicit discourse on data for policy feature this orientation.

- H2. Data-centric policymaking is a process realized through the work and practice of actors involved in policymaking, that seek to leverage non-traditional data for diverse needs connected to policy (e.g., derive insights). These processes demand the involvement of several actors in the process of policymaking. Because of this high level of interaction, data-centric policymaking can lead to insurgence of policy-relevant knowledge among the actors involved.

- H3. Data-centric policymaking would lead differently to policy-relevant knowledge in different contexts, both on the basis of data-related factors (e.g., accessibility, sharing) and policy contexts, processes and individuals involved.

61 On the base of themes found in Section 1.3.3.

It must be said that the concept of data-centric policymaking and the relative hypotheses here proposed are meant to purposefully orient the research based on the interest in "design for policy". One of the main elements discussed by the design for policy field is the capacity of enacting collaborative governance toward the solution of public problems and including forms of knowledge that are contextual (Ansell & Torfig, 2014; Hermus et al, 2018).

The interpretation of the innovation of data in policymaking through the lens of data-centric policymaking is advanced as a possible way to connect data and design for policy, which still appear distant. Exploring this connection has been one of the purposes of this doctorate (see Publications Section) and will be carried on as an original contribution of the thesis (see Chapter 5). Therefore the fourth hypothesis assumes that:

- H4. It would be relevant to converge data for policy and design for policy approaches into data-centric policymaking practices.

1.6. Research questions

The hypotheses are operationalized in one main research question[62], pursued through three sub-questions (Table 5).

Table 5. Research questions

MRQ	**How can data-centric policymaking be realised?**
SRQ1	How is the surrounding discourse on data-centric policymaking characterised?
SRQ2	Does data-centric policymaking affect policy-relevant knowledge at regional and local governance levels?
SRQ3	What factors affect policy-relevant knowledge in data-centric policymaking at the regional and local governance levels?
SRQ4	How can we converge a design for policy and data for policy approach into data-centric policymaking?

62 The main research question starts with "*How...?*", thus it defines an explorative type of research typical of qualitative research (Maxwell, 2012; McGregor, 2018, pp. 246–247; Merriam & Tisdell, 2016, pp. 77–78). "How..." questions are formulated with the broadest possible angle on the phenomenon inquired, normally without specific reference to variables. These research questions are expected to change as relevant elements for exploration emerge during data collection and analysis (McGregor, 2018, p. 246).

The MRQ is explorative in nature and thus principally meant to develop empirical understanding in an emerging research field where scarcity of empirical cases is reported (see Secton 1.3.3.3). The questions adopt the concept of data-centric policymaking, which include aspects relevant both to data for policy in general and to this research in particular, as emerged from the review of the field (see Section 1.5). The SRQ1 decides to develop the inquiry on data-centric policymaking by firstly addressing the discourse surrounding data for policy. This seems the correct approach given the influence of undergoing debates of scientific and governmental communities on the topic of data and innovation (see Section 1.2), and the apparent specificity of data for policy in this debate (see Section 1.3.3). The SBRQ2 and SBRQ3 indicate the theoretical decisions of the research on how to address the "policy-data" interaction in data-centric policymaking, namely through policy-relevant knowledge. As said, the theoretical justification for this choice will be provided in Chapter 2.

Finally, SBRQ4 intends to explore the role of design for policy approaches and methods in data-centric policymaking, as well as their practical connection. As said, design, rather than emerging from the field, was carried to the investigation as a research interest and because of that the notion of data-centric policymaking was advanced. The value of an integration has been proposed in the hypothesis and will be explored starting from what emerges from the other SBRQs.

1.7. Research Goals

The overall purpose of this doctoral research is to argue for the notion of data-centric policymaking and explore it empirically. In doing so, the research aims to provide an original theoretical lens and empirically-based insight on the field of data for policy, where these appear as much needed. This research work could be of value to scholars of different disciplines, public officials and practitioners in government. To public officials and practitioners the research intends to also offer guidance for ground work through integrative models of practice, bringing elements of the field of data for policy together with elements of the field of design for policy.

In light of the thesis' purpose, the thesis's statements are:

1. To advance a conceptual framework of data-centric policymaking

The research will advance an original way to address empirical examples in the field of "data for policy" in a data-centric perspective, based on theoretical literature on policymaking.

2. To collect and analyse primary and secondary data on the data for policy field, in terms of its discourse and practices

The research will collect and synthetise data from literature and actors in data for policy, to provide an updated picture of discourse and experimentation in this field.

3. To collect empirical data on the connection between data-centric policymaking practices and policy-relevant knowledge

The research will collect first-hand empirical data on sampled cases considered relevant to understand the phenomenon inquired. Cases will be analyzed to understand the impact of data-centric policymaking practices on policy-relevant knowledge, and to understand how contextual variables play a role in these cases.

4. To advance understanding on the potential value of integrating data-centric policymaking with design for policy approach and methods and to offer practical guidance for this integration.

Chapter 2. Theoretical background and conceptual framework

"Whether or not one sees the existence of multiple accounts as problematic or not – and if not, how one sees the relationships between them – ultimately determines how one conceptualises policy and policy making." (Hoppe, 2010, p. 48)

Within the literature considered relatable to the field of "data for policy", agreement exists that the use of non-traditional data for policymaking is still a new-born phenomenon. The infancy of this field has been attributed to technological factors (e.g., the siloed design of ICT information systems in public organizations) as well as to a general lack of capacity and culture of data analytics/data science in public sector. The use of non-traditional data sources seems also somehow limited by the inherent nature of public policymaking process. It is suggested that public policies require data collected with the highest technical/scientific standards and — perhaps even more importantly — from sources that are legitimated and trusted by decision-makers. Because of these limitations, a lack of empirical cases is reported (see Section 1.3.3.3). As part of its interpretation of the field, this research argued that this "empirical" lack might also depend by the theoretical/conceptual perspectives currently applied to interpreting the relation between uses of non-traditional data in public sector and policymaking[63] (see Section 1.3.4).

In this vein, it is worth reminding that literature in policy studies is clear in proposing policy and policymaking not as given elements, but rather as abstract conceptual constructs used as interpretive lenses of the acts of governments (Colebatch, 2010, p. 31; Page, 2006). It follows that to decide which empirical instances (among all potential uses of data in public sector) could be considered cases of data for policy depends by how policy is conceptualized. It has been argued that, considering the nature of policy problem), some uses of data in public sector might not be representative (see Section 1.3.4)

Most of the authors reviewed seems to adopt the policy cycle model to relate innovative applications of data and data analytics to the policy-

63 What has been labeled as the "*policy-data*" interaction (Verhulst et al., 2019).

making stages (Suominen & Hajikhani, 2021) (see Section 1.3.3). These theoretical choices appear to be mostly producing prospective inferences on the potential of data and data analytics, rather than empirical and contextual analyses on the use of non-traditional data (see Section 1.3.3.1). For all the reasons exposed in Chapter 1, contextualized analyses would be of the outmost importance for data for policy. The view of the policy process underlying the policy cycle model therefore might be regarded as problematic (see Section 1.3.3.5).

The notion of *data-centric policymaking* proposed in Chapter 1 attempts to move beyond the underlying conceptualization of policy entailed by the policy cycle; a view known as *authoritative instrumentalism* (H. K. Colebatch & Hoppe, 2018a; Turnbull, 2018) (see Section 1.3.4). The scope of this Chapter 2 is to define the sensitizing concept of *data-centric policymaking* into a theoretical/conceptual framework, starting from a critique of authoritative instrumentalism, which is regarded as the traditional view of policy and policymaking (see Section 2.2). This view will be shown as faulted in giving realistic account of the nature of policy-relevant knowledge and policy problems (see Section 2.3; Section 2.4; Section 2.5). After this initial *pars destruen;* the chapter will move to a *pars construens* by starting to consider the challenge of reading the practices of using of non-traditional data in relation with policymaking (see Section 2.6).

By considering individual policy-relevant knowledge on policy problems as the potential link, the Chapter will adopt the conceptualization of policy from the **"policy work"** perspective and propose to read the innovation of non-traditional data through the concept of **policy learning** (see Section 2.7). These two elements will be combined into a theoretical/conceptual framework of data-centric policymaking to be used in the empirical research (see Section 2.8).

2.1. Premise: to work with concepts in this research

Concepts are both abstract and general: they do not represent reality, but the essential aspects of a potential class of empirical realizations (Toshkov, 2016). Analytical definitions of concepts provided by disciplines are meant to literally delimit the boundaries of concepts in their potential empirical realizations.

The interdisciplinary nature of this doctoral research, given by the perspective and treated topics, compels to develop a work capable of enacting an interdisciplinary dialogue (Pacheco et al., 2017, p. 308). At the risk of being pedantic, working with concepts seems thereby necessary for a thoroughly explanation, not only intended to provide basic definitions, but to problematize the disciplinary views on concepts.

A good part of this research demanded to clarify concepts across definitions, research traditions and communities. After reviewing the definitions and related literature, it seemed clear that these terms often entailed very different conceptualizations. Such differences could be noted both inside and across disciplines. The very concept of "policy" could entail very different conceptualization and analytical perspectives (see Section 2.2). To work with concepts appeared necessary both to explain across disciplines but also to instruct any solid empirical analysis. Consequently, the effort of constructing a theoretical/conceptual framework should not be seen as a purely theoretical reflection (Green, 2014) but as an inevitable step within the interpretative effort here pursued, to empirically inquiry data for policy through the concept of data-centric policymaking.

2.2. Authoritative instrumentalism as the traditional view of policy

As anticipated, the policy cycle (see Section 1.3.2.2) entails a conception of governmental action as rational problem solving (Hoppe, 1999; Jann & Wegrich, 2007; Turnbull, 2018). This view largely influenced the study of policies and coincides with the birth of the policy science approach in the early 50s in the United States[64] (DeLeon, 2006; Dunn, 2017).

Several political scientists, today widely regarded as the modern fathers of the field[65] (Dunn, 2017), advanced a research agenda inspired by the American Pragmatist philosophy[66]. In line with it, they advocated for the scientific inquiry as central in supporting governments's actions and fostering democratic values (DeLeon, 2006). These propositions were advanced during an historical period, the post-war, that saw the transformation of the government's role in several Western democratic society. Governments started to become not only bodies for political representativeness, but organizations which should intervene on social groups problems and needs (Hoppe, 2010, pp. 10–12). The launch of large-scale social planning initiatives during the post-war decades sparkled the demand of scientific expertise that could inform governments (DeLeon, 2006; Dunn, 2017). Sociologist and political scientists, who have traditionally studied governments politics and political communication, began to analyse the governmental action through the lens of policymaking (Colebatch & Hoppe, 2018a).

64 Dunn (2017) retraces this birth to the publication of "*The Policy Sciences: Recent Developments in Scope and Method*" edited in 1951 by Daniel Lerner and Harold Lasswell.

65 The *policy science* approach was firstly advocated through the early writings of American political scientists Harold Lasswell, Daniel Lerner, Abraham Kaplan and others (DeLeon, 2006; Dunn, 2017)..

66 In particular, Harold Lasswell would refer to the figure of John Dewey (Dunn, 2017).

The emerging professional figures of policy analysts (DeLeon, 2006; Dunn, 2017) emerged and started to be employed in the public sector for developing systematic and rigorous research on policy and for policy. In coherence with the structure of American political systems, where political elites and administrative professionals are starkly divided (Jann & Wegrich, 2007), policy analysis seek to advise politicians to "*speak truth to power*" (Hoppe, 1999).

The notion of policy — under these influences — emerged as a lens to systematically explain, operationalize and advise governmental action (Colebatch & Hoppe, 2018a). Policy as notion was greatly affected by the *problem-orientation* perspective that connoted the policy sciences (Brunner, 1991; Turnbull, 2018). The view entailed with policy saw the political systems as a cybernetic systems, receiving inputs and feedbacks from society and outputting policies (Easton, 1957; Radaelli, 1995). Thus, policy and policymaking as concepts remained fundamentally characterised from the influential work of this period that— while largely discussed in the following seventy years — still seems to have remain relevant to contemporary research on public policy (Turnbull, 2018). Manuals appears to have inherited the main attributes that accounted for policy in the view of policy science (Dye, 2013; Howlett & Cashore, 2020). Accordingly, a textbook definition of policy would define them as institutional politico-administrative actions, undertaken by the governments on the basis of its authority for consciously solving public issues (Howlett & Cashore, 2020). Policy makers are usually identified either with politicians or ministers (Kohoutek et al., 2018, p. 253) or senior office-holders (Colebatch & Hoppe, 2018a, p. 5). The official decisions taken by these figures are policies, which are then realized through the policy process.

The most famous model of *policy process* — the policy cycle (see Section 1.3.2.2) — is also attributed to one of the founding fathers of political sciences, Harold Lasswell. In line with the vision of policy science, he attempted to interpret political systems activities so that scientific empirical knowledge could be applied to improve governments' action (Dunn, 2017, p. 44, 2018; Turnbull, 2018). Lasswell proposed that policymaking could be considered as a process made of a series of purposeful functions (Dunn, 2017, p. 44). His view was later re-interpreted as a "stage" model of policymaking — i.e., the policy cycle (Jann & Wegrich, 2007). As anticipated, the policy cycle became perhaps the most widely applied interpretation of the policymaking process (Knill & Tosun, 2012). However, as explained, authors discussing policy cycle (or, in general, a stage view of policy process) might usually present only as a useful model for heuristic thinking, meant to simplify the policy process complexity (Bridgman & Davis, 2003; Jann & Wegrich, 2007). The policy cycle's strength appears to reside in its capacity of interpreting social actors' activities by relating them to a certain stage of policy process (Knill & Tosun, 2012). Further, the model had become a way through which organizing the separated field of policy studies, which would be connected to one of the stages (Jann & Wegrich, 2007).

Given these *caveat* on its use, it has been argued that the policy cycle still retains the original connotation of policy and policymaking that characterised its origins (Turnbull, 2018). This underlying view of policy and policymaking has been labelled as *authoritative instrumentalism* (Colebatch & Hoppe, 2018b):

> "Authority is perhaps the ultimate core value of policy: having the right to choose what others must do. [...] Authority is demonstrated by making this directive choice, and there is often a link to another core modernist value: instrumentality — that is, that the choice is made to accomplish some known purpose. So authority and instrumentality form part of the 'signature': the mark activity as policy." (H. K. Colebatch & Hoppe, 2018b, p. 109)

2.3. Critiques to authoritative instrumentalism

Critiques were moved against the underlying view of policy and policy process implied with policy cycle — synthetised by the label of *authoritative instrumentalism* (Colebatch & Hoppe, 2018b). These critiques span in various theoretical reflections throughout policy studies. Without the pretence of covering the totality of them, the main critical points relevant for building a theoretical/conceptual framework of data-centric policymaking are here considered. The critiques essentially challenge several attributes given to policy under authoritative instrumentalism, that derive from the dual "authority" and "instrumentality" defining this traditional account of policy (Colebatch & Hoppe, 2018a).

2.3.1. Critique of policy as choice: incrementalism

Central to authoritative instrumentalism view stands the view of policies as intentional choices to address an issue (Turnbull, 2018). The notion of choice not only suggests a deliberate action, but capacity to exactly frame a problem and moving with consciousness toward a solution. For this reason, the idea of *policy as choice* is often criticized through questioning the possibility of rational decision-making in policy[67] (Hoppe, 2018a).

67 Substantial rationality in policy decisions regards the achievement of best possible outcomes on the basis of scientific knowledge (Andrews, 2007). As said, the inclusion of systematic scientific knowledge in the policy process was also central to the policy science approach (DeLeon, 2006) and has been reflected in policy analysis, that developed in close connection to it, and, in its tradition, largely employed quantitative methodologies (e.g., benefit-cost analysis) (Andrews, 2007).

The view of policy as rational choice had been notably challenged by the perspective of *incrementalism* in policy studies (Hoppe, 2018a). Early proponents of incrementalism[68] build upon the notion that the rationality of decision-makers was bounded by available information (Simon, 1979). Therefore, incrementalists suggested policies could not be seen as the result of applying fully comprehensive knowledge to reach optimal goal (since this was seen as unattainable), but as the continuous application of political and practitioner wisdom for constantly moving toward a ameliorative status (Lindblom, 1959). Accordingly, policies would manifest as incremental adjustments of previous programs and practices (F. S. Berry & Berry, 1999). With this perspective, the incrementalism school profoundly impacted later theorization of policy process:

> "Incrementalism revealed a deep, ineliminable tension in policy process theorizing between ideal and practical thought, between policymaking as a series of science-informed choice opportunities and as a continuously evolving process of practice-informed adjustment and change" (Hoppe, 2018a, p. 398)

The limits of rational choice connects, in the incrementalism view, with the limits of central coordination and the recognition of the multiplicity of actors involved in policymaking (Lindblom, 1979). In this view, policy actors are not neutral truth-seekers, but partisans dwelling in a political arena made of conflicting interests (Dunlop et al., 2018, pp. 6–7) — therefore: "*decision-makers, pressure groups, experts and civil society organizations make policies because they have objectives of power, influence, prestige or epistemic authority in society*" (Dunlop et al., 2018, p.7-8). The recognition of this multiple interests, however, was not meant to reduce policymaking to a matter of political calculation or power plays. The incrementalism view suggests that, in pluralist democracy, not a single agency can drive policymaking instrumentally. (Hoppe, 2018a). Since many reasons and points of view exists in the political space, instead of instrumental choices, policies might result from *partisan mutual adjustments* (Lindblom, 1979, p. 522):

> "Policymaking is an interaction between democratic politics, on open and pluralist civil society, and a capitalist, market-driven economic system, which is not dominated by a single actor [...] Interpunctions (like in the orthodox stages model), as attempts to detect or impose order or regularity on the process, only obscure that policies emerge from serial,

68 Among them stands the seminal works of political scientists Charles Lindblom and Aaron Wildavsky, who developed their research starting from empirical analysis on the decision-making processes and work of policy practitioners and experts in public administration in United States (Hoppe, 1999, p. 206)

ongoing work on day-to-day problems and slow hardening of particular rules and routines [...] In this sense, the making of policies as strategic, longer-term guidelines is not just serial interaction, but also epiphenomenal, that is, not the purpose or focus, but a by-product of daily dealing with practical problems" (Hoppe, 2018a, p. 398)

2.3.2. Critique of policy as problem solving: the significance of policy workers

By recognizing the limits of scientific rationality in policy decision-making — and the complex social interactivity of the political sphere — incrementalism questioned the view of policy as the product of a unitary choice (Hoppe, 2018a). This recognition also fundamentally shatters the anthropomorphic Cartesian image of government suggested by authoritative instrumentalism, wherein the "thought" (or mind) is separated and gives legitimacy to the "action" (or body) (H. K. Colebatch & Hoppe, 2018a, p. 7). This separation would be represented, in the stage view of policy process, by the division between early stages of policy cycle (agenda setting, policy formulation, decision-making) from policy implementation. Critics of authoritative instrumentalism consider that this division obfuscates the multiplicity of actors contributing to policy-relevant knowledge; and limits policy to a matter of biased vision of problem solving (Bartels, 2018; H. K. Colebatch, 2005; Page, 2006; Turnbull, 2018). These arguments start from the views of *policy as practice,* often more specifically defined as *policy work* (Bartels, 2018; Turnbull, 2013). These views intend to develop a conceptualization of policy that gives account of this concept starting from the experience of those involved in the work that makes policies — i.e., the policy workers (H. K. Colebatch, 2010; H. K. Colebatch et al., 2010b). Policy workers are proposed as a category that includes a variety of professionals working within and together public administrations (Kohoutek et al., 2013, p. 32):

"They may be employed by the government, or one of a range of bodies concerned about how the authority of government can be brought to bear on problems: think tanks, interest groups, professional bodies, community associations, international organizations, etc. They may be activists, not employed in this sector at all, but committed to policy as a major part of their lives [...]" (H. K. Colebatch et al., 2010a, pp. 10–11)

The account of policy as practice of policy workers thus aims to acknowledge policies as the intentional activity of several actors working with a problematic and the knowledge they mobilize to address it (H. K. Colebatch et al., 2010a; Kohoutek et al., 2013; Turnbull, 2013). The more explicit strand of research on policy as practice and policy workers appears relatively recent (Bartels, 2018; Kohoutek et al., 2018; Turnbull, 2013). Although it seems to have notable antecedents in studies on implementation[69] (Lipsky, 2010; Pressman & Wildavsky, 1984) that outlined how the activity of civil servants and front-line public workers essentially constituted policymaking in the view of citizens and other social actors (Lipsky, 1971), and how these activities were drastically affecting the intended action of governmental decisions they were supposed to realize (Pressman & Wildavsky, 1984 in Bartels, 2018). For this reason some advanced these practices could be seen as policies without an agenda[70] (Page, 2006).

The perspective of policy as practice particularly focuses on the role of policy workers in creating policy-relevant knowledge. Against a positivistic traditional view in the field[71] this perspective advances policy-relevant knowledge is not only created exogenously to policy process by external experts and then fed back into it (Radaelli, 1995). Policy workers might not be in the position of modifying the high-level architecture of policy directions, but they are themselves retainers of practical and political intelligence and mobilize a variety of non-scientific and lay sources for turning these directions into policies (H. K. Colebatch, 2015; Maybin, 2015). Therefore, the perspective of policy as practice (or policy work) argues that policy workers create policy-relevant knowledge (H. K. Colebatch et al., 2010b; Turnbull, 2013):

> "Policy knowledge" is generated in use; it does not precede the action, but is part of it. Different sorts of knowledge may be mobilized, reflecting the nature of participation in the issue, and the relationship between knowledge and participation can cut both ways: those who establish their right to participate bring with them their own expertise and this shapes the sort of knowledge that is recognized; but conversely, the way in which the policy question is framed, and therefore what knowledge is appropriate, indicates who has this knowledge, and therefore should be participating. (H. K. Colebatch, 2015)

The key insights brought by this critique of authoritative instrumentalism perspective seems to be that, instead of being considered as problem-solving activity, policymaking should be seen as *problem finding* (Hoppe, 2010, p. 25; Turnbull, 2018). Through their practices policy workers re-articulate given interpretations of a problematic (i.e., policy problems)

69 These actors have been defined by Lipsky (2010) as street-level bureaucrats.

70 Once more showing a contrast within authoritative instrumentalism, these policy without agenda could *non-actions*, e.g., the laissez-faire decisions of not strictly enforcing a regulation within the degree of discretion allowed by the law.

71 The one proposed by a more positivist strand of policy analysis (Hoppe, 1999).

into new problems that are politically and operationally contextualized (Maybin, 2015; Turnbull, 2013).

2.4. Nature of policy-relevant knowledge and policy problems

The two main critiques to authoritative instrumentalism considered seem to pave the way for conceptualizations of policy other than the one connected with authoritative instrumentalism. On the one hand, the argument of incrementalism advances the concept of policy as *structured interaction* (Hoppe, 2018a); on the other, in the perspective of *policy as practice* (or policy work), policy coincides with the endeavour of policy workers, as these practitioners mobilize policy-relevant knowledge for making sense and working toward a problematic. Policymaking, in this latter view, is advanced as a process of social construction (H. K. Colebatch & Hoppe, 2018a). The two theoretical perspectives seem to connect in acknowledging the multiplicity of actors in policymaking and, for this reason, both provide interesting ground for further developing a theorization of data-centric policymaking (see Hypotheses). However, they also seem to entail slightly different path for further theorization. The incrementalism's argument looks as the right terrain for developing research on expert decision-makers' heuristics and their individual political manoeuvring (Considine, 2012). The perspective of policy workers, instead, seems to bring this type of manoeuvring outside the individual dimension, by considering the collective and knowledge-intensive dimension of policy work. This latter perspective appears of more interests for this research. Before relating the perspective of policy as practice of policy workers to data-centric policymaking, it seems necessary to deepen two concepts that are essential to it; namely policy-relevant knowledge and policy problems.

Dunn (2017) proposes a scheme (Fig. 8) of different types of policy-relevant knowledge, produced through the application policy analysis methodologies to policy components. In this scheme, policy-relevant knowledge regards:

- The assumption underlying policy problems in public agenda, to be obtained by the problem-structuring (more below on policy problems).
- The consequences of adopting (or not) certain policy measurers, via forecasting.
- The value/utility of expected policy outcomes, that can be provided through recommendations based, for example, on benefits-costs analysis.
- The consequences of policy adoption, obtained by observing and measuring policy outcomes through monitoring.
- The discrepancies between expected and actual policy performance, which can emerge from evaluation of policy outcomes.

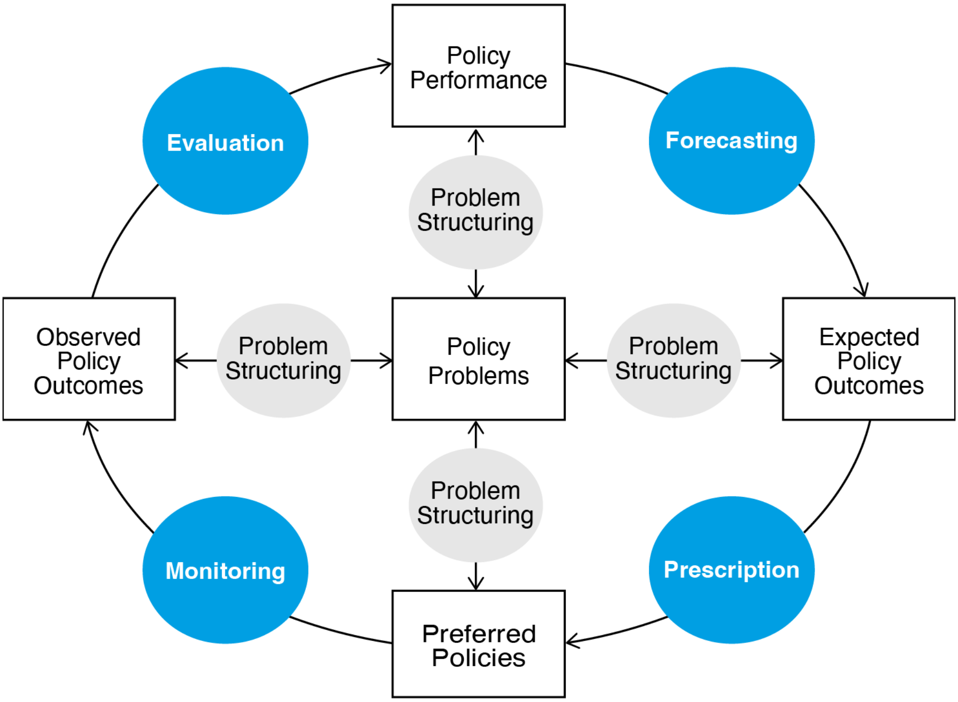


Fig 8. Types of policy-relevant knowledge derived by the application of policy analysis methods (Dunn, 2017).

Dunn's scheme (2017) usefully highlights the centrality of policy problems within any systematic creation of policy-relevant knowledge, since all forms of policy-relevant knowledge depend by how policy problems are structured (Hoppe, 2010). In Dunn's perspective policy problems are "[...] *unrealized needs, values, or opportunities for improvement*" (Dunn, 2017, p. 69), systematically structured through dedicated methodologies. It is thereby suggested that policy problems are not the same as public issues in general, and already represent a form of policy-relevant knowledge, i.e., an interpretation of what constitutes a problem and the solution to address them (Dunn, 2017; Hoppe, 2010). This structuration depends by the nature of the public issue. Hoppe (2010) provides a typology of policy problems structures (fig. 9), based on *the level of certainty over means and expertise for solving them; the level of ambiguity over the norms and value underlying these problems*.

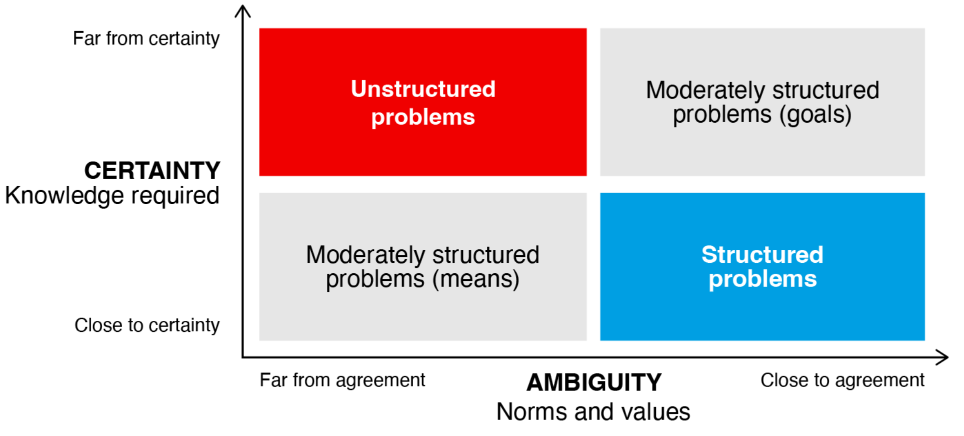


Fig 9. Typology of policy problems structures based on Hoppe (2010, p. 73) and Dunlop and Radaelli (2018b).

According to this typology, the more structured problems are those pertaining administrative and professional routine, for which high certainty on the solution exists and the underlying problematic is not extremely divisive. Conversely, the unstructured problems are those for which there is no certainty on the knowledge or body of expertise required to solve them, and there is also great ambiguity on norms at stake and values (Hoppe, 2010, p. 73; Peters, 2005). This latter category acquired much interest in policy research under the label of *wicked problems.* The term originally meant to highlight the inadequacy of a techno-rationalistic approach to most of social problems (Rittel & Webber, 1973), which depended by "*political judgement rather than scientific certitude*" (Alford & Head, 2017, p. 399). However, it has been ultimately recognized that all policy problems possess some essential wickedness (Newman & Head, 2017; Peters, 2017; Turnbull & Hoppe, 2019) or, otherwise said, a dimension of ambiguity that strictly depends by stakeholders' different interpretations.

Given this ambiguous nature of policy problems, some authors questioned if public policymaking could even be considered a problem-solving activity, given public issues do not lend themselves to a definitive answer or solution (Turnbull, 2006). Instead of providing a complete answer, policymaking could be considered as a questions-answer dynamic that continuously privileges certain frames and suppress others to move forward in engaging with the public issues. This view might be well-captured by an example from Turnbull and Hoppe (2019):

"When a policy actor responds to a problem by excluding some interpretations – e.g. drugs policy is about crime, not or less so about harm reduction – it structures the problem by limiting its scope. This is a repressing and constraining answer, but it makes progress in the sense of making possible deeper and more detailed probing. But, in remaining a (new, different, adjusted) problem, it is also a 'weak' answer that expresses the continued problematic and invites further questioning. Both properties coexist and allow policy workers to question their problems in context, i.e. with regard to their own problems." (Turnbull & Hoppe, 2019, p. 322)

In light of the nature of policy problems, Dunn's scheme seems somehow limiting, as it confines problem structuring (and policy-relevant knowledge) to the realm of policy analysis, thus a domain of experts. This view strikes with other accounts, which contend that the ambiguous and public nature of policy problems makes them subject to framing that could be advanced by several types of actors (Hoppe, 2010). It is often maintained that there is an interdependency between shared interpretation of problems at the more institutional level; and how these are considered individually (6, 2018; Surel, 2000). At the broadest institutional level, these interpretation are called policy paradigms; i.e., shared "ideas and standards" on policy problems and solutions (Hall, 1993). At the individual level, these interpretation are commonly defined as *frames* (6, 2018; Surel, 2000). Frames are a mix of assumptions/perception based on cognitive and normative knowledge (respectively knowledge on facts and value-laden beliefs) (Schön and Rein, 1994 in Hoppe, 2018b; Surel, 2000). Following the perspective of policy work, interpretations of a policy problems might fuel the practices of policy work that demand different types of knowledge to be used.

2.5. Knowledge and evidence utilization in policymaking

Research focused on policy work has shown how that policy-relevant knowledge is not only produced by experts (e.g., policy analysts), researchers or scientists (Maybin, 2016). Policy workers tend to rely on a mixture of scientific, practical-technical knowledge and political judgement for their activity (Tenbensel, 2006). This was partially already highlighted by dedicated researches[72] focused on the use of scientific/experts knowledge and evidence in the public sector (Sanderson, 2002; Weiss, 1979; Weiss, 1999). Four main modalities of using scientific/experts knowledge knowledge/evidence in policymaking were subsequently theorised (Dunlop & Radaelli, 2020):

72 The field of knowledge utilization has a long history in policy studies and had been particularly relevant during the 70s/80s (Radaelli, 1995). It seemed to have been revived by the evidence-based policy movement (EPM), which advocated for the use of scientific evidence in policymaking (Strassheim, 2018).

a) *Instrumental mode*.
Knowledge directly applied to specific decisions.
b) *Conceptual mode.*
Knowledge used as inspiration/influence for decisions.
c) *Symbolic mode*.
Knowledge is used to deliver and confirm pre-existing preferences.
d) *Imposed mode*.
Knowledge is forcefully adopted over the imposition of a higher authority.

It has been suggested that, in practice, the instrumental mode is quite rare, in respect to the conceptual mode[73] (Weiss, 1977). This seems explained by the multiple interests of the political sphere, in which *"[...] decision makers have a plurality of audiences to whom they must attend and appeal*" (Dunlop & Radaelli, 2020, p. 210). Actors involved in policymaking show willingness to incorporate evidence whenever possible (Strassheim, 2018, pp. 93–94), but the use of scientific/experts knowledge is often limited because of nature of policy problem and tight time constraints for deciding and acting (Strassheim, 2018). Under such circumstances, certain policy decisions might be informed by scientific/experts knowledge (Andrews, 2007) while others privilege other types of knowledge/evidence (Head, 2008; Wesselink et al., 2014). For these latter contexts, policy workers might privilege non-scientific but more accessible sources (Pawson, 2002; Strassheim, 2018). This has been suggested, for example, by Maybin's anthropological research (2015, 2016) in the Department of Health in United Kingdom. Maybin highlighted how civil servants in that context preferred to rely on the experiences of other colleagues and front-line workers to do their work[74] and "*get a set of proposal established as a policy*" (Maybin, 2015, p. 290). In practice, thanks to their experience and practical *policy know-how*, they were able to use the knowledge retrieved to build connections, align the stakeholders' interests, and build legitimacy so to make policy happen (Maybin, 2015).

Overall, research suggests that policy-relevant knowledge remains hard to describe in general and neutral terms. What counts as relevant policy knowledge/evidence is highly dependent by specific contexts and policy problems under question, as well as the strategy of knowledge utilization of policy workers (Colebatch, 2015; Wesselink et al., 2014). This

73 This recognition of using knowledge/evidence in the instrumental mode has been often advanced in response to the evidence-based policymaking movement (Wesselink et al., 2014). According to Strassheim (2018) origin of EBPM can be traced back to UK Labour Government reforms for the modernization of public administration, that led to several white papers publications around 1999.

74 Conversely, Maybin (2016) showed how scientific research, while accessible to civil servants in library databases, appeared as the least used.

recognition seems to resonate with part of the discourse in data for policy (see Section 1.3.3.5).

2.6. Theorising the link between data practices and policymaking

The critiques of authoritative instrumentalism and the literature on knowledge utilization provide several useful insights on the interactive nature of policy-relevant knowledge, the role of policy practice and policy workers in actively creating it, and the centrality of policy problems in these processes. These accounts appear particularly valuable for creating a framework around the concept of data-centric policymaking — that is the scope of this chapter — since they suggest a way to relate the use of non-traditional data at the level individual/group practices with the institutional and structural dimension that denotes policymaking (see Section 1.3.2.1). In fact, the establishment of this link cannot be taken for granted and it appears necessary for the analysis.

Plenty of theoretical frameworks have been proposed in policy studies (Carlsson, 2017) to help identify dynamics internal and external to governments that affect policymaking. For the most, these frameworks aim to model actors and their relation on *an account of policy based on authority and output* (Kohoutek et al., 2018). This means, in the research practice, to consider institutional outputs (e.g., publication of official documents, the establishment of an office, etc) are reference so to interpret the empirical (observed activities) or historical data as relevant for analysing policy. This approach would also essentially imply a return to the authoritative instrumental view already described, by considering the activity of policy workers in relation to an institutional decision[75]

Conversely, to adopt the view of policy *as the practice* of policy workers implies to refuse the idea that policy is an output detached by these activities, reconsidering the "policy process" as "*policy-as-process*" (H. K. Colebatch & Hoppe, 2018a). However, proponents of this view also seems aware of the *"[...] need to do better than to offer a rambling list of practices which may 'involve' policy*" (H. K. Colebatch & Hoppe, 2018a, p. 7). Therefore, the view of policy work implies methodological challenge — i.e., choices and trade-offs— common to the whole field of policy studies and well-synthetized by the renown political scientists Paul Sabatier:

> "Given the staggering complexity of the policy process, the analyst

75 Arguably, this view could translate into a solid methodology only by working retrospectively, because of the dependence by an official output that already established as official.

must find some way of simplifying the situation in order to have any chance of understanding it. One simply cannot look for, and see, everything." (Sabatier, 2007, p. 4)

This methodological conundrum do not seem completely resolved in the view of policy work, also because the aspiration of producing an universal theoretical account of practices would arguably be in contrast with their contextual nature (Kohoutek et al., 2013). However, as said, literature considered demonstrates awareness on this analytical issue and it has been proposed to consider the orientation of toward policy problem as a salient aspect accounting for policy work (H. K. Colebatch & Hoppe, 2018a; Kohoutek et al., 2013). Following this direction, on the base of what exposed above (see Section 2.4) the *knowledge of policy workers about a policy problem* could represent a potential link between the micro-level of individual practices and the macro-level of institutional dimension. Through an "agency vs. structure" dichotomy, it is here argued that:

- *Practices at the micro level interact and shape policy problems from "bottom-up.* What is considered a policy problem and how to address it (the policy paradigm) is shaped, over long period of time, by continuous collective processes of reflection on a public issue (Hall, 1993).
- *Institutions, shared norms, and rules constrain individual agents on their frames of policy problems from "top-down".* What any single individual know about a policy problem is limited, and its beliefs depends by the influence of broader institutional contexts (Dunlop & Radaelli, 2017; Grin & Loeber, 2007; Moyson et al., 2017).

In essence, a focus on policy problems might specifically identify the activity of policy workers (Turnbull & Hoppe, 2019), as intrinsically connected to policy. What remains necessary to understand is how — from this focus on problems — the innovation of non-traditional data in the activity of policy workers (i.e., in policymaking) could be understood.

2.7. Theorising the innovation of data-centric policymaking: policy learning

Various theoretical perspectives have explored the link between the ideas, knowledge, and beliefs of individual/groups and existing policy directions and programs (Table 6). For example, the Schneider and Ingram (1993), in what is considered a foundational work on the social construction of policy problems (Radaelli, 1995), showed how pre-existing ideas of decision-makers about target groups affected policy design choices and ended up privileging certain population instead of other.

Table 6. Examples of notable theoretical framework that connected the micro-level (individual/groups) to the macro-level dimension of policymaking (policies) through ideas, knowledge, beliefs

Theoretical framework	**Insight**	**Key research**
Social Construction of Policy Design	The cultural characterization of target populations affects the decisions of public officials, ultimately shaping agendas, policy design and style of participation.	(Schneider & Ingram, 1993)
Policy Paradigms	Three levels of change could be achieved by social learning within a state. Among them, third order change modify policy paradigms (i.e., existing notions of policy problems and solutions).	(Hall, 1993)
Advocacy Coalition Framework	Coalitions defines groups sharing the same system of beliefs. These coalitions compete with other coalitions (with different beliefs) in subsystem dedicated to a specific policy issue. Coalitions attempts to influence policies on the base of their beliefs.	(Sabatier & Jenkins-Smith, 1993)
Multiple Stream Framework	Specific advocates, defined as policy entrepreneurs, can promote a given policy ideas and drive institutional change once the right 'window of opportunity' presents.	(Kingdon, 1984)

While they appear too far from the phenomenon here considered (i.e., data-centric policymaking), these perspectives seem to offer useful concepts to be incorporated into a conceptual framework of data-centric policymaking. In particular, the concept of *policy learning*, that has been generally defined as "*the updating of beliefs based on lived or witnessed experiences, analysis or social interaction*" about public policy (Dunlop & Radaelli, 2013, p. 599, 2018b). The theoretical perspectives based on learning, consequently, *"consider changes in ideas as a central factor in understanding policy change*" (Grin & Loeber, 2007, p. 5).

It is advanced here that policy learning could represent the interesting variable to read the innovation of data-centric policymaking. Before arguing for this choice, the concept needs to be discussed — and so the related concepts of policy change and innovation. Policy innovation regards, in policy studies, the adoption of existing policy measures across jurisdictions (e.g., nations or local governments), rather than their creation *ex novo* (F. S. Berry & Berry, 1999; Freeman, 2006). Policy change, on the other hand, has been defined as incremental adjustments to existing policies in the form of modification or amendments to regulations (Bennett & Howlett, 1992). Traditionally policy innovations were studied *ex-ante* to understand the interdependencies of governments systems (e.g., federal states, local governments). For explaining these macro-level dynamics, policy learning across political systems was conceptualized as "lesson drawing", "policy transfer" and "policy diffusion" (Karch, 2021; Moyson & Scholten, 2018; Rose, 1991).

Policy learning, intended as social learning at the micro-level of group and individuals[76] has been studied and considered as a driver of policy change (Hall, 1993; Heclo, 1974; Sabatier & Jenkins-Smith, 1993). Policy change would then not be only driven by dynamics of power, but by network of actors "puzzling" over policy problems (Hall, 1993; Heclo, 1974). A notable example to illustrate how learning is used as the *explicans* of policy change, could be found in the Advocacy Coalition Framework (ACF) (Sabatier & Jenkins-Smith, 1993). According to ACF, change on policies depends by the influence of individuals/groups (advocacy coalitions) that gather around a shared set of beliefs on policy problems and solution, thus giving shape to a policy subsystem (i.e., a part of the political space focused on a certain issue). As these coalitions change their existing beliefs because of internal discussions over policy problems, or "shocks" external to subsystem[77] they develop *policy learning* at various level of beliefs eventually affecting policies.

Thanks to these type of studies, policy learning became a well-established field of the policy studies[78] and it has been proposed as interpretative lens of policy processes and their dynamics[79] (Dunlop & Radaelli, 2018a; Heikkila & Gerlak, 2013a; Moyson et al., 2017). Typologies of policy learning outcomes (Bennett & Howlett, 1992; May, 1992) and explanatory dimen-

76 At the organizational level, policy learning has been considered mostly in terms of analytical capacities of government of acquiring and using evidence/knowledge for policy (Borrás, 2011; Howlett, 2009).

77 In the Advocacy Coaliton Framework, for example, despite learning taking place in a policy network, innovation at the level of core beliefs might be provoked by factors which are totally exogenous.

78 The historical review of Dunlop, Radaelli and Trien (2018) traced it back to the philosophical tradition of pragmatism, which desired to break away from ideological approaches to public policy and analyse the mechanisms "that worked" so to learn from them (Dunlop et al., 2018, p. 4).

79 In contrast with knowledge utilization, policy learning intended to explain also unintentional dynamics of knowledge within networks of actors involved in policy (Heikkila & Gerlak, 2013b).

sions (Dunlop & Radaelli, 2013) were proposed accordingly. Policy learning, however, appears to swing between positive and negative accounts. Positive perspectives consider it a useful concept and theoretical framework not only for reading knowledge utilization, but the diverse dynamics of policymaking in a governance perspective (Dunlop & Radaelli, 2018a) — that is to say, in a multi-actor and complex perspective, that acknowledges policymaking depends by a multiplicity of stakeholders.

On the other hand, negative accounts (Bennett & Howlett, 1992; Goyal & Howlett, 2019) find it too theoretically fragmented to be employed in explanatory researches. The main critical argument moved against policy learning as explanatory variable is that it remains difficult to isolate what exactly causes learning in the complexity of policy settings, or even to define when learning does not occur (Goyal & Howlett, 2019). As consequence, the link between policy learning and policy change remains investigated by many, but never presented as obvious (Moyson et al., 2017) Nevertheless, policy learning seems to be thriving through a great variety of theoretical and empirical studies which shown a great level of coherency in terms of methodologies applied (Squevin et al., 2021). Perhaps because of its emphasis on the macro-/micro-level link, the concept also appears to have travelled into public sector innovation literature as part of definitions of policy innovation[80] (Windrum, 2008).

2.8. Proposing a theoretical-conceptual framework

The theoretical review of this chapter intended to substantiate a theoretical-conceptual framework that could further define the sensitizing concept of data-centric policymaking (see Section 1.5). The use of non-traditional for policymaking appears as a new phenomenon with scarce theoretical development (see Section 1.3.3.4). The various perspectives reviewed from policy studies do not seem to deal with technological innovation; and they usually consider change and innovation on a decades-long time frame (e.g., ACF), that makes hard to apply them not retrospectively. However, they represent a useful contribution, in terms of concepts

80 Some authors lamented that policy studies have neglected various dimensions of policy innovation, by equating it to political reforms (Windrum, 2008). Among authors within the field of public sector innovation, the concept of policy innovation has been defined, for example, as: "new policy directions and initiatives" (Mulgan & Albury, 2003, p. 3); "a specific kind of innovation that involves the formulation, implementation and diffusion of new visions of what a good society is, concrete goals inspired by these visions, and strategies for moving society in the desired direction" (Agger & Sørensen, 2014, p. 189) or "the change of values and knowledge in a policy network" (Windrum, 2008, p. 10).

and links, to reinforce this research theoretically, given that, at this stage of data for policy, no choice would be completely optimal.

This second Chapter attempted to isolate elements and perspectives useful for conceptualizing policy and policymaking in a way that would allow the application of data-centric policymaking framework to experimental practices of non-traditional data in public sector. This resulted in the adoption of two main theoretical perspective, here articulated:

Choice (A) Data-centric policymaking is conceptualized as the *practices of policy workers* using non-traditional data to address policy problems.

This choice allows to escape the theoretical limitations individuated in data for policy (see Section 1.3.4). Without questioning the policy cycle as device for heuristics, the intention here is to move beyond most of the prospective views it seem to have generated in data for policy (see Section 1.3.3.4) and, conversely, to employ a conceptualization of policy that can provide the field of data for policy with contextual view focused on practices (see Section 1.7). To avoid the risk of ending up into a purely anedoctical and idiosyncratic analysis because of the variety of different practices, the attempt is to focus on the dimension of policy problems, which remains a common account of policy work throughout different contexts. Further, a way to reach insights on the innovation data might represent across different contexts of data-centric policymaking is proposed through the second theoretical choice (next point).

Choice (B) The potential innovation brought by non-traditional data in data-centric policymaking is conceptualized as policy learning.

Change ascribable to policy dimension entails and usually is regarded as having an institutional dimension (what is often called the macro-level) (Moyson et al., 2017). The time scale of this change therefore can be decades-long, and it could be hardly detectable if not retrospectively. Why conceptualizing policy as practice partially resolves this problem, it might remain complicated to understand the “policy-data interaction” in contemporary cases.

In this research, policy learning — intended as social learning at the micro-level of policy workers — is seen as a possible interesting, while well-established, theoretical perspective to understand this change; also, in line with data-centric innovation and the aspects in data for policy that this concept seeks to enlighten (see Section 1.5). The knowledge of policy workers is connected to broader organizational and institutional framing of what is perceived as a policy problem, and this remains, in the policy work perspective, a constant of different policy practices. To consider variations in individual policy learning across different practices of data-centric policymaking is not intended to draw a causal link between the use of data and institutional change, but to explore how different forms of data-centric policymaking affects policymaking differently. In essence, considering policy learning is not intended to describe the innovation of non-traditional data in absolute terms, but the different forms this innovation can take.

Choice B implies, in coherence with choice A, to consider policy innovation as *innovation in the process of policymaking*, rather than in new policies or policy adjustments (Vaz & Prendeville, 2019). Finally, choice B seems to follow the line traced by some authors in the data for policy field, that started to consider the concept of learning in relation to policymaking. Among them, some positive accounts proposed that the availability of non-traditional data could transform the whole policymaking process into a social and experimental process of learning, where evidence — being more widely available — could be used throughout time beyond expert advices and the political/symbolic use (Concilio & Pucci, 2021, p. 8). Conversely, other authors hold negative perspectives on learning and the use of non-traditional data for policy, as the necessary outsourcing of data analytics competences would hamper, instead of reinforce, the capacity of governments to develop effective policymaking (Giest, 2018). These views suggest the choice of investigating policy learning might be a valuable contribution to the field of data for policy.

The conceptual framework proposed (fig. 10) advances that policy workers collecting, sharing, and processing non-traditional data define *data ecosystems* (i.e., a series of actors connected by data sharing and processing) (Oliveira & Lóscio, 2018; Parsons et al, 2011). Their practices as data ecosystems account for a form of policy practice that, in the view of policy work, equates with policymaking — i.e., they engage with data-centric policymaking. As result of these practices (and other potential factors), their existing frame of policy problems, defined at the micro-level by cognitive and normative knowledge about an issue, might change. This change, if compared across different cases of data-centric policymaking, can highlight how different uses of data affect policymaking. A more detailed view of actors, practices and knowledge change in data-centric policymaking will be defined in Chapter 3, as the framework will be operationalized through the methodology.

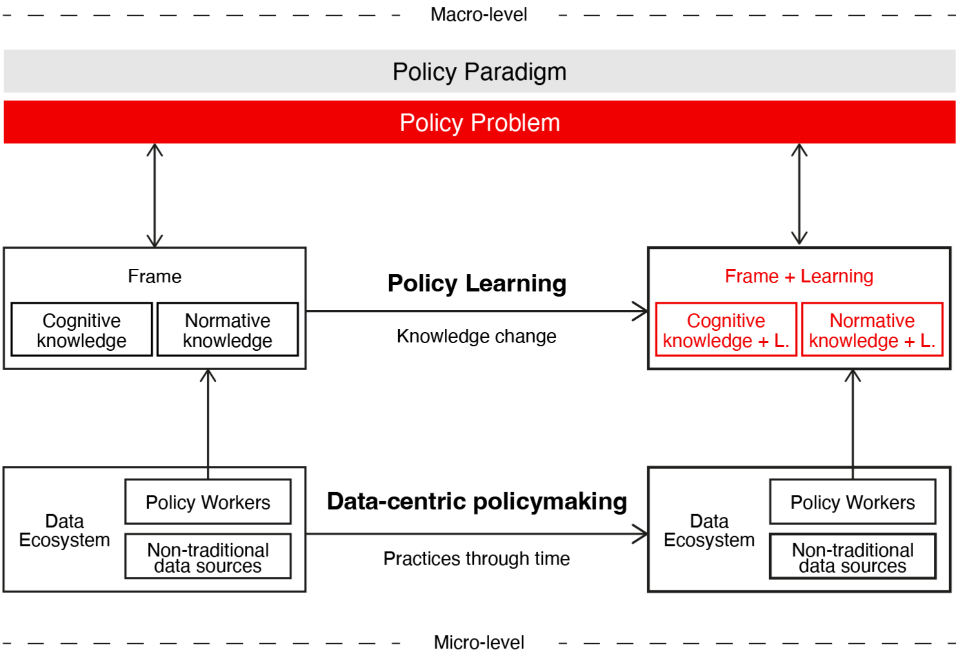


Fig 10. Conceptual/Theoretical Framework of data-centric policymaking. The new knowledge acquired by policy works on non-traditional data on policy.

Chap. 3. Methodology

This chapter describes the methodological framework of this doctoral research (fig. 11), starting from its epistemology, Critical Realism (CR). This epistemology was chosen in coherece with the research design adopted, that aimed to consider both the existing discourse and the practices in data for policy (see Section 3.1). It addressed them respectively through sub-research question 1 (that refers to discourse) and the sub-research question 2 and 3 (that refer to practices and contexts). Accordingly, a two-phase research design was imagined, with each phase respectively dedicated to the two dimensions.

The "discourse" of the emerging field of data for policy — whose main topics were identified and reviewed in Research Setting (see Section 1.3.4) — is considered relevant as an element surrounding the practices of data-centric policymaking. Therefore, the first phase of research design, called "*Understanding the Data for Policy Discourse*", aimed to refine, and further expand, the list of topics identified and to understand how the overall discourse is balanced (see SRQ1). The first phase thereby employed qualitative methods to interpret the discourse, either in spoken or written form, such as qualitative interviewing and literature analysis.

The second phase aim to shift the research inquiry from the "discourse" surrounding data-centric policymaking to those practices considered to account for this concept empirically (see Chapter 2). The phase was labelled "*Understanding data-centric policymaking practices*"[81]. During this phase, desk research was used to create a list of cases based on secondary data based on theoretical assumptions. Further, primary data collection, via key agents interviews and survey, was pursued for those cases in the list that were willing to participate to the research.

The methodology chosen for phase 2 intended to embrace the rich diversity of cases, while not remained confined to individual idiographic cases. To address this challenge the research decided to experiment with a comparative set-theoretic methodology called Qualitative Comparative Analysis (QCA). In line with QCA, the data from cases were interpreted as configurations of condition variables a single outcome variable (Rihoux & Lobe, 2009). Based on the hyphotheses formulated, the chosen conditions variables were divided in *structural factors*, intended to describe the macro-level contextual conditions that would enable/hinder non-tradi-

81 The word "practices" used here intends to reinforce the division "discourse-practices" of the research design rationale. Admittedly, it might sound redundant since data-centric policymaking is conceptualized in the research as practices.

tional data use in the cases; and the *learning factors*, that would describe the meso/micro level conditions which affected practices of data-centric policymaking. Policy learning was identified as the outcome variable (or dependent variable) to read cases of data-centric policymaking, In line with the theoretical-conceptual framework.

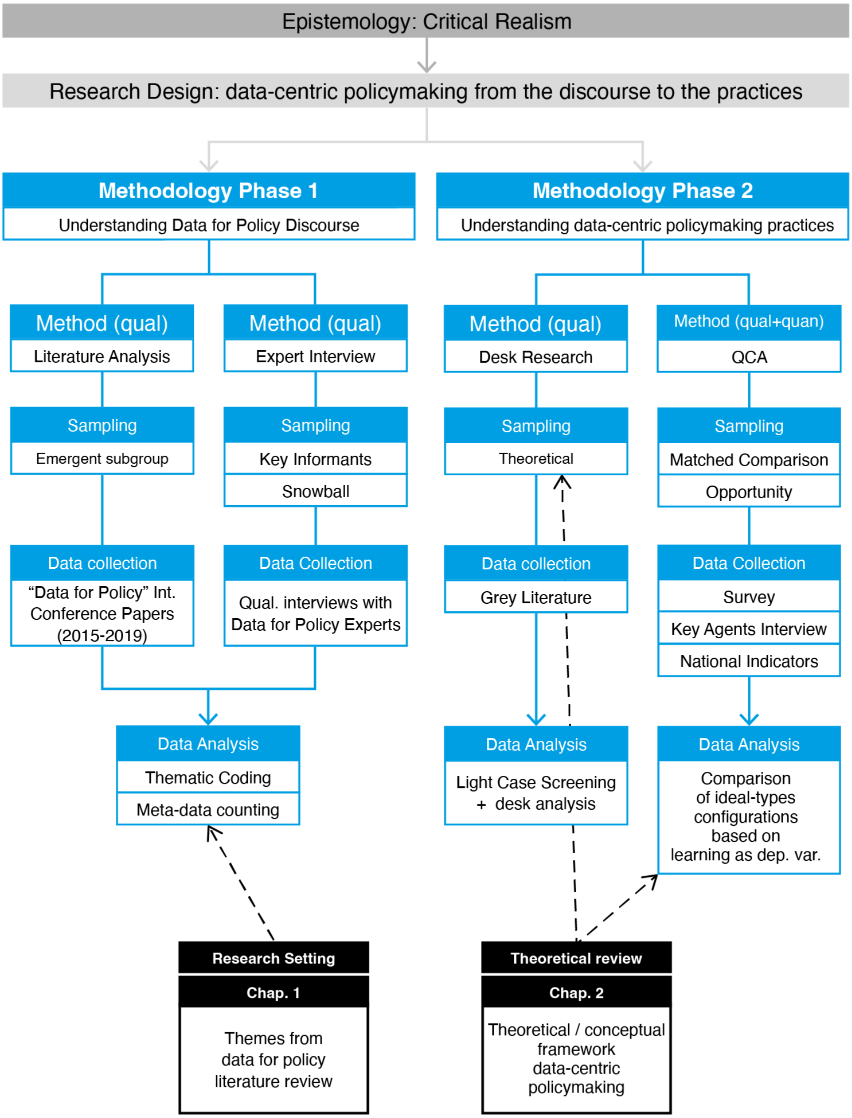


Fig 11. Methodological framework and methods

3.1. Epistemology

The research crafted its methodology throughout time to address the RQs in the most sensible way. Critical Realism (CR) (Cruickshank, 2003; Hartwig, 2016) was used as a reference throughout this construction process, for maintaining methodological coherence both toward the phenomenon inquired and across the method adopted. CR was originally advanced by Roy Bhaskar (Fletcher, 2017), stemming from the aspiration that philosophy could be an integral part of the world it aims to describe (Hartiwig, 2016). The scope of CR as epistemological view is to provide a philosophical *under-labouring* of the scientific practice (Hartiwig, 2016). This means that, rather then a prescriptive philosophical vision, CR offers a meta-theoretical understanding of phenomena; a conceptual terrain from which to derive theoretical and methodological elements (Hartwig, 2016; Svensson & Nikoleris, 2018, p. 256). In this sense, rather than being tacitly adopted, CR and its contribution to single researches are often explicitly referenced within scientific writings (Gerrits & Verweij, 2013). The value of CR mostly seems to regard a satisfying view of complexity that also accommodates for an appreciation of casual mechanisms (Gerrits & Verweij, 2013). This capacity might depend by how CR conceives ontology; as well as from its reconsideration of the traditional separation between ontology and epistemology (Hartwig, 2016, p. 6). CR partly inherits elements from the philosophical doctrine of Realism that, generally speaking, considers what is real to exist independently from our knowledge of it (Schwandt, 2007). Nonetheless, CR distances itself from the more rigid positions of Realism (known as direct realism or naïve realism) that would consider " [...] *things in the world just are as they appear*" (Schwandt, 2007, p. 256). In this distancing, CR does comply with an *ontological realism*, that thus conceive phenomena as distinct and independent from empirical observations (Fletcher, 2017). At the same time CR also accepts *epistemological constructivism*, which implies "[...] *our understanding of this world is inevitably our construction, rather than a purely objective perception of reality*" (Fletcher, 2017, p. 52). The ontological stance of CR pivots on a view of reality as an open system of complex underlying structures, which can be investigated only in their emergent aspects through partial and competing explanations (Gerrits & Verweij, 2013; Svensson & Nikoleris, 2018). CR thus maintains a "stratified view" of reality, from which it follows that what is observable results from underlying causal interplays across its different layers. The inner level constitutes the "empirical", i.e., what we observe and experience; the "actual" level regards the phenomena that exist, either if observed or not; and the "real" world level encompasses enduring mechanisms and structures which manifest themselves on the other two levels (Harvey, 2009; Svensson & Nikoleris, 2018).

Adopting CR as overarching epistemological perspective in this research most practically demanded a careful appraisal of two points that emerged while crafting the methodology: first, what value should be given to data emerging from the subjective experience (e.g., from interviews); secondly, what use of case-based methodology would be coherent in a CR perspective.

The review of the overall data debate (see Section 1.2) indicated that progressive narrative — albeit allegedly biased by techno-optimism — could affect the discourse on data and innovation to its very core. In this respect, CR encourages researchers to a lucid vision that separates the level of discourse from what is "*real*". Appreciating this separation also implies not prioritize one dimension over the other but to understand their interplay. Accordingly, CR is not incongruent with qualitative methods that focus on language, discourse, or other manifestations of subjective experience (e.g., as qualitative interviews) (Fletcher, 2017). What the CR stance most essentially rejects is the notion that the world is socially constructed (Cruickshank, 2003). Thus, according to CR, subjective accounts cannot be held as fully representative in themselves but, as insurgences of underlying phenomena of interest, they offer potential routes for understanding them (Smith & Elger, 2014). In line with that, the use of theoretical perspectives is a common trait researches informed by CR, since what is not empirical has to be interpreted through the empirical world (Fletcher, 2017). This work also explored the field of data for policy through the personal experience of individuals involved in data for policy and data-centric policymaking, either acting as producers of this discourse or actors within practical experiences, while integrating these accounts with quantitative approaches.

The second critical point, toward which the CR view had cautioned this research, concerned the use of explorative case studies, a method that was foreseen as sensible mean to investigate an emerging field as data for policy (Yin, 2018). While CR appears very well appreciated in combination with case-based research (Byrne & Ragin, 2009), it seems almost never mentioned in relation to single cases or purely qualitative methodologies (Fletcher, 2017). CR seems to compel the researcher to treat what is empirical (as the case studies) as contingent and contextual manifestations of an emergent reality (Gerrits & Verweij, 2013; Harvey, 2009). Hence, a reality of interest can manifest across different contexts, through different combinations of factors, and remains only partly observable (Harvey, 2009). Consequently, cases are to be considered as bounded objects not fully holding the phenomenon, but contingent instances that suggest something about it. Case studies are intended as configurations to be deconstructed and interpreted, by keeping a vision always broader than the single case (Gerrits & Verweij, 2013). This essentially means that CR is an epistemology that carries conjunctural causation (Thomann & Maggetti, 2020) within its vision of reality as an open system (Harvey, 2009). This view is welcomed by case-based comparative and variable-based methodologies; and for this reason, this work decided to experiment with one of them in the second stage of the research design.

3.2. Research Design

The rationale and assumptions of Research Design are here described by referring to its two phases separately (Fig. 12). At this stage, it is worth reminding that the MRQ (see Section 1.6) intended to drive the research through the concept of data-centric policymaking (which synthesized both hypotheses on the field and research interests). To answer the MRQ, it was assumed as necessary first to investigate the discourse surrounding data-centric policymaking (SRQ1); and then use it to analyze the types and contexts of practice that account for it (SRQ2, SRQ3). It seemed imaginable that the first dimension (the discourse) would take place at the level of data for policy as a field (see Section 1.3.1), while the second dimension (the practices) was to be investigated within the possible uses of non-traditional data in the public sector to address policy problems.

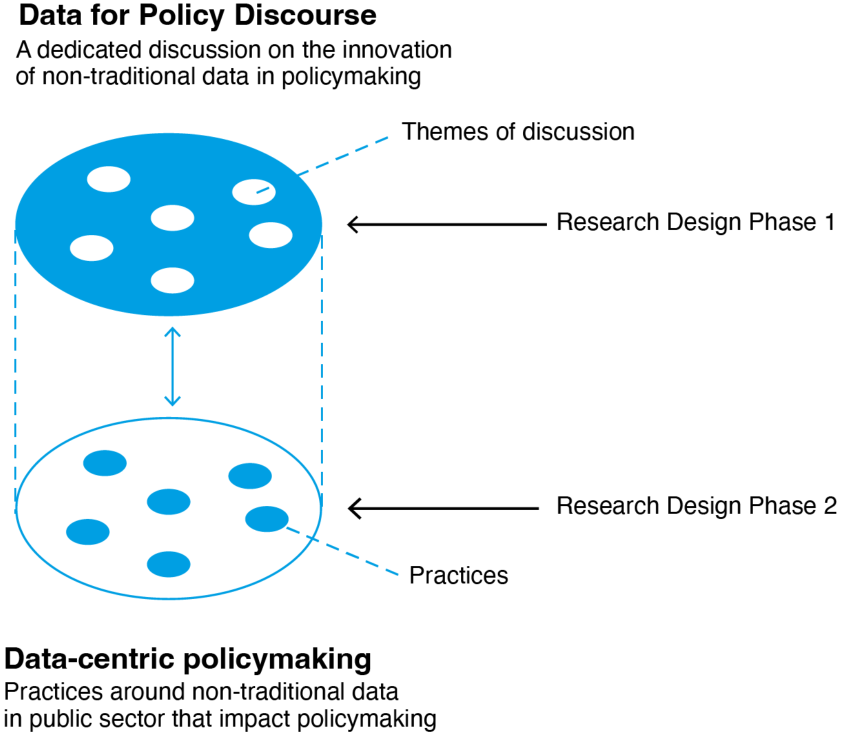


Fig 12. Scheme of research design

3.2.1. Phase 1 - Understanding the Data for Policy Discourse

The review of the field of data for policy (see Section 1.3.3) suggested existing friction between high expectations on the impact of non-traditional data in policymaking and realistic accounts of its current concrete realizations (Klievink et al., 2017; Poel et al., 2018). The discourse dedicated to data for policy appeared influenced by narratives similar to those characterizing the data debate more generally (Vydra & Klievink, 2019) (see Section 1.2). At the same time, different perspectives seem to be emerging in the literature, which cautioned the enthusiastic accounts of the specificities of policymaking (see Section 1.3.3.5).

Therefore, getting a sense of how the discourse is data for policy oriented — and which topics are perceived as relevant — has been considered a sensible step toward understanding data-centric policymaking. Hypothesis H1 (see Section 1.5) assumed this discourse to engage with the innovation represented by non-traditional data in a different way than the dominant view of the data debate (i.e., datafication) (see Section 1.2) and closer to alternative views (e.g., Critical Data Studies). The research expected to encounter a discussion more oriented on policy actors and the policymaking process and less on data and technology.

Understanding the extent of the former theme remains a major research interest, exemplified by the concept of data-centric policymaking, given several critical points that emerged in the review (see Section 1.3.4). Assuming data for policy represents a coherent and independent field of research and practice (Section 1.3.1) with its topics (Section 1.3.3), the first phase intended to understand the internal orientation of the discourse in data for policy by looking at its internal topics.

3.2.2. Phase 2 - Understanding data-centric policymaking practices

After the first phase, the research decided to move away from the "discourse" surrounding data for policy to focus on "practices" in this field. This goal essentially implied developing an empirical investigation through the concept of data-centric policymaking, following the theoretical-conceptual framework presented in Chapter 2.

The primary motivation behind this phase was to address the lack of empirical cases in this field (see Section 1.3.3.3). Hypothesis H1 assumed data-centric policymaking cases could be found in the practices centred on non-traditional data in the public sector. It also assumed different actors (i.e., policy workers) involved in these practices would gain new policy-relevant knowledge through the social interaction process enabled by non-traditional data. Hypothesis H2 assumes that these practices of data-centric policymaking are highly contextual and depend on several factors. Assumptions of H1 and H2 were translated into SRQ1 and SRQ2, which aimed at reading data-centric policymaking through types of practices and contextual factors.

Therefore, the methodological choices of phase 2 followed the goal of isolating cases of data-centric policy through typologies dependent on contextual factors. On the other hand, the theoretical-conceptual framework was operationalized in this second phase methodology to read these practices broadly and from a policy perspective. Most practically, this phase goal was to consider micro-level practices through the lens of policy learning, hence understanding which types of practices non-traditional data enabled by integrating into the puzzling and learning dynamics of policy workers toward policy problems.

3.3. Methodology of Phase 1

Sociological interpretations of technological innovation highlight how the networks of interests that gathers around any emerging innovation actively contributes to its development (Akrich et al., 2002). In line with that, experimental practices of data-centric innovation could be considered shaped by all potential forums explicitly discussing the innovation of data in policymaking, forming a broad general discourse on data for policy. As a sample of this broader discourse the research choose to analyze the "*Data for Policy Conference*", as an explicit example of a venue of discussion in this field. This author engaged with this conference as attendee and author during the doctoral path (edition 2019, 2020). Networks other than this one may indirectly/implicitly address *data for policy* within similar but differently focused arena of discussion (e.g., e-government). The conference, however, arguably represents the most explicit one for the phenomenon inquired.

Hence, the research decided to investigate the "data for policy discourse" starting from this community through document analysis and expert interviews. The data collected through the two methods were then jointly analyzed following the principle of triangulation (Flick, 2018) through a coding approach (Saldaña, 2013) that employed the themes found in the first review of the field (see Section 1.3.3).

3.3.1. Qualitative Interviewing

Qualitative interviewing (Brinkmann, 2013) was deemed as the logical method to reach an understanding of the data for policy discourse through the accounts of those involved. Interviewing (often referred as qualitative interviewing) is one of the most diffuse research methods in the history of social research, and still today paramount within qualitative methods (Brinkmann, 2013). Interviewing is a qualitative data collection method based on conversation. By the act of interviewing, the researcher can investigate a phenomenon of interest with two epistemological approaches toward conversation. In the first approach, the researcher collects data by eliciting the interviewee's personal experience; in the second approach, interviewing allows the researcher to develop knowledge by analyzing the discourse on a certain topic. It should be noted that, in both the approaches, the *subjective experience takes the central stage in this research method*. This reveals the well-recognized and extensively discussed double-edged nature of interviewing, wherein its main strength (i.e., subjectivity) is also considered its main weakness, due to the difficulty of generalize results (what would be called external validity in the experimental method or, simply, validity)[82] On this point, Brinkmann, who developed great methodological work on interviewing, says:

> "Contrary to widespread criticisms that qualitative research is too subjective, one might argue [...] that qualitative interviewing is in fact the most objective method of inquiry when one is interested in qualitative features of human experience, talk, and interaction (at least if objectivity means being adequate to a subject matter)" (Brinkmann, 2013, p. 4)

Since the method of qualitative interview essentially deals with *the subjective dimension* (Brinkmann, 2013), the possibility to understand complex intersubjective phenomenon through this method should not be taken for granted. Further, methodological textbooks caution researchers about the uncritical adoption of qualitative interviews as data collection method. Brinkmann (2013, 2018) affirms that researchers often will not engage in critical reflection when adopting qualitative interviews, because they are very diffuse and apparently easy method to implement. A preliminary reflection should always relate interviewing with the research's epistemological orientation and the nature of the object of analysis (Brinkmann, 2013, 2018). Given the epistemology adopted (see Section 3.1) and the nature of the phenomenon under inquiry, the use of qualitative interviewing was critically reviewed. As part of the development of methodological tools for this research, the use of qualitative interviews for researching complex innovation phenomena was supported from a systematic literature review of 26 studies (Annex 1).[83]

82 The external validity issue has always been considered the Achille's heel of qualitative research, especially for those to adhere to the epistemological tradition of naturalism (Moses & Knutsen, 2019). Others epistemological paradigms have emerged against this "conventional wisdom" of science; advancing that a qualitative scientific inquires based on small samples (e.g., case study) are suitable for advancing theoretical knowledge (Flyvbjerg, 2006).

83 The systematic review analyzed articles from the scientific database Scopus, with queries aiming at selecting the use of qualitative interviewing within stud-

To mitigate the biase of the subjective dimension, the review suggested that interviews could be integrated with other methods based on quantitative analysis. This insight translated in the choice of considering recurring themes in the other method employed in this phase, the literature analysis (see Section 3.2.3), and to connect the two sources of data through coding. The review also highlighted the importance of designing interviews depending by the type of interviewees. In fact, methodological literature suggests that different types of actors might possess different type of knowledge and therefore the researcher should consider their positionality in respect to the phenomenon investigated, so to understand how to elicit information through interviewing. In particular, the review highlighted three typologies of qualitative interviewing: the expert interview (Bogner & Menz, 2009), the key informant interview (Marshall, 1996) and the key agent interview (Döringer, 2020).

In light with this, the research prefigured two main typologies of actors potentially to be encountered in this investigation: i) *the data for policy Expert;* ii) *the Key Agent in data for policy practices*.

Categories of interviewees identified

Qualitative interviewing supports the researcher by giving him access to the knowledge possessed by the interviewee. Literature on this method suggests that not all actors have the same type of knowledge on a subject matter; thus, who is considered "acknowledgeable"will depend by the research's focus (Bogner & Menz, 2009). Hence, the importance of keeping types of interviewees separated, since the analysis of interview for each type could contribute differently to the research goals. In line with the overall rationale that drove the Research Design, actors to be potentially interviewed were divided in two categories[84]:

ies whose areas could be comparable to the one addressed in this PhD research. Articles were retrieved through a series of queries that could address the following areas: a) "public sector", b) "public sector innovation", c) "technological innovation", and "public sector"; d) "governance" and "public sector". For each area, the ten most cited articles were selected. Excluding duplicates and those out of scope, the final review focused on 26 studies (note: the query of area "c" retrieved only five articles). For a complete review of the queries used for each area and their description, see Annex 1.

84 These two typologies of actors are based on the author's notion of data for policy as field, also on the basis of first-hand experience as participant at the Fourth and Fifth International Data for Policy Conference (during 2019 and 2020).

The data for policy Expert

The definition of expert in qualitative interviewing does not only identify actor's possessing above-average cognitive knowledge on a subject, but those actors with a privileged view on a phenomenon under transformation (Bogner & Menz, 2009; Döringer, 2020). Accordingly, it would be impossible to expect someone to be an expert on *data for policy* strictly-speaking, as this field appears currently under-development[85] Nonetheless, several actors could be imagined possessing a broad perspective about the innovation represented by non-traditional data in government; and the specific topics and challenges this implies. It appears legitimate to consider them "*data for policy experts*", on the base of their privileged view on this evolving landscape (Pfadenhauer, 2009), made of culture and past experiences. These individuals could be working within different types of affiliation: the public sector, academia, or the private sector. The research thus defines that a *data for policy Expert* is an individual with an explicit involvement in data for policy as a field (thus he/she has been working in connection with this area); and with years-long experience (at least 5 years, which appears reasonable considering the usual terms limits of political mandates in most jurisdictions). These actors could be seen as relevant entry point for the phenomenon of data-centric innovation, since theory on innovation confirms the role of social networks in shaping technological innovation (Akrich et al., 2002). In line with Critical Realism, this does not imply that data for policy is a socially constructed phenomena, but that the discourse of these Experts might exert an influence on practices.

The Key Agents in data for policy practices

The division between "discourse" and "practice", which underlies the research design, posits that those who are involved in practices might not be part of the discourse. While interviewing experts is valuable, from a researcher point of view, because of their positionality over a phenomenon in general; the *key agents* are usually interviewed because of their agency and influence on a phenomenon in a specific context (Döringer, 2020). These types of actors could be found in literature on public sector innovation described as innovation *promoters and champions* (Bankins et al., 2017). They are usually identified with senior top/middle-level managers, which either advocate for innovations with the highest-tier decision-makers or struggle to realize them concretely within public organization (Borins, 2001). In line with this general typology, the research

85 In line with Bourdieu, this means that there is not yet an hegemonic perspective on what expertise count with this area of specialization (see Section 1.3.1).

considers *key agent in data for policy practices* to be public managers and civil servants that are actively involved in innovation related to the use of non-traditional data in public sector. These actors are therefore affiliated with the public sector and — while they might have knowledge on ICT, digitalization, and data — they are primarily knowledgeable on a policy or public service area, as they are concerned with bringing change in that context (Döringer, 2020). The positionality of this type of actors suggests they might be less interested/involved in a data for policy discourse in comparison to experts.

While the division of typologies of actors in the data for policy discourse seems important for the preparation of data collection (i.e., the interview format), in reality is imaginable that those categories are not mutually exclusive. Therefore, a *data for policy Expert* might also be a *key agent in data-centric policymaking*, whereas the opposite is possible but not always necessarily true (because of what explained above). What identify an expert — in comparison to a specialist — is the *coupling a long experience with consciousness* (Pfadenhauer, 2009, p. 82). In other words, the expert (compared to the key agent) possesses a comprehensive knowledge/overview about the nature of the problems he/she deals with. In *data for policy* an expert should indicate awareness of the field and position it in the broader panorama of public sector innovation.

Sampling of interviewees

The interviewees were sampled through slightly different strategies. The data for policy Experts were sampled in line with a sampling strategy called "reputational sampling" (Patton, 2014, p. 430), in which the sampling depends by the researcher's judgement on who is to be considered a knowledgeable figure. As expressed above, this might be a difficult judgement for the data for policy field, which is still emerging as an area of specialization. The strategy was to start from the "Data for Policy Conference" which was considered a relevant and explicit expression of the data for policy discourse (see Section 1.3.1). Profiles from universities (e.g., scholars) were privileged because assumed to have developed a higher level of reflection about data for policy as a field, also possible to verify by their writings. As already explained above, expert knowledge is characterized by the breadth of their perspective and by the comprehensiveness of their knowledge (i.e., understanding the deeper causes). Then, the interview design was meant to evaluate if the interviewees would meet several criteria that could be attributed to an expert: *the past experience* (i.e., how many years the interviewee state to have worked on this topic); *the affiliation* (e.g., academia, government, etc.); and the personal opinion on *data for policy* field. This last aspect was addressed through a specific question in order to understand the degree of awareness or critical sense the interviewee had. Therefore, the experts were initially contacted among researchers with a prominent role in the Data for Policy Conference in terms of publications.

For what concerns *Key Agents in data for policy practices*, the approach was mostly based on *snowball or chain sampling* (Patton, 2014, p. 450). This means that *key agents* were suggested by data for policy Experts, to whom the interview asked to suggest relevant cases. This strategy seemed logical because of the positionality of experts and their capacity to point at more exemplificatory cases. Not secondarily, this was also a way through which the research could start selecting cases for the phase 2, together with desk research (see Section 4.2.1). Some of the Key Agents interviews, whenever the person contacted was available and the case relevant, developed into one or more detailed interviews about the case (thus becoming a data source for case studies in phase 2).

For both types of interviewees, the sample size could be only hypothesized. On this topic, Beitin (2012) indicates that different sampling ranges could be adopted in qualitative interviewing. Authors propose ranges from six to twelve; from five to twenty-five and from two to ten (Beitin, 2012, p. 244). To decide when a sample becomes representative is a common issue in qualitative research, and many interview studies solve the problem by progressively analysing interviews until *saturation* (Beitin, 2012) — i.e., the sample is satisfactory once no more themes emerge from the data (Corbin & Strauss, 2008). This approach was attempted also in the sample of interviewees. However, it should be clear that the breadth of data for policy makes difficult to ensure representativeness. Past researches in this field that faced this problem (Poel et al., 2018) ultimately resorted to a convenience sample, considered sufficient for an initial exploration.

Interview Format and Data collection

On the base of two types of interviewees, two different types of interview formats were proposed (see Annex 2) differing on several aspects (Table 7).

Table 7. Synthesis of differences between Interviews Format proposed for phase 1

	Format 1	**Format 2**
Type of actors	**Data for policy Experts**	**Interview with key agents of data-driven policymaking**
Goals	To understand to what extent the these found in literature are relevant. To receive indications about relevant actions for comparative cases study (*snowball sampling*).	To understand how dynamics of data for policy landscape were translated into the design and realization level To receive indications about relevant practices that can account as data-centric policymaking cases in Phase 2
Sampling	Reputational (starting from the Data for Policy Conference, privileging scholars)	Snowball (on suggestion from Experts).
Structure	Semi-structured	Semi-structured
Data analysis	Open and Axial Coding with Atlas. ti software	Open and Axial Coding with Atlas.ti software
Contribution to research	Phase 1	Phase 1 and 2

Interview Format 1 is intended to access data for policy Experts' perspectives on to the evolving field of "*data for policy*" and understand its discourse. The interview Format 1 aims to collect data about:

- how interviewees understand and frame the field of *data for policy*
- how interviewees relate *data for policy* field to the landscape of innovation in government
- what interviewees consider the most relevant aspects to consider in the data for policy discourse (e.g., privacy regulation, government technological readiness, etc)
- what interviewees believe to be a likely evolution for this field and why
- what interviewees consider cases and key agents in data for policy

Interview Format 2 is intended to understand the perception of key Agents on the field starting from their involvement in practices. By referring to specific initiatives that used data these Key Agents were involved, the interview Format 2 aims to collect data about:

- the main goals, beneficiaries and expected outputs
- subjects involved and roles
- the time-frame

- the administrative scale
- the general policy area
- relation with existing policy schemes (e.g., on national level) and regulations
- the political support
- the level of institutionalization (e.g., policy, pilot, small scale experimental project)
- the possible development of artefacts (e.g., new services, information systems, etc)
- the main enabling technologies those actions relies on (e.g., specific softwares)

The interviews were of semi-structured type, that is the most widely used typology of qualitative interview (Brinkmann, 2013). In semi-structured interviews, the interviewer partly steers the conversation with specific inputs to access the interviewee's perspective. Format 1 was therefore developed around questions eliciting the expert's view and opinions. The interviewees received guidelines (shared beforehand) that explicit stated the importance of having an open and direct discussion, wherein personal opinions are valued. After providing an initial definition of *data for policy,* the initial questions (1, 2, 3) the Expert is asked to broadly express his mind about the field. Toward the end the questions aim to open for a more speculative orientation of the conversation. Finally, the Expert is asked specific indications about real-case examples.

Format 2 proposes a set of questions that led the interviewee to talk about the aspects about which I seek to collect data (see the point about Goals above in this section).

Data analysis of interviews

Coding is chosen as a technique to analyze the interview's outputs (i.e., verbatim transcripts). Coding is the act of dividing the data collected into codes:

"A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data." (Saldaña, 2013, p. 3)

Coding, in essence, "*labels segments of data with terms to summarize, categorize, and account for these segments*" (Charmaz et al., 2018, p. 740). Coding was chosen because it appeared to be a pragmatic technique to isolate recurrent topics and compare them throughout different data sources. This choice was also supported by the systematic review of qualitative interviewing literature in several studies that also employed coding to triangulate insights from interviews with document analysis and other written sources (e.g., notes from observations).

The analysis of interviews followed an open coding approach with the intention to integrate the broad topics emerged in the first analysis of the field (see Section 1.3.3.1), used themselves as coding, and the codes emerged from a focused literature analysis that constitute the supporting method in this phase.

3.3.2. Literature Analysis

The literature analysis employed in this phase attempted to isolate themes and topics relevant in the data for policy discourse starting from textual sources in data for policy. Secondly, it intended to connect what emerged with topics identified in the first narrative literature review of the field (see Section 1.3.3.1) and what emerged from the interviews. Some authors define this method as Qualitative Content Analysis or systematic literature analysis (Mayring, 2004), standing in between literature review and content analysis methodologies.

The method was adopted not to drive interpretivist analysis on semantics or structure embodied in the texts. Instead, coding was used to discover themes starting from a document generated by my sample group (Saldaña, 2013, pp. 175–176). It was intended as a explorative method and a preparation for the experts interview.

Sampling and data collection

The body of work selected for the literature analysis was sampled according to an emergent subgroup sampling (Patton, 2014, p. 253). This sampling is typical of qualitative inquiries, where the researcher might identify a specific subgroup throughout the research process, that is significant to understand the phenomenon under investigation. (Patton, 2014, p. 253). Since literature analysis was meant to support and integrate qualitative interviewing with data for policy Experts and key agents, the body of texts sampled was collected within the Proceedings of the "Data for Policy International Conference" (editions 2015, 2016, 2017, 2019), which were all the available proceedings at the time of the analysis.

Analysis strategy

The literature analysis was developed through an open coding approach (Saldaña, 2013). The scope, as for the qualitative interviews, was to find themes in the data in the most open exploratory way (Saldaña, 2013, pp. 175–176), and understand how these themes could integrate the topics identified in the previous part of the research (i.e., literature review, qualitative intereviews), basically performing a triangulation of collected qualitative sources (Flick, 2018).

3.4. Methodology of Phase 2

This second phase intended to shift the overall inquiry from the level of *discourse* surrounding data-centric policymaking — identified with the field of data for policy in the first phase — to the level of "*practices",* i.e., empirical realizations of the data-centric policymaking. In light of this focus the second phase adopted a case-based methodology (Byrne & Ragin, 2009) and the two main methods employed aimed at the sampling, construction and interpretation of cases. Desk research was initially used to screen public sector initiatives that seemed centred on the use of non-traditional data sources. Then, the method of Qualitative Comparative Analysis (QCA) intended to analyse by systematic comparison those of them that could account as cases of data-centric policymaking. Case-oriented methodology is relatively new in the history of social inquiry (Ragin & Becker, 1992). Through case studies the researcher aims at exploring contemporary and complex phenomena with deepness, by encapsulating its empirical instances into carefully constructed theoretical boundaries (Ragin, 1992). As said, the use of a case-based methodology in this research intended to balance the will to achieve qualitative insights with maintaining a reasonable degree of generalization. These two goals might be conflicting when using case studies, exactly because of the method implies great familiarly (or empirical intimacy) with cases, which achievement is highly time-consuming (Ragin, 2014, p. 50; Yin, 2018).

In exploring the emerging field of data for policy by using the concept of data-centric policymaking, this doctoral research intended to acknowledge the breadth of the phenomenon and the importance of its contextualities. Epistemologically and methodologically, this proposition demanded coherence in employing case studies in a middle ground between describing single instances and providing generalization.

3.4.1. Desk research

Desk research defines a method that systematically relies on secondary sources for research, which essentially are: "*reports of a phenomenon of interest by those who have not directly experienced the phenomenon*" (Merriam & Tisdell, 2016, p. 178). Desk research provides an uncostly way to explore international phenomena and develop meta-analysis, thus synthesizing new inquiry fields by relying on other's studies (van Thiel, 2014, pp. 102–117). Desk research often uses grey literature (i.e., not academic literature) as source (e.g., government and policy documents). These documents are deemed as an easy-accessible yet authoritative source for insight for policy-related research (Godin et al., 2015).

Sampling and Data collection

Desk research took place from the beginning of the research and continued throughout its later stages. While it initially intended to explore the field of data for policy through its practices (and understand how these could be related with policymaking), in the later stages it became a way to sample cases of data-centric policymaking for an in-depth analysis based on primary data.

The sampling strategy followed a stratified or nested purposeful sampling to isolate these cases (Patton, 2014, pp. 462–465) — in other words, a series of non-random sequential steps (Fig. 13).

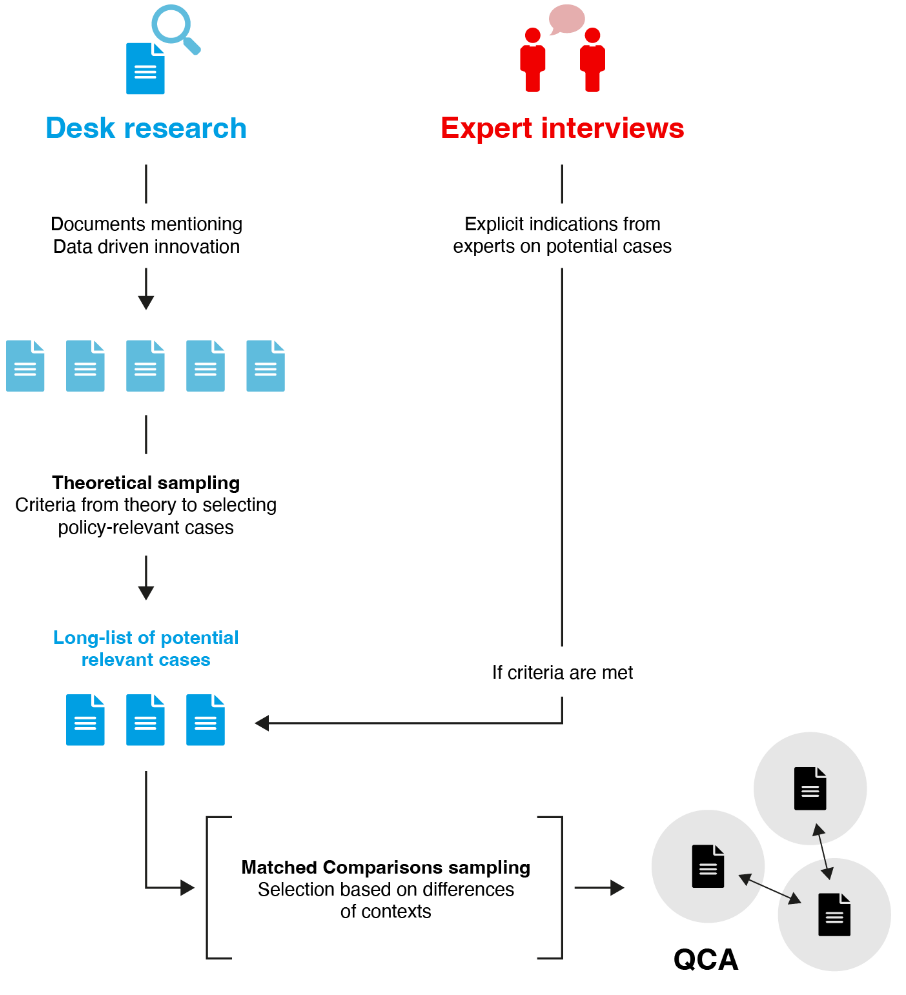


Fig 13. Sampling strategy for cases

The desk research collected and reviewed different sources of grey literature featuring “data-driven innovation” as keywork. The term was chosen as first way to gather material as it seemed widely used in policy documents (OECD, 2015) (see Introduction). The sources considered were prevalently reports and roadmaps by supra-national organizations (e.g., United Nations, World Bank, OECD, specific Directorates within the European Commission) and national governments on data-driven innovation and digitalization. Conveniently, these sources referred to examples in these fields (e.g., dedicated initiatives and programs) which were further investigated through integrative sources (e.g., other documents, websites).

This first screening was further refined through theoretical sampling. This type of sampling isolates empirical cases that represent concepts or constructs derived from theory (Patton, 2014, pp. 436-439). To apply this sampling intended to narrow down my list to only cases that could include policy-relevant aspects. The eligibility criteria thus aligned with the theoretical choices illustrated in Chapter 2, as well as considerations on non-traditional data in the Research Setting. In line with those, the eligibility criteria adopted were the followings:

- *Public sector/governmental organization as protagonist.*

All potential cases included featured an explicit and relevant involvement of at least one governmental body and agency, that should be traceable in documents. This choice was based in the theoretical view of governance that sees public policymaking as defined by governance settings with several stakeholders, but not in absence of government (Pierre & Peters, 2020). The perspective of policy work also aligns with this point. Not all relevant actors involved in policy work might be affiliated with government, but usually are hired or work in partnership with it (Kohoutek et al., 2013, p. 32). According to this sampling criterion it was possible to include ministerial and non-ministerial departments/offices, national agencies, public financial and non-financial corporations (e.g., national research centers), and local authorities. Instead, research experimentations initiated by universities or private sector initiatives were excluded.

- *The use of non-traditional data had to be traceable and be employed to address the substantive aspects of a policy problem.*

All potential cases included should indicate the use of one of administrative or non-traditional data sources identified as from the list in the Research Setting (see Section 1.3.2.2). Further, the cases had to suggest an attempt to use data for addressing a substantive issue, i.e., a specific policy problem. Public initiatives and actions centered on data that, instead, were more focused on the procedural aspects of policymaking (e.g., enhancing participation, transparency, etc...) or employed as part of public administration optimization (see Section 1.3.2.2) were not included. This sampling criterion therefore excluded from governments items such as: e.g., data frameworks, single ICT-based innovations, and e-services (chatbots, government web portals or web applications, GIS portals) (although these were included if monitoring was present) and digital solutions for governance (e.g., e-petition/e-participation platforms and portals or other digital solutions for public consultation and reporting).

Desk research data analysis

To derive some initial insights on the dataset that was emerging from desk research, various analytical lenses were tentatively applied to the list of potential cases individuated. The strategy implied to cluster subjects and activities through already existing categorizations. For example, the main subjects involved in potential cases were sorted through a typology of public sector organizations provided by the European System of National Accounts (Eurostat, 2021). Another tentative filter was applied to what appeared to be the nature of the substantive issues in the actions (i.e., the policy problems), divided according to the policy domain (May et al., 2006). Finally, the dataset was interpreted through the stages of policy cycle model connected with the perspective of policy work, as proposed by Wellstead and Stedman (2015, p. 57)[86] (Table 8). This did not imply the research would change its theoretical perspective, that was later adopted in the comparative analysis of primary data from selected cases. Instead, it meant to gain a first understanding of the potential cases at hand. Also, it was part of an attempt to understanding the limits of the policy cycle found in the literature of data for policy. Since the policy cycle identified stages which were too broadly defined, it was applied in conjunction with a list of policy work and tasks of policy work expected at each stage; in order to better interpret how the use of data in desk research cases would be applied in relation to policymaking.

86 The work of Wellstead and Stedman (2015) stemmed from the policy capacity framework (X. Wu et al., 2015), which carries the policy work perspective at the organizational level (Kohoutek et al., 2018). Policy capacities are defined as the combination of skills and resources necessary to perform policy functions (Wu et al., 2015). The policy capacity dimensions regard *analytical policy capacity*; related to information gathering and evidence-building; *operational policy capacity*; related to resource management for implementation and *political policy capacity*; related to seeking political support (Wu et al., 2018).

Table 8. List of policy work and tasks based on the policy cycle.

Policy Cycle	Policy work and tasks
1. Agenda-setting	- Identifying policy issues - Identifying policy options - Environmental scans - Consulting with public
2. Policy Formulation	- Appraising policy options - Collecting policy-related data - Collecting policy-related information - Conducting policy-related research - Negotiating with stakeholders - Preparing position papers
3. Decision Making	- Comparing policy options - Decision matrices - High-level briefing - Negotiating with central agencies - Department planning
4. Implementation	- Implementing/delivering - Consulting stakeholders - Legal analysis - Negotiating with program managers - Legal Analysis
5. Evaluation	- Policy evaluation skills - Risk-based tools/techniques - Evidence-based policy

3.4.2. Qualitative Comparative Analysis

In the expectation to approach a selection of highly complex cases sampled from desk research, while seeking for some degree of generalization, the research decided to employ a comparative method called Quantitative Comparative Analysis (QCA) (Berg-Schlosser et al., 2012). QCA is a well-established and highly diffuse method within the family of comparative methodologies, that bridges its qualitative and quantitative strands. It has been appreciated for its potential of generalizing knowledge across multiple cases while maintaining a rich and complex perspective on each single cases (Gerrits & Verweij, 2016). QCA works on the comparison of a set of cases by identifying the presence or absence of an outcome variable and a series of conditions variables that might have caused that outcome. Instead of looking for direct causality (i.e., one variable causes the effect), QCA supports the identification of different typologies of causal configuration (i.e., equifinality), by grouping together all cases which present sufficient and necessary conditions for the outcome condition. Conditions are conceptualized by researchers from theory and empirical engagement with the cases.

Introducing the rationale of QCA

The comparative method represents a cornerstone of social inquiry[87] (della Porta, 2008; Rihoux & Ragin, 2009). In social research, comparative methods and comparative analysis are often mistakenly equated: in fact, the comparative methods is just one approach — together with the experimental and statistical approaches — in the broader methodological family of comparative analysis (della Porta, 2008)[88]What most distinguishes comparative analysis approaches is the adoption of a case-oriented strategy, typical of the comparative method, in contrast to a variable-based strategy, typical of experimental and statistical approaches (della Porta, 2008). A variable-based strategy focuses on comparing variables among the maximum possible number of empirical units or observations (cases). In contrast, the case-oriented strategy would place cases (and their internal variables) at the center of comparison (C. Ragin, 1992). Further, a variable-based strategy often entails the collection of a large or medium number of cases (commonly refer to as large-N or intermediate-N sample) and a quantitative approach, whilst case-oriented strategy defines studies with a small number of instances, each possessing a lot of internal variables (Lijphart, 1971).

The case-oriented strategy emerged to address the need to balance a nomothetic stance of knowledge production (i.e., driven by the need to generalize knowledge from single instances) and an ideographic one (i.e., driven by the need to describe the uniqueness of single instances) (Ragin,

87 As a formalised form of scientific inquiry, the comparative methods date back to John Stuart Mill's book "A System of Logic" (1891). Mill advanced that a methodology based on systematic comparison could be placed at the same level as the experimental method. According to Mill, these methods essentially share the "same logical design" and epistemological perspective toward discovering hidden causal regularities (Moses & Knutsen, 2019). For example, in randomised controlled trials (RCTs), an epitomic experimental method, relies on comparing two groups, wherein the researcher observes the relationship between independent/dependent variables as an intervention is applied.

88 Given that comparison remains a common thread among all three approaches, they differ in many other regards. One difference is about data collection. In the experimental approach, researchers typically create their data during experiments. Instead, the statistical and comparative methods commonly employ already existing data (e.g., national statistics), which are then linked to variables and theoretical constructs to infer new knowledge (e.g., socio-economic structures. This type of operation is usual in comparative politics, where empirical research requires a continuous "back and forth" between the conceptualisation of variables and their quantification (cf. Lijphart, 1971; Sartori, 1970).

2008). The case-oriented strategy therevy intertwines comparative analysis with the case study method[89] (Gerring, 2004, p. 351; C. Ragin, 1992; Rihoux et al., 2011). A case is a bounded system (Yin, 2018), representing the studied phenomenon, which possesses a defined place and time (C. Ragin, 1992). Empirically, a case is "*an instance, incident, or unit of something and can be anything—a person, an organization, an event, a decision, an action, a location like a neighborhood, or a nation-state*." (Schwandt & Gates, 2018, p. 600). Hence, cases are not independent existing entities: a researcher must construct the case through the process of casing (Ragin, 1992). Quantitative Comparative Analysis (QCA) was proposed by sociologist Charles Ragin in 1987 as a method to address the tension between variable-oriented and case-oriented approaches. By proposing QCA, Ragin intended to transcend the traditional qualitative/quantitative divide of social science (Ragin, [1987] 2014). The goal of QCA therefore is:

> "[...] to allow systematic cross-case comparisons, while at the same time giving justice to within-case complexity, particularly in small- and intermediate-N research designs." (Rihoux & Ragin, 2009, p. xviii)

General use of QCA

Many articles are explicitly dedicated to explain QCA and its protocol[90] (Gerrits & Verweij, 2016; Rihoux et al., 2011; Rubinson et al., 2019; Schneider & Wagemann, 2009, 2010; Thomann & Maggetti, 2020). QCA is usually intended either as "an approach" to case construction or as "an umbrella-term" for the three data analysis technique (see below)

89 The use of multiple case studies as research strategy resurfaced from the 1990s onward in social research and has been particularly appreciated in policy-related research since these studies often deal with a limited number of observed instances (for example, a study on nation states would include tens of cases rather than hundreds) (Rihoux et al., 2011). Before that (' 60s-'70s), intermediate-N or large-N quantitative studies that focused on macro-level dependent variables (e.g., socio-economic factors), adopting a theory-building perspective across various nations, prevailed in the field (della Porta, 2008). During that period, specific fields as comparative politics started to emphasise the importance of case studies and their potential when used in combination with comparative analysis (Lijphart, 1971). The quantitative and qualitative fracture persists to the present days in public policies' studies, that still divide between the original macro-quantitative tradition and the more recent qualitative analysis of in-depth cases (Knill & Tosun, 2012, p. 288).

90 In the following points, the QCA protocol is summarised drawing from these sources and in particular from Gerrits and Verweij (2016). The summary referes to a recent study from Ruhlandt et al. (2020b), who used QCA to investigate the adoption of data analytics in cities.

(Berg-Schlosser et al., 2012; Rubinson et al., 2019). Rihoux and Ragin inscribe QCA into configurational comparative methods — i.e., a method that approaches case studies as a “complex configuration of properties” (Rihoux & Ragin, 2009, p. 6). Instead, Schneider and Wagemann (2009) consider QCA as a set-theoretic method that interprets social reality in terms of membership to a set of data:

> “[...] the data on which set-theoretic methods operate are membership scores of cases in sets which represent social science concepts. For instance, France is an element of the set of European Countries whereas the USA is not. France’s set membership score in this set is therefore 1, while that of the USA is 0.” (Schneider & Wagemann, 2009, p. 3)

In essence, QCA as an approach implies that researchers see cases as the configuration of multiple variables (i.e., configurational perspective). The researcher identifies cases by inscribing them under these variables — or sets — with various membership degrees (i.e., adopting a set-theoretic perspective). These variables are defined as conditions and outcome variables[91].

A researcher who seeks to employ QCA is encouraged to derive variables from theory, which has a decisive role in the process (Berg-Schlosser et al., 2012), and from substantive knowledge on the topic supported by empirical research (Gerrits & Verweij, 2016). These variables are then calibrated through three techniques: crisp-set QCA (csQCA), multi-value QCA (mvQCA) and fuzzy-set QCA (fsQCA). “Calibration” refers to different choices of values that define the membership thresholds of a variable. While csQCA allows a variable to be set either as present (1) or absent (0), mvQCA and fsQCA consider a more nuanced set of anchor points: values that define the membership thresholds (see below for calibration in this research).

91 For example, Ruhlandt et al. (2020b) identify as their outcome variables “the level of utilization of data and analytics within the city” (Ruhlandt et al., 2020b, p. 5), which also represents the phenomenon studied across the different cases. In a previous study, they defined 11 different conditions variables (Ruhlandt et al., 2020a) they use to compare all cases.

Data Matrix

CASES	VAR A	VAR B	VAR C	OUTPUT
X	1	0	1	1
Y	0	0	1	0
Z	1	1	1	0

Truth table row

VAR A	VAR B	VAR C	OUTPUT	CASES
1	0	1	1	X, Y, Z

Fig 14. Data Matrix and truth table row

Once the calibration thresholds are defined, the researcher interprets each case as a configuration of these variables, creating a data matrix that compares all variables for all cases. At this point, the researcher utilizes QCA software[92] to produce the so-called truth table, that shows all the possible figurations in the data matrix in relation to an output, finally proceeding with logical minimization (see Fig. 14).

Use of QCA in this research

QCA is usually regard as optimal method to tackle causal mechanisms in terms of sufficiency and necessity (Thomann & Maggetti, 2020). For this purpose, in the later phases of QCA procedure, algorithmic analysis is applied to the data matrix to derive the truth table. From truth table the necessary and sufficient conditions generated the output are shown in relation to cases, and expressed as parsimonious explanations in logical algebra (a process called logical minimization) (Rubinson et al., 2019). These

92 Several of these are listed on the site of COMPASSS (COMPArative Methods for Systematic cross-caSe analySis), a network of scholars dedicated to set-theoretic methods: https://compasss.org/software/.

explanations are normally used to prove or disprove preliminary hypotheses and typologies. Therefore, the most optimal use of QCA would depend by a theoretical-based understanding of the phenomenon expressed in terms of sufficiency and necessity (Thomann & Maggetti, 2020).

In the context of this research, hypotheses of this type could not be made, as the field of data for policy appeared still theoretically under-developed. Rather, some conjectures based on literature analysis and the specific interests of this research emerged; which proposed a way to explore data for policy, rather than attempting to hypothesize about its dynamics as a phenomenon. This implies that QCA was used in this research as an exploratory approach and for inductive reasoning toward the data collected in cases — i.e., to synthetize and summarize the data about the cases.

While this use of QCA is perfectly legitimized by literature (Berg-Schlosser et al., 2012, p. 15), admittedly it only partly exploit the potential that the method could offer. The choice was also driven by much necessary precaution, given the preliminary knowledge of the author and the experimental nature of this method in data for policy. Therefore, the use of QCA in this work intended to exploit the method not much as an analytical tool for parsimonious explanations, but as an incentive both for rigorous data collection and an iterative back-and-forth from assumptions to cases (Pagliarin et al., 2022; Rihoux & Lobe, 2009; Thomann & Maggetti, 2020).

Defining QCA Conditions variables

To compare the cases of data-centric policymaking through QCA, the research proposes a series of variables that could describe the cases, dividing them between the macro-level and the meso/micro-level. These variables, in line with the theoretical-conceptual framework, describe the cases in relation to policy learning (which is considered the output), while — as pointed out in the previous section — they do not imply any preliminary causal hypotheses. The macro-level conditions were seen as structural elements of data-centric policymaking, thus concerning the enabling condition of using non-traditional data in the different public sector contexts (see Section 1.3.3.2). The meso/micro-level conditions, on the other hand, were advanced as pertaining the level of social group involved in data-centric policymaking. This choice was based on the interest on data-centric policymaking as practice (in line with hypotheses, see Section 1.5), and by the focus on policy learning (see next section).

The cases' macro conditions for QCA were derived from two existing national-level indicators developed by the OECD; the *OURdata Index* (ODI) (OECD, 2020b) and the *Digital Government Index* (DGI) (OECD, 2020a). The ODI was first piloted in 2015, then published officially in 2017 and 2019, as an national-based indicator of progresses in digital innovation with a focus on government data re-use assessed on three main dimensions (Lafortune & Ubaldi, 2018, p. 5). The DGI, on the other hand was piloted more recent (2019), to evaluate how governments were progressing toward digital government (Ubaldi & Okubo, 2020). The DGI is based on a composite score of six dimensions considered to characterize the ideal digital government. Data used in ODI and DGI are collected through survey administrated to high-level public officials in OECD countries governments and other qualitative sources (Lafortune & Ubaldi, 2018; Ubaldi & Okubo, 2020). These indicators are both proposed as the first to focus on how governments make data and digital government central elements for public value creation and innovation (Lafortune & Ubaldi, 2018). The three macro-level structural conditions included this research QCA were respectively based on the second and third dimensions of ODI[93], while the third was based on the data-driven government dimension of DGI[94] In order these were:

Data Accessibility (ACC)
The extent of provision of government data and associated metadata in open and re-usable formats within a national government.

Data Culture (CULT)
The extent to which a national context promotes the re-use of government data inside and outside the public sector.

Data Governance (GOV)
The degree of presence of legal frameworks, specific regulations and responsible roles/organizations for government data sharing and re-use.

Moving to the meso/micro-level conditions, these were identified at the level of the social group involved in the cases of data-centric policymaking. To identify these conditions, the research considered the factors usually associated with policy learning in networks (i.e., learning as result of social interaction at the micro-level) (more below) (Riche et al., 2021).

This literature could not provide a definitive list of factors causing policy learning but advanced several individual and intra-individual char-

93 Data availability, the first dimension of ODI, describes the government progresses in the adoption of formal requirements to promote open government data seemed out of scope as too much focused on open data.

94 The DGI's dimension called data-driven public sector; examines "the extent to which governance, management and use of data informs and approaches the design, delivery and monitoring of public policies and services" mainly by considering the presence of rules and institutional role for data management in public sector (OECD, 2020a, p. 30).

acteristics of the social group that influence it[95]. Starting from these characteristics, and on the base of practical considerations about which conditions would be feasible to measure, the following meso/micro-conditions were advanced:

Political Support (POL)
Duration and quantity of social interactions are correlated to policy learning (Resh et al., 2014; Riche et al., 2021). In data-centric policymaking it is imaginable that the prolongated interaction of policy workers over non-traditional data depend by the *political support* to data-centric policymaking, in form of endorsement, dedicated budget or personnel time.

Leadership (LEAD)
The presence of leaders or respected mediators, who can overview the process and mediate across actors, is suggested as an important elements for policy learning network which features several types of actors (McFadgen & Huitema, 2017; Resh et al., 2014; Riche et al., 2021).

Experience (EXP)
Policy learning depends by the previous competences and types of knowledge which are brought by policy actors in a learning network. The presence of knowledgeable figures, from which the others actors can learn is supposed to influence learning.

Diversity (DIV)
The diversity of the profiles and background of the learning group is seen as influential factor for learning. Similarity of these views (homophily) can facilitate learning (Riche et al., 2021) but can also hinder it (Resh et al., 2014).

Conditions POL, LEAD, EXP and DIV were based on insights and data internal to the cases.

These came from interviews with one or more key informants involved in each case. A first contact with key informants resulted from suggestions

95 General the conditions of policy learning in networks of policy actors the literature indicates: the presence of experts or knowledge holders with good reputation; the level of trust; the balance between dynamics of cooperation and competition; the presence of clear leadership over the process; the presence of a respected mediator/knowledge broker; the time and frequency of social interactions during the process; the size of the learning group; the dynamics of expressing preferences (voting systems or consensus-driven); the level of homophily (i.e., degree of people sharing the same vision) and the balance between informal/formal exchange (Ishii & Okubo, 2014; Resh et al., 2014; Riche et al., 2021).

of Experts in Phase 1 (see Section 3.2.1); or by the identification of potentially relevant cases through desk research and attempt to contact the case directly. Key Informants were asked about the case in line with the interview structure imagined in Phase 2. They were also asked to provide integrative information sources[96] on cases in the form of documents (e.g., reports, internal management documents, etc.). Moreover, the key informants further supported the data collection on the case study and on policy learning through a survey (see next section). The data collected through the survey were used to calibrate the final outcome condition — individual policy learning (LEA) — which was also considered at the meso/micro level (Dunlop & Radaelli, 2013; Moyson et al., 2017) (see next sections).

Survey design on policy learning 1: characteristics of sample

QCA can be applied to micro-level cases that identify "individuals who possess a certain set of characteristics relevant for a given research, provide an extensive amount of primary information, gathered through multiple sources, qualitative and quantitative" (Rihoux & Lobe, 2009, p. 227).

The unit of analysis for QCA in this work were the cases of data-centric policymaking (see Section 1.5). Data about these cases were collected both externally to cases and internally to them from individual policy workers involved. The conceptual/theoretical framework provided in Section 2.8 intended to instruct the theoretical view for the analysis so to read data-centric policymaking in relation to policymaking (see Section 2.8).

In line with the theoretical perspective, data-centric policymaking consists of the practical activity of policy workers involved in a data ecosystem (Oliveira & Lóscio, 2018; Parsons et al., 2011) (see Section 1.4.3). For this research, it was interesting to explore data for policy through the practices generated around non-traditional data, as these practices are suggested to develop through an experimental and explorative fashion (Concilio & Pucci, 2021) to require the involvement of different types of actors, with different competences and knowledge (Giest, 2018).

Through these notions from literature, and from an initial engagement with cases, it was decided to organize the data collection within the cases through a survey that collected data on both on the groups involved in the data-centric policymaking case and on their individual learning.

These two aspects were reflected in two sections of the survey. The first part collected data by dividing the sample between different types of actors, depending from the role they bring to the practices of data-centric

96 In one of the cases eventually individuated it was also possible to integrate the date with direct online observation of project meetings and an English translation of minutes.

policymaking in terms of knowledge/expertise (which seemed sensible given the focus of the research on policy work and learning). In line with that, three types of actors could be imagined to be part of the cases of data-centric policymaking, based on their expertise and knowledge:

Data practitioners
These were considered all the actors with a more technical profile that contributed to data-centric policymaking with their direct work on non-traditional data. The exemplary workers in this sense would be the data scientist, but it could regard also other type of workers that added digital and technical competences (e.g., general ICT, UX experts).

Actors from policy subsystem
Side to the data practice in data-centric policymaking it was sensible to imagine another group of actors: public managers, civil servants, government pundits and consultants. The would be those actors who bring knowledge and expertise on a policy problem and its specific area, i.e., the policy subsystem (Sabatier & Jenkins-Smith, 1993).

Project managers
Project managers are considered to have a mediating role between the other two categories in the data-centric policymaking process. Project managers can supervise the data-centric policymaking process, thus they have to enter in the work of data practicioners, but also remain in contact with the policy subsystem and the broader political space around the data-centric policymaking process.

This division operates a simplification that assumes a separation of roles based on knowledge and positionality in respect to the policy problem addressed. Accordingly, a data scientist would have a much direct involvement and agency in data practices but will possess limited knowledge of the policy system these data describe. On the contrary, a civil servant involved in data-centric policymaking would well understand the dynamics of a policy subsystem but will not be involved in direct work with data. These categories of actors were reflected into the data collection strategy for cases, in order to understand the sample and their role. Moreover, the division attempted to understand how different workers were engaged with data, considering the overall activity of data-centric policymaking as a process of data science (Crisan et al., 2021) (see Fig. 15).

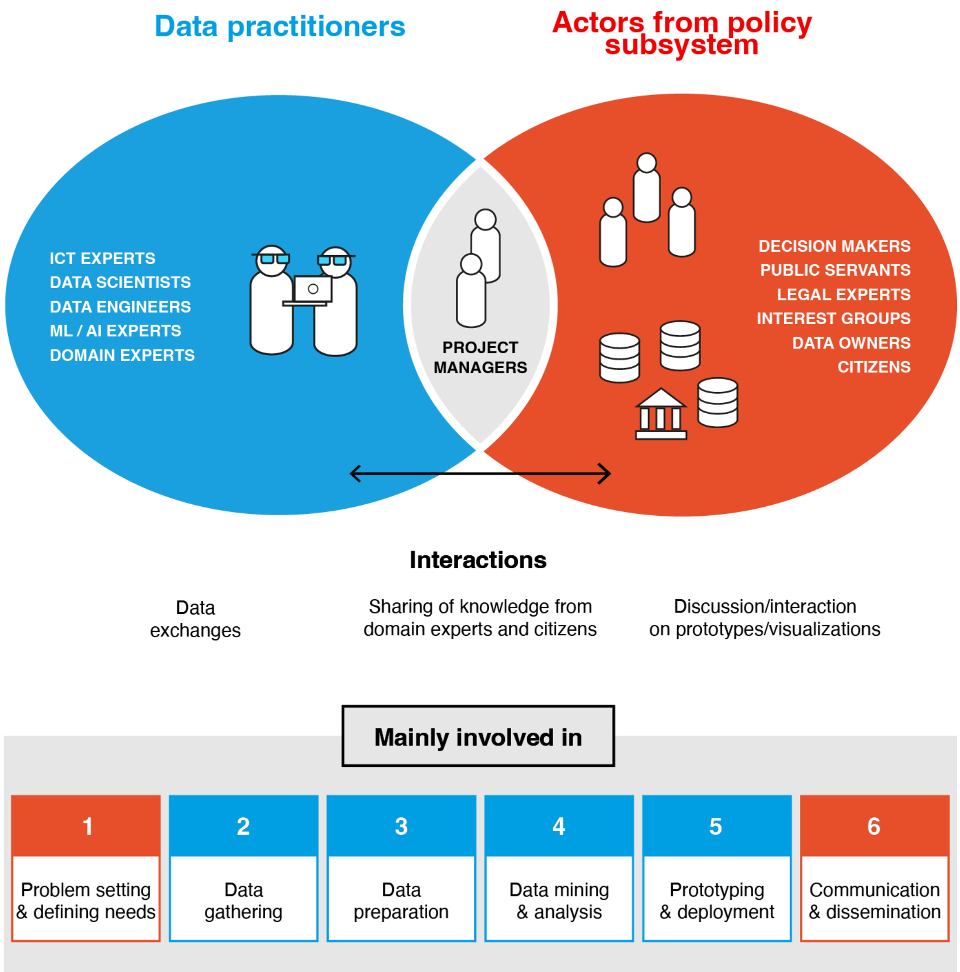


Fig 15. Scheme of data-centric policymaking actors and roles

The second part of the survey instead aimed at collecting data on self-reported learning. These data were then used to calibrate the output variable in QCA (see below).

Survey design on policy learning 2: self-reported policy learning

To collect data on characteristics of sample and on individual policy learning within the cases, the research opted for a direct measurement of policy learning through a survey distributed at the individual level. The survey was forwarded to persons involved in the case with the support of the key informants. After being presented with the intentions and goals of the study, the key informants were asked to individuate the relevant actors involved in the case to which the survey could be sent. The data collection took few months and was considered completed when more than half of the potential respondents compiled.

The surevey intended to obtain a self-assessment of individual policy learning — i.e, the individual perception of learning obtained during the process of data-centric policymaking.

This methodological choice aligns with a relevant part of the current research that study policy learning as social learning[97]. In line with the definition of policy learning (i.e., an updated of knowledge and beliefs) (Dunlop & Radaelli, 2013), the direct approach to policy learning research concentrates on cognitive and normative updates (Squevin et al., 2021, p. 155). It assumes a conception of policy learning as result of social learning at the micro-level, developed through acquisition of new information and social interactions and social interaction (Dunlop & Radaelli, 2017). Within a multi-level theoretical perspective, this dynamics of learning is defined as micro-foundations (Dunlop & Radaelli, 2017) and it assumes learning is individual (i.e., people have the cognitive capacities to learn, not institutions). In essence, the micro-level studies on learning "*focus on policy-making as a process of 'puzzling' among individual policy actors dealing with ideas and uncertainty*" (Moyson et al., 2017, p. 162). Learning happening at the micro-level might then be aggregated and institutionalized; so to reach the levels of public organizations and or society at large (Dunlop & Radaelli, 2017).

In applying direct approach to measuring learning through cognitive and normative outcomes, researches on policy learning usually adopts a differentiation between types of learning and rely on self-reported learning (McFadgen & Huitema, 2017). This clearly implies that data collected should be treated not as objective (e.g., as would be in a cognitive test), but as an indication on the individual perception. In this work, data collection on learning adopted two main typologies, following other similar attempts (McFadgen & Huitema, 2017):

Cognitive Learning

This type of learning concerns the acquisition of new information or factual knowledge on policy-relevant topics. Cognitive learning is deemed to affect the more instrumental aspects of policymaking, as the type of policy solutions and tools used to address a policy problem (May, 1992). The update of cognitive knowledge thus prevalently regards acquisition of new information.

97 E.g., in a range between threshold "0" and "0.33" a case with scoring below "0.165" with their index normalized would be given "0" for that variable; or "0.33" if above "0.165".

Normative Learning

Several theories of learning advance normative learning as the main effect of social learning and the main driver behind large scale policy change (Hall, 1993; Sabatier & Jenkins-Smith, 1993). Normative learning pertains a change of beliefs/preferences on policy-relevant topics. It is suggested that a learner goes through this type of learning once he/she updates his/her existing mental/social frames about a given policy problem, the solution adopted and its underlying paradigm. This implies to challenge existing institutional and personal frames (P6, 2018).

The individual perceptions of participants in the cases of data-centric policymaking about these two types of learning were measured through the second section of the survey, by relating each type of learning to the policy problem that each case was dealing with (see Annex 4 for the structure).

Calibration of variables

After all conditions and outcome variables were defined and data was collected, the QCA procedure required a procedure called calibration. As a set-theoretic method, QCA describes cases according to *set membership.* Each of the variables identified (both conditions and outcome) constitute a *set*, i.e., a concept that either describes the case or not (C. Q. Schneider & Wagemann, 2009, p. 24).

The set membership for each variable is thereby proposed by the researcher through the calibration procedure, through which the researcher aims to establish thresholds that define the possible degree membership. Depending by the QCA technique employed (see secton above) the number of thresholds can range from “0” (indicating non-membership) to “1” (indicating full-membership). Several practical indications and techniques exist in literature on how to perform calibration according to the best standards. Generally, it is suggested that calibration should be determined by the researcher substantive knowledge of the case and theoretical assumption on the relation between conditions and outcome, therefore calibration highly depends by each research (Berg-Schlosser & Meur, 2012). The researcher is encouraged to reflect thoroughly about the meaning of each variable and what it means for a case to be fully a member of that set or not (Rubinson et al., 2019). Whenever possible, calibration should be based on data that are external to the pool of cases considered (C. Q. Schneider & Wagemann, 2009), which was done for structural enabling conditions of cases (see below). Calibration could also be based on qualitative data collected within the cases, as was done for the other variable conditions. This work employed direct calibration through a fuzzy-set 4-points threshold (0; 0.33; 0.67; 1) for each of the variable used, as this scale seemed adaptable both to quantitative and qualitative data and widely adopted (Basurto & Speer, 2012). For the calibration of the macro-level structural variables (GOV, ACC, CULT), the score of each case national contexts was considered from the indexes used as referral. The score for each nation in the index was normalized on a "0-1" scale (with 0 being the lowest country in the ranking and the 1 the highest). The score for each variable in each case country was also normalized in the same way and approximated to the closest threshold[98]. The process was not mechanical but critical, based on the index but also on the insights from key agents interviewed. Regarding micro/meso variables these, as said, were based on insights from cases:

The calibration of Political Support (POL) was based on the reported presence or absence of explicit support to the data-centric policymaking cases, which was explicitly asked to the main actors involved. Support implied that the case received dedicated budget, assigned staff or explicit endorsement. This variable was either present or absent.

The calibration of leadership (LEAD) depended on the presence of actors or organization which clearly were regarded as leading figures within the cases and carried the other actors alongside in the process. These figures had also a central role in mediating relationships of policy workers over the use of non-traditional data. This variable was either present or absent.

The calibration of Experience (EXP) was based on years of working experience in government reported through the survey by respondents. The percentage of respondents that had "from 5 to 10 years" and "more than 10 years" of working experience on the total of sample was used to assign the threshold.

The calibration of Diversity (DIV) was based on the number of affiliations reported by respondents in the survey. Since this variable might be quite relative, the case with highest number of discernible affiliations was used as maximum score for the variable (value of 1), and the other cases were defined accordingly.

Finally, **the calibration of learning (LEAR)** was based on the number of possible statements that could be expressed by each group on the eight questions concerning cognitive or normative learning about the policy problem and related policy tools, public services and actors (see Annex 3 section 2). For example, in a case with 10 individuals the total of potential statements that could be expressed for these questions would be 80. The highest number of preferences for each grade of the Likert scale was used to assign the score for the learning variable in that case.

98 E.g., in a range between threshold "0" and "0.33" a case with scoring below "0.165" with their index normalized would be given "0" for that variable; or "0.33" if above "0.165".

Table 9. QCA Conditions and calibration

Code	Name	Set description Full membership (1) in this set indicates...	Data Source for assigning membership	Fuzzy-set value definition
Macro level variables – structural enabling conditions				
GOV	Data Governance	National ecosystem with a mature data governance framework for value-creation from public sector data (e.g., data policy, data governance frameworks, responsible public bodies).	Ubaldi & Okubo, 2020	0 = data governance is lowly developed; 0.33 = data governance is below average; 0.67 = data governance is above average; 1 = data governance is very highly developed;
ACC	Data accessibility	National ecosystem where public data are widely accessible in open and re-usable formats.	Lafortune & Ubaldi, 2018	0 = data accessibility is very low; 0.33 = data accessibility is below average; 0.67 = data accessibility is above average; 1= data accessibility is very high;
CULT	Data Culture	National ecosystem with high data literacy and where the re-use of data inside and outside government is incentivized.	Lafortune & Ubaldi, 2018	0 = data culture is very low; 0.33 = data culture is below average; 0.67 = data culture is above average; 1= data culture is very high;"
Meso/Micro-level variables – group level conditions				
POL	Political Support	The case had dedicated budget and resources / is supported by institutions. Therefore, the process has been carried out for long time.	Insights from key informants	0 = absent political support (no dedicated budget, short time) 1 = high political support (dedicate budget or staff, long time)
LEAD	Leadership	Presence of leading figures with strong commitment and clear vision on data and policy problems. These leaders supervise the project and mediated across actors.	Insights from key informants	0 = absent leadership 1 = presence of leadership

EXP	Experience	The case had an high percentage of actors with a long experience of working in government on the total of individuals involved.	Survey	0 = case with zero or few actors with past working experience in government 0.33 = more low than high presence of actors with past working experience in government; 0.67 = more high than low presence of actors with past working experience in government; 1 = high level of past working experience in government
DIV	Diversity	The case had a high number of organizations involved with respect to the sample	Survey	0 = case with the minimum number of organizations involved with respect to the sample; 0.33 = case with low number of organizations involved with respect to the sample; 0.67 = case with high number of organizations involved with respect to the sample; 1 = case with the maximum number of organizations involved with respect to the sample;
Dependent/outcome variable – self-reported individual policy learning				
LEAR	Learning	The majority of respondents “Strongly Agree” on having learned (gained new information, changed beliefs) by involvement in the case	Survey	0 = the majority of respondent “Strongly Disagree” on having learned (gained new information, changed beliefs) by involvement in the case; 0.33 = the majority of respondent “Disagree” on having learned (gained new information, changed beliefs) by involvement in the case; 0.67 = the majority of respondent “Agree” on having learned (gained new information, changed beliefs) by involvement in the case; 1 = rhe majority of respondents “Strongly Agree” on having learned (gained new information, changed beliefs) by involvement in the case

Chap. 4. Results and Discussion

In this chapter, the data and insights of the research are presented and discussed. The multiple analysis conducted throughout the research intended to explore the field of data for policy and the concept of data-centric policymaking through primary and secondary data. Retrospectively, it could be said that several had been tentative analysis (in particular, phase 1 and desk research of phase 2) to define a clear aspect to inquiry with some deepness.

Therefore, the process that will be described in this chapter never was a straightforward and direct movement from previous knowledge to the empirical world, but a dialogical interaction between the area investigated, divided into two ideal dimensions by the Research Design. As more data was collected, it interplayed with the knowledge from literature and the methodological approach was refined accordingly.

Phase 1, as said, was largely exploratory, fundamentally starting with a conjecture on the independency of the field of data for policy (see Section 1.3.1), to be then explored through a dialogue with the Experts and Key Agents involved in the field. Similarly, phase 2 required a casing process (Ragin, 1992) made of back-and-forth between cases and the theoretical/conceptual framework. Opportunistic sampling certainly played a major role, although potential cases were always checked against the framework of data-centric policymaking proposed. Concrete opportunities to perform first-hand data collection relied on direct and personal connection; either granted by connection through the Experts or by personal initiative.

For this reason, it was useful to resort to the "Data for Policy Conference", both as a venue for engagement and a concrete reference for explaining the field of data for policy. The results showcased in chapter 4 describe this complex exploration, moving from the discourse in data for policy to the practices of data-centric policymaking.

4.1. Results of phase 1

4.1.1. Expert interviews findings

A total of thirteen interviews were conducted from December 2020 to March 2021 with profiles considered Experts. In line with the sampling decided for this phase, the profiles were contacted primarily on the base of their involvement with the "Data for Policy Conference". Five interviewees held a position in this venue as conference chairs or authors. The other interviewees were suggested by the experts or autonomously individuated (e.g., authors of widely cited papers).

Among the thirteen interviews conducted, nine ultimately were considered relevant for the final scope of this Phase 1 and analyzed. Those excluded were mostly with public managers and practitioners, who provided useful information on cases within the public sector but did not seem to possess a relevant awareness about data for policy as a field (see Section 3.2.1.). Their insights therefore contributed to the inclusion of cases to the desk research phase.

The sample featured a good diversity of perspectives. Six of the nine profiles were academics or independent researchers (e.g., from think tanks). The other three profiles worked in national governments (ministries, public agencies) or international governmental organizations. All the profiles except one operated in Europe. Their expertise regarded the broad area of digital technologies in government, with various degree of variation depending on the individual topics of interest (see Table 10).

The interview developed around the semi-structure interview format (see Annex 2), but divergences and discussions were welcomed (Bogner & Menz, 2009). The author developed the interviews by purposefully showing interest and awareness on the topic addressed, although any opinion was presented as personal. Moreover, the interviewees were reminded they were contacted in quality of Experts and encouraged to freely share their views on the subject.

All the interviews were conducted on-line with the video conferencing software Zoom and lasted from 45 minutes to 1 hour. An introduction of the scope of the research and the interview format was sent to the interviewees beforehand. In the final format (Annex 2) a definition of data for policy as a field by the researcher was provided as a starting point for the discussion.

Table 10. Expertise and topics of interest of Experts interviewed

Code	Expertise
INT-1	Data-driven public sector, Public Services, Co-creation, Social experiments with data, AI and Data Ethics
INT-4	Digital twins, Urban modelling, Urban Planning
INT-5	Data-driven public sector, Open Data, E-gov, Digital Government
INT-7	Data and policymaking, Co-creation, E-gov, Digital Government, Data ethics, AI
INT-8	Privacy, Data ethics, Digital twin, Data Visualization
INT-9	Data and policymaking, E-gov, Digital Government, ICT for social development, AI
INT-11	Data and policymaking, Political Science, Data Ethics
INT-12	Data and policymaking, Remote sensing, Parliamentary advising
INT-13	Public information systems, Evidence-based policymaking, Environmental data

The interviews were analyzed with open coding to individuate commonalities (Saldaña, 2013, p. 213). The first list of codes was then confronted with the topics individuated in literature review (see Section 1.3.3.1). Further code grouping was operated if needed. Several codes could be connected to this previous list. Interviewees were asked specifically an opinion about the field of data for policy, that therefore emerged as an independent set of codes. The Themes emerging from these interviews (each based on a codes group) are presented here as findings.

Theme 1 – The perception of data for policy as field

Table 11. Expert interviews' excerpts for Theme 1

INT-1	I think that the way that I see data for policy is when there is understanding of an act commitment to data as a resource in Policymaking. And when policymakers are understanding the alignment from accumulation of data processing of data and decision making.
INT-4	I usually struggle with this anyway and so to be honest, data for policy sounds like the supplier side of evidence-based policies [...] I'm not sure to what extent it is actually a new thing or is it just referring to a different way, which is very much in line with like ICTs, of producing evidence for policy.
INT-5	So, I think policy as a broad field, where it assumes a disconnect, or mostly it assumes that you are set apart, and you are looking at the problem, and you are developing the answer to that problem. And therefore you must bring data into that conversation. There's an element to which that is a that is an unhelpful model, because it assumes that you are set apart, but actually, policy turns into delivery, and it turns into an ongoing impact on people's lives. And so it's very artificial to say that I can just use data over here to develop the policy without worrying about everything that happened. [...] there's no alternative other than for data to become fundamental to the practice of the public sector. In not just policymaking, but in delivery and in operating government.
INT-11	I think oftentimes, when we define data for policy, it sounds like a new thing. So it sounds like it's something sort of innovative or ground-breaking. But I think it builds on layers of previous research around evidence based policymaking, obviously [...] But I do think that there's something distinctly different about the data policy movement, that's sort of perpetuated by new technical possibilities, and also the awareness around what's possible with data. So what we can learn from data, and what we can do if we start linking data sets, or if we start combining, for example, micro data with sort of national statistics. So I think the awareness changed, but I think the data that's being used, especially administrative data is not new, in that sense. It's just new technical opportunities to utilize it in new ways.
INT-13	I don't think we can talk in a consistent way about data for policy across different domains, is so massively heterogenous across different domains [...]] there's probably a lot more use of data for programs, than for policy. So, policy is the bigger lever, and in programs to how you kind of, you know, move the car forward, if you will. So I'd say, you know, quite often what, what maybe is badged as and bundled under data for policies is actually data for programs.

The definition of data for policy provided to the interviewees (see Annex 2) proposed it as an independent field among others investigating digital technology in governments. The definition proposed was intentionally straight-forward for generating opinions and dialogue. The notion that the field could be regarded as independent and totally innovative was received with some skepticism, with only two Experts agreeing on that note. Six interviewees would instead advance that the field presents only partial elements of novelty, depending on the perspective adopted. Two interviewees explicitly referred to the evidence-based policymaking movement as antecedent of data for policy, i.e., advancing that, in a policy perspective, data for policy could not be considered a new thing. For these Experts, the data for policy field did not seem particularly different from past ICT initiatives meant to bring evidence in government.

Nonetheless, by reflecting on "data for policy", some interviewees seemed to start considering what it exactly meant to use data in connection with "policy" and how this was a different type of use of data in government. These reflections addressed the meaning of policy. One Expert highlighted how "policy" is often used for defining all decision-making in government, while a line should be drawn between decision for policy and decision for public programs. Using data for policy, according to two interviewees, required to use data to understand the reason behind public issues and setting a fundamental line of governmental direction toward a public problem.

In the sharp words of one interviewee, "policy" was to "public programs" what "steering" was to "rowing". Different interviewees agreed that these two dimensions were not to be conflated, at the same time they agreed that data collected by program monitoring could be used to support policy decisions. On this note, two of the interviewees' visions converged. As the nature of data would change drastically depending on the context in which they are collected, that context had to be considered when using the data for policy. What appeared implied, in essence, was the proposition that using data for policymaking demanded to consider the broad spectrum of the policymaking process.

Theme 2 – Political factors in data for policy

Table 12. Expert interviews' excerpts for Theme 2

INT-5	Elected representative versus the public servant, the think tank versus the practitioners, I think that's where the biggest gap lies, in terms of really being able to sit down and say: What do we take away from the data? And how do we use the data to change what we're doing, and better understand how it's working on an ongoing basis? Because it exposes you, you cannot hide, if you have set yourself up to say: Here's our baseline, here's our intervention, here's what we expect to change in a positive way.
INT-9	If you cannot exchange data, or if you don't want to exchange data, and then you use the excuse of privacy or, you know, technical problems, or legal barriers that sometimes are not necessarily there
INT-7	And in this sense, a lot of public administrations do not want to open their data. Because, actually, data is wealth for them, data is power. So if they share their data, they kind of lose, you know, some of their power, and that is why they are a bit reluctant.
INT-11	For example, the cities themselves are very protective of their data. So they don't actually share their data often with the national statistics or the national level ministries for that fact. So what happens is that they have knowledge about their own sort of city, but they're unable to connect that knowledge to a larger sort of database. And that means there, they might oversee things or patterns that come from actually linking that microdata to larger statistics. But they're afraid of doing that, because then it's revealed basically what their performance is.

Among the interviewees, there was agreement that political factors were to be regarded as influential in data for policy. It would be correct to say that the sample aligned with a policy-pessimistic perspective on the use of non-traditional data for policy (i.e., contrary to a techno-enthusiastic perspective) (Vydra & Klievink, 2019). This opinion was discernible in eight out of nine interviews transcripts. The political factors were presented as contrasting factors to data sharing and use. Among the possibly more pessimistic accounts, the reasons behind these practices were to be regarded as the fear of public authorities to lose political power or become accountable and publicly exposed to undesirable levels. The political factor would also connect to the specificity of using evidence in policymaking (see next point).

Theme 3 – The nature of evidence in policymaking

Table 13. Expert interviews' excerpts for Theme 3

INT-4	Yes, we can integrate data, but that still has to somehow relate to the organization, the social organizational context in which that evidence or that data has to provide reasonable evidence or usable evidence
INT-9	And sometimes it's not enough to have, you know, the entitlement, the benefits, the data of the benefits, but you need to understand the conditions underpinning the context. Why some persons are in troubles? Is it enough to give them some money, and then what they do with their money [...] But as we well know, then the decision making process is not necessarily based on the data, because either the data, I mean, doesn't necessarily show you the causality links. So, you may have some data, but you're not sure if this is because of a certain... it's the effect of a certain policy or activity.
INT-12	[...] data reports, or dashboards... if they're just too much of them, too thick, too heavy, too inaccessible, not relevant enough, too generic... Yeah, it's just not helping [...] A politician is always eager to give an answer, to sort of demonstrate that there's no problem, or to come up with a solution, that sort of political reality. And it's in the culture of politics. [...] So if it's clear, accessible, relevant, up to date, high trust, then we use it.
INT-13	Ah, look, frankly, I think we overestimate the role that information plays in decision making process. Let me just share this with you. So I've spent my entire career, 30 years doing... trying to improve the delivery of information into decision making processes, only to see that information repeatedly ignored.

In line with insights from knowledge utilization literature (see Chapter 2.5), several authors challenged the notion that more access to data would directly connect to a more diffuse or better use of evidence for policymaking. One of the Experts that worked in the central government highlighted that its standard work as a political advisor depended on very tight timing, and only a few weeks could be granted by law to collect evidence. Too many data sources, in this interviewee's opinion, might even be counter-productive for decision-making (or simply being ignored). The same interviewee also stressed how data could be used only when coming from highly trustable sources and not being generic (e.g., statistics) (see Section 1.3.3.2). Other Experts noted how non-traditional data could often be of little use without proper contextualization. To understand the causes underlying policy problems is a common need of public policy, which can be hardly solved by data detached from context or other sources for sense-making.

Theme 4 – The importance of data literacy and culture

Table 14. Expert interviews' excerpts for Theme 4

INT-1	Another example, I mean, I think one, I think that data literacy, in some ways, is more important than technological novelty
INT-7	[…] because inside public administration there is still a kind of lack of people that is really able to use the data and to treat the data.
INT-8	Well, to start with, there's the skills problem. The people you want are very gainfully employed in industry
INT-9	Ah, we need more data science. Yes, we need more data science. But also we need more capacity to understand what we need. And for instance, also interdisciplinary approaches, social logical, or MDG. Engineers talking to trying to find a way to communicate. […] And that's... more than capacity, is a cultural issue. So the culture of understanding the data is something that should be basics in the school, the primary school.
INT-12	If you have policymakers who have no understanding of the data, and how it is produced, what the limitations are, how it could be improved, what other sources of information there are alongside the data, then you get into a risky area.

When asked about the potential barriers to the use of non-traditional data in policymaking, the lack of internal capabilities and competences emerged across several interviewees (5 out of 9), thus confirming the literature. Also, other barriers usually to be found in literature were mentioned (see Section 1.3.3.3), e.g., the lack of interoperability, data governance and legal frameworks.

However, some interviewees appeared rather optimistic regarding these problems, stating that they would have been increasingly addressed as they were becoming more well-known to government. Conversely, Experts seemed to agree that the lack of data literacy and culture in the public sector was a more intrinsic problem. By asserting that, Experts appeared to imply that the need of a more holistic knowledge and appreciation of data and evidence was to be regarded as a challenge in data for policy — possibly greater than the inclusion of technical competences.

Theme 5 – Data Ethics

Table 15. Expert interviews' excerpts for Theme 5

INT-1	And I think the first thing we need to do is realize, okay, the way that they handle data for decision making is based on a commercial logic, we need to understand, number one, what is data for us? Because for Google, and Facebook, and so on, I mean, the data is equals money, basically. But in a democracy, it doesn't necessarily do that. […] we have algorithms that decide if you get social benefits or not, that the programmers have decided certain variables within those algorithms. That should actually be a political decision. But it's too complex for politicians to understand how you how these variables interact with different aspects of the code.
INT-7	Artificial intelligence is the most important, because artificial intelligence basically allows you to take decisions in an unprecedented velocity and allows you to automatize policymaking in a sense. […] Horizon 2020 projects, that are just starting, consider the ethic dimension as very, very important.
INT-9	Who are you, government, to anticipate that I will have a problem in the future, if I commit a crime today? Or if a child that is with parents, poor parents, drug addicted, they will become a criminal? Of course, he has more chances to do that, but there are more issues that need to be addressed.
INT-11	And I think slowly, I think it becomes clear also with these scandals, these AI driven scandals, where people say, well, it's not the algorithm itself, it's the data that was underneath the algorithm, people understand that the data is not neutral. It's how you sort of what do you do with the data? How you analyse it, how you link it?
INT-12	[…] if you get a level higher, to the policy of, let's say, a national educational policy […] what then can you learn actually from these new data sources, or these new combinations of data? Can you learn that specific schools are doing well and others are not? And then can you intervene and maybe supporting the weak schools a little bit, or training, or helping them? If that's an outcome, that to me, it sounds very… well, feasible, and also quite okay. But if it comes about some sort of ranking, and measuring, and punishing, and naming, and shaming… Where some schools are based in very tough cities and very tough neighbourhoods, and some other schools are based in very posh, suburban areas with a lot of Teslas on the driveway. Well, what does it say then, if a certain school scores a little bit better?

The topic of Artificial Intelligence emerged quite naturally and was repeatedly addressed in the interviews. Different Experts acknowledged the potential of these technologies and the fact they were under the political spotlight, although they had different opinions on their impact. When asked about the several advanced analytics cases in public administration (found by the author in the literature) (see Section 1.3.3.3, 1.3.3.4), Experts seemed to consider them far from public policy. Discussion about AI linked to the topic of data ethics, as well as the social implication of digital technologies for public decision-making (e.g., modelling and forecasting). These implications were reconducted to at least two main ethical aspects: firstly, the approach of governments to the value represented by data (driving choices in public information systems); secondly, socially harmful outcomes due to data misuse and the potential lack of representativeness when non-traditional data were to be adopted for policy decisions. Regardless of slightly different angles on the topic, most of the Experts raised the ethical implication as central in data for policy.

4.1.2. Literature analysis

The document analysis performed in phase 1 consisted of a qualitative review of a sample of 74 conference papers, presented during the first four editions of the "Data for Policy" Conference (2015, 2016, 2017, 2019) (i.e., all the available articles at the time of the analysis). To connect relevant themes across interviews and this body of literature, coding was used through the qualitative analysis software Atlas.ti.

In line with the approach described in the methodology, the goal of coding was only to individuate the main themes in this body of literature, i.e., *theming* the data (Saldaña, 2013). For each conference paper, a choice was made about which topics were to be coded and included as themes of the analysis. This choice was operated by evaluating a maximum of three main topics for each paper that appeared to be the most important for the paper's argument, mainly derived from the information in the title, abstract, key words and conclusions. This work implied two rounds of review. The affiliation of the author (or group of authors) was also considered.

Most authors had an academic affiliation (54 of 74 papers), and the remainder was divided between public sector, international organizations, private sector and research centres. Through the analysis 97 individual topics/codes emerged. Instead of merging them, the individual codes were grouped according to the researcher's interpretation. The several groups emerging were considered to individuate different distinct themes (see Annex 4). This choice intended to maintain transparency on the topics as they were originally found in the texts analyzed. The topics isolated in the review of data for policy (see Section 1.3.3.1) were used as a reference. In some cases, they could not be applied or had to be merged. For example, the topic/code "privacy" was used to discuss ethical implications about data use or the legal aspect of data sharing. As a consequence, groups as "data ethics" and "privacy" (originally intended separated) (see Section 1.3.3.1) were merged into one group.

The biggest codes group was identified as "*Data Ethics/Privacy*", with 29 topic/codes individuated. "*Privacy*" was the most recurring code in the group (9 times) and the second most recurring in the whole code set. "*Data ethics*" and "*Transparency*" (3 times respectively) are the second most used codes for this group. The second biggest group was named "*Specific topics, technologies and applications*" (28 codes) and individuated a series of specialized themes and technologies discussed in the papers. The third group contained 23 codes about "*Data (sources and types)*" indicated in the papers. Administrative data was the most recurring data source/type (9) and the topic/code with the highest counting in the code set. The second most mentioned data source was "open government data" (7). The following groups encompassed sets of topics/codes brought together into groups as "*Data science and analytics)*" (15); "*Artificial Intelligence/Machine Learning*" (14) and "*Analytics and Modelling*" (13).

The next code group of almost equal size contained themes relatable to "*Citizens engagement and participation*" (13). Finally, the least numerous codes groups emerged were defined as "*Policy-related topics*" (11); "*Data Culture*" (8) and "*Data Governance*" (7).

The analysis was integrated with a counting of all the keywords used in proceedings of all the four editions of the Data for Policy Conference (edition 2015, 2016, 2017, 2019, 2020, 2021). The keywords could be scraped from an API end-point in the conference's open access repository on Zenodo[99]. After the data from the exported were cleaned with the software Open Refine, a dataset of 539 single keywords resulted from the analysis. The counting of instances in this dataset was operated again with the software Atlas.ti in order to understand the most recurring ones. Most of the keywords recurred only once, except for "data governance" and "privacy (Fig. 16). However, this analysis was meant to be only indicative as both the time frame and volume of proceedings were different from the group analyzed manually.

Fig 16. Word cloud based on keywords from papers of "Data for Policy Conference" on Zenodo platform (2015-2021)

99 The Data for Policy Community on Zenodo is accessible at https://zenodo.org/communities/dfp17/?page=1&size=20

4.1.3. Discussion of phase 1 results

Table 16. Comparison of topics across three sources in the data for policy discourse

Literature review	Expert interviews	Literature analysis
Data governance		Data governance
Data Culture	Data Culture	Data Culture
Data science and analytics		Data science and analytics
Data quality		Data quality
Privacy		Privacy
Data ethics	Data ethics	Data ethics
	AI/ML	AI/ML
Use of Evidence in Policy	Use of Evidence in Policy	
	Data Literacy	Citizens engagement and participation
	Political factors	Specific technologies and areas of application
	The perception of data for policy as field	Analytics and Modelling

The research hypothesized “data for policy” as an autonomous field (see Section 1.3.1) since several themes could be identified among authors discussing the innovation of non-traditional data for policymaking (see Section 1.3.3.1). These themes emphasized aspects not usually addressed in the broader data debate, e.g., the importance of data quality and the relevance of the policy context for using non-traditional data (see Section 1.3.3.2, 1.3.3.5). In synthesis, this discourse was way more oriented toward the context in which non-traditional data were used and its characteristics, rather than data or technology per se (see Section 1.3.4). Given this specificity, it was sensible to imagine similar insights to emerge either in the scientific discourse explicitly connected to data for policy or from the perspective of experts operating in this field.

The results of Phase 1 suggest that an orientation toward policymaking was more detectable through Experts interviews than in scientific literature on the topic. The Experts were not necessarily convinced of data for policy being something new or different from other strands of research and practices addressing the phenomenon of digital technologies in government. However, once prompted by the notion of “data for policy” to reflect on the meaning of “policy”, they agreed that a set of considerations was to be made in considering non-traditional data for policy. These considerations seemed to link back to the topics found in literature.

The centrality of Data Ethics (i.e., the ethical use of data) and Data

Culture (i.e., the recognition of the value of data by the public sector) in the view of Experts essentially suggest an acknowledgment of the normative political dimension of using data, which remains above any technical and technological consideration (Table 16). The political dimension and its influence on the use of data were also very clear in the vision of Experts, and resonated with literature on knowledge utilization for policy and evidence-based policymaking (see Section 2.5) and more recent accounts of non-traditional data for policymaking (Kettl, 2016). While some Experts might have advanced the influence of political factors with a shade of resignation, they did not appear to present politics as intrinsically bad for non-traditional data in policymaking, but simply an intrinsic aspect of governing. For the research, this confirmed that it would have been complicated to find a direct and rational use of non-traditional data in policymaking, as suggested also by literature (van der Voort et al., 2019).

If we were to regard these Experts' perspectives as a signal of how the field is orienting itself (Bogner & Menz, 2009), it could be affirmed that the recognition of the specificity and political nature of policymaking is still not fully reflected in the dedicated scientific literature, of which the body of papers analyzed was taken as an example. While these writings address many specialized topics — and Data Ethics and Data Culture were among them — it also appears still largely concerned with themes that mostly regard technological applications/innovations and regulations. Papers clearly taking issue with the policy dynamics of using non-traditional data in policymaking (Longo et al., 2017) were not recognized. In the papers analyzed, privacy emerged as a relevant topic, but often addressed as a matter of regulation. On the other hand, the analysis also interestingly shown a concrete interest of this community for citizens participation and engagement. In general, the impression left to the author from this analysis is that the conference, up to that point, dealt more with policies for data than data for policy. The venue might be working as a basin of confluence for the many existing topics in the sphere of digital technology for government, thus it still seems fragmented and in search of a clear identity (Verhulst et al., 2019), possibly as the whole field it seeks to represent.

4.1.4. Limitations of phase 1

A main limitation of the phase 1 regards the representativeness of the samples considered for describing a field certainly huge and multifaceted. Using as a reference the "Data for Policy Conference" limited the issue since the venue could be arguably taken as an explicit representation of this discourse. However, as for any other conference, the subjects treated in the various editions were likely depending on the editorial choices made during these years. In fact, the keywords counting (that contemplated also more recent years) was already balancing the topic in the discourse differently.

Another limitation was the bias of the author in conducting the coding analysis. As the starting coding scheme was based on a literature review of topics also conducted by the author, the risk of ending up in a self-fulfilling interpretation cannot be totally dismissed. However, that is partly unavoidable in qualitative research, which depends and thrives on subjectivity (Brinkmann, 2013). In qualitative research, the subjectivity brought by the researcher to the context is intrinsically connected to the findings (Patton, 2014, p. 134). The research attempted to mitigate that aspect by including several sources in the coding analysis, in line with what was suggested by the systematic literature review on the topic of interviewing. Overall, the analysis of the discourse of phase 1 provides an exploration into a yet fragmented field (Suominen & Hajikhani, 2021), and was meant to engage with it and improve the overall orientation of this research for its later phases.

4.2. Results of phase 2

4.2.1. Desk research analysis

The desk research resulted into a list of 20 case studies. Four were later included in the QCA analysis, while the other sixteen are here synthetised (Table 17). Desk research case studies were developed and analysed through secondary sources. For few cases reached through snowball sampling (i.e., direct suggestion from the Experts/Key Agents engaged in phase 1), it was possible to integrate the secondary data with an interview developed according to the Key Agent format (Annex 2). A list of desk research case studies with full description and sources is made available in Annex 5.

Table 17. Summary table of desk research case studies

ID	Case Name	Main subject	Country	Organization type	Date	Specific type of org
1	Non-compliance identification and risk assessment in tax system	HM Revenue and Customs	United Kingdom	Public financial corporations	2017	Non-ministerial Department
2	Piano casa Italia	Presidency of the Council of Ministers	Italy	Central government	2016	Ministerial Department/Office
3	Their Future Matters	Department of Communities and Justice - New South Wales Government	Australia	State government	2016	Ministerial Department/Office
4	Housing Benefit Matching Service	Department of Work and Pensions	United Kingdom	Central government	1996	Ministerial Department/Office

5	Kenya Livestock Insurance Program	Department of Agriculture - State Department for Livestock	Kenya	Central government	2014	Ministerial Department/Office
6	The NEAR Program	Department of Industry, Science, Energy, and Resources	Australia	Central government	2017	Ministerial Department/Office, National Agency
7	Program to Calculate Deforestation in the Amazon (PRODES)	National Institute for Space Research	Brasil	Central government	1988	National Agency
8	Tackling opioid crisis through public health data	U.S. Department of Health and Human Services	United States	Central government	2017	Ministerial Department/Office
9	RapidSMS	Rwandan Ministry of Healthcare	Rwanda	Central government	2009	Ministerial Department/Office
10	Kennisnet supports the educational system use of learning analytics	Kennisnet - Ministry of Education, Culture, and Science	Netherlands	Public non-financial corporations	2011	Ministerial Department/Office
11	The Norwegian Agency for Public Management use transactional data to optimise digital public procurement	Agency for Public Management and eGovernment (Norwegian Ministry of Government Administration and Reform)	Norway	Central government	2012	Public Agency
12	Management of public administration personnel data at the national scale	Directorate for Information Systems and Innovation (DSII) . Ministry of Economic and Finance	Italy	State government	2016	Ministerial Department/Office
13	A Mobility DataLab in the city of Bergen	Bergen Municipality	Norway	Local government	2017	Municipality
14	Helsingborg Data Lab	Helsingborg Municipality	Sweden	Local government	2018	Municipality
15	City of Ghent uses mobile phone data to identify the habitational patterns of students	Data and Information Service of the Ghent City Council	Belgium	Local government	2017	Municipality
16	xKRP – Community Experience Data Lab Kronoparken	Research Institutes of Sweden (RISE)	Sweden	Public non-financial corporations	2016	Agency

Few provisional analyses were developed on this dataset, in line with the methodology (see Section 3.4.1). As expected, applying the policy cycle to perform a neat categorization proved challenging, even if the model was integrated with the additional layer of policy work/task for each phase (Wellstead & Stedman, 2015). No cases could be interpreted as being part of agenda-setting activities, whereas examples connected with implementation could be easily recognized as such. For instance, the use of predictive analytics in fraud detection ([1],[4]) or public service interventions monitoring in health policies ([8],[9]). Several cases could be reasonably connected to a policy formulation stage, as they identify an explorative use of non-traditional data to appraise various policy options. Piano Casa Italia [2] could be connected to the decision-making stage, as it regarded the use of governmental and administrative data for calibrating existing policy tools within a urban development and risk-assessment policy. Some cases might be applied to several stages. In the case "Their Future Matters" [3] the Australian government developed a financial model based on an administrative dataset to predict future use of welfare services by children and families. The case might be considered relevant for the formulation stage (since new policy options might be informed by these forecasting) but it also entails implications for monitoring (as the system works through routinely collected service data). Its financial model also offer chances for evaluation purposes.

In short, the analysis faced several limits in discerning cases by using the chosen interpretative lens and secondary data (Table 18). In the perspective of policy tools (Hood & Margetts, 2007), most cases could be linked with information policy tools (i.e., government tools for gathering or publishing information) often in conjunction with organizational tools (i.e., the development of new organizations)(Howlett, 2019). Piano Casa Italia [2], for example, was constituted with a mission to employ non-traditional data sources of data for addressing the existing policy problem. In this sense, either at the national or at the local authority level (the two main types of subject in this first list), a mechanism of delegation of non-traditional data use could be inferred by the presence of dedicated internal units, labs or experimental spaces ([12], [13], [14], [16]), possibly to compensate a lack of internal skills (Giest, 2018). In terms of policy problems addressed there seemed to be some recurrences: e.g., public personnel management ([11], [12]), mobility and citizens flow ([13], [15]) and healthcare ([8], [9]).

Table 18. Provisonal analysis of desk research case studies

Policy Cycle Stage	Policy work and tasks	Desk case (tentative grouping)
Agenda Setting	• Identifying policy issues • Identifying policy options • Environmental scans • Consulting with public	No cases
Formulation	• Appraising policy options • Collecting policy-related data • Collecting policy-related information • Conducting policy-related research • Negotiating with stakeholders • Preparing position papers	[3] Their Future Matters [6] The NEAR Program [7] PRODES [10] Kennisnet [14] Helsingborg Data Lab [13] A Mobility DataLab [15] City of Ghent uses mobile phone data [16] xKRP

Decision-making	• Comparing policy options • Decision matrices • High-level briefing • Negotiating with central agencies • Department planning	[2] Piano casa Italia
Implementation	• Implementing/delivering • Consulting stakeholders • Negotiating with program managers • Legal Analysis	[1] Non-compliance identification [3] Their Future Matters [4] Housing Benefit Matching Service [5] Kenya Livestock Insurance Program [8] Tackling opioid crisis [9] RapidSMS [11] The Norwegian Agency for Public Management [12] Management of public administration personnel data
Evaluation	• Policy evaluation skills • Risk-based tools/techniques • Evidence-based policy	No cases

4.2.2. Sample of data-centric policymaking cases for comparison

The four cases for QCA were selected from the initial list of 20 desk research case studies. The sampling was based on the intention of matching comparable cases, but that featured diversity on multiple characteristics (Patton, 2014, p. 423), as this is considered a good way to develop generalization. The cases were chosen either at the local or regional government scale (Hooghe & Marks, 2016), given the interests of this research (see Section 1.4.3). An opportunistic sampling strategy was also followed, based on the willingness of case studies' actors to participate to the data collection (i.e., key agent interview, survey on learning) (see Section 3.4.2).

The four cases selected took place during the last five years in four different European countries (Sweden, Belgium, Italy, Netherlands). The national difference already constituted a first dimension of comparison and diversity across cases in terms of structural and political conditions of data use. Among the 33 countries ranked in the Digital Government Index 2019 (OECD, 2020a, p. 54),Italy and Netherlands ranks fairly better in the dimension of data-driven public sector (13th and 18th respectively) than Sweden and Belgium (23th and 28th respectively), with Italy being over the OECD average. On the other hand, among the 34 countries considered in the OURData Index (OECD, 2020b, p. 21), the Netherlands ranks above Italy and Belgium but without great distance (respectively 13th, 17th and 19th position in the index), while Sweden remains almost at the bottom of the ranking (31st position). According to these sources, it is possible to say that Netherlands, Belgium and Italy are average national contexts for using data in the public sector, while Sweden features structural conditions that are way less favourable (Fig 17). A dimension of difference in case studies sampled could also be retraced in the type of governance structures that emerged (Micheli et al., 2020) and the policy problems addressed.

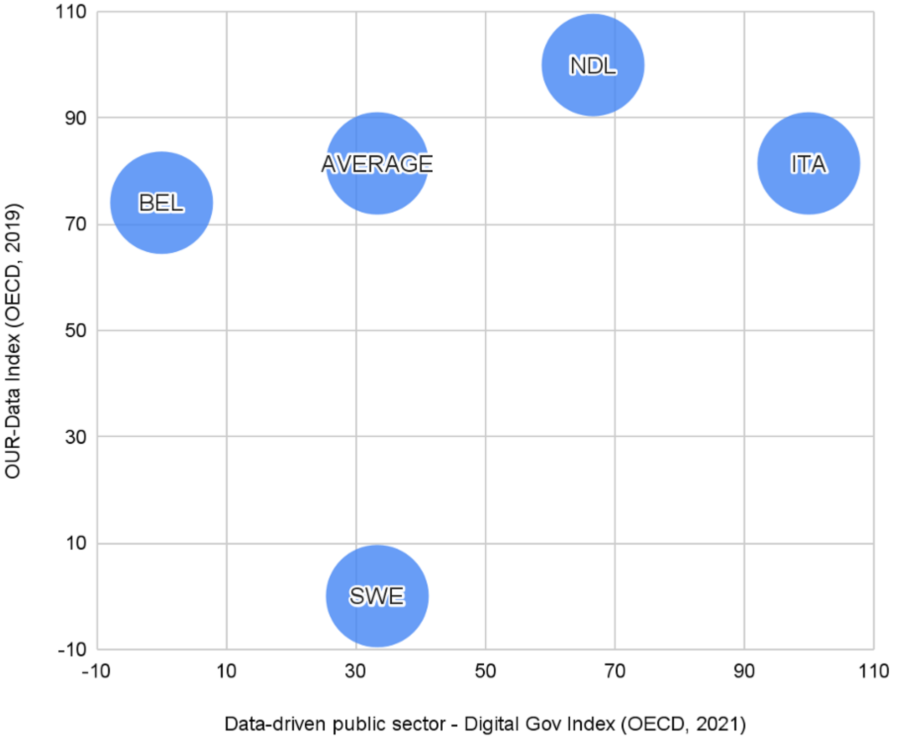


Fig 17. Comparison of case studies based on ODI and DIG indexes (OECD)

4.2.3. Description of cases

Case study 1 (SWE)

Understanding fragmented services pathways with administrative data to support vocational rehabilitation (SWE)

The Swedish public agency Samordningsförbundet Centrala Östergötland (SCÖ) (Coordination Agency of Östergötland Region in English) wanted to develop a new Digital Welfare Guide that could support regional and local bodies in vocational rehabilitation services. The agency team partnered with a network of experts from Sweden, Iceland, and the United Kingdom to work with administrative and public data. They intended to investigate how citizens have used the different services (i.e., the service pathways), thus addressing the problem of coordinating several vocational rehabilitation welfare services in the Region.

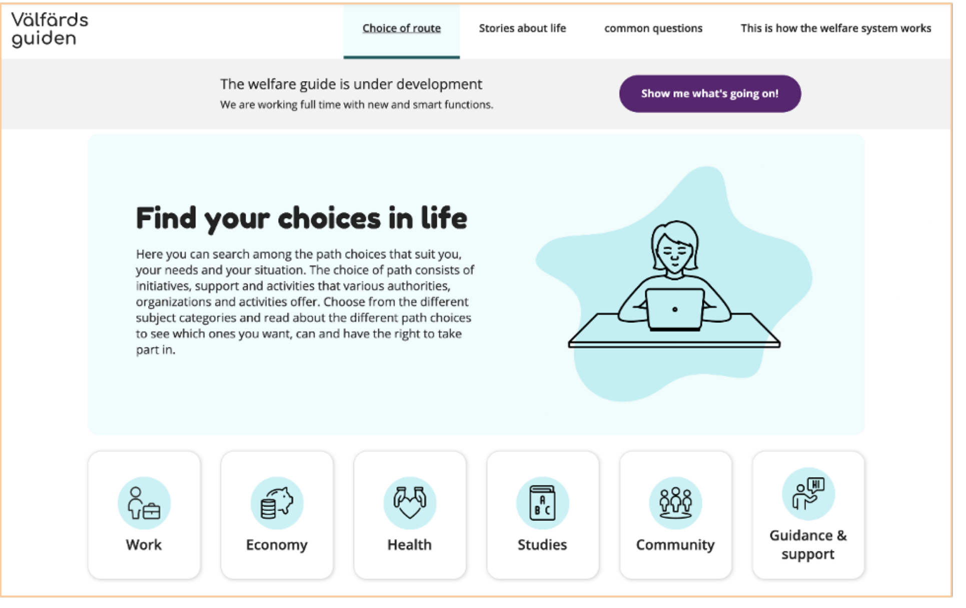


Table 19. Summary table of Case study (SWE)

Country	Period	Main public subjects involved	Type (Eurostat, 2021)	Sources
Sweden	2018-2022	Samordningsförbundet Centrala Östergötland	Autonomous administrative authority	- 2 key informant interviews with project manager at SCÖ - Online attendance to 2 working meetings with data scientists researcher (held in English) - Online attendance to 1 SCÖ board meeting (held in Swedish, I was provided with live transcript/notes of meeting translated in English) - Document analysis (project memos, notes, grant proposal documents, planning documents) - Desk research - Survey
		Östergötland Region	Local Government	
		Swedish Social Insurance Agency	Central Government	
		Swedish Public Employment Service	Central Government	
		Swedish Municipalities (Kinda, Linköping, Åtvidaberg)	Local Government	

Policy problem/area

Vocational rehabilitation

Type of data used

- Administrative data about beneficiaries of rehabilitation public services (age, sex, etc.)
- Administrative data on rehabilitation public service attendance
- Public data on rehabilitation public service offering

Policy context

SCÖ operates as a partnership association that connects the Municipality of Kinda, Linköping and Åtvidaberg, plus the regional chapter of Swedish Social Insurance Agency and Swedish Public Employment Service (both government agencies). SCÖ is one of the Coordination Agencies established in 2004 by the Swedish Act on Financial Coordination of Rehabilitation Initiatives (Finsam), advanced by the Swedish Parliament (Riksdag)[100] in 2003.

Vocational rehabilitation, the main target addressed by Finsam, regards "*a multidisciplinary intervention to help individuals to return to work after an occupational injury, or a period of unemployment or sickness [...] usually includes different health and social services, occupational health services, employment services, and social or private insurance*" (Andersson et al., 2011, p. 2). The relevant stakeholders individuated by the Act are the public authorities at the city and regional local levels, the Swedish Social Insurance Agency and the Swedish Public Employment Service (at the central government level).

100 The background of the act was a series of discussion and evaluative experiments on how to address sickness absence in Sweden. What emerged from them was that sickness absence could not be imputed to more individuals taking sickness leave days, but longer uninterrupted period of sickness from singles individuals. The reason was attributed to previous reforms and fragmentation of the landscape of Swedish public services. Due to this fragmentation, main rehabilitation services could not be any more accessible to several target groups, whose needs were falling in between the organizational boundaries (Andersson J., 2016; *Finsam – a Follow-up of Nancial Coordination of Rehabilitation Measures*, 2014). The Act took inter-organizational collaboration as main solution to this problem and issued a set of pooled financial resources to be shared between stakeholders in central government and local authorities. The goal of inter-organizational collaboration was to enhance services for vocational rehabilitation to improve work capacity of unemployed citizens. From 2012 to 2015, about €530 million were allocated to budgets by central government and local authorities in compliance with Finsam law. The European Social Fund also was used to finance specific projects within Finsam framework (between €35 and €43 million per year).

Against this legal background, the Coordination Agencies are constituted as separated legal entities that can contribute and access to the pooled budget distributed by the law. In 2015, Coordination agencies were about 82. Coordination Agencies do not coincide with local government administrative areas, therefore more CAs can operate in the same local authority area, or one CA can extend across diverse counties. Each agency establishes a joint board that includes members appointed by participating counties and cities, and civil servants from the central government. CAs do not have statutory authority from the central government and member organizations can decide to join or retire at their will. Finsam Act purposely established a good degree of freedom and independent budget for CAs to be managed with local intervention. This delegation of policy capacity is motivated as CA are meant to support coordinated initiative for vocational rehabilitation on the local level. CAs mainly act to enhance collaboration by coordinating the actions of its member organizations through information sharing, management and creation of joint projects that directly affect services for vocational rehabilitation. The CAs represents a relevant example of an organizational tool for policy implementation, being an organization with mandate to address policy relevant issues by working both at the procedural (enhancing coordination among subjects) and substantive (addressing better service quality) level (Howlett, 2019).

Description of the data-centric policymaking process

Around 2018, SCÖ launched a project to develop a new Digital Welfare Guide that could provide access to public services by integrating administrative and public data in a single digital solution. Up to 2021, the first project output has been an official prototype of the Welfare Guide up to the level of initial testing (cf. the Societal Readiness Levels proposed by Innovation Fund Denmark (2018). The release of the prototype in 2021 was the last step of a collective process of designing, centered on the analysis of public and administrative data sources. The internal project team included ICT and machine learning experts, UX and service designers and content editors (about 14 people). As the project progressed, the internal team was complemented with an international group of researchers in the field of data ethics, data science and digitalisation. The international participation was established with a consortium, which added competences in machine learning and healthcare to SCÖ. The international participation was a fundamental step to the experimental employment of data in the project given the context of data governance and privacy in which SCÖ operated. The goal of the Welfare Guide was the development of an automated recommendation system based on existent administrative data, that could propose service paths for vocational rehabilitation across various welfare services to the citizens.

To develop this project, SCÖ could rely on two main sources of data.

Firstly, anonymized samples of administrative data at the individual level provided by the Swedish Social Insurance Agency. These data describe the attendance and identity (e.g., age, sex, etc.) of citizens who used vocational rehabilitation services of any type (e.g., training, counseling, etc.). Secondly, data on existent public services on the territory, which were collected in another national-level project of SCÖ called the "Activity Catalogue". The data from Activity Catalogue represent existing services and their offering and could be shared by public subjects willing to participate without any privacy concerns. The data from the Swedish Social Insurance Agency, on the contrary, would have been illegal to use and share in Sweden. The team worked with these data (e.g., cluster analysis, correlational matrix) to understand if service pathways could be identified across vocational services provided. The intention was to understand how the systems worked and how the policy problem of vocational rehabilitation could be addressed.

Case study 2 (ITA)

Investigating mobility and citizens' flows to face COVID-19 pandemic

The Covid-19 pandemic gave impulse to a series of experimentations with non-traditional data, especially mobile phone data (Oliver et al., 2020). As part of a local government plan for returning to normal urban life after the first lock-down periods, the Municipality of Milan decided to experiment with several non-traditional data. The occasion to experiment with non-traditional data to understand citizens' flows was leveraged as part of a European initiative.

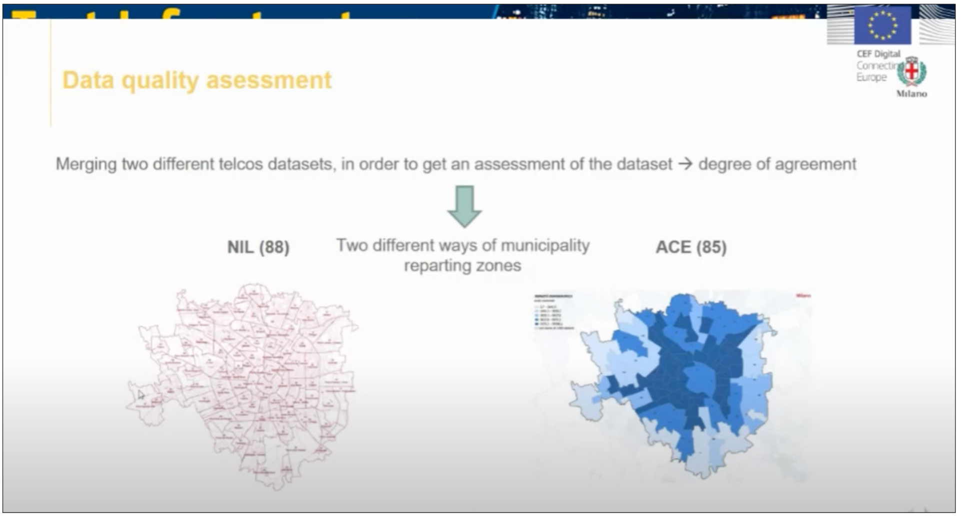


Table 20. Summary table of Case study (ITA)

Country	Period	Main public subjects involved	Type (Eurostat, 2021)	Sources
Italy	2020-2021	Municipality of Milan	Local Government	- 1 key informant interviews with data scientist at Comune di Milano - Desk research - Survey
		CEF Digital	Supra-national Government	

Policy problem/area

Urban mobility and citizens' flow

Type of data

- Telecommunications data, Call Detail Records
- Commercial activities data
- Devices Connection to public wi-fi hotspots

Policy context

In April 2020, the Municipality of Milan launched a strategy and public consultation document for responding to the first lockdown and quarantine imposed by Covid-19 pandemic. The Milan 2020 Adaptation Strategy has been a formal vision by the Municipality for a gradual return to new ordinary urban life. At that time, restrictions originally imposed during March 2020 in the city were expected to be gradually alleviated in the upcoming months. As a form of public hearing, the Adaptation Plan represented an organizational policy tool for collecting initiatives and projects from citizens (Howlett, 2019, p. 51). The Adaptation Plan described several goals and three main areas of intervention: maximizing flexibility of urban transportation services to avoid overcrowding; re-valuing social activities and accessibility of social spaces; enhancing digital infrastructures (to support smart working). Addressing the policy problems of urban mobility appeared central in the Plan (Pucci et al., 2020), as well as understanding the citizens' flows in the city to better manage reduced capacity of public transportation as lock-downs softened.

Description of data-centric policymaking process

In the context above, two internal offices "*Direzione Sistemi Informativi e Agenda Digitale*" and "*Direzione Progetto Città Resilienti*" of the Municipality of Milan decided to participate to a European initiative developed by CEF Europe, called "*Big Data Test Infrastructure*". This initiative aimed to support local administrations in Europe by offering competences and cloud infrastructures for the analysis of heterogenous data sources in pilot projects. During summer 2020, the two offices involved appointed an internal team of 4 persons to follow the project (2 department managers, 1 office manager and 1 civil servant with data science competences). The team decided to use the data to focus on urban mobility and citizens' flows, in line with the interest outlined by the Adaptation Plan. The data used were both from private and public sources. Private data were Call Detail Records (CDR), i.e., meta-data recorded by providers as single phones or devices interact with network infrastructure cells (e.g., during a call or browsing the web). These data could be geolocated to understand citizens' position in the city during the day. The private data sources used (CDR) belonged to the Municipality from previous projects with telecommunications providers. The public data were instead of two types: data on commercial activities, offered by public bodies as the Chamber of Commerce, and data on device connection in public wi-fi hotspots (which could also provide the devices' position). Private data had been previously purchased by the city from a telecommunication provider, as part of a precedent initiative. The private provider sold the data anonymized and in aggregated form, according to privacy regulation. The work done with data mostly focused on data science techniques to understand how to integrate and normalize different data sources, with analysis on granularity.

Case study 3 (NDL)

Researching organized crime and abuse of property in Dutch cities through national micro-data platform

The project Zicht Op Ondermijning[101] is a national-scale initiative promoted by the Dutch Government of Minister of the Interior and Kingdom Relations. The initiative intends to implement an analytics platform and service, based on high-quality e micro-data on crime (drug-related problems and abuse of private property) collected at the national statistical office. The scope of Zicht Op Ondermijning is to help Dutch Municipalities to better understand their local crime patterns and develop better policies, which are connected by this project through a network.

101 Cf.: https://www.zichtopondermijning.nl/

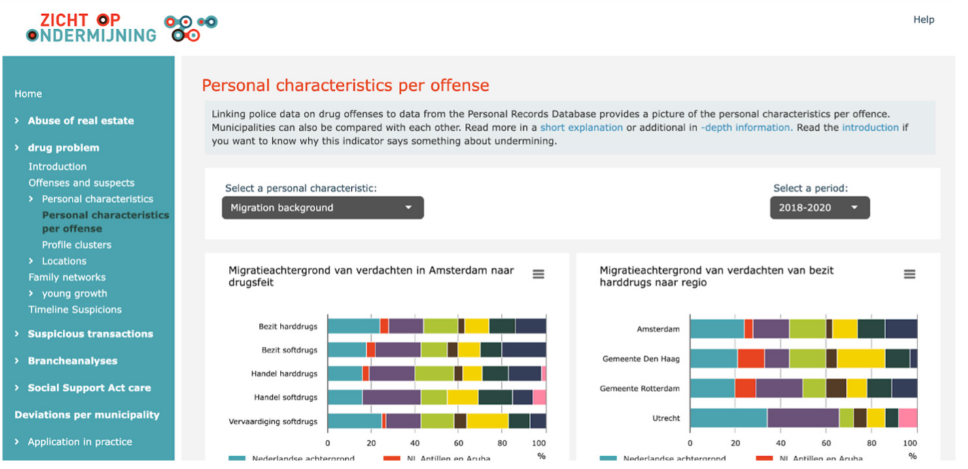


Table 21. Summary table of Case study (NDL)

Country	**Period**	**Main public subjects involved**	**Type (Eurostat, 2021)**	**Sources**
Netherlands	2017-on going	Minister of the Interior and Kingdom Relations	Central Government	1 key informant interview with project leader ICTU 1 key informant interview with two researchers in criminology involved Document analysis (project reports) Desk research Survey
		Ministry of Justice and Security	Central Government	
		ICTU	Non governmental public organization	
		Statistics Netherlands	Autonomous administrative authority	
		14 Dutch Municipalities	Local Government	

Policy problem/area

Organized crime (drug-related issues and abuse of property)

Type of data

- Administrative data on locations and type of drug offences
- Personal data (Personal Record Database)
- Cadastre data
- Real estate data (region and municipalities)
- Purchases and transactions (amount, country of origin)

Policy context

In 2017, the Scientific Council for Government Policy (WRR), an independent advisory body of the Dutch Government, published a document on '*Big Data in a free and safe society*" (Broeders et al., 2017). The document exposed possibilities and risks for using data analysis to prevent criminal behaviors and frauds. In response to the document, the Dutch Government of Minister of the Interior and Kingdom Relations (BZK) sponsored a pilot initiative for the use of national data held by Statistics Netherlands (CBS) called the "*City Deal*". The City Deal was further supported through a formal letter by the Minister of Justice in 2018, that indicated how the initiative had been part of the broader effort of the Ministry to fight organized crime. The City Deal has been conceived as a multi-level governance collaboration between ministries and central government bodies (BZK, Ministry of Justice and Security, CBS, Tax and Customs Administration, Public Prosecution Service) and several Dutch municipalities (Amsterdam, The Hague, Rotterdam, Tilburg, Utrecht and others) and relative police forces on territories. The initiative intended to share micro-data for researching and prevention of specific phenomena as drug-related problems and abuse of private property. The project had assumed the name of Zicht Op Ondermijning. It is managed by the independent non-profit organization ICTU, hired by BZK. In 2019 the project was renewed and is currently at the second stage of piloting.

Description of data-centric policymaking process

The central actor in the innovative use of data is the Zicht Op Ondermijning team (a team of about 21 persons), internal to ICTU. ICTU includes a staff mostly made of data scientists, domain experts and researchers and connects, as part of the initiative, with city managers in the municipalities. To manage the network of stakeholders, the initiative developed its own agile methodology. The process usually starts from research questions posed by cities which regularly send questions to ICTU through a dedicate platform. The questions concern specific issues perceived as relevant, often starting from a specific phenomenon or hypothesis that cities formed on the basis of their local knowledge. All partners cities can see the questions proposed in the platform and prioritize them through a voting system. The most voted are considered the more relevant and selected by data scientists and analysts of the core team of Zicht Op Ondermijning in ICTU. The team decides the design of data analytics and selects the relevant datasets for the question. After the analytics is produced, ICTU shares the results with domain experts and researchers for a quality check, to understand the coherence and soundness of the analysis produced in respect to scientific theories and by considering usability in cities indicators.

The final results are published in the Zicht Op Ondermijning dashboard, which is publicly accessible online. The dashboard shows various visualizations based on predictive analytics and official figures from the Statistical Office. Criminal patterns shown in the platform are at the regional, municipal, district and even neighborhood level.

Case study 4 (BEL)

Analysing city environmental data to understand urban green spaces and citizens' well-being

The City of Leuven in Belgium supported a citizen science project through its Smart City Department, in close collaboration with other city stakeholders. As part of that experimental initiative, the City supported the deployment of 98 monitoring stations for the collection of real-time environmental data (Beele et al., 2022), in order to understand the effects of urban green spaces on citizens' well-being.

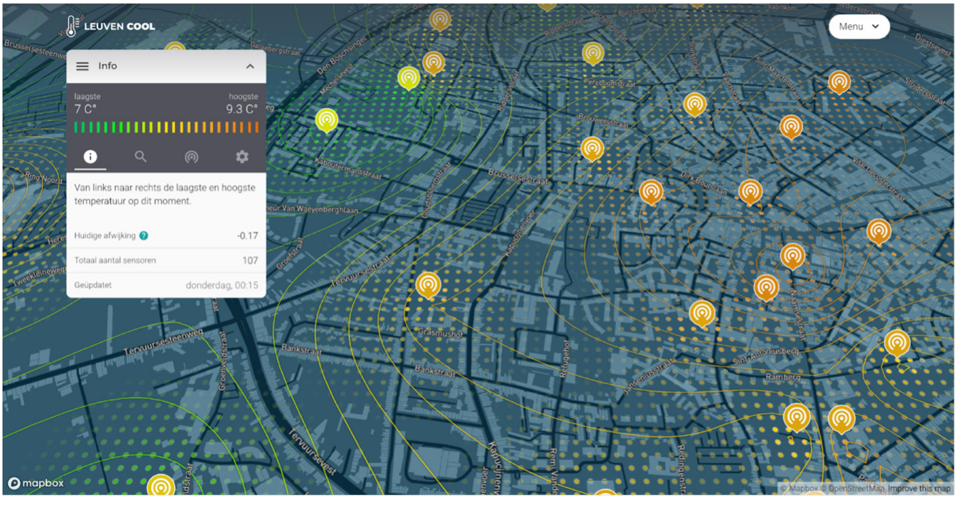


Table 22. Summary table of Case study (BEL)

Country	Period	Main public subjects involved	Type (Eurostat, 2021)	Sources
Belgium	2017-on going	Municipality of Leuven	Local Government	1 key informant interview with Municipal Officer 1 key informant interview with lead researcher in the team Desk research Survey
		KU Leuven	University	
		Royal Meteorological Institute	Federal Agency	
		Leuven2030	Non-profit organisation	

Policy problem/area

Urban heat islands and urban green spaces effect on citizens' well-being

Type of data

- Real-time Environmental data (e.g., humidity, solar radiance, wind speed, temperature, UV radiation)

Policy context

The City of Leuven develops its Smart City Program through a dedicated Department in the Municipality. The Program aims to leverage on innovative technologies and data to improve the life of the city, both with long-term policy vision and specific projects. The five main pillars of the Smart City Program are "a better city experience, stimulating talent, smarter services, optimizing flows (focus on people and energy) and smart(er) health and healthcare." (City of Leuven Annual Report 2020)[102]. The Leuven.cool[103] project has been one of the several projects developed by the Department, in close collaboration with other stakeholders in the city, developed from 2019 to 2022. The intention of the municipality was to understand how green interventions in public and private gardens could tackle the issue of heat zones and urban micro-climate. The city interest

102 Annual Report City of Leuven 2020, p. 369 (translated with DeepL). Available at: https://www.leuven.be/sites/leuven.be/files/documents/2021-07/Jaarverslag_Stad_2020.pdf (Dutch only)

103 https://www.leuven.cool/

also regarded how these factors could affect the well-being and comfort of citizens, with regard to the more vulnerable population.

Description of data-centric policymaking process

The project Leuven.cool has been developed as an experimentation oriented to research and citizen science (Beele et al., 2022), centered on the installation of environmental data in the urban area and suburbs. The city of Leuven promoted the initiative, which was realized in close collaboration with the university KU Leuven (Division of Forest, Nature and Landscape) , the non-profit association Leuven2030 and the Dutch Royal Meteorological Institute.

The project has been mainly implemented by a small team of environmental engineers in KU Leuven, which initially piloted the experimentation by installing 20 sensors in various parts of the city and then scaled up to 98 weather stations. The environmental data from the station is provided in real-time and are owned and analysed by the KU Leuven staff throughout the initiative. On several occasions during the project the data were asked by the city, which used them as an evidence source for the urban interventions (e.g., road paving).

4.2.4. Survey results

Before proceeding with the comparative analysis, the main results from the survey will be discussed in the following sections to understand characteristics and types of actors involved in the cases. The results on learning will be also commented, in general and in relation to the types of actors.

4.2.4.1. Sample characteristics

The survey was answered by a total of 31 individuals across the four case studies. For each case, the data collection started once at least 50% of the actors involved in the cases had answered the survey (Table 23). The size of the sample was provided by the key actors involved.

Table 23. Response rate across cases

SWE		ITA		BEL		NDL	
Tot	Response rate	Tot	Response rate	Tot	Response rate	Tot	Response rate
16	10 (62.5%)	4	3 (75%)	8	4 (50%)	20	14 (70%)

Perception about the centrality of data in the case

To understand the perceived role of non-traditional data for each case, participants were asked if they considered that the innovative use of public/administrative data was the central element in the project. The vast majority of the sample either agree with the statement (18 respondents, 58.1% of sample) or strongly agree (8 respondents, 25.8% of sample). In particular, ITA reports a 100% agreement, with 2 respondents that strongly agree and 1 that agrees. These responses suggest the appropriateness of the sample for studying the concept of data-centric policymaking.

Affiliation and working position

To understand to which degree the case was connected with government/public sector (which was considered a relevant criteria of data-centric policymaking), the respondents were asked to indicate their affiliation. Most respondents across the four cases (13 respondents, 41.9% of sample) declare to have been affiliated with a City/Municipality at the time of their involvement in cases. This type of affiliation is almost totally present in NDL (9 respondents) and ITA (3 respondents). The second biggest group across cases declares its affiliation with a public University (7 respondents, 22.6% of sample). Respondents affiliated with universities are almost equally distributed across the SWE (3), BEL (2) and NDL (2) cases. The rest of respondents declare not to be affiliated with government or public sector (4 respondents, 12.9% of sample); to have worked for a Ministerial department or Office (3 respondents, 9.7% of sample); to have been affiliated with a Public Agency (3 respondents, 9.7% of sample) and to have worked for a Regional Council or agency (1 respondent, 3.2% of sample).

When asked to describe their job position at the time of involvement in the cases, more than half of the respondents declared to be employed as public servants (16 respondents, 51.6% of sample). In line with data on affiliation, the second biggest employment position recorded from the survey is university staff (7 respondents, 22.6% of sample), followed by a private company (6 respondents, 19.4% of sample). Public servants are therefore highly represented in NDL (9), ITA (3) and BEL (2). University staff is present in SWE (3), BEL (2) and NDL (2). The vast majority of those who declared to be employed by a private company belong to the SWE case (5).

These data indicate that the sample is fairly representative of public policy workers as category.

Working experience within government

Given the scope of the survey (investigate policy learning) the survey intended to map the previous knowledge and experience of participants with government and policy problems. Therefore, respondents were asked to express a claim about their knowledge of their own country politico-administrative systems, by identifying themselves with four suggested categories:

- *Full Expert* – possessing deep knowledge of own government/ public sector and expertise in specific policy areas or public services
- *Specialist* – possessing specific knowledge and expertise in only one policy area or public service
- *Generalist* – possessing deep knowledge of their own government system, but no specific expertise on policies or services
- *Average knowledge* – not having any general or specific knowledges about government or the public sector

Most respondents declare themselves as Specialists (10 respondents, 32.3% of sample); the second biggest group declared to have an Average knowledge (9 respondents, 29%); other respondents identified themselves as Generalist (7 respondents, 22.6% of sample) and Full experts (5 respondents, 16.1 of sample%). The two biggest groups are almost fully divided in NDL (8 Specialist) and SWE (8 Average knowledge). The majority of those who declared themselves as Generalist belong to the NDL case (4). Most of the Full experts are divided between ITA (2) and NDL (2).

Previous experience and frequency of working in public sector

In line with the intention of understanding the sample characteristics and affiliation with public sector, respondents were asked if they had previous experience of working with governments, public sectors or public authorities, and how many years of experience they declared.

The majority (13 respondents, 41.9% of sample) declared to have more than 10 years' experience of working within the government. The two following biggest groups have the same size (both 6 respondents, 19.4% of the sample) and declared respectively a working experience spanning from 5 to 10 years and no experience at all. The remaining respondents (6, 19.4% of the sample) overall declared less than 5 years of working experience with government or the public sector. Looking at the specific cases, those who have more than 10 years of working experience are mostly in the NDL case (7), and then equally distributed across cases (2 per each). The second more experienced group (from 5 to 10 years' experience) is divided between NDL (3), SWE (2) and BEL (1). Almost all of those who declared not to have working experience with the government belong to the SWE case (5).

Respondents were also asked to declare with which frequency they have worked in or for the government in the past. Overall, they declared to have often worked for the government (12 respondents, 38.7% of the sample); always (9 respondents, 29% of the sample); never (5 respondents, 16.1% of the sample). The remaining 5 (16.1% of the sample) declared to have occasionally or rarely worked for the government. The majority of those who responded “often” are in NDL (7) and SWE (4). Of those who declared to have always worked for the government, 5 are in NDL and the others equally distributed between ITA and BEL (2). Those who have never worked for the government are almost all belonging to the SWE case (4).

These results indicate that the sample was prevalently made of senior profiles with good past working experience in government.

Presence of technical profiles (data/ICT expertise)

The sample was expected to have a high presence of technical profiles due to the need of working with data. It was deemed important to clarify how many profiles with pure technical expertise were present in the sample, so to understand how this could relate to policy learning (as these profiles were expected to have little policy knowledge). In the sample, 16 respondents (51.6% of the sample) declared not to have a technical profile, while 14 respondents (45.2%) declared to recognize themselves as profile with such technical background description. The majority of non-technical profiles are found within the NDL (9), SWE (5) and BEL (2) cases. Technical profiles are mostly in NDL (5), SWE (4), and ITA (3).

Involvement in the project

The survey intended to define the activities and work done by individual (policy workers) involved in the data ecosystems as part of the cases. To understand their learning, it was important to understand which degree of participation they had.

Most of respondents declared to have had a high (19 respondents, 61.3% of sample) or very high (7 respondents, 22.6% of sample) degree of participation in the project.

The respondents were then asked to indicate which position they had in the process (referring to the division of roles proposed in Section 3.4.2) and to indicate the reason behind their involvement. In line with the notion of policy work, respondents could indicate if they worked with the data as part of the development team, or as part of the broader policy subsystem. Most of them were directly involved in the work of the development team (14 respondents, 45.1% of sample), with different tasks and roles. Their participation in the project was mainly due to their skills in UX/UI, Development, ICTs, Data science, Communication. A total of other 10 respondents (32.2% of the sample) declared that their reasons for involvement were mainly related to their previous experiences with that policy area and policy subsystem, together with their political position or affiliation (Table 24).

Table 24. Involvement and work of actors in the cases of data-centric policymaking

The respondent declares himself/herself part of...	**Count/percentage**	**Reason of involvement (work/knowledge)**
Development team	14 (45.1%)	Verification/test and accessibility knowledge UX/UI Communication General ICT competencies Project manager Networking and coordination Developer Data science Policy knowledge General research competencies Logistical coordination and support
Public/government organization that directly promoted the initiative	5 (16.1%)	Experience with public services Experience in working in public/government projects Political position/affiliation Experience on the policy problem Affiliation with organization owning the data
An organization partner to the project	5 (16.1%)	Experience in gov tech or gov data projects Expertise and experience in policy problem Political position/affiliation Experience in working in public/government projects
An external advisor that directly supported the development teamwork	3	Data science AI/ML Policy knowledge
An organization interested in the project development and outcome	3	Data and AI/ML ethics Experience in gov tech or gov data projects Political position/affiliation Interest of organization of affiliation
None of the above	1	

Looking at the process imagined in the cases (see section 3.4.2), the Data practitioners, i.e., those who were involved in the project as part of the development team, mainly participated in the first phases of the project, contributing to the Problem setting, to the Data preparation, mining and analysis, and to the Prototyping/deployment phases. The actors involved in the policy subsystem gave their contribution in the first and last phases of the project, being mainly involved in the Problem setting and then in the Communication and dissemination phase (Table 25). These results aligned with what was imagined in defining actors and roles for the sample (see section 3.4.2).

Table 25. Stages of involvement by types of actors

	Problem setting & defining needs	Data gathering	Data preparation	Data mining & analysis	Prototyping & Deployment	Communication & Dissemination	None of the above
Data Practitioners	7	3	4	7	5	1	3
Project Management	2	1	0	0	1	1	1
Actors in Policy Subsystem	8	3	0	1	2	6	4

4.2.4.2. Results of Individual policy-relevant learning

This section presents the results of the second part of the survey dedicated to individual policy learning (see Section 3.4.2).

Self-assessment on previous knowledge on problem

When asked if they could be considered experts in relation to the policy problem addressed in the case, most of respondents either disagreed (9 respondents, 29% of sample) or strongly disagreed (6 respondents, 19.4% of sample). A relevant part neither agreed or disagreed with the statement (8 respondents, 25.8% of sample), while only a smaller part agreed (6 respondents, 19.4% of sample) or strongly agreed (2 respondents, 6.5% of sample).

Self-assessment on policy learning: COGNITIVE TYPE

Most respondents self-reported a gain in new information throughout their involvement in the case. Four different questions in the survey were dedicated to this type of cognitive learning, and approached the topic dividing it into four main subjects:

- New information about policy problems
- New information about policy tools
- New information about public services
- New information about system actors connected to the specific area of the case.

In total, slightly more than half of the possible 124 preferences for these four questions resulted in an "agree" answer (64 answers, 51.6% of total), suggesting that an overall learning process was perceived by most of the sample. 23 answers were given to the "strongly agree" option (18.5% of total); 21 to the "neither agree nor disagree" one (16.9% of total); 12 to "disagree" (9.7% of total) and 4 to the "strongly disagree" option (3.2% of total).

With a closer look to the single questions, the response was divided as follows:

- Policy problems: 19 respondents answered with “agree” (61.3% of the question’s total) and 8 with “strongly agree” (25.8% of the question’s total)
- Policy tools: 12 respondents answered with “agree” (38.7% of the question’s total); 7 with “strongly agree” (22.6% of the question’s total) and 7 with “neither agree or disagree” (22.6% of the question’s total)
- Public services – 17 respondents answered with “agree” (54.8% of the question’s total) and 6 with “neither agree or disagree” (19.4% of the question’s total)
- System actors – 16 respondents answered with “agree” (51.6% of the question’s total) and 7 with “neither agree or disagree” (22.6% of the question’s total).

Self-assessment on policy learning: NORMATIVE TYPE

Unlike the results on cognitive learning, data regarding the change of beliefs in participants were highly indecisive. The four questions that inquired about this type of learning regarded the following topics:

- Change of beliefs on what is considered a public problem
- Change of beliefs on what are the right solutions
- Change of beliefs on the existing governance structures
- Change of beliefs on the personal vision on how the case area should be addressed in the future.

In total, of the possible 124 preferences, the majority was given to the “neither agree nor disagree” answer (53, 42.7% of total); 35 answers were given to the “disagree” option (28.2% of total); 24 to “agree” (19.4% of total); 8 to “strongly disagree” (6.5% of total) and the remaining 4 to the “strongly agree” option (3.2% of total).

Regarding the single questions, the response was divided as follows:

- What is considered a public problem – 13 respondents answered with “neither agree nor disagree” (41.9% of the question’s total); 10 responded with “disagree” (32.3% of the question’s total) and 6 with “agree” (19.4% of the question’s total)
- What are the right solutions – 16 respondents answered with “neither agree nor disagree” (51.6% of the question’s total) and 8 with “disagree” (25.8% of the question’s total)
- Existing governance structures – 11 respondents answered with “neither agree nor disagree” (35.5% of the question’s total); 9 with “disagree” (29% of the question’s total) and 6 with “agree” (19.4% of the question’s total)
- Personal vision on how the case area should be addressed in the

future – 13 respondents answered with "neither agree nor disagree" (41.9% of the question's total); 8 with "agree" (25.8% of the question's total) and 8 with "disagree" 8 (25.8% of the question's total).

As stated, data about normative learning resulted highly indecisive, with most of the preferences (42.7% of total) for the "neither agree nor disagree" answer, thus denoting the highest indeterminate point in the Likert scale. This was considered in the conclusion but could not be used in the final comparative analysis. Therefore, the QCA comparison was only based on cognitive learning.

Learning based on previous respondent knowledge

When cross-referencing data regarding the respondents' knowledge of their own country's politico-administrative system with the ones derived from the cognitive learning analysis, the following results emerged.

Of the 5 respondents that declared themselves as Full Experts, 2 stated that they "neither agreed nor disagreed" about possessing previous knowledge on the policy problem; 2 of them disagreed and 1 agreed. On being asked if they have gained new information, on a possible total of 20 preferences, 9 were given to the "agree" answer (45% of total); 8 to the "disagree" option (40% of total); 3 to the "neither agree nor disagree" one (15% of total).

Of the 10 respondents that identified as Specialists, 5 answered "agree" about possessing previous knowledge on the policy problem; 2 said "neither agree nor disagree"; 2 "disagree" and the remaining one "strongly agree". On being asked if they have gained new information, on the total of 40 possible preferences, 25 were given to the "agree" answer (62.5% ot total), and 10 to the "strongly agree" one (25% of total).

Of the 7 respondents that declared themselves as Generalists, 3 said that they "strongly disagreed" about possessing previous knowledge on the policy problem; 2 disagreed with the statement and 2 said that they "neither agree nor disagreed" with it. On being asked if they have gained new information, out of the possible 28 preferences, 15 were given to the "agree" option (53.6% of total) and 5 to the "strongly agree" one (17.9% of total).

Of the 9 respondents that declared to have an Average knowledge about their country's politico-administrative system, 3 strongly disagreed about possessing previous knowledge on the policy problem; 3 of them disagreed; 2 neither agreed nor disagreed and 1 strongly agreed. When asked if they have gained new information, on a possible total of 36 preferences, 19 were given to the "agree" option (52.8% of total); 10 to the "neither agree nor disagree" one (27.8% of total).

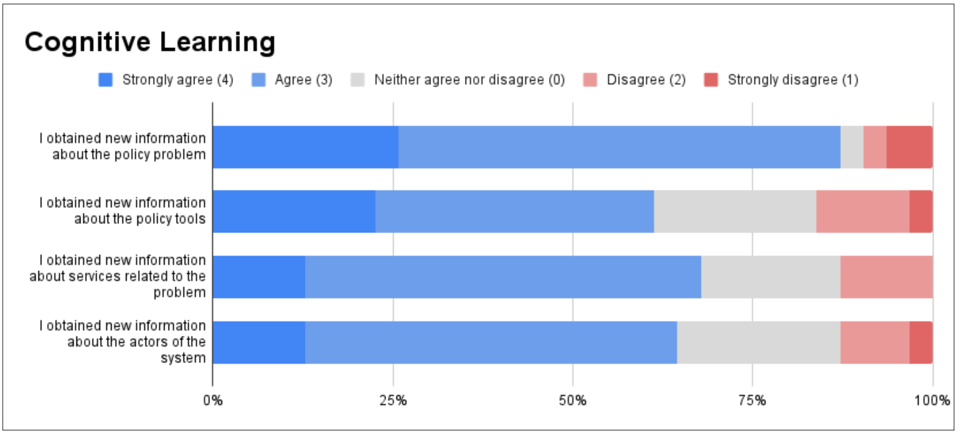
In general, these results suggest that those who identified as Full experts seemed to have gained more information about policy problems and policy tools. The Specialists appear to be the category to have learned the most among the four previous knowledge profiles, and they subject upon which they learned new information seem to be equally distributed across the options provided. The Generalists appear to have a balanced learning across the four options. Those with a previous Average knowledge seemed to have learned about policy problems, policy services and system actors.

4.2.4.3. Discussion on results from survey

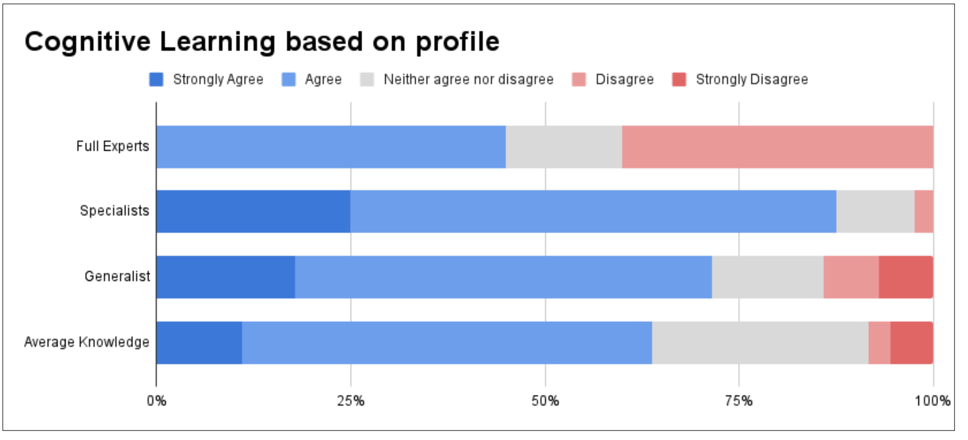
The results coming from the survey returns an image of data-centric policymaking as a process of coordination — and only partly of collaboration — starkly divided between data practitioners and the actors from the policy subsystem. Competencies related to non-traditional data (data practitioner) work separately from those who know about the political-administrative systems (actors of policy subsystem).

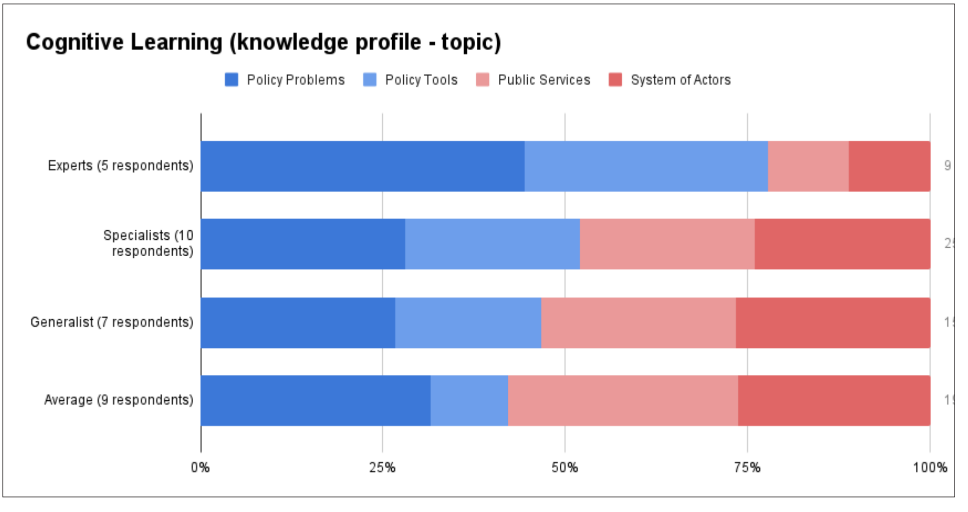
Data practitioners and actors of policy subsystem collaborate at the beginning and at the end of the data-centric policymaking processes (defining problems/needs and communicating results), but the central parts concerning data gathering, preparation and analysis are delegated to data practitioners (as expectable because of the difference of competences).

For the goal of this research, it can be claimed that people involved in the processes affirm to have gained new information on policy relevant topics by being involved in the process. This learning took place in the sample of 31 respondents which was almost equally distributed in terms of self-perceived knowledge of the policy problem or about the politico-administrative system. This learning is interesting since — despite of previous knowledge on the problem or public sector— most of the sample is made of professional with several years of working experience for the government (either internally or externally affiliated).



As the level of knowledge profiles, Full Experts did not provide fully clear data on their learning. The Specialists instead appeared as the category that learned the most among the four knowledge profiles, with equal learning across all the subjects presented. The Generalists also perceived this type of balanced learning. Those with a previous Average Knowledge of government reported cognitive learning but, in comparison to the other categories, they seem to have learned more about public services and actors in the systems.





Overall, these results tell that individuals involved in processes centred on non-traditional data perceive to have gained new policy-relevant knowledge (i.e., knowledge about policy relevant topics). They affirm that the gain of knowledge particularly regarding the policy problem/area addressed in the cases. However, what is interesting is that an overall learning was perceived by people with different competences and previous knowledge, which worked only partly together in a process with separated functions.

Considered together with the lack of clarity that the sample reported on their change of beliefs, what seems to emerge is that being involved in these processes actors reinforced their previous knowledge on policy-relevant topics. Experts of policy subsystems learned more about the policy problems, while people with average knowledge (mostly working in the development team) improved their learning in general.

4.2.5. QCA Comparison and comparative interpretation of cases

The final QCA data matrix presented here (Table 26) has been used as reference to derive the comparison. As expressed above, the normative learning (i.e., change of beliefs) could not be included as those survey data were not clear, so the comparison relies only on the self-reporting of cognitive learning (i.e., gain of new information).

Table 26. Data matrix of cases of data-centric policymaking

CASE	GOV	ACC	CULT	POL	LEAD	EXP	DIV	LEAR
SWE	0.33	0	0.33	1	1	0.33	1	0.67
BEL	0	0.33	0.33	0	0	0.67	0.67	1
ITA	0.67	0.67	0.33	0	0	1	0	0.33
NDL	0.33	0.67	0.67	1	1	0.67	1	0.67

The same self-reported score of cognitive learning for cases in SWE and NLD might appear surprising given the almost opposite structural conditions for data use characterizing these national contexts. In NDL, the use of data was facilitated both by the structural conditions and by the advanced system of data sharing and analysis that connotated the case. It should not be surprising, given this mixture, that most of actors in NDL case asserted to have gained new information on the policy problem they were facing. Further, the NDL group obtained new knowledge even if most people were specialists and experienced public servants. The NDL case thus identifies and advanced example of data-centric policymaking in terms of technology and expertise, potentially very impactful (especially for in the cities involved).

However, SWE (a group of similar size) also reported the same degree of learning, although it operated in a context whose structural conditions were the least favorable across the four cases. The case also adopted a very different approach to the use of data. In fact, the Swedish case could not freely integrate administrative data due to strict national regulations on privacy (OECD, 2019a). The actors had to resort to experimentations with data scientists on an anonymized sample, provided from the national government in Sweden or from associations involved.

The SWE group also featured a data ecosystem with several profiles that were not affiliated with the public sector — unlike the NDL case. SWE shows, as NDL, the presence of political support, leadership and diversity. This suggests that these meso/micro-level conditions were influential on policy-relevant knowledge, despite the very different structural conditions for data use.

This insight, taken together with what emerged from the survey, suggests that — in different conditions — the cases were able to gain knowledge out of data because of a good separation of roles and coordination, favored by each stakeholder interest and political support. Every individual was adding a competence and working in an organized system, designed differently depending by the needs and conditions.

A similar comparison, but on a different scale, might be made between BEL and ITA, that featured both small groups involved in data-centric policymaking at the city level. Both cases were collecting and analyzing data to increase information on a topic, rather than addressing a substantive issue. ITA was the case where learning was the least reported. Instead, almost all people in BEL said that new information on the policy problem could be gained.

Following what emerged in the other two cases, the reason might be found in political support, leadership and diversity. The ITA case included experienced public servants and took place within relatively favorable structural conditions for using data. However, the time of the process was tight and mostly confined to the staff of the Municipal department. For the BEL case — while no relevant financial investments were made to support the action —more time and public visibility were granted (also since it was a citizen science project and in partnership with a university). There was possibly a sense of larger investment of the municipality in the action, which fueled the efforts of the actors involved.

This reinforces the point that very contextual micro/meso conditions might enable form of policy-relevant knowledge even at a small, experimental scale. In the SWE case experimenting with data was not an aim per se but helped public organization to pursue their goals and develop an initiative intended to gain political support and momentum. The access to data was limited, so the data ecosystem had to develop an entrepreneurial approach to drive innovation by using non-traditional data. In this sense, some lesson can also be learned from the learning reported in the BEL case, which — also in adverse structural conditions — could collect data nonetheless useful to the municipality.

This suggests that, despite conditions, use of non-traditional data can create value if it aligns with the pre-existing will and action of public organizations (Klievink et al., 2017).

4.2.6. Overall discussion of Phase 2

As said, the most important result of phase 2 is that most of the individuals across the different cases analyzed perceived to have gained new information on policy-relevant topics while participating in the processes of data-centric policymaking.

It is therefore possible to say that policy work processes centered on the use of data for addressing a policy problem (i.e., data-centric policymaking processes) can generate policy-relevant knowledge. This knowledge seems to emerge differently in different contexts depending by a series of factor. These factors, the results suggest as a second important contribution of this phase 2, are prevalently happening at the micro/meso-level and seem to have less to do with the structural conditions of data sharing/use.

When learning was perceived, relevant conditions appear to be the leadership, the heterogeneity of the group, the policy support, and the possibility to work with data (and people) for a prolonged time. These elements might describe processes of using non-traditional data for policy characterized by a conscious political commitment toward making data the center of collective learnining process in institutional settings (Heikkila & Gerlak, 2013; Riche et al., 2021).

However, the unclarity of normative learning reported remains somehow an open question.

The stark division of role and competences and the relevance of leadership and coordination might suggest that these cases identified forms of data-centered coordinated actions, rather than spaces for exchange, social learning and reflection based on data. In a policy perspective this might be problematic, since existing framing are not challenged (Schneider & Ingram, 1993). Also, the fact that the specialists affirmed to have learned much more about the policy problem than any other topic questions the degree to which these cases of data-centric policymaking could bring new knowledge about services and actors. In essence, it is legitimate to ask: have experts of policy subsystem involved in the cases received any contextual information on the administrative/non-traditional contexts from where the data came from? These aspects appear relevant of further investigation and will be addressed in the next Chapter.

4.2.7. Limitations of phase 2

The research of phase 1 had proven challenging both because of the theoretical complexity of casing (Ragin, 1992) and the practical difficulties of collecting primary data. The challenge in empirical research — which remained concretely present for the entire process — was two-fold: from one side, the individuation of the appropriate examples that could reasonably account for cases of data-centric policymaking (among the myriad of public initiatives that experimented with non-traditional data); on the other side, the successful engagement of actors involved in these cases, to convince them participating to the data collection. This limitation resulted in few cases to be analyzed, which is a suboptimal circumstance for a comparative approach through QCA. To cope with scarcity of data and scarcity of theory, QCA was used as a data synthesis tool.

Other limitations were from understanding learning and practices (which are contextual) through a survey, which was a necessary condition given the scope of the research and the contemporary research circumstances. Survey is also notoriously a data collection method prone to social desirability bias (Bradburn et al., 2004). Methodologically, studying policy learning also presents the issue of defining the null-hypothesis (i.e., defining when learning does not happen) (Goyal & Howlett, 2018).

Chap. 5. Designing in Data-centric Policymaking

The reading this thesis brought to the field of data for policy was intentionally driven by the contemporary interest of design in policymaking, as explained in Chapter 1. The concept of data-centric policymaking, which the research used throughout its work, had the scope to understand the use of non-traditional data for policy under what appears to be an important perspective for both the design for policy and data for policy fields.

The empirical research had, on the other hand, highlighted some dimensions that appear relevant both in the discourse and in the practices of this field. Holding this knowledge, in this fifth Chapter I wish to abandon the impersonal tone recommended for scientific writing and adopt the first-person pronoun to propose how these two fields can converge (see Section 1.4.4). This is intended to signal that the arguments proposed here are prospective and personal. To the best of my knowledge, the field of data for policy appears to me still both young and far from the field of design for policy.

This fifth chapter is based on a paper I have first presented at the 5th Data for Policy International Conference, held online in 2020 (Leoni, 2020). That paper intended to answer the several comments I received during the official reviews by the Department of Design committees. Confronted with my work, professors were, perhaps understandably, concerned about the disciplinary positioning of the research. In short, they could see policymaking and data: but where was design in this PhD work in design?

To answer that question, the challenge to me was starting from the set of values I knew were discussed in design for policy (on the account of what I knew about "design for policy" through readings and my experience as part of the Design Policy Lab[104]) and connect these values with the empirical knowledge from discourse and practices in data for policy. To shed light on the value of design for policy, I will first clarify design as *designing*.

104 See http://www.designpolicy.eu/

5.1. Defining *designing* in design for policy

"*Design*", as used in "*design for policy*", carries a semantic which remains distant from the two primary meanings intrinsic to its etymology (Cassin et al., 2014). In my view, when design is discussed in "*design for policy*", it is usually not intended as the *output* of a design activity (e.g., a carefully devised architectural plan) or the *representation of that output* (e.g., a technical drawing of the building). Instead, it would be more correct to say that *design* in "design for policy" acquired the meaning both of *design process* (Friedman, 2000) and *designing* (Redstrom, 2017) (more below). In this sense, the proposition of "design for policy" implies a shift of meaning, from design as a problem-solving activity devoted to an output; to design as the exploration of a problem space (Bason, 2016). This perspective identifies a traditional line of investigation about design as a cognitive activity. The foundational reference to that would be the work of Herbert Simon and the bounded rationality approach, which also exerted a great deal of influence in policy studies (Peters & Zittoun, 2016, p. 8).

Simon advanced that actors in complex decision-making settings, as policymaking, have limited information and capacity to envisage solutions, therefore complete rational decision-making is impossible. What is maintained in this view is that the process of design unfolds *through* and *in relation with* the bounded empirical experience of the individual actors (Maffei, 2010).

The situatedness and path-dependency of the design action — with a strong focus on its specific cognitive processes and methods (i.e., design thinking) — were further explored in the seminal work of many authors such as Horst Rittel, Christopher Alexander, Buckminster Fuller and Donald Schön (Bousbaci, 2008; Legeby et al., 2018; Maffei, 2022). It could be argued that their works represent fundamental stepstones for "*design for policy*". In fact, they did not just outline the specificity of design as a process for problem-solving in complex settings, but also how design worked as a form of problem setting and inquiry (Cross, 2001). In line with this new line of inquiry, design scholars investigated the cognitive styles and steps of creative problem solving (Dorst & Cross, 2001), highlighting how the design process features a particular form of co-evolution between the problem at hand and the solutions proposed (Dorst, 2019). Design considered as a "reflective practice" enables an iterative passage from the problem to the solution space through abductive thinking and *reframing* of problems (Dorst, 2019). And while the thesis that *design thinking* could represent a separated or superior cognitive style was criticised (Bousbaci, 2008), the potential of design as reframing (van der Bijl-Brouwer, 2019) and heuristic in solving complex problems remained relevant subjects to "design for policy" (Bason & Austin, 2021; Considine, 2012).

5.2 The core values of "design for policy"

Design "as a reflective practice" and reframing both highlight another important meaning of design, implicit within "design for policy": design as *designing.* If in the problem-solving conception of design thinking early authors (above) intended design as a process to produce an outcome, *designing* instead more broadly refer to "*the overall orientation of the effort that produce that outcome*" (Redstrom, 2017, p. 39). The notion of *designing* — that emerged diffusely in the disciplinary discussion of design — implies that one or more paradigms drive the action of design and connect it with a desired result. These paradigms are shaped by the socio-technical and cultural conditions in which the design action takes place, where individual agency can result in several modes of design (Manzini, 2015, p. 40).

For "design for policy", this element is central because it implies that the design approach could make evident existing forms of "*designing*" in the public sector and policymaking (Bailey & Junginger, 2014). From a purely phenomenological perspective, this implies that new approaches and practices are brought into public sector settings:

> "Many governments are experimenting with design labs and are applying design methods or design thinking to their primary processes of policymaking, service delivery and decision-making. This new use of 'design' is characterised by a process of creativity and participation. The latter implies that more ideas from different sources are included. This can be labelled as design for policy. In this process, design is considered a way to better understand and structure a policy problem, rather than finding solutions for predefined goals" (van Buuren et al., 2020, p. 7)

It might be argued that this perspective implies a normative view about the capacity of a design to bring positive change in the public sector and policymaking. I do not wish to take any issues with these aspects here, as they would demand an entire dissertation to be debated thoroughly. Instead, I will focus on which are the values of that vision, as they are currently discussed in the design for policy field starting from experimental practices in the public sector. These values of "design for policy" could not be found in an explicit list, but have been subject of discussion from several scholars from different disciplinary angles — also outside the discipline of design (Ansell & Torfing, 2014; Buuren et al., 2020; Hermus et al., 2020). Here, I will address three potential contributions of design to policymaking[105]:

105 These areas were first advanced in the paper upon which this chapter is based (Leoni, 2020) and were further refined in an article (currently under peer-re-

Envisaging long-term transitions and future scenarios.
Design can support policymakers in the work of collaborative prefiguration (Maffei et al., 2020). Policymakers engaged in futures designing can envision the "radically new socio-economic and political paradigms" necessary to make policy decisions toward long-term transition (Irwin, 2015). This practice can highlight the construction of problems and publics, revealing layers of the system, mediating diverse expertise and data, and inviting broader participation (Kimbell, 2019).

Promoting governance innovation.
Design has revealed the potential of new ways of governance that are increasingly "relational, networked, interactive and reflective"; central to this is a reconceptualization of the relationship between governments and the people they serve (Ansell & Torfing, 2014; Bason, 2017).

Connecting decision-making to public services.
Design offers an opportunity to more tightly connect policies and services by conceiving services as more than the implementation of policy, but rather as the primary mechanism through which policies are realised and experienced (Junginger, 2013a). Further, by seeing policy instruments as the object of design, they can become 'meta-interfaces of policy delivery', revealing the potential consequences of policy instrument choices (Trippe, 2021). The materiality of design also offers a way to make otherwise abstracted systems of government – and their attendant social, cultural, and historical origins – tangible to citizens (Tunstall, 2007).

These three values are part of the design for policy proposition to bring innovation to policymaking. In the next section, I will link them with knowledge emerged from the research of Chapter 4 and substantiated them with examples.

5.3. Areas of convergence

Chapter 1 highlighted how data for policy seems to be taking the distance from a reified view of data inherited by the discourse on datafication. As part of that distancing topics, data quality and the contextuality of policy process are discussed. Chapter 4 highlighted that the discourse in data for policy seems to emphasize two main aspects: data ethics and data

view) wrote together with my team at the Design Policy Lab (Martina Carrato, Erin McAuliffe and prof. Stefano Maffei).

culture. Data ethics defines the ethical and responsible use of data from organizations (Hasselbalch, 2019). Data culture has been used to define, in general, both the capacity of a public sector organization to utilize data (Giest, 2017b) or several contextual attitudes that public sector organizations develop toward data and evidence, which shapes the practices of data use (Falk, 2021). Together with what emerged from the cases, especially by thinking about the unclarity of normative learning, I have reasons to believe that the most crucial contribution of designing in data-centric policymaking would be three-fold:

> **Challenging its existing framings**
> Data-centric policymaking should not only be a process of self-fulfilling prophecy centered on data, neither only a way to acquire public consensus and political support with an alluring technological initiative. It should primarily be a way to redefine, through evidence, how policies are made.
>
> **Making it more open (internally/externally)**
> Data-centric policymaking should not only be a coordinated action of a data ecosystem, but a platform for discussing proficiently about policy problems.
> This implies improving the internal dialogue and, more importantly, deciding which exogenous elements can be included.
>
> **Connecting its materiality and contextuality**
> Data-centric policymaking should be a lever to reconsider the stark division between decision and realization in policymaking.

For each of these needs I will articulate one area of convergence, presented below. Since the thesis has adopted learning as its conceptual-theorical lens (thus it shaped the concept of data-centric policymaking proposed), the areas are intended as learning areas.

5.3.1. Learning from Data-centric Anticipatory Governance

The first area I am proposing considers the potential of design to discuss possible futures in collaborative settings. Traditionally, governments use foresight and horizons scanning to understand the potential threats and unintended outcomes and to be better prepared to react to them (Ramos, 2014). In contrast with the more strategic and abstract approach of these methodologies, the design approach to futures discusses situated experiences through diegetic prototypes and visualization, in order to develop normative visions in participatory settings (Hines & Zindato, 2016). Kimbell (2019) connects the design approach to futures (i.e., *designing futures*) to *design for policy*, offering several examples on how the former contributes to make policy problems graspable and understandable, also by mediating between the expert and lay knowledge — and making visible existing discourses and framings (Kimbell, 2019, p. 134). In my review of the data for policy conference, the only paper related to design that I could find until 2020 discussed a case of design fiction method to understand public acceptance of IoT technologies deployment (Fig. 16). This approach can bring value in terms of public legitimization and sense-making with the stakeholders involved in the process and the probing of the publics' opinion during data technologies development. In the perspective of using non-traditional data for policymaking, these activities are not only relevant for public consultation, but can define pivotal questions affecting the design of these systems, as the authors themselves highlight (Jacobs et al., 2019):

> "It is important to ask such questions at the start of the process and as data are being collected, and consider why data needs to be collected at all, rather than just collecting it because it is there with usefulness to be decided later." (Jacobs et al., 2019, p. 5)

Fig. 18 Design fiction method for Participatory IoT research[106]

Predictive analytics is gaining ground in the public sector for public service delivery and operations. While some of these technological applications already provoked big political scandals because of algorithmic biases, causing national governments to lose money and public trusts

106 Picture from TrustsLens project (Imagination Lancaster). https://imagination.lancaster.ac.uk/update/made-up-rubbish-design-fiction-as-a-tool-for-participatory-internet-of-things-research/

(Henman, 2017), they will likely continue to be used for uncontroversial public operations, where they might also be useful. Arguably, these systems will never be able to address the prescriptive and value-based dimension required by policymaking, which demands not only to know what could happen, but also what should be done (Höchtl et al., 2016). However, in a data-centric policymaking perspective, these predictions and simulations can become an element for discussing policy problems with multiple stakeholders. Design for policy, realized in its futures designing approach, could contribute to rethink the relation between data and futures within an open and dialogical perspective. The creation of futures would not equal with prediction, but predictive systems could bring together different stakeholders' visions and improve the making of policies. Regardless of the few experimental practices in this sense, it is a vision of anticipatory governance which appears yet largely unrealized for policymaking as "[…] *because the key design challenge is to accomplish wide participation while overcoming inherent technological barriers posed by data usage (e.g. the technical competences and data literacy)."* (Maffei et al., 2020, p. 9). Design professionals could be employed in this area of convergence to add communication, data visualization and digital design skills to the visual exploration of data. Through these competences they can lower the threshold of different data literacies employing data as part of aesthetic narrative artifacts. As they work with simulations and tools devoted to informing policy workers in their practices, they can concretely contribute to a design for policy approach to data-centric policymaking.

For example, designers can support data visualization and collective visual exploration processes for forecasting tools and policy modelling, as the one presented by Dutt et al. (2019) in the Data for Policy Conference 2019 — a scenario-based simulation platform called "*Simulogue*", that integrates quantitative data (e.g., land-use) and qualitative data (e.g., interaction between stakeholders) to improve policy decision-making, enabling dialogue and reflection on futures strategies.

5.3.2. Learning from local/contextual knowledge

The second area I propose considers the value of design for policy to include local/contextual knowledge in the processes of data-centric policymaking. To gather user/local knowledge in an inspirational approach — i.e., to incorporate the perspective of users into policies/services (Hermus et al., 2020) has been considered a value of design for policy. In this sense, design for policy inherits the tradition of participatory and user-centred methods that are part of the design disciplinary tradition[107].

107 This is better exemplified by the tradition of Participatory Design (PD) and Co-design. PD emerged in Scandinavia During the 70s, as an approach to the de-

By carrying these approaches and methods into public sectors, design for policy enables a participatory setting that encourages — and seeks to capture — the individual vision of participants about future policies and services (van Buuren et al., 2020). For governments that strive to develop policies in complex governance systems and to maintain high level of legitimization and trust among their citizens, this approach appears much valuable.

I consider citizens participation/engagement as a fruitful common ground between "design for policy" and "data for policy", given the space this topic has in the data for policy discourse (Section 4.1.2). Its presence there should not surprise, given that digital technologies and data have, since years, been brought together in many ways and the possibility of collecting data through public engagement has been regarded as a non-traditional data source for policy (Ponti & Craglia, 2020). Citizens can become "activists" and collect data on relevant issues as part of participatory sensing activities or by being involved in data collection for scientific projects (Longo et al., 2017).

Design for policy can help to nurture data-centric practices by designing the socio-technical conditions that enable value co-creation in these practices (Ciuccarelli & Elli, 2019; Morelli et al., 2017). In this sense, the most specific way in which design can contribute to support this area of convergence is through design professionals with expertise and fieldwork experiences in co-design, participatory design and or meta-design for open innovation processes (Menichinelli & Valsecchi, 2016). To support these public settings can become a way through which design contributes to changing the narrative of datafication that dominates the data debate, showing new ways of approaching data value-creation and fostering new forms of data democracy (Morelli et al., 2018). Although ambitious, developing these different narratives can change data culture also in the public sector, which would otherwise design its approach to data based on private companies models (Falk, 2021).

velopment of information and management systems in the workplace (Ehn, 2017; Simonsen & Robertson, 2013). PD arise as an collaborative action-research approach, involving professional information scientists and workers in the Scandinavian Unions, with the political intention to allow their perspective affect the design of these systems (Holmlid, 2009). PD impacted several disciplines and defined a specific Scandinavian approach to information systems and has been regarded as one of the seminal experiences where design was brought into public settings. Co-design instead initially emerged from the private sector, particularly in relation to ergonomics, human factors, and user-centered design approaches applied to the development of products and digital interfaces (Sanders & Jan Stappers, 2017). At its initial stages, co-design was not concerned with the political dimension that characterized PD, but seek to improve market research by involving end-users (Sanders & Jan Stappers, 2017). Later it became a relevant aspect in the co-production paradigm (Cantu & Selloni, 2013).

It is my opinion that designing new forms of participatory collaboration and practices around data should be done also for more trivial reasons than enhancing democracy.

The knowledge of citizens or local communities should not be incorporated only because participation is considered ethical and desirable, but because local and contextual knowledge might be of the utmost importance to understand policy problems and their underlying causes in process of data-centric policymaking. As the NDL case presented in this research shows, the use of data in the public sector is being increasingly given a lot of political support. Even if we can expect these data not to be instrumentally used for policymaking (see Section 2.4), they exist as a base of evidence. This implies that the model of problem behind the data collected — and its potential bias — will remain part of an infrastructure of evidence. As said, policy problems depend on suppression of options of what constitutes a public issue (see Section 2.4).

In the case of organized crime, it might not be easy to decide what to measure. Should the government use an indicator that measures the effect of crime (e.g., arrests) or an indicator that measures the factors which might have caused crime (e.g., poor schooling conditions) (Stuurman et al., 2020)? Possibly, the government will just use the available data, which are the ones increasingly collected and processed by data-centric processes of data ecosystem.

In a design view of data-centric policymaking, the inclusion of knowledge from communities, street-level bureaucrats and citizens can be used to shape information systems for policymaking and address the bias on what constitutes a policy problem. The design competences mentioned above can contribute to that and drive practical experimentations for participatory indicators — a practice that has been already experimented (Van Den Homberg & Susha, 2018).

An illustrative example in that sense is proposed by Edwards et al. (2017), that worked in a participatory setting with ontologies (i.e., the formal conceptualizations that underlie an information system). The interdisciplinary research group Ensemble presented a paper in the Data for Policy Conference 2019 for a new methodology for flood management called "*Flooding Data Walk*". The intention of the project was to rethink the formalization of flooding management systems as a problem, by incorporating local/contextual knowledge to the model:

> “"The advantage of using semantic integration in the construction of a scenario library goes beyond the ability to interrogate data from different perspectives. It also allows for the incorporation of novel data, qualitative data and localized data with existing data sources to present richer, nuanced picture of places with the potential for more refined models of risk and uncertainty.”" (Edwards et al., 2017, p. 3)

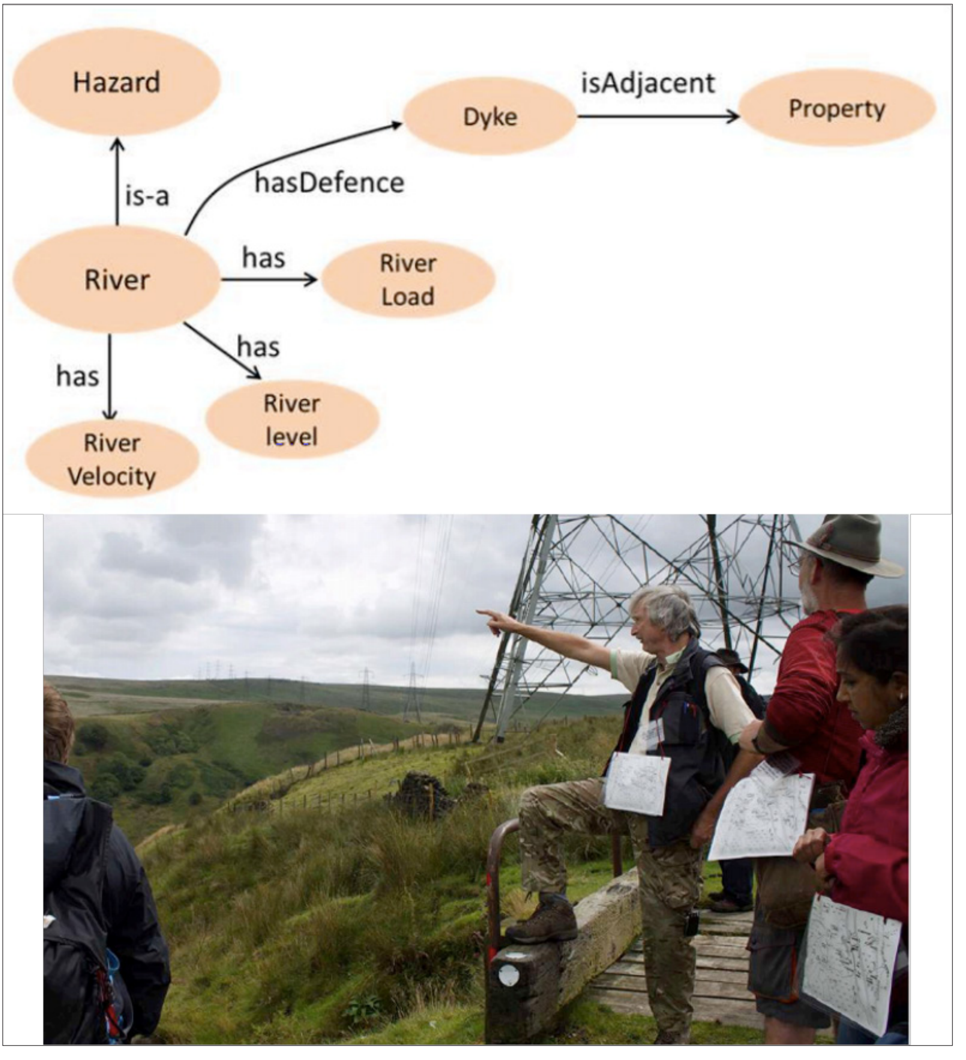


Fig. 19 A participatory methodology of semantic integration for flood management “Flooding Data Walk”[108]

108 Photo from Flooding Data Walk (Ensemble, Lancaster University) https://www.ensembleprojects.org/wp-content/uploads/2017/11/Conference-slides-Data-for-policy-conference-Sept-2017-1.pdf

5.3.3. Learning from Data-driven Service Systems

Digitalized services are a valuable source of data for policy, and administrative and services data are widely recognized as a relevant topic in data for policy (see Section 4.1.2) because of the characteristics of administrative data (see Section 1.3.2.3). For example, the Australian program "*Their Future Matters*", developed by the New South Wales Government Stronger Communities Investment Unit, illustrates how data can be included in ambitious plans for governmental action.

The program aimed to deliver improved outcomes for vulnerable children, young people and their families through a model based on a linked administrative dataset that could forecast the beneficiaries' categories most needing support, according to data collected by various government departments during service delivery. The project model allocates the financial resources based on that forecast and modifies services delivery accordingly. In this model, data are collected during service implementation and then monitored and service data feedback in the system.

Fig. 20. Their Futures Matter[109], NSW Communities & Justice

109 Src: https://www.theirfuturesmatter.nsw.gov.au/Home

The example shows the potential of service data to connect the various decisions and activities of policymaking. The perspective of analyzing data from services thus supports a crucial conceptual shift:

"Data analytics enable a closer working relationship between policy design and service delivery activities with a resulting shift from top-down implementation of public services to a user need led approach to design and delivery, based on an end to end understanding of a particular service journey, which can consequently increase its reach and effectiveness." (Ubaldi et al., 2019b, p. 21).

However, the data needed are often not available or in the same database, making data integration and further data collection necessary. Malomo and Sena (2017) described how Kent County Council Children's Service (UK) developed an integrated data model that could give insights into children's behaviours from how services were used. Service data (e.g., access to public libraries) can both lead to intervention on services (e.g., closing libraries to cut expenses) or addressing policy problems (e.g., lack of interest in reading). An inquiry into public data-driven service systems to collect and integrate the necessary data for policy could become an innovative practice of data-centric policymaking. The need to integrate data across public service systems represents an important opportunity for design for policy given its human-centred perspective of policymaking, that considers policies from the point of view of citizens' experience and public services (C. Buchanan et al., 2017; C. Buchanan & Junginger, 2014). The competencies of service designers would therefore be of value to the exploration of service systems and to support sense-making activities among stakeholders in data ecosystems (Gwilt et al., 2017). Further, service design and data science can be used to research and explore various dimensions of a public issue (Kunneman & Alves da Motta Filho, 2020). Such integrative practice appears to be interesting, and some experimental practices in public innovation labs indicates they could be further explored[110].

However, other than sense-making, I think that a design for policy perspective — realized through service design — would bring tremendous value to data for policy and the use of service data by considering the socio-technical and contextual materiality of the services from which data are collected. In many cases, data of interest for policy might not be available because standard data collection procedures are not in place. For these cases, the integration of data collection into existing services

110 See for example the work of UN Pulse Lab Jakarta (https://www.unglobalpulse.org/2021/09/applying-service-design-to-support-data-analytics-for-decision-making/) and other exampled in Leoni (2020).

could be realized with relatively cheap technologies by considering both specific policy needs and the contextualities of services. In my opinion, this was exemplified by the project "*Food Market 4.0 Dashboard*"[111], that seamless incorporated a digital tracking system into the daily operations of Municipal markets in the city of Milan to collect data on material flows in the urban food system, in line with the city policies on the subject (Bianchini & Maffei, n.d.).

In brief, the mutual reinforcement of design and data for policy in learning from service systems could be two-fold. On the one hand, the design for policy approach (through service design) provides competencies to map services' journeys starting from their users, thus allowing a data integration centred on the user and its interaction with public services.

It can also support a sense-making activity across emerging data ecosystems involved in data-centric policymaking, to align the internal intentions and support data integration. On the other hand, a data-centric policymaking perspective could be used to develop and integrat data collection as part of new or existing service systems, to become a source of evidence for policymaking.

5.4. Discussion of areas proposed

The three areas proposed above are prospective and based on the idea of transdisciplinary convergence (Morton et al., 2015). In other words, they define three ideal paradigms of designing in data-centric policymaking, that can unify different disciplines toward a common goal. The areas can be used to interpret existing practices or give indication on how new ones should be defined. Needless to say, how these practices will concretely unfold will depend on several contextual factors: the needs of the stakeholders involved, the extent to which it is possible to experiment, the financial resources, the available data, the technological means, the competences involved.

The micro-level practices happening under these areas are informed by a design for policy approach and give shape to forms of learning that contribute in different ways to data-centric policymaking. Ideally, if all the practices could be realized freely, each learning types would reinforce each other. Therefore, practices under area 1 will affect *policy goals and problems* for which data is used; practices under area 2 will provide *new measures and indicators* that non-traditional data can

111 The research project was developed by Polifactory (Department of Design), a group to which I am currently affiliated as researcher, as part of the EU-funded project Reflow (https://reflowproject.eu/).

support; and practices under area 3 will provide new and *more accurate data through improved data sharing and data collection* (see fig. 21).

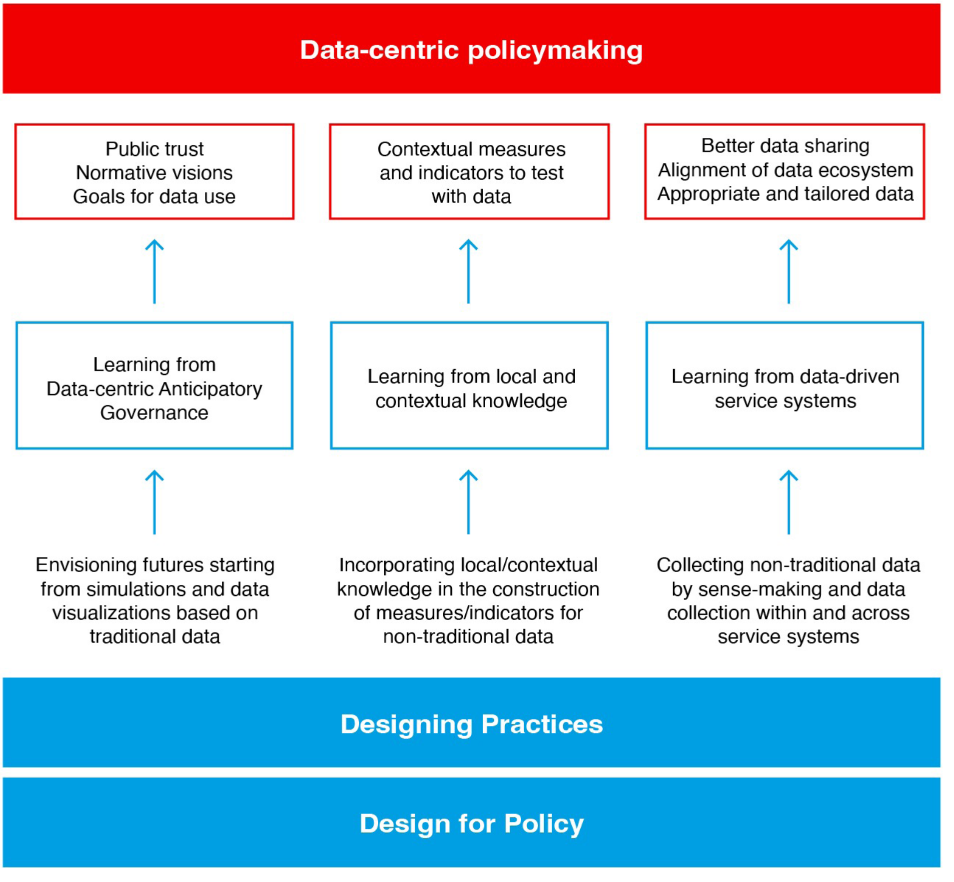


Fig. 21. Proposal of convergence for designing in data-centric policymaking

Although, it is unlikely that all these areas could be implemented together through relative practices in the same setting or experimentation, since it would require a total lack of constraints.

It might be imaginable that a design for policy approach can be realized — through the various design competences and expertise exposed above — at one or several specific stage of the data-centric policymaking process. Designing in data-centric policymaking would then have to work with the specific goals, conditions and actors of that phase of working with non-traditional data. Based on these notions, the proposal is further detailed with a focus on the stages of data-centric policymaking used for the empirical analysis of case studies in Chapter 4[112]. I will consider, based on the knowledge that the analysis gave about the actors involved in these stages, how designers can enable practices by considering the areas above and how they should work with the other actors at each stage:

Phase 1: Problem setting & defining needs

In this phase, the goal is to challenge existing frames of policy problems and drive new ideas on how data can be used for policymaking. Designers, data scientists and public servants will work together. The group effort will be devoted to developing practices by aligning the existing needs, political constraints and pressures, and the available non-traditional data. In this early phase, the definition of the policy problem (or how to address it) might still be affected to a certain degree. In other words, at this stage designers are confronted with the interest of a public sector organization in using non-traditional data — perhaps as it sees an opportunity to use available data sources, or because it has received funding/support to do so. However, the specific way the data-centric policymaking will unfold is not yet totally defined.

Designers can thus work to enable reframe through competences of Area 1. Data visualization, simulations and visual explorations of available datasets can be used to start group reflection about the policy problem and perspectives that are not considered. Competences of communication, visual and digital design could be used to develop these visualizations and facilitate the visual exploration. If these data are not available, low-fidelity diegetic prototypes can be created for the same purpose, by bringing to the table futures designing work. These prototypes should not be well-refined but aim at showing what are the existing frames of public problems in the mind of civil servants and data scientists. Regardless of the specific competences, the scope will be to shed light on aspects of the public issue that have not been considered, before the process of data-centric policymaking gets fixed on a certain framing.

Phase 2: Data gathering

The goal in this phase is to include domains of knowledge which would have been otherwise excluded by the process of data-centric policymaking. It is therefore a particularly delicate stage in terms of data ethics. In this phase designers, data scientists and public servants will still work together, although the group will shrink, because not all public servants of phase 1 will participate. Designers should use design research skills to understand what is relevant to measure, before the process gets more engineered. The aim is to avoid data gaps (Giest & Samuels,

112 Themselves based and on an ideal process of data science work (Crisan et al., 2021).

2020)— i.e., not letting underrepresented categories and forms of knowledge outside the process of data-centric policymaking. Competences of Area 2, co-design, and of Area 3, service design, can be used to holistically map the phenomenon that data will describe. The research can start from the policy beneficiaries, by researching which data collection points in a system would be sensible to include given their interaction with public services. Service design could also be used to map and bring the relevant data owners into the project, through co-design and participatory settings that highlight to them what value exists in creating data collaboratives with governments (Susha et al., 2017). Moreover, to develop co-design sessions that include people at close contact with citizens and several aspects of a policy problem (e.g., street-level bureaucrats) can help to tap into the local intelligence of these actors, to understand what are the policy-relevant dimensions of the problem that the data will have to describe.

Phase 3: Data preparation

This phase aims to support the data preparation process (data wrangling) (Crisan et al., 2021) and to understand, given the data collected in previous phases, which use can be made of the data collected from a policy perspective. This phase is normally delegated to the data scientists, whose job is to "clean" data and understand them. While these professionals might have a good understanding of the domain the data describe, and will be provided with requirements, the empirical research of this thesis shows that they might only possess a low knowledge of policy subsystems. Therefore, design competences from area 1 should be employed to support a collaborative explorative analysis of data, also involving the actors in the policy subsystem. Data visualization will be needed here, but — contrary to phase 1 — the scope will not be to challenge framing but to explore collectively the data to clarify their potential (Verstraete et al., 2021, p. 76).

Phase 4: Data mining and Analysis

In this phase, the goal is to define the most appropriate model of analysis, based on the technical advice of data scientists and the policy needs from actors in the policy subsystems.

As the previous phase, this phase is usually under the expertise of data scientists or other technical profiles. However, there are examples of experimentations that involve domain experts and even citizens in the co-creation of analysis models (as policy modeling) (Ronzhyn & Wimmer, 2021). Given some degree of open collaboration is possible in this phase, design competences from Area 2 can make accessible and transparent what happens in this phase to actors in policy subsystems.

Phase 5: Prototyping and deployment

In this phase, the goal is to support the realization of prototypal artifacts that turns data analysis into something usable and tangible: data visualizations, dashboards, a digital service. The point therefore is not just to read the data but to create long-lasting value for policymaking, by valorizing the work done in previous phases. In this sense,

design competences from area 3 could be employed to support the realization of a service that considers what is known about knowledge utilization in policymaking. The new service should make clear to policy decision-makers where the data comes from, the degree to which the non-traditional source is trustable, and make them easy to access and use. Designers will have to work with data scientists to understand the feasibility of engineering this prototype (also by trying to involve other technical competences in the conversation) and with public servants to understand the end-users needs. The work of this phase could bring momentum to the project and allow new public funding for scaling up the solutions. To do so, it would be important for designers to support and understand which are the organizations, in the public sector or in civil society, that can benefit from this solution.

Phase 6: Communication and Dissemination

In this phase, the goal is to share the work done throughout the data-centric policymaking process.

The group involved will be the same that started the process in phase 1. This phase can thus become a learning occasion for the main organization that drove the action, but also other public sector organizations that are interested in the outcome of the project. Here the designer, prevalently with competences from area 1, should once again challenge existing frames on what is the best solution, highlighting not only what was done, but "what could be done". It is therefore a stage that could be redefined from being just the sharing of a good practice or success story. It might become a phase of involvement, using the work done to understand the vision of other actors in the public system and how the work done can be brought in their contexts without ending up in dynamics of isomorphism.

Table. 27. Designing in data-centric policymaking (stages and goals)

Designing in data-centric policymaking	
Stages	**Goal**
Problem setting & defining needs Activities to understand how data can solve specific public problems or address public needs.	To challenge existing frames of policy problems and driving new ideas on how data can be used for policymaking.
Data gathering Activities for accessing, selecting and collecting relevant public/administrative datasets among the available ones.	To include domains of knowledge (local/contextual) which would have been otherwise excluded by the process of data-centric policymaking.
Data preparation Activities to prepare public/administrative datasets for future data analysis.	To explore collectively what use can be made of the data collected and what value exists for policy
Data mining and analysis Activities for data exploration and analysis through programming and techniques to obtain original information from data.	To define the most appropriate model of analysis, based on the technical advice of data scientists and the policy needs from actors in the policy subsystems
Prototyping and Deployment Activities for realizing data-centred outputs featuring innovative use of public/administrative data. As outputs, we can consider web/digital services, visualizations/simulations, a dashboard, etc.	To support the realization of prototypal artifacts that serves the needs of policy decision-makers at the various level. To drive new political commitment and build momentum around the work done in the whole process.
Communication and Dissemination Activities of discussion and reflections start from a data-centred output produced in the project. As output, we can consider web/digital services, visualizations/simulations, a dashboard, etc.	To turn a moment of results presentation into a moment of learning. To understand how the data-centric policymaking process can be adapted to other public sector contexts and needs.

Chap. 6. Conclusions

This concluding chapter retraces the path of the thesis and recaps its original knowledge contributions.

6.1. Research rationale and process

The rationale of this thesis was to develop an investigation under a different lens from the one entailed by the narrative of datafication that dominates the data debate (Couldry, 2017; Kallinikos, 2013). The reasons to develop other ways of looking at the potential of digital data for policymaking was driven by literature review. In fact, authors are increasing questioning that the innovation of data — as presented in the broader data debate — can unfold in the public sector and for policymaking (Durrant et al., 2018; Giest, 2017b; Klievink et al., 2017).

This discussion seems to be developing in an emerging field — defined in the research as data for policy — which debates the specificity of using of non-traditional data for policymaking and several associated challenges (e.g., ensuring data quality). The review interpreted this specificity as regarding the contextual, social and political nature of policymaking processes; and function of evidence within them (see Section 1.3.3). In general, the field also appeared to be lacking empirical evidence on practices, which was imputed to the theoretical perspective adopted, that seemed to be mostly driving prospective analysis, rather than contextual.

In light of this, the research proposed the concept of data-centric policymaking and adopted it as sensitizing concept upon which constructing a conceptual/theoretical framework, with the final intention to drive the empirical investigation. The concept was proposed out of an interest of this thesis (design for policy) and with the goal of providing guidelines for a converge of design for policy into data-centric policymaking. The empirical research was conducted through a two-phase research design, informed by the epistemology of Critical Realism. On the one level, the discourse of data for policy was analyzed to understand its orientation. (see Section 4.1) One the other level, the cases of data-centric policymaking were isolated and analyzed with primary data through the conceptual-theoretical framework of data-centric policymaking (see Section 4.2).

6.2 Research questions

The research intended to explore an emerging field with an original perspective. Therefore, the hypotheses driving the research questions were based on a first understanding of main themes in data for policy, and the interpretation of them through the research interests.

The hypotheses were supported by the concept of data-centric policymaking, which synthetised a certain stance of the research, based on an interpretation of the themes in data for policy (i.e., a specific perspective on what was the locus, the value, and the potential of innovation in data for policy). It seemed interesting to carry on the research thorugh three hypotheses:

- The discourse surrounding data-centric policymaking (H1)
- The relation between data-centric policymaking and policy-relevant knowledge (H2)
- The conditions affecting the relation between data-centric policymaking and policy-relevant knowledge (H3)

These three hypotheses intended to guide the research investigation and the employment of the concept of data-centric policymaking. Additionally, given the initial research interest, a fourth hypothesis formalised the will to investigate how design for policy and data for policy could converge into data-centric policymaking. The four sub-research questions formulated were therefore based on these four hypotheses, and were the followings:

- SRQ1. How is the surrounding discourse on data-centric policymaking characterised?

- SRQ2. Does data-centric policymaking affect policy-relevant knowledge at regional and local governance levels?

- SRQ3. What factors affect policy-relevant knowledge in data-centric policymaking at the regional and local governance levels?

- SRQ4. How can we converge a design for policy and data for policy approach into data-centric policymaking?

The answers to these fours SBRQs directly contributed to answer the main research question which was:

- MRQ. How can data-centric policymaking be realised?

The thesis answered to SRQ1 with the findings of Phase 1 (see Section 4.1.3). The discourse surrounding data-centric policymaking — identified with the discourse in data for policy and the voice of experts in this field — features Data Ethics (i.e., the ethical use of non-traditional data) and Data Culture (i.e., both the capacity of use data in public sector and the cultural approach to their use) as relevant and recurring elements. This suggest that such discourse, while still largely fragmented, might be acknowledging the specific challenges of data for policy. Both these topics arguably pertain to a dimension that is eminently value-laden and normative, addressing what is the right approach and use of non-traditional data for policymaking. At the same time, the field appears still largely fragmented and divided by several interests, of which a large part does not seem concerned with the influence of policymaking to the use of data and technologies, but with technological applications and their impact on public sector and government.

Then, this thesis answered to SRQ2 with the findings of Phase 2 (see Section 4.2.4.4). It was possible to collect data indicating a gain of new knowledge across 31 participants within four case studies of data-centric policymaking, which compiled a survey on learning self-assessment. The results show that cognitive learning (the gain of new information) on policy-relevant topics (e.g., the policy problem the group was dealing with) was perceived by participants as they took part to the cases. The strength of the findings relies in the diversity of the sample. Learning was perceived across the four cases even if they developed in different contexts and involved different profiles. However, it was impossible to determine clearly if normative learning happened (i.e., a change of pre-existing beliefs). The findings also suggest that the learning happened might have only increased information gain, depending on the initial knowledge of actors. Experts of the policy problem gained new knowledge and increased their expertise (on policy problems); while individuals without pre-existing knowledge gained a general increase of knowledge. This might suggest that data-centric policymaking process risk to become process of where existing beliefs gets reinforced. Literature in policy learning considers this scenario as a condition of non-learning (Pattison, 2018).

The thesis answered to SRQ3 with the findings of Phase 2 (see Section 4.2.4). The most important finding of this phase regarded the recognition that the gain of knowledge reported was not negatively influenced by the structural conditions of the national context in which the processes of data-centric policymaking took place. The comparison of cases have shown that conditions at the micro/meso level might have been the most influential. These included the presence of a stakeholder with a leadership role, the possibility and time to experiment given by the political support (not necessarily coupled by relevant financial investments). This reinforces how the appropriate use of data for policymaking highly depend by contextuality of policy process and its alignment with the goals of the organization collecting the data (Durrant et al., 2018; Klievink et al., 2017).

Finally, the thesis answered SRQ4 with an original proposal of areas of convergence between data for policy and design for policy in data-centric policymaking (see Chapter 5). The three area presented identify paradigms under which practices of designing can merge into data-centric policymaking, with clear indication on goals, roles, technologies; and the support of illustrative examples.

6.3 Research contributions

In the process of addressing the policy goals stated in section 1.7, the research developed several knowledge contributions, listed here:

A formalization of data for policy

This thesis proposed a formalization of data for policy and a first characterization of it, in line with other recent attempts (Suominen & Hajikhani, 2021). This proposal appears relevant for those accounts who have started to question a technocratic approach toward data in policymaking, as the one proposed in the datafication narrative (van der Voort et al., 2019). The formalization can be modified or challenged to discuss the extent to which data for policy is developing as an autonomous field.

Proposal of the concept of data-centric policymaking and articulation into a theoretical/conceptual framework

The thesis proposed a concept for reading practices in data ecosystems in relation to policymaking, starting from theory in political and policy studies (Chapter 2). The proposal was meant to address what were perceived as deficiencies of existing theoretical interpretation in the data for policy field, and their capacity to read policymaking contextually. The framework has been built upon the concepts of policy work, policy as practice and policy learning — relevant theoretical lenses of policymaking since long time (Dunlop et al., 2018). The framework offers an interpretative lens that can unify different disciplinary perspectives for understanding the relation between micro-level practices and technological innovation under a policy perspective.

Experimentation with QCA methodology for comparative multilevel analysis of data-centric policymaking cases

The thesis experimented with QCA, a comparative methodology well-known among scholars in political science, public administration and urban planning for its capacity to read complexity (Gerrits & Verweij, 2016). The methodology is usually not adopted in design research, and the way it was applied in this work was clearly documented as knowledge base for future research.

Analysis of the discourse in data for policy with multiple qualitative sources

The thesis developed a qualitative analysis of the discourse in data for policy, through several methods and sources (literature review, literature analysis and experts interviews). The codes groupings that emerged isolate a series of topics emerged in the literature analysis and provide a picture of this discourse. These groupings are made available in the Annexes.

Development of a list of 20 cases studies of innovative use of data in government/public sector at the regional and local governance level

Through desk research, the thesis developed a list of 16 desk research case studies of governmental initiatives centered on data in government/public sector, providing detailed information and data on each case. The list is made available in the Annexes. Four additional case studies were described in depth and also presented in the main section of thesis empirical research.

Proposal and articulation of four areas of convergence between data for policy and design for policy in data-centric policymaking

Based on the empirical knowledge from case studies, the thesis developed 3 areas of convergence between data and design for policy (Chapter 5), with detailed recommendations on how design practitioners can work together with other actors in data-centric policymaking process, under different paradigms relevant for design for policy.

Scientific writing

During its development, the thesis advanced the connection between data for policy and design for policy within publication in official scientific venues relevant for policy, design and data for policy (Maffei et al. 2019; Leoni, 2020).

6.4. Research Limitations

Difficulty in retrieving empirical data

The perspective taken by the thesis, and its focus on data for policy as made of contextual practices, would have demanded a better and more empirical intimacy with the case studies. This could be seen as a general trade-off of comparative analysis: the attempt to generalization implies more case studies and empirical data, thus more data collection. This overall process is therefore extremely time consuming, evermore during an historical period when public official have been burdened by the recent Covid-19 pandemic. Nonetheless, given more time and resources are provided, the comparative approach here proposed (if not part of its design) could be further developed by future research.

The novelty of the conceptual-theoretical framework

The framework proposed in Chapter 2 was crafted after a series of considerations on literature review and the research interests. This does not diminish its validity to read empirical data, but — if it will have to be carried into future research — it might benefit of being linked to other theoretical perspectives. Most of the existing frameworks appear to inquiry data for policy as a matter of organizational/technical barriers to technology adoption; or meso-level technological innovation. Critical data studies offer an incredibly interesting lens, but they seem to lack a specific focus on policymaking and a specific theoretical grounding. In synthesis, if more work has to be done on data for policy, researchers might need better theories that link micro-level socio-technical behaviours with the macro-level of institutions, policy ideas and paradigms. That challenge certainly will be interdisciplinary.

The breadth, novelty and interdisciplinarity of data for policy

This research was not dealing with one specific technology, technological application, or technological adoption/innovation but with data for policy. Keeping this breadth was necessary to link data for policy and design for policy, without being biased toward design.

The novelty and the interdisciplinary breadth of the topic limited the thesis, and it demanded several theoretical/methodological choices that could be certainly be questioned.

6.5 Recommendations for future research

Three main recommendations are made here for future research:

Develop micro-level qualitative research on policy workers and their use of data

The entrepreneurial nature of some cases shown how policy workers can be incredibly skillful in developing innovation, driving change, and gathering political support in unfavorable conditions. At the same time the findings suggested that their perspective might be not prone to change. This might be expectable since the tight dynamics of public sector work do not allow much space for reflection on existing frames. Future research should be made to understand how policy workers can be engaged in reframing and how this reframing be infrastructure and drive broader change. It is not hard to see why this topic greatly matters for design research (van der Bijl-Brouwer, 2019; Vink et al., 2021).

A broader investigation of cases through the QCA approach

A broader application of QCA methodology to a larger number of cases can be used to develop typologies of data-centric policymaking. More data will be needed, and the comparative framework could be modified by other theoretical perspectives.

Researching and testing practices under areas of convergence proposed The areas proposed should be further defined and tested through design research. Work between area 2 and area 3 can offer in particular great potential to understand how to bring design for policy out from the interesting, but possibly limited (Tõnurist et al., 2017) environments of public sector innovation lab. Further it can be used to focus on one of the strongholds of the design tradition, which is material culture. A work on data-driven public services would be also relevant given current public concerns about algorithmic decision-making.

6.6 Recommendations for policy workers

My recommendation for policy workers, in other words, the broad group of professionals and figures which work, even if they may not recognize it, contributes to the making of policy, would be three-fold. The use of words and phrases pertaining to the datafication narrative — as big data — is an understandable choice of communication, but should be done with critical sense. When words developed under a certain logic (private sector) (Diebold, 2013) travel outside their original domain, they might convey framing which do not necessarily serve the needs of domain in which they are landing. For data and digital technologies in public sector this implies potential threats to democratic values (Falk, 2021). The current interest on digital tracing technologies — sparked by Covid-19 pandemic — certainly awaken the political interest in data. It is therefore of the outmost importance that those involved in the field are cognizant of the different semantic frameworks and develop practices under narratives that make sense in their contextuality (Calzada & Almirall, 2019). This brings to the second recommendation that is to avoid isomorphism, or the tendency to replicate solutions which worked elsewhere. Under financial constraints it might make sense to replicate something that worked elsewhere, but every context is different and might strive to develop different solution. In line with that, the public sector should use what it has in abundancy: authority and time (at least until the end of the political mandate). This time should be used to focus on specific experimentations toward a political goal. Once good demonstrators have been built, they can be used as levers to obtain more funding and deliver public value.

From data-driven innovation, to data-centric policymaking, to people-centred data practices

We should be reminded that information and data are independent by their support; and their quality can be understood only in use (Floridi, 2014). While the mainstream narrative of the data debate might have decided that the bits and bytes transmitted and stored in contemporary technologies are data — and of the utmost relevance — the modern meaning of the word tells us that they can be considered as such only once they become part of a knowledge-creation process (Rosenberg, 2013). In contemporary socio-technical systems, the relevant data are those people decide to collect for their measurement. In policy, these decisions are— or should be — part of a political and social learning process (Hall, 1993) where is decided what is a policy problem. This research suggested to start considering these elements in our discussion on policy, design and data.

Bibliography

6, P. (2018). Frames and framing in policymaking. In H. K. Colebatch & R. Hoppe (Eds.), Handbook on policy, process and governing (pp. 275–294). Edward Elgar Publishing.

Agar, J. (2003). The government machine: a revolutionary history of the computer. MIT press.

Agger, A., & Sørensen, E. (2014). Designing collaborative policy innovation: Lessons from a Danish municipality. In Public Innovation Through Collaboration and Design. https://doi.org/10.4324/9780203795958

Akrich, M., Callon, M., & Latour, B. (2002). the Key To Success in Innovation Part I: The Art of OF Interessement. International Journal of Innovation Management, 06(02), 187–206. https://doi.org/10.1142/s1363919602000562

Alford, J., & Head, B. W. (2017). Wicked and less wicked problems: A typology and a contingency framework. Policy and Society, 36(3), 397–413. https://doi.org/10.1080/14494035.2017.1361634

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine, 16(7), 7–16.

Andersson, J., Ahgren, B., Axelsson, S. B., Eriksson, A., & Axelsson, R. (2011). Organizational approaches to collaboration in vocational rehabilitation—An international literature review. International Journal of Integrated Care, 11(4). https://doi.org/10.5334/ijic.670

Andrews, C. J. (2007). Rationality in policy decision making. In Handbook of Public Policy Analysis (pp. 161–171). Routledge.

Androutsopoulou, A., & Charalabidis, Y. (2018). A Framework for Evidence Based Policy Making Combining Big Data, Dynamic Modelling and Machine Intelligence. ACM International Conference Proceeding Series, 575–583. https://doi.org/10.1145/3209415.3209427

Ansell, C., & Torfing, J. (2014). Public innovation through collaboration and design. Routledge.

Arnaboldi, M., & Azzone, G. (2020). Data science in the design of public policies: dispelling the obscurity in matching policy demand and data offer. Heliyon, 6(6), e04300. https://doi.org/10.1016/j.heliyon.2020.e04300

Athey, S. (2017). Beyond prediction: Using big data for policy problems. Essay Special Section Prediction, 485(February), 483–485. https://doi.org/10.1126/science.aal4321

Atto, J., Lord, J., & Potter, C. (2015). Predictive analytics: the science of non-compliance. https://quarterly.blog.gov.uk/2015/01/27/predictive_analytics/

Azzone, G. (2018). Big data and public policies: Opportunities and challenges. Statistics and Probability Letters, 136, 116–120. https://doi.org/10.1016/j.spl.2018.02.022

Baffour, B., King, T., & Valente, P. (2013). The Modern Census: Evolution, Examples and Evaluation. International Statistical Review, 81(3), 407–425. https://doi.org/10.1111/insr.12036

Bailey, J., & Lloyd, P. (2016). The introduction of design to policymaking: Policy Lab and the UK government. DRS2016: Future-Focused Thinking, 9(Christensen 2015), 1–14. https://doi.org/10.21606/drs.2016.314

Bailey, S., & Junginger, S. (2014). Designing vs Designers.

Bangemann, M., Fonseca, E. C., Davis, P., Benedetti, C., Gyllenhammer, P., Hunsel, L., Lescure, P., Maragall, P., Thorn, G., & Velazquez-Gastelu, C. (1994). Bangemann report recommendations to the European Council. Europe and the Global Information Society, Brussels, Belgium.

Bankins, S., Denness, B., Kriz, A., & Molloy, C. (2017). Innovation Agents in the Public Sector: Applying Champion and Promotor Theory to Explore Innovation in the Australian Public Service. Australian Journal of Public Administration, 76(1), 122–137. https://doi.org/10.1111/1467-8500.12198

Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., & Viboud, C. (2016). Big data for infectious disease surveillance and modeling. Journal of Infectious Diseases, 214(Suppl 4), S375–S379. https://doi.org/10.1093/infdis/jiw400

Barbero, M., Coutuer, J., Jackers, R., Moueddene, K., Renders, E., Stevens, W., Toninato, Y., van der Peijl, S., & Versteele, D. (2016). Big data analytics for policy making. In European Commission & Deloitte. https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf

Barocas, S., & Selbst, A. D. (2018). Big Data's Disparate Impact. SSRN Electronic Journal, 671, 671–732. https://doi.org/10.2139/ssrn.2477899

Bartels, K. P. R. (2018). Policy as practice. In Handbook on Policy, Process and Governing (pp. 68–88). https://doi.org/10.1177/089124196025003001

Bason, C. (2016). Design for policy. Routledge.

Bason, C. (2017). Leading public design: Discovering human-centred governance. Policy Press.

Bason, C., & Austin, R. D. (2021). Design in the public sector : Toward a human centred model of public governance. Public Management Review, 00(00), 1–31. https://doi.org/10.1080/14719037.2021.1919186

Bason, C., & Schneider, A. (2014). Public design in global perspective: Empirical trends. In Design for policy (pp. 23–40). Gower Publishing.

Basurto, X., & Speer, J. (2012). Structuring the Calibration of Qualitative Data as Sets for Qualitative Comparative Analysis (QCA). Field Methods, 24(2), 155–174. https://doi.org/10.1177/1525822X11433998

Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274–279. https://doi.org/10.1177/2043820613513390

Batty, M. (2017). Data about cities: Redefining big, recasting small. Data and the City, 31–43. https://doi.org/10.4324/9781315407388

Beele, E., Reyniers, M., Aerts, R., & Somers, B. (2022). a citizen science weather station network in Leuven , Belgium. 2050(May), 1–43.

Beitin, B. K. (2012). Interview and sampling: How many and whom. In J. F. Gubrium, J. A. Holstein, A. B. Marcasti, & K. D. McKinney (Eds.), The SAGE Handbook of Interview Research: The Complexity of the Craft (Second Edi). https://doi.org/10.4135/9781452218403.n17

Bekkers, V., Edelenbos, J., & Steijn, B. (2011). Linking innovation to the public sector: Contexts, concepts and challenges. In Innovation in the public sector (pp. 3–32). Springer.

Bennett, C. J., & Howlett, M. (1992). The lessons of learning: Reconciling theories of policy learning and policy change. Policy Sciences, 25(3), 275–294.

Berg-Schlosser, D., & Meur, G. De. (2012). Comparative Research Design: Case and Variable Selection. In B. Rihoux & C. Ragin (Eds.), Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques (pp. 19–32). SAGE Publications. https://doi.org/10.4135/9781452226569.n2

Berg-Schlosser, D., Meur, G. De, Rihoux, B., & Ragin, C. (2012). Qualitative Comparative Analysis (QCA) as an Approach. In B. Rihoux & C. Ragin (Eds.), Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques (pp. 1–18). SAGE Publications. https://doi.org/10.4135/9781452226569.n1

Berry, D. M. (2011). The computatonal turn: Thinking about the digital humanities. Culture Machine, 12, 1–22.

Berry, F. S., & Berry, W. D. (1999). Innovation and diffusion models in policy research. Theories of the Policy Process, 169(5).

Bettencourt, L. M. A. (2014). The uses of big data in cities. Big Data, 2(1), 12–22. https://doi.org/10.1089/big.2013.0042

Bianchini, M., & Maffei, S. (n.d.). Food markets as circular digital hub. Prototyping new enabling ICT infrastructure solutions and city-scale data practices for urban food systems. International Journal of Food Design.

Blomkamp, E. (2018). The Promise of Co-Design for Public Policy. Australian Journal of Public Administration, 77(4), 729–743. https://doi.org/10.1111/1467-8500.12310

Bogner, A., & Menz, W. (2009). The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction. In U. Flick (Ed.), Interviewing Experts. SAGE Publications. https://doi.org/10.1057/9780230244276_3

Borins, S. (2001). Encouraging innovation in the public sector. Journal of Intellectual Capital, 2(3), 310–319.

https://doi.org/10.1108/14691930110400128

Borrás, S. (2011). Policy learning and organizational capacities in innovation policies. Science and Public Policy, 38(9), 725–734. https://doi.org/10.3152/030234211X13070021633323

Boschetti, A. (2006). Bourdieu's Work on Literature: Contexts, Stakes and Perspectives. Theory, Culture & Society, 23(6), 135–155. https://doi.org/10.1177/0263276406069779

Bousbaci, R. (2008). "Models of man" in design thinking: The "bounded rationality" episode. Design Issues, 24(4), 38–52. https://doi.org/10.1162/desi.2008.24.4.38

Bowen, G. A. (2019). Sensitizing concepts. SAGE Research Methods Foundations. SAGE Publications. Doi, 10(9781526421036788000).

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information Communication and Society, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Bradburn, N., Sudman, S., & Wasink, B. (2004). Asking questions: The Definitive Guide to Questionnaire Design— For Market Research, Political Polls, and Social and Health Questionnaires, Revised Edition. John Wiley & Sons. https://doi.org/10.1016/0198-9715(90)90052-u

Brennen, J. S., & Kreiss, D. (2016). Digitalization. In K. B. Jensen & R. T. Rothenbuhler, E.W. Pooley, J.D. Craig (Eds.), The International Encyclopedia of Communication Theory and Philosophy (pp. 329–338). https://doi.org/10.4324/9780203736319-36

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? Journal of Econometrics, 65(1), 83–108.

Bridgman, P., & Davis, G. (2003). What Use is a Policy Cycle? Plenty, if the Aim is Clear. Australian Journal of Public Administration, 62(3), 98–102. https://doi.org/10.1046/j.1467-8500.2003.00342.x

Brinkmann, S. (2013). Qualitative Interviewing. In Qualitative Interviewing. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199861392.001.0001

Brinkmann, S. (2018). The Interview. In N. K. Denzin & Y. S. Lincoln (Eds.), The SAGE Handbook of Qualitative Research (Fifth Edit, pp. 997–1038). SAGE Publications.

Broeders, D., Schrijvers, E., & Hirsch Ballin, E. (2017). Big Data and Security Policies: Serving Security, Protecting Freedom. WRR Policy Brief. https://www.wrr.nl/publicaties/policy-briefs/2017/01/31/big-data-and-security-policies-serving-security-protecting-freedom

Brunner, R. D. (1991). The policy movement as a policy problem. Policy Sciences, 24(1), 65–98.

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in numbers: How does data-driven decision-making affect firm performance? International Conference on Information Systems 2011, ICIS 2011, 1, 541–558. https://doi.org/10.2139/ssrn.1819486

Buchanan, C., & Junginger, S. (2014). (Service) Design and Policy Making. In D. Sangiorgi, A. Prendiville, & A. RIckett (Eds.), Mapping and developing Service Design Research in the UK (pp. 48–49). Arts & Humanities Research Council.

Buchanan, C., Junginger, S., & Terrey, N. (2017). Service design in policy making. Designing for Service: Key Issues and New Directions. London.

Buchanan, R. (1992). Wicked Problems in Design Thinking. Design Issues, 8(2), 5–21. http://www.jstor.org/stable/1511637

Byrne, D., & Ragin, C. C. (2009). The Sage handbook of case-based methods. Sage Publications.

Cairney, P. (2019). Understanding public policy: theories and issues. Bloomsbury Publishing.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. Transportation Research Part C: Emerging Technologies, 26, 301–313. https://doi.org/10.1016/j.trc.2012.09.009

Calzada, I., & Almirall, E. (2019). Barcelona's Grassroots-led Urban Experimentation: Deciphering the 'Data Commons' Policy Scheme. SSRN Electronic Journal, 2019, 11–12. https://doi.org/10.2139/ssrn.3359268

Cantu, D., & Selloni, D. (2013). From engaging to empowering people: a set of co-design experiments with a service design perspective.

Caplan, N. (1975). The Use of Social Science Knowledge in Policy Decisions at the National Level: A Report to Respondents.

Carlsson, L. (2017). Policy Science at an Impasse: A Matter of Conceptual Stretching? Politics and Policy, 45(2), 148–168. https://doi.org/10.1111/polp.12196

Cashore, B., & Howlett, M. (2007). Punctuating which equilibrium? Understanding thermostatic policy dynamics in Pacific Northwest forestry. American Journal of Political Science, 51(3), 532–551. https://doi.org/10.1111/j.1540-5907.2007.00266.x

Cassin, B., Apter, E., Lezra, J., & Wood, M. (2014). Disegno. In Dictionary of Untranslatables. A Philosophical Lexicon. (pp. 224–226). Princeton University Press.

Castaldi, C., & Dosi, G. (2010). Technical change and economic growth: some lessons from secular patterns and some conjectures on the current impact of ICT. In Innovation and Economic Development. Edward Elgar Publishing.

Ceschin, F. (2014). Sustainable Product-Service System. Between Strategic Design and Transition Studies. In International Business Research (Vol. 7, Issue 7). https://doi.org/10.5539/ibr.v7n7p62

Charmaz, K., Thornberg, R., & Keane, E. (2018). Evolving grounded theory and social justice inquiry. In The SAGE Handbook of Qualitative Research.

Chong, C. Y., & Kumar, S. P. (2003). Sensor networks: Evolution, opportunities, and challenges. Proceedings of the IEEE, 91(8), 1247–1256. https://doi.org/10.1109/JPROC.2003.814918

CISCO. (2020). Cisco Annual Internet Report (2018–2023).

Ciuccarelli, P., & Elli, T. (2019). Beyond Visualization.

Clarke, A., & Craft, J. (2019). The twin faces of public sector design. Governance, 32(1), 5–21. https://doi.org/10.1111/gove.12342

Colebatch, H. K. (2005). Policy analysis, policy practice and political science. Australian Journal of Public Administration, 64(3), 14–23. https://doi.org/10.1111/j.1467-8500.2005.00448.x

Colebatch, H. K. (2010). Giving accounts of policy work. Working for Policy, 31–43.

Colebatch, H. K. (2015). Knowledge, Policy and the Work of Governing. Journal of Comparative Policy Analysis: Research and Practice, 17(3), 209–214. https://doi.org/10.1080/13876988.2015.1036517

Colebatch, H. K., & Hoppe, R. (2018a). Introduction to the Handbook on Policy, Process and Governing. In Hal K. Colebatch & R. Hoppe (Eds.), Handbook on Policy, Process and Governing (pp. 1–13). Edward Elgar Publishing Limited.

Colebatch, H. K., & Hoppe, R. (2018b). Policy as Authoritative Choice. In Handbook on Policy, Process and Governing (pp. 109–111). Edward Elgar Publishing.

Colebatch, H. K., Hoppe, R., & Noordegraaf, M. (2010a). Understanding policy work. Working for Policy, 1, 11–25.

Colebatch, H. K., Hoppe, R., & Noordegraaf, M. (2010b). Working for Policy. In Working for Policy. Amsterdam University Press. https://doi.org/10.5117/9789089642530

Concilio, G., & Pucci, P. (2021). The Data Shake: An Opportunity for Experiment-Driven Policy Making. In G. Concilio, P. Pucci, L. Raes, & G. Mareels (Eds.), The data shake. Opportunities and obstacles for urban policy making: Vol. PolimiSpri (pp. 3–18). Springer.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. Social Science Research, 59, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Considine, M. (2012). Thinking Outside the Box? Applying Design Theory to Public Policy. Politics and Policy, 40(4), 704–724. https://doi.org/10.1111/j.1747-1346.2012.00372.x

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in Society, 1(1), 104–126. https://doi.org/10.1007/BF03177550

Cooper, R. (2019). Design research – Its 50-year transformation. Design Studies, 65(xxxx), 6–17. https://doi.org/10.1016/j.destud.2019.10.002

Corbin, J., & Strauss, A. (2008). Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory. In Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory. SAGE Publications. https://doi.org/10.4135/9781452230153

Couldry, N. (2017). The Myth of Big Data. In M. T. Schäfer & K. van Es (Eds.), The Datafied Society: Studying Culture through Data (pp. 235–239). Amsterdam University Press. https://doi.org/10.25969/mediarep/12467

Crato, N., & Paruolo, P. (2019). The Power of Microdata: An Introduction. In Data-Driven Policy Impact Evaluation (pp. 1–14). Springer, Cham.

Crisan, A., Fiore-Gartland, B., & Tory, M. (2021). Passing the Data Baton : A Retrospective Analysis on Data Science Work and Workers. IEEE Transactions on Visualization and Computer Graphics, 27(2), 1860–1870. https://doi.org/10.1109/TVCG.2020.3030340

Crompvoets, J., Vancauwenberghe, G., Ho, S., Masser, I., & De Vries, W. T. (2018). Governance of national spatial data infrastructures in Europe. International Journal of Spatial Data Infrastructures Research, 13, 253–285.

Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science. Design Issues, 17(3), 49–55. https://doi.org/10.1162/074793601750357196

Cross, N. (2018). Developing design as a discipline. 4828. https://doi.org/10.1080/09544828.2018.1537481

Cruickshank, J. (2003). Introduction. In J. Cruickshank (Ed.), Critical Realism: The difference that it makes. (pp. 1–14). Routledge.

d'Albergo, E. (2010). Governance, Participation and in-between. Pôle Sud, n° 32(1), 93. https://doi.org/10.3917/psud.032.0093

Dahlbom, B. (2002). From systems to services. Internet document available at http://www. viktoria. se/~ dahlbom.

Dalton, C., & Thatcher, J. (2014). What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data.' Society and Space, 29.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings, 1644, 97–104. https://doi.org/10.1063/1.4907823

DeLeon, P. (2006). The historical roots of the field. In The Oxford handbook of public policy (pp. 39–57).

della Porta, D. (2008). Comparative analysis: Case-oriented versus variable-oriented research. In D. della Porta & M. Keating (Eds.), Approaches and Methodologies in the Social Sciences: A Pluralist Perspective. Cambridge University Press. https://doi.org/10.1017/CBO9780511801938.012

Deloitte & The Lisbon Council. (2020). Study on data tools and technologies used in the public sector to gather, store, manage, process, get insights and share data. https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2020-07/Study on data tools and technologies used in the public sector to gather%2C store%2C manage%2C process%2C get insights and share data – ANNEX Case studies.pdf

Deloitte & The Lisbon Council. (2021). Study on public sector data strategies policies and governance. Data analytics for Member States and Citizens. https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2020-07/Study on public sector data strategies policies and governance.pdf

Deloitte, & the Lisbon Council. (2021). Study on public sector data strategies policies and governance.

Department for Digital Culture Media & Sport. (2019). National Data Strategy. https://www.gov.uk/guidance/national-data-strategy

Department for Digital Government. (2019). NL DIGITAAL: Data Agenda Government. https://www.nldigitalgovernment.nl/wp-content/uploads/sites/11/2019/04/data-agenda-government.pdf

Desouza, K. C., & Jacob, B. (2017). Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration and Society, 49(7), 1043–1064. https://doi.org/10.1177/0095399714555751

Desrosières, A. (1998). The politics of large numbers: A history of statistical reasoning. Harvard University Press.

Diebold, F. X. (2013). A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline, Second Version. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2202843

Döringer, S. (2020). 'The problem-centred expert interview'. Combining qualitative interviewing approaches for investigating implicit expert knowledge. International Journal of Social Research Methodology, 00(00), 1–14. https://doi.org/10.1080/13645579.2020.1766777

Dorst, K. (2019). Co-evolution and emergence in design. Design Studies, 65, 60–77. https://doi.org/10.1016/j.destud.2019.10.005

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. Design Studies, 22(5), 425–437.

Dubois, V. (2015). The Fields of Public Policy. In M. Hilgers & E. Mangez (Eds.), Bourdieu's Theory of Social Fields: Concepts and Applications (pp. 199–220). Routledge.

Dunleavy, P. (2016). 'Big data' and policy learning. Evidence-Based Policy Making in the Social Sciences: Methods That Matter, January, 143–168. https://doi.org/10.2307/j.ctt1t89d4k.15

Dunlop, C. A., & Radaelli, C. M. (2013). Systematising Policy Learning: From Monolith to Dimensions. Political Studies, 61(3), 599–619. https://doi.org/10.1111/j.1467-9248.2012.00982.x

Dunlop, C. A., & Radaelli, C. M. (2017). Learning in the bath-tub: The micro and macro dimensions of the causal relationship between learning and policy change. Policy and Society, 36(2), 304–319. https://doi.org/10.1080/14494035.2017.1321232

Dunlop, C. A., & Radaelli, C. M. (2018a). Does Policy Learning Meet the Standards of an Analytical Framework of the Policy Process? Policy Studies Journal, 46(00), S48–S68. https://doi.org/10.1111/psj.12250

Dunlop, C. A., & Radaelli, C. M. (2018b). The lessons of policy learning: Types, triggers, hindrances and pathologies. Policy and Politics, 46(2), 255–272. https://doi.org/10.1332/030557318X15230059735521

Dunlop, C. A., & Radaelli, C. M. (2020). Policy Learning in Comparative Policy Analysis. Journal of Comparative Policy Analysis: Research and Practice, 00(00), 1–22. https://doi.org/10.1080/13876988.2020.1762077

Dunlop, C. A., Radaelli, C. M., & Trein, P. (2018). Introduction: The family tree of policy learning. In Learning in public policy (pp. 1–25). Springer.

Dunn, W. N. (2017). Public Policy Analysis. An Integrated Approach. Routledge.

Dunn, W. N. (2018). 'Stage'theories of the policy process. In Handbook on policy, process and governing. Edward Elgar Publishing.

Durrant, H., Barnett, J., & Rempel, E. S. (2018). Realising the benefits of integrated data for local policymaking: Rhetoric versus reality. Politics and Governance, 6(4), 18–28. https://doi.org/10.17645/pag.v6i4.1586

Dutt, Vaibhav; Sil, Srijan; Krishna, Harsha; Palavalli, B. (2019). Imagining Futures – A generative scenario-based methodology to improve planning and decision-support systems for policymakers. Data for Policy. https://doi.org/10.5281/zenodo.3066348

Dye, T. R. (2013). Understanding Public Policy. Pearson.

Easton, D. (1957). An Approach to the Analysis of Political Systems. World Politics, 9(3), 383–400. http://www.jstor.org/stable/2008920

Ehn, P. (2017). Learning in participatory design as I found it (1970–2015). In Participatory Design for Learning (pp. 7–21). Routledge.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Ratandeep Suri, V., Tsou, A., Weingart, S., & Sugimoto, C. R. (2014). Big Data, Bigger Dilemmas: A Critical Review. Journal of the Association for Information Science and Technology, 66(8), 1523–1545. https://doi.org/10.1002/asi

El-Taliawi, O. G., Goyal, N., & Howlett, M. (2021). Holding out the promise of Lasswell's dream: Big data analytics in public policy research and teaching. Review of Policy Research, 38(6), 640–660. https://doi.org/10.1111/ropr.12448

Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2020). Data-driven urban management: Mapping the landscape. Journal of Urban Management, 9(2), 140–150. https://doi.org/10.1016/j.

jum.2019.12.001

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001

European Commission. (2020). A European Strategy for Data. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN

European Parliament. (2019). Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN

Eurostat. (2021). The European System of Accounts ESA 2010 — interactive version. https://doi.org/10.2785/793816

Falk, P. (2021). Towards a Public Sector Data Culture: Data as an Individual and Communal Resource in Progressing Democracy. In The Data Shake (pp. 35–45). Springer, Cham.

Feast, L. (2018). Constitutional realism and sustainability: Lessons learned from a systemic design investigation of New Zealand's democratic system.

Federal Data Strategy Team. (2019). Federal Data Strategy 2020 Action Plan. https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf

Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. International Journal of Social Research Methodology, 20(2), 181–194. https://doi.org/10.1080/13645579.2016.1144401

Flick, U. (2018). Triangulation in Data Collection. In U. Flick (Ed.), The SAGE Handbook of Qualitative Data Collection (pp. 257–544). SAGE Publications.

Floridi, L. (2008). Data. In W. A. Darity (Ed.), International Encyclopedia of the Social Sciences. Macmillan.

Floridi, L. (2012). Big data and their epistemological challenge. Philosophy and Technology, 25(4), 435–437. https://doi.org/10.1007/s13347-012-0093-4

Floridi, L. (2014). Big Data and information quality. In The philosophy of information quality (pp. 303–315). Springer.

Floridi, L. (2016). Mature Information Societies—a Matter of Expectations. Philosophy and Technology, 29(1), 1–4. https://doi.org/10.1007/s13347-016-0214-6

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative Inquiry, 12(2), 219–245. https://doi.org/10.1177/1077800405284363

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (2016). Big Data and Social Science. A Practical Guide to Methods and Tools. In Statistics in the Social and Behavioral Sciences Series. CRC Press. https://doi.org/10.18637/jss.v078.b02

Freeman, R. (2006). Learning in public policy. In The Oxford handbook of public policy.

Friedman, K. (2000). Creating design knowledge: from research into practice. IDATER 2000 Conference, May, 05–32.

Fuchs, C. (2005). The internet as a self-organizing socio-technological system. Cybernetics & Human Knowing, 12(3), 37–81.

Fuglsang, L., & Pedersen, J. S. (2011). How common is public sector innovation and how similar is it to private sector innovation? In Innovation in the Public Sector (pp. 44–60). Springer.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Garson, G. D. (2010). Public information technology and e-government: a historical timeline. In C. M. Shea & D. Garson (Eds.), Handbook of Public Information Systems (pp. 7–28). CRC Press Taylor & Francis Group.

Gelli, F., & Tedesco, C. (2001). Goverments and modes of governance in the U.S. and European cities. The change in urban policies. Area-Based Initiatives in Contemporary Urban Policy, May.

Gerrits, L., & Verweij, S. (2013). Critical realism as a meta-framework for understanding the relationships between complexity and qualitative comparative analysis. Journal of Critical Realism, 12(2), 166–182. https://doi.org/10.1179/rea.12.2.p663527490513071

Gerrits, L., & Verweij, S. (2016). Qualitative Comparative Analysis as a Method of Evaluating Complex Cases. Zeitschrift Für Evaluation, April, 7–22.

Giest, S. (2017a). Big data analytics for mitigating carbon emissions in smart cities: opportunities and challenges. European Planning Studies, 25(6), 941–957. https://doi.org/10.1080/09654313.2017.1294149

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

Giest, S. (2018). Policy learning in times of big data analytics: The challenges of skill-based outsourcing. In Knowledge, Policymaking and Learning for European Cities and Regions. Edward Elgar Publishing.

Giest, S., Miotto, J. M., & Kraaij, W. (2021). The value of data matching for public poverty initiatives: a local voucher program example. Data & Policy, 3(May). https://doi.org/10.1017/dap.2021.2

Giest, S., & Mukherjee, I. (2018). Behavioral instruments in renewable energy and the role of big data: A policy perspective. Energy Policy, 123(May), 360–366. https://doi.org/10.1016/j.enpol.2018.09.006

Giest, S., & Ng, R. (2018). Big Data Applications in Governance and Policy. Politics and Governance, 6(4), 1–4. https://doi.org/10.17645/pag.v6i4.1810

Giest, S., & Samuels, A. (2020). 'For good measure': data gaps in a big data world. Policy Sciences, 53(3), 559–569. https://doi.org/10.1007/s11077-020-09384-1

Godin, K., Stapleton, J., Kirkpatrick, S. I., Hanning, R. M., & Leatherdale, S. T. (2015). Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Systematic Reviews, 4(1), 138. https://doi.org/10.1186/s13643-015-0125-0

Göransdotter, M. (2021). Transitional Design Histories: Present-ing History in Design. Nordes, 187–196.

Goyal, N., & Howlett, M. (2018). Lessons Learned and Not Learned: Bibliometric Analysis of Policy Learning. In Learning in Public Policy (pp. 27–49). Springer International Publishing. https://doi.org/10.1007/978-3-319-76210-4_2

Goyal, N., & Howlett, M. (2019). Framework or metaphor? Analysing the status of policy learning in the policy sciences. Journal of Asian Public Policy, 12(3), 257–273. https://doi.org/10.1080/17516234.2018.1493768

Gray, J. (2015). Towards a Genealogy of Open Data. SSRN Electronic Journal, 1–38. https://doi.org/10.2139/ssrn.2605828

Greckhamer, T., Koro-Ljungberg, M., Cilesiz, S., & Hayes, S. (2008). Demystifying interdisciplinary qualitative research. Qualitative Inquiry, 14(2), 307–331. https://doi.org/10.1177/1077800407312049

Green, H. E. (2014). Use of theoretical and conceptual frameworks in qualitative research. Nurse Researcher, 21(6), 34–38. https://doi.org/10.7748/nr.21.6.34.e1252

Grin, J., & Loeber, A. (2007). Theories of Policy Learning: Agency, Structure, and Change. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), Handbook of Public Policy Analysis (pp. 201–220). CRC Press Taylor & Francis Group.

Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. Global Journal of Flexible Systems Management, 18(3), 203–229. https://doi.org/10.1007/s40171-017-0159-3

Gwilt, I., Prendiville, A., & Mitchell, V. (2017). Making sense of Data through Service Design: opportunities and reflections.

Hagen, L., Keller, T. E., Yerden, X., & Luna-Reyes, L. F. (2019). Open data visualizations and analytics as tools for policy-making. Government Information Quarterly, 36(4), 0–1. https://doi.org/10.1016/j.giq.2019.06.004

Hall, P. A. (1993). Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain. Comparative Politics, 25(3), 275. https://doi.org/10.2307/422246

Hamelink, C. J. (1997). New information and communication technologies, social development and cultural change (Vol. 86). United Nations Research Institute for Social Development Geneva.

Hand, D. J. (2011). Statistics: A Very Short Introduction. Journal of Applied Statistics, 38(4), 865–865. https://doi.org/10.1080/02664760903075598

Hand, D. J. (2020). Dark Data. Why what you don't know matters. Princeton University Press.

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L., & Goldstein, H. (2017). Challenges in administrative data linkage for research. Big Data and Society, 4(2), 1–12. https://doi.org/10.1177/2053951717745678

Hartwig, M. (2016). On the presuppositions and origins of the philosophy of critical realism. In Enlightened Common Sense (pp. 19–40). Routledge.

Harvey, D. L. (2009). Complexity and case. The SAGE Handbook of Case-Based Methods, 15–38.

Hasselbalch, G. (2019). Making sense of data ethics. The powers behind the data ethics debate in European policymaking. Internet Policy Review, 8(2), 0–19. https://doi.org/10.14763/2019.2.1401

Head, B. W. (2008). Three lenses of evidence-based policy. Australian Journal of Public Administration, 67(1), 1–11. https://doi.org/10.1111/j.1467-8500.2007.00564.x

Head, B. W. (2022). Debates in Public Policy—Problem Framing, Knowledge and Interests. In Wicked Problems in Public Policy (pp. 7–19). Springer.

Heclo, H. (1974). Modern social politics in Britain and Sweden. Yale University Press.

Heikkila, T., & Gerlak, A. K. (2013). Building a conceptual approach to collective learning: Lessons for public policy scholars. Policy Studies Journal, 41(3), 484–512. https://doi.org/10.1111/psj.12026

Henman, P. (2017). The computer says "DEBT": Towards a critical sociology of algorithms and algorithmic governance. Data for Policy 2017: Government by Algorithm?, 1–6. https://espace.library.uq.edu.au/view/UQ:8963b02

Hermus, M., van Buuren, A., & Bekkers, V. (2020). Applying design in public administration: A literature review to explore the state of the art. Policy and Politics, 48(1), 21–48. https://doi.org/10.1332/030557319X15579230420126

Hilbert, M., & Lopez, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(April), 60–65.

Hilgers, M., & Mangez, E. (2015). Introduction to Pierre Bourdieu's theory of social fields. In M. Hilgers & E. Mangez (Eds.), Bourdieu's Theory of Social Fields: Concepts and Applications (pp. 1–36). Routledge.

Hillgren, P.-A., Seravalli, A., & Emilson, A. (2011). Prototyping and infrastructuring in design for social innovation. CoDesign, 7(3–4), 169–183.

Hines, A., & Zindato, D. (2016). Designing Foresight and Foresighting Design: Opportunities for Learning and Collaboration via Scenarios. World Future Review, 8(4), 180–192.

Hirschheim, R., & Klein, H. K. (2011). Tracing the history of the information systems field. The Oxford Handbook of Management Information Systems: Critical Perspectives and New Directions, 16–61.

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

Holmlid, S. (2009). From participatory design to service design. First Nordic Conference on Service Design and Service Innovation, 14.

Hood, C. C., & Margetts, H. Z. (2007). The Tools of Governments in the Digital Age. Palgrave Macmillan.

Hooghe, L., & Marks, G. (2016). Community, scale, and regional governance: a postfunctionalist theory of governance, Volume II. Oxford University Press.

Hoppe, R. (1999). Argumentative Turn. Science and Public Policy, 26(3), 201–210.

Hoppe, R. (2010). The governance of problems: Puzzling, powering, participation. In The governance of problems. The Policy Press.

Hoppe, R. (2011). The governance of problems: Puzzling, powering, participation. In The governance of problems. Policy Press.

Hoppe, R. (2018a). Choice v. incrementalism. In Handbook on Policy, Process and Governing. Edward Elgar Publishing.

Hoppe, R. (2018b). Rules-of-thumb for problem-structuring policy design. Policy Design and Practice, 1(1), 12–29. https://doi.org/10.1080/25741292.2018.1427419

Howlett, M. (2009). Policy analytical capacity and evidenceAbased policyAmaking: Lessons from Canada. Canadian Public Administration, 52(2), 153–175.

Howlett, M. (2014). From the “old” to the “new” policy design: Design thinking beyond markets and collaborative governance. Policy Sciences, 47(3), 187–207. https://doi.org/10.1007/s11077-014-9199-0

Howlett, M. (2019). The Policy Design Primer. Routledge. https://doi.org/10.4324/9780429401046

Howlett, M., & Cashore, B. (2020). Public policy: definitions and approaches. In A Modern Guide to Public Policy (pp. 10–21). Edward Elgar Publishing.

Howlett, M., & Giest, S. (2013). The policy-making process. In Routledge Handbook of Public Policy (pp. 17–28). Routledge.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. Big Data and Society, 3(2), 1–7. https://doi.org/10.1177/2053951716674238

Irwin, T. (2015). Transition design: A proposal for a new area of design practice, study, and research. Design and Culture, 7(2), 229–246.

Ishii, A., & Okubo, A. (2014). Path dependence and paradigm shift: How Cetacean scientists learned to develop management procedures that survived the controversial Whaling regime. Review of Policy Research, 31(4), 257–280. https://doi.org/10.1111/ropr.12076

Jacobs, N., Edwards, P., Markovic, M., Cottrill, C. D., & Salt, K. (2019). Public Sector Internet of Things Deployments: Value, Transparency, Risks and Challenges.

Jann, W., & Wegrich, K. (2007). Theories of policy cycle. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), Handbook of Public Policy Analysis. Theories, Politics and Methods (pp. 43–62). Taylor & Francis.

Janssen, K., & Dumortier, J. (2003). Towards a European framework for the re-use of public sector information: A long and winding road. International Journal of Law and Information Technology, 11(2), 184–201. https://doi.org/10.1093/ijlit/11.2.184

Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. Journal of Organizational Computing and Electronic Commerce, 26(1–2), 3–13. https://doi.org/10.1080/10919392.2015.1124005

Jasanoff, S., & Kim, S. H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. Minerva, 47(2), 119–146. https://doi.org/10.1007/s11024-009-9124-4

Jégou, F., & Manzini, E. (2008). Collaborative services. Social innovation and design for sustainability (Vol. 1). Polidesign.

Jia, J., Durrani, T. S., & Chen, J. (2018). The Innovation Waves in Mobile Telecommunication Industry. IEEE Engineering Management Review, 46(3), 63–74. https://doi.org/10.1109/EMR.2018.2863253

Jones, P. H. (2014). Social Systems and Design (Vol. 1). https://doi.org/10.1007/978-4-431-54478-4

Julier, G. (2017). Economies of Design. SAGE Publications.

Junginger, S. (2013a). Design and innovation in the public sector: Matters of design in policy-making and policy implementation. Annual Review of Policy Design, 1(1), 1–11.

Junginger, S. (2013b). Design and Innovation in the Public Sector: Matters of Design in Policy-Making and Policy Implementation. Annual Review of Policy Design, 1(1), 1–11. http://ojs.unbc.ca/index.php/design/article/view/542

Junginger, S. (2014). Towards Policy-Making as Designing: Policy-Making Beyond Problem-Solving and Decision-Making. Design for Policy, 57–69.

Junginger, S. (2017). Design Research and Practice for the Public Good: A Reflection. She Ji, 3(4), 290–302. https://doi.org/10.1016/j.sheji.2018.02.005

Kallinikos, J. (2013). The allure of big data. Mercury Magazine, 2(3), 40–43.

Kamateri, E., Panopoulou, E., Tambouris, E., Tarabanis, K., Ojo, A., Lee, D., & Price, D. (2015). A comparative analysis of tools and technologies for policy making. In Policy Practice and Digital Science: Integrating Complex Systems, Social Simulation and Public Administration in Policy Research. https://doi.org/10.1007/978-3-319-12784-2_7

Karch, A. (2021). Innovation and Diffusion: Connecting Theory and Method. In Methods of the Policy Process (pp. 181–207). Routledge.

Kettl, D. F. (2016). Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. Governance, 29(4), 573–579. https://doi.org/10.1111/gove.12211

Kimbell, L. (2016). Design in the Time of Policy Problems. DRS2016: Future-Focused Thinking, 9(November). https://doi.org/10.21606/drs.2016.498

Kimbell, L. (2019). What if there were more policy futures studios? Journal of Futures Studies, 23(4), 129–136. https://doi.org/10.6531/JFS.201906_23(4).0014

Kimbell, L., & Bailey, J. (2017). Prototyping and the new spirit of policy-making. CoDesign, 13(3), 214–226.

Kingdon, J. W. (1984). Agendas, alternatives, and public policies (Vol. 45). Little, Brown Boston.

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. Big Data and Society, 1(1), 1–12. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. SAGE Publications. https://doi.org/doi:10.4135/9781473909472

Kitchin, R. (2014c). The real-time city? Big data and smart urbanism. GeoJournal, 79(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kitchin, R. (2021). The Data Revolution: A critical analysis of big data, open data and data infrastructures. Sage.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data and Society, 3(1), 1–10. https://doi.org/10.1177/2053951716631130

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. Information Systems Frontiers, 19(2), 267–283. https://doi.org/10.1007/s10796-016-9686-2

Knill, C., & Tosun, J. (2012). Public Policy: A New Introduction. In Palgrave Macmillan (Vol. 73, Issue 1). https://doi.org/10.1111/j.1467-9299.1995.tb00817.x

Kohoutek, J., Nekola, M., & Novotný, V. (2013). Conceptualizing policy work as activity and field of research. Central European Journal of Public Policy, 7(1), 28–58.

Kohoutek, J., Nekola, M., & Veselý, A. (2018). Policy work and its significance for public policy. Handbook on Policy, Process and Governing.

Kroes, N. (2013). Big data for Europe. Retrieved from European Union Website: Http://Europa. Eu/Rapid/Press-Release_SPEECH-13-893_en. Htm.

Kunneman, Y., & Alves da Motta Filho, M. (2020). Data Science for Service Design: An exploration of methods. DRS2020: Synergy, 5, 11–14. https://doi.org/10.21606/drs.2020.331

Kuonen, D. (2004). Data Mining and Statistics: What is the Connection? The Data Administration Newsletter 30.0, 2004(October), 1–6. www.TDAN.com

Lafortune, G., & Ubaldi, B. (2018). OECD 2017 OURdata Index. 30. https://doi.org/https://doi.org/https://doi.org/10.1787/2807d3c8-en

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety. In Gartner. META Group Research Note,. https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...

Lanza, G. (2021). Data-Related Ecosystems in Policy Making: The PoliVisu Contexts. The Data Shake Opportunities and Obstacles for Urban Policy Making, 91.

Legeby, M., McAleenan, P., Andersson, H., & Holmlid, S. (2018). Guiding the welfare state towards a cocreative and explorative mindset: When a crisis is an opportunity. Service Design Proof of Concept, June.

Leoni, F. (2020). Design for policy in data for policy practices. Exploring potential convergences for policy innovation. Data for Policy 2020 5th International Conference. Zenodo. https://doi.org/http://doi.org/10.5281/zenodo.3988176

Linåker, J., & Runeson, P. (2021). How to enable collaboration in open government data ecosystems: A public platform provider's perspective. EJournal of EDemocracy and Open Government, 13(1), 1–30. https://doi.org/10.29379/jedem.v13i1.634

Lindblom, C. E. (1959). The science of "Muddling" through. Public Administration Review, 12(2), 79–88. https://doi.org/http://www.jstor.org/stable/973677 .

Lindblom, C. E. (1979). Still Muddling, Not Yet Through Author. Public Administration Review, 39(6), 517–526.

Lipsky, M. (1971). Street-Level Bureaucracy and the Analysis of Urban Reform. Urban Affairs Review, 6(4), 391–409. https://doi.org/10.1177/107808747100600401

Lipsky, M. (2010). Street-level bureaucracy: Dilemmas of the individual in public service. Russell Sage Foundation.

Longo, J., Kuras, E., Smith, H., Hondula, D. M., & Johnston, E. (2017). Technology Use, Exposure to Natural Hazards, and Being Digitally Invisible: Implications for Policy Analytics. Policy and Internet, 9(1), 76–108. https://doi.org/10.1002/poi3.144

Lowi, T. J. (1964). American business, public policy, case-studies, and political theory. World Politics, 16(4), 677–715.

MacFeely, S. (2018). Big Data and Official Statistics (Issue December). https://doi.org/10.4018/978-1-5225-7077-6.ch002

Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. International Review of Administrative Sciences, 83(1_suppl), 120–135. https://doi.org/10.1177/0020852316640058

Maffei, S. (2010). Il progetto è una pratica clinica. La prospettiva del design. In T. Fabbri (Ed.), L'organizzazione: concetti e metodi (pp. 265–282). Carocci.

Maffei, S. (2022). Expanding the Galaxy. Designing More-than-Human Futures. Diid, 1(75), 12–19. https://doi.org/10.30682/diid7521a

Maffei, S., Leoni, F., & Villari, B. (2020). Data-driven anticipatory governance. Emerging scenarios in data for policy practices. Policy Design and Practice, 3(2), 123–134. https://doi.org/10.1080/25741292.2020.1763896

Malomo, F., & Sena, V. (2015). Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy and Internet, 9(1), 7–27. https://doi.org/10.1002/poi3.141

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The next frontier for innovation, competition, and productivity. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Manzini, E. (2015). Design, when everybody designs: An introduction to design for social innovation. MIT press.

Marchi, G. De, Lucertini, G., Tsoukiàs, A., Marchi, G. De, Lucertini, G., Tsoukiàs, A., & Policy-making, F. E. (2012). From Evidence-Based Policy-Making to Policy Analytics. 236(1), 15–38.

Margetts, H. (1999). Information technology in government: Britain and America. Routledge.

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. Nature Publishing Group.

Margolin, V. (2016). Design Research: What is it? What is it for? DRS2016: Future-Focused Thinking, 1, 1–11. https://doi.org/10.21606/drs.2016.9

Markussen, T. (2017). Disentangling 'the social' in social design's engagement with the public realm. CoDesign, 13(3), 160–174. https://doi.org/10.1080/15710882.2017.1355001

Marshall, M. N. (1996). The key informant technique. Family Practice. https://doi.org/10.1093/fampra/13.1.92

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly, 37(3), 101284. https://doi.org/10.1016/j.giq.2018.01.006

Maton, K. (2008). Habitus. In M. Grenfell (Ed.), Pierre Bourdieu: Key Concepts (pp. 49–65). Acumen.

Maxwell, J. (2012). A Realist Approach for Qualitative Research. In A Realist Approach for Qualitative Research.

May, P. J. (1992). Policy Learning and Failure. Journal of Public Policy, 12(4), 331–354. https://doi.org/10.1017/S0143814X00005602

May, P. J., Sapotichne, J., & Workman, S. (2006). Policy coherence and policy domains. Policy Studies Journal, 34(3), 381–403. https://doi.org/10.1111/j.1541-0072.2006.00178.x

Maybin, J. (2015). Policy Analysis and Policy Know-how: A Case Study of Civil Servants in England's Department of Health. Journal of Comparative Policy Analysis: Research and Practice, 17(3), 286–304. https://doi.org/10.1080/13876988.2014.919738

Maybin, J. (2016). Knowledge sources. In Producing Health Policy. Knowledge and Knowing in Government Policy Work (pp. 38–63). Palgrave Macmillan.

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work and Think. Houghton Mifflin Harcourt.

Mayring, P. (2004). Qualitative content analysis. A Companion to Qualitative Research, 1(2), 159–176.

Mcafee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review, October, 1–9.

McFadgen, B., & Huitema, D. (2017). Stimulating learning through policy experimentation: A multi-case analysis of how design influences policy learning outcomes in experiments for climate adaptation. Water (Switzerland), 9(9). https://doi.org/10.3390/w9090648

McGann, M., Blomkamp, E., & Lewis, J. M. (2018). The rise of public sector innovation labs: experiments in design thinking for policy. Policy Sciences, 51(3), 249–267. https://doi.org/10.1007/s11077-018-9315-7

McGregor, S. L. T. (2018). Introduction and Research Questions. In Understanding and Evaluating Research: A Critical Guide. SAGE Publications.

Menichinelli, M., & Valsecchi, F. (2016). The meta-design of systems: how design, data and software enable the organizing of open, distributed, and collaborative processes. Systems&design: Beyond Processes and Thinking, 518–537.

Mergel, I. (2016). Big Data in Public Affairs Education. Journal of Public Affairs Education, 22(2), 231–248. https://doi.org/10.1080/15236803.2016.12002243

Merriam, S. B., & Tisdell, E. J. (2016). Qualitative Research A guide to Design and Implementation. In The Jossey-Bass Higher and Adult Education Series.

Meyer, E. T., Crowcroft, J., Engin, Z., & Alexander, A. (2017). Data for Public Policy. Policy and Internet, 9(1), 4–6. https://doi.org/10.1002/poi3.147

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. Big Data and Society, 7(2). https://doi.org/10.1177/2053951720948087

Milan, S., & Velden, L. van der. (2016). The Alternative Epistemologies of Data Activism. Digital Culture & Society, 2(2), 57–74. https://doi.org/10.14361/dcs-2016-0205

Mitchell, T. M. (2006). The discipline of machine learning (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning ....

Moore, M. H. (1995). Creating public value: Strategic management in government. Harvard university press.

Morelli, N., Aguilar, M., Concilio, G., Götzen, A. De, Mulder, I., Pedersen, J., & Torntoft, L. K. (2017). Framing Design to support Social Innovation: The Open4Citizens Project. Design Journal, 20(sup1), S3171–S3184. https://doi.org/10.1080/14606925.2017.1352823

Morelli, N., de Götzen, A., Mulder, I., & Concilio, G. (2018). Editorial: Designing with Data, Democratisation Through Data. DRS2018: Catalyst, 4. https://doi.org/10.21606/drs.2018.001

Morozov, E., & Bria, F. (2018). Rethinking the smart city. Democratizing Urban Technology. New York, NY: Rosa Luxemburg Foundation, 2.

Mortati, M. (2019). The Nexus between Design and Policy: Strong, Weak, and Non-Design Spaces in Policy Formulation. Design Journal, 22(6), 775–792. https://doi.org/10.1080/14606925.2019.1651599

Mortati, M., Christiansen, J., & Maffei, S. (2018). Design craft in Government. SerDes2018 - Service Design Proof of Concept, June, 561–562. http://www.servdes.org/wp/wp-content/uploads/2018/07/46.pdf

Morton, L. W., Eigenbrode, S. D., & Martin, T. A. (2015). Architectures of adaptive integration in large collaborative projects. Ecology and Society, 20(4). https://doi.org/10.5751/ES-07788-200405

Moses, J. W., & Knutsen, T. L. (2019). Ways of Knowing: Competing Methodologies in Social and Political Research. (Third). Red Globe Press.

Moyson, S., & Scholten, P. (2018). Theories on policy learning: Existing approaches and future challenges. In Knowledge, policymaking and learning for European cities and regions. Edward Elgar Publishing.

Moyson, S., Scholten, P., & Weible, C. M. (2017). Policy learning and policy change: Theorizing their relations from different perspectives. Policy and Society, 36(2), 161–177. https://doi.org/10.1080/14494035.2017.1331879

Mulgan, G., & Albury, D. (2003). Innovation in the Public Sector. Strategy Unit Cabinet Office. https://doi.org/10.4337/9781849809757.00026

Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. Information Systems Journal, 16, 317–342.

Mureddu, F., Osimo, D., Misuraca, G., & Armenia, S. (2012). A new roadmap for next-generation policy-making. ACM International Conference Proceeding Series, 62–66. https://doi.org/10.1145/2463728.2463743

Mureddu, F., Osimo, D., Misuraca, G., Onori, R., & Armenia, S. (2014). A living roadmap for policymaking 2.0. In Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling (Issue January 2014). https://doi.org/10.4018/978-1-4666-6236-0.ch022

Mureddu, F., Schmeling, J., & Kanellou, E. (2020). Research challenges for the use of big data in policy-making. Transforming Government: People, Process and Policy, 14(4), 593–604. https://doi.org/10.1108/TG-08-2019-0082

Negreiro, M. (2019). EU Legislation in Progress Re-use of public sector information. European Parliamentary Research Service.

Newman, J., & Head, B. W. (2017). Wicked tendencies in policy problems: Rethinking the distinction between social and technical problems. Policy and Society, 36(3), 414–429. https://doi.org/10.1080/14494035.2017.1361635

Niederman, F., Ferratt, T. W., & Trauth, E. M. (2016). On the co-evolution of information technology and information systems personnel. Data Base for Advances in Information Systems, 47(1), 29–50. https://doi.org/10.1145/2894216.2894219

Nochta, T., Wan, L., Schooling, J. M., & Parlikad, A. K. (2021). A Socio-Technical Perspective on Urban Analytics: The Case of City-Scale Digital Twins. Journal of Urban Technology, 28(1–2), 263–287. https://doi.org/10.1080/10630732.2020.1798177

Nohra, C. G., Pereno, A., & Barbero, S. (2020). Systemic design for policy-making: Towards the next circular regions. Sustainability (Switzerland), 12(11). https://doi.org/10.3390/su12114494

OECD. (2015). Data-Driven Innovation: Big Data for Growth and Well-Being. https://doi.org/10.1787/9789264229358-en

OECD. (2019a). Digital Government Review of Sweden. https://doi.org/https://doi.org/https://doi.org/10.1787/4daf932b-en

OECD. (2019b). Government at a Glance 2019. https://doi.org/doi.org/10.1787/8ccf5c38-en

OECD. (2020a). Digital Government Index. 03. https://doi.org/https://doi.org/https://doi.org/10.1787/4de9f5bb-en

OECD. (2020b). Open, Useful and Re-usable data (OURdata) Index: 2019. 01. https://doi.org/https://doi.org/https://doi.org/10.1787/45f6de2d-en

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? ACM International Conference Proceeding Series. https://doi.org/10.1145/3209281.3209335

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., & Cattuto, C. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. In Science advances (Vol. 6, Issue 23, p. eabc0764). American Association for the Advancement of Science.

Pacheco, R. C. S., Manhães, M., & Maldonado, M. U. (2017). Innovation, Interdisciplinarity, and Creative Destruction. In R. Frodeman, J. Thompson Klein, & R. C. S. Pacheco (Eds.), The Oxford Handbook of Interdisciplinarity, Second Edition (pp. 303–318). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198733522.013.25

Page, E. (2006). The Origins of Policy. In M. Moran, M. Rein, & R. E. Goodin (Eds.), The Oxford Book of Public Policy (pp. 207–227). Oxford University Press.

Pagliarin, S., La Mendola, S., & Vis, B. (2022). The “qualitative” in qualitative comparative analysis (QCA): research moves, case-intimacy and face-to-face interviews. Quality & Quantity, 0123456789. https://doi.org/10.1007/s11135-022-01358-0

Papanek, V. (1972). Design for the real world. Thames and Hudson London.

Papyshev, G., & Yarime, M. (2021). Exploring city digital twins as policy tools: A task-based approach to generating synthetic data on urban mobility. Data & Policy, 3, 1–18. https://doi.org/10.1017/dap.2021.17

Parisi, L. (2018). Computational Turn. In R. Braidotti & M. Hlavajova (Eds.), Posthuman critical theory (pp. 88–90). Bloomsbury Publishing.

Parsons, M. A., Godøy, Ø., Ledrew, E., De Bruin, T. F., Danis, B., Tomlinson, S., & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. Journal of Information Science, 37(6), 555–569. https://doi.org/10.1177/0165551511412705

Pattison, A. (2018). Factors Shaping Policy Learning: A Study of Policy Actors in Subnational Climate and Energy Issues. Review of Policy Research, 35(4), 535–563. https://doi.org/10.1111/ropr.12303

Patton, M. Q. (2014). Qualitative Research & Evaluation Methods: Integrating Theory and Practice. SAGE Publications.

Pawson, R. (2002). Evidence-based Policy: The Promise of `Realist Synthesis’. Evaluation, 8(3), 340–358. https://doi.org/10.1177/135638902401462448

Perez, C. (2010). Technological revolutions and techno-economic paradigms. Cambridge Journal of Economics, 34(1), 185–202.

Peters, B. G. (2005). The problem of policy problems. Designing Government: From Instruments to Governance, 6988(2005), 77–105. https://doi.org/10.4337/9781786431356.00007

Peters, B. G. (2017). What is so wicked about wicked problems? A conceptual analysis and a research program. Policy and Society, 36(3), 385–396. https://doi.org/10.1080/14494035.2017.1361633

Peters, B. G., & Zittoun, P. (2016). Contemporary approaches to public policy. Theories, Controversies and Perspectives. UK: Palgrave Macmillan.

Pfadenhauer, M. (2009). At Eye Level: The Expert Interview — a Talk between Expert and Quasi-expert. In A. Bogner, B. Littig, & W. Menz (Eds.), Interviewing Experts. Palgrave Macmillan. https://doi.org/10.1057/9780230244276_4

Philip Chen, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Pierre, J., & Peters, B. G. (2020). Governance, Politics and the State (Second Edi). Red Globe Press.

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big Data for Policymaking: Great Expectations, but with Limited Progress? Policy and Internet, 10(3), 347–367. https://doi.org/10.1002/poi3.176

Ponti, M., & Craglia, M. (2020). Citizen-generated data for public policy. European Commission, Ispra.

Pressman, J. L., & Wildavsky, A. (1984). Implementation: How great expectations in Washington are dashed in Oakland; Or, why it’s amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals (Vol. 708). Univ of California Press.

Privy Council Office. (2018). Report to the clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service. https://www.canada.ca/content/dam/pco-bcp/documents/clk/Data_Strategy_Roadmap_ENG.pdf

Pucci, P. (2013). DATI DI TRAFFICO TELEFONICO E POLITICHE PER LA MOBILITÀ. TeMA Journal of Land Use, Mobility and Environment, 6(3), 1689–1699.

Pucci, P., Lanza, G., & Vendemmia, B. (2020). Distributing, De-Synchronizing, Digitalizing: Dealing with Milan Transport Inequalities in Post-COVID-19 Society. 1–11.

Pulse Lab Jakarta. (2019, November). Repositioning Pulse Lab Jakarta. Medium. https://medium.com/pulse-lab-jakarta/repositioning-pulse-lab-jakarta-ce66fca4d48f

Radaelli, C. M. (1995). The role of knowledge in the policy process. Journal of European Public Policy, 2(2), 159–183. https://doi.org/10.1080/13501769508406981

Raghavan, P. (2014). It's time to scale the science in the social sciences. Big Data and Society, 1(1), 1–4. https://doi.org/10.1177/2053951714532240

Ragin, C. C. (1992). "Casing" and the process of social inquiry. In C. C. Ragin & H. S. Becker (Eds.), What is a case? Exploring the foundations of social inquiry. (p. 217—226). Cambridge University Press.

Ragin, C. C. (2014). The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies. University of California Press.

Ragin, C. C., & Becker, H. S. (Eds.). (1992). What Is a Case? Exploring the Foundations of Social Inquiry. Cambridge University Press. https://doi.org/10.2307/3322115

Ramos, J. M. (2014). Anticipatory governance: Traditions and trajectories for strategic design. Journal of Futures Studies, 19(1), 35–52.

Redstrom, J. (2017). Making design theory. MIT Press.

Resh, W., Siddiki, S., & Mcconnell, W. R. (2014). Does the network centrality of government actors matter? Examining the role of government organizations in aquaculture partnerships. Review of Policy Research, 31(6), 584–609. https://doi.org/10.1111/ropr.12101

Ricci, D., Brilli, A., & Tassi, R. (2019). Repurposing Digital Methods for Human-Centered Design. DRS2018: Catalyst, 4(Manzini 2016). https://doi.org/10.21606/drs.2018.308

Riche, C., Aubin, D., & Moyson, S. (2021). Too much of a good thing? A systematic review about the conditions of learning in governance networks. European Policy Analysis, 7(1), 1–18. https://doi.org/10.1002/epa2.1080

Rieder, G. (2018). Tracing Big Data imaginaries through public policies: the case of the European Commission. In The politics of big data: big data, big brother? (pp. 89–109).

Rihoux, B., & Lobe, B. (2009). The case for qualitative comparative analysis (QCA): Adding leverage for thick cross-case comparison. The Sage Handbook of Case-Based Methods, 222–242.

Rihoux, B., & Ragin, C. (2009). Introduction. In B. Rihoux & C. Ragin (Eds.), Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques. SAGE Publications. https://doi.org/10.4135/9781452226569

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. Policy Sciences, 4, 155–169. https://doi.org/10.1080/01636609209550084

Roberts, L. G. (1978). The Evolution of Packet Switching. Proceedings of the IEEE, 66(11), 1307–1313. https://doi.org/10.1109/PROC.1978.11141

Ronzhyn, A., & Wimmer, M. A. (2021). Research directions in policy modeling: Insights from comparative analysis of recent projects. Data & Policy, 3. https://doi.org/10.1017/dap.2021.8

Rose, R. (1991). What is Lesson-Drawing? Journal of Public Policy, 11(1), 3–30. https://doi.org/10.1017/S0143814X00004918

Rosenberg, D. (2013). Data before the Fact. In L. Gitelman (Ed.), "Raw Data" Is an Oxymoron (pp. 15–40). MIT

Press. https://doi.org/10.7551/mitpress/9302.003.0003

Ross, J., & Short, J. (2012). Finding Value in the Information Explosion Intelligence. 53(53408). https://media.proquest.com/media/pq/classic/doc/2704279311/fmt/pi/rep/NONE?cit%3Aauth=Beath%2C+Cynthia%3B-Becerra-Fernandez%2C+Irma%3BRoss%2C+Jeanne%3BShort%2C+James&cit%3Atitle=Finding+Value+in+the+Information+Explosion&cit%3Apub=MIT+Sloan+Management+Revi

Rubinson, C., Gerrits, L., Rutten, R., & Greckhamer, T. (2019). Avoiding Common Errors in QCA: A Short Guide for New Practitioners. Sociology, 9, 397–418. https://doi.org/10.1177/2059799119862110.A

Sabatier, P. A. (2007). The need for better theories. In P. A. Sabatier (Ed.), Theories of the policy process (pp. 3–17). Westview Press.

Sabatier, P. A., & Jenkins-Smith, H. C. (1993). Policy change and learning: An advocacy coalition approach. Westview press.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. Big Data and Society, 6(1), 1–12. https://doi.org/10.1177/2053951718820549

Salamon, L. M. (2002). The New Governance and the Tools of Public Action: An Introduction. In The Tools of Government, A Guide to the New Governance. Oxford University Press.

Saldaña, J. (2013). The Coding Manual for Qualitative Researchers (2nd Ed.). SAGE Publications Inc.

Salö, L. (2017). Pierre Bourdieu: view points and entry points. In The sociolinguistics of academic publishing (pp. 29–42). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-58940-4_3

Sanders, E. B.-N., & Jan Stappers, P. (2017). Co-creation and the new landscapes of design. Design: Critical and Primary Sources, March, 1–16. https://doi.org/10.5040/9781474282932.0011

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. Public Administration, 80(1), 1–22. https://doi.org/10.1111/1467-9299.00292

Schneider, A., & Ingram, H. (1993). Social Construction of Target Populations : Implications for Politics and Policy Author (s): Anne Schneider and Helen Ingram Source : The American Political Science Review , Vol . 87 , No . 2 (Jun ., 1993), pp . 334-347 Published by : American Politic. The American Political Science Review, 87(2), 334–347.

Schneider, C. Q., & Wagemann, C. (2009). Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis. In Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis. https://doi.org/10.1017/CBO9781139004244

Schroeder, R. (2014). Big Data and the brave new world of social media research. Big Data and Society, 1(2), 1–11. https://doi.org/10.1177/2053951714563194

Schwandt, T. A. (2007a). Realism. In The Sage Dictionary of Qualitative Inquiry (Third Edit, pp. 256–259). SAGE Publications.

Schwandt, T. A. (2007b). Sensitizing Concept. In The SAGE dictionary of qualitative inquiry (3rd Editio, p. 274). SAGE Publications.

Shapiro, A. (2018). The Urban Stack. A Topology for Urban Data Infrastructures. TECNOSCIENZA: Italian Journal of Science & Technology Studies, 8(2), 61–80. http://www.tecnoscienza.net/index.php/tsj/article/view/318

Simon, H. A. (1979). Rational Decision Making in Business Organizations. American Economic Review, 69(4), 493–513. http://www.jstor.org/stable/1808698

Simonsen, J., & Robertson, T. (2013). Routledge international handbook of participatory design (Vol. 711). Routledge New York.

Siodmok, A. (2020, December). Lab Long Read: Human-centred policy? Blending 'big data' and 'thick data' in national policy. Policy Lab. UK Government Cabinet Office. https://openpolicy.blog.gov.uk/2020/01/17/lab-long-read-human-centred-policy-blending-big-data-and-thick-data-in-national-policy/

Smith, C., & Elger, T. (2014). Critical Realism and Interviewing Subjects. In P. K. Edwards, J. O'Mahoney, & S. Vincent (Eds.), Studying Organizations Using Critical Realism. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665525.003.0006

Squevin, P., Aubin, D., Montpetit, É., & Moyson, S. (2021). Closer than they look at first glance: A systematic review and a research agenda regarding measurement practices for policy learning. International Review of Public Policy, 3(2). https://doi.org/10.4000/irpp.2083

Strassheim, H. (2018). Policy as a body of expertise. In Handbook on Policy, Process and Governing. Edward Elgar Publishing.

Studinka, J., & Guenduez, A. A. (2018). The Use of Big Data in the Public Policy Process: Paving the Way for Evidence- Based Governance. Smart Government Lab, Institute for Public Management and Governance. https://www.alexandria.unisg.ch/255680/1/Studinka and Guenduez - The Use of Big Data in the Public Policy Process- Paving the Way for Evidence-Based Governance .pdf

Stuurman, J., Kolthoff, E., Van den Tillaart, J., & Kokkeler, B. (2020). Datagedreven zicht op ondermijning in woonwijken: Een verkenning van de mogelijkheden om indicatoren te ontwikkelen om zicht te krijgen op ondermijning in woonwijken. Tijdschrift Voor Veiligheid, 2020(1), 17–50.

Suominen, A., & Hajikhani, A. (2021). Research themes in big data analytics for policymaking: Insights from a mixed-methods systematic literature review. Policy and Internet, October 2020, 1–21. https://doi.org/10.1002/poi3.258

Surel, Y. (2000). The role of cognitive and normative frames in policy-making. Journal of European Public Policy, 7(4), 495–512. https://doi.org/10.1080/13501760050165334

Susha, I., Janssen, M., & Verhulst, S. (2017). Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. Proceedings of the 50th Hawaii International Conference on System Sciences (2017), 2691–2700. https://doi.org/10.24251/hicss.2017.325

Svensson, O., & Nikoleris, A. (2018). Structure reconsidered: Towards new foundations of explanatory transitions theory. Research Policy, 47(2), 462–473. https://doi.org/10.1016/j.respol.2017.12.007

Tenbensel, T. (2006). Policy knowledge for policy work. The Work of Policy: An International Survey, 199–216.

Thomann, E., & Maggetti, M. (2020). Designing Research With Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools. Sociological Methods and Research, 49(2), 356–386. https://doi.org/10.1177/0049124117729700

Thomson, P. (2008). Field. In M. Grenfell (Ed.), Pierre Bourdieu: Key Concepts (pp. 67–86). Acumen.

Tõnurist, P., Kattel, R., & Lember, V. (2017). Innovation labs in the public sector: what they are and what they do? Public Management Review, 19(10), 1455–1479. https://doi.org/10.1080/14719037.2017.1287939

Toshkov, D. (2016). Concepts and Operationalization. In Research Design in Political Science (pp. 28–51). Red Globe Press.

Trippe, H. P. (2021). Policy Instrumentation : The Object of Service Design in Policy Making Helena Polati Trippe. 37(3), 89–100. https://doi.org/10.1162/DESI

Trist, E. L. (1978). On socio-technical systems. Sociotechnical Systems: A Sourcebook, 43–57.

Tsoukias, A., Montibeller, G., Lucertini, G., & Belton, V. (2013). Policy analytics: An agenda for research and practice. EURO Journal on Decision Processes, 1(1–2), 115–134. https://doi.org/10.1007/s40070-013-0008-3

Tunstall, E. D. (2007). In Design We Trust: Design, Governmentality, and the Tangibility of Governance. Iasdr07, 1, 1–16. http://www.sd.polyu.edu.hk/iasdr/proceeding/papers/in design we trust_ design, governmentality, and the tangibility of governance.pdf

Turnbull, N. (2006). How Should We Theorise Public Policy? Problem Solving and Problematicity. Policy and Society, 25(2), 3–22. https://doi.org/10.1016/s1449-4035(06)70072-8

Turnbull, N. (2013). The questioning theory of policy practice: Outline of an integrated analytical framework. Critical Policy Studies, 7(2), 115–131. https://doi.org/10.1080/19460171.2013.776501

Turnbull, N. (2018). Policy as (mere) problem-solving. In H. K. Colebatch & R. Hoppe (Eds.), Handbook on Policy, Process and Governing (pp. 53–67). Edward Elgar Publishing. https://doi.org/10.4337/9781784714871.00009

Turnbull, N., & Hoppe, R. (2019). Problematizing 'wickedness': a critique of the wicked problems concept, from philosophy to practice. Policy and Society, 38(2), 315–337. https://doi.org/10.1080/14494035.2018.1488796

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives.

Ubaldi, B., & Okubo, T. (2020). OECD Digital Government Index (DGI). 41. https://doi.org/https://doi.org/https://doi.org/10.1787/b00142a4-en

Ubaldi, B., Van Ooijen, C., & Welby, B. (2019a). A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance. OECD Working Papers on Public Governance, 33, 1–59. https://doi.org/10.1787/09ab162c-en

Ubaldi, B., Van Ooijen, C., & Welby, B. (2019b). A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance. OECD Working Papers on Public Governance, 33, 1–59. https://dx.doi.org/10.1787/09ab162c-en

UN. (2019). United Nations National Quality Assurance Frameworks Manual for Official Statistics. Including recommendations, the framework and implementation guidance. https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAFManual-WEB.pdf

UN. (2020). UN E-Government Survey 2020.

van Buuren, A., Lewis, J. M., Peters, B. G., & Voorberg, W. (2020). Improving public policy and administration: Exploring the potential of design. Policy and Politics, 48(1), 3–19. https://doi.org/10.1332/030557319X15579230420063

Van Den Homberg, M., & Susha, I. (2018). Characterizing data ecosystems to support official statistics with open mapping data for reporting on sustainable development goals. ISPRS International Journal of Geo-Information, 7(12). https://doi.org/10.3390/ijgi7120456

van der Bijl-Brouwer, M. (2019). Problem framing expertise in public and social innovation. She Ji, 5(Forthcoming), 1–20. https://doi.org/10.1016/j.sheji.2019.01.003

van der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? Government Information Quarterly, 36(1), 27–38. https://doi.org/10.1016/j.giq.2018.10.011

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. Surveillance and Society, 12(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., & Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government, 13(2), 1–27. https://doi.org/10.29379/jedem.v13i2.644

van Thiel, S. (2014). Research Methods in Public Administration and Public Management. An introduction. Routledge.

Vaz, F., & Predeville, S. (2019). Design as an Agent for Public Policy Innovation. Conference Proceedings of the Academy for Design Innovation Management, 2(1). https://doi.org/10.33114/adim.2019.06.231

Verhulst, S. G., Engin, Z., & Crowcroft, J. (2019a). Data & Policy : A new venue to study and explore policy–data interaction. Data & Policy, 1, 1–5. https://doi.org/10.1017/dap.2019.2

Verhulst, S. G., Engin, Z., & Crowcroft, J. (2019b). Data & Policy: A new venue to study and explore policy–data interaction. Data & Policy, 1, 1–5. https://doi.org/10.1017/dap.2019.2

Verstraete, J., Acar, F., & Pucci, P. (2021). Turning Data into Actionable Policy Insights.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big Data: What It is and Why You Should Care. IDC White Paper, 7–8. http://www.tracemyflows.com/uploads/big_data/IDC_AMD_Big_Data_Whitepaper.pdf

Vink, J., Wetter-Edman, K., & Koskela-Huotari, K. (2021). Designerly Approaches for Catalyzing Change in Social Systems: A Social Structures Approach. She Ji: The Journal of Design, Economics, and Innovation, 7(2), 242–261. https://doi.org/10.1016/j.sheji.2020.12.004

Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. Government Information Quarterly, 36(4), 101383. https://doi.org/10.1016/j.giq.2019.05.010

Wagner-Pacifici, R., Mohr, J. W., & Breiger, R. L. (2015). Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. Big Data and Society, 2(2), 1–11. https://doi.org/10.1177/2053951715613810

Ward, J. S., & Barker, A. (2013). Undefined By Data: A Survey of Big Data Definitions. http://arxiv.org/abs/1309.5821

Weber, I., Imran, M., Ofli, F., Mrad, F., Colville, J., Fathallah, M., Chaker, A., & Ahmed, W. S. (2021). Non-Traditional Data Sources: Providing Insights into Sustainable Development. Communications of the ACM, 64(4), 88–95. https://dl.acm.org/doi/10.1145/3447739%0Ahttps://cacm.acm.org/magazines/2021/4/251357-non-traditional-data-sources/fulltext

Webster, F. (2014). What is an information society? In Theories of the information society (pp. 8–31). Routledge.

Wegrich, K., & Jann, W. (2006). Theories of the Policy Cycle. 43–62. https://doi.org/10.1201/9781420017007.pt2

Weible, C. M., & Sabatier, P. A. (2018). Theories of the policy process. Routledge.

Weiss, Carol H. (1977). Research for Policy's Sake: The Enlightenment Function of Social Research. Policy Analysis, 3(4), 531–545. https://www.jstor.org/stable/42783234

Weiss, Carol H. (1979). The Many Meanings of Research Utilization. American Society for Public Administration, 39(5), 426–431.

Weiss, Carol Hirschon. (1999). The Interface between Evaluation and Public Policy. Evaluation, 5(4), 468–486.

Wellstead, A., & Stedman, R. (2015). Mainstreaming and Beyond: Policy Capacity and Climate Change Decision-Making. Michigan Journal of Sustainability, 3(20181221). https://doi.org/10.3998/mjs.12333712.0003.003

Wesselink, A., Colebatch, H. K., & Pearce, W. (2014). Evidence and policy: discourses, meanings and practices. Policy Sciences, 47(4), 339–344. https://doi.org/10.1007/s11077-014-9209-2

Wetter-Edman, K., Vink, J., & Blomkvist, J. (2018). Staging aesthetic disruption through design methods for service innovation. Design Studies, 55, 5–26. https://doi.org/10.1016/j.destud.2017.11.007

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

Windrum, P. (2008). Innovation and entrepreneurship in public services. In Innovation in Public Sector Services: Entrepreneurship, Creativity and Management (pp. 3–20).

World Bank. (2017). BIg Data in Action for Government Big Data Innovation in Public Services, Policy and Engagement.

Wu, X., Ramesh, M., & Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. Policy and Society, 34(3–4), 165–171. https://doi.org/10.1016/j.polsoc.2015.09.001

Wu, Xindong, Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Wyatt, S. (2008). Technological determinism is dead; long live technological determinism. The Handbook of Science and Technology Studies, 3, 165–180.

Yildiz, M. (2007). E-government research: Reviewing the literature, limitations, and ways forward. Government Information Quarterly, 24(3), 646–665. https://doi.org/10.1016/j.giq.2007.01.002

Yin, R. K. (2018). Case Study Research and Applications. In Case Study Research and Applications: Design and Methods. SAGE Publications.

Zarsky, T. Z. (2013). Transparent predictions. University of Illinois Law Review, 2013(4), 1503–1569.

Annex 1

List and description of queries used for the systematic literature review of qualitative interviewing method

Research area	Query used on Scopus	Description of the query
Public sector (general)	(TITLE-ABS-KEY ("interviews" AND "public sector")) OR (TITLE-ABS-KEY ("interviewing" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviewing" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviews" AND "public sector")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))	The query retrieves all entries that contain in the title, abstract or keywords the following couples of words: “interviews” and “public sector”; “interviewing” and “public sector”; “qualitative interviews” and “public sector”; “qualitative interviewing” and “public sector”. Entries can be either Journal or Conference Articles, from 2016 to 2021.
Public sector innovation	(TITLE-ABS-KEY ("interviews" AND "public sector innovation”)) OR (TITLE-ABS-KEY ("interviewing" AND "public sector innovation")) OR (TITLE-ABS-KEY ("qualitative interviewing" AND "public sector innovation")) OR (TITLE-ABS-KEY ("qualitative interviews" AND "public sector innovation")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))	The query retrieves all entries that contain in the title, abstract or keywords the following couples of words: “interviews” and “public sector innovation”; “interviewing” and “public sector innovation”; “qualitative interviews” and “public sector innovation”; “qualitative interviewing” and “public sector innovation”. Entries can be either Journal or Conference Articles, from 2016 to 2021.
Technological innovation in public sector	(TITLE-ABS-KEY ("interviews" AND "technological innovation" AND "public sector")) OR (TITLE-ABS-KEY ("interviewing" AND "technological innovation" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviewing" AND "technological innovation" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviews" AND "technological innovation" AND "public sector")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))	The query retrieves all entries that contain in the title, abstract or keywords the following triplets of words: “interviews” and "technological innovation" and “public sector”; “interviewing” and "technological innovation" and “public sector”; “qualitative interviews” and "technological innovation" and “public sector”; “qualitative interviewing” and "technological innovation" and “public sector”. Entries can be either Journal or Conference Articles, from 2016 to 2021.
Governance in public sector	(TITLE-ABS-KEY ("interviews" AND "governance" AND "public sector")) OR (TITLE-ABS-KEY ("interviewing" AND "governance" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviewing" AND "governance" AND "public sector")) OR (TITLE-ABS-KEY ("qualitative interviews" AND "governance" AND "public sector")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))	The query retrieves all entries that contain in the title, abstract or keywords the following triples of word: “interviews” and "governance" and “public sector”; “interviewing” and "governance" and “public sector”; “qualitative interviews” and "governance" and “public sector”; “qualitative interviewing” and "governance" and “public sector”. Entries can be either Journal or Conference Articles, from 2016 to 2021.

Annex 2

Interview formats (Expert Interview / Key Agent Interview)

Guidelines for Expert Interview

Dear interviewee,
The following sections will guide you in the interview process.
We kindly ask you to read them before the scheduled interview.

1. ABOUT THE INTERVIEW

1.1. — What do we want to find out?
The interview aims to share information and opinion about *data for policy*, as a growing international field of studies and governmental initiatives at the intersection between the use of digital data and policy-making. We intend to empirically learn how this field is evolving by adopting the disciplinary perspective of design. Our inquiry asks: how do practices through which data get institutionalized into knowledge for policy affects governance structures and policy design?

1.2. — What is the output of this research?
The funding university (Politecnico di Milano) entrust his researchers with the publication of study results as part of scientific open access venues. Results must also be published as part of a doctoral thesis, publicly available on Politecnico di Milano digital repository (www.politesi.polimi.it).

1.3. — Who could benefit from it?
The research seeks to support policy-makers and public servants who carry the responsibility of harnessing new digital data for solving complex societal issues while maintaining public trust and citizens' needs at the centre of their actions. Moreover, it aspires to support scholars who inquire about the socio-technical innovation of using data in the public sector and how this results in innovative government and governance models.

1.4. — How to deliver the interview?
We kindly ask you to have a direct interview online via Microsoft Teams or Zoom.
The interview will be held in English by a member of our research team and they wil be recorded.
Before starting, the interviewee will be asked consent for recording.
The expected duration is 45/60 minutes.
The transcript verbatim will be shared afterwards to the interviewee for final approval, together with a detailed consent form.
In case some relevant aspects remains unclear, we ask your availability for follow-up questions.

2. ABOUT DATA FOR POLICY
Please, read this brief introduction on data for policy before proceeding with questions.

2.1. — What is data for policy?
Data for policy is a field encompassing actors interested in building knowledge about the socio-technical co-evolution of governments/public sector and new information systems.

More specifically, *data for policy* is interested in dealing with the *data revolution*, by asking:
How to harness the unprecedented digital data abundance through data-driven approaches to policy? What are the consequences for government and governance?

Researchers who so far investigated *data for policy* agree on an ambivalent scenario.
There is consensus that the government's use of public data (for example administrative data) will be

increasingly important. However, great expectations are so far unmatched by practical applications. *The use of data in policy-making* seems to be hampered by current organizational barriers, the lack of capabilities and absence of appropriate ethical/legal frameworks for ensuring data privacy.

While still barely utilized for policy-making, the treatment of large datasets with analytics techniques appears a common practice both in the public services provision and management.
These include, for example: police patrolling, planning of health inspections, detection of non-compliance in tax collection, detection of fraud in application for welfare services (e.g., housing).

3. INTERVIEW QUESTIONS

The following questions aim to understand your perspective on data for policy.
They intend to support an interactive conversation between you and the researcher.
Therefore, we equally value the information you will be willing to share by answering to them and articulating your thoughts starting from them. Keeping in mind they do not represent a fixed structure, we welcome any opinions you feel comfortable sharing.

Pre-interview

At the beginning of the interview recording you will be asked some general information about yourself and your past experience with innovative practices in government and policy.

List of questions

1. **What are your considerations about *data for policy*?**
   You are invited to share your considerations starting from the brief introduction proposed above. Please, be direct and critical. You can accept, integrate or totally refuse what stated.

2. **In your opinion, what are the characterising elements of *data for policy* in contrast with other data practices in the public sector (e.g., Open data)?**
   Please, indicate what you think are the aspects that specifically characterize *data for policy* as a field and tell us why. For instance, these elements can be drivers (potential of new technologies) or challenges (ensuring privacy).

3. **What do you foresee as changing factors for this field in the next years?**
   Describe what you think will be game-chaning factors in data for policy in the next future and why. You can think this as recommendations for policy-makers or intermediaries.

4. **Can you describe an example of public policy, governmental initiative or small-scale experimental project that you consider relevant for *data for policy*?**

In conclusion, *we will briefly introduce you three case studies* that we consider relevant for our specific disciplinary perspective. After the presentation, you are invited to share your thoughts about them.

CLOSING REMARKS

We'd like to thank you again for you time and effort in contributing to this research.
For any doubts or questions that were not covered in these pages, please contact:
francesco.leoni@polimi.it

Annex 2

Interview formats (Expert Interview / Key Agent Interview)

Guidelines for Key Agents Engagement

Dear interviewee,
The following sections will guide you in the interview process.
We kindly ask you to read them before the scheduled interview.

1. ABOUT THE STUDY

1.1. — What do we want to find out?
The interview aims to share information and opinion about specific *data-driven public action,* which we consider relevant cases to investigate the *data for policy field*. This is a growing international field of studies and governmental initiatives at the intersection between the use of digital data and policy-making. We intend to empirically learn how this field is evolving by adopting the disciplinary perspective of design. Our inquiry asks: how do practices through which data get institutionalized into knowledge for policy affects governance structures and policy design?

1.2. — What is the output of this research?
The funding university (Politecnico di Milano) entrust his researchers with the publication of study results as part of scientific open access venues. Results must also be published as part of a doctoral thesis, publicly available on Politecnico di Milano digital repository (www.politesi.polimi.it).

1.3. — Who could benefit from it?
The research seeks to support policy-makers and public servants who carry the responsibility of harnessing new digital data for solving complex societal issues while maintaining public trust and citizens' needs at the centre of their actions. Moreover, it aspires to support scholars who inquire about the socio-technical innovation of using data in the public sector and how this results in innovative government and governance models.

1.4. — The research design
Your interview will support the development of a case study analysis: a very closed analysis on a specific context (also called bounded system). Such analysis will be performed by accessing multiple primary and secondary data sources. We will integrate your perspective with other information on your broader context (e.g., on your national context).
If you consider it possible, we would also hope to directly access the implementation activities of your initiative. We would appreciate any help you can provide in this sense.

1.5. — How we would like to engage you
We kindly ask you to have one or two information sharing meetings online via Microsoft Teams or Zoom (depending on your convenience). These will be held in English by a member of our research team and they will be recorded. Before starting, the interviewee will be asked consent for recording. The expected duration is 30/45 minutes for each of the two meetings or 1 hour for an unique meeting. The two meetings activities are described below.
The transcript verbatim and any visual support outputs produced during the session will be shared afterwards to the interviewee for final approval, together with a detailed consent form. In case some relevant aspects remains unclear, we ask your availability for follow-up questions.

DESIGN POLICY LAB

www.designpolicy.eu
www.dipartimentodesign.polimi.it

Annex 2

Interview formats (Expert Interview / Key Agent Interview)

2. INFORMATION SHARING SETUP

The activities described below aim to understand the *data-driven public action* in which you are involved thorugh one or two meetings of information sharing.
During the first meeting you will be asked questions, in a regular interview format.
During the second meeting you will be involved in a co-design activity called *system mapping.*
In case you prefer to have an unique meeting, the two activities will be held in sequence during the same session.

2.1. — Activity 1: Interview

At the beginning of the interview recording you will be asked some general information about yourself and your role within the data-driven public action your are involved (henceforth simply called *action*).

List of questions

Please, describe the action by referring to the questions below.

- What are the main goals, beneficiaries and expected outputs of the action?
- Which subjects are involved in the action and with what roles?
- What type of political support the action received?
- When did the action start and when it is expected to end?
- What is the specific policy area, service system or government function of this action? (e.g., welfare)
- How the action is positioned in respect to existing policy schemes and regulations?
- Can you describe the technological components of this action?
  Specifically, what is the role of data? What type of data sources are involved?

2.2. — Activity 2: Visualizing the action System Map

This activity will use visualization as a technique to display the data-driven public action as a networked system. In order to understand the main elements of the system, you will be asked to participate and help to build its visual representation. In case you do not feel comfortable in engaging with visual techniques, the researcher will act as visual facilitator and represent what you describe. The goal is to display the *data-driven public action* in its main components.
It should be possible to understand: what are the data sources and data mining/collection methods, how data processing and information building is handled, which actors contribute to this process and how.

CLOSING REMARKS

We'd like to thank you again for your time and effort in contributing to this research.
For any doubts or questions that were not covered in these pages, please contact:
francesco.leoni@polimi.it

DESIGN
POLICY
LAB

www.designpolicy.eu
www.dipartimentodesign.polimi.it

Annex 3

Self-assessment of learning in data-centric policymaking

Section 1 - Characteristic of sample (learners)

Indicator	Question	Scale
1A. Working position	Which of the following options describes your job position when you were involved in [case name]?	☐ I was working as a public servant, employed as part of a ministerial department, public agency/office or local authority (city, region, province) ☐ I was working for a company/agency owned by the government/public sector ☐ I was working for a private company ☐ I was working for a non-profit ☐ I was working as university staff ☐ None of the above applies If none of the above applies, please describe what was your job position during during the time you were involved in [case name]: [open answer]
1B. Affiliation	Which of the following options defines your affiliation with the government/public sector when you were involved in [case name]?	☐ I worked for a ministerial Department/Office ☐ I worked for a Non-ministerial Department (e.g., the central bank) ☐ I worked for a Public agency ☐ I worked for a City/Municipality ☐ I worked for a Regional council/agency ☐ I worked for a public university ☐ I was not affiliated with government/public sector If none of the above applies, please describe with your own words your affiliation with government/public sector during the time you were involved in [case name]? [open answer]
1C. Claim on knowledge of government and public sector	Please, consider your past working experience with the government and public sector. Based on that, how would you define your knowledge of the politico-administrative system (e.g., in terms of politics, regulations, responsible roles and organizations, policies and services)?	☐ <u>I am a full expert on government.</u> I developed a deep knowledge of my country's government/public sector. Additionally, I am an expert in one or more specific policy areas or public services (e.g. social care). ☐ <u>I am a specialist.</u> I consider myself an expert in one or more specific policy areas or services (e.g. social care) because of my previous specific experience. My knowledge about other areas of government is not equally profound. ☐ <u>I am a generalist.</u> I have deep knowledge of how government and the public sector function. However, I did not develop any specialist knowledge about one specific policy area or service. ☐ <u>I have an average knowledge of government.</u> I do not possess any general or specialist knowledge on government or public sector functioning, although I am involved in government projects,
1D Sorting out the technical profiles	Please consider your past working experience inside or for the government/public sector.	☐ Yes ☐ No ☐ I don't know

Annex 3

Self-assessment of learning in data-centric policymaking

Section 1 - Characteristic of sample (learners)

	Can you be described as a technical profile with generic or specialized expertise in a technological field? (For example, ICT)	
1E. Years of past working experience in/for government	How many years have you worked in or for the government/public sector/any public authority (a city or a region)?	☐ More than 10 years ☐ From 5 to 10 years ☐ From 1 to 5 years ☐ Less than 1 year ☐ Never
1F. Frequency of experience in/for government	Considering your career, how would define the frequency you have worked in or for the government/public sector/any local authority (as a city or a regional council)?	☐ Constant ☐ Sporadic
1G. Degree of participation in case	Thinking of time and effort, how would you define your participation as part of [case name]?	☐ Very HIgh ☐ High ☐ Don't know ☐ Low ☐ Very Low
1H. Perception about role of data/technology	How much do you agree with the following statement: Innovative use of public/administrative data was the central element of the *[case name]*.	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
1I. Interaction with data	During your participation in *[case name]*, how would you define your position regarding the project?	☐ I was part of the development team (LINKS TO QUESTION 8) ☐ I was part of a public/government organization that directly promoted the initiative (LINKS TO QUESTION 8B) ☐ I was part of an organization partner to the initiative (LINKS TO QUESTION 8B) ☐ I was part of an organization interested in the project development and outcomes (LINKS TO QUESTION 8B) ☐ My affiliation was none of the above (LINKS TO QUESTION 9)
1L. Competences added by data practitioners	What were the primary competencies you added to the team that worked in [case name]? (indicate max. 2)	☐ General ICT competencies ☐ Data science ☐ AI/ML ☐ UX/UI ☐ Service design ☐ Project manager ☐ Communication ☐ General research competencies ☐ Legal advice ☐ General support

Annex 3

Self-assessment of learning in data-centric policymaking

Section 1 - Characteristic of sample (learners)

		☐ Policy knowledge
1M. Reason of involvement (actors from policy subsytem)	Please, indicate the primary reasons that motivate your involvement in the [case name].	☐ I provided legal advice on public data sharing and privacy ☐ I provided advice on ethics of data use and processing techniques (AI/ML) ☐ My experience in gov tech or gov data projects ☐ My experience in citizen engagement ☐ My experience in working in public/government projects ☐ My experience in policymaking ☐ My experience with public services ☐ My expertise and experience in vocational rehabilitation policymaking ☐ My expertise and experience in vocational rehabilitation services ☐ I was involved because of working for one of the organizations that own the data used in the project ☐ I was involved because of my position in the organization supporting the project ☐ I was involved because of working for an organization interested in the project ☐ I independently joined the initiative during a public event ☐ I was invited to a workshop/focus group (GOES TO 1O)
1O. Stage of involvement	[case name] aimed to employ public/administrative data innovatively to address a public issue or improve public services. Concerning that specific aspect, at which stage of activity you were mostly involved?	☐ Problem setting & defining needs Activities to understand how data can solve specific public problems or address public needs. ☐ Data gathering Activities for accessing, selecting and collecting relevant public/administrative datasets among the available ones. ☐ Data preparation Activities to prepare public/administrative datasets for future data analysis. ☐ Data mining & analysis Activities for data exploration and analysis through programming and techniques to obtain original information from data. ☐ Prototyping or Deployment Prototyping or Deployment - Activities for realizing data-centred outputs featuring innovative use of public/administrative data. As outputs, we can consider web/digital services, visualizations/simulations, a dashboard, etc. ☐ Communication & Dissemination Activities of discussion and reflections start from a data-centred output produced in the project. As output, we can consider web/digital services, visualizations/simulations, a dashboard, etc.

Annex 3

Self-assessment of learning in data-centric policymaking

Section 2 - Learning Self-assessmen

Indicator	Question	Scale
2.A Statement on previous knowledge	To what extent do you agree with the following statement: *I am an expert in [policy problem of this case study] intended as general areas of public issues and solutions.*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.B Policy Problem (Cognitive Learning)	To what extent do you agree with the following statement: *By being involved in [case name] I obtained new information on [policy problem of this case study].*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.C Policy Tools (Cognitive Learning)	To what extent do you agree with the following statement: *By being involved in [case name], I obtained new information on government policies and measures in the area of [policy problem of this case study].*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.D Public Services (Cognitive Learning)	To what extent do you agree with the following statement: *By being involved in [case name], I obtained new information on the public systems and services in the area of [policy problem of this case study].*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.E. Policy Subsystem and actors (COGNITIVE)	To what extent do you agree with the following statement: *By being involved in [case name], I obtained new information about responsible public organizations and actors in the area of [policy problem of this case study] in my country.*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.F Policy Problem (Normative Learning)	To what extent do you agree with the following statement: By being involved in *[case name]*, I changed my *fundamental* beliefs about what is considered a public problem in the area of *[policy problem of this case study]*.	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.G Policy Tools (Normative Learning)	To what extent do you agree with the following statement: *By being involved in the [case name], I have changed my fundamental beliefs on what are the right solutions in [policy problem of this case study].*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.H Governance (Normative Learning)	To what extent do you agree with the following statement: *By being involved in [case name] I have changed my fundamental beliefs on existing governance structure in the area of [policy problem of this case study].*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree
2.I Normative Vision (Normative Learning)	To what extent do you agree with the following statement: *By being involved in [case name] , I have changed my personal vision on how government should address [policy problem of this case study] in the future.*	☐ Strongly disagree ☐ Disagree ☐ Neither agree nor disagree ☐ Agree ☐ Strongly agree

Annex 4

Codes groupings - literature analysis

Code Group	Codes	Count	Total count by group
Data Ethics/Privacy	Privacy	9	29
	Data Ethics	3	
	Transparency	3	
	Data Ownership	2	
	Personal Data Stores	2	
	Trust	2	
	GDPR	2	
	Digital Divide	1	
	Data Protection	1	
	Informational Wellbeing	1	
	Pervasive Data profiling	1	
	Transparency of Model	1	
	Data protection by Design	1	
Specific topics, technologies and applications	Smart Contracts	2	28
	Automatic Number Plate Reader	1	
	Digital Credit	1	
	Drones	1	
	e-services	1	
	Internet of Things	1	
	Remote sensing	1	
	Location Identification	1	
	Multi Party Computation	1	
	Semantic-based legal research	1	
	Signal Processing	1	
	Smart Cities	1	
	Policing	1	
	Disaster Management	1	
	Smart Policing	1	
	Smart Statistics	1	
	Risk Management	1	
	National Identity Programs	1	
	Geo-orientation	1	
	Augmented Humans	1	
	Data-driven innovation	1	
	Augmented Humans	1	
	Future Scenarios	1	
	Design Fiction	1	
	Proxy measure	1	
	Typology Classifier	1	
	Block Chains	1	

Annex 4

Codes groupings - literature analysis

Data (sources and types)	Administrative Data	9	23
	Open government data	7	
	Micro data	2	
	Crowd-sourced data	2	
	Public procurement data	1	
	Electronic Health Records	1	
	Land Records	1	
Data science and analytics	Data Visualization	5	15
	Data Science	4	
	Data Mining	2	
	Data Analytics	1	
	Data Streams	1	
	Data Model	1	
	Integrated Data Model	1	
AI/ML	Machine Learning	5	14
	Natural Language Processing	4	
	Algorithmic Governance	2	
	Artificial Intelligence	1	
	Supervised Learning	1	
	Text mining	1	
Analytics and Modelling	Agent-based Modelling	4	13
	Policy Analytics	1	
	Policy Modeling	1	
	Predictive Analytics	1	
	Sentiment Analysis	1	
	Policy Monitoring	1	
	Topic Modeling	1	
	City Modelling	1	
	Real-time insights	1	
	Real-time simulation	1	
Citizens engagement and participation	Citizen engagement	3	13
	Citizen Empowerment	2	
	Citizen Science	1	
	Civic Responsibility	1	
	Collaboration	1	
	Community Engagement	1	
	User-generated data	1	
	Policy Co-creation	1	
	Participatory Data Collection	1	
	Participatory Sensing	1	
Policy-related	Evidence-based policymaking	4	11
	Local Government	3	
	Indicators	2	
	Evaluation of task demand and	1	

Annex 4

Codes groupings - literature analysis

	workload		
	Sustainable Development Goals	1	
Data Culture	Research Data Centers	2	9
	Data Literacy	1	
	Data Access	1	
	Data infrastructure	1	
	Data Integration	1	
	Interoperability	1	
	Data Management	1	
	Data Quality	1	
Data governance	Data Governance	4	7
	Data brokerage	1	
	Data trust	1	
	Data Commons	1	

Annex 5

Desk research case studies

1 — Non-compliance identification and risk assessment in tax system

Main subject involved
HM Revenue and Customs - (United Kingdom)

Type of organization
Public financial corporations - Non-ministerial Department

Date	Scale	Policy field
2017	National	Taxation

Description
The HM Revenue and Customs (HMRC) is a non-ministerial department within the UK public sector responsible for tax collection. HRMC has since long time used data mining, analytical tools and analytical techniques in a strategy to "maximise revenues and bear down on avoidance and evasion" (HM Revenue & Customs, 2017, p.2). HMRC uses data at a strategic level with the specific aims to reduce the tax gap (i.e., a statistical figure that measures how much should be paid in taxes compare it with how much it is actually paid) and in relation to compliance risk assessment.
In relation to the latter, HMRC uses predictive analytics models to anticipate the probability of specific categories of taxpayers being non-compliant, plus to get other information (e.g., which economic sector non-compliance categories belong to).

"Our approach is underpinned by cutting-edge data analysis to identify where tax is most at risk of not being paid, so we can design tailored, targeted and proportionate interventions" (HM Revenue & Customs, 2020).

These models are used to develop insight for the national treasury, to tackle evasion, and to address frauds and errors in the systems. HMRC declares to experiment analytics with nudging techniques to develop some targeted communication toward specific taxpayers, for example, to support shifting from paper to online filing.

Type of data
Data is collected through the interaction of citizens with the department's services (e.g., submitting income tax), third parties (e.g., employers, other public departments, banks) and also publicly available sources. Data collected regards the following entities: members of the public, customers and clients, businesses, suppliers and service providers, advisers, consultants and other professional experts, complainants and enquirers, agents and representatives, relatives, children, guardians, dependents and associates, offenders and suspected offenders, employees. (HM Revenue & Customs, 2021, February). Data collected regards: name, title, addresses, telephone numbers, personal email addresses, gender, marital status and dependents, National Insurance number, bank account details, information about your income, information about your employment, information about your business activities, information about your domestic and business properties, passport and driving licence information, biometric data (voice recognition data).

How data is used
Prevent non-compliance, block fraud, prevent mistakes, prevent fraudulent claims, personalise online services, automate calculations

Sources considered for the analysis

Atto, J., Lord, J & Potter, C. (2015, January). Predictive analytics: the science of non-compliance . Her Majesty's Revenue and Customs UK Government. https://quarterly.blog.gov.uk/2015/01/27/predictive_analytics/

Bae Systems (ND) Using Technology to Help HMRC Find More Tax Fraud.
https://www.baesystems.com/en/cybersecurity/feature/using-technology-to-help-hmrc-find-more-tax-fraud

Devereux, R. (Department of Work & Pensions) & Thompson, J. (HM Revenue & Customs). (2016, June) Fraud and Error Stocktake. Letter to the House of Commons.
https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf

European Commission. (2017). Good Practice Fiche – UK: Data mining tools and methods to tackle the hidden economy in the UK. http://ec.europa.eu/social/BlobServlet?docId=18525&langId=en

HM Revenue & Customs. (2021, February). HMRC Privacy Note.
https://www.gov.uk/government/publications/data-protection-act-dpa-information-hm-revenue-and-customs-hold-about-you/data-protection-act-dpa-information-hm-revenue-and-customs-hold-about-you#kind-of-information

HM Revenue & Customs. (2020). 2019 to 2020 Annual Report and Accounts.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/629941/HMRC-Strategy.pdf

2 — Piano casa Italia

Main subject involved
Presidency of the Council of Ministers - (Italy)

Type of organization
Central government - Ministerial Department/Office

Date	Scale	Policy field
2016	National	Urban development

Description
Piano Casa Italia is “a risk prevention policy aimed at reducing the vulnerability of buildings in case of earthquakes in Italy” (Azzone, 2018, p. 118). The Italian Presidency of the Council of Ministers launched it in September 2016, in response to the earthquakes that struck the central regions of Italy during the summer of that same year. The policy is implemented through the institution of an homonymous Department called “Casa Italia” (http://www.casaitalia.governo.it/it/), with the specific mission to achieve its goals.
The policy priority underlying Piano Casa Italia is the conservation of fragile buildings (e.g., public buildings, as schools) by collecting and integrating informative sources and allowing households to renovate those considered at risk.

Type of data
The department integrated public data on geology and hydrogeology (e.g., historical data on earthquakes) coming from other institutins and research centers (e.g., Istat, Ingv e Ispra) with census and administrative data by municipalities (e.g., number of inhabitants, number of public heritage buildings) and data on individual buildings (e.g., energy class) from a digital cadastre.

How data is used
The data analysis was intended to steer a publicly-funded “building diagnosis”, directed to specific residential buildings that the study showed as more at risk. In other words, private householders would be able to apply for public funds for a structural assessment of their households.

Sources considered for the analysis
Azzone, G. (2018). Big data and public policies: Opportunities and challenges. Statistics & Probability Letters, 136, 116-120.

3 — Their Future Matters

Main subject involved
Department of Communities and Justice - New South Wales Government - (Australia)

Type of organization State government - Ministerial Department/Office

Date	Scale	Policy field
2016	National	Vulnerable families and childrens

Description
Their Futures Matter (TFM) is an initiative from New South Wales Government Stronger Communities Investment Unit (https://www.theirfuturesmatter.nsw.gov.au/) aiming "to deliver improved outcomes for vulnerable children, young people and their families" (Taylor Fry 2019, p. 27). Started in 2016, in response to an independent governmental review on the state of out of home care system (i.e., the system of services that support children and young people who can't live in their family home). The review pointed out the poor quality of outcomes despite important investments. To address the issue of performance and demand for evidence, TFM developed a model "to help define vulnerable groups […] expected to require a high level of government services and supports in the future" (Taylor Fry 2019, p. 13). The model uses a linked administrative data set called Human Service Data Set (HSDS) to forecast how much specific beneficiaries will cost in terms of services delivery, making individual projection up to when children will be forty years old. The HSDS combines data on how families have interacted with the government (i.e., service use pathways) from different agencies during 27 years, and focuses on child/young people born after 1990 and their relatives, guardians or carers. Data were collected among a series of governmental agencies from the NSW Police Force (e.g., custody data) to the NSW Ministry of Health (e.g., Public hospital admissions). Data are also collected during the phase of service delivery through a framework that translates TFM's goals into a set of measurable indicators, with the overall purpose of monitoring the initiative's results.

Type of data
The HSDS combines data on how families have interacted with the government (i.e., service use pathways) from different agencies during 27 years, and focuses on child/young people born after 1990 and their relatives, guardians or carers. According to an independent audit, this dataset is “unprecedent” in NSW, with “over seven million records, from more than 60 frontline data sets in 11 government agencies.” (Audit Office of New South Wales, 2020, p.28). The Centre for Health Record Linkage (CHeReL) apparently played a central role. Data are collected among a series of governmental agencies from the NSW Police Force (e.g., custody data) to the NSW Ministry of Health (e.g., Public hospital admissions). Data are also collected during the phase of service delivery through a framework that translates TFM's goals into a set of measurable indicators, with the overall purpose of monitoring the initiative's results.TFM collaborate with NSW Privacy Commissioner to establish a Public Interest Direction and a an Health Public Interest Direction (specific legal frameworks) to allow the temporary share of data for the initiative.

How data is used
To address the issue of performance and demand for evidence, TFM developed a model "to help define vulnerable groups […] expected to require a high level of government services and supports in the future" (Taylor Fry 2019, p. 13). The model uses a linked administrative data set called Human Service Data Set (HSDS) to forecast how much specific beneficiaries will cost in terms of services delivery, making individual projection up to when children will be forty years old.

Sources considered for the analysis
Audit Office of New South Wales (2020, July). Performance Audit Their Futures Matter. https://www.audit.nsw.gov.au/sites/default/files/documents/Their%20Futures%20Matter%20-%20PDF%20Report.pdf
Taylor Fry. (2019). Forecasting Future Outcomes. Stronger Communities Investment Unit — 2018 Insights Report. New South Wales Government. https://www.theirfuturesmatter.nsw.gov.au/__dat a/assets/pdf_file/0003/67 3284/Forecasting-Future-Outcomes-Stronger-Communities-Investment-Unit-2018-Insights-Report.pdf

4 — Housing Benefit Matching Service

Main subject involved Department of Work and Pensions - (United Kingdom)		
Type of organization Central government - Ministerial Department/Office		
Date 1996	**Scale** National	**Policy field** Housing benefits

Description
The Department of Work and Pensions of the UK government adopts an automatic mechanism for detecting frauds and errors in the application for house benefits and universal credit. The system has been running and used since 1996.

Type of data
The Housing Benefit Matching Service compares data on citizens from local authorities (i.e., city councils) with application data possessed by DWP and other benefit systems (HMRC, HM Prison Service and Royal Mail) to identify discrepancies.

How data is used
Once a discrepancy is detected in application by the HBMS, the DWP notifies it to the local authories that can decide if pursue an investigation.

Sources considered for the analysis
UK Government. (2018). Housing Benefit General Information Bulletin. Department for Work and Pensions.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/701453/g4-2018.pdf

5 — Kenya Livestock Insurance Program

Main subject involved Department of Agriculture - State Department for Livestock - (Kenya)		
Type of organization Central government - Ministerial Department/Office		
Date 2014	**Scale** Regional	**Policy field** Agriculture

Description
The Kenyan Government developed an index-based livestock insurance program to support livelihood and pastoralist communities that are affected by severe droughts caused by global warming. The plan — launched in 2014 by the Government, actors from the national private sector insurance world, and other international partners — adopt a predictive index-based insurance model (which is common in agricultural insurance) based on satellite data.

Type of data
"KLIP design is based on satellite data of the vegetation cover to assemble an index of seasonal forage availability/scarcity, referred to as Normalized Differenced Vegetative Index (NDVI) that is used to determine when payouts are made" (State Department of Livestock Government of Kenya, 2018, July, p.9).

How data is used
More specifically, satellite imagery analyzes the green vegetation available to livestock and predict the mortality rate. As this rises above a certain threshold, an automatic payment is given by Government to affected beneficiaries through a payment app.

Sources considered for the analysis
Bett, K. (2019). Agriculture Data Shaping Policy and Changing Lives in Kenya & Tanzania. http://www.data4sdgs.org/sites/default/files/services_files/Ag%20Data%20Shaping%20Policy%20and%20Changing%20 Lives_Case%20Study.pdf

State Department of Livestock Government of Kenya. (2018, July). Executive Seminar on Index Based Livestock Insurance Summary report.

World Bank Group Financial Protection Forum. (2017, June). Kenya Livestock Insurance program (KLIP

6 — The NEAR Program

Main subject involved		
Department of Industry, Science, Energy, and Resources - (Australia)		
Type of organization		
Central government - Ministerial Department/Office, National Agency		
Date	**Scale**	**Policy field**
2017	National	Energy

Description
The National Energy Analytics Research is a program developed by the Department of Industry, Science, Energy, and Resources of Australian Government, together with the national research center CSIRO and the private Australian Energy Market Operator. NEAR's goal is to improve the government's analytics capacity and provide better data about energy consumption, thus allowing forecasting and planning in the policy.
The program – originally called Energy Use Data Model – firstly went through a 3 years-long pilot with 6 millions dollars, then received more funding in the 2017-2018 budget (13 millions dollars).

Type of data
The portal already has 150 datasets on the following data: demographics, household characteristics, building features, energy efficiency measures, appliance uptake and usage. thermal comfort, space heating and cooling, solar panels and battery storage, energy sources and fuel switching, hot water systems, swimming pools and spas, transport methods, electric vehicles, electricity and gas bills, changing energy plans. Data are, for example, collected from the transforming stations of the electric system grid. Furthermore, to gather more granular data on individual consumption, the program integrated surveyed data by volunteers citizens through an app called "Energise".

How data is used
NEAR seeks to connect and link existing data on energy consumption with other factors that determine it. One of the goals of NEAR is not only to gather and link data, but also to enable new research capacities: the web portal of the initiative presents data analysis and visualizations.

Sources considered for the analysis
Australian Government Department of Industry, Science, Energy and Consumption. (ND)
https://www.energy.gov.au/government-priorities/energy-data/national-energy-analytics-research

International Energy Agency. (2019, Jun). Case Study: The National Energy Analytics Research (NEAR) Programme in Australia. https://www.iea.org/articles/case-study-the-national-energy-analytics-research-near-programme-in-australia

James Hill. (2019, Oct). National Energy Analytics Research Program (NEAR).
http://storage.rdbk.com.au.s3-ap-southeast-2.amazonaws.com/events/REED/2019-All-Energy/ENERGY%20EFFICIENCY/ENERGY%20TRANSFORMATON%20THEATRE-MARKET%20TRENDS%20AND%20DEVELOPMENTS/Thu%20ETT%201035%20James%20Hill.pdf

7 — Program to Calculate Deforestation in the Amazon (PRODES)

Main subject involved
National Institute for Space Research - (Brasil)

Type of organization
Central government - National Agency

Date	**Scale**	**Policy field**
1988	Regional	Forestry

Description
PRODES (Programa de Monitoramento da Amazônia e Demais Biomas) is an operative program to monitor deforestation rates in the Amazon forest. PRODES is one of three similar operative programs (together with DETER and Terraclass) in the initiative "Monitoring Program for the Amazon and Other Biomasses" (Programa de Monitoramento da Amazônia e Demais Biomas). The initiative is developed by the National Institute for Space Research (INPE) of Brazil Government which started PRODES in 1988.

Type of data
PRODES relies on a remote sensing system, getting images from various satellites about the state of the Amazon forest with a 30 meters resolution every 16 days. From that, it calculates the annual rate of deforestation.

How data is used
This indicator supported the proposal and evaluation of public policies, specifically regarding certifications in agricultural businesses, intergovernmental agreements, and serving as an information base for charity.

Sources considered for the analysis

Australian Government Department of Industry, Science, Energy and Consumption. (ND) https://www.energy.gov.au/government-priorities/energy-data/national-energy-analytics-research

International Energy Agency. (2019, Jun). Case Study: The National Energy Analytics Research (NEAR) Programme in Australia. https://www.iea.org/articles/case-study-the-national-energy-analytics-research-near-programme-in-australia

James Hill. (2019, Oct). National Energy Analytics Research Program (NEAR). http://storage.rdbk.com.au.s3-ap-southeast-2.amazonaws.com/events/REED/2019-All-Energy/ENERGY%20EFFICIENCY/ENERGY%20TRANSFORMATON%20THEATRE-MARKET%20TRENDS%20AND%20DEVELOPMENTS/Thu%20ETT%201035%20James%20Hill.pdf

8 — Tackling opioid crisis through public health data

Main subject involved U.S. Department of Health and Human Services - (United States)		
Type of organization Central government - Ministerial Department/Office		
Date 2017	**Scale** National	**Policy field** Epidemic

Description
The high overdose mortality rates due to excessive prescription and abuse of opioid-based medicaments urged the U.S. Department of Health and Human Services (the main national health department) to launch a national strategy in 2017 to tackle the so-called "Opioid Crisis." One of the main points in the strategy advocates for a more timely data collection, encouraging data sharing practices among agencies to support policy-makers from various public agencies. To do so, the various public health agencies gathered and made publicly available their data through reports, official websites, and interactive visualizations. Among the specific efforts developed as part of this strategy, we can mention two. The Centers for Medicare & Medicaid Services (CMS) (administrative agency for healthcare in the U.S) developed a web-based tool to visualize opioids prescribing rates among different states. The Agency for Healthcare Research and Quality developed dashboards and other specific evidence-based resources on opioids through data on individual patients, for example, through out-patient and surgery services.

Type of data
Data mostly come from health and law enforcement. Sources of health data are: PDMPs, state coroners/medical examiners, death certificates (e.g., information about the cause of death, the drugs involved, and demographic information, such as age, sex, marital status, race/ethnicity, and education level), Medicaid and other claims, insurance providers, and state emergency services (e.g., number of overdoses, treatment admissions, emergency department admissions, and the number of 911 and poison control calls).
Law enforcement data sources are: e state forensic labs, drug arrests, drug seizures, crime or incident reports, and urinalysis results.

How data is used
Building better evidence for the HHS response and enhance the evidence-base

Sources considered for the analysis

Dullabh P, Dhopeshwarkar R, Heaney-Huls K, Hovey, L, Rajendran N, Moriarty E, Steiner C. (2019) Building the Data Capacity for Patient-Centered Outcomes Research: The 2018 Annual Report. https://aspe.hhs.gov/system/files/pdf/259016/2018PortfolioReport.pdf

Martinez, C. (2018). Cracking the code: Using data to combat the opioid crisis. The Journal of Law, Medicine & Ethics, 46(2), 454-471. https://journals.sagepub.com/doi/pdf/10.1177/1073110518782953

Sanders, E., Dullabh P. & Dhopeshwarkar R. (2019, October) Addressing the opioid epidemic with better data: an overview of HHS priorities and projects to expand data capacity for patient-centered outcomes research on opioids. University of Chicago. https://aspe.hhs.gov/system/files/pdf/263391/AddressingOpioidEpidemic_PCOR.PDF

U.S. Department of Health and Human Services. (2018, May). Better Data. https://www.hhs.gov/opioids/about-the-epidemic/hhs-response/better-data/index.html

9 — RapidSMS

Main subject involved Rwandan Ministry of Healthcare - (Rwanda)		
Type of organization Central government - Ministerial Department/Office		
Date 2009	**Scale** National	**Policy field** Health

Description
RapidSMS is a governmental program promoted by the Rwandan Ministry of Healthcare to enhance a public health service delivery framework. It aimed at supporting an already established network of volunteers, called community health workers (CHWs), involved as part of a broader national policy to tackle high mortality rates of mothers and newborns. The government provided mobile phones to 15.000 CHWs to monitor mothers' and children's health and send timely information to local hospitals in case of emergency. UNICEF firstly piloted the initiative in 2009, and then it was developed on the national scale, becoming part of the health care system infrastructure.

Type of data
“The second mHealth component, mUbuzima, uses interactive voice response (IVR) technology to enable CHW team leaders in each village to submit data on a monthly basis relating to indicators for case management of sick children, nutritional status, vaccinations, supervision, maternal health and deaths at home." (World Health Organization, 2013, p.1)

How data is used
"Together, these two components facilitate realtime decision-making through the aggregation of data into charts and dashboards, and contribute to the national monitoring of the MDG indicators for maternal and child health” (World Health Organization, 2013, p.1).

Sources considered for the analysis
World Health Organization. (2013). Assisting community health workers in Rwanda: MOH’s RapidSMS and mUbuzima. https://apps.who.int/iris/bitstream/handle/10665/92814/WHO_RHR_13.15_eng.pdf;sequence=1

10 — Kennisnet supports the educational system use of learning analytics

Main subject involved
Kennisnet - Ministry of Education, Culture, and Science - (Netherlands)

Type of organization
Public non-financial corporations - Ministerial Department/Office

Date	Scale	Policy field
2011	National	Education

Description
Kennisnet is a public organization fully-funded by the Dutch Government's Ministry of Education, Culture, and Science. Within the Dutch educational systems, "the government sets goals for schools and provides direct funding, but schools are free to decide for themselves how to achieve these goals" (Ferguson et al., 2016, p. 111). For this reason, more than in other European countries, Dutch educational institutions can choose to rely on external vendors of technological/e-learning solutions for implementing their activities.

Type of data
The use of digital solutions allows the collection and analysis of data on students learning paths (e.g., test scores), which is called "learning analytics."

How data is used
Kennisnet has the mission to help schools adopt digital and ICT solutions, and identified learning analytics has a relevant field to develop. The Ministry supported, through Kennisnet, the use of learning analytics to establish the standards of education.

Sources considered for the analysis
Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). Research Evidence on the Use of Learning Analytics - Implications for Education Policy . R. Vuorikari, J. Castaño Muñoz (Eds.). Joint Research Centre Science for Policy Report.
https://publications.jrc.ec.europa.eu/repository/bitstream/JRC104031/lfna28294enn.pdf

11 — The Norwegian Agency for Public Management use transactional data to optimise digital public procurement

Main subject involved Agency for Public Management and eGovernment (Norwegian Ministry of Government Administration and Reform) - (Norway)		
Type of organization Central government - Public Agency		
Date 2012	**Scale** National	**Policy field** Public procurement

Description
The Agency for Public Management and eGovernment (DigDir) working within the Norwegian Ministry of Government Administration and Reform has launched a pilot project to develop a value chain linking data collection via PEPPOL BIS (a European project for data standardization and interoperability for digital procurement) with analysis by an end user with the aim to improve the public procurement process (especially related to ICT services and contracts). The pilot has the support of the Big Data Test Infrastructure developed by the EU. In particular the BDTI collects raw data on transactions (since 2012) and transforms them into data ready for further analysis.

Type of data
The data collected are business transactional data from Norwegian Public Administration.

How data is used
Data analysis is meant to identify areas of inefficiency and bottlenecks in order to improve general operations and e-procurement performances.

Sources considered for the analysis
Russo, F. (2021, March). Improving public procurement processes with data analysis. Big Data Test Infrastructure Community Portal, European Commission.
https://ec.europa.eu/cefdigital/wiki/display/BDUC/Improving+public+procurement+processes+with+data+analysis

12 — Management of public administration personnel data at the national scale

Main subject involved Directorate for Information Systems and Innovation (DSII) . Ministry of Economic and Finance - (Italy)		
Type of organization State government - Ministerial Department/Office		
Date 2016	**Scale** National	**Policy field** Human Resources

Description
The Department of General Administration of personnel and services (DAG) is an internal office of the Italian Ministry of Economics and Finance (MEF) responsible for human resources management in Italian public administration. DAG manages NoiPA, the current human resources system of central and local Public Administration and the associated welfare and tax compliance systems. The Direction of Informative Systems and Innovation (DSII), one of the five DAG's departments, regularly manages and publicly release various dataset that are used by DAG and MEF for their work. DSII recently launched the Cloudify NoiPA program, that will implement the new system where all the information related to each Italian public employee can be found.
The program complies with the "PON Governance and Capacità Istituzionale 2014-2020". From this funding it received 99.000.000€ for the years 2016-2023.

Type of data
The data includes administrative data shared by public administration agencies: for example, anagraphic information about the public administration employees, public administration organizations (administrative structures), working contracts, absences from work, taxes applied, etc.

How data is used
DAG uses administrative data to manage and control performances of human resources in public administration on national level.

Sources considered for the analysis

Agenzia per la Coesione Territoriale, (ND). Cloudify NoiPA – Il sistema di gestione del personale pubblico. http://www.pongovernance1420.gov.it/it/progetto/cloudify-noipa-il-sistema-di-gestione-del-personale-pubblico/

Dipartimento dell'Amministrazione Generale, del Personale e dei Servizi. (ND). Open Data NoiPA. Ministero Econoomia e Finanza Governo Italiano.https://dati-noipa.mef.gov.it/cl/web/open-data/dataset

Direzione dei Sistemi Informativi e dell'Innovazione. (2017). Cloudify NoiPA Il Progetto di trasformazione digitale di NoiPA. Dipartimento dell'Amministrazione Generale, del Personale e dei Servizi. Ministero Econoomia e Finanza Governo Italiano.https://www.cloudifynoipa.it/documents/20143/672653/Presentazione-Forum-PA/68b1a36f-92fc-0535-16d4-4bac14851d03?version=1.1

ETAPAS Project, (2021). Use case 1: Ethically Responsible Big Open Data (MEF). https://www.etapasproject.eu/usecases/post-1/

13 — A Mobility DataLab in the city of Bergen

Main subject involved
Bergen Municipality - (Norway)

Type of organization
Local government - Municipality

Date	**Scale**	**Policy field**
2017	City	Mobility

Description
(COPY) The purpose of MUST DataLab is to contribute to insight and innovation using available transport and mobility data. The database in MUST contains, for example, passenger data from public transport which says something about which routes most people travel, how many they are and when in the day the passengers are in motion. Correspondingly, there is data on city bikes in Bergen and on car traffic. It points out where the use and load is great, and when in the day we see this. When we then combine this data with information about, for example, road work and weather and driving conditions, opportunities are created for new insights and new services.

Type of data
Type of data
Bergen City Bike
Shuttle (passenger bus / train number per line and stop, etc.)
Public APIs from the Norwegian Public Roads Administration (traffic data, motor vehicle register, travel times, etc.)
Air quality from NILU
The car sharing
Parking data from Bergen Parkering
Employee register from NAV (basic district home / work)
Air traffic from Avinor
Mobility data from Telia (NB! Can not be shared openly.

How data is used
Develop new insights

Sources considered for the analysis
Bergen Municipality. (2021, March). What is MUST Data Lab? http://mustlab.no/must-datalab

14 — Helsingborg Data Lab

Main subject involved
Helsingborg Municipality - (Sweden)

Type of organization
Local government - Municipality

Date	Scale	Policy field
2018	City	Welfare

Description
Helsingborg Social Services (Helsingborg Municipality) is currently developing a project that focuses on using different digital data sources for developing new approaches to social services design. The project is an experimentation, intended to provide evidence to policy-makers, and it has the political support of the Municipality.

The project is called Data Lab and will run until 2022. It is a project not linked to a particular policy, but it is part to the H22 Initiative of Helsingborg (https://h22.se/).

The main challenge of the Data Lab faces is to find insights on individual needs by gathering aggregated data from multiple sources, mostly administrative data (i.e., collected by the administrative offices of the city). The project decided not to focus on data on individual level, but on aggregated data. Nonetheless legal issues must be addressed: the project should construct the appropriate legal framework.

The project asks: how can data be used so that the citizens experience services in a seamless way throughout various subjects?

Type of data
Data comes from the welfare services people apply to.

How data is used
Data are gathered to support the design and development of better social services within the city

Sources considered for the analysis
Information coming from direct interview project manager (key agent)

15 — City of Ghent uses mobile phone data to identify the habitational patterns of students

Main subject involved
Data and Information Service of the Ghent City Council - (Belgium)

Type of organization
Local government - Municipality

Date	Scale	Policy field
2017	City	Mobility

Description
(COPY from official websti) In a EU project pilot, the city of Ghent would like to understand how big data sources can provide valid and accurate information for policy decisions, which may serve as a complement to traditional statistics, but also as entry point to phenomena inaccessible until now by analysing only the traditional registration data. One such example is the large population of students in Ghent, which are not registered in the local population register. Ghent has the biggest student population of all the Flemish Cities (70.000 students). But only 14% of them are registered residents of Ghent; others stay in student lodgings during the week or commute to Ghent. So, the exact number of students in lodgings or their address is unknown to the city. Their mobility behaviour is not known either, this makes it hard to make urban planning and mobility policy around their needs. The Data and Information Service of the Ghent City Council has extensive experience in providing data-analyses and insights to support operational and strategic decisions, and recognises that big data from mobile phones provides a promising source for enhancing accurate population, migration and mobility information to gain new insights. Understanding the mobility behaviour of students will allow for experimentation with this data to create more effective mobility and urban planning policy that meets everyone's needs.

Type of data
Available datasets include mobile data (but a calibration point is needed), Registry, The student lodging dataset, The number of resident non-registered students, and Their mobility behaviour. Other data that can be leveraged for policy visualisations includes information about traffic accidents, traffic density, is available from current data providers. Static traffic accident data and data coming from mobile phone providers. Additional available data includes; Data to check policy implementations like low emission zones. ANPR cameras used for access control of low emission zones. This data can be used to measure the traffic into the city centre; Data to define “trajected control” areas where speed is measured in a zone instead of one spot.

How data is used
Gent will focus on the policy design stage, using data visualisations to identify needs and options. Mobile phone data was used to identify the habitational patterns of students.

Sources considered for the analysis

PoliVisu European Project Consortium (2019). Deliverable 3.9: Experiment driven policy making: pitfalls and suggestions for Public Administrations.
https://05a0e658-0382-45bd-be34-80e9fa8750e9.filesusr.com/ugd/68109f_a4e8693b70b441fb83eefe0c5a7406a4.pdf

16 — xKRP – Community Experience Data Lab Kronoparken

Main subject involved
Research Institutes of Sweden (RISE) - (Sweden)

Type of organization
Public non-financial corporations - Agency

Date	Scale	Policy field
2016	City	ND

Description
(COPY from website) Community Experience Data Lab Kronoparken (xKRP) was a conceptual and mobile open data lab, focused on developing, testing and evaluating visualization, interaction and use of data, by the local community.

The project manager is Research Institutes of Sweden and RISE Service Labs. Project partners are the County Council of Värmland, The Swedish Consumer Agency, Karlstad University, Karlstad Innovation Park, the NGO Ett öppnare Kronoparken, Canwz AB, Thindermaps AB and RISE Interactive Institute

Type of data
ND

How data is used
This initiative involved developing, testing and evaluating visualization, interaction and use of data where the local community is the user. The investment itself took place in the Kronoparken district in Karlstad, a so-called million program area. Within the framework of the data lab, everything from interactive screens and locally developed applications to new ways of sharing data were tested.

Sources considered for the analysis

Frick, A. (2019). Datalabb – för ett smartare Sverige. VINNOVA.
https://blogg.vinnova.se/wp-content/uploads/2019/12/E-bok_om_datalabben.pdft.

VINNOVA. (2016, March). xKRP - Community Experience Data Lab Kronoparken.
https://www.vinnova.se/p/xkrp---community-experience-data-lab-kronoparken/