European Commission

JRC SCIENCE FOR POLICY REPORT

DRAFT

DigiTranScope: The governance of digitally-transformed society

Craglia M., Scholten H., Micheli M., Hradec J., Calzada, I., Luitjens, S., Ponti M., Boter J.

2020



Joint Research Centre

EUR XXXXX XX

This publication is a Science for Policy report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information [optional element]
Name:
Address:
Email:
Tel.:

EU Science Hub
https://ec.europa.eu/jrc

JRCXXXXXX

EUR XXXXX XX

PDF	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX
Print	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX

Luxembourg: Publications Office of the European Union, 20XX [if no identifiers, please use Brussels: European Commission, 20XX or Ispra: European Commission, 20XX or Geel: European Commission, 20XX or Karlsruhe: European Commission, 20XX or Petten: European Commission, 20XX or Seville: European Commission, 20XX depending on your unit]

© European Union/European Atomic Energy Community, 20XX [Copyright depends on your directorate, delete as applicable: European Atomic Energy Community for Dir. G, European Union for rest of JRC]



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union/European Atomic Energy Community [Copyright depends on your directorate, delete as applicable: European Atomic Energy Community for Dir. G, European Union for rest of JRC], 20XX, except: *[page XX, artist's name, image #], Year. Source: [Fotolia.com]*
(unless otherwise specified)

How to cite this report: Author(s), *Title*, EUR (where available), Publisher, Publisher City, Year of Publication, ISBN 978-92-79-XXXXX-X (where available), doi:10.2760/XXXXX (where available), JRCXXXXXX.

Contents

Abstract 4

Acknowledgements 5

Executive summary 6

1 Digitranscope: An Introduction 7

1.1 About Digitranscope and the Centre for Advanced Studies 7

1.2 About Digitranscope 7

1.3 The Shifting political landscape 8

1.4 Structure of the project and of the book 10

2 For the benefit of all!? Some notes on the role of government in the digital transformation. 12

2.1 Introduction 12

2.2 Three dominant strategies 13

2.3 A European way? 14

2.4 Action perspectives 16

2.5 Conclusion and closing remarks 18

3 Data governance models emerging from the practices of social actors 22

3.1 Introduction 22

3.2 Data governance: A social science-informed definition 23

3.3 Research Strategy 23

3.4 Emerging data governance models 24

3.4.1 Data sharing pools 24

3.4.2 Data cooperatives 25

3.4.3 Public Data Trusts 25

3.4.4 Personal data sovereignty 26

3.5 Discussion 27

4 Seeing Platform and Data Co-operatives Through European Pandemic Citizenship 31

4.1 European Pandemic Citizenship at Stake 31

4.2 Co-operatives as a Collective Resilient Response to the Pandemic Crisis 31

4.3 Objective of the study 32

4.4 The Research Question and Two Main Hypotheses 32

4.5 Co-operatives and Pandemic Citizenship 33

4.6 Towards a Taxonomy for Platform and Data Co-operatives 34

4.7 Final Remarks 36

5 Citizen-generated data for policy: A review of EU projects 41

5.1 Introduction 41

5.1.1 Citizen-Generated Data Projects 41

5.1.2 Governance of Data and Data for Governance 42

5.2 The key-features of European CGD projects 43
5.2.1 Purpose of data collection 43
5.2.2 Categories of projects and types of data 43
5.2.3 Use of CGD by the public sector 44
5.3 Findings from interviews 45
5.4 Discussion 46
5.4.1 New forms of public participation and emerging roles for citizens and local governments 46
5.4.2 Rethinking relationships between citizens and local governments 47
5.4.3 Generating richer and trusted data to address current challenges 47
5.5 Conclusion 48
6 Private sector data for the public interest: Data sharing practices in twelve EU cities 51
6.1 Introduction 51
6.2 How local administrations access to private sector data 51
6.3 Methodology 52
6.4 Results 53
6.4.1 Data donorship: “Free lunch?” 53
6.4.2 Public procurement of data: Resistance and caution 54
6.4.3 Data partnerships: Win-win collaborations 55
6.4.4 Tender obligations: Data sovereignty 55
6.5 Discussion 57
7 Probabilistic Synthetic Population modelling for EU policy support 60
7.1 Introduction 60
7.2 Review of Methods 60
7.3 Applications 61
7.4 Conclusions 64
8 Semantic Text Analysis Tool (SeTA) 67
8.1 Introduction: What is SeTA and why it is important 67
8.2 The technology 68
8.3 A different approach to summarising text 70
8.4 Stepping into the future 71
8.5 Conclusions 71
9 Digitranscope Experiments: Digital Twins and Smart Cities, case studies of Amsterdam and Duisburg 73
9.1 Introduction 73
9.2 Amsterdam case study 73
9.2.1 Problem Statement 73
9.2.2 Technological readiness 74
9.2.3 Implementation 74
9.2.4 Partners 75

9.2.5 Citizens 75
9.3 Duisburg case study 76
9.3.1 Problem Statement 76
9.3.2 Technological readiness 76
9.3.3 Implementation 76
9.3.4 Citizens 77
9.4 Findings 77
10 Involving children in the renewable energy transition: exploring the potential of the digital transformation 80
10.1 Involving children in the energy transition 80
10.2 The Warsaw Experiment 81
10.3 The Amsterdam Experiment 82
10.4 Using gaming in digital twin setting: opportunities and limitations 84
11 Conclusions 88
List of boxes 91
List of figures 92
List of tables 93
Annexes 94
About the Authors 94

Abstract

This volume presents the key outcomes and research findings of the DIgitranScope research project of the European Commission Joint Research Centre. The project set out to explore during the period 2017-2020 the challenges and opportunities that the digital transformation is posing to the governance of society. We focused our attention on the governance of data as a key aspect to understand and shape the governance of society. Data is a key resource in the digital economy, and control over the way it is generated, collected, aggregated, and value is extracted and distributed in society is crucial. We have explored the increasing awareness about the strategic importance of data and emerging models to distribute the value generated more equitably in society. These findings have contributed to the new policy orientation in Europe on technological and data sovereignty and social inclusion. The digital transformation, and the rise of artificial intelligence and the Internet of Things, offer also new opportunities for new forms of policy design, implementation, and assessment providing more personalised support to those who need it and being more participative throughout the policy cycle. The use of digital twins, gaming, simulation, and synthetic data are just at their beginning but promise to change radically the relationships among all the stakeholders in governance of our society.

Acknowledgements

The authors are particularly grateful to Mr. Vladimir Sucha, former Director General of the European Commission Joint Research Centre for believing in the scope of the project and his encouragement in setting it up. We are also grateful to Jutta Thielen-Pozo, head of the Scientific Development Unit, and all her wonderful colleagues for her leadership and support, and to Alessandro Annoni, former head of the Digital Economy Unit for his help in connecting to policy colleagues and transitioning the activities of Digitranscope into his Unit to develop further in the years to come.

A special thanks goes to Prof. Michael Blakemore for his important contribution in the formative year of the project, and Prof. Peter Lloyd for his constant curiosity, engagement, and insights.

Authors

Massimo Craglia, European Commission Joint Research Centre

Henk Scholten, Free University of Amsterdam and GEODAN

Marina Micheli, European Commission Joint Research Centre

Jiri Hradec, European Commission Joint Research Centre

Igor Calzada, European Commission Joint Research Centre

Steven Luitjens, Ministry of the Interior and Kingdom Relations, Government of the Netherlands

Marisa Ponti, European Commission Joint Research Centre

Jaap Boter, Free University of Amsterdam.

Contributors

Anna Berti Suman, Dutch Research Council

Margherita Di Leo, Arcadia SIT srl.

Corentin Kusters, GEODAN

Magdalena Grus, Free University of Amsterdam

Sanne Hettinga, Free University of Amsterdam

Executive summary

This volume presents the key outcomes and research findings of the DIgitranScope research project of the European Commission Joint Research Centre. The project set out to explore during the period 2017-2020 the challenges and opportunities that the digital transformation is posing to the governance of society.

Policy context

The development of the project coincided with an increasing recognition of the importance of Artificial Intelligence (AI) to master the increasing volumes of big data available on a daily basis. The control of AI and of the data underpinning its development are strategic for the future development of society, and the focus of an increasing geopolitical competition. The European Union has identified technological and data sovereignty as key priorities for Europe and developed several policy initiatives to strengthen its regulatory framework and increase its preparedness to address the dual digital and green deal transformations.

Key conclusions

There are many dimensions to address the governance of a digitally-transformed society and the project focussed on the governance of data as a critical aspect. Data is a key resource in the digital economy, and control over the way it is generated, collected, aggregated, and value is extracted and distributed in society is crucial. We have explored the increasing awareness about the strategic importance of data and emerging models to distribute the value generated more equitably in society. These findings have contributed to the new policy orientation in Europe on technological and data sovereignty and social inclusion.

At the same time, the digital transformation, and the rise of artificial intelligence and the Internet of Things, offer also new opportunities for new forms of policy design, implementation, and assessment providing more personalised support to those who need it and being more participative throughout the policy cycle. The use of digital twins, gaming, simulation, and synthetic data are just at their beginning but promise to change radically the relationships among all the stakeholders in governance of our society.

Main findings

With respect to the governance *of digital data*, we identified four emerging models of data governance: data sharing pools, data cooperatives, public data trusts, and personal data sovereignty. The goal of our analysis has been to investigate to what extent they support different, more balanced, power-relations between actors compared to the dominant one of datification, in which few dominant actors get most of the value. From the analysis we conclude that data cooperatives and personal data sovereignty are particularly promising in trying to redistribute the value generated from data more equitably across actors and society. Both models however need policy support and incentives to become established and scale-up.

With respect to the governance *with the digital transformation* we have explored the use of synthetic population, digital twins, and gaming environments. The development of synthetic populations through AI and machine learning methods results in an artificial set of individuals, families, and households with the same characteristics and behaviour of the true population. This allows the design, modelling and testing of citizen-centred policies, targeted to those who need intervention most without the use of personal data. We found that the opportunities are very significant, and for this reason many governments and statistical agencies are becoming interested in this methodology. The concept of digital twins has been known and applied for many years in manufacturing, creating a digital replica of an artefact for testing and assessment before going into production. The increased availability of data, and processing power at declining costs, makes it now possible to develop digital twins for entire cities and nations. We discuss the use of digital twins in Amsterdam and Duisburg to address local problems and found them effective tools to communicate with all the stakeholders involved from government officials to business and the public. We tested also the combined use of digital twins and gaming environments to engage school children in the energy transition and urban planning, and found this combination as having many opportunities to get the citizens of today and tomorrow to have a say in shaping their environment. We conclude that technological change is much faster than the ability of governments to react. Therefore, is necessary to anticipate and shape the future direction of development through foresight studies, qualitative research, and experimenting with new technology and methods, rather than trying to fix the present that too quickly becomes the past. Governments play a key role, but it is ultimately up to all of us to shape our futures.

1 Digitranscope: An Introduction

M. Craglia and H. Scholten

1.1 About Digitranscope and the Centre for Advanced Studies

Digitranscope is a three-year research project of the Centre for Advanced Studies at the European Commission Joint Research Centre. The Centre for Advanced Studies (CAS) of the European Commission's Joint Research Centre (JRC) was created in 2016 improve the connections between science and policy and help inform better and influence the regulatory frameworks needed to address the new and emerging societal challenges confronting the European Union and our societies as a whole. The CAS aims at creating the conditions necessary for innovative and interdisciplinary research, as well as offering a creative and generative space in which ideas and knowledge in emerging thematic fields across different scientific and technological disciplines can thrive and flourish. As such, the CAS has become an incubator for formal inquiry, stimulating ideas and activities and providing the JRC with new insights, data projections and solutions for the increasingly complex medium and long-term challenges facing the EU, especially in the fields of artificial intelligence, demography, big data and digital transformation.



1.2 About Digitranscope

DigiTranScope originated from the JRC Strategy 2030. The strategy identified ten strategic topics on which the JRC should concentrate to anticipate future policy requests. One of these topics was Data and Digital Transformation to address which the JRC set up a transversal project on Artificial Intelligence and Digital Transformation led by the Digital Economy Unit, as well as a research project, more exploratory in nature, on digital transformation to be implemented in the CAS. The project originally proposed to address two key issues: i) how the information glut triggered by the digital transformation reverses the cognitive balance between humans and machines, and ii) the impact of digital information technology on the rules and institutions that guide modern societies. The proposal led to the establishment in 2017 of two projects: Humaint on human behaviour and machine intelligence and DigiTranScope, on digital transformation and the governance of human society.

The objectives of DigiTranScope are:

- to explore the changing flows, ownership, quality and implications of digitised data and information;
- to identify the key policy challenges relating to massive interconnection (the Internet of Things - IOT) and the associated opportunities and risks;
- to determine what skills are needed to live fulfilling and healthy lives in a digitally transformed society, and to explore how to offer all citizens the opportunity to develop these skills, and;
- to explore innovative forms of governance for Europe leveraging the characteristics of digital transformation.

At the early stages of the project, we recognised that the governance of digitally transformed societies revolves to a large extent around the governance of data: those who control the production, integration, use and dissemination of data have formidable levers of power in today's digitised society. What we did not realise at the start of the project is that what we thought was going to be an exploratory project looking 5-10 years ahead turned instead into one providing already direct input to policy as policy priorities shifted much faster than we anticipated. We highlight some of these shifts in the next section.

1.3 The Shifting political landscape

On Data Governance: The most significant event that occurred during the lifetime of the project was the emergence of Artificial Intelligence (AI) as a key geopolitical battleground, particularly between the US and China. This brought AI also at the forefront of the European political attention with an initial strategy on AI adopted by the Commission in April 2018 (EC. 2018a), followed by a Coordinated Plan with the Member States in December 2018 (EC, 2018b), the establishment on a High-Level Expert Group[1] to advise on the development of ethical guidelines for AI and priority areas for investment in 2019, and an AI White Paper in February 2020 (EC, 2020a) setting the framework for a consultation on a risk-based regulatory framework for AI. Why is this important? Because it immediately became clear that data is the absolutely key asset underpinning the development of AI, and that to govern the future development of the technology it is necessary also to govern better European data. Technological and data sovereignty became key objectives of the new Commission which took office in November 2019.

As a result of this increased attention to data, the Commission stepped up its work on Business-to-business (B2B) and Business-to-government (B2G) data sharing that had started in 2017 with the Communication 'Building a European Data Economy' (EC, 2017) and then the COM Towards a Common Data Space (EC, 2018c). These Communications had provided a set of guidelines but now it seemed appropriate to develop them further into a regulatory framework. The Commission therefore, organised several workshops and studies on data governance which contributed to the European Strategy for Data published (EC, 2020b) in February 2020 which establishing several European common data spaces in different thematic domains (e.g. environment, health, agriculture, automotive, finance, etc.) including both public sector and commercial sector data, and a Regulation on the governance of data (Data Governance Act[2]) launched in December 2020. The latter extends the categories of public sector data available for reuse, creates the framework for the sharing of business data, facilitates the reuse of personal data via data intermediaries or on altruistic grounds.

From data sharing to data analytics: It is important to note that the evolution of technology, and the lessons learned from the bog commercial platforms, has brought also a change not just in *what* is shared but *how* data is shared. In the past 20 years data publication for reuse, via catalogues and portals, was seen as the end of the process of data collection, analysis and use by the (public sector) data custodian. It was often also perceived as a burden because the organisation publishing the data was not a direct beneficiary of the value subsequently generated by third parties or accrued by society as a whole in terms of greater transparency and accountability. Observing the big data platforms at work it became noticeable that to them data publication was the beginning of the value-creation chain, not the end! In fact, social media platforms and search engines, do not even create the original data, they let the users do so. They then integrate the users' data, add value through analytics, repackage into products or services, and sell to third parties thus monetizing the added value created. This shift in the datafication paradigm (Mayer-Schoenberger and Cukier, 2014; Ericsson, 2014) has led to an increasing call for public authorities to also shift from publishing datasets in portals to open access via machine-to-machine Application Protocol Interfaces (APIs) interfaces and add value to the data they publish by adding the intelligence via analytics on who uses the data for what purpose (Vaccari et al. 2020).

From data analytics to digital twins: There are many definitions of digital twins but a generic one is provided by El Saddik (2018) as follows: *A digital twin is a digital replica of a living or non-living physical entity. By bridging the physical and the virtual world, data is transmitted seamlessly allowing the virtual entity to exist simultaneously with the physical entity.*

Digital twins have been used in industry for several decades, largely as an extension of computer-aided design. They allow simulation and testing of artifacts before moving into production. With the development of Industry 4.0 and the vast increase in sensors networks and computer processing digital twins they have seen a significant growth in every sector, as show in this classification by ISO.

1 https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
2 https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act

Table 1.1. Industries and applications of Digital Twins

Industry	Applications
Manufacturing	• Design verification • Predictive maintenance • Process optimization • Safety management • Equipment utilization, etc.
Energy	• Power monitoring and management • Failure analysis • Grid operation and maintenance, etc.
Smart cities	• Transportation monitoring • Urban planning • Strategy evaluation • HVAC (heating, ventilation and air conditioning) control, etc.
Farming	• Planting optimization and monitoring, • Machine and products tracking • Pesticide monitoring, etc.
Building	• Progress monitoring • Budget control and adjustment • Building quality assessment • Worker safety monitoring • Resource allocation and waste tracking, etc.
Healthcare	• Health monitoring • Personalized medicine • Medical resource allocation, etc.

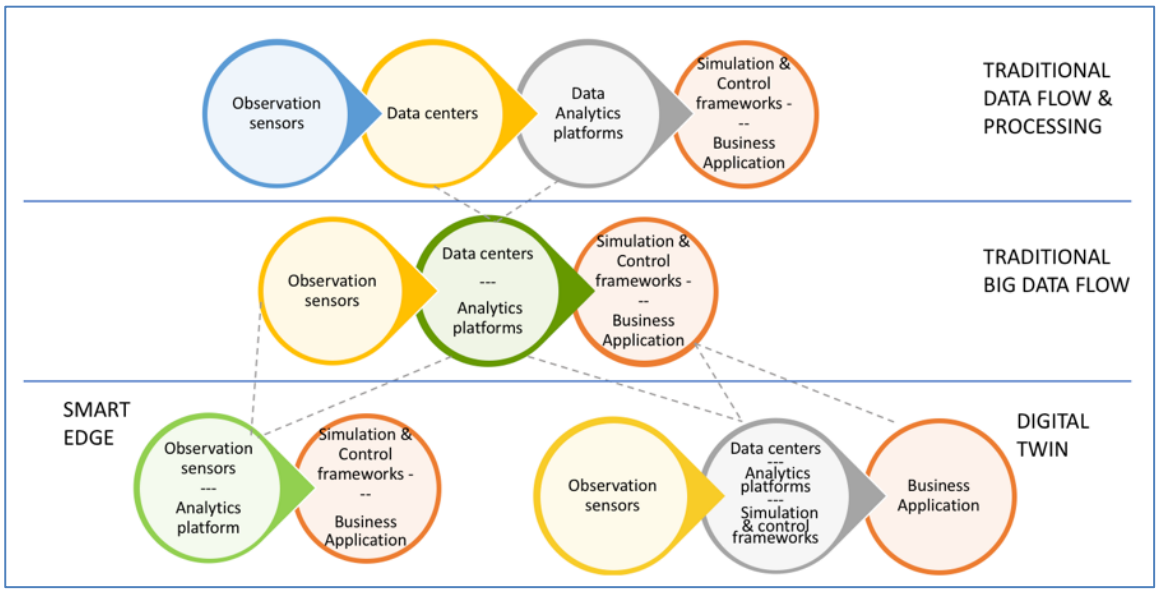
Source; Nativi, Craglia, Delipetrev, 2020

With the much-increased availability of high-resolution data from space and airborne instruments, sensor networks, public administrations, and the general public, we have seen a rapid development of digital twins also for urban areas, and in environmental applications. A survey of digital twins in the environmental domain by Nativi, Craglia and Delipetrev (2020) indicates that indeed urban management and earth system modelling are the two more prominent areas. European policy supports these developments with new initiatives to develop a smart cities' ecosystem[3] and Destination Earth[4] aiming at developing a very high-precision digital model of the Earth to monitor and simulate natural and human activity, and support European environmental policies.

With the deployment of 5G networks and the diffusion of the Internet of Things we are likely to see a step change in the diffusion of the digital twin paradigm, which represent the evolution of the big data analytics discussed earlier. This evolution is depicted in Fig 1 below. As shown, the shift from traditional data processing to big data already required a closer integration of data processing and analytics into a single (virtual) platform. The development of the IoT and edge computing brings analytics and processing already at the level of the sensor collecting the data while the development of digital twins integrates simulation into the data processing and analytics platforms strengthening the move towards dynamic and interactive environments based on data streams and feedbacks-loops (see Nativi et al. 2020)

Figure 1.1: Evolutionary landscape in data analytics

[3] https://ec.europa.eu/digital-single-market/en/smart-cities-smart-living
[4] https://ec.europa.eu/digital-single-market/en/destination-earth-destine



Source: Nativi S. and Craglia M. (2020)

On new forms of policy design: significant policy shifts have also emerged in this area during the last few years and become increasingly mainstream. Notably, the increasing use of big data analytics to profile and nudge voters following the example of the commercial sector recognised not only the power of data but also the emotional side of decision making. The mantra of evidence-based decision-making that was all the rage in the 1990s has come under increasing scrutiny together with the scientific method when applied to social and political phenomena (Funkowitz et al. 2016). We have seen therefore a greater acknowledgement of the multi-faceted dimensions of rationality, decision-making, and post-normal science. Communication, participation, and the use of narratives have gained currency exploiting also the new opportunities of the digital transition, from the booming of citizen-generated content for science and policy to the development of digital twins for policy simulation, co-creation, and communication.

1.4 Structure of the project and of the book

Against the background of the rapidly-evolving technological and policy landscape highlighted above, we decided to structure the Digitranscope project on two main tracks.

A. The first track investigated issues around the *governance of digital data* including the role of government in emerging models of data governance; citizen-generated data for public policy; citizenship and data cooperatives, and the perspectives of city governments in accessing and using data held by the commercial sector.

B. The second track investigated new forms of *governance with digital data* including experimenting with the use of publicly available data for profiling and the design of policies targeted to specific needs and groups, and used this in the context of energy transition and the risk of infection in the COVID-19 crisis. We have organized experiments to involve children in energy transition through digital twins in controlled gaming environments. We used the emerging lessons from the deployment of the Internet of Things and digital twins for "smart" cities to develop a City Operating System and used AI methods to extract knowledge from policy documents and apply it in the context of impact assessment.

The book is organised following these two tracks. The first track includes four chapters addressing the data governance models emerging from the practices of social actors by Marina MIcheli and colleagues, a deep dive into one of these models, that of data and platform cooperatives by Igor Calzada, followed by a review of EU projects on citizen-generated data for policy by Marisa Ponti, and an analysis of the practices of data sharing between the commercial sector and local government in twelve cities by Marina Micheli.

The second track explores the use of probabilistic synthetic populations for policy modelling by Jiri Hradec and colleagues, and the use of semantic text analysis for policy assessment also by Jiri Hradec, followed by the reports of two experiments carried out in the project, one on the case studies of Amsterdam and Duisburg's digital twins by Coren Kusters and Henk Scholten, and the other on the use of gaming to Involve

children in the renewable energy transition by Jaap Boter and colleagues. We conclude the book with a chapter reflecting on the key messages and lessons learned.

Before diving into these two worlds of the policy of, and with, the digital transformation Steven Luitjens, a senior programme manager in the Dutch government in charge of its digital transformation programme, helps us to put all the contributions that follow into context with a personal reflection on the role of government in the current digital transformation.

References

El Saddik A. Digital Twins: The Convergence of Multimedia Technologies *IEEE Multimedia* 25(2), 2018. https://ieeexplore.ieee.org/document/8424832

Ericsson. The impact of datafication on strategic landscapes. Ericsson report, 2014.

European Commission 2017. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Building A European Data Economy*. COM/2017/09 final. Publications Office: Luxembourg

European Commission. *Communication: Artificial Intelligence for Europe*. 2018a. https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

European Commission. *Coordinated Plan on Artificial Intelligence*. 2018b. https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence

European Commission 2018c. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Towards a Common Data Space*. COM/2018/232 final. Publications Office: Luxembourg. https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-232-F1-EN-MAIN-PART-1.PDF

European Commission. *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. COM(2020)65final. 2020a. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *European Strategy for Data* COM/2020/66 final https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

Funtowicz S, Ravetz J. Science for the post-normal age. *Futures* 31: 739–755, 1993. http://www.uu.nl/wetfilos/wetfil10/sprekers/Funtowicz_Ravetz_Futures_1993.pdf.

Mayer-Schönberger, V., and Cukier, K. *Learning with Big Data; The Future of Education*. Eamon Dolan/Houghton Mifflin Harcourt. 2014

Nativi S. Craglia M. and Delipetrev B. Destination Earth: Survey on "Digital Twins" technologies and activities, in the Green Deal area. Publications Office of the European Union. 2020. https://publications.jrc.ec.europa.eu/repository/handle/JRC122457

Nativi S. and Craglia M. *Destination Earth*. Presentation at the Kick-Off Meeting, March 6th 2020.

Nativi S. Kotsev A., Scudo P. et al. IoT 2.0 and the INTERNET of TRANSFORMATION (Web of Things and Digital Twins): A multi-facets analysis. Publications Office of the European Union. 2020.

Vaccari L. Posada M., Boyd M., et al. 2020. *Application Programming Interfaces in Governments: Why, what and how*, EUR 30227 EN, Publications Office of the European Union, Luxembourg. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/application-programming-interfaces-governments-why-what-and-how

2 For the benefit of all!? Some notes on the role of government in the digital transformation.

Steven Luitjens[5]

2.1 Introduction

In the autumn of 2014 I was invited to reflect on the findings of two U.S. consultants during their study tour to Silicon Valley exploring the latest trends. At the time I was head of the agency Logius, responsible for the generic digital infrastructure of the Dutch government for service delivery to citizens and companies. I was surprised to see that my strategy was very much in line with the common approach in Silicon Valley: launching new open standard software solutions, industrializing and commoditizing them and making them the *de facto* standard for all to use before anyone else does it (Wardley, 2014). My reflection resulted in adding an extra chapter to the report, sketching what Logius was doing especially in the field of the generic authentication solution DigiD. It was the only chapter that was not about private initiatives and this was probably the justification for opening the chapter with an astonishing question for me: "What role, if any, should the government have in developing, overseeing and when necessary regulating the explosive development of the Matrix?" (Moschella, and Mead, 2015).

In January 2019 a quite similar study tour to Silicon Valley was made by a group of Dutch civil servants. Software-as-a-service was no longer the issue, it was all about artificial intelligence and big data. One of the most intriguing findings was that by now enormous amounts of data -including location data on individuals, companies and strategic resources- were commercially for sale on an already quite mature market for everyone who is interested. When they asked whether the companies collecting and selling these data feel and indeed have any responsibility for what is being done with it, the reaction was always the same: "That's a very good question!". The companies made it absolutely clear that they take no responsibility whatsoever for the usage. The established data economy in Silicon Valley appeared to be a simple, unregulated and unsupervised market of demand and supply[6].

The popular phrasing of data being the new oil is at least true in the sense that the current datafication phase[7] in the ongoing digitization has resulted in the lightning-fast rise of a whole new generation of companies that are tirelessly inventive in discovering new applications for their technology almost every day, thus broadening their scope and impact to all kinds of different sectors that don't know what hits them and how to react. Datafication rewrites the rules of almost every game in the book, without society fully recognizing and understanding it and without government effectively acting on it. In the even more popular anonymous tripartite division of organization types, government is certainly not the one making things happen in the datafication age, but at best the one watching things happen and mostly just the one wondering what's happened.

There is growing consensus, at least in the EU and individual Member States, that governments urgently have to play the full role in the information society that they traditionally had in industrial society[8]. This role is not just fixing market failure, as often suggested by the private sector. Particularly in the datafication era, much is at stake that requires government attention and intervention. The COVID-19 crisis, for one, has proven not only to experts that our information position now fully depends on the Big Tech firms, although the algorithms they use are a black box and the data they provide is data we ourselves apparently have given them. So, rhetorically speaking, are we sure that we're in the right game to begin with?

5 The author wishes to express his gratitude to emeritus professor Peter E. Lloyd for his comments on the draft version of this chapter.

6 https://www.geonovum.nl/uploads/documents/Geonovum_Silicon_Valley_Studiereis_Rapport_eindversie%20%28080519%29.pdf

7. See also https://en.wikipedia.org/wiki/Datafication.

8 See for example https://www.rathenau.nl/en/digitale-samenleving/urgent-upgrade; and https://www.rathenau.nl/en/digitale-samenleving/directed-digitalisation.

2.2 Three dominant strategies

Continued digitization is now systemically disrupting socio-economic structures, revenue models and the power balance in societies worldwide. Especially since the datafication phase and the emergence of the Internet of Things, we are confronted with hyper-complexity and hyper-connectivity. The main feature of the current situation is, that a small number of platforms own unbelievable amounts of data on almost everyone and everything, anytime and anywhere. These platforms are revolutionizing the future of societies in ways that are not all beneficial to say the least.

In the DigiTranScope project we identified three dominant strategies in datafication:

1) to maximize financial profits: datafication as the easiest way in history for privately-owned platforms to make unprecedented amounts of money - the dominant paradigm in the US;

2) to maximize political control: datafication as the easiest way in history for government-controlled platforms to gain unprecedented state power over society - the dominant paradigm in China;

3) to maximize confusion: datafication as the easiest way in history for everybody who has an interest in creating massive insecurity, uncertainty, distrust and outright chaos and to destabilize society wherever and whenever they like - the dominant paradigm for all sorts of individuals and groups, including state actors.

All three strategies use the same technology, combined with insights from psychology and sociology. It is behavioural economics at work in the fullest sense of the word (Aldred, 2019), skilfully making use of concepts from crowd or mob psychology like the deindividuation theory and the bystander effect. Central in datafication is the need to collect and combine, in one way or the other, as many personalized data as possible in order to make all sorts of assumptions and the best possible predictions. Gathering ever more data is essential for data platforms. In fact, they are really insatiable hungry.

When we take the first strategy, following the money -seeing how the major players make their profits- is the most direct way to understand what's happening. In recent years, the evidence is piling up that at least the US Big Tech firms are really only in IT for the money and that they go very far indeed to maximize profits for their shareholders and nobody else. They do this with all their might, showing great techno-optimism and using a very instrumental image of citizens as simply docile consumers, that they can manipulate to do whatever they want them to do (Aldred, 2019; Zuboff, 2019). Their products and services are addictive-by-design[9] and the sales philosophy is largely the drug dealer model[10]. Furthermore, they do not shy away from actively discrediting and even sabotaging efforts from governments to regulate them. As far as at least some are concerned, the very existence of government will be redundant in a mature information society (Zuboff, 2019). Apparently they have alternative facts when it comes to the vital role governments have played and still play in risky financing basic research and in constantly developing the basic infrastructure they all use and benefit from (Mazzucato, 2018a).

The second strategy is essentially the same when comparing the underlying principles and insights, but the objectives differ. Where the first strategy is aimed at creating a perfect surveillance society run by a few private companies (Zuboff, 2019), the second strategy is aimed at creating a perfect surveillance society run by government. For most Europeans this seems unthinkable. At the same time, we see a fast-growing number of cameras in our own streets and other, less visible monitoring devices in all EU countries. So is it a discussion of scale on which we permit surveillance or is there more to our dislike of what happens outside the EU zone? The Chinese government is, compared to us, at least quite transparent to its citizens about what they're doing.

At first, we did not distinguish the third strategy as a separate one. But looking at its clearly different objectives, its impact and the professional and industrialized way in which it is deployed by now, we decided it should definitely be added to the other two. The third strategy is the most sinister one, with many faces that constantly keep changing. It has become an arms race with quite a colourful collection of weapons. We have seen already for years daily hacking attempts and successful cyberattacks to disrupt social and economic structures. And we know about the existence of the dark web for all sorts of criminal trades, or the market of DDoS attacks-as-a-service for sale by anyone who pays a reasonable price. In

9 https://www.theguardian.com/technology/2018/jan/08/apple-investors-iphone-addiction-children.

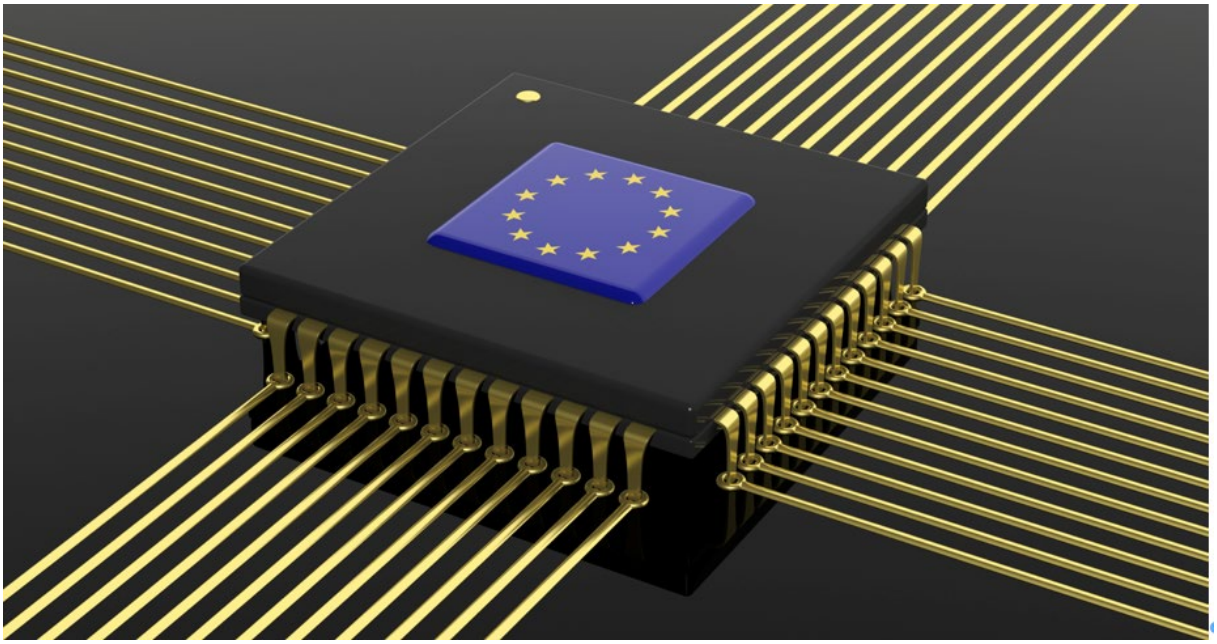
10 Earning revenues in three steps: first you make life easy for people, let them try and give it for free, second you make them addictive, and third you start taking their money (or data!).

addition to these increasingly grim threats, the datafication age has brought us some new challenges that have become very serious, such as the enormous spread of fake news, the rise of influencing elections, of cyberbullying, or of using social media for online shaming, criminalizing and cancelling people [3]. Fostering conflict and accepting inequality as a given is a prescription for social breakdown. The current technologies easily accelerate this. Going wider, it is an opening to social destabilization and populist-nativist movements and thus a serious danger to democracy.

The conclusion about these three strategies seems crystal clear. Each has, in addition to its disruptive effects and amongst other things, serious consequences for feeding inequalities, for undermining solidarity, for the future of work (although it is still very much in debate how this will precipitate in our daily lives exactly) and for exercising our fundamental human rights (like the right to liberty, the right to privacy and the right to self-determination). The days are long gone, that digitization was the promising fresh way of innovative techies to improve service delivery and realize cost cuttings. The real focus of especially the major innovators and frontrunners has dramatically shifted in the datafication age, and as a result we are now trapped in a completely different situation. Analysing them in terms of objectives, underlying principles and effects from the perspective of our European public norms and values, each strategy is essentially abject and reason for serious concerns. You don't have to be a seasoned pessimist, a left-wing activist or technophobic to be at least a bit cautious[11]. The third strategy is the most imminent and acutely dangerous. With so many lies, half-truths and deliberately false messages, we lose our sense of morality. This is not just a philosophical point; it is practical. People have died because they don't know what to believe with COVID-19 news. Experts like Francesca Bria and her group stand out by acknowledging this already for some time, but the overwhelming power of the data monopolists makes it a sideshow. In view of this it is really high time, and hopefully not too late, for the EU and their individual member states to jointly take a clear stand and to start acting more energetically, coherently and boldly. What is *our* strategy? What is our *European way*?

2.3 A European way?

When devising an alternative strategy, there are in essence three questions. The first is, of course, what kind of society we want to live in? The second is, how to keep this society prosperous in the datafication age and beyond, while at the same time defending this prosperity when in peril? And the third is, what we are going to *do*? When trying to answer these questions, this chapter focuses on the role of government. But let's start somewhat more in general.



[11] See for example the warnings of Stephen Hawking (https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge) and the joint concerns of the Pope, IBM and Microsoft on AI on (https://www.bbc.com/news/technology-51673296).

The technological developments are astonishing and have already resulted in huge and useful progress in various fields to seize opportunities for much requested improvements and to realize solutions for difficult-to-solve problems that we really need to deal with in our society. At the same time it is clear, that in all strategies up till now, the benefits are very unevenly distributed. Datafication has so far been a "winner takes all" economy. Even the COVID-19 crisis has not resulted in any setback or at least a serious slowing down in revenue growth for Big Tech; on the contrary. It is the niche-players with less generic platforms that are hit[12], strengthening Big Tech even further. Many suggest that this is in fact the whole point of the game we're in and thus something of a law of nature so to speak. But is this really true?

A European way should not start from the recurring simplification that value creation is only about making money. In the DigiTranScope project we wanted to go beyond that. We have been exploring and discussing how "a better alignment between risks and rewards, across public and private actors, can turn smart, innovation-led growth into inclusive growth", and how to ensure that "public value […] is not created exclusively inside or outside a private-sector market, but rather by a whole society" (Mazzucato, 2018b, p. 263 and 265). As shown in other chapters, we have collected several interesting cases and done some fascinating experiments ourselves, using the same technology as in the dominant strategies but at the same time really challenging the idea that it is inevitable that just a happy few benefit. Just look at the promising steps we have taken with big data analyses on absolutely non-personalized open data about 'synthetic people', enabling municipalities to implement neighbourhood-specific policy interventions (see Section 9)

Characteristic of most cases and experiments we studied, is that active digital citizenship is stimulated and that both economic and social returns occur -without one pushing the other away- in settings where civil society is seriously taken on board[13]. In that sense there is definitely a relation to the plea for reenergizing the role of what has been called the Third Pillar (Rajan, 2019) -the community- that, together with the state and the market, supports prosperous and resilient societies. Involving the local level and the community seems critical. There are wonderful examples of creative things when choosing this perspective. It might turn out that communities can especially be helpful to have a sharp eye for situational differences when implementing solutions, instead of imposing the same standardized developments everywhere, regardless of the real issues and interests locally at stake. During one of the DigiTranScope workshops there was the interesting observation that this is one of the points Big Tech again and again misses and thus one of the reasons why their solutions are not equally popular in the cultures of the U.S., China and Europe[14]. The old problem was the challenge of scaling up small initiatives. IT can be useful here if it is used wisely by all cooperating parties. Big Tech has shown that horizontal network extensions can be done at zero cost (that's how they make their super profits). With the right leadership, horizontal social networks -communities of interest in the interest of communities- can maybe also be extended at zero cost.

In the DigiTranScope project we concluded that, all in all, a proactive *European way* to move forward in the datafication age on its own terms can best be built on the following foundations:

1. Government focusing on public value driven innovation, *for the benefit of all* and grounded in the European Convention of Human Rights as its moral compass (CoE, 1950);
2. Opting for open innovation 2.0[15], co-created between public partners, private partners, communities and science;
3. Creating a prosperous and resilient society of self-reliant citizens and competitive companies, based on solidarity and fuelled by a strong European public-private R&D agenda;
4. Economic and social development being equally important and going hand-in-hand, with room for customization that relate to situational differences, priorities and preferences;

---

12 For example Booking.com. See https://www.ft.com/content/64716675-b461-4fd3-ae0f-836973c68f12.

13 Not an entirely new idea as such, but the scale and the technology certainly are.

14 https://ec.europa.eu/jrc/communities/sites/jrccties/files/report_workshop_digitranscope_connectingpolicydevelopersanddecisionmakerswithdigitranscopeexperiments.pdf.

15 This means going beyond 'traditional' open innovation as originally promoted by Chesbrough, that was mostly confined to private innovation.

5. Embracing the benefits of the newest technology while at the same time keeping our eyes open for the downsides and acting firmly against them if necessary.

   The European way should certainly not be reactive or even defensive, based on fear of the evil associated with the datafication era when looking at where current strategies are headed. The European response should be forward-looking with the overall objective of creating and maintaining a proper balance between opportunities and threats, recognizing what is appropriately expressed as the Janus face of technology (Moschella, 2011). Technology innovations are here to stay and will inevitably go even further. Government should accept this as a fact and encourage an investigative, exploratory and experimental attitude ('permanent beta') towards what is happening, knowing that it is an adventure where mistakes are made and where there is no "first time right"[16].

   The COVID-19 crisis has been the ultimate wake-up call -and in that sense a gift- for everyone to fully understand what hyper-complexity and hyper-connectivity means in practice, and the perfect demonstration of how far Big Tech's influence extends, how heavily we depend on it, and how this dictates our actions. Despite our own efforts to register data, only Google data and algorithms can help us to monitor effectively the spread of the virus. And we can't get around Google and Apple with the strict and for them beneficial conditions they set, when experimenting with the use of apps to support crowd control[17]. COVID-19 has definitely proven that the *European way* should stem from the conviction that it is really urgent for Europe to abandon its innocence and naivety. This does not mean that we must make a complete break with our approach of the last years. This certainly goes for legislation. The GDPR has really been an excellent impetus to reset the way Big Tech should handle data, and understandably set an example for the state of California, among others. And let us also continue our multi-stakeholder dialogues on open standards and frameworks, or on transparency in algorithms and on ethical guidelines. But in view of all the fast changes, there should certainly be room for frequent additions and fresh accents in the policies and for stronger leadership in enforcing directives. Datafication is, as argued, on the one hand a new evolutionary stage in the ongoing digitization. But on the other, it is a creeping revolution when we take the trouble of seriously studying the underlying strategies. And these studies should be done incessantly. Policy-making, particularly in this area, is by nature not just a matter of setting goals and monitoring their achievement; it requires permanent alertness of the ongoing changes, in order not to miss the point [18]. There are no quick fixes; major difficult decisions have to be made fast with far-reaching consequences (Harari, 2020) in the knowledge that powerful organizations and institutions -Big Tech, but also the major players in international finance or in auditing for example- have huge interests in preserving the world as it is today, and that they are very determined and creative in finding ways to ensure that. This means they will resist structural changes with all their might: the system fights back!

2.4 Action perspectives

The third question we asked ourselves is what a *European way* means in terms of what it is that we actually should *do*. There is an abundance of reports on the current developments. But at the same time, there is a remarkable shortage of proposals for concrete action and, looking at the role governments should play, politicians and top officials in the civil service show a serious shyness and embarrassment to act.

In the DigiTranScope project, we have therefore conducted not just desk research on the challenges that governments face and on how these relate to new governance models more in general[19]. In addition, we also organized intensive multidisciplinary workshops to (a) collect and compare experiences with the digital transformation of governments themselves and (b) explore ways to create an effective interplay between the various national and transnational layers, as well as between governments, industry, civil society and science. Apart from identifying several tips and tricks, we searched for real game-changers to consider as pillars for the *European way* to fundamentally shift the course of the current system. We are working out three.

---

16 See Make it happen!, a report of the Dutch Information Society and Government Study Group, established by the Dutch government. https://www.digitaleoverheid.nl/wp-content/uploads/sites/8/2017/09/Make-it-Happen.pdf

17 https://www.forbes.com/sites/zakdoffman/2020/06/19/how-apple-and-google-created-this-contact-tracing-disaster/

18 I am somewhat paraphrasing Charles Leadbeater here, who once answered during a conference on how governments perform in digitization: "I'm sure they're hitting all their targets, but they're missing the point".

19 For an example of EU studies on this; see https://ec.europa.eu/jrc/en/publication/future-government-2030.

Data sovereignty

The probably most talked about game-changer at the moment is to put real control over personal data into practice. Everyone's right to control his or her personal data is one of the principles of the GDPR. But in general people have no clue how to exercise this right and even if they do have a clue, it is almost impossible to actually be in control. First, transparency is still at best fragmented when it comes to proactively letting people know who has personal data on them, who uses them and how, and with whom they are shared and for what reasons[20]. Secondly, there is also almost nowhere any way for an ordinary person to correct personal data or have them corrected, or to explicitly either refuse or give consent for sharing personal data.

Despite the GDPR, a huge amount of personal data is still continuously collected and used online by many private as well as public parties in the EU without people giving consent or even knowing that it is happening. Of course, many are glad not to have to give their data every other second for endless reasons and happy for personalized advertisements that suit them. But the way in which things are happening at the moment is serious food for thought. It is not only structurally against the law. It fuels uncertainty and mistrust of being followed and monitored every step of your life.

Estonia has been the first country to empower its citizens to really exercise their right to control personal data and there are indeed examples of Estonian government officials being convicted for viewing personal data without a plausible reason. Other countries, like for example the Netherlands, are discussing and experimenting to take the same step.

Of course, the real game-changer will be to give people this right not just for governmental data collecting and usage. It should certainly also include private companies. In most cases, this is for now not in scope. But it most certainly must be.

Inclusive growth

A second game-changer to consider is to create a breakthrough to drive true inclusive growth. This goes far beyond allowing everybody to fully participate in the information society. But must be the point of departure.

Despite the popular idea that it is just the elderly that can't keep up with the changes, the fact is that there is abundant proof that the digital divide is growing in all sections of the population and showing a significant relation to social inequality in the more general sense (van Dijk, 2020). When the objective of a European way is -as proposed- first and foremost digitization and datafication for the benefit of all, this trend must be broken. Once again, this is not a call to action for just governments to deliver accessible services as already agreed (EC, 2016) but also for companies. And it is not only about accessibility or using comprehensible words, but also about an active multichannel approach because for some e-mail is so old-fashioned by now that they don't use it anymore, whereas others have still to learn the basics. It's about informing people about their obligations as well as their rights as digital citizens, about being transparent and building up their trust on what is being done with their data and what they may expect at what moment, et cetera. And last but not least, it is not just a matter of pious promises, but also of really tackling organizations that don't comply.

People won't adjust when they miss the appropriate tools, designed according to their capabilities[21]. But it takes more than just fixing this problem to have an inclusive information society. When people are not involved in society in general, when they feel left out, have no control over their lives, and watch everything that happens full of fear and distrust, they will also not be part of the digital changes that are flooding them. As long as that is the case, socio-economic innovations that really last won't materialize. Realizing a European way as we propose will then fail.

An effective game-changer for inclusive growth is not easy to find and deserves further investigation. However, one thing is clear. Inclusion is not just an ideological issue, it is a matter of common sense. If you don't strive for an inclusive information society because it's the right thing to do, then do it because we

[20] We consciously stay away from the term 'ownership' of data.

[21] https://news.un.org/en/story/2019/06/1040131.

can't afford the dropout, which according to some estimates has now reached 20 to 25 percent of the population in the Netherlands, for example[22].

Stricter and more comprehensive compliance and law enforcement

The third game-changer we propose is stricter and more extensive compliance and law enforcement, including more systematic supervision on how the datafication market as a whole is developing. We think this is the most seriously missing link at the moment and an indispensable addition to current legislation and the other two game-changers if we want them to be effective and make a real difference.

All modern organizations, whether they like it or not, are also moral actors. Analytically, their behaviour can be regulated on three levels (by self-discipline, by self-regulation in sectors or professions, and by legislation) (Kimman, 2011). When it comes to datafication, we seem very liberal in addressing each other and very reluctant in actually enforcing codes of conduct, contracts or even laws. The EU and its member states mainly limit themselves to demanding fair tax payments from the Big Tech companies and imposing fines for breaking the rules, even if this is done systematically and not incidentally.

During the DigiTranScope project, we touched on compliance and law enforcement many times. We give here two situations to reflect on the challenges we see with the existing approach:

- The first concerns addressing cybersecurity weaknesses that are structural and have significant implications. As an example, think back on the worldwide ransomware attack in June 2017 that struck several major companies, simply because they hadn't updated their software in time. The attack spread like wildfire around the world. The Danish container giant Maersk was one of the first victims, but it soon turned out that many more companies were affected. American pharmaceutical company Merck, Russian oil giant Rosneft, British advertising company WPP, Spanish food company Mondelez: computer systems were going down everywhere. In Ukraine, the airport near Kiev was hit, as were several banks. This was not the first or last time that major chain effects and even global disruptions have occurred as companies are negligent in their cybersecurity and do not consider simple things like timely patches necessary. How should we respond to these kinds of situations? It seems that we limit ourselves to a firm conversation or perhaps a fine. Apparently, serious measures are not being considered, even if the consequences are enormous and critical processes are disrupted for days. What about revoking operating permits? Why is this not part of industry self-regulation schemes, nor government enforcement tools?

- The second is the enforcement of the GDPR. If, as argued, the competition in datafication for Big Tech is essentially about who holds the largest amount of personalized data, their operation fundamentally collides with the heart of the GDPR. For private data platforms, working within the GDPR means the end of their very lucrative revenue model and ultimately the end of their business. So what are we trying to achieve by imposing fines for violations? Again, revocation of operating licenses comes to mind as probably a more effective approach. Or, if their market power and their power in general are indeed so overwhelming, why not at least consider forcing the split of these companies, just as we did a few years ago when banks and insurance companies merged on a large scale? In this sense, we agree that there is no such thing as the "iron law of the market" and that inequality, monopolistic control or ownership of property are political choices (Piketty, 2019).

The third game-changer would be to seriously start acting in this line of thinking. This does not automatically mean more regulations and/or administrative burden. The Dutch Council of State (the supreme advisory body on legislation and the highest general administrative court) has convincingly argued that -in view of the hyper-complexity and the hyper-connectivity we live in and the speed of change- it is best to downsize laws to the core values at stake instead of losing ourselves in detailed rules that cloud what we essentially want to regulate[23]. This helps us much better to identify system failures and to intervene firmly if necessary.

2.5 Conclusion and closing remarks

There is a Dutch saying that in literal translation says "gentle healers make stinky wounds" which means that it is better to treat a problem thoroughly, even if the treatment is painful, otherwise it can get worse. Despite all the miracles that we clearly like and benefit from (Warley, 2014), the datafication has

[22] https://www.government.nl/documents/reports/2018/06/01/dutch-digitalisation-strategy.

[23] See (in Dutch) https://www.raadvanstate.nl/@112661/w04-18-0230/

progressed so far that this saying comes to mind. We are well on the way to completely losing the human touch in datafication, and in that sense things are really getting out of hand. If the question at the beginning of the chapter was "And what, if any, is the role of government in all of this?", the answer seems now clear enough. Big Tech and the market it has created will not change by itself because it is much too profitable, and the community cannot force the market to do so. So it is up to the government to do what the government exists for. That role is not just facilitating and supportive by repairing social inequalities as inevitable collateral damage of datafication as we now know it. That role is to proactively protect public interests based on the type of society we want. It is the natural role of the State to condition the game by determining the rules, explicitly based on human rights and fundamental freedoms, and by systemically enforcing the laws about playing the game.

The systemic changes we've seen over especially the last decade are not easy to adjust. As mentioned, major interests are at stake. The proposed game-changers are interventions that will only be effective if deployed at EU level. There, the guiding principles must be formulated and maintained. This does not mean that the implementation cannot start with a number of front runners or that everything has to be implemented in exactly the same way and at exactly the same time. Digitization and datafication are not as hot political topics as climate change, health care or social security, on which political parties may differ in major ways. For the benefit of all, the interplay between the layers of government must be determined on a situation-by-situation basis, and politically this seems quite possible.

Of course, after this exploratory DigiTranScope study, the work is not over yet. This should also include topics that we have not focused on so far. Looking at what we touched in this chapter, it would make sense to explore the role of government to make Europe less dependent on the US and East Asia in this ongoing competition. Another issue would be how to improve the internal organization of public administration to act in a more coherent and consistent manner. The silos as we know them all work from their own perspective. One moment, Big Tech companies are asked to invest and establish their data centers in the EU; the next moment their integrity is questioned. So what do we really want from them?

A recent observation is that the COVID-19 crisis, at least in the EU, has led to an interesting reappraisal in society of the role of both government and science. This also applies to digitization and datafication. The EC and the other layers of government in the EU would do well to use this momentum together to really implement the European way as recommended. At the same time, the mistrust we now feel about Big Tech, but also about other private sectors, does not necessarily lead to a renewed confidence of society in government. In the EU, for example, governments also use AI in incomprehensible, hidden and less-than-rigorous ways that rightly scare people. A successful European path depends on the clear profile, transparency and unambiguous intentions of the government. It needs a government that is not part of the problem, but undeniably just part of the solution.

References

Aldred J. License to be bad: how economics corrupted us. Allen Lane. 2019

Cocking D. and van den Hoven J. *Evil online*, Wiley Blackwell. 2018

Council of Europe. *European Convention on Human Rights*. 1950. https://www.echr.coe.int/Pages/home.aspx?p=basictexts

Cukier K. and Mayer-Schoenberger V. *The Rise of Big Data: how it's changing the way we think about the world,* Foreign Affairs, May/June 2013. https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data

European Commission. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. Publications Office of the European Union. 2016.

Harari Y. N. *The world after coronavirus*, 2020. https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75.

Kimman E. *Understanding compliance*, Rozenberg Publishers 2011.

Mazzucato M. The entrepreneurial state: debunking public vs private sector myths, Rev. Ed., with a new foreword, Penguin Books 2018a.

Mazzucato M.. The value of everything: making and taking in the global economy, Allen Lane. 2018b.

Moschella D. The next generation of digital game-changers, Leading Edge Forum. 2011.

Moschella D. and Mead K. Leveraging 'The Matrix'- digital ecosystem dynamics and strategies, Leading Edge Forum. 2015.

Piketty T. *Capital and ideology*, Harvard University Press. 2019.

Rajan R. The third pillar: how markets and the state leave the community behind, William Collins. 2020.

van Dijk J. *The digital divide*, Polity Press. 2020.

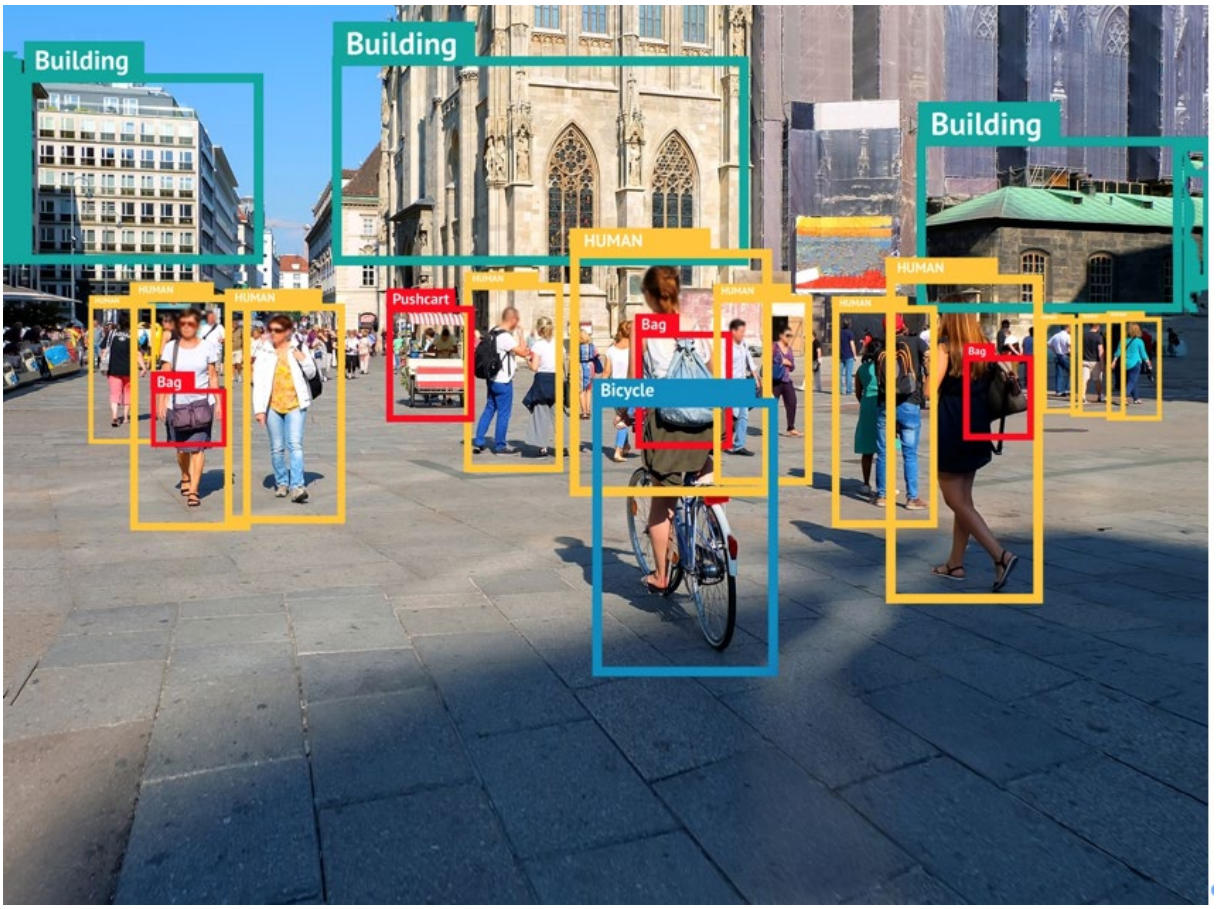
Wardley S. *Of wonders and disruption*, Leading Edge Forum. 2014.

Zuboff S. The age of surveillance capitalism: the fight for a human future at the new frontier of power, Profile Books. 2019.

PART A:
on the Governance of Digital Data

3 Data governance models emerging from the practices of social actors

Marina Micheli, Marisa Ponti, Massimo Craglia, Anna Berti Suman



3.1 Introduction

This chapter illustrates the results of our research on the emerging data governance models. The study has examined current discourses and practices on the governance of data. In particular, it scrutinised approaches for accessing, controlling, sharing and using data in today's platform economy and derives four emerging models of data governance. These models could be understood as inventive practices that problematize current arrangements and reassemble them in accordance to the interests of the actors involved.

In the last years, following scandals like Cambridge Analytica and new regulations for the protection of data like the GDPR, there is mounting attention concerning how data collected by big tech corporations and business entities might be accessed, controlled, and used by other societal actors. Scholars, practitioners and policy makers have been exploring the possibilities of agency for 'data subjects', as well as the 'alternative data regimes' that could allow public bodies to use such data for their public interest mission. Yet, the current circumstances, which are the result of a tradition of 'corporate self-regulation' in the digital domain and an overall laissez-faire approach (albeit increasingly divergent by geopolitical context), see the hegemonic position of a few technology corporations that have established de-facto quasi-data monopolies. In terms of data governance, this is reflected in an asymmetry of power between those corporations, which hold most of the decision-making power over data access and use, and other stakeholders. **With this research we wish to increase knowledge about the practices for data governance that are currently developed by various societal actors - beyond 'big tech' - emphasizing these actors' power to control how such personal data is accessed and used to produce different kinds of value.**

3.2 Data governance: A social science-informed definition

The term governance has been extensively used in the last two decades but its meaning is still ambiguous (Colebatch, 2014). Our understanding is informed by existing debates in the political science and risk scholarship (Kooiman, 2003; Colebatch, 2014) where, for example, governance has been framed as "the multitude of actors and processes that lead to collective binding decisions" (Van Asselt and Renn 2011: 431). Governance broadly refers to the web of actors involved, with different roles, in the process of governing a system. The term stresses a discontinuity from so-called "command-and-control" by the State, and acknowledges that a broader set of actors and institutions are (also) involved in managing societies (private sector, civil society and other non-government entities) (Kooiman, 2003). As Wolf (2002) puts it, the governance phenomenon takes place within *horizontally*-organized structures where both state and non-state actors (including citizens) interact. Yet this is easier in theory than in practice as power disparities among actors continue to exist and matter. Market actors often benefit from these more fluid allocations of power and responsibilities (DeNardis, 2019; Srnicek, 2017), at the detriment of less (economically) powerful actors such as citizens, communities and civil society organizations (Heeks and Shekhar, 2019).

Based on this understanding of governance, we examine in this chapter ways in which personal data collected through datafication processes is and could be governed. This contribution, however, adopts a social science-informed perspective of data governance that complements other framings, such as those of platform governance or privacy and data protection law. Our perspective on data governance draws in particular from science and technology studies and critical data studies, which informed our work through concepts of data infrastructure (Kitchin and Laurialt, 2014) and data politics (Ruppert et al., 2017). Informed by these scholarships and concepts, we understand data governance as ***the power relations between all the actors affected by, or having an effect on, the way data is accessed, controlled, shared and used, the various socio-technical arrangements set in place to generate value from data, and how such value is redistributed between actors.*** Such social science-informed definition allows moving beyond concerns of technical feasibility, efficiency discourses and 'solutionist' thinking. Instead, it points to the actual goals for which data is managed, emphasising who benefits from it, the power un(balances) among stakeholders (or lack of thereof), the kind of value produced, and the mechanisms (including the underlying principles and system of thoughts) that sustain these approaches. Based on this conceptualisation, the chapter examines four emerging data governance models.

3.3 Research Strategy

We delved into grey and academic literature, as well as news articles and websites of recent projects and initiatives to look for emerging practices of data governance. The collection of resources started in preparation of a workshop held in October 2018 on data governance with seventeen invited experts from academia, public sector, policymaking, research and consultancy firms (Micheli et al., 2018). We adopted a flexible search strategy and used a snowballing approach including progressively new sources according to their relevance to the theme of interest. The initial sources considered were identified for the preparation of the workshop and from the inputs provided by the workshop participants. A subsequent step was to review related work, which addressed similar issues or was directly linked to the sources examined. Simultaneously, we kept track of new publications and on-going projects or initiatives. The review strategy proceeded iteratively, until the typology of the models was consolidated. The review covered documents, publications, news and websites in English that addressed emerging practices for the governance of data with a focus on the European context. On the whole, it included 72 academic articles, 16 book chapters, 63 reports and policy documents, and 22 websites of projects/initiatives. The resources were collected in the time span from October 2018 to July 2019, with 9 documents added during the review process. Most of these are recent, as 74% have been published from 2017 onwards.

To guide our analysis and description of the emerging models of data governance we used the analytical dimensions, drawing in particular from Winter and Davidson (2018) and Abraham, Schneider and vom Brocke (2019) (Table 3.1).

Table 3.1: Analytical dimensions

Dimension	Definition
Stakeholders	The individuals, institutions, organisations or groups who are affected by, or have an effect on, the way data is governed and the value created.
Governance goals	The objectives held by actors who influence how data is governed.
Value from the data	The resources expected to be generated from the use of data and how these are distributed among actors and across society.
Governance mechanisms	The different instruments adopted to achieve specific governance goals, including the underlying principles.
Reciprocity	The power relation between stakeholders for data access and use.

3.4 Emerging data governance models

This section describes the data governance models identified following the five dimensions described above. These models should be understood as ideal types. They are abstract constructs. They are not intended as an exhaustive description of the state of the art, but as a contribution in synthetizing emerging data governance models. The four models described are labelled: data sharing pools (DSPs), data cooperatives (DCs), public data trusts (PDTs), personal data sovereignty (PDS).

3.4.1 Data sharing pools

Different actors join a DSP to 'analyse each other's data, and help fill knowledge gaps while minimizing duplicative efforts' (Shkabatur, 2019: 30). By creating these partnerships, they ease the economic need for exclusive rights and obtain limited co-ownership stakes in the resulting data pool. Data is treated and exchanged as a market commodity with the aim of producing data-driven innovation, new services, and economic benefits for all the parties involved (Carballa Smichowski, 2019). **DSPs are horizontal joint initiatives among data holders to aggregate data from different sources to create more value through their combination** (Mattioli, 2017; Shkabatur, 2019). Their overall ratio- nality is attuned with dominant discursive regimes of Big Data (Kitchin, 2014) and lies in the assumption that 'the greatest advantages of data sharing may be in the combination of data from multiple sources, com- pared or 'mashed up' in innovative ways' (Mayer- Schonberger and Cukier, 2013 cited in Mattioli, 2017: 184).

A key mechanism for DSPs is the contract, a legal and policy framework, that defines the modalities for data sharing, how data can be handled, and for which purposes. These contracts could be 'repeatable frameworks of terms and mechanisms to facilitate the sharing of data' between entities, which are especially useful for organisations that do not have the know- how and legal support to leverage data (Hall and Pesenti, 2017; Hardingens and Wells, 2018). Although these frameworks have been referred to as data trusts, there is not a full consensus whether they could be assimilated to actual legal trust structures or a 'marketing tool' facilitating the responsible sharing of data (Delacroix and Lawrence, 2019: 242). The assumption of this kind of contracts is that all parties benefit since the DSP enables them to obtain easily data that would otherwise be inaccessible. There is reciprocity between partner organisations, but only data holders are involved, as data subjects tend to be excluded from the relation and are at best depicted as passively benefiting from it. A practical limitation for data sharing pools consists in the transaction costs, such as data preparation, ensuring privacy and interoperability challenges, which put small businesses and under-funded entities at a disadvantage (GovLab, 2018). A further limitation is that often there is one

dominant partner (Carballa Smichowski, 2019). Therefore, although involving potentially many actors beyond big tech platforms, the relations are not necessary as horizontals (and sustainable) as claimed.

3.4.2 Data cooperatives

DCs distribute data access/rights among actors like DSPs, but differently from those, provide higher involvement of data subjects and are guided by different goals. **DCs enable a de-centralised data governance approach in which data subjects 'voluntarily pool their data together, to create a common pool for mutual benefits'** (Ho and Chuangt, 2019: 204). Participants of DCs share data while retaining control over it, having a say on how it is managed and put to value, and not submitting to the extractive logic of digital capitalism (Borkin, 2019; Ho and Chuangt, 2019). Therefore, data subjects are key stakeholders within DCs. By establishing a relationship of trust with the cooperative that manages data on their behalf, they preserve democratic control over their data and might demand an equitable share in the benefits produced (Borkin, 2019; Delacroix and Lawrence, 2019). This model is characterised by high reciprocity since 'all parties are stakeholders and are equally affected and bound by the governing rules they discuss, negotiate and then agree upon' (Ho and Chuangt, 2019: 203).

The underlying principles of DCs stem from the co-operative movement, established in UK and France in the 19th century, and from the more recent platform cooperativism (Scholz, 2016). The cooperative movement promotes fairer conditions of value production, in a non-monopolistic and transparent setting, alternative to the dominant capitalist model (Pazaitis et al., 2017). Analogously, **DCs address the power unbalances of the current data economy and are an explicit attempt to rebalance the relationship between data subjects, data platforms and third-party data users.** Enabling mechanisms for DCs are 'bottom-up data trusts' (Delacroix and Lawrence, 2019): agreements and contracts that provide the means for citizens to be informed, express their preferences and concretely decide how to share their data and for which purpose.

DCs need to generate sufficient income for their maintenance and development, but are not based on profit-maximising objectives. They often aim to create public value across society, including promoting social change and addressing societal issues, for instance by fostering equality, digital rights, environmental causes or medical research (Carballa Smichowski, 2019; Sandoval, 2020). Many DCs are 'commons-based' and open, blurring the distinction between the notion of data commons and DCs ('open cooperativism') as data is shared with an open license and made public (Carballa Smichowski, 2019; Ho and Chuangt, 2019; Pazaitis et al., 2017; Sandoval, 2020).

Examples of DCs operating with health data are MIDATA.coop and Salus Coop that let citizens donate their personal health information for scientific research. Although there is a growing interest in DCs for ethical approaches to data sharing and use (e.g. Ilves and Osimo, 2019), at the moment there are only few small examples, since this model struggles to compete and scale up against big tech that are advantaged by their monopolistic position, their critical mass of users, and greater financial resources (Sandoval, 2020).

3.4.3 Public Data Trusts

PDTs refer to a model of data governance in which a public actor accesses, aggregates and uses data about its citizens, including data held by commercial entities, with which it establishes a relationship of trust (Delacroix and Lawrence, 2019; Hall and Pesenti, 2017; Mulgan and Straub, 2019). Several stakeholders might be involved in this model, including city administrators, managers of public institutions, platform companies, trusted data intermediaries, research institutions, start-ups, and SMEs. Public administrations may also invite third-parties to access their data sources and develop data-driven services or to offer guidance on data sharing (Hall and Pesenti, 2017; Morozov and Bria, 2018). A key goal of PDTs is to integrate data from multiple sources to inform policy-making, promote innovation and address societal challenges, while adopting a responsible approach to the use of personal data (Bass et al., 2018; Morozov and Bria, 2018).

In PDTs, public actors assume the role of trustees that guarantee citizens' data is handled ethically, privately and securely. Thus they imply the establishment of a relationship of trust between citizens and public bodies: citizens must be reassured that public actors are capable to keep their personal information safe and secure and that they will use data for the public interest (Collinge, 2018). To earn trust from citizens, public bodies might engage in citizens' consultations and living labs, or require the intervention of external independent organisations that act as trusted intermediaries (Collinge, 2018; EC, 2020c; Mulgan

and Straub, 2019). These trusted intermediaries are new institutions that are allegedly held to account for securely managing data, preserving citizens' privacy, and maximising the public value of data (Mulgan and Straub, 2019). These entities will be independent and unrelated to for-profit firms and big tech corporations, and guarantee that data is managed without abuses through strong accountability and standards. Even if citizens are mostly seen as recipients who benefit from services and policies developed through PDTs, they might be explicitly involved through 'trust building' governance mechanisms such as living labs, public consultations and civic society initiatives.

An underlying assumption of PDTs is that all data with a public interest component is part of a nation infrastructure, therefore the information it affords should be 'socialised' to produce value for citizens and society as a whole (Cardullo, 2019; Morozov and Bria, 2018). Whilst at present PDTs are largely limited to small pilot projects, a key enabler would be a legal framework mandating private companies to grant access to data of public interest to public actors under conditions specified in the law (Shkabatur, 2019). This was considered by the EC (2020c), which then appointed a High-Level Expert Group on Business-to-Government data sharing. The issue has also been discussed at national level in Europe. For instance, French Member of Parliament Belot proposed creating the legal concept of 'territorial interest data' to give local governments the power to demand access to data (Carballa Smichowski, 2019).

Figure 3.1 - Vignette of the data governance models examined in the chapter.



3.4.4 Personal data sovereignty

The PDS model is characterised by data subjects having greater control on their data, both in terms of privacy management and data portability compared to the current dominant model. The label comes from the broader principle of technological sovereignty, which concerns subjects, public administrations, or governments regaining control of technology, digital content and infrastructures – thus reducing the influence of IT commercial enterprises and of foreign States in which these companies reside (Villani, 2018).

This model promotes a different and fairer data economy, echoing critical accounts of the dominant model of surveillance capitalism (Lehtiniemi, 2017). Data subjects are envisioned as key stakeholders together with digital service providers – which deliver the means for subjects to control, use and share their data – and re-users with whom data subjects decide to share their data (Ilves and Osimo, 2019). This governance model pursues two goals: it increases individuals' self-determination, granting more opportunities to access, share and use personal data, and engendering a more balanced relationship between users and digital platforms; and it is expected to foster a socially beneficial usage of data through the development of new data-driven services centred on user needs (Ilves and Osimo, 2019; Lehtiniemi, 2017).

Among the main mechanisms enabling PDS are personal data spaces, like Digi.me, Citizen-me or Meeco, which consist of 'intermediary services' allowing users to store their personal data, collecting data disseminated in different platforms, and control their sharing with third parties (Lehtiniemi, 2017). These services, which appeared in early 2000s, have been strengthened by Art. 20 of the GDPR (data portability). They are expected to remove obstacles for individuals wanting to exchange their data for research or other purposes, acting as trusted intermediaries and improving citizens' ability to make choices about their data (Delacroix and Lawrence, 2019).

PDS has been especially encouraged within the context of MyData, an international movement and a community of activists, non-profit organisations, think-tanks as well as commercial actors, start-ups and SMEs. **PDSs are expected to produce value in the form of data subjects' self-determination, knowledge, and public interest, but at the same time foster economic growth through an eco- system of new commercial services supporting them**. A limit of this model lies in its dependence on personal data spaces as these are currently adopted by only a niche of users and often fail to scale beyond pilots (Ilves and Osimo, 2019). As, as business entities, they may have interested in how to 'nudge' users and a few personal data spaces might gain more power in the market (Lehtiniemi, 2017). Furthermore, citizens have limited awareness about platforms' use of personal data for profit and the need for alternative models of value production, and the majority would not be capable, nor have the time to, take advantage of the opportunities offered by these intermediary services (e.g. Andrejevic, 2014). Envisioning citizens as 'market agents' (Lehtiniemi and Haapoja, 2020) free to choose from an ecosystem of personal data spaces might not fully address the asymmetries of power of the current data landscape.

3.5 Discussion

In this chapter we contribute to the policy debate on data governance using a socio-technical perspective to describe four emerging models of data governance: Data sharing pools (DSPs), Data cooperatives (DCs), Public data trusts (PDTs) and Personal data sovereignty (PDS). The models are abstract conceptualisations (Kvist, 2006) that do not necessarily represent discrete implementations of data governance. Nonetheless, they provide a foundation for discussion on alternative approaches or "desirable futures" for accessing and sharing data in the age of datafication (Jasanoff, 2015). **All models highlight a concern for redressing the structural power imbalances between corporate big data platforms and other actors, such as data subjects, public bodies, third parties, civil society and researchers.** There are nonetheless substantial differences regarding which stakeholders exert influence over data, and what value is pursued through data use.

With respect to the kind of value pursued, DSPs focus on producing economic value, while other forms of value gradually "chime in" in the remaining models, such as social change, public interest, fairness, and data subjects' self-determination. For the most part, these models could be found in niche initiatives or pilot projects, and there is still limited research concerning the value they generate and their sustainability over time (Borkin, 2019). Value production and redistribution, thus, could be assessed more at the level of the imaginary, than from evaluations of tangible outcomes. **In DSPs, data is a "market commodity" and economic value is redistributed horizontally among data holders who join the partnership.** PDSs put forward important innovations for data subjects' exerting digital rights, but do not question the datafication and commodification (Van Dijck et al., 2018) mechanisms of big data platforms. They are oriented towards the creation of value for the individual (self-determination) and new commercial actors (data services), with public interest as a by-product of these. The remaining models expressively pursue the public interest: **DCs allow data subjects to collect and aggregate their data for the public interest, while PDTs act on behalf of citizens, aggregating and analyzing different data sources to inform policy-making and address societal challenges**. If in DCs a cooperative has to be trusted, in PDTs is a public body. Yet, in the latter, it might also be that a trusted external independent organisation acts as a data

intermediary between citizens and a public body; this demonstrates how the abstract models can easily overlap in practice. PDTs represent a form of public-driven governance that could significantly redistribute the value of data and increase fairness, but requires the support of a new legal framework mandating access to data for public interest. Similarly, DCs are a fairer alternative of surveillance capitalism, but struggle to find financial sustainability and to reach a critical mass of users. Therefore **we did not find a single model to be "recommended" or "promoted" for a fairer data landscape. Instead, a combination of all these models should be envisioned for a "desirable future"** (Jasanoff, 2015). In particular, to oppose the privatization of internet governance (DeNardis, 2019), and the resulting dominant model of data governance stirred by big data platforms, it is advisable to look at the inventive data practices of civic society and public bodies as it is from these actors that we have found more interest in the redistribution of value generated through data.

An important dimension to discuss is the extent to which these models democratize data governance. To answer it we turn our attention to three models that involve data subjects. In all cases, data subjects can choose a trusted intermediary for their data, being it a commercial service from an ecosystem of personal data spaces (PDS), a cooperative that allow to keep democratic control over data and share responsibilities (DCs), or a public body that is entrusted by citizens to use (their) data ethically and for the public interest (PDTs). Involving subjects in the governance of data is a key strategy to address, and avoid, many of the possible negative consequences of data governance, such as dataveillance, function creep, technocratic governance, etc, (Kitchin, 2014). **The more powerful data subjects are in a data governance model, the greater accountability is required to the data holders, which in turn limits risks and data misuses.** At the opposite end, DSPs are only accessible to data holders and/or those in a position to pay for data. How does that model guarantee that needs and interests of data subjects (citizens at large and marginalized groups) are accounted for? To address this, and for good data governance, it may be advisable to combine DSPs with the others models that offer more guarantees, at least in principle, in terms of accountability.

The findings of this study highlight that the same "buzzwords" can be associated to different rationalities of data governance, since the notion of data intermediaries and data trusts is included somehow in all models. This underlines how important it is to think critically about data infrastructures as socio-technical products, moving beyond mere instrumental and technical aspects. Data trusts might be powerful means to reduce the power unbalances of the current data economy if adopted within DCs, while they may foster very different aims in DSPs. Indeed, in the first case these would be "bottom-up data trusts" that act in behalf of citizens' interests and preferences (Delacroix and Lawrence, 2019), while the latter would be repeatable frameworks of terms and mechanism to facilitate the sharing of data (Hall and Pesenti, 2017). Conversely, data trusts could also be a service offered by the public sector in a top-down manner to earn trust and foster the public interest, as in PDTs.

A final consideration concerns the intertwined relationship between the data practices we have examined and the regulatory frameworks in which they exist. **These data governance models can only develop further if they are sustained by appropriate legal frameworks**, such as the GDPR for personal data or a new legal act to mandate access to commercial data of public interest. With the recent developments in data policy (EC, 2020), the European Commission is strengthening its role as transnational regulator of technology with repercussions on a global scale. In doing so, it will be crucial to engage with the wider set of stakeholders identified in our research including local administrations and many actors from civil society who have an important role in shaping the emerging forms of data governance that address the asymmetries of the current data landscape.

References

Abraham R, Schneider J and Vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management* 49: 424–438. 2019.

Andrejevic M. The big data divide. *International Journal of Communication* 8: 1673–1689. 2014.

Bass T, Sutherland E and Symons T. *Reclaiming the Smart City. Personal data, trust and the new commons*. Report, NESTA, London, UK, July. 2018.

Borkin S. *Platform co-operatives. Solving the capital conundrum*. Report, NESTA and Co-operatives UK, London, UK, February. 2019.

Carballa Smichowski B. Alternative data governance models: Moving beyond One-Size-Fits-All solutions. Intereconomics. *Review of European Economic Policy* 54(4): 222–227. 2019.

Cardullo P. Smart commons or a "Smart Approach" to the commons? In: Cardullo P., Di Feliciantonio C. and Kitchin R. (Eds.) *The Right to the Smart City*. Bingley: Emerald Publishing Limited, pp. 85–98. 2019.

Collinge A. A 'new deal' for city data? In: London Data Store. 2018. Available at: https://data.london.gov.uk/ blog/a-new-deal-for-city-data/ (accessed 18 June 2020).

Delacroix S. and Lawrence N. D. Bottom-up data trusts: Disturbing the 'one size fits all' approach to data gover- nance. *International Data Privacy Law* 9(4): 236–252. 2019.

DeNardis L. The social media challenge to internet governance. In: Graham M. and Dutton. W. H. (Eds.) Society and the Internet: How Networks of Information and Communication are Changing Our Lives. Oxford: Oxford University Press, pp. 389–402. 2019.

European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Shaping Europe's Digital Future.* COM(2020) 67 Final. Publications Office, Luxembourg. 2020.

GovLab. Data Stewards: Data leadership to address 21st century challenges. In: *Medium*. 2018. Available at: https://medium.com/@TheGovLab/data-stewards-data-leadership-to-address-21st-century-challenges-a5e555cd4235 (accessed 18 June 2020).

Hall W. and Pesenti J. *Growing the artificial intelligence industry in the UK*. Independent report, UK. 2017.

Hardingens J. and Wells P. Defining a 'data trust'. In: *Open Data Institute Blog*. 2018. Available at: https://theodi.org/ article/defining-a-data-trust/ (accessed 18 June 2020).

Heeks R. and Shekhar S. Datafication, development and marginalised urban communities: An applied data jus- tice framework. Information, *Communication & Society* 22(7): 992–1011. 2019.

Ho C. and Chuangt T. Governance of communal data sharing. In: Daly A, Devitt K. and Mann M. (Eds.) *Good Data*. Amsterdam: Institute of Network Cultures, pp. 202–213. 2019.

Ilves L. K. and Osimo D. *A roadmap for a fair data economy*. Policy Brief, Sitra and the Lisbon Council. Helsinki, Finland, April. 2019.

Jasanoff S. Future imperfect: Science, technology, and the imaginations of modernity. In: Jasanoff S. and Kim S. H. (Eds.) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press, pp. 1–33. 2015.

Kitchin R. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. London: Sage. 2014.

Kitchin R. and Laurialt T. Towards critical data studies: Charting and unpacking data assemblages and their work. In: Thatcher J., Eckert J. and Shears A. (Eds.) *Thinking Big Data in Geography*. New Regimes, New Research. Lincoln/London: University of Nebraska Press, pp. 83–94. 2014.

Kooiman J. *Governing as Governance*. London: Sage. 2003.

Lehtiniemi T. Personal data spaces: An intervention in surveillance capitalism?. *Surveillance & Society* 15(5): 626–639. 2017.

Lehtiniemi T. and Haapoja J. Data agency at stake: MyData activism and alternative frames of equal participation. *New Media & Society* 22(1): 87–104. 2020.

Mattioli M. The data-pooling problem. *Berkeley Technology Law Journal* 32: 179–236. 2017.

Micheli M., Blakemore M., Ponti M., et al. *The governance of data in a digitally transformed European Society*. Report of the second Workshop of the DigiTranScope Project, European Commission, JRC114711.

Morozov E. and Bria F. *Rethinking the Smart City: Democratizing Urban Technology*. New York, NY: Rosa Luxemburg Stiftung. 2018.

Mulgan G. and Straub V. *The new ecosystem of trust: How data trusts, collaboratives and coops can help govern data for the maximum public benefit*. Nesta. 2019. Available at: https://www.nesta.org.uk/blog/new-ecosystemtrust/ (accessed 18 June 2020).

Open Data Institute *Data trusts: Lessons from three pilots*. Report, Open Data Institute, London, UK. 2019. Available at: https://theodi.org/article/odi-data-trusts- report/ (accessed 20 June 2020).

Pazaitis A., Kostakis V. and Bauwens M. (Digital economy and the rise of open cooperativism: The case of the enspiral network. *Transfer*: *European Review of Labour and Research* 23(2): 177–192. 2017.

Ruppert E., Isin E. and Bigo D. *Big Data & Society* 4(2): 1–7. 2017.

Sandoval M. Entrepreneurial activism? Platform cooperativism between subversion and co-optation. *Critical Sociology* 46(6): 801–817. 2020.

Scholz T. *Platform cooperativism. Challenging the corporate sharing economy*. Report for the Rosa Luxemburg Foundation, New York, USA. 2016.

Shkabatur J. The global commons of data. *Stanford Technology Law Review* 22: 1–46. 2019.

Van Asselt M. and Renn O. Risk governance. *Journal of Risk Research* 14(4): 431–449. 2011.

Wolf K. D. Contextualizing normative standards for legitimate governance beyond the state. In Grote J. R. and Gbikpi B. (Eds.) *Participatory Governance: Political and Societal Implications*. Opladen: Leske und Budrich, pp. 35–50. 2002.

Van Dijck J., Poell T. and De Waal M. *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press. 2018.

Villani C. For a meaningful artificial intelligence: Towards a French and European strategy, Mission assigned by the Prime Minister Edouard Philippe. 2018.

Winter J. S. and Davidson E. Big data governance of personal health information and challenges to contextual integrity. *The Information Society* 35(1): 36–51. 2019.

4 Seeing Platform and Data Co-operatives Through European Pandemic Citizenship

Igor Calzada

4.1 European Pandemic Citizenship at Stake

In Europe, many global citizens will likely be unemployed in the course of COVID-19 (McKinsey, 2020). The coronavirus does not discriminate, yet its economic and social impact are unevenly distributed even across and within state borders by producing a new pandemic citizenship regime that surfaces health, socio-economic, cognitive, and even digital vulnerabilities (Calzada, 2020a). By contrast, the COVID-19 pandemic has shown as well that digital platform economy and work can offer opportunities—alongside also exposing citizens to pervasive multifaceted vulnerabilities and precarious labour conditions—to continue working and earning even during times of crisis. But how can job quality be ensured for all platform workers while also creating further democratic socio-economic *platformised* alternatives to revert the algorithmic and *data-opolitic* extractivist business-as-usual hegemonic paradigm (Barns, 2020; Belloc, 2019; De Marco et al., 2019; Digital Future Society, 2019; Fairwork Foundation, 2020; Helberger et al., 2018; Kilhoffer et al., 2019; Lane, 2020; Riso, 2020; Taylor, 2020)?

At a glance, over the last decades, globalisation has led to a new class of global citizenship for workers (Calzada, 2020b). While the access to this global citizenship remains uneven, many have enjoyed unlimited freedom to move, work, and travel. However, COVID-19 has drastically slowed down this global citizenship regime and introduced a new level of ubiquitous vulnerabilities in global affairs by provoking the emergence of new pandemic citizenship regime, where citizens—regardless of their locations—share fear, uncertainty, and risks (Taylor, 2020). Furthermore, COVID-19 is pervasively related to data and artificial intelligence (AI) governance issues, which expose citizens' vulnerabilities in a potential surveillance state and market (Hintz, Dencik and Wahl-Jorgensen, 2017; Kitchin, 2020; Morozov, 2020; Zuboff, 2019).

Actually, the responses to this pandemic emergency have been extremely different from location to location, even within the same state. Among the resilience strategies adopted by various European governments, collective intelligence stemming from a proactive citizenship response has been highly considered to further avoid dystopian measures that could exacerbate existing social inequalities and techno-political vulnerabilities among European pandemic citizens (Bigo, Isin and Ruppert, 2019). As such, a particular collective intelligence response emerging in Europe is the creation of digital co-operatives (Borkin, 2019; Cherry, 2016; McCann and Yazici, 2018) also known as *platform co-operatives* (Scholz, 2016; Schneider, 2018; Scholz and Schneider, 2017; 2015) and *data co-operatives* (Blassimme et al., 2018; Hardjono and Pentland, 2019a; Hafen, 2019; Pentland et al., 2019).

There is a growing consensus in Europe, that it is urgent for governments to start filling in the same role in the information society that they have traditionally taken in the post-industrial society: not only fixing market failure, but also regulating the digital power relations and supervising its actual economic interplay among stakeholders. This means not just demanding fair tax payments by the big tech companies and imposing fines when they violate the GDPR or when they abuse their market power. There is much more - and more fundamental - at stake that calls for government attention beyond intervention. The COVID-19 crisis has clearly shown that citizens are highly dependent on data and the economic value that creates. COVID-19 crisis has led thus to a necessary revaluation in society of the roles of both government and citizens explicitly through extending to digitization and datafication to economic and socially innovative alternatives (Moulaert and MacCallum, 2019). We are referring to co-operatives (Beckett, 2019).

4.2 Co-operatives as a Collective Resilient Response to the Pandemic Crisis

Historically, co-operatives have been created when people work together—now with the help of technology—to respond with collective resilience to complex crises and to mobilise a wider range of information, ideas, labour, and insights to address social structural transformations through economic disruptive innovations (Calzada, 2013). The co-operative movement began in the UK and France back in the 19th century. Remarkably, though, several unique regionally rooted experiences with strong

communitarian identity have flourished in Europe since then, respectively in the 1950s and in late 1970s, such as the Mondragon case in the Basque Country (Spain) (Bengu, 2018; Clamp and Alhamis, 2010; Ellerman, 1984; Gupta 2014; Heales et al., 2017) and the Emilia Romagna case (Italy) (Battilani and Zamagni, 2012; Borzaga and Galera, 2012; Gonzales, 2010).

At present, against the pandemic fragile backdrop, European citizens working in tourism, the arts, retail and education, and all informal workers are the hardest hit (Kilhoffer et al., 2019; Gramano, 2019). Further, low-income, marginalized, working-class citizens and immigrants are more adversely affected than the average of the standard population (Eubanks, 2017). Income inequality is growing, confidence in governments is eroding and increasingly more people are embracing populism (Dyer-Witherford, 2020). Workers may lose power and a sense of agency over their lives and consequently their own data (Delacroix and Lawrence, 2019). To this end, how can working citizens organize, regain control of their data, and participate in the building of socio-economic alternatives to alter the existing data governance extractivist conditions to protect pandemic European citizens' digital rights (Calzada and Almirall, 2020)? Actually, how does European citizenship (reacting and therefore self-organising) to challenge data extractivism (Morozov, 2019) and surveillance capitalism? Is there any alternative response to big tech AI-driven *data-opolies*? What will be next?

In this direction, new possibilities for how Europe could be better, moving beyond considering the citizen a simple resource, have already been claimed in the widely spread manifesto '#DemocratizingWork: Democratize, Decommodify, Remediate' (#DemocratizingWork, 2020) signed by relevant academics worldwide.

4.3 Objective of the study

The study aimed to decipher the rationale behind the platform and data co-operatives by providing evidence-based research and policy analysis, and by responding to how new forms of co-operatives using digital technologies can provide a framework to rethink, renew, and offer alternatives for the way policies on digital transformations and AI can help enhance pandemic citizens' well-being and thus to improve the post-COVID-19 working conditions of vulnerable and/or already empowered citizens.

At a moment when the concentration of big tech companies is skyrocketing, platform and data co-operatives do not stop at an analysis of surveillance capitalism; they might equip citizens to succeed and build an alternative as cooperative platform entrepreneurs or activists in the fast-growing gig economy (Alosi, 2016; Hayes, 2019; Lutz, 2019). They might give members of the co-operative the opportunity to analyse and get involved with a generation of citizens who are experimenting with innovative power-building strategies rooted in cooperative ownership of digital platforms and data storages. Could we imagine an Uber owned by the drivers or Twitter being owned by its users? Ultimately, advocates of platform and data co-operatives suggest that a shift from a *sharing economy* to a genuinely participatory, democratically owned economy might be possible.

4.4 The Research Question and Two Main Hypotheses

The ongoing post-General Data Protection Regulation (GDPR)/COVID-19 Europe may require reshuffling the way data is already affecting the nature of citizens' work and the overall way digital economy might function more democratically and more locally rooted in cities and regions (Calzada, 2019). Thus, the analysis in this chapter will focus on a citizen-driven resilient post-COVID-19 response. Consequently, **the research question in this chapter revolves around whether it is possible to alter the existing data governance extractivist conditions for incentivising the emergence of platform and data co-operatives to further democratise and thus protect pandemic European citizens' labour and digital rights**?

The main hypotheses of this chapter are twofold:

(i) It seems that a shift from the centralised and extractive platform-knows-best capitalist model towards the experiment-driven **'platform co-operatives'** is emerging, albeit this phenomenon is rather embryonic and still too weak to be consolidated.

(ii) There is an associated policy challenge in how to consolidate a pan-European post-GDPR AI through **'data co-operatives'** now that the Gaia-X project has been launched (GaiaX, 2020).

Hence, both 'platform co-operatives'—which address alternative communitarian business/social models—and 'data co-operatives'—which aim to customize protected data stores for their members while generating a non-profit social value from doing so—share the same underlying issues: citizen-centric collective ownership, decentralised self-governance, transparency, and being an alternative to platform capitalism (Bastani, 2019; Beardman, 2012; Beckett, 2019; Blok et al., 2017; Como et al., 2016; Srnicek, 2017).

4.5 Co-operatives and Pandemic Citizenship

Overall, co-operatives could be considered a movement of citizens in Europe which accounts for 130,000 enterprises in Europe, in basically all economic sectors, with 127 million members, more than 4 million employees and nearly €990 billion annual turnover (Cooperatives Europe, 2020). However, the understanding, practices, and perceptions of co-operatives significantly vary from state to state—and particularly now in the post-COVID-19 era—by showing a substantial path-dependency regarding contextual factors deriving from the history and current digital and socio-political transformations. At a glance and generalising historical trends, two distinct citizenship regimes have framed and shaped perception around co-operatives. First, in the Western European co-operative tradition, sometimes there is a concern that many of small and local platform and data co-operatives seem to be approaching the co-operative model mainly for ideological and value reasons, clearly underestimating the economic dimension of a required sustainable business model in a global competitive context. This approach creates exposure to a high risk of failure beyond the purely altruistic, volunteer, civilian, and grassroots-driven initiatives far removed from the formal professional entrepreneurial institutions. The second regime appears in the Eastern European co-operative tradition. In essence, the communist legacy left a generalised distrust of the population in the co-operative concept, a distrust which is still linked to the memory of past communist collectives, and has been clearly replaced by an individualistic pattern where the idea of private ownership of assets is generally preferred over sharing or direct exchange with other peer citizens.



The notion of co-operatives (Castilla-Polo and Sánchez-Hernández, 2020) in the digital era—eminently as a transnational phenomenon—cannot be dissociated from the emerging pandemic citizenship regime in Europe that are shared among different states' borders (Calzada, 2020a). Further, such regime is actually the seed for creating co-operative forms in the digital economy and society that aim to protect citizens' digital rights, such as platform and data co-operatives.

Hence, at this stage, any approach to citizenship in Europe needs to be analysed through the lenses of the post-COVID-19 aftermath. Citizenship encompasses concepts of not only identification and belonging, but also power, control, and techno-politics. Long before COVID-19 swept the globe, insecurity and social vulnerabilities were already ubiquitous. Countless people have faced housing, health, and food insecurity. Prior to the advent of *social distancing,* people hid behind doors, locks, gates, and border walls, afraid of

both public space and one another. Meanwhile, online, people have long fretted over information security, devising passwords to access passwords, fearful they might be hacked or exposed. People are insecure in their jobs, homes, and relationships, and on social media. Not to mention, people feel insecure about themselves as well. Co-operatives may thus mean building a more secure economy and society for everyone. It is a challenge Europe cannot afford to ignore (Taylor, 2020).

However, co-operatives' far-reaching aspirations are related to the idea that digital revolutionaries should reshape everything but the central institution of modern life: the market. As such, while big data and AI do not naturally favour non-market activities, they do make it easier to imagine a post-neoliberal world where production is automated and technology underpins universal healthcare and education for all in the post-COVID-19 era—a world where abundance is shared by peers, not appropriated (Bastani, 2019; Dyer-Witheford, 2019; Riso 2020). Today's debate on the right technological response to COVID-19 revolves around, on the one hand, the trade-offs between privacy and public health on the one hand (Kitchin, 2020) and, on the other, the need to promote innovation by start-ups. Why are there no other options? It is because we have let digital platforms and telecom operators treat our entire digital universe as their fiefdom (Khan, 2017)? They run it with just one goal in mind: keep the micro-targeting going and the micro-payments flowing. As a result, little thought has gone into building digital technologies that produce macro-level anonymous insights about the collective behaviour of non-consumers. The digital platforms, as they are known hegemonically today, are the sites of individualised consumption, not of mutual assistance and solidarity (Sandoval, 2019; Siapera and Papadopoulou, 2016). Thus, could digital platform innovation in Europe be led by an asymmetric network of co-operative SMEs (De Marco et al., 2019; Helberger, Pierson and Poell, 2018)? This emerging European pandemic citizenship regime is currently shaping the potential for platform and data co-operatives formation (Calzada, 2020a).

Since March 2020, coronavirus has mocked immigration controls, biometrics, digital surveillance, and every kind of data analytics, and struck hardest—thus far—in the richest, most powerful states of the world. Consequently, the significance of European citizenship might be rapidly shifting through a sort of pandemic citizenship adjustment, with different consequences for citizens depending on which state they call home and their living conditions. What might be called a shared pandemic citizenship—where citizens in Europe share exactly the same fears—seems to be here to stay. This trend has different levels of techno-political implications as it intersects with another global trend—how algorithms are increasingly shaping everyday life.

Arguably, the current pandemic crisis and democracy are pervasively related to data governance issues, exposing citizens' vulnerability in a potential surveillance state (Calzada, 2020b; Lucas, 2020). Should European governments protect citizens from being infected even if doing so might mean establishing a new digital non-privacy norm? Will this pandemic crisis become an algorithmic crisis, with serious side-effects for governments in Europe? Could these rapidly changing times for European citizenship be seen as an opportunity to foster digital co-operatives in Europe in pursuit of a *Tech New Deal* to allow citizens and communities to own and govern their own data and platforms (Bauwens and Pazaitis, 2018; Calzada, 2013; Hardjono and Pentland, 2019a, 2019b; Pentland et al., 2019; Schneider, 2020; Scholtz, 2016; Scholz and Schneider, 2015, 2017)?

4.6 Towards a Taxonomy for Platform and Data Co-operatives

According to Bauwens (Bauwens and Vasilis, 2014; Bauwens, Kostakis and Pazaitis, 2019; Bauwens and Pazaitis, 2018) and Scholz (2016), **data co-operatives** could be seen as a subcategory inside **platform co-operatives**. But generally speaking, **data co-operatives** arguably focus merely on data stores, while **platform co-operatives** revolve around the whole business model of workers, services, and products, which also includes data.

Table 4.1. Definitions: Platform Co-operatives and Data Co-operatives

Platform Co-operatives	Data Co-operatives
A platform cooperative, or platform co-op, is a cooperatively owned, democratically governed business that establishes a computing platform, and uses a website, mobile app or a protocol to facilitate the sale of goods and services. Platform cooperatives are an alternative to venture capital-funded platforms insofar as they are owned and governed by those who depend on them most—workers, users, and other relevant stakeholders.	Cooperative structures could enable the creation of open data and personal data stores for mutual benefit; they could rebalance what many perceive as an asymmetric relationship between data subjects (people with personal data) and data users (people who use data to develop services and products). Members of a community voluntarily pool their data to create a commons pool for mutual benefits. This common pool of data acts as a commons resource of collective ownership upon a framework which is collectively discussed and agreed upon.

There is a diverse set of taxonomies (Scholz, 2016):

- Generally speaking, **platform co-operatives** focus on business models, while **data co-operatives** mutualise and store data.
- Regarding the flow, **platform co-operatives** manage labour exchange and distribute content while aggregating the data of a bunch of members/citizens.

Finally, platform co-operatives consist of four typologies:

- Consortia Worker Platforms:
    - Co-operatively owned online labour brokerages and market places: This co-operative is the most common platform where workers/citizens own the company, receive dividends and have a voice in running the company.
    - Union-backed labour platforms: Unionised workers/citizens can create their own companies as a result of the collaboration between unions and workers.
- **Produser-led Platforms:** Users and producers own the platform through which producers can sell their work.
- Multistakeholder/Community Platforms:
    - City-owned platforms: This model could be created as a collaboration between a large number of cities that would pool their resources to create a software platform for any kind of service: short-term rentals, utilities and so on (e.g., CCDR).
    - Co-operatives from within: This case exists when workers/citizens from a sharing economy platform like Uber use the technical infrastructure of the company to run their own enterprise. Worker co-operatives form inside the belly of the sharing economy (Mensakas).
- Regarding Data Co-operatives, they could be considered a sub-typology of Platform Co-operatives, also known as, **Data Consortia Platform**.

To further simplify the analysis, four typologies for platform co-operatives are defined as follows:

- **Worker:** This typology refers to the flow of labour exchange and revolves around co-operativising work (stemming from mobility services).
- **Produser** (as a merge of users and producers): This typology refers to the flow of content distribution and revolves around co-operativising the outcome resulting in the exchange between users and producers (stemming from culture, agriculture, food, software, websites, hosting, start-up support, videoconferencing, etc.).
- **Multistakeholder:** This typology refers to the flow of content distribution and revolves around co-operativising the community services (stemming from healthcare, delivery riders, media, rental, housing, land, etc.).

- **Data** (this fits into data co-operatives): This typology refers to the flow of data aggregation and revolves around co-operativising and mutualising data (particularly data related to finance, health, security, etc.).

Table 4.2. Taxonomy for Platform Co-operatives and Data Co-operatives

Oriented to			Flow	Typologies	
Business	**Platform Co-operatives**		Labour exchange	1. Consortia Worker Platform 1.1. Co-operatively Owned Online Labor Brokerages and Market Places 1.2. Union-Backed Labor Platforms	**1. Worker**
			Content distribution	Produser-led Platform	**2. Produser**
				3. Multistakeholder/Community Platform 3.1. City-Owned platforms 3.2. Co-operatives from Within	**3. Multistakeholder**
Store		**Data Co-operatives**	Data aggregation	4. Data Consortia Platform	**4. Data**

Table 4.3 illustrates a classification of each platform co-operative according to its typology (https://ioo.coop/directory). Several cases could be included in multiple typologies, but the identification process aimed to include each case in only one typology.

4.7 Final Remarks

This ongoing exploratory research intended to decipher the rationale behind platform and data co-operatives amidst a new citizenship regime unfolded and presented by this chapter as *European pandemic citizenship*. This chapter concludes with three aspects that frame a potential future research and policy agenda for platform and data co-operatives.

First, experts' analyses, cases' identifications and preliminary conducted fieldwork action research have demonstrated that the post-COVID-19 world is reigniting the need to reactivate European civil societies by further experimenting with digital socio-economic innovations, such as platform and data co-operatives—but marginally and at a small scale.

Second, consequently, procurement and public incentives are required to push ahead, enhance, and reinforce platform and data co-operatives beyond extremely marginal experiments aligned with data donation and altruism.

Third, ultimately, initiatives around platform and data co-operatives need to find their own strategic pathway amidst the digital and social economy policy agenda of the European Commission. The ongoing institutional arrangements around H2020-Smart Cities and Communities have already gathered 17 lighthouse projects (encompassing 46 lighthouses and 71 fellow cities), and the still-experimental Digital Innovation Hubs follow the new European Data Strategy altogether. In all likelihood, these initiatives can firmly aid and foster city-regional connections between social entrepreneurs and urban activists in pursuit of scaling up platform and data co-operatives to better articulate the uneven European pandemic citizenship.

Table 4.3. Case Identification by Typology

Platform Coops			Data Coops
Worker (29; 18%)	**Produser** (38; 24%)	**Multistakeholder** (49; 31%)	**Data** (39; 25%)
Co-operativising Work	***Co-operativising Exchange***	***Co-operativising Community Services***	***Co-operativising Data***
Mobility	Culture, agriculture, food, software, websites, hosting, start-up support, videoconferencing	Healthcare, delivery riders, media, rental, housing, land	Health, finance, security
1. www.greentaxico-op.com 2. www.Fairmondo.de 3. www.Loconomics.com 4. www.upandgo.coop 5. www.start.coop/ 6. www.coopcycle.org/en/ 7. www.eva.coop/ 8. www.enspiral.com/ 9. www.reservation.alphataxis.fr/ 10. www.co-optaxi.com/ 11. www.aarhusmakers.com/ 12. www.applicolis.com 13. www.casa-comuna.coop 14. www.codesolid.com 15. http://www.nesta.org.uk/sharelab-fund-meet-grantees/driver-co-op 16. www.eyemole.io 17. www.crowdfunder.co.uk/faircab 18. www.gildedsplinters.coop 19. www.indycube.community 20. www.lowimpact.org 21. www.means.tv 22. www.themobilityfactory.eu 23. www.modo.coop 24. www.noncorporate.org 25. www.ridygo.fr 26. www.staffing.coop 27. www.taxiapp.uk.com 28. www.wechange.eco 29. www.wordjammers.com	1. www.Resonate.is 2. www.stocksy.com 3. www.foradoeixo.org.br/ 4. www.smart-ib.coop/ 5. www.agrilyst.com 6. www.ampled.com 7. www.ampliativeart.org 8. www.cleanenergycu.org 9. www.en.goteo.org 10. www.ccor.org 11. www.cosmos.coop 12. www.darkpeak.org 13. www.docservizi.it 14. www.dorg.tech 15. www.driversseat.co 16. www.drutopia.org 17. www.eten.com 18. www.guerrillatranslation.org 19. www.hcoop.net 20. www.interchanges.io 21. www.joinus2eat.be 22. www.kostaki.id 23. www.marketers.coop 24. www.mayfirst.coop 25. www.membersmedia.net 26. www.org.meet.coop 27. www.openfoodnetwork.org/ofn-local/open-food-network-scandinavia/ 28. www.about.openfoodnetwork.org.uk 29. www.originclub.org 30. www.partago.be 31. www.thephone.coop 32. www.smart-eu.org 33. www.social.coop 34. www.sofcoop.org 35. www.vngrd.online 36. www.webarchitects.coop 37. www.webhosting.coop 38. www.webtv.coop	1. www.saltspacecoop.co.uk/ 2. www.formandfunction.coop/ 3. www.graphics.coop/ 4. www.taskrabbit.com/ 5. www.fairbnb.coop/ 6. www.equalcare.coop/ 7. www.mensakas.com/ 8. www.savvy.coop/ 9. www.thenews.coop/global/ 10. www.banyanproject.coop/ 11. www.libretaxi.org/ 12. www.ethoscharity.co.uk/ 13. www.coopdescommuns.org/en/the-coop/ 14. www.snowdrift.coop/ 15. www.ampliativeart.org/ 16. www.fair.coop/ 17. www.incubator.coop 18. www.affinity.works 19. www.anyshare.coop 20. www.bhive.coop 21. www.brave.coop 22. www.coopsource.org 23. www.collective.tools 24. www.cooby.io 25. www.networks.coop 26. www.pittsburgh.covivi.us 27. www.demcra.com 28. www.doma.city 29. www.encode.org 30. www.francebarter.coop 31. www.freedomcoop.eu 32. www.gebiedonline.nl 33. www.highplainsfood.org 34. www.kabelan.id 35. www.nearnow.org.uk/stories/re-imagining-childcare-introducing-kidoop 36. www.knowledgeatlas.com 37. www.mediacoop.ca 38. www.newscoopyyc.coop 39. www.lesoiseauxdepassage.coop 40. www.parti.coop 41. www.positivenews.org.uk 42. www.pueblo.global 43. www.share.coop 44. www.signco.io 45. www.sommobilitat.coop 46. www.tapazz.com 47. www.vientos.coop 48. www.weco.io 49. www.wehelpen.nl	1. www.culedger.com 2. www.MiData.coop 3. www.Salus.coop 4. www.cozy.io 5. www.mydex.org 6. www.aqdatacommons.org 7. www.Openhumans.org 8. www.decodeproject.eu 9. www.decidim.org 10. www.meta.decidim.org 11. www.mydex.org 12. www.opendatamanchester.org.uk 13. www.thegooddata.org 14. www.healthbank.coop 15. www.ubiquitouscommons.org 16. www.datacommons.coop 17. www.nordicdei.org/ 18. www.givememydata.com 19. www.waze.com 20. www.citydataexchange.com 21. www.airbus.com/aircraft/support-services/skywise.html 22. www.agproexchange.com 23. www.bankofthecommons.coop 24. www.cupay.coop 25. www.datavest.org/ 26. www.divvydao.org 27. www.market.fair/coop 28. www.find.coop 29. www.gisc.coop 30. www.ledgerback.coop 31. www.mainst.market 32. www.moeda.in 33. www.opencredit.network 34. www.patientcritical.com 35. www.privacyco-op.com 36. www.rchain.coop 37. www.robinhoodcoop.org 38. www.somconnexio.coop 39. www.zerodark.coop

Source: elaborated from https://ioo.coop/directory

References

Alosi, A. (2016). Commoditized Workers: Case Study Research on Labor Law Issues Arising From a Set of on-Demand/Gig Economy Platforms, *Comparative Labor Law & Policy Journal 37*(3), 653-690.

Barns, S. (2020). Platform Urbanism: Negotiating Platform Ecosystems in Connected Cities. Springer Singapore: Singapore.

Bastani, A. (2019). *Fully Automated Luxury Communism*. Verso: London.

Battilani, P. and Zamagni, V. (2012). The Managerial Transformation of Italian Co-Operative Enterprises 1946–2010, *Business History 54*(6), 964-985.

Bauwens, M. and Vasilis K. (2014). From the Communism of Capital to Capital for the Commons: Towards an Open Cooperativism, *Triple C: Communication, Capitalism and Critique 12*(1), 356-361.

Bauwens, M., Kostakis, V. and Pazaitis, A. (2019). *Peer to Peer: The Commons Manifesto*, University of Westminster Press: London.

Bauwens, M., and Pazaitis, A. (2018). *P2P Accounting for Planetary Survival*, P2P Foundation, Guerrilla Foundation and Schoepflin Foundation.

Beardman, S. (2012). Altruism and the Experimental Data on Helping Behavior, *Ethical Theory and Moral Practice 15*(4), 547-561.

Beckett, A. (2019). *The New Left Economics: How a Network of Thinkers Is Transforming Capitalism.* The Guardian. Retrieved from: https://www.theguardian.com/news/2019/jun/25/the-new-left-economics-how-a-network-of-thinkers-is-transforming-capitalism

Belloc, F. (2019). Why Isn't Uber Worker-Managed? A Model of Digital Platform Cooperatives. CESifo Working Papers, 7708.

Bengu, R. (2018). *What Can South Africa Learn from Mondragon? The Answer Lies in Community Ownership.* Retrieved from: https://www.dailymaverick.co.za/article/2018-10-26-what-can-south-africa-learn-from-mondragon-the-answer-lies-in-community-ownership/

Bigo, D., Isin, E. and Ruppert, E. (2019). *Data Politics*, Routledge: London.

Blasimme, A., Vayena, E. and Hafen, E. (2018). Democratizing Health Research through Data Cooperatives, *Philosophy & Technology 31*, 473-479.

Blok, A., Courmont, A., Hoyng, R., Marquet, C., Minor, K., Nold, C. and Young, M. (2017). Data Platforms and Cities. *Tecnoscienza: Italian Journal of Science & Technology Studies 8*(2), 175-219.

Borkin, S. (2019). Platform Co-Operatives - Solving the Capital Conumdrum, NESTA: London.

Borzaga, C. and Galera, G. (2012). The Concept and Practice of Social Entreprise. Lesson from the Italian Experience, *International Review of Social Research 2*(2), 85-102.

Calzada, I. (2013). Knowledge Building and Organizational Behavior: The Mondragon Case from a Social Innovation Perspective, In Moulaert, F., MacCallum, D., Mehmood, A. and Hamdouch, A. *International Handbook on Social Innovation: Collective Action, Social Learning and Transdisciplinary Research*. Edward Elgar: Cheltenhman, UK and Northampton, MA, USA. pp. 219-229. DOI: 10.13140/RG.2.2.30748.69766.

Calzada, I. (2015). Benchmarking Future City-Regions beyond Nation-States, *Regional Studies, Regional Science 2*(1), 351-362. DOI: 10.1080/21681376.2015.1046908.

Calzada, I. (2019). Technological Sovereignty: Protecting Citizens' Digital Rights in the AI-Driven and Post-GDPR Algorithmic and City-Regional European Realm, *Regions eZine*, 4, DOI: 10.1080/13673882.2018.00001038.

Calzada, I. (2020a). Emerging Citizenship Regimes and Rescaling (European) Nation-States: Algorithmic, Liquid, Metropolitan and Stateless Citizenship Ideal Types. In Moisio, S., Koch, N., Jonas, AEG, Lizotte, C., and Luukkonen, J., *Handbook on the Changing Geographies of the State: New Spaces of Geopolitics*. Edward Elgar: Cheltenhman, UK and Northampton, MA, USA. DOI: 10.13140/RG.2.2.17301.6832/1.

Calzada, I. (2020b). Will Covid-19 Be the End of the Global Citizen?, *Apolitical*. Retrieved from https://apolitical.co/en/solution_article/will-covid-19-be-the-end-of-the-global-citizen

Calzada, I. and Almirall, E. (2020). Data Ecosystems for Protecting European Citizens' Digital Rights, *Transforming Government: People, Process and Policy 14*(2), 133-147. DOI: 10.1108/TG-03-2020-0047.

Cherry, M. (2016). *Legal and Governance Structures Built to Share: The Rise of Platform Co-Operatives.* Legat Studies Research Paper Series, Saint Louis University School of Law.

Clamp, C. A. and Innocentus A. (2010). Social Entrepreneurship in the Mondragon Co-Operative Corporation and the Challenges of Successful Replication, *The Journal of Entrepreneurship 19*(2), 149-177. DOI: 10.1177/097135571001900204.

Como, E., Mathis, A., Tognetti, M. and Rapisardi, A. (2016). Cooperative Platforms in a European Landscape: An Exploratory Study, *ISIRC Conference*.

Cooperatives Europe (2020). Retrieved from https://coopseurope.coop/resources/news/covid-19-coop-response

De Marco, C. E., Di Minin, A., Marullo, C. and Nepelski, D. (2019). *Digital Platform Innovation in European SMEs: An Analysis of Sme Instrument Business Proposals and Case Studies*, Publications Office of the European Union: Luxembourg.

Delacroix, S. and Lawrence, N.D. (2019). Bottom-up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance, *International Data Privacy Law*, DOI: 10.1093/idpl/ipz014.

DemocratizingWork: Democratize, Decommodify, Remediate (2020). Retrieved from www.democratizingwork.org

Digital Future Society (2019). The Future of Work in the Digital Era: The Rise of Labour Platforms. DFS: Barcelona.

Dyer-Witheford, N. (2020). Left Populism and Platform Capitalism, *Triple C: Journal for a Global Sustainable Information Society 18*(1), 116-131.

Ellerman, D.P. (1984). Entrepreneurship in the Mondragon Cooperatives, *Review of Social Economy 42*(3), 272-294. DOI: 10.1080/00346768400000025.

Eubanks, V. (2017). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press: NYC.

Fairwork Foundation (2020). *How Is the Platform Economy Responding to COVID-19?* Open Democracy, Retrieved from: https://www.opendemocracy.net/en/oureconomy/how-platform-economy-responding-covid-19/

GaiaX (2020). Retrieved from www.data-infrastructure.eu

Gonzales, V. (2010). Italian Social Cooperatives and the Development of Civic Capacity: a Case of Cooperative Renewal? *Affinities: A Journal of Radical Tehory, Culture, and Action 4* (1), 225-251.

Gupta, C. (2014). The Co-Operative Model as a 'Living Experiment in Democracy', *Journal of Co-operative Organization and Management 2* (2), 98-107. DOI: 10.1016/j.jcom.2014.09.002.

Hafen, E. (2019). Personal Data Cooperatives – a New Data Governance Framework for Data Donations and Precision Health. In Krutzinna, J. and Floridi, L. (Eds), *The Ethics of Medical Data Donation*. Springer International Publishing, Cham, pp. 141-149.

Hardjono, T. and Pentland, A. (2019a). Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management. *arXiv e-prints*.

Hardjono, T. and Pentland, A. (2019b). Empowering Artists, Songwriters & Musicians in a Data Cooperative through Blockchains and Smart Contracts, *arXiv:1911.10433v1*.

Hayes, R. (2019). Worker-Owned Apps Are Trying to Fix the Gig Economy's Exploitation. Motherboard, Tech by Vice.

Heales, C., Hodgson, M. and Rich, H. (2017). *Humanity at Work: Mondragon, a Social Innovation Ecosystem Case Study*, The Young Foundation: London.

Helberger, N., Pierson, J. and Poell, T. (2018). Governing Online Platforms: From Contested to Cooperative Responsibility, *The Information Society 34* (1), 1-14.

Hintz, A., Dencik, L. and Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance society, *International journal of Communications 11*, 731-739.

Kilhoffer, Z., Pieter de Groen, W., Lenaerts, K., Smits, I., Hauben, H., Waeyaert, W., Giacumacatos, E., Lhernould, J-P. and Robin-Olivier, S. (2019). *Study to Gather Evidence on the Working Conditions of Platform Workers*, European Commission: Brussels.

Kitchin, R. (2020). Civil Liberties or Public Health, or Civil Liberties and Public Health? Using Surveillance Technologies to Tackle the Spread of Covid-19, *Space and Polity*, 1-20. DOI: 10.1080/13562576.2020.1770587.

Lane, M. (2020). Regulating platform work in the digital age, *Going Digital Toolkit Policy Note*, No. 1. Retrieved from https://goingdigital.oecd.org/toolkitnotes/regulating-platform-work-in-the-digital-age.pdf

Lucas, E. (2020). *Pandemic Democracy*, Retrieved from: https://www.cepa.org/pandemic-democracy

Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data, *Human Behavior and Emerging Technologies 1* (2), 141-148. DOI: 10.1002/hbe2.140.

McCann, D. and Yazici, E. (2018). Disrupting Together: The Challenges (and Opportunities) for Platform Co-Operatives, NEF: London.

McKinsey (2020). *The future of work*. Retrieved from https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-in-europe?cid=soc-web#

Morozov, E. (2020). The Tech 'Solutions' for Coronavirus Take the Surveillance State to the Next Level'. Retrieved from https://www.theguardian.com/commentisfree/2020/apr/15/tech-coronavirus-surveilance-state-digital-disrupt

Morozov, E. (2019). Digital Socialism? *New Left Review,* 116/117, March & June.

Mos, E. (2020). Digital platforms and the (re)organization of solidarity. Retrieved from https://platformlabor.net/blog/digital-platforms-and-the-reorganization-of-solidarity

Moulaert, F. and MacCallum, D. (2019). *Advanced introduction to social innovation*, Edward Elgar: Cheltenham.

Pentland, A., Hardjono, T., Penn, J., Colclough, C., Ducharmee, B. and Mandel, L. (2019). *Data Cooperatives: Digital Empowerment of Citizens and Workers*, MIT Connnection Science: Boston.

Riso, S. (2020). Digital Age: Mapping the Countours of the Platform Economy, Eurofound: Dublin.

Sandoval, M. (2019). Entrepreneurial Activism? Platform Cooperativism between Subversion and Co-Optation, *Critical Sociology*. DOI: 10.1177/0896920519870577.

Schneider, N. (2020). Tech New Deal: Policies for Community-Owned.

Schneider, N. (2018). An Internet of Ownership: Democratic Design for the Online Economy. *The Sociological Review 66* (2): 320-340. DOI: 10.1177/0038026118758533.

Scholz, T. (2016). Platform Cooperativism: Challenging the Corporate Sharing Economy, Rosa Luxemburg Stiftung: New York.

Scholz, T. and Schneider, N. (2017). *Ours to Hack and to Own*, OR Books: NY.

Siapera, E. and Papadopoulou, L. (2016). 'Entrepreneurialism or Cooperativism?', *Journalism Practice 10* (2), 178-195. DOI: 10.1080/17512786.2015.1125760.

Taylor, A. (2020). The People's Platform: Taking Back Power and Culture in the Digital Age, Metropolitan Books: NYC.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, Profile: London.

5 Citizen-generated data for policy: A review of EU projects

Marisa Ponti and Massimo Craglia

5.1 Introduction

The rapid embedding of digital technologies into everyday practices and spaces has made it possible for citizens to generate and share data about local community-relevant problems, ranging from air quality measurement to reporting street problems. **The growth of data generated by citizens can give the public sector new opportunities for addressing critical social and economic issues and inform policies**. For example, mobile-equipped citizens can complement digital sensors for real-time reporting and situational awareness, providing public authorities with opportunities for data-driven decision making, improved performance management, and heightened accountability (Linders, 2012). Thus, digital technologies constitute the socio-technical means citizens can use to participate on issues that affect their lives, for example by producing data and statistics and creating a new political subjectivity (Ruppert, 2018). **This chapter provides an overview of European projects involving citizen-generated data (CGD).** We aim to understand how CGD makes it possible to experiment with new forms of public participation, rethink relationships between citizens and local governments and explore new emerging roles for citizens and local governments.



5.1.1 Citizen-Generated Data Projects

CGD has been defined as "data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them. It is actively given by citizens, providing direct representations of their perspectives, and an alternative to datasets collected by governments or international institutions" (DataShift, 2015, p. 1). This definition denotes two main characteristics. One is the voluntary participation of the public in collecting data on community-relevant problems. CGD can be considered a form of user-generated data collected explicitly for tackling those problems, such as improving local infrastructures, tracking environmental issues, or collecting spatial data. The other one is the creation of alternative datasets that can complement official data, offering the opportunity for citizens to make their voices heard within democratic processes at the local level of government (DataShift, 2015). CGD can help gain new perspectives, involving communities in surfacing and responding to issues that affect them.

CGD efforts are typically organised as projects. Motivations for setting up a CGD project can be very different. For example, citizens can be stirred by a lack of accurate data from the public sector or a lack of trust between public sector authorities and citizens. In other cases, citizens collect data to raise awareness of a topic that does not receive enough attention from institutions or to complement data granularity to institutional sources. By providing a means for citizens who want to make their voices heard, CGD projects can make valuable contributions to understanding and addressing social and economic problems. As reported by Lämmerhirt, Jameson and Prasetyo (2016), **CGD projects often result from and depend on partnerships between citizens and several organisations, including civil society organisations,**

community-based organisations, public sector, and businesses. These partners play a decisive role to provide resources, support, and knowledge to citizens. In return, they can tap into the data generated by citizens. Thus, citizens and their partners can gain mutual benefits from the application of CGD approaches.

5.1.2 Governance of Data and Data for Governance

Involving citizens in producing data can help to experiment with new forms of participation in data production and its governance that, in turn, may lead to new types of relationships between citizens and public institutions (Ruppert, Isin & Bigo, 2017). This involvement may lay the ground for expanding the public sphere and may create an opening for thinking about what Castells (2008) called the "Network State" that is "characterised by…greater diversity in the relationship between governments and citizens" (p. 88). This argument resonates with the more recently argued-for need for governance systems involving multi-stakeholder collaboration (British Academy and Royal Society, 2017, p. 55). This collaboration implies that not only the public sector, but also businesses, academia, and citizens can provide data publicly (Meijer & Potjer, 2018) to address common concerns and meet public good.

Digital technologies generate opportunities for producing, managing and using data that citizens may want to take, resulting in different forms of data governance. Technologies are inextricable components of data governance, as also explicitly indicated in the definition of this concept provided by the British Academy and Royal Society (2017). In their report, they refer to data governance as the processes of governing data management, data use, and the technologies involved in these processes "to inform the extent of confidence" in these processes (p. 1).

Dealing with data is not just about producing data. Much discourse on data governance focuses on the capacity of data for knowing and representing the world. While data certainly is a representational resource, it can also shape the way we see and think about the world (Gray & Marres, 2018). Two assumptions underpin this performative view of data. Fist, data is not neutral. The creation, extraction, and analysis of data, involved in the governance of data, are deployed taking into account specific objectives, needs and capacities. Therefore, data "construct" the world following different visions and interests (Grey, Gerlitz, & Bounegru, 2018). Second**, data is a social and political practice engaging participants who are not only "objects of data", about which data is produced, but also "subjects of data", as they drive the how and why data is produced** (Ruppert, Isin & Bigo, 2017). CGD can expand what gets measured, how, and for what purpose. Data generation can create opportunities for citizens to play a more active role in examining a situation and taking action, such as, for example, in local development and collaborative strategies for monitoring, auditing, planning and decision-making (Lämmerhirt, Gray, Venturini, & Meunier, 2019). Lämmerhirt and colleagues (2019) noted that CGD projects at a local level could help facilitate engagement of citizens. For example, concerning the achievement of the Sustainable Development Goals, localisation has been acknowledged to connect the broad global dimensions of SDGs with their contextual relevance. This connection would help engage residents because it could provide relevance to the intended actions, add value, and create local ownership.

In the following sections, we first present a brief overview of the main features of 18 European projects CGD in five areas: environment**,** public health**,** energy**,** transport, and infrastructure. Then, we summarise findings from five interviews related to data governance, setup and development of projects, project impact, project sustainability, and CGD use by the public sector.

Box 1. Selection of Citizen-Generated Data projects in a nutshell
To select relevant cases of CGD projects, we used the following inclusion criteria (Ponti and Craglia, 2020):
The project had to be about data actively generated by citizens around issues concerning them.
The project had to involve citizens generating data in partnerships with the public sector and community-based organisations at the local level.
The collection of data had to serve the public good primarily (e.g., collect data on air and water quality), and inform policy and create public services.
The project had to be at a local scale (neighbourhood, municipality, and city-scale) in the European Member States.
The technologies used by citizens to collect data had to be digital devices (e.g., cellular telephones, calculators, sensors).
Besides collecting information about such projects through desk research, we also conducted interviews with five people responsible for the development of five of the above projects.

5.2 The key-features of European CGD projects

5.2.1 Purpose of data collection

The sampled projects are distributed across countries as follows: Germany (two cases), Netherlands (four cases), Belgium (one case), Switzerland (one case), Italy (one case), Spain (four cases), and UK (five cases). Half of the projects were ongoing, while eight where a pilot and one completed. They tended to be of limited scale in time and place, although several of them had the potential for replication and expansion. A slight majority of the 18 sampled projects aimed at collecting data for environmental monitoring and environmental decision making, as shown in Table 5.1.

Table 5.1. Purpose of data collection in the selected projects

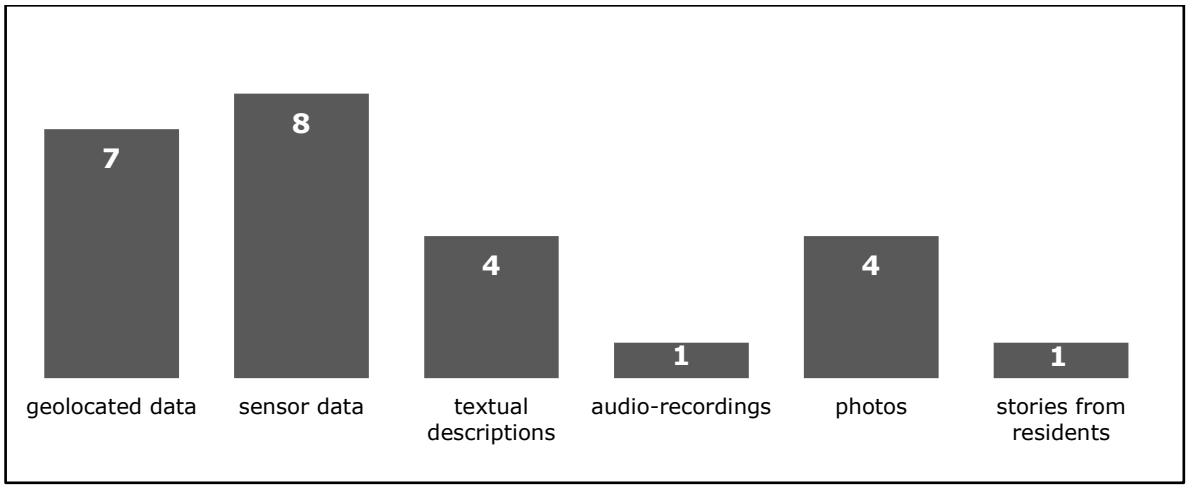
Examine the effects of climate change
Identify where building new homes
Improve local infrastructure
Improve transportation
Make a neighbourhood a healthier place
Map accessibility of public places for disabled
Map cycling conditions
Map favourite local places to influence development plans
Map quiet areas
Measure air quality
Measure noise pollution
Measure odour pollution
Report street problems
Share info about policy proposals
Use renewable energies

5.2.2 Categories of projects and types of data

Figure 5.1 shows the relative frequency of the primary categories of the selected projects. Projects aimed at measuring air quality or noise pollution were categorised as ‘passive sensing’ because they depend on participants using a resource that they are provided with, or they own (e.g., a home sensor), for automatic sensing. Thirteen projects were categorised as ‘crowdsourcing’ because we stressed their reliance on a

large number of contributors. Citizens collected different types of data, as shown in Figure 1, using a variety of methods and devices, including sensors, online platforms, mobile phones, and maps. Data generation involved automatic sensing, plotting urban places using mobile apps, taking pictures with mobile phones, writing observations using mobile apps, and locating street problems on maps, among the others. In relation to forms of engagement, citizens were enrolled as sensors, monitors, reporters, observers, platform users and co-creators of sensors. Reporting information sometimes occurred in a very structured way to ensure quality and consistency.

Figure 5.1. Types of collected data



5.2.3 Use of CGD by the public sector

Regarding the use of CGD by the public sector, we could find this information for ten of the sampled projects. Public sector organisations were interested in CGD for public service planning and facilities improvement, or reporting and planning environmental actions. Table 5.2 shows how CGD in the sampled projects was used by public sector organisations.

Table 5.2. Reported uses of CGD by public organisations

Project	Use of CGD	Public Organisation
Hush City	Partial update of the Berlin Plan of Quiet Areas 2018-2023 for the Berlin Action Plan of Noise Reduction	Berlin City Council - Berlin Senate Department for the Environment, Transport and Climate Protection
Curious Noses	Reporting, planning and model improvement	Flemish Environmental Agency
Samen Meten	Official air quality monitoring	Dutch Institute for Public Health and the Environment (RIVM)
Decoro Urbano	Planning maintenance and repair of street problems and resource allocation	Italian local councils
FixMyStreet	Report management	UK local councils
Catch!	Identify existing problems and develop solutions	Coventry City Council, Ipswich Borough Council, Oxfordshire County Council, Leeds City Council, Newcastle City Council

Project	Use of CGD	Public Organisation
D-Noses	Potential integration in official noise pollution measurements	Saõ João da Madeira Municipality and the Municipality of Sofia, and the Intermunicipal Waste Management of Greater Porto
Botellon no me deja dormir	Implementation of solutions co-designed with residents	Barcelona City Council
Cycle Hackney	Prioritise investments into cycling or designing road infrastructures	Hackney Council
Southwark New Homes	Prioritise investments into housing	Southwark Council

5.3 Findings from interviews

Data governance: CGD was accessible and findable in all the five projects. Data was open and reusable in three out of the five projects. Different reasons for not opening the data included the need to allow citizens to govern their data, or the need to restrict the use of data to avoid purposes not connected to the public interest. In all the projects, CGD was not interoperable, although it could be but not without overcoming some challenges. The lack of definition of data standards and metadata in Citizen Science was considered a problem. Problematic was also the tension between the interest of the data platform used in a project to keep the data and make it compatible with specific standards, and the citizens' right to keep the data.

Setup and development of projects: Projects were set up for different reasons but shared two aspects: **CGD was collected to address public issues** – for example, air and noise pollution, or urban infrastructural problems - **and expected to be used for a public purpose, such as raising awareness, helping local authorities make better decisions, or influence local governments to respond to community-relevant problems**. The choice of a participatory approach also differed, ranging from leveraging a vast community of citizens using sensors, to using a qualitative approach involving citizens' perceptions of urban problems, to the need for human "noses" to capture a specific phenomenon like odours. To involve stakeholders, including the public sector and community-based organisations, different modalities were used. These modalities are reflected in the roles played by public organisations. These organisations had different roles, ranging from being clients paying a fee for the service, to co-creating and implementing solutions to reduce noise pollution, to providing institutional support and offering mentorship to the project. One project also applied a quadruple helix model (the public, the industry, scientific community and the policy influencers) at local, national and global levels.

Project impact: **The projects were reported to have effects on citizens, collaborating stakeholders, and political agendas**. Citizens benefited from participating in the projects in different ways. For example, they felt empowered from being able to participate and generate data that can be used directly by a city council. In another case, the careful reports submitted by citizens on public street issues suggested that they believed in the usefulness of contributing their data to their representatives in city councils to improve local problems. Regarding political agendas, CGD can have an impact as it brings a critical problem – e.g., air quality – to the attention of politicians more strongly than official reports produced by institutions. Thus, CGD can point politicians to matters that their constituencies are interested in. Benefits for the collaborating stakeholders were also reported. For example, the city council of a big city partially adopted the framework used in a project to update their plan on quiet areas. In another case, city councils without financial resources to hire dedicated staff that can collect and process citizen reports can use valuable and accurate reports generated by citizens and made available through an external platform. Councils can also use the same platform to explore detailed statistics, for example, to see how many

reports they have received and how many they have processed. These figures can also be used to support political agendas because city councils can show how many reported problems have been fixed over time.

Project sustainability: Informants described both financial and non-financial support mechanisms to ensure the project's viability, continuity, and scalability. **Financial support mechanisms included micro-grants or tax exemptions to pay for server maintenance and software development, and European contributions for the creation of data platforms to store, manage and make available collected data even after the end of projects.** Other financial mechanisms used in the projects included the creation of a start-up to replicate project results and use the same methodology to tackle environmental challenges, and the charge of a service fee to be collected from city councils which signed a contract with the data platform owners. The fee was necessary to cover the increasing costs of software upgrading and reporting moderation service. Non-financial support mechanisms included ways in which the EU could play a role. For example, an informant wished the EU would help CGD projects gain more credibility and fairly acknowledge and reward citizen contribution to data generation.

CGD use by the public sectorPublic sector organisations use CGD for solving local problems, or for reporting and planning environmental actions. Building trust and credibility among public authorities regarding the accuracy of CGD remains a primary challenge. Public sector organisations may not trust how citizens generate data – for example, how they use sensors to measure air quality – and question the quality of the collected data. Quality remains a critical aspect of CGD. An important distinction needs to be made between high-quality data for scientific purposes and the fit-for-purpose use of data. Quality of data for scientific purposes means that data collected are analysed, elaborated, and validated according to rigorous scientific standards set for the given purpose of a study. Data quality for non-scientific purposes needs to refer to trustworthiness in terms of sources and used devices, as in the examples presented here. Solutions to provide data of sufficient quality included validation mechanisms along with the development of a new methodology for data collection, strict data collection procedures to ensure the consistency of data, and sensor calibration methods.

5.4 Discussion

5.4.1 New forms of public participation and emerging roles for citizens and local governments

Digital technologies, such as interconnected sensors and mobile applications, constitute an important driver of participation, by allowing citizens to generate data on community-relevant problems. Thus, CGD projects become techniques allowing citizens to exert their rights by producing data to evidence local problems and generate better living environments (Gabrys, 2019). By collecting data that local governments can use to derive relevant insights and informing action, citizens can be more actively involved in improving and maintaining the quality of their living environment. Generating data can offer the opportunity for citizens to raise and amplify their voices within democratic processes at the local level of government (DataShift, 2015).

Environmental sensors and information and communication technologies seem to act as "focal devices" (Haklay, 2015). This means they *hold the potential* to change the way citizens look at their living environments and facilitate data creation as a focal practice, a purposeful and meaningful social activity (Haklay, 2015). Most of the sampled projects suggest the existence of focal practices, as they engage citizens with their local environments in various ways, from recording sounds and odours to taking photos of street problems and plotting places where new houses should be built. This process of recording and mapping can become a focal practice (Haklay, 2015).

CGD projects as focal practices hold the potential to bring back agency and control to citizens who become "subjects of data" and not only "objects of data", about which data is produced (Ruppert, Isin, & Bigo, 2017). Following Haklay (2015), we see the act of mapping and recording in itself as "an act of asserting presence, rights to be heard or expression of personal beliefs in the way that the world should evolve and operate" (p. 4). **However, this new agency of citizens cannot just be assumed because of the use of technologies**, as grassroots CGD projects can just be populated with citizens primarily contributing data. **Opportunities must be actively created through closer collaborations between local authorities and citizens** where citizens can also contribute to driving "the how and why data is produced" (Ruppert, Isin & Bigo, 2017). However, this closer collaboration should not be expected or mandatory. It should not be intended as an attempt to downplay the role of government agencies, but as a way to redesign governance

to integrate citizen's efforts in broader institutional settings (Lam, 1996), with citizens bringing something essential to the table that would be lacking otherwise.

5.4.2 Rethinking relationships between citizens and local governments

Promoting the agency of citizens through data collection can open up the possibility to "achieve" citizenship, rather than receiving it (Hintz, Dencik & Wahl-Jorgensen, 2019), through a reconfiguration of the relationships between citizen and local authorities. For example, citizen involvement in reporting street problems makes a difference for local city councils in terms of cost, staff time savings, and efficient reporting. Citizen reports help initiate interventions on street problems and hold local governments accountable for their promises. While data and technologies in the sampled projects place responsibilities on bureaucracies and politicians who have a role to play in fixing urban problems, they also displace responsibilities onto citizens – in particular, to take more responsibility for their urban environment, rather than merely "blaming others." In these projects, collecting data becomes a way of taking up responsibility as individual citizens and can go a long way in solving urban problems and achieve active citizenship.

The act of mapping and recording becomes both an act of asserting citizens' presence or their rights to be heard, and an act of sharing responsibility for the governance of the public good – term used here as shorthand signal for the shared benefit at a societal level (Morrell, 2009). In this respect, data collection practices within participatory approaches hold the potential to open up more systematic interactions between local authorities and citizens, as well as between local authorities, citizens and businesses. In this "two-way traffic" model of governance (Kooiman, 2003), the boundaries between local governments and citizens become more porous. Local governments become more open to gather data and information external to their organisations from multiple actors, including citizens. No single actor, either public or private, has all the information and knowledge necessary to address and solve complex problems in a fast-changing and diverse society. The interaction between local governments and other stakeholders, including citizens, are often based on the recognition of interdependencies (Kooiman, 2003).

CGD projects may enrich already produced data, published by governments as open data or otherwise (Lämmerhirt et al., 2019). One of the most potent actions from public authorities to boost the integration of CGD with official datasets would simply be to open CGD data (and in a second step, to do that using shared standards or APIs) so that third-party applications can be created. Having applications based on a mix between CGD and authoritative data (instead of only CGD) would increase their reliability and impact.

5.4.3 Generating richer and trusted data to address current challenges

Most of the sampled projects involve citizens using low-cost sensors and accessible digital technologies to conduct indicative monitoring and generate data over a wider spatial area or more extended periods. However, this data may not be at the same level of precision or accuracy as data produced for regulatory compliance (Gabrys, Pritchard & Barratt, 2016). The methods used by citizens are different from those used by official organisations, and citizens usually have no legal mandates to standardise reporting within or across countries (Lämmerhirt et al., 2019). CGD may not comply "with established conventions to obtain the quality, interoperability and verifiability of data and sometimes abide by 'good enough' standards for operational use, different from those of established official professional statistics" (Lämmerhirt et al., 2019, p. 8).

Data quality is a critical issue in a policy context where alignment with monitoring requirements and regulatory standards is paramount (Brenton, von Gavel, Vogel & Lecoq, 2018). Some policymakers have advocated the notion of fitness for purpose approach – at times used interchangeably with fitness for use – where key aspects like data quality, scale, cost, interoperability and data format must be taken into account when evaluating the value of CGD for a particular policy question (Holdren, 2015). While the analysis of the interview data show that projects developed and applied validation processes to ensure data reliability and trustworthiness, it also indicates that others aspects such as interoperability and data format are not yet fully implemented. Two measures to help address this problem could be the development of citizen science data and metadata standards and the use of technologies aligned with regulatory requirements (Bonn et al., 2018). For example, while some environmental monitoring sensors may align with regulatory standards, others may not (Volten et al., 2018). In this respect, according to some key informants, the EU could have a role to play by enforcing that technologies used by citizens to collect data are assessed by certified bodies and deemed to meet EU specifications. Last, CGD projects

would benefit from adhering to open practices and fully documenting such practices (e.g., objectives, data collection protocol and analysis techniques), to ensure the trust of participating citizens and other stakeholders. This aspect also brings to another critical point, which is the openness of software (and hardware) used to generate data. Using open-source software in CGD would allow full control (e.g., reproducibility) of the procedures and workflows, which is crucial to ensure the reliability of results/measures.

5.5 Conclusion

The examined projects have been driven by community-relevant problems that affect citizens' quality of life, and the need to provide evidence for local authorities to take action. These projects indicate that CGD can have an impact by enabling different relationships with the public sector. They can provide the opportunity to find novel ways of interaction and open up channels of communication between policymakers and citizens. This chapter shows that types of CGD data, types of uses of data and types of technologies used in data collection are interrelated. The interweave of data, data collection practices and forms of use indicate how citizens can generate data in various social contexts reflecting their lived experiences.

The chapter points to three main aspects of CGD reconfiguring the relationship between citizens and the public sector.

1. First, the potential of CGD to act as a focal practice. Digital technologies can change the way citizens look at their living environments and facilitate data creation as a focal practice, a purposeful and meaningful social activity. In turn, CGD projects as focal practices hold the potential to bring back agency and control to citizens, moving them closer to the role of agents of change in the places where they live. This potential does not develop automatically, though: opportunities must be actively created for forms of closer collaboration between local authorities and citizens where citizens can also contribute to driving "the how and why data is produced." This process can be challenging because it implies ways of shifting agency, accountability and responsibility towards citizens.
2. Second, the potential of CGD to enable citizens to “achieve” citizenship, rather than receiving it. Citizen reports help initiate interventions on problems affecting the places where citizens live and hold local governments accountable for their promises. Collecting data becomes a way of taking up responsibility as individual citizens and can go a long way in solving urban problems and achieve active citizenship. However, “achieving” citizenship is challenging because it requires a “culture shift” such that citizens and communities become active participants.
3. Third, the issue of data quality in policy contexts. Most CGD is collected using low-cost sensors and accessible digital technologies. This data may not be at the same level of precision or accuracy as the data produced for regulatory compliance. However, it could raise different concerns and possibilities useful to describe "data stories" together with citizens and integrate the representation of reality provided by official data.

The sample size of 18 projects examined for this chapter is only a portion of currently active CGD projects, and new projects surface regularly. Our goal was selecting representative entities rather than achieving exhaustiveness or statistical analysis while making a smaller sample size appropriate for this review. The number of CGD projects in Europe is growing. This trend is expected to continue, along with more influence on decision-making at the local government level. **CGD has great potential to be used as a resource for the public good. However, this will happen if the public sector, citizens, and other stakeholders will work together to match policy needs, data sources, technological solutions, and standards**. National/regional legal and policy frameworks could be useful to guide the public sector to use unofficial data together with official data and private repositories. This could lead to a digital ecosystem which can provide evidence that can be understood and used by decision-makers, businesses, and citizens alike (cf. United Nations for the Environment, 2018).

References

Bonn, A., Hecker, S., Bowser, A., Makuch, Z., Vogel, J., & Haklay, M. (2018). Citizen science to foster innovation in open science, society and policy. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A.

Bonn, (Eds.), *Citizen science: Innovation in open science, society and policy* (pp. 1-23). London: UCL Press. DOI: 10.14324/111.9781787352339

Brenton, P., von Gavel, S., Vogel, E., & Lecoq, M-E. (2018). Technology infrastructure for citizen science. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn, (Eds.), *Citizen science: Innovation in open science, society and policy* (pp. 63-80). London: UCL Press. DOI: 10.14324/111.9781787352339

British Academy and Royal Society (2017*). Data management and use: Governance in the 21st century: A joint report by the British Academy and the Royal Society*. Available at https://royalsociety.org/-/media/policy/projects/data-governance/data-management-governance.pdf

Castells, M. (2008). The new public sphere: Global civil society, communication networks, and global governance. *The ANNALS of the American Academy of Political and Social Science*, *616*(78), 78-92.

DataShift (2015). *What is citizen-generated data and what is the DataShift doing to promote it?* Available at: http://civicus.org/images/ER%20cgd_brief.pdf (accessed 14 February 2019.).

Gabrys, J. (2019). Data citizens: how to reinvent rights. In Bigo D, Isin E, Ruppert E (Eds) *Data Politics: Worlds, Subjects, Rights* (pp. 248-266). London: Routledge Studies in International Political Sociology.

Gabrys, J., Pritchard, H., & Barratt, B. (2016). Just good enough data: figuring data citizenships through air pollution sensing and data stories. *Big Data & Society*, *3*(2).

Gray, J., & Marres, N. (2018). *The data as public experiment*. London Ideas. Centre for London. Available at: http://jonathangray.org/wp-content/uploads/2018/07/GrayMarres_DataCityPublicExperiment.pdf

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, July-December 2018, pp. 1-13.

Habermas, J. (1996). Between facts and norms: Contributions to a discourse theory of law and democracy. Cambridge, MA: MIT Press.

Haklay, M. (2015). *Beyond quantification: a role for citizen science and community science in a smart city*. Prepared for the Data and the City workshop, 31st August and 1st Sept 2015, Maynooth University, Ireland. Available at: https://discovery.ucl.ac.uk/id/eprint/1470344/

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital citizenship in a datafied society*. Cambridge, UK: Polity Press.

Holdren, J.P. (2015). *Addressing societal and scientific challenges through citizen science and crowdsourcing*. Memorandum to the Heads of Executive Departments and Agencies. Washington DC: White House Office of Science and Technology Policy. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/holdren_citizen_science_memo_092915_0.pdf.

Kooiman J. (2003). Societal governance. In I. Katenhusen and W. Lamping (Eds), *Demokratien in Europa* (pp. 229–250). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Lam, W. (1996). Institutional design of public agencies and coproduction: A study of irrigation associations

in Taiwan. *World Development*, *24*(6), 1039–1054.

Lämmerhirt, D. Jameson, S., & Prasetyo, E. (2016). *Making citizen-generated data work*. Available at: http://civicus.org/thedatashift/wp-content/uploads/2017/03/Making-Citizen-Generated-Data-Work_short-report_.pdf

Lämmerhirt, D., Gray, J., Venturini, T., & Meunier, A. (2019). Advancing sustainability together? Citizen-generated data and the sustainable development goals. Available at: http://www.data4sdgs.org/sites/default/files/services_files/Advancing%20Sustainability%20Together%20CGD%20Report_0.pdf

Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, *29*(4), 446–454. Available at: https://www.sciencedirect.com/science/article/pii/S0740624X12000883

Meijer, A., & Potjer, S. (2018). Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly, 35*, 613-621-

Ruppert, E. (2018). *Socio-technical imaginaries of different data futures. An experiment in citizen data*. 3e Van Doornlezing. Erasmus University Rotterdam. Available at: https://www.eur.nl/sites/corporate/files/2018-06/3e%20van%20doornlezing%20evelyn%20ruppert.pdf

Ruppert, E., Isin, E., & Bigo, D. (2017). Data politics. *Big Data & Society*, 1-7. DOI: 10.1177/2053951717717749

UN Environment Science Policy Business Forum (2018). The case for a digital ecosystem for the environment: Discussion paper. Bringing together data, algorithms and insights for sustainable development. Available at: https://un-spbf.org/

Volten, H., Devilee, J., Apituley, A., Carton, L., Grothe, M., Keller, C., Kresin, F.,… Wesseling, J. (2018).Enhancing national environmental monitoring through local citizen science. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn, (Eds.), *Citizen science: Innovation in open science, society and policy* (pp. 337-352). London: UCL Press.

6 Private sector data for the public interest: Data sharing practices in twelve EU cities

Marina Micheli

6.1 Introduction

Privately held personal data collected by mobile phone operators, social media platforms, transport services, accommodation websites, energy providers, and so on, might offer insights and valuable opportunities to public authorities supporting their efforts to address both societal challenges and emergencies. This study looks at how local governments in Europe are (or are not) gaining access to privately held data examining the perspectives of city's managers and project leaders (e.g. chief data officers) that are working in the field of data, technology and urban innovation. Through twelve interviews, we collected experiences and opinions about the various operational modes for accessing privately held personal data. ***Public bodies, especially including local governments, might have a key role in the current European data landscape, as they could help promoting a more balanced data economy based in which the value produced with personal privately held data is redistributed across society*** (Couldry & Powell, 2014; Morozov & Bria, 2018; Bass et al., 2018; Adalovelace Institute, 2019).

6.2 How local administrations access to private sector data

Privately held data is often described as part of the "urban data landscape" (Kitchin, 2018). Yet, notwithstanding the expectations, the practice of data sharing between businesses and governments is currently sporadic and lack sustainability (HLEG, 2020; Martens & Dutch-Brown, 2020). The European Commission's High Level Expert Group on Business-to-Government (B2G) data sharing - established to provide recommendations on how to enhance the sharing of privately held data for the common good - identified several challenges. The lack of governance frameworks means that private companies have to face various uncertainties when sharing their data - in relations to liability regimes (who is responsible if inaccurate or biased data is shared that leads to discrimination), intellectual property, and competition law. Furthermore, companies face operational and technical challenges for the preservation of sensitive commercial information and the protection of customers' personal information (HLEG, 2020). In a study on the economics of B2G data sharing, Martens and Dutch-Brown (2020) identified the following economic barriers: monopolistic data markets (companies can charge high prices for data), high transaction costs and perceived ex-post risks for data providers, and lack of incentives for private firms to contribute to the public good if it might affect them negatively (e.g. competition, market regulation). Finally, public bodies lack a "culture" on data sharing (e.g. how to create value with it), have limited resources and deficiency of skills and also limited trust from both the private sector and citizens on public bodies accountable use of data (HLEG, 2020). The regulations currently available for privately held data sharing vary by EU country and are sector specific, while at the EU level regulations are currently being discussed (EC, 2020).

Notwithstanding the circumstances, some cities are developing their own strategies to access private data. The operational models adopted by local administrations to access privately held data are diverse and denote very different relations between these actors. Private companies, for instance, might share data with public bodies at no cost on a voluntary basis as corporate social responsibility, such as during an emergency or to support initiatives for the public interest. This mode for accessing private data is referred to as data donorship (Huyer & Cecconi, 2019; HLEG, 2020). Otherwise, public administrations might purchase data through public procurement (HLEG, 2020): triggered by specific needs, public bodies request to acquire a specified set of data, or data-driven insights, from a data supplier through a contract (Huyer & Cecconi, 2019: 16). A different approach is that of data sharing pools (Shkabatur, 2019; Micheli et al, 2020) in which a public authority establishes a partnership with other actors to pursue mutual interests, and commercial companies, government entities, data platforms, and/or research institutions exchange data in a collaborative way. They "analyse each other's data, and help fill knowledge gaps while minimizing duplicative efforts" (Shkabatur, 2019: 30). A related mode is that of data research partnerships, that is when public bodies collaborate with research/scientific institutions for a project of mutual interest to analyze privately held data that the latter have at disposal (HLEG, 2020). Another relevant mode for

accessing data of private sector consists in the introduction of data-sharing obligations as part of subcontracted services (HLEG, 2020): cities might include clauses in their tenders "specifying that a service provider must make any data that may be of public value available to the city council in machine-readable format" (Bass et al., 2018: 28). All modes could lead to access different types of data: raw, pre-processed (e.g. cleaned, re-sampled, normalised), processed (aggregated and combined) or insights derived from the data (HLEG, 2020). Private companies might be more willing to sell (or donate) insights deriving from internal data analysis ("intelligence sharing", such as dashboards, apps, reports), instead of actual datasets, as a way to keep control of information and reduce risks (Shkabatur, 2019; HLEG, 2020; Micheli et al., 2020). Furthermore, data could be shared in various technical solutions, such as Application Programming Interfaces (APIs), limited release of data under conditions stipulated in a contractual agreement, remote access by a trusted intermediary, etc. (HLEG, 2020). The study presented in this contribution contextualizes the different operational modes for accessing privately held data described in this section in the experiences and opinions of European city's managers, directors and project leaders.

6.3 Methodology

This study examines how the socio-technical practice of accessing privately held data is "constructed" throughout relationships between actors. ***Data sharing is not examined as a technical issue, neither as an economic activity, but as a socio-technical practice.*** The methodological approach is informed by the tradition of research in media domestication and the social shaping of technology (Silverstone & Haddon, 1996; Lievrouw, 2006). This method is adopted to understand how public actors envision their 'power to set the terms' on how privately held data is shared and what strategies they put forward to facilitate access. The study focuses in particular on the perspectives of specific actors from the public sector: cities' Chief Technology Officers, Chief Data Officers, or project leaders working on a city's innovation/smart city agenda. Twelve semi-structured interviews have been conducted with representative of as much cities during the course of 2019. A combination of purposive and snowball sampling procedures has been adopted for the selection of cities to be included. The participants were chosen in a way to have a diversified group by city size, area in Europe, and tradition of innovation. Together with the "usual smart cities" the list includes cities of different European areas and size (Table 6.1).

Table 6.1 – List of cities involved in the study[24]

City	Country	Size	EU zone
Amsterdam	NL	Large	North West
Barcelona	ES	Large	South
Ghent	BE	Small	West
Ljubljana	SI	Small	Central East
London	UK	Large	North West
Milan	IT	Large	South
Rennes	FR	Small	West
Rjeka	CR	Small	East
Tallin	EE	Mid-size	North East
The Hague	NL	Mid-size	North West
Vilnius	LT	Mid-size	North East
Zaragoza	ES	Mid-size	South

Large: > 1 million inhabitants; Mid-size: Between 300.000 and 1 million; Small: < 300.000.

The interviews, which lasted 70 minutes on average, have been conducted through VoIP (Voice over Internet Protocol) technologies that allow interviewing research participants using voice and video across the Internet via a synchronous connection. Conducted by the author, the interviews investigated the "concrete realities" of working in this area, digging into the actual experiences of these professionals, and simultaneously analyzing discourses and imaginaries, investigating their opinions on the topic. The transcriptions did undergo a qualitative thematic analysis through manual coding; the documents have been coded following the main themes of the interviews. The findings discussed in this contribution focus in particular on the analysis of the codes that refer to access to private data, which were labeled:

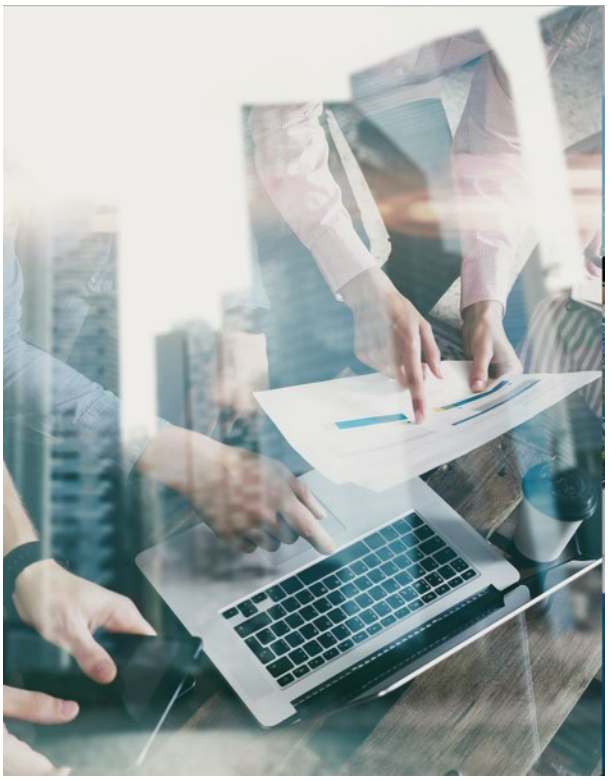
[24] To protect participants' identities, the numbered cities in the quotes does not correspond to the order of cities in this table.

"Operational modes of access", "Discourses and perspectives", "Relationships between actors", "Power to set the terms", "Strategies to negotiate power". The analysis aims to delineate common trends in the experiences and discourses, as well as key differences and how these relate to the specific context. During the interviews it was taken into account that digital innovation is a highly marketed issue for cities, for instance by explicitly asking about obstacles and unrealized projects, and being self-reflexive during the conversations.

6.4 Results

Local administrations' access to privately held data is a rather sporadic activity. All respondents stated that accessing privately held data is challenging. Data companies often have no interest in selling data, and neither in sharing it with a municipality (cfr. Section 2). The practices mentioned in the interviews are often pilot projects, activities at the "early stages", if not still in preparation ("figuring out"). In a few cities the topic was rather novel, as access to private data was not part of current/planned activities. Companies with mobility data were cited more often, both as potential, past or actual data providers, highlighting how access to privately held data might be promising especially for this sector.

The types of data providers most interesting according to those who participated in the study are: (1) utilities companies, and (2) online platforms and telecom operators. Utility companies are very interesting for municipalities. These are depicted as the ideal candidate for access to privately held data, but also as a difficult one to deal with, due to lack of human resources and interest in data sharing. Online platforms and telecom operators feature less prominently. According to a participant, big platforms are difficult to reach because they do not have representatives working at the local level. Furthermore, they seem to have less to gain from engaging with a municipality. Several other actors, beyond data holder companies, are involved for enabling public bodies to access (and use) privately held data. Start-ups, public universities, research institutions and civic organizations also have an important role in this context, helping with data stewardship and analytics. Collaborations with research institutions or PhD students, is mentioned as an enabler for working in this field, especially by cities with less economic resources.



In the remaining sections we contextualize the operational modes for accessing private data in the discourses and experiences of the professionals of twelve cities (See Table 2 for a summary). Hopefully these findings could shed light on the hindrances, as well as the promising avenues, for public bodies' access to privately held data.

6.4.1 Data donorship: "Free lunch?"

Some respondents recognized the possibility to access privately held data or information at no cost, as companies/data providers occasionally make it available for free on a voluntary basis. This *data donorship* mode, however, was often associated to a specific discourse. Instead of being described as a philanthropic move, it was acknowledged as a marketing strategy used by private companies, which favored already privileged "smart cities". Companies that freely share data with cities – and collaborate with them to develop products or services valuable for the municipality without asking anything in return - do so because this allows them to market new products and services to other cities in the future.

"A company will approach us and say "hey, we've got this…", but it will be on a pilot form only, because they want to say that they work with the, you know, the Mayor of cityX, in order to market their products

in other places (...) In this case they were gaining some free promotion from these experimental samples (...) And we, yes, we didn't pay them any money." (city09)

Companies use the reputation of (smart) cities as promotional material. Thus, being a well-known city seems to be a key enabler for such form of access to private data. This creates a double source of disadvantage for smaller cities because they lag behind and are proposed the same service to a price. This phenomenon emerged as an 'ethical dilemma' in a couple of interviews in which managers questioned their city's position (its privilege or lack of thereof) in relation to that of others:

"They (companies, ndr.) can say to other cities, "hey cityX did this use case, our data is very valuable, so our product is also more valuable, so you can pay more". This is for us a way to work with these companies. But again, there is the ethical question, do we want to have a free lunch if others are paying for it?" (city02)

[During the meeting of a national group of cities] "He said, 'Okay, for us, in cityX, the conditions under which we deal with the great companies is that we deal for nothing. They come and they develop some solutions, and we work together in partnerships, and it's free for us'. And the other one in the room, they said, 'Okay, it's free for you, but it's not fair. You have money, more than we have'. And when they get to us, they say, 'Okay, we developed a solution with cityX.'" (city06)

This operational mode for access tend to be associated to smaller or one-time-only projects; one interviewer described that as an "incidental partnership" to stress the volatile nature of the initiative.

6.4.2 Public procurement of data: Resistance and caution

While almost all interviewees discussed the possibility to acquire data directly through *public procurement*, most were against this solution and the remainders experienced it with great circumspect. This operational access mode is contrasted with ideological arguments: (1) data produced in public spaces should be accessed by public bodies and not be treated as a commodity; (2) local administrations have to serve the public interest and should not invest economically is acquiring data; and (3) it is important that cities keep sovereignty over data, becoming a buyer to a private platform (especially if a big corporation) might undermine their autonomy.

"I'm very reluctant to pay for data (…) first of all; we need to keep a certain amount of independence from third parties when it comes to information on your city. Because data is not neutral and if we become very dependent on a tracker, we know there's not a lot of competition on the market, because the technology is expensive, scaling up is expensive, the knowhow is a long process (…) I never heard that a product is becoming cheaper over the years." (city04)

"One of the most important things in the equation is that we are not putting money in it, so if we bought the data that would be easy (…) sometimes it is their business model, so they don't want to give the data for free, they want to have money. And well, we don't have that kind of money and it's also some kind of a principle discussion that the data has been collected public space. Data collected in public space is from everyone, it's not just from the company who happens to put a sensor" (city05)

Those engaging in public procurement of data also questioned its effectiveness. Companies, in fact, often sell data packaged with limited options for personalization, they send finished products (such as dashboards or PDFs) that curtail the possibility of intervention on data. Furthermore, they are not transparent regarding data quality and representativeness ("we've also found companies over-promise").

"In the best scenario, we receive data in a PDF, but not in an XML format, so it's very difficult for us to process the data in our systems and well, basically it's not of any use at all if we receive a PDF." (city10).

Respondents who obtained data through public procurements stress the experimental setup of such activities, described as "evaluation phases" to assess the quality of the data and the opportunities it affords. This mode for access is criticized mostly because it does *not* allow municipalities to be involved in defining the information needed and the kind of analysis performed, leaving great decision-making power in the hands of the company.

"The company has set the rules and we're not at a stage yet where we've set the rules in any of the projects I've been involved in. (...) But until you agree to their terms, you can't get your hands on the data.

So, we've had quite a few discussions that have gone round and round and round, maybe for a year, maybe for a year and a half." (city09).

Some respondents suggested that cities should do collective bargaining to deal with the issues of the long negotiations and the high prices enforced by private companies. This strategy, which consists in cities creating a coalition and relating with companies altogether, would allow municipalities to strike better deals.

"We think that cities have a real role in basically collective bargaining on this and telling companies what they'll pay for it, rather than the other way round." (city08)

"With a collective, in a sort of a collective effort with other stakeholders, to share also the cost (…) we could also imagine that the data, which is for sale, we can go and buy it under a collective." (city06)

6.4.3 Data partnerships: Win-win collaborations

A completely different attitude emerged when respondents claimed that they established (or wish to establish) partnerships with private companies to cooperate and solve challenges together. In this operational mode local administrations identify shared interests with the private companies holding data, seeking a win-win collaboration. Both parties are involved in the project and in the analysis, at time also sharing the objective for which the data is used. Local administrations eventually give administrative data in exchange, creating a data sharing pool with the private company (Shkabatur, 2019; Micheli et al., 2020), or are simply a partner in the development of a product (e.g. a public service).

"We try to talk to them, like, 'what are your incentives and how can we help you?' We didn't go to them and say, okay, we need your data. We say 'how can we work together?'" (city02)

According to the respondents, this approach allows establishing productive relations with private companies and to address societal challenges more effectively. This relation is described as a form of "co-creation" in opposition to "buying data" and being "just a client" of data holder companies. A key enabler seems to be to work with people already in ones network with whom a personal relationship has already been established.

"We have a history with the people. I mean the people working in companyX, I know her for, I've knowing her for now five years maybe. Had discussion on different topics, and now I know where she wants to go. She knows where we want to go. We know where we could go together- it is easier. With companyY, it is the same. We are working with them on data since 2010."

Another important enabler is the societal relevance of the projects on which these collaborations are based. According to some respondents, the new generations of developers are interested in working on socially relevant issues. Therefore, establishing personal relationships with them (for instance during hackatons) pave the way for future collaborations. Occasionally respondents highlighted that a partnership originated from a common goal (between the municipality and a private company) to impede the dominance of big tech corporations in a certain sector, such as Google Maps as mobility app. Yet, private companies join partnerships to develop a business model or a commercial product to offer to other cities/clients. Therefore these partnerships might also lead to inequalities between cities (such as for data donorship), as municipalities with advanced knowledge and expertise are more likely to find companies willing to collaborate with them. The more "experienced" a city is, the more it has to offer to private companies in terms of data and support.

"The collaboration so far is more that it's a win-win, that they give us what they have, and they see what we do with it, how we enhance it, which makes their product better. So, it's that iteration." (city02)

6.4.4 Tender obligations: Data sovereignty

Another way to access private data consists in putting clauses within tender contracts with suppliers so that data collected as by-product has to be accessible to the city council. A minority of respondents has put this in place, while a few others were seriously considering it. This mode can allow accessing data of suppliers and, theoretically, of any company that has a contractualised relationship with the municipality (public transport, waste collection, avenues rented for events, telecommunication infrastructures, etc.).

"We have done that [paid for data] in a couple of situations where it was not specified well in the tender (…) What we try to do now is prevent that by making our contracts better and have a warrant in our contract that says all the data being used in something we buy, belongs to the city of X." (city02)

"We are thinking about something more systematic, like how to introduce data questions in our contracts, in our agreements on different public policies. It's quite a different perspective. It's not how to reach new partners on data, but how to introduce data with our historical partners." (city06).

Table 6.2. Summary of the operational modes for accessing private data contextualized in the discourses and experiences of twelve local administrations.

Operational mode	Discourses	Public/private relations	Strategy to support access
Data donorship 	"Free lunch" "Incidental partnership"	The promotional city	Reputation
Public procurement of data 	"Reluctant to pay for data to keep independence" "Data as a product"	City as a client	Negotiations Evaluation phases Collective bargaining
Data partnerships 	"Win-win collaborations" "Societal relevance"	City as a partner	Data in exchange Experience Personal relations Hackatons
Tender obligations 	‘‘Data as a public good" "Responsible use of data for the public interest" "Independence from companies"	The sovereign city	Legal framework Collective of cities

The reasoning behind this is analogous to the explanations for not using public procurement. Data collected for public services, should be available to public bodies as data sovereignty allow directing the digital transformation at the service of public interest.

"Thinking about contract services, tenders, saying that you are providing services as you were the city council, so it's not your business to collect data about the city. Okay, your business is to provide a service that you are contracted for, so the data you are collecting within the service needs to be available for everyone to provide, or for the city to provide, or to improve the service." (city12)

A strategy put forward by a few respondents to enhance such mode of access consists in working collectively with other cities and jointly define the same contractual framework to be used with private companies:

"we are working together with the association of cities and we want to come up with a model contract in which we can come up with the juridical text where we can use that to make a contract with these businesses upfront. So, there is no discussion about a data, but it would be every city in the country is using the same contract, so it's no use to go shopping to another city because it's very similar."(city10)

6.5 Discussion

This short chapter summarizes the findings from a qualitative study with innovation/data managers of twelve European cities. Access to privately held data has been examined as a socio-technical practice that is still "in-the-making", as no established praxis is fixed yet. The research contextualized four operational modes for accessing private data in the experiences and discourses of cities innovation/data managers. From the results we learned that ***bigger and smart cities seem to have more chances to access privately held data. Their reputation, their professional network, their resources and expertise, put them in a favorable position to be contacted by private companies (for* data donorship*) or be saluted as partners (for* data sharing pools*).*** Private companies, in fact, use such "use cases" to market their services and products to other cities. Further research could see to what extent a divide between cities regarding their chances to access data and use exists and with what implications. Another underlying issue, emerging from the discourses of participants, is that of data sovereignty: ***respondents are wary in buying data through public procurement, both because that would place them in a dependent position (*economically*), and because there is a lack of transparency regarding privately held data quality and limited possibilities to control how it is configured and used***. To preserve control, respondents envision a form of collective bargaining to strike better deals when acquiring data, but most often they look towards other modes of access. Among the best ways to keep control of data is establishing collaborative relations (co-creation) through *data partnerships* with private companies and, when possible, including *obligations clauses in tender contracts* with suppliers. ***The strategies*** described by the respondents ***to facilitate access to private data are collective efforts in which cities join forces for the cause: from collective bargaining, to develop a common contractual framework to use with businesses for tenders or partnerships***. These tactics could help to level the playing field, lessening the inequalities described above, and increasing cities' strength in demanding access to privately held data with a public interest.

The study provided qualitative insights in the lived experiences of cities' innovation and data managers in relation to access to private data as a way to reflect on the role of public bodies in the current European data ecosystem. Yet, this a short chapter and only provide some short hints about the subject. Further avenues of research in this area include: an analysis of how cities are using or planning to use privately held data; a systematic review of the resources available to local administrations and how they relate to access and use of private data; how citizens, and public trust, are taken into account when gaining access to privately held personal data.

References

Adalovelace Institute (2019). Rethinking data. Changing the data governance ecosystem. Report (pp. 1-26): Adalovelace Institute.

Bass, T., Sutherland, E. & Symons, T. (2018). Reclaiming the Smart City. Personal data, trust and the new commons. Report (pp. 1-76): NESTA.

Couldry, N., & Powell, A. (2014). Big Data from the bottom up. Big Data & Society, 1(2), 1-5.

European Commission (EC) (2020) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "a European Strategy for Data" COM(2020) 66 Final. Luxembourg, Publications Office.

Huyer, E., Cecconi, G., & Capgemini Invent (2019). Analytical Report 12. Business-to- Government Data Sharing. Report (pp. 1-24) European Data Portal.

HLEG (2020). Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing (pp.1-112) European Union.

Kitchin, R. (2018). Data-driven urbanism. In R. Kitchin & T. P. Lauriault, Data and the city (pag. 44-56). New York: Routledge.

Lievrouw, L. A. (2006). New Media Design and Development: Diffusion of Innovations v Social Shaping of Technology. In L. A. Lievrouw & S. Livingstone (Eds.), Handbook of New Media (pp. 246– 265). Thousand Oaks: Sage.

Martens, B., & Duch-Brown, N. (2020). The Economics of Business-to-Government Data Sharing. JRC Technical Report, JRC Digital Economy Working Paper 2020-04 (pp. 1-31), European Commision.

Micheli, M., Ponti, M., Craglia, M., and Suman A.B. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 1-15. DOI: 10.1177/2053951720948087

Morozov, E., & Bria, F. (2018). Rethinking the Smart City: Democratizing Urban Technology. New York, NY: Rosa Luxemburg Stiftung.

Shkabatur, J. (2019). The global commons of data. Stanford Technology Law Review, 22, 1–46.

Silverstone, R. & Haddon, L. (1996). "Design the Domestication of Information and Communication Technologies: Technical Change and Everyday Life". In R. Mansell and R. Silverstone (Eds.), Communication by Design: The Politics of Information and Communication Technologies (pp. 44 -74). Oxford: Oxford University Press.

PART B:
on the Governance with Digital Data

7 Probabilistic Synthetic Population modelling for EU policy support

Jiri Hradec, Margherita Di Leo, Massimo Craglia

7.1 Introduction

This stream of activity of the project started from the recognition that there is a widening gap between the accuracy and timeliness of the data available to the big commercial platforms to personalise services to their customers and influence their behaviours via recommendation algorithms, and the slow pace of official statistics released at area-level aggregates and often years out of date. Governments of course do collect and have at their disposal individual-level data about their citizens and residents but most of the times are not allowed to use it and link it to other individual-level data because as a society we value our privacy and confidentiality (from government). For this reason, government policy is normally based on the ecological fallacy of assuming that people living in the same arbitrary spatial units at which statistical data is released all share the same characteristics. In Digitranscope we started therefore to consider how we could use some of the techniques of the commercial sector of consumer profiling and targeting and apply them to publicly available administrative data to develop more "personalised" policies targeted to those who need it most. The key idea to develop these personalised profiles without having to deal with real personal data was to create a probabilistic synthetic population from disaggregated official statistics.

Synthetic populations in the form of disaggregated individual data represent the main input entities to several multi-agent models and micro models, used in many different contexts in policy design and evaluation, from economic simulations (labour market policies, tax benefit, poverty-reduction policies, multi-country microsimulations, etc.) to agricultural and environmental policies, education, health, demographic and social well-being, just to mention a few. In policy decision-making, the levels of complexity that come into play include the heterogeneity of the population and its often under represented different subgroups; the behavioural response of individuals; and difficulty of assessing ex-ante and ex-post the impacts on different subgroups, over time and space. To try and disentangle such a level of complexity, it is often necessary to recur to models (Aaberge et al. 2014).

7.2 Review of Methods

Micro models based on synthetic population can simulate the effects of proposed policy implementation on subgroups, as well as estimating program costs and caseload (National research Council, 1991). The spatial dimension is important to consider whenever the model is spatially targeted. Datasets with a spatial component are usually available at census area (or coarser level of detail). The input data to the models may be rich in contextual data related to persons or households and lacking on the spatial information, or the sample size may be too limited to be representative at a fine spatial scale, or vice-versa, data at fine-grain spatial resolution may have gaps in contextual data (Aaberge et al., 2014).

Synthetic populations are a powerful tool because they can be extremely informative without breaking the privacy of citizens, but still reflect the complexity of the structure of the population, and the characteristics of the individuals that influence their behavioural response. A synthetic population is designed to reflect the heterogeneity of the real population, including minorities and under-represented individuals that would not be characterized considering just the general statistics of the population.

In Digitranscope we aimed at reconstructing a synthetic population that would serve as a baseline to be used as an input to different kinds of models. This baseline should be flexible enough to be successively enriched and updated whenever more data becomes available. The population has to carry all possible information until a model is chosen and the relevant features can be selected accordingly. The baseline is characterised by all features present in the source.

Any individual is characterised by a set of variables, or features: I = <f_1, f_2, ..., f_N>. Some are independent, like age and gender. Others are correlated, e.g. education, family composition, place where

they live etc. The system is characterised by a state that can be updated with ancillary data and constraints deriving from additional data sources.

Several methods have been proposed for generating synthetic populations (Farooq et al. 2013; Ye et al. 2009; Ironmonger et al. 2000; Jorosz, 2013; Antoni et al. 2017; Arentze et al. 2007; Lenormand and Deffuant, 2012; Gargiulo et al. 2010; Delhoum et al. 2020; Thiriot and Sevenet 2020; Sajjad et al. 2016; Stevens et al. 2015; Namazi-Rad et al. 2014). Most methods tackle the generation of the synthetic population as a fitting problem. Two main families of techniques are Synthetic Reconstruction (SR) and Combinatorial Optimization (CO) (Farooq et al. 2013; Namazi-Rad et al. 2014). Both families of methods start from a very limited data sample that represents from 1% to 5% of the population of disaggregated data available at individual level. These data are also known as PUMS (Public Use Microdata Sample), or simply microdata. This is an anonymized subset of census data that Statistical Offices put at disposals of researchers, after making sure of removing every location detail and blurring other information that may allow to reverse engineer the identity of the individuals.

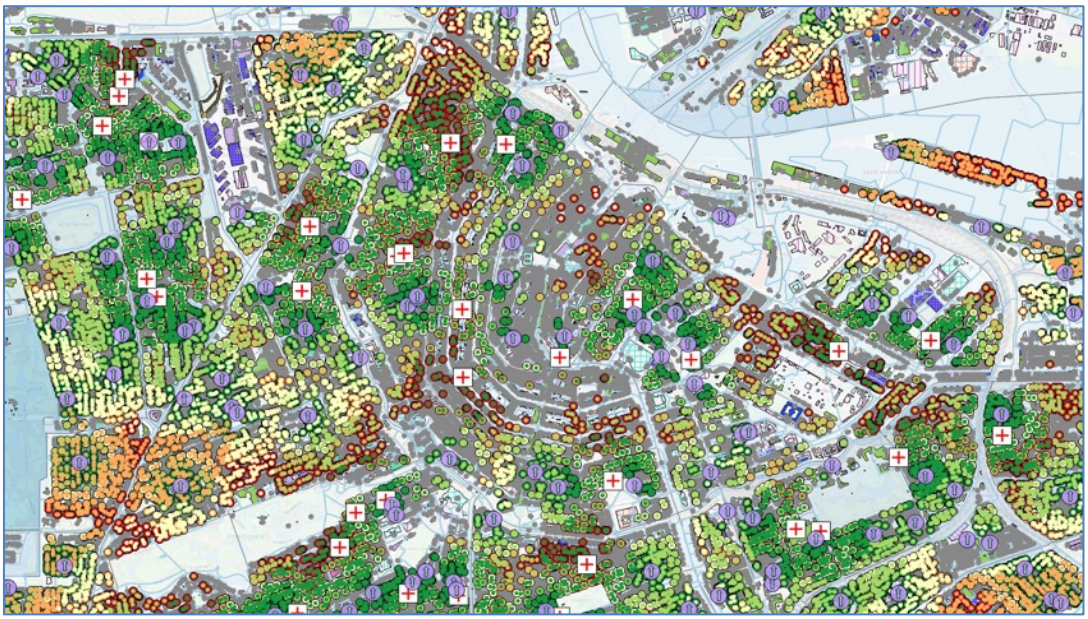
Synthetic Reconstruction is based on microdata and cross-classification tables, also released by the Statistical Offices, that present the statistical figures of the population at various levels of details of 1 – 3 attributes, like for example unemployment by level of education by gender. From such tables it is possible to derive conditional probability for a certain co-occurrence of attributes (“marginals”). Starting from the microdata (also called “seed”), both SR and CO methods reconstructs the missing records (individuals) using the “marginals” as constraint, making sure that the statistical figures of the population are reflected in the modelled population, within some level of accuracy. It is normally necessary to operate a selection of the features that are important to the problem under investigation, neglecting or relaxing the constraints on the remaining attributes (Namazi-Rad et al. 2014).

The most popular SR methods are Iterative Proportional Fitting (IPF) and Iterative Proportional Updating (IPU) techniques. Among Combinatorial Optimization methods, Hill Climbing (HC) is one of the more frequent, proposing a random solution and iteratively tries to improve it maximizing an objective function, measuring the performance at each loop. Hilltop is reached when the predefined errors are lower than a certain predefined threshold. Each method has its strengths and weaknesses, and for a full review see Hradec et al. 2021.

7.3 Applications

In Digitranscope we used the IPF/IPU techniques to generate the synthetic population of Amsterdam, assigning it by household to properties, and identifying vulnerable groups such as elderly people living alone or single parents with small children that should be the priority target for policies supporting access to healthcare and education (see Fig 7.1), or the energy transition (insulation, sustainable energy sources).

Figure 7.1: Access to education and health for single-parent families and elderly people



Note: Properties in green are well serviced while those in yellow and red and not.
Source: JRC

Having prepared the base to target these policies, we planned a survey of people' attitudes to the energy transition in collaboration with Housing Europe, a federation of social housing organisations in Europe, to identify the policy options more appropriate to trigger the support towards this transition. This survey has been delayed due to the Covid pandemic and we will therefore complete this usecase in 2021 after the end of the Digitranscope project.

Another application we developed was in the context of the JRC Corona Virus Task Force supporting the European Commission and the EU member states to assess the relative risks of reopening different economic sectors after the lock-down period in Spring 2020. During the initial lock-down in the Spring of 2020, which was more or less stringent in different EU countries, only essential eservices were kept running at all times (e.g. utilities, food production and distribution, pharmaceuticals, essential infrastructures). As the peak of the first wave was reached and passed, there was a need to identify which economic sectors to open first to allow the restart of the economy whilst reducing risk of second waves of infection.

To answer this question, several steps were needed:

1. Create a model of the likely number of daily contacts of each person based on both economic and social activities,
2. For the economic activities, identify the relative number of daily contacts of each worker by economic sector, taking also into account for each sector the potential for telework and the proportion of workers commuting daily by public transport in "normal" circumstances.
3. Assess the socio-economic impact of risk (by gender and income)
4. Assess the spatial distribution of risk, based on socio-economic characteristics of different regions, and commuting patterns.

Steps 1-3 are described in Craglia et al. (2020). Here we discuss briefly step 4 which involved the creation of a synthetic population for the whole of France, as data from the French Statistical Institute (INSEE) was found to be the most readily available for this task. It was not necessary to use any of the methods reviewed in Section 2 to create this synthetic population as the INSEE makes available anonymised detailed files that make it possible to carry out exploratory analyses of data, to model behaviours, or simply to tabulate on a particular subpopulation defined according to certain criteria: belonging to a geographical area and / or statistical unit presenting certain characteristics.

The key datasets used in this usecase were:

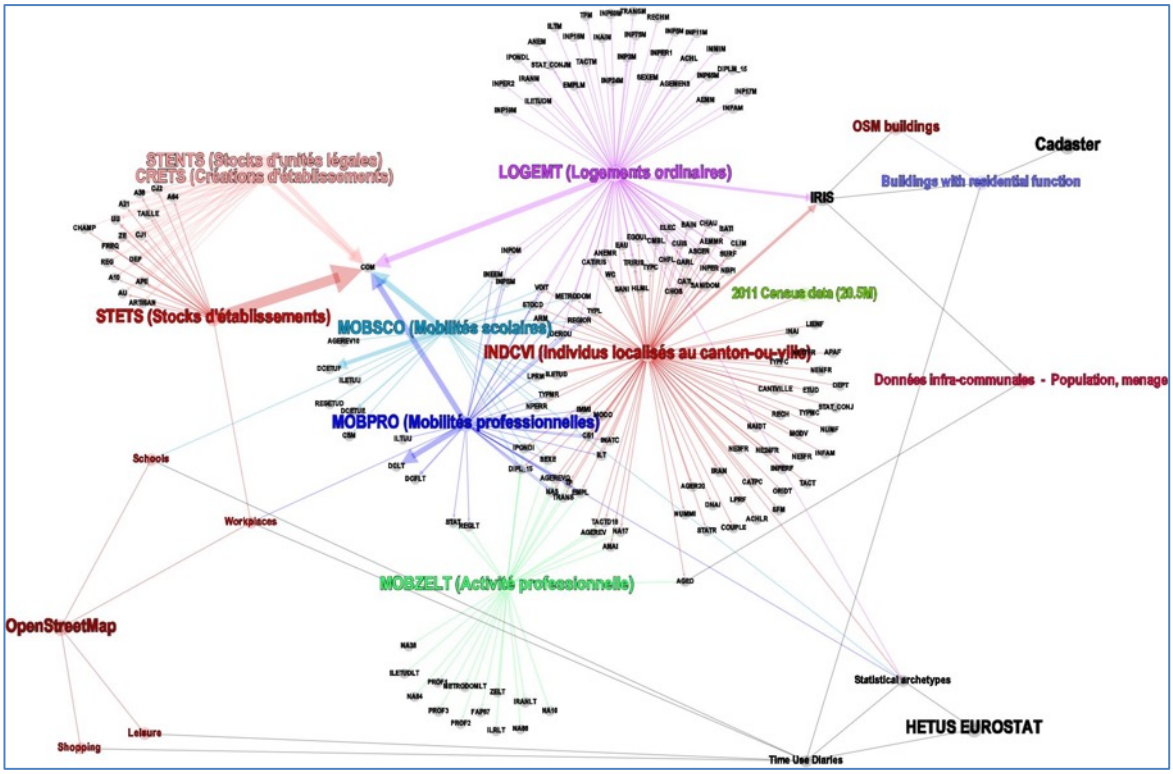
- INDCVI (Individus localisés au canton-ou-ville): Table containing the characteristics of the individuals, such as age, sex, level of education, household composition, etc. Each record contains attributes of an individual aggregated by means of a weight (IPONDI) that gives a measure of the frequency with which every record (profile of the individual) is found in the population in a certain area. Records also contain the 5-digit code of the census area (IRIS code). Small census areas, where the privacy is at stake, are grouped into larger areas and the attributes are given at a coarser resolution.
- LOGEMT (Logement); Every record in the table corresponds to an ordinary dwelling described according to its location, its characteristics (category, type of construction, comfort, surface area, number of rooms, etc.), and the socio-demographic characteristics of the household residing there. Household information is provided only when the accommodation is occupied as the main residence. Information are aggregated by the weight.
- MOBPRO (Mobilités professionnelles): Table containing the information about professional mobility. Each record in the file corresponds to an individual described according to the characteristics of his trips to go to work (home-work trips), his main socio-demographic characteristics, as well as those of the household to which he belongs.
- MOBSCO (Mobilités scolaires): Table containing the information about the mobility related to education. Each record in the file corresponds to an individual described according to the characteristics of his trips to go to attend an education institute (home-study trips), his main socio-demographic characteristics, as well as those of the household to which he belongs.

- MOBZELT (Fichier Activité professionnelle des individus (localisation à la zone d'emploi du lieu de travail). Each record in the table corresponds to an individual located at the workplace described according to the characteristics of their trips to work (home-work trips) as well as their main socio-demographic characteristics. All active individuals with a job, aged 15 or over, registered and working in France are taken into account.

Additional datasets included the map of the census tracks used by the INSEE, and data from the cadastre about properties cross linked with geographic data form the French Geographic Institute (IGN) and OpenStreetMap to create as detailed a map with the distribution of dwellings by type.

Data about the location of educational establishments was extracted from the Ministry of Education[25] while the location of economic activities was obtained by cross referencing the data from MOBZELT which covers 64 different economic activities with the buildings for the OpenStreetMap database. The linkages between the datasets are shown in Figure 7.2.

Fig 7.2 Conceptual model of the synthetic population for France.



Source: JRC

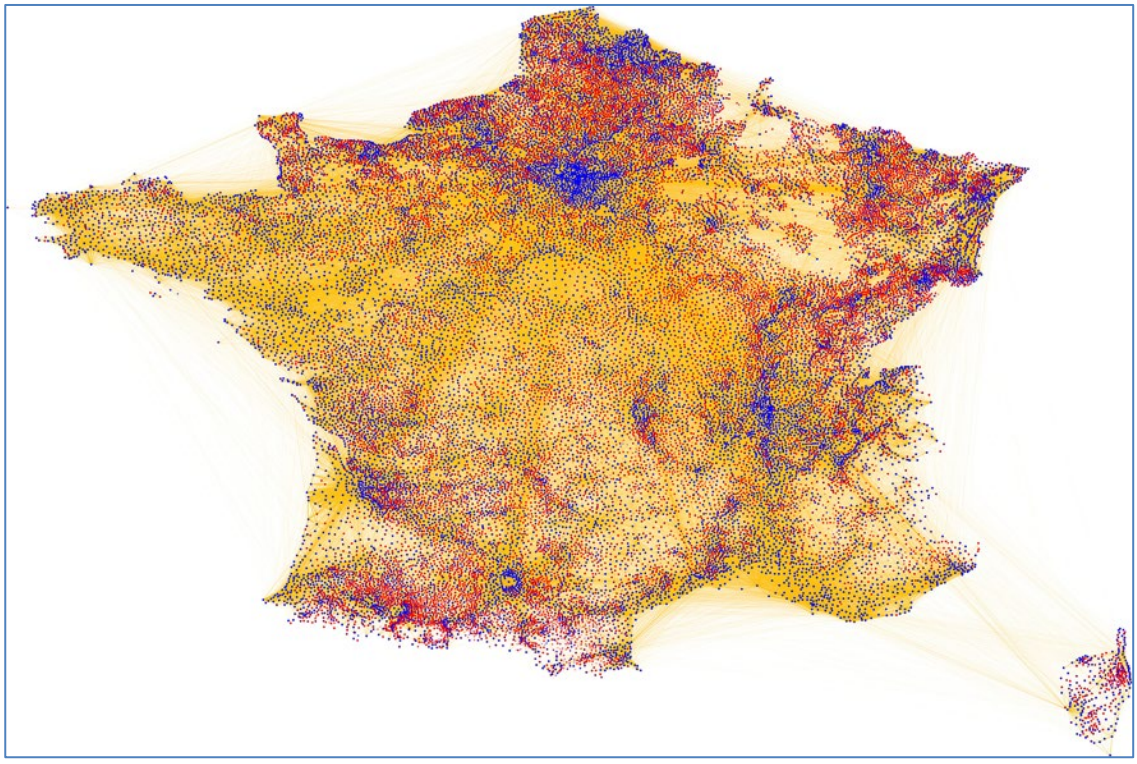
By linking the datasets above it was possible in the first instance to create families and households and then to attribute them to individual buildings. This combinatorial optimization is known as the Variable Size Multiple Knapsack Problem. Some authors (Thiriot, S., and Sevenet, M., 2020) propose a probabilistic approach to pair households to housing. This problem can be tackled in different ways, no solution is perfect but there is always a trade-off between precision and computational intensity. Aiming at a better precision is only possible when the input data add up useful information. Sometimes least computationally intensive solutions offer reasonable results as well. In our case, having any additional attribute to houses, e.g. year when built, would make people positioning much more precise. Another source of uncertainty is

[25] https://data.education.gouv.fr/explore/dataset/fr-en-adresse-et-geolocalisation-etablissements-premier-et-second-degre/table/?disjunctive.nature_uai&disjunctive.nature_uai_libe&disjunctive.code_departement&disjunctive.code_region&disjunctive.code_academie&disjunctive.secteur_prive_code_type_contrat&disjunctive.secteur_prive_libelle_type_contrat&disjunctive.code_ministere&disjunctive.libelle_ministere

that, in the absence of better information, we assumed that larger families would inhabit larger housing surfaces, which is obviously not always the case[26].

Notwithstanding these limitations, we were able to model the synthetic population of 63 million people, in 35 million households allocated in 10 million houses in France including their travel to work and study behaviour, and the proportion of people using public transport by economic sector. The figure 7.3 below shows the model of the commuting patterns of 26 million French in 2016 with blue/red scale showing commuting balance (blue = positive influx and red = outflux).

Fig 7.3: Influx-outflux of French commuters 2016



Source: JRC

Assuming the same commuting patterns by public transport by sector as those of France (because of lack of data in other countries) it was possible to arrive at a cumulative estimation of the relative risk or reopening the economy by sector, and the social and geographical distribution of potential impacts. This helped to informed the European Commission and the EU member states about possible policy options. Ultimately, the choice of what to open, where and how is political as it needs to balance the health vs. the economic and social risks, but this example shows the opportunities offered by the application of AI methods on available official data to support policy.

7.4 Conclusions

In this chapter we have introduced the concept of probabilistic synthetic population and shown applications at both urban and national/European level. The clear advantages of using synthetic data is that we are able to retain the full richness of the geographic and socio-economic/demographic distributions at the individual level without encroaching into personal data on the one hand nor loosing detail through the traditional approaches of aggregation. Pseudo-anonymization (i.e. by masking or obfuscating) does not protect against de-anonymization as advances in machine learning lead to ever smarter reidentification attacks.

[26] For a full discussion of the method and limitations see Hradec et al. (2021)

We have recreated the synthetic population as a directed network graph where synthetic individuals belong to synthetic families and households and live in real houses with real workplaces, schools and shopping and leisure places. Such a population is fit for epidemiological studies, where either attributes or graph properties help design communities of common features (e.g. same city in simplest example) and meta-population connects these communities. Other sectors and policies may benefit accordingly. We are now working with the Central Bureau of Statistics of the Netherlands to replicate and validate this approach to synthetic population modelling in the Netherlands. We plan then to work with other statistical agencies and then EUROSTAT to bring this approach EU-wide and have a robust population base for comparative modelling and policy design.

One of the key future opportunities is then to link the synthetic population with behaviour as inferred from either EU-wide data collections like the Eurobarometer[27] series or national/local survey, like the one we are currently doing with Housing Europe on the attitude to the energy transition in social housing. A further possibility is to enrich the synthetic population with the highly detailed and comprehensive behavioural models developed by the large web platforms for marketing purposes, if they were made accessible, or to encourage the public to "adopt" their synthetic personal digital twin and endow it with their behavioural responses to stimuli or problems, for example in a gaming environment. The possibilities are endless, but the outcome could be a very rich and detailed framework for agent-based modelling, where unified population used in different domains will lead to comparability of the results. It can be consistently used to better understand the dynamics of an infectious disease outbreak, socioeconomic impacts and policy response. This framework would allow to design "personalised" policies that are target to those who need support most and could be location-specific. In this way, also the feedback on policy outcomes could be localised and personalised and allow a much richer understanding of what worked well and what can be improved, opening the door in the future for policies-that-learn by design.

In another line of work, presented in the next chapter, we have used artificial intelligence tools to identify commonalities and patterns in the European legal and technical documents, helped bridge the domain jargons and started extracting and verifying facts from texts. This tool has helped us to move towards analysis of legislation on all levels of the European administrations, from the local and regional bylaws to national legal frameworks to the European legal umbrella. The existing legal framework settled into a certain local optimum. We started discovering where our simulated population has needs while the legal analysis shows the boundary conditions to be able to satisfy these needs. Ultimately, we may be able to completely redesign the policy cycle so that it is built bottom-up with people really at the centre of government intervention.

References

Aaberge, R., Bach, S., Bargain, O., Burlacu, I., Buslei, H., Capéau, B., ... & Decoster, A. (2014). Handbook of Microsimulation Modelling. Emerald publishing.

Antoni, J. P., Vuidel, G., & Klein, O. (2017). Generating a located synthetic population of individuals, households, and dwellings. Luxembourg Institute of Socio-Economic Research (LISER) Working Paper, (2017-07). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972615

Arentze, T., Timmermans, H., & Hofman, F. (2007). Creating synthetic household populations: problems and approach. Transportation Research Record, 2014(1), 85-91. URL: https://journals.sagepub.com/doi/10.3141/2014-11

Craglia M. (Ed.), de Nigris S., Gómez-González E., et al. 2020. *Artificial Intelligence and Digital Transformation: early lessons from the COVID-19 crisis*. EUR 30306 EN, Publications Office of the European Union, Luxembourg.

Delhoum, Y., Belaroussi, R., Dupin, F., & Zargayouna, M. (2020). Activity-Based Demand Modeling for a Future Urban District. *Sustainability*, 12(14), 5821. URL: https://www.mdpi.com/2071-1050/12/14/5821

Farooq, B., Bierlaire, M., Hurtubia, R., & Flötteröd, G. (2013). Simulation based population synthesis. *Transportation Research Part B: Methodological*, 58, 243-263. URL: https://econpapers.repec.org/article/eeetransb/v_3a58_3ay_3a2013_3ai_3ac_3ap_3a243-263.htm

[27] https://www.europarl.europa.eu/at-your-service/en/be-heard/eurobarometer

Gargiulo, F., Ternes, S., Huet, S., & Deffuant, G. (2010). An iterative approach for generating statistically realistic populations of households. *PloS one*, 5(1), e8828. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008828

Hradec J, Craglia M., and Di Leo M., 2021 *Multipurpose Probabilistic Synthetic Population for Policy Applications*. JRC Technical Report (forthcoming)

Ironmonger, D., Jennings, V., & Lloyd-Smith, B. (2000, October). Long term global projections of household numbers and size distributions for LINK countries and regions. In *Project LINK meeting*, Oslo, Norway on (pp. 3-6). URL: https://pdfs.semanticscholar.org/832a/8931ae225cf0148e564efaf8d5f29dc67d7a.pdf

Jorosz, B. (2013). Estimating Households by Household Size Using the Poisson Distribution. In *Population Association of America*. URL: https://paa2013.princeton.edu/papers/130204

Lenormand, M., & Deffuant, G. (2012). Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. *arXiv preprint* arXiv:1208.6403. URL: http://jasss.soc.surrey.ac.uk/16/4/12.html

Namazi-Rad, M. R., Mokhtarian, P., & Perez, P. (2014). Generating a dynamic synthetic population–using an age-structured two-sex model for household dynamics., *Plos One* 9(4), e94761. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094761

National Research Council. (1991). Improving Information for Social Policy Decisions--*The Uses of Microsimulation Modeling: Volume I, Review and Recommendations (Vol. 1).* National Academies Press.

Sajjad, M., Singh, K., Paik, E., & Ahn, C. W. (2016). A data-driven approach for agent-based modeling: simulating the dynamics of family formation. *Journal of Artificial Societies and Social Simulation*, 19(1), 9. URL: http://jasss.soc.surrey.ac.uk/19/1/9.html

Soille, P., Burger, A., De Marchi, D., Kempeneers, P., Rodriguez, D., Syrris, V., & Vasilev, V. (2018). A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Generation Computer Systems,* 81, 30-40. URL: https://www.sciencedirect.com/science/article/pii/S0167739X1730078X

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2), e0107042. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0107042

Tanton, R. (2018). Spatial microsimulation: developments and potential future directions. International *Journal of Microsimulation*, 11(1), 143-161. URL: https://www.microsimulation.org/IJM/V11_1/IJM_11_1_4.pdf

Thiriot, S., & Sevenet, M. (2020). Pairing for Generation of Synthetic Populations: the Direct Probabilistic Pairing method. *arXiv preprint* arXiv:2002.03853. URL: https://arxiv.org/pdf/2002.03853.pdf

Ye, X., Konduri, K., Pendyala, R. M., Sana, B., & Waddell, P. (2009, January). A methodology to match distributions of both household and person attributes in the generation of synthetic populations. In 88th *Annual Meeting of the Transportation Research Board*, Washington, DC. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.537.723&rep=rep1&type=pdf

8 Semantic Text Analysis Tool (SeTA)

Jiri Hradec

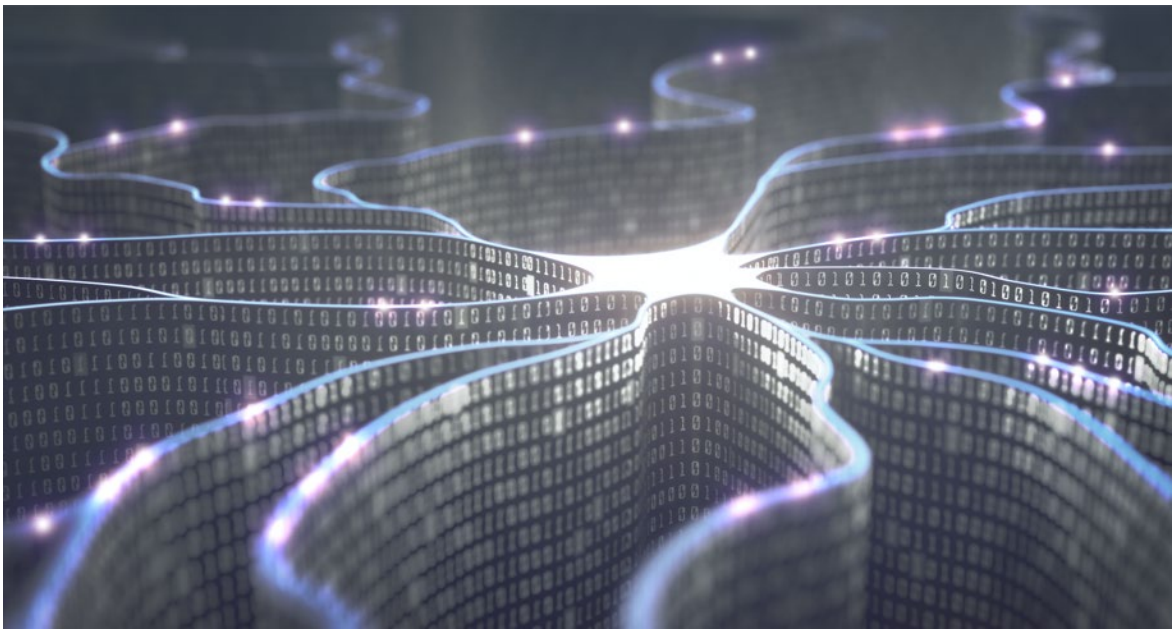
8.1 Introduction: What is SeTA and why it is important

Searching for information is the story of our lives. It is like trying to find your house keys on a Monday morning after you've spent the weekend with your kids in the mountains. You try desperately to find it in the huge pile of stuff which, as tired parents, you have yet to clean up after a late Sunday night arrival. You usually put it next to the front door in its designated box, but where on earth could it be this time?

Information extraction suffers from exactly the same problem: usually organised, sometimes not; high-value facts are usually stored in designated databases, sometimes not. However, 'usually' is not enough, especially in situations that require quick action and a comprehensive overview. What we need is a powerful assistant that provides us with this overview, and gives us quick access to the information we need.

The idea to develop SeTA, or Semantic Text **A**nalysis tool, came from brainstorming in Digitranscope on how we could use AI and related technologies to develop new forms of more responsive and citizen-centred policies. Policies that could "learn" from the feedback provided by the recipients of the policy intervention and that would be dynamic and flexible in achieving the objectives agreed at policy level.

These initial ideas (see Craglia, Hradec and Troussard, 2020) resonated with the very practical needs of colleagues in the JRC Competence Centre for Modelling who have responsibility for the policy impact assessment of European policies. In their work it is important to understand the relationships between different related policies and identify approaches, models, and data that might help assess the contribution of each policy intervention to the overall impact measured. Their needs provided an excellent use-case to apply AI methods to structure and extract knowledge from the body of legislation of the EU and develop an important building block of this concept of "policies-that-learn".



SeTA is now a web application that is accessible on the European Commission's network to provide support for its staff. It depends on a set of neural networks[28] that have been trained on the practically complete corpora of public European Commission documents (EURLEX, the EU Bookshop, etc.) since 1953. We have chosen English-language texts due to the sheer volume of available material and because the

[28] Word2vec and FastText word embedding are explained here: https://towardsdatascience.com/word-embedding-with-word2vec-and-fasttext-a209c1d3e12c

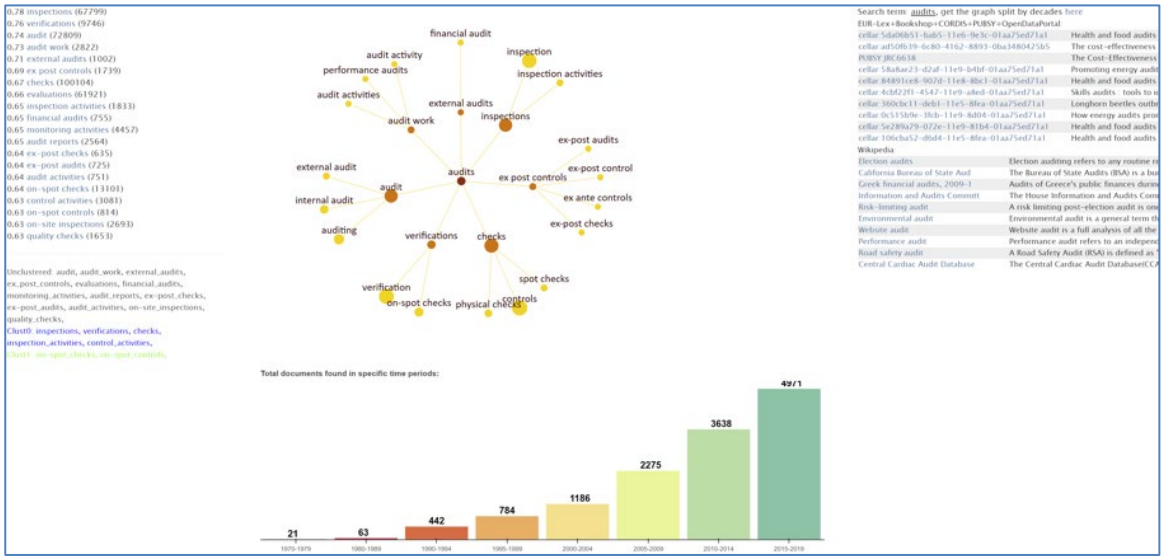
majority of important documents exist in an English version. The key advantage of SeTA – which is currently a prototype – is its ability to grasp the meaning of terms, and the changes in those meanings over time. Using this ability, it builds up a comprehensive ontology which makes it possible to carry out semantic searches in more than 500,000 Commission documents. A dedicated Technical Report contains a detailed description of the application, together with a host of examples of its application in real-life policy support scenarios (Hradec et al. 2019)

8.2 The technology

The English linguist J. R. Firth (1957) postulated that 'You shall know a word by the company it keeps'. SeTA is a vector space – imagine a cube – where the position of words and phrases determines their meanings and similarities. Thus, the most similar terms to the word *auditor* are *external auditor* and *internal auditor*. As we have trained the neural network using our policy-related document corpus, the most similar term to *eca* is *European Court of Auditors*. And while Google gives *hat*, *lid* or *limit* as synonyms for *cap*, it obviously means *Common Agricultural Policy* to us.

Our approach goes beyond mere term similarities. A well-known application of vector logic trained on general texts is *queen-woman+man,* which gives *king*. Although this does not work in our vector space as our texts are generally gender-neutral, *Water Framework Directive – water + waste* yields *Waste Framework Directive*. And if we know that *Water Framework Directive* is actually a *directive* and we are interested in *indicators* that are linked in the vector space, we get *water indicators, water quality indicators*, but also *spatial indicators* and *ecosystem service indicators.* Sounds like science fiction? No: just the application of a mathematical principle. We learn by learning to ask, and the learning curve can be pretty steep (see **Figure 8.1**).

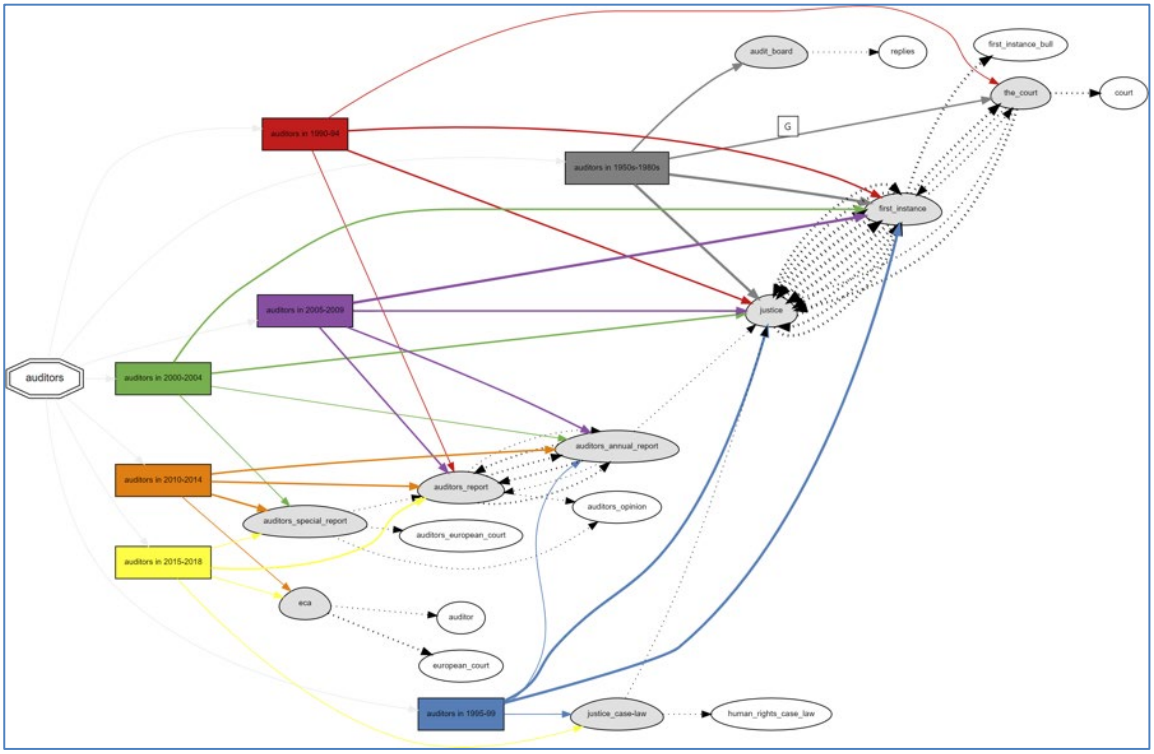
Figure 8.1 - On-the-fly generated semantic map surrounding the term 'audits'



Source: JRC

By training networks by half decades to capture the language specific to each new Commission, we also extract how the meaning of a term changes over time in a given policy context (see **Figure 8.2**).

Figure 8.2 Evolution of the term 'auditors' over decades, reflecting contextual change in policy documents and technical reports



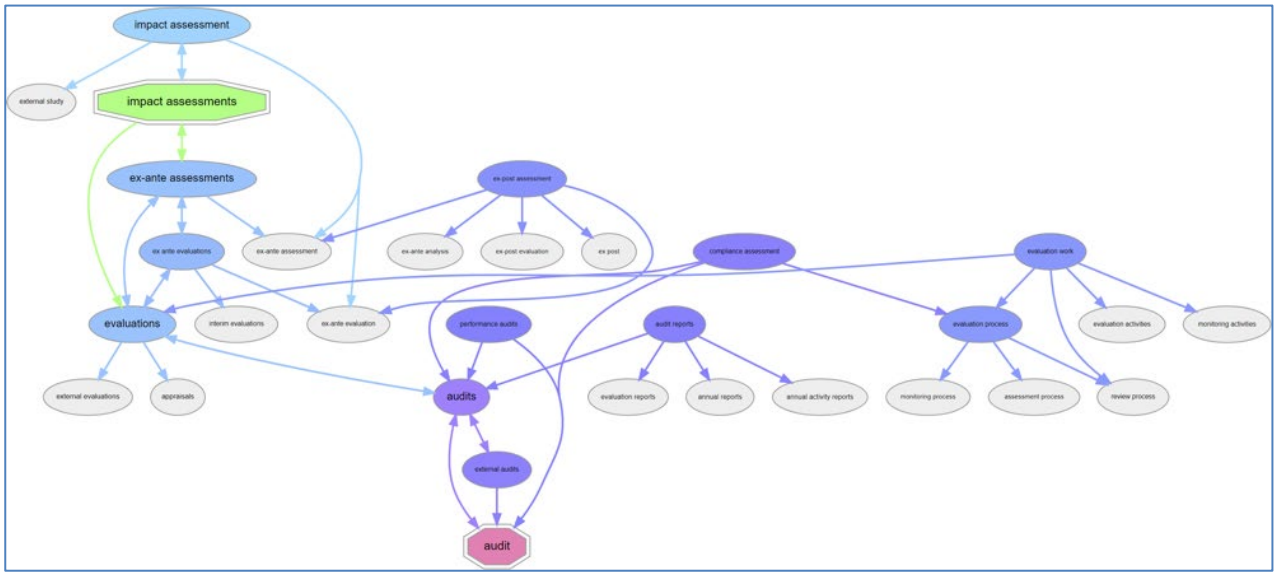
Source: JRC

SeTA's key users are Commission experts working in the field of policy impact assessments, where an ability to learn quickly about a new domain and all terms used within it (including slang[29]), plus fast access to relevant documents, are prerequisites for allowing experts to concentrate on generating added value, instead of performing mundane tasks. However, SeTA's potential user community is much broader: archivists, policy analysts, anyone responsible for checking a large document collection, or even the leader of a newly established cross-domain team trying to help people understand their interlocutors' slang would benefit from an instrument that explains the meaning behind Commission policy terms and their context.

One of the most significant side-effects of building SeTA was finding that a digital assistant can support domain experts by offering comprehensive coverage of their domains. Thus, SeTA provides on-the-fly ontology creation, where the expert's job is to say exactly where to stop. This approach also helps to explain and explore the relationships between different terms, as in Figure 3, which simply asks: *What do impact assessments have in common with audit?* In less than two seconds, SeTA finds similarities, components and variants directly from the vector space, with no need to search for anything in the texts. It also satisfies the frequently expressed desire for a more holistic viewpoint, as comparing terms from different domains will unearth relationships with which the reader may have been unfamiliar.

[29] According to Wikipedia, 'slang exists because we must come up with ways to define new experiences that have surfaced with time and modernity'. The meanings of words develop in different ways, depending on whether they are used by politicians, scientists, engineers or policy makers. However, context and frequency remain the same, so SeTA can tell that waste water treatment plant is very similar to municipal sewage treatment facility.

Figure 8.3 – Features common to impact assessments and audit



Source: JRC

8.3 A different approach to summarising text

There are several methods for extracting meaningful information from a large text without having to read everything. For instance, we can combine word embedding with the *TextRank* algorithm to extract the most important words or sentences from a document. In theory, we can train a deep neural network on a large corpus of text/abstracts to obtain a solid model for generating summaries from a document. However, we have taken a different approach that is best suited to the needs of policy analysts: claim extraction and fact checking.

Imagine being tasked with checking a draft impact assessment about population exposure to PM2.5. The search "PM2.5 exposure population in health risk" yields 220 results on the Publications Office website. About two weeks' reading time, you might think?

We have split our EU corpus into roughly 500 million sentences, identifying all phrases (e.g. *audit trail*) and mentions of date, location or institution, etc. The neural network predicted whether the *2000* in the sentence is a year, a quantity or part of a phrase (e.g. *Natura 2000*).

This approach allowed database queries such as [*eurovoc: health risk AND sentence:PM2.5 AND sentence:exposure AND entities:(PERCENT)]* that yields 18 sentences in 10 documents. We can refine the query results further by adding *children*, which returns one sentence from the final report of the JRC project SINPHONIE on school indoor pollution. The whole procedure took less than 30 seconds. Why *search* for information when you can simply *find* it?

Such an approach therefore helps us to construct a knowledge graph as the key to yet another and more important advance: automated fact checking. Natural language-processing algorithms can easily parse the sentence *GDP growth in Belgium in 2007 was 3.3%* into GDP growth [phrase], Belgium [location], 2007 [date] and 3.4% [quantity, percentage]. However, sentences are usually written by real people and, as paraphrased in the very first sentences of this article, are often too complex and difficult for an algorithm to comprehend: 'The Federal Statistical Office released figures for GDP growth and the general government deficit in 2005, at 0.9% and 3.3% of GDP, respectively'.[30] Where an analyst can get slightly confused for a second, the machine still fails miserably to understand how the numbers connect to the terms … so far. But we are working on it.

[30] CELEX:32006D0344

8.4 Stepping into the future

One of the main reasons we are working on knowledge extraction is automated fact checking. While the first sentence in the previous paragraph can easily be structured into a Eurostat database RESTful service query to obtain information that Belgian GDP growth in 2007 was actually 3.4%, the second sentence requires the attention of an expert.

Automating fact extraction and verification will help us in the very near future to build an AI assistant that will offer policy analysts the facts they need while they are actually writing their sentences. GDP growth in Ireland in 2015 was … And, *ping*, the computer automatically inserts the eye-watering figure of 25.6%.

However, more importantly, such information can be extracted to create a database of facts and claims. Computational models (e.g. for GDP forecasts) can benefit hugely from having comprehensive information for back-casting to improve models. Knowing what claims have previously been made about a topic would help avoid contradictions. Ex-post analysis will be much simplified if all facts and claims upon which policy formation is based are already available. New problems where information quality yields a totally different meaning will result from this process.

There are several limitations to our approach we know of:

- First, it is based on technology established in 2015-16, eons ago in AI development. Representation learning has evolved beyond simple word vectors but still has not redefined the purpose – understanding the word in context so analyst can learn ten times faster.
- Second, it has been trained only on the European corpus and only on English texts. Multilingual embedding was not available at the time of the SeTA design, a task still ahead of us. Without proper multilinguality, where Luftverschmutzung, znečištění ovzduší and air pollution are the very much the same thing.
- Third and most important aspect is how many use cases the SeTA covers. Originally trained to solve less than twenty problems, the new domains will significantly improve what all analysts can extract from the text.

SeTA has been trained and continuously improved in order to support analysts exactly where we think it will help them the most. Finding all facts, citations and relevant documents will reveal fundamental truths. We may not yet have reached a situation of on-the-fly data extraction from published news and Member State reports, but we are working on it. We may still have a low fact-extraction ratio, but 80% is better than nothing. PDF is a print format and a lack of information on text/sentence/paragraph flow means that extracting text from PDFs is a real headache. To compensate, we have also written de-hyphenators, spellcheckers and phrase formatters.

8.5 Conclusions

In this chapter we have highlighted some of the key features and ideas behind the development of SeTa. For more details see Hradc et al, 2019.

Extending our scope to national, regional and local levels of policy and legal documents will bring new insight into imbalances, inconsistencies and inequalities hiding in gaps introduced in all layers of administration when top down European legislation transposition process was meeting the bottom up national laws and even the practical detailed local bylaws.

SeTA's tools, even in their current state, either as a web application or a web service for information system integration, provide a major opportunity for alleviating the burden for the institution's *analysts/generalists*, who can understand the facts when they see them instead of reading reams of self-confirming text.

Our ultimate goal along the simulation of synthetic population described in the previous chapter is to create synthetic legislation that understands needs and human behaviour and calibrate intervention to respond to the stated needs. If then augmented with user feedback and interaction, it can really open up the possibilities of a new contract between the citizen and his/her community. In the words of artificial intelligence community, huge multi-domain policy shifts may help go beyond local optimum achievable by

legislation evolution but by finding global optimum, of fair resilient society. We are not there yet but experimenting and getting feedback is the way in which we can make progress.

References

Craglia, M, Hradec, J**.** and Troussard, X. (2020) "The Big Data ad Artificial Intelligence: Opportunities and Challenges to Modernise the Policy Cycle". In Soucha V. and Sienkewicz (Eds.) *Science and Policy Handbook* Elsevier. 2020. https://www.sciencedirect.com/book/9780128225967/science-for-policy-handbook.

Firth, J. "A Synopsis of Linguistic Theory 1930-1955," in *Studies in Linguistic Analysis*, Philological Society, Oxford. 1957.

Hradec J., N. Ostlaender, C. Macmillan, S. Acs, G. Listorti, R. Tomas, X. Arnes Novau, *Semantic Text Analysis Tool: SeTA*, EUR 29708 EN, Publications Office of the European Union, Luxembourg, 2019, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC116152/kjna29708enn_1.pdf

9 Digitranscope Experiments: Digital Twins and Smart Cities, case studies of Amsterdam and Duisburg

Corentin Kusters and Henk J. Scholten

9.1 Introduction

At the present time, around 80% of the population in Europe and North America lives in cities. Asia and Africa are estimated to host 2/3 of their population in urban areas by 2050. Those figures are expected to increase in the next 30 years with a population growth rate reaching 5% a year or more in certain urban agglomerations around the globe. In that perspective, there are growing concerns on how to accommodate such a growing population and maintain a sustainable environment and a great quality of life in urban areas.

In this context, cities' officials are facing an increasingly complex environment. Preoccupation is now towards the design and operation of sustainable cities, cities that combine environmental, social and economic balance and it can be hard for decision makers to respond to such requirements in an effective way. We must provide them with the necessary tools that will help them apprehend this complex environment and build efficient processes.

Smart City projects are at the crossroad of those imperatives using digital solutions and advanced analytics to tackle the growing concerns mentioned above. It is believed to provide a greater understanding of a city's dynamic and help decision-makers in their tasks in various domains. Smart city projects leverage the growing pool of ICTs and data to benefit inclusive, sustainable and liveable cities.

In this study, we investigated the potential of smart city projects and digital twins to address issues on mobility, liveability and energy. Two case studies form the base for this research:

- the Amsterdam case study
- the Duisburg case study

Both case studies explore key aspects of digital twining and of the smart city movement at large with the inclusion of technologies such as the Internet of Things (IoT), Big Data, Artificial Intelligence (AI) and Machine Learning. Beyond technological implementation, we also aim to demonstrate the importance of open standards for the development of flexible and interoperable smart cities.

The experience of these case studies has provided Digitranscope with practical insights into the complex relationships between data, technology, social and economic problems in urban environments and stakeholders perspectives. It has allowed us to see the opportunities of the digital transformation but also its limitations and complemented insights from the activities of the project described in the other chapters.

9.2 Amsterdam case study

9.2.1 Problem Statement

Amsterdam's population is growing and is estimated to host nearly 100 000 additional citizens in the next 10 years. Beyond its population, the city attracted approximately 19 million tourists in 2018, roughly 3 million more compared to 2017. In this context, the city is facing challenges as to accommodate such a dense population, especially in terms of mobility. Indeed, the transportation network is more and more pressured resulting in traffic congestion, public transport delays, safety issues and more. In addition, in 2018, the Dutch government has initiated the National Climate Agreement which aims at reducing greenhouse gas emissions in the Netherlands by 49% by 2030 compared to 1990 levels. It also includes an action plan to facilitate the energy transition as much as possible. This policy considers a regional approach where public authorities such as the Amsterdam Metropolitan Region have a big role to play.

Facing those challenges, the Municipality of Amsterdam has chosen to invest in the implementation of digital twins and smart city dashboard. The objective here is to develop a platform that convey relevant insights for assets and services management. Indeed, beyond infrastructure investment, a greater understanding and management of the city is believed to a have a positive impact on GHG mitigation, mobility issues and energy consumption.

Figure 9.1. Intelligent Dashboard Amsterdam: Mobility view



Source: authors

9.2.2 Technological readiness

For a bit more than a decade, the city of Amsterdam has well integrated the smart city vision in its development agenda. The municipality has deployed efforts to create an inclusive, communities-centered city where citizens, learning institutions and private partners can actively participate in the city's digital transformation. Indeed, the 2016 Europe's Capital of Innovation has created structures such as the Amsterdam Smart City that facilitates the transition toward a data-driven city. Those efforts resulted in well-ground infrastructures and protocol for data collection and storage. In the Netherlands we can cite for instance the Nationale Databank Wegverkeersgegevens (NDW), a consortium of 19 authorities that work together to collect, store and distribute road traffic data. The technological readiness of the city of Amsterdam greatly contributed to the development of smart city projects, digital twin and dashboards.

9.2.3 Implementation

For this case study, we collaborated with the city of Amsterdam and the Johan Cruyff ArenA stadium, and used the Arena and its surrounding as a living laboratory. With hundreds of thousands of visitors each year, this entertainment, shopping and business area perfectly fit the scope of the smart city. After consultation with the different actors, stakeholders and also citizens, three core themes have been selected: Smart Mobility, Smart Energy and Sustainable environment.

Data has been collected from diverse data providers (both public and private) in and around the Arena. Information on traffic flow, parking occupancy, public transports as well as solar energy production, wind turbines energy production, building energy labels, air quality and pollutant concentration are collected, cleaned and analysed. Following the requirements and based on the datasets available, key performance indicators (KPIs) have been designed with the objective of actionable insights (that we can act upon).

That information is then integrated within a digital twin of the city of Amsterdam. This faithful representation of Amsterdam, in 3D, has been made possible by the use of open BIM and GIS standards

that aimed at digitalising building and city environment. Beyond cosmetic, this 3D model is now enhanced with the information about the city and its mechanisms.

The digital twin is then displayed in a dashboard, presenting the different KPIs and features in a user-friendly manner. This 3D interface enables an intuitive navigation and helps decision makers in taking effective and immediate actions.

Note that aside the formal requirements, room has been made for research and innovation and the possibility to develop AI solutions, new KPI's, improved 3D environment. For instance, machine learning algorithms have been developed to predict traffic flow, games engines have been used to develop virtual reality interfaces and efforts have been made to integrate point cloud data in the system.

Finally, this software solution is supported by cloud-based technologies that allows us to store a large quantity of data and perform analytics tasks in an efficient way. We developed GOST, an IoT platform that can be deployed with IoT devices, data and applications on the Web. GOST is an open-source, certified, OGC standard, that allows the storage and processing of the real-time data and makes them available to parties.

9.2.4 Partners

In this project, we partnered with the City of Amsterdam and in the Johan Cruyff ArenA for the development of our solution. This type of public and private partnership is set to be the basis of any smart city projects. Indeed, public organisations are in the best position to lead the smart city movement but cannot do it on their own. They need strong private partners with a good expertise on ICT and IT infrastructure to fill the gaps. In the frame of our project, our cooperation has been built on the basis of a common desire to develop impactful solutions that will serve citizens. Beyond simple client-provider agreements, we jointly designed solutions. This calls for stronger inter-organisations ties with a great emphasis on communication and accessibility. For example, we worked following an Agile methodology which encourages regular communication over the tasks and requirements and re-enforces cooperation between partners.

9.2.5 Citizens

In 2017, the municipality of Amsterdam invited strategic stakeholders, including residents and associations to contribute to the elaboration of the vision and ambition plan for the Amstel III/ArenAPoort area. Thanks to this participative approach, the Johan Cruijff ArenA has seized this opportunity to address the need for a holistic transformation of the area with an improved inclusion of socio-economic, environmental and digital requirements. It has resulted in the creation of a coherent and inclusive development plan that surely grants the project its "smartness". Moreover, in an effort to bring digital twins to the public, Geodan has launched in close cooperation with the Vrije Universiteit and the JRC "Ecocraft" (see chapter 10): a computer game with an educational angle, based on the popular Minecraft. The virtual world in Ecocraft is equal to the digital representation of Amsterdam. Such tools allow the youngest generation to familiarize themselves with the digital representation through an interface and game that they know and love. And perhaps more importantly, through play they can participate in designing their future neighbourhood and community. This tool has been used to digitalise the entire Netherlands and presented to citizens in events such as the WeMakeTheCity festival for greater outreach of the work being made in our communities.

9.3 Duisburg case study

9.3.1 Problem Statement

Duisburg has one of the biggest inland harbours in Europe and a direct connection via train and road to Eindhoven and Rotterdam. Furthermore, the city is part of the New Silk Road, an International Freight railway that connects China and Europe. The city being an important hub for freight transport in Europe, it is facing congestion of the ring roads due to heavy traffic in and out the harbour, especially during peak hours. Moreover, Duisburg is concerned by the effect that the heavy load of lorry vehicles have on the environment, both in terms of GHG emission and noise.

9.3.2 Technological readiness

In 2018, the city has initiated a masterplan for digitalisation with the aim to become a leading smart city in Germany and Europe. Duisburg won the silver medal, in 2019, for "Best digitalization project in cities/ regions" with the Smart City concept. Following this masterplan, the city of Duisburg has launched an IoT platform based on RhineCloud, a flexible cloud infrastructure, and has extended its Wifi and soon 5G accessibility across the city. Moreover, efforts have been deployed for the creation of a digital model of the city via the collection of point cloud data taken by Lidar technology.

Beyond infrastructures and the ongoing effort to collect more information, Duisburg has made its datasets available through an Open Data platform, the Smart City Duisburg open Data that aims to ease access to relevant information, invite citizens and organisations to participate and fuel innovation.

The technological and organisational readiness of Duisburg grounded a solid basis for the development of smart city projects and digital twining.

9.3.3 Implementation

The first step of this case study was to engage conversation with the different stakeholders in order to prioritise some key aspects that needed to be addressed. Due to its important harbour, mobility is the prime challenge Duisburg is facing. The number of lorries transiting through the city is important which can cause strong disturbance in traffic. Therefore, the prime aspects to be considered were toward traffic flows, parking' availability, public transport and EV charging stations. Along with mobility, the municipality was also interested in investigating the environmental impact of such traffic. Consequently, some environmental conditions such as the temperature, air quality or noise are equally part of the project's specifications.

Following the requirements, we have investigated the different datasets available that could help us in the design of our key performance indicators. Live data on traffic flow, EV stations availability, parking spot availability as well as environmental recording on air quality, noise and temperature have been included in our digital solution. Point cloud data and 3D generated GIS models have also been used for the design of our digital twin's 3D environment.

In this project, we have taken advantage of the RhineCloud to host our IoT solution. Disperse data are collected and centralised into this platform where they can be transformed and analysed by our services.

Innovation has taken a great place in this project. Indeed, cutting edge machine learning algorithms have been developed in order to predict traffic flow a day-ahead. Additionally, new KPIs such as the congestion index (the % of congestion over the city road network in km/km) has been implemented. Those additional features fit perfectly the scope of the study and are believed to enhance the understanding on mobility and help in making more efficient decisions. This is a great example of how open data can fuel innovation and how we can design relevant data analytics from such projects.

Finally, the whole has been integrated within a dashboard (Figure 9.2) that enables a user-friendly experience and an instinctive understanding of the KPIs. The gathering of information that was so far distributed gives a global view of the issue of mobility and greatly helps stakeholders in making effective decisions. For instance, agents can track congestion in real-time and even predict upcoming congested road segments. They can therefore deploy efforts into redirecting some motorists in alternative paths. More, with environmental data at hand, they can spot in an effective manner (both temporally and spatially) the effect of those measures on GHG emission and/or noise. In our perspective, a digital twin is

not solely a visualisation tool but is seen as an information “bank”. Its smartness is therefore bound to the information it conveys to the stakeholders.

Figure 9.2 Smart City Platform Duisburg



Source: Authors

9.3.4 Citizens

Before engaging in infrastructure development and project definition, the first step has been for the city and its partners to gather ideas. This has been done through a series of workshops with citizens, public partners, universities and firms. The workshops have resulted in the creation of 271 ideas from which the 18 most relevant have been selected as core requirements in the project. In an effort of continuous innovation, an online platform has equally been deployed where everyone can send new ideas. Those are continuously evaluated to integrate the scope of the smart city project.

Beyond consulting citizens, Duisburg has partnered with the University of Duisburg-Essen to create a Smart City program to train current and future administrative workforce to the digitalisation of our environment. Indeed, training workforce on the new digital technologies is essential for their penetration within public organisations.

9.4 Findings

The use cases of Amsterdam and Duisburg have grounded a number of good practices that we think are essential in the good implementation of smart city projects. Hence, our findings will come in the form of recommendations:

- ***Beyond technology, smart is about people***

The definition of smartness in the city is inevitably linked to one’s understanding of the city needs. On that topic, Amsterdam is cited as a good example to follow:

“Without the engagement of stakeholders, a city can never be Smart, no matter how much ICT shapes its data … The starting point of [Amsterdam Smart City] is not the (technical) solutions, but the collaboration, co-creation, and partnering of stakeholders within the city of Amsterdam”

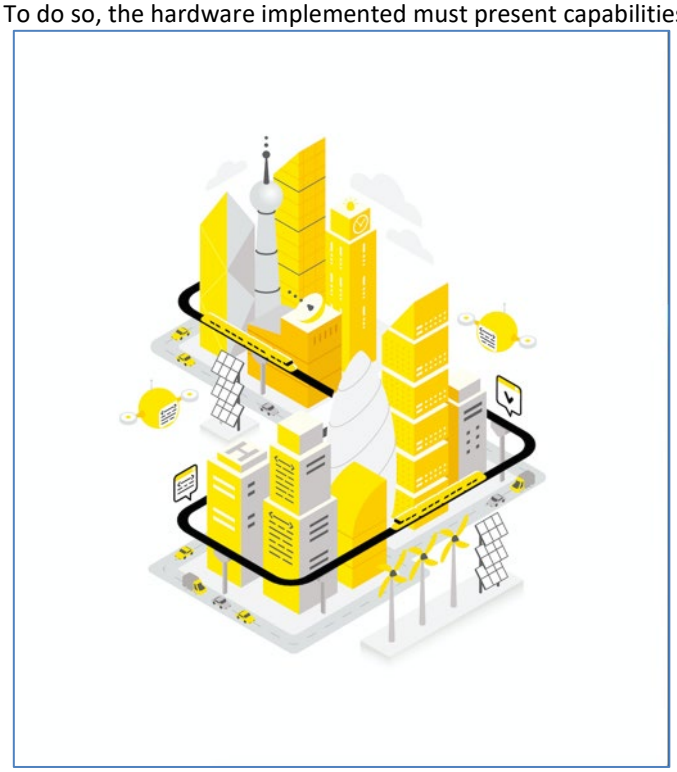
Consequently, the “smartness” of a project is directly related to the level and nature of participation from people. One must offer the possibility to all parties to actively participate in its smart city project development, inviting citizens, business owners, research organisations to collaborate. This can be done

for instance via the deployment of urban platforms and/or schemes that will trigger participation and co-creation.

- ***Think ubiquity and seek infrastructures ecosystem***

Data collection, transmission, storage and processing are the core components of the IoT. However, those parts are rarely handled by a single organization. The IoT covers a large range of industries and is applied to cases of various natures, scales and capabilities. Trying to merge all those different actors and technologies as one interoperable system is certainly the biggest challenge that the industry is currently facing. Indeed, a crucial characteristic of the smart city paradigm is its ubiquity, the creation of a single homogenous system.

To do so, the hardware implemented must present capabilities to easily operate with other systems. This is also important for scalability as it eases the progressive integration of new components and services. This is equally true for software solutions and data. The adoption of open standards enables flexibility and keeps solutions vendor-independent. Adopting common standards and protocols at every infrastructures' levels that will ensure a seamless ecosystem.



In addition, consortiums and alliances have emerged intending to unify the IoT industry landscape as well as software domains such as GIS, BIM, big data, machine learning etc.

Moreover, smart city projects certainly benefit from standardized technologies as it prevents monopoly and unlocks the market capability, leaving the door open to smaller players and entrepreneurs. It is an efficient way to break limitations and fuel innovation.

- ***Innovative Data Analysis from the cloud***

The requirements and scale of a smart city project call for the adoption of cloud-based solutions. For instance, the SensorThings API is an open-source, certified, OGC standard, that allows the storage and processing of the real-time IoT data and makes it available to parties over the web.

Data integration is an important step to unify heterogeneous sources into a single, homogeneous system. A successful smart city project should be able to scale up and progressively integrate new features and data sources. KPIs should be designed based on the existing IT infrastructure, available documentation and gathering all relevant data. A strong emphasis is on the notion of actionable insights which serves as a guideline when it comes to developing valuable data analysis. In such a perspective, data format and metadata must follow standardized frameworks to deploy an interoperable digital environment.

Finally, research and innovation should not be undermined and opportunities should be given to developers to innovate upon the data whether internally or from external sources. For instance, a research group could be assigned with tasks such as developing AI solutions, new KPI's, improved 3D environment and more.

- ***Care for visualization***

One should not neglect the value of dashboard design. It is essential to present the end-user a user-friendly interface and to invest in front-end development. The dashboard should be interactive and

straightforward with an easily understandable display of KPI's. Besides, a dashboard can benefit from 3D digital twins of the city for enhanced user experience.

- ***Be transparent and ensure privacy***

Organisations that process data must prove they are accountable. To do so, full transparency is a key element. When seeking personal data, consent must be freely given, specific, informed and unambiguous which means that it is the organization's accountability to clearly specify the use made of personal data and all parties involved.

Measures should be taken to preserve the privacy of personal data. Pseudonymization and encryption, regularly testing, assessing and evaluating IT systems as well as organisational systems, and following strict guidelines from the GDPR is a must.

Finally, when developing a smart city project, one must keep in sight the prime reasons for such development which are to lessen our environmental impact, reach social harmony, help decision-makers, and help citizens in living a good life. To do so, it is central to favour citizen sovereignty over data, data openness, transparency, privacy protection and to mitigate data monetization and technological lock-in.

- ***Build strong partnerships***

PPPs must follow the smart city vision with long term, coherent strategies and collaboration plans. A first step in the creation of such a collaborative perspective would be simply to re-considered the status of each party. We must step away from the traditional client-supplier schema toward a more "organic" approach where every party is considered as a partner that pro-actively serves common objectives.

The different parties must jointly participate in setting clear objectives, schedules, communication protocols, arrangements for data management and sharing within partners, knowledge transfer, risk and revenue sharing and policies compliance.

- ***Involve citizens***

Lastly, a great number of professionals in the field, academics as well as business leaders have stressed the importance to integrate citizens within the loop. The city of Amsterdam, for instance, has promoted inclusive development, inviting citizens to have an active role in data collection, strategies definition and advisory inputs. Same goes for the city of Barcelona that has put citizens at the core of the decision-making process and by giving citizens sovereignty over the data produced. Overall, initiatives to engage citizens in digital life making should flourish. It is a crucial element to build democratic smart cities.

The practice-oriented recommendations above provide useful elements of reflection together with the more theory-informed findings identified in Chapters 3 and 4. They add to the variety and richness of the Digitranscope project and will be discussed further in the concluding chapter of this book.

10 Involving children in the renewable energy transition: exploring the potential of the digital transformation

Jaap Boter, Magdalena Grus, Sanne Hettinga, Henk Scholten



10.1 Involving children in the energy transition

In the European Union, the built environment is responsible for 40% of the energy consumption (IEA, 2017). To reach internationally agreed climate goals, we not only need energy saving measures for individual buildings, but also must address the issues at a city or neighbourhood scale. At this level, including all stakeholders in the planning process, creates more support for implementation of renewable energy technologies and minimizes NIMBY effects. There are several strategies to engage these different stakeholders in spatial planning, including Geodesign (Steinitz, 2012) and Public Participation GIS (PPGIS) (Brown & Kyttä, 2014; Reed et al., 2018; Sieber, 2006), that allow large groups of stakeholders to be involved.

A digitally transformed European society offers new opportunities to fuel many of the steps of a participatory process. Not only are vast amounts of data more readily available, but the digital transformation also offers new possibilities for integrating and representing big data. Rather than a large, two-dimensional map with abstract symbology, now a virtual representation of the real world can be created, based on accurate geospatial data: a *digital twin* of reality. Herein, one can visualize and model all kinds of processes as well as relations between these different processes, as well as the impact of future scenarios, facilitating multi stakeholder collaboration and citizen participation.

Several authors mention the inclusion of children as one of stakeholders in cases (e.g., Drazkiewicz et al., 2015; Eiter & Lange Vik, 2015). Children are advocated as ‘adults of tomorrow’, who need to learn to take responsibility for their living environment; or even as ‘full citizens of today’, with the right to be heard (Levine, 2015). With a different perspective on their local environment, their involvement may also lead to improved decision making and more innovative solutions. Finally, recent activism on a global scale, such as climate protests initiated by children like Greta Thunberg, suggest a great need to build trust from children. As a non-negligible side effect, children’s’ involvement in participatory planning may indirectly also engage others, such as parents, family, and the wider public.

Involving children in a public participatory process, however, is not self-evident, as many of the principles in the participatory planning literature have been developed in the context of adult decision making. Stakeholder engagement, one of the primary drivers in the process, for instance, is likely an important differentiator; with children requiring (quantitative) information to be presented in a more accessible and entertaining format to capture engagement. Like in regular participatory planning, representativeness is also an issue with children, if not more so; with a voluntary approach potentially leading to self-selection. Minors can more easily be reached through certain channels such as schools, but this does require specific consideration of time and group dynamics (Buchy & Hoverman, 2000) of the context, for example, in the form of a lesson plan.

Given how digital twins resemble the idea of contemporary computer game environments, the question is whether a digital twin in computer game format can be a particularly accessible way for children to participate. To test this idea, two digital twins of European neighbourhoods, one in Warsaw and one in Amsterdam, have been developed by local parties utilizing the environment of the popular computer game Minecraft™. In Minecraft, users are supplied with a large array of different types of building blocks, to create houses, gardens, waterworks, and other planning elements. Converting real-world land use data and 3D point cloud data, an empty Minecraft world can be populated with a representation of the real (built) world, referred to as 'Geocraft' by Scholten et al. (2017). Children can then use this familiar environment to generate data driven future scenarios by simulation and visualization of future developments and their implications. As the two experiments focused on the energy transition, the environment was renamed as 'Ecocraft' by the organisers. The two experiments have somewhat different aims, use different strategies in execution, and rely on different tooling; allowing to explore the effects of context, aim and supporting activities.

10.2 The Warsaw Experiment

In Warsaw, smog, as side effect of household heating, is a pressing issue. In the old Praga-Północ district of Warsaw, on the eastern side of the Vistula river, for instance, a relatively high number of houses use coal as source of heat and especially in the winter it has a large impact on air quality in the city. Awareness about renewable sources of energy as cleaner alternatives is still limited, certainly among the middle-aged group of citizens. The aim of the campaign "Ciepło dla Pragi" ("Heat for Praga") is to inform and engage all residents of the Warsaw Praga-Północ district who want to actively work on their neighbourhood and who are not indifferent to issues related to environmental protection and energy saving. The project wants to stimulate creative thinking about a "dream district", with sustainable energy, green areas, and friendly environment. While trying to reach all residents, the organisers see children as vital stakeholders. The organizers particularly want to create a sense of engagement.

The Warsaw experiment had a strong "campaign" stance. Technical tooling and infrastructure were kept at a basic level, with most of the effort put into a large competition, outside the classroom, with prize money available. Technically, participants had access to a basic three dimensional Minecraft model of the buildings, streets, and green objects in Praga district. Information was confined to parameters such as height and size of the buildings, basis infrastructure and available greenery. Unlike the Amsterdam experiment, information on current energy use was not available in the game.

The experiment was run as an out-of-school competition that ran over an eight-month period, starting in August 2018 and ending mid March 2019. Parents and/or teachers were approached to submit participation forms for their children or pupils. Then, in November, participants were prepared for the challenge, focusing on both background knowledge on energy and environmental themes, as well as the Ecocraft environment. The organization had setup several support activities to achieve this, such as educational materials, a mini-audit to be carried out on the own household, workshops, and online meetings with experts to get help.

Using the Ecocraft environment, they then designed and implemented their ideas for a residents-friendly and energy effective area for a chosen part of the district. Additionally, teams had to prepare a final presentation of the project in the form of short movie, with the authors explaining the proposed idea and changes in the district. A jury reviewed the projects and chose ten finalists who presented their projects during a "Ciepło dla Pragi" Open Day (March 9). The winning teams received a money prize to implement some of their proposals. During the Open Day event, also older generations of inhabitants were invited to

discuss issues concerning electric efficiency with the experts and to get advice on energy efficiency at home.

Teachers, volunteers, and local experts worked with 80 school children in total, divided up into 18 teams from five elementary schools. There were several items of the project in the local newspaper "Mieszkaniec" as well as other Warsaw newspapers; estimated total newspapers circulation is about 40,000. The campaign also had a dedicated website with information about the competition as well as on energy efficiency in general (www.cieplodlapragi.pl), which has been visited by 992 unique users to date. The final event saw more than 200 visitors.

In line with the 'campaign' stance taken in the Warsaw experiment, the key insights generated from the experiment all centre around 'engagement':

1. Choice of problem area. The organizers observe that part of the success depends on the familiarity and attachment participants and other stakeholders already have with the project area, whether familiarity with its physical appearance, or emotional attachment. Attachment to the area helps to increase the engagement. It thus may also help to focus on a relatively small area, that participants know well.
2. Involving other stakeholders early on. While children are an eventual link to reach older generations, such as parents and grandparents, these other generations could already be involved in some way early on in the project, before a final presentation of the results. In a next project in Poland, for instance, the organization is looking for the possibility to create family groups (with parents, grandparents etc), rather than just the children as participants. If anything, parents' involvement from the beginning of the project seems highly beneficial.
3. Real problems, real agency. It also helps if children are working on real problems of their district and see the follow-up of their work, such as implementation of the proposed ideas/solutions. This requires upfront commitment of authorities for change. Here, UNEP/GRID-Warsaw, together with Veolia as well as experts from the Praga-Północ district office and the authors of the winning projects, will choose, finance and implement elements of the proposed by children ideas. One of the ideas presented that will be implemented is the facade of the building growth by plants; to be realized on the one of the schools of the participants.

On a more practical note, the organizers observe that using a digital twin ideally is supported by good, 24/7 IT support, with dedicated people and control and enough workshops. Also, while teachers don't have to be eco-experts, it would be good to have very detailed online tutorials available.

10.3 The Amsterdam Experiment

Like many large European capitals, Amsterdam has a high diversity in income, ethnicity, educational level, and quality of buildings. This is particularly true for the district of Amsterdam South-East; a rather complex district with a relatively large proportion of flats, undergoing structural redesign. The City of Amsterdam and local social housing corporations (owning about half of the housing in the city) have made serious commitments to more energy efficient housing, and to increase clean and renewable production of energy consumed by households. Any change, particularly for high-rise buildings will have to be done as larger project per block. Also, Dutch legislation requires a minimum of 70% of the tenants approving any change that increases monthly rent. So here, a successful strategy to improve housing in energy efficiency and renewable production of energy must rely on soliciting grassroots support as well as careful selection of building blocks, based on metrics like costs and gains.

In contrast to the campaign stance of the Warsaw experiment and open enrolment, here a specific group of school children was chosen in a class setting; with a strong focus on education on how energy efficiency and renewable energy work, and the use of an extended version of Ecocraft that allows for calculation of real metrics for each real building. While the Warsaw experiment thus had a limited technical design and was elaborate in the procedure, here the procedure is more straight-forward (part of a lesson plan), with the technical design and governance as more important. Key question here is whether these children are not just engaged, but also become a valuable source for co-decision making, thus allowing for participatory decision-making and generation of support despite the technical nature.

In Amsterdam, the participants were provided with instant quantitative feedback on their ideas through a plugin. The plugin, an addition to the virtual environment, was developed on the foundation of the Geocraft environment of Scholten et al. (2017) and deals with the energy situation *of a specific (real) location*. As there is a one-to-one relationship between this virtual world and the real world, clicking on a virtual house allows for identification of the real house and model any choices on the real situation; such as the angle of the (real) roof top and the subsequent (real) efficiency of a solar panel. This was incorporated by linking to another spatial data infrastructure described by Hettinga et al. (2018). The modelling capabilities described in that paper are used, as well as data on for instance energy labels, solar potential and energy consumption. To make the material more suitable to work with children, the insulation step was simplified to include individual insulation measures – floor, roof, wall, and window. Additionally, a dashboard keeping track of the total investment cost, electricity produced, gas consumption avoided, and CO2 emissions avoided are displayed, as well as a small dashboard that appears whenever a player alters a building, showing the costs and benefits of that individual change.

The children involved in this project were in the second year of the Dutch secondary school system and about 14 years old. The project was embedded in geography classes, with children of a highly diverse educational level, as well as from the diverse backgrounds that the district South-East is known for. The teachers provided a booklet with information and assignments to the children and gave time in class to work through. They were also available to answer questions and to provide help when needed. A total of 31 teams of three children each worked on designing plans in Ecocraft. The children were first given time to come up with a strategy to make their plan. Afterwards, they were provided with 4 lessons to construct it in the game, supervised by the teacher.

Key question here was whether the children are a valuable source for co-decision making. Thus, allowing for participatory decision making and generation of support, despite the technical nature. The conclusion of this project is, therefore, not an open day, but an analysis of the results. Table 10.1 shows the descriptive statistics of the different results per technology choice for the groups. The differentiation presented in this Table implies that the children had their own individual visions and ideas for the area and use different approaches. There are, for instance, groups that have only placed solar panels and have not applied any insulation. Furthermore, there is a group that made changes to a total of 521 buildings within the allotted time, while another group has only selected eight buildings.

Table 10.1: Overview of the spread in different indicators between the groups participating in the public participatory planning session.

Indicator	**Mean**	**St. Dev.**	**Min.**	**Max.**
Number of solar panels	1,400	1,570	38	7,612
Buildings with solar panels	62	63	6	270
Average number of solar panels per building	28	30	2	143
Buildings insulated	71	87	0	344
Buildings addressed	132	135	8	521
% buildings addressed with solar panels	47	27	19	100
% buildings addressed with insulation	53	27	0	81

Table 10.2 shows several indicators of the plan for solar panels, for insulation and for a summary of all technologies. It shows that the investment cost of solar panels is 6.5 million euros and replaces 3.2 million kWh of fossil electricity with electricity from solar panels. The locations for solar panels the children have selected have an average payback period of 5.3 years. The investment in insulation is significantly higher, and mainly has a relatively high payback period compared to the payback period of solar panels: 15.3 years. However, when analysing the payback period of insulation for the buildings in Amsterdam South-East, it has indeed an average payback period of 15 years. This indicates that the children have not made irrational or random choices, but that the payback period of insulation is significantly higher than the payback period of solar panels. The fact that the children have also selected insulation as a technology

despite its worse economic performance, indicates the added value the children see in this technology, for instance the added benefit of comfort or noise reduction, as was explained in the lessons the children received before they obtained access to the environment.

Table10.2: Impact analysis of the overlaid plan of the children

	Investment cost (€)	Reduction in gas consumption (m3)	Electricity production (kWh)	Reduction in CO2 emission (kg CO2-eq)	Payback period (years)
Solar panels	6,512,100	-	6,423,219	3,211,609	5.3
All insulation	81,908,719	8,878,861	-	17,580,145	15.3
All technologies	88,420,819	8,878,861	6,423,219	20,791,754	12.6

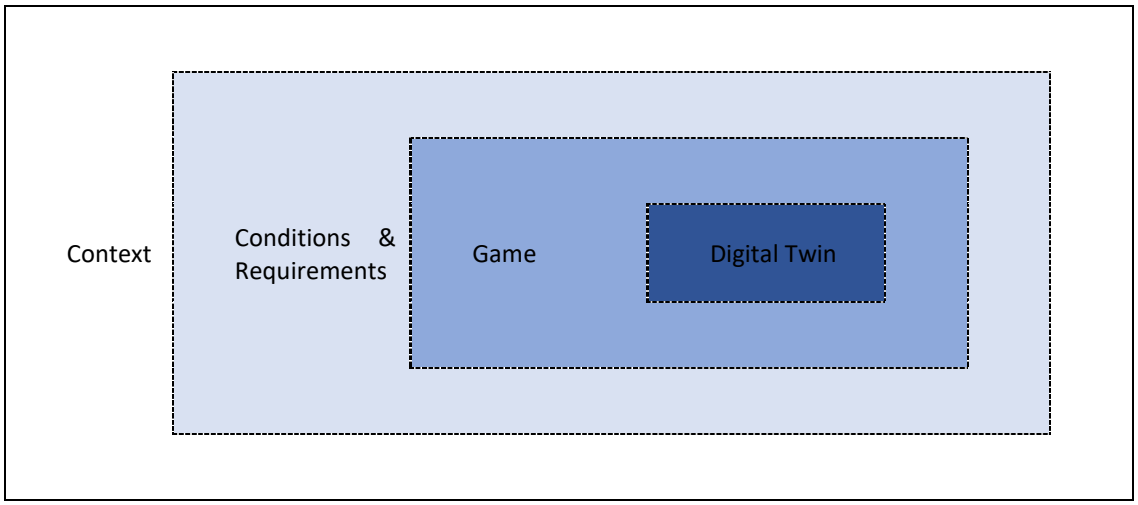
The impact analysis shows that the children can make reasonable choices for the area around their school, which they are familiar with. The payback periods of the buildings and technologies the children selected are in line with the average payback periods of the district. Additionally, they have chosen buildings specifically for certain technologies, showing that this was a well-considered choice.

With the Amsterdam approach a wide range of children of different background, learning style and learning level was reached, enabling children of all backgrounds and levels to participate. Consequently, this case included children that have never been included in the public participation planning process, but also those that never wanted to be a part of it or have no affinity with the topic of the energy transition. It was found in this study that the Ecocraft environment was usable by all children included in the pilot. First, the lesson plan provided enough knowledge and skills to teach them about climate change and the renewable energy transition to participate in the planning process. Second, the decision support system in the familiar gaming environment with the gaming element managed to engage all children to share at least some of their ideas and plans. The differentiation between the plans in terms of technologies applied, buildings selected, etc. indicate that they designed the plans by their own insights.

10.4 Using gaming in digital twin setting: opportunities and limitations

In line with the forward-looking nature of the DigiTranScope project, the two experiments were not aimed at empirically proving literature-based hypotheses from well-established streams of research; but rather attempting to explore the possibilities of new technologies and fields, and the new questions they raise. Given how relatively new gaming and digital twinning are as objects of academic interest, the exploratory nature is well-justified, but then as qualitative research may also benefit from post-hoc reflection on how this can be understood and what lessons can be drawn.

Figure 10.1. Context, conditions, game and digital twin as hierarchical set of elements



Source: Authors

As visualised in Figure 10.1, the setup of both experiments may be seen as an interplay between the four elements of context, conditions, game, and the digital twin. While definitions of games are plenty and evolving, core elements often include, for instance, tools or tokens, rules and aims, players, or a story. The digital twin is, like a chess board, only the platform or board on which the game takes place, with tokens. The rules and aims, the type of players, and the story – the other elements of a game – determine how the digital twin is used. Effective play of the game – certainly with concrete aims like creating engagement or soliciting consultation – is dependent on a range (practical) conditions and requirements; ranging from the availability of data to build the local digital twin; the availability of computers to access the digital twin as board; the availability of time in or outside school to play; the capabilities of the umpire or coach (staff); or the knowledge and skills of the players. Finally, the project is dependent on a local context of the theme, the political and urgency of the theme, or whether there is a culture of public participation; all driving the motivation of players, design of the game, and quality of conditions. Note that these elements are:

Hierarchically embedded, such that the digital twin is a part of the game; the game is dependent on several conditions and requirements to be successful; and these all operate in the context of a particular theme, in a particular city, in a particular country and culture.

Mutually dependent on each other; they influence each other in both directions and/or can compensate for each other; as symbolised by the dashed line. E.g., an advanced digital twin with interactive, quantitative data may allow for a game design with point scoring and rewards. But a lack of such features in the digital twin may be compensated by other game elements, such as rounds with expert feedback.

Since Arnstein (1969), different levels of public participation have commonly been distinguished, such as 'awareness', 'engagement', 'consultation', and 'partnership (plus power, control)'; with the latter implying more participation than the first. On the negative side, low levels of public participation have been labelled as either 'manipulation', 'decoration', or 'tokenism'. The intended level of public participation will drive the design of the project and hence the role of the digital twin and use of game elements surrounding it. The role of the digital twin in a public participation project may be complementary to the other game elements:

Table 10.3. Possible contribution of game elements vs digital twin at different levels of public participation

Participation level	**Possible game elements**	**Possible features digital twin**
Awareness	Rewards and punishment	Visualisation of the issue
Engagement	Competition	Realism/familiarity
Consultation	Field trip	Interaction
Partnership, Power & Control	Community	Reporting, dashboard
Manipulation, etc	Rules	Impression management

Compared to a traditional public participation (GIS based) setup, here, the map has been replaced by a high-tech, 3D representation with much more possibilities for input (building blocks) and (in the case of Amsterdam) direct feedback. The process, traditionally a group discussing around a map, has been replaced by a game environment, where individuals or teams mark proposed changes; or even alter and (re)design the environment. In the process, the digital twin may contribute in visualising the issue; engage more in its realism; allow for interaction like a multiplayer online game; and serve as a presentation platform of a design or ongoing dashboard in a monitoring situation.

But as the experiments show, a vital part of the process also happens outside the digital twin. Setting rewards or punishments ('rules of the game') will steer particular participant behaviour and give feedback on right or wrong; judges and prizes will fuel participant behaviour; consultation may happen as field trips and interaction in real life; and ongoing partnership may be established in community forms, such as meetings, newsletters, voting, and town hall meetings.

In combination, a digital twin and game elements can greatly support the public participation process with children, but as stated above, both are interrelated and mutually dependent. It is a combination of digital twin and game elements that needs to be designed for a particular situation and level of participation. The

two cases are particularly exemplary in how different situations and levels of participation may drive the design of the two parts.

There are many smaller and bigger differences in how the two experiments have piloted the use of Ecocraft in children's participation in energy transition. The two key distinguishing factors, however, are the underlying difference in aims and the balance between game elements versus digital twin, as visualised in Table 10.4.

Table 10.4. Warsaw (green) vs Amsterdam (blue)

Participation level	Possible game elements	Possible features digital twin
Awareness	Rewards and punishment	Visualisation of the issue
Engagement	Competition	Realism/familiarity
Consultation	Contribution	Interaction
Partnership, Power & Control	Community	Reporting, dashboard
Manipulation, etc	*Rules*	*Impression management*

Both cases pay ample attention to preparing the participants in providing information and a priori knowledge building. There are brochures, lesson plans and, in the case of Warsaw, home assignments and field trips. Both cases try to ensure sufficient awareness of the participants. However, in the next stages of the experiments, the two diverge.

The Warsaw case strongly focuses on achieving *engagement*, challenging volunteers – outside school hours – to imagine a 'dream neighbourhood', that is not only 'green' in the use of sustainable energy, but also in its living conditions (greenery). There is a strong competition element, with an extensive judging process and audience vote as well as monetary prizes; and a large celebratory, concluding event. Judging by the very professional video presentations of the teams online, the Ecocraft digital twin with its familiar Minecraft look has certainly helped the children to present their ideas, but true participatory consultation does not seem the ultimate aim.

The Amsterdam case has no competition element. There is no judging beyond teachers grading assignments and there was no prize to be won. The project did not lead up to a public event with presentations and school children were required to participate, as this was part of a regular lesson plan. Here, also in absence of field trips or broader aims of improving the neighbourhood, the digital twin was a key factor in the concrete assignment to try out various forms of insulation and solar panel in the virtual environment with quantitative feedback on the quality of the choices; with a report as deliverable instead of a video impression. The Amsterdam case strongly focused on *consultation*, and hence was also evaluated by the organisers in terms of the quantitative quality of the advice (e.g., payback period of proposed solar panel/insulation solutions). The different reports were not evaluated against each other with a 'best report' presented to external parties, but the focus was on aggregating the quantitative input for a next decision making level. The children enjoyed the use of the Minecraft environment, though some found it somewhat monotonous.

Several reasons have been mentioned to choose for a broader, more inclusive decision-making process (e.g., Gluck et al., 2013; Reed et al., 2018); including *normative motivations* on democratic rights of individuals and communities; expectations on *improved decision-making* when more are heard, local knowledge is leveraged, or more innovative solutions are developed; the *creation of trust* in policies and other stakeholders as a result of the cooperation, but also as more knowledge is exchanged in the process (learning). While the cases share some of the rationale, the Amsterdam case leaned more towards ideas of inclusive decision-making, whereas the Warsaw case seemed more focused on normative motivations of activating children to make sure they are heard.

Note that this does not imply anything about the role of public participation for either city administration in general, but only these two specific experiments on the use of Ecocraft in involving children in the energy transition. Also, in spite of the 'ladder with rungs' metaphor often used to describe the different levels of participation, higher is not necessarily better and, hence, neither case, Warsaw nor Amsterdam, is per se 'better'. They are different, and these differences point to the design parameters relevant in considering the use of a digital twin with gaming elements.

On a more abstract level, the two experiments seem to have a different choice in type of attitude formation among children. A 'standard learning hierarchy', such as in the Amsterdam case, is based on cognitive information processing. In this learning hierarchy, attitude starts with learning facts and information about energy and a resulting effect on what is 'right', eventually affecting behaviour. One might view the Warsaw strategy of involving children as focusing more on an 'experiential attitude formation', where attitude formation starts with an emotional response (affect rather than cognition), and behaviour, later followed by (cognitive) learning about the topic, if and when relevant. Both learning hierarchies have their place in consumer behaviour, dependent on the type of object (product) and situation. Should involvement in energy transition and participatory planning follow a form of standard (cognitive) learning or a form of experiential learning? Facts or feelings? As always, more research is needed to start formulating an answer to this question.

References

Arnstein, S.R. 1969. A ladder of citizen participation. Journal of the American Institute of Planners 35 (4), 216-224.

Brown, G., Kyttä, M. 2014. Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. Applied Geography 46, 122-136.

Buchy, M., Hoverman, S. 2000. Understanding public participation in forest planning: a review. Forest Policy and Economics 1 (1), 15-25.

Drazkiewicz, A., Challies, E., Newig, J., 2015. Public participation and local environmental planning: Testing factors influencing decision quality and implementation in four case studies from Germany. Land Use Policy 46, 211-222.

Eiter, S., Lange Vik, M., 2015. Public participation in landscape planning: Effective methods for implementing the European Landscape Convention in Norway. Land Use Policy 44, 44-53.

Hettinga, S., Nijkamp, P., Scholten, H.J. 2018. A multi-stakeholder decision support system for local neighbourhood energy planning. Energy Policy 116 (May), 277-288.

IEA, 2017. World Energy Outlook. Paris: International Energy Agency.

Levine, P. 2015.The Future of Democracy: Developing the Next Generation of American Citizens (Civil Society: Historical and Contemporary Perspectives). University Press of New England, Lebanon, NH.

Reed, M.S., Vella, S., Challies, E., De Vente, J., Frewer, L., Hohenwallner-Ries, D., Huber, T., Neumann, R.K., Oughton, E.A., Sidoli del Ceno, J., van Delden, H. 2018. A theory of participation: what makes stakeholder and public engagement in environmental management work? Restoration Ecology 26 (1), 7–17.

Scholten, H., Fruijtier, S., Dias, E., Hettinga, S., Opmeer, M., van Leeuwen, W. S., Linde, M., Bos, S., Vaughan, R., van Kaam, H., van Manen, N., Fruijtier, C., 2017. Geocraft as a means to support the development of smart cities, getting the people of the place involved - youth included. Quality Innovation Prosperity 21 (1), 119-150.

Sieber, R. 2006. Public participation geographic information systems: a literature review and framework. Annals of the Association of American Geographers 96 (3), 491-507.

Steinitz, C. 2012. A framework for geodesign: changing geography by design. ESRI Press, Redlands, CA.

United Nations, 2015. Paris Agreement. United Nations, Paris.

11 Conclusions

Craglia M., Scholten H., Micheli M., Hradec J., Calzada, I., Luitjens, S., Ponti M., Boter J.

DIgitranScope set out to explore the challenges and opportunities that the digital transformation is posing to the governance of society. We focused our attention in Part A on the governance of data as a key aspect to understand and shape the governance of society. Data is a key resource in the digital economy, and control over the way it is generated, collected, aggregated, and value is extracted and distributed in society is crucial. We have explored the increasing awareness about the strategic importance of data and emerging models to distribute the value generated more equitably in society. These findings have contributed to the new policy orientation in Europe on technological and data sovereignty and social inclusion.

The digital transformation, and the rise of artificial intelligence and the Internet of Things, offer also new opportunities as shown in Part B for new forms of policy design, implementation, and assessment providing more personalised support to those who need it and being more participative throughout the policy cycle. The use of digital twins, gaming, simulation, and synthetic data are just at their beginning but promise to change radically the relationships among all the stakeholders in governance of our society.



As indicated in the Introduction in Chapter 1, we did not realise at the start of the project is that what we thought was going to be an exploratory project looking 5-10 years ahead would turn instead into one providing already direct input to policy as policy priorities shifted much faster than we anticipated.

On Data Governance: The most significant event that occurred during the lifetime of the project was the emergence of Artificial Intelligence (AI) as a key geopolitical battleground, particularly between the US and China. This brought AI also at the forefront of the European political attention and with it came an increasing recognition of the strategic importance of data as the key asset necessary to develop AI applications. Technological and data sovereignty became key objectives of the new Commission which took office in November 2019.

The new focus on AI and data brought several new policy initiatives reviewed in Chapter 1 including a European Strategy for AI, a plan of investments in AI coordinated with the EU Member States, a European Strategy for Data, a Data Governance Act, and in 2021 a risk-based regulatory framework for AI.

Digitranscope contributed to raising the attention on the strategic importance of data governance through the JRC Flagship report on “Artificial Intelligence: A European Perspective” published in December 2018, which highlighted also the need for proactive policies at the local level to develop local data ecosystems, applications, and training addressing local issues to foster social inclusion. These recommendations were adopted by the Coordinated Plan for AI published shortly afterwards.

Digitranscope contributed directly also to the proposed regulation on European data governance by sharing the research results reported in Chapters 3-6 with the colleagues in the Commission in charge of these data policy initiatives and thus informing about emerging data governance models including data altruism, platform and data cooperatives, and agreements to share business data with local governments

which are supporting the case for the creation of new common data space for public administrations as part of the European strategy for data.

On new forms of policy design: significant policy shifts have also emerged in this area during the last few years and become increasingly mainstream. Notably, the increasing use of big data analytics to profile and nudge voters following the example of the commercial sector recognised not only the power of data but also the emotional side of decision making. The mantra of evidence-based decision-making that was all the rage in the 1990s has come under increasing scrutiny together with the scientific method when applied to social and political phenomena. We have seen therefore a greater acknowledgement of the multi-faceted dimensions of rationality, decision-making, and post-normal science. Communication, participation, and the use of narratives have gained currency exploiting also the new opportunities of the digital transition, from the booming of citizen-generated content for science and policy to the development of digital twins for policy simulation, co-creation, and communication. Within this changing landscape, Digitranscope has contributed in three main ways:

SeTa, the semantic text analysis tool developed in Digitranscope described in Chapter 7 started as a pilot project but has now been institutionalised in the JRC by the Unit responsible for text and data mining and has been made available throughout the European Commission to support the work of all colleagues required to do an impact assessment, ex-ante or ex-post of European policies. Its enormous value is to have turned hundreds of thousands of separate documents into a coherent and usable repository of the European Commission's knowledge.

The Probabilistic Synthetic Population modelling described in Chapter 8 contributed directly to the work of the JRC Corona Virus Task Force which advises the Commission on potential policies and strategies to address the pandemic and its epidemiological and socio-economic effects. A new project has now started with the Dutch Central Bureaus for Statistics to develop the model further, validate it against the statistical data held by the Bureau and provide advice to the Dutch government on covid19-related policies. The ambition is then to extend this collaboration between the JRC and statistical agencies further, involving also EUROSTAT, to create a Europe-wide synthetic population base for policy simulation and analysis.

Digital Twins: the project was able to experiment with the digital twins of Amsterdam and Duisburg as desisted Chapter 9 and gain a deeper understanding through practice and direct interactions with local administrators and industry the man facets of the so-called smart cities. We were also able to leverage the digital twin of the Netherlands and the Ecocraft plug-in developed by the Dutch EduGIS Foundation to raise the awareness of young adults on the trade-offs needed in the energy transition in two schools in Amsterdam and Warsaw as described Chapter 10. Moreover, Digitranscope was able to contribute to a big event in the stadium of the Ajax football team in Amsterdam where 500 kids used the digital twin of their city to design a new sustainable neighbourhood. In that occasion, UN Environment Program (UNEP) signed a partnership agreement with the Dutch EduGIS Foundation. Under the agreement, geospatial data tools will allow the game to map territories around the globe and simulate environmental challenges related to achieving the Sustainable Development Goals.

"Of course, it is best for children to be outside playing in nature. But if they do sit behind the computer, this game can tap their creativity to design future living spaces in harmony with the environment. The next generation of urban planners will see how much our lives depend on this balance," said UNEP's Europe Director Jan Dusik at the 'Liveable Smart Cities by Design' event held in the Dutch capital.

Environmental education is vital to raising awareness on and achieving the Sustainable Development Goals. A target for citizens to participate more in sustainable urban settlement planning is included under Goal 11 on 'sustainable cities and communities,' while Goal 7 aims for 'affordable and clean energy'. With these experiments and the follow-up promoted by UNEP worldwide, Digitranscope was able to show the value of digital twins and gaming as key assets for policy co-creation and testing, and for engaging the new generations of citizens in the decision-making of today that will affect above all their futures.

As the Digitranscope project comes to its conclusion we are conscious that there is still much work to do. We are only at the beginning of the digital transition of society, and at the early stages of equipping ourselves with the necessary theoretical frameworks, regulatory instruments, and networks of partnerships and international alliances necessary to try and shape our futures effectively. As Steven Luitjens reminded us in Chapter 2 governments needs to step up their actions to help guide the process,

build capacity inside public administrations and society through education and investments in research and innovation, develop greater capacity for foresight studies to try and anticipate change and foster a culture of experimentation without fear of making mistakes. The necessary conditions for this are however openness, transparency and inclusiveness. These important principles are not a given in the current digital transformation. On the contrary, we witness increasing polarization in society and the political discourse, and growing inequality between rich and poor, and among different regions, nations, and continents. The covid-19 pandemic has illustrated these dangers well with the effects of both the health crisis and the increasing transitions towards digital platforms and services hitting the most vulnerable groups (the elderly, children, migrants, minorities) worst. How we can channel the digital transformation so that it helps reduce inequality and injustice rather than increase them remains a key challenge. We were only able to address these issues partially in Digitranscope looking for example at emerging models to redistribute more equitably the added value of data, or ways to engage citizens and children in particular in taking the advantage of digital tools to shape their future. There is much more work to do, but we are fortunate that we will be able to continue the work started in Digitranscope at the European level thorough the work of the JRC, at the national level through the new projects we have started with the Dutch geographic council and the statistical agency, and locally through the network of wonderful and committed colleagues we have developed during the three years of the project.

If there is a take-away message from the Digitranscope journey we have described in this volume is that the governance of our digitally-transforming society is challenging and complex, full of opportunities and pitfalls, but that ultimately it is up to all of us to shape it, we cannot afford to leave it to others.

List of boxes

Box 1. Selection of Citizen-Generated Data projects in a nutshell 43

List of figures

Figure 1.1: Evolutionary landscape in data analytics 9
Figure 3.1 - Vignette of the data governance models examined in the chapter. 26
Figure 5.1. Types of collected data 44
Fig 7.2 Conceptual model of the synthetic population for France. 63
Fig 7.3: Influx-outflux of French commuters 2016 64
Figure 8.1 - On-the-fly generated semantic map surrounding the term 'audits' 68
Figure 8.2 Evolution of the term 'auditors' over decades, reflecting contextual change in policy documents and technical reports 68
Figure 8.3 – Features common to impact assessments and audit 70
Figure 9.1. Intelligent Dashboard Amsterdam: Mobility view 74
Figure 9.2 Smart City Platform Duisburg 77
Figure 10.1. Context, conditions, game and digital twin as hierarchical set of elements 84

List of tables

Table 1.1. Industries and applications of Digital Twins 9
Table 3.1: Analytical dimensions 24
Table 4.1. Definitions: Platform Co-operatives and Data Co-operatives 35
Table 4.2. Taxonomy for Platform Co-operatives and Data Co-operatives 36
Table 4.3. Case Identification by Typology 37
Table 5.1. Purpose of data collection in the selected projects 43
Table 5.2. Reported uses of CGD by public organisations 44
Table 6.1 – List of cities involved in the study 52
Table 6.2. Summary of the operational modes for accessing private data contextualized in the discourses and experiences of twelve local administrations. 56
Figure 7.1: Access to education and health for single-parent families and elderly people 61
Table 10.1: Overview of the spread in different indicators between the groups participating in the public participatory planning session. 83
Table10.2: Impact analysis of the overlaid plan of the children 84
Table 10.3. Possible contribution of game elements vs digital twin at different levels of public participation 85
Table 10.4. Warsaw (green) vs Amsterdam (blue) 86

Annexes

About the Authors

Massimo Craglia is a Senior Expert at the European Commission Joint Research Centre, Digital Economy Unit, responsible for projects addressing Digital Transformation and the socio-economic impacts of Artificial Intelligence in different economic sectors, new forms of governance in digitally-transformed societies, and the evolution of the space data economy and the geospatial sector. He is the coordinator of the Digitranscope project.

Henk Scholten is professor in Spatial Informatics at the School of Business and Economics of the Vrije Universiteit Amsterdam. He is also founder and CEO of Geodan Holding, one of the largest European companies specialised in Geo-IT. He is the lead scientist of the Digitranscope project at the Centre for Advanced Studies of the European Commission's Joint Research Centre. Further, he is programme director of UNIGIS Amsterdam, distance-learning MSc. in GIS.

Marina Micheli is Scientific Project Officer at the European Commission's Joint Research Centre (JRC) Centre for Advanced Studies since 2018. Previously she was a senior researcher and teaching associate at the Institute of Communication and Media Research of the University of Zurich and a post-doctoral fellow at the Department of Sociology and Social Research of University Milano-Bicocca, where she earned her PhD in 2013. Her research revolves around technology innovation, adoption and use examined with a socio-cultural and critical perspective.

Jiri Hradec is Scientific Project Officer at the European Commission's Joint Research Centre (JRC) Centre for Advanced Studies exploring personalised policies through data mining, machine learning and econometric modelling. He specialises in the use of text mining, machine learning, and deep neural networks in support to policy across a range of projects and departments at the JRC. Between 2005 and 2012 he was Director of CENIA, the Czech Environmental Information Agency, and between 2000 and 2004 Head of the Informatics Department of the Czech Ministry of the Environment.

Igor Calzada, is Scientific Project Officer at the European Commission's Joint Research Centre (JRC), Centre for Advanced Studies. Since 2012 he is a senior researcher at the University of Oxford, Urban Transformations ESRC and Future of Cities Programmes at COMPAS, leading several research projects on smart cities and city-regions: (i) EU-H2020-Smart Cities and Communities-SCC-01-2015-691735-Replicate (2016<), (ii) ESRC-Urban Transformations (2016-2019), and (iii) City-Regions project funded by Ikerbasque (2012-2014) and the RSA (Regional Studies Association) (2014-2015). His main research interest draws on how digital transformation processes driven by AI disruption in the post-GDPR current context are altering techno-political and democratic conditions of data governance for the emergence of new algorithmic citizenship regimes in European (smart) cities and regions.

Steven Luitjens is program director within the Dutch government in a variety of digital transformation projects. Formerly he was head of the department of information society and government within the Ministry of Interior Affairs and Kingdom Relations, and of an agency responsible for a range of generic solutions for public service delivery. Within the DigiTranScope project, Steven especially helps to bridge the gap between politicians, policy makers and scientists.

Marisa Ponti is Assistant Professor at the Department of Applied IT, University of Gothenburg, Sweden and was a scientific officer at the JRC in 2018 contributing to the activities of the project with respect to data governance and citizen-generated content.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

XX-NA-XXXXX-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:XX.XXXX/XXXXXX

ISBN XXX-XX-XX-XXXXX-X